Causal risk models of air transport

Comparison of user needs and model capabilities

Alfred Roelen

Causal risk models of air transport

Comparison of user needs and model capabilities

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. dr. ir. J.T. Fokkema,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op maandag 10 november 2008 om 15:00 uur

door

Alfred Lambertus Cornelis ROELEN

ingenieur in de Luchtvaart en Ruimtevaart
geboren te Vught

Dit proefschrift is goedgekeurd door de promotor:
Prof. dr. A.R. Hale

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. A.R. Hale	Technische Universiteit Delft, promotor
Prof. dr. B.J.M. Ale	Technische Universiteit Delft
Prof. dr. ir. M.J.L. van Tooren	Technische Universiteit Delft
Prof. dr. A. Mosleh	University of Maryland
Prof. dr. ir. A.C. Brombacher	Technische Universiteit Eindhoven
Dr. H.A.P. Blom	Nationaal Lucht- en Ruimtevaartlaboratorium

ISBN

Keywords: Safety, risk modelling, aviation, safety management.

Published and distributed by IOS Press under the imprint Delft University Press

Publisher & Distributor
IOS Press
Nieuwe Hemweg 6b
1013 BG Amsterdam
Netherlands
fax:+31-20-687 0019
email: info@iospress.nl

Distributor in the USA and Canada
IOS Press, Inc.
4502 Rachael Manor Drive
Fairfax, VA 22032
USA
fax: +1-703-323 3668
e-mail: sales@iospress.com

PRINTED IN THE NETHERLANDS

“People in those days fancied, as people generally fancy when they catch sight for the first time of a new problem, that it was far easier and simpler than was actually the case; they did not know till experience taught them how painfully they would be compelled to advance from step to step, and to unravel the intricate chain of causes which have gone to bring the earth into its present shape; and still less how one principal result of the enquiry would prove that the most interesting questions lay outside the reach of human knowledge”.

Leslie Stephen, The playground of Europe, Fredonia Books, Amsterdam, The Netherlands, 2004, reprint from the 1910 edition.

Acknowledgements

A PhD study has been compared with running a marathon, where every ten kilometres symbolise one year of research [Roelen 1997]. Although this is an interesting comparison, I believe that a PhD study is better compared with climbing a mountain; even though the route in general might be known, you'll have to find your way while you go along and some bits are easy going, but other sections require all your strength and technical abilities. You'll often think you have the peak in sight, only to discover the mountain continuing beyond the ridge you believed to be the top. Perhaps more importantly, running a marathon is a solitary effort, but climbing a mountain is a group endeavour. The rope group is vitally important for the probability of success. The connecting and life-saving rope allows the weaker to find support from the stronger member of the team [Harrer 1988]. A commonality between running, climbing and a PhD study is of course the inverse relation between 'distance remaining' and 'desire to finish'. But then, even a bit unexpectedly, you'll find yourself near the summit. 'A few more whacks of the ice-axe in the firm snow and we stood on top' [Hunt 1953]. It is here, on the top of the mountain, that one shakes hands with all members of the rope-group, acknowledging their contribution to the success. First of all many thanks to Andrew Hale for persuading me to start this endeavour for being such a thoughtful and kind promotor. John Lapointe and Kathy Fazen of the FAA Tech Center in Atlantic City and Hok Goei of the Transport and Water Management Inspectorate provided the first opportunity for me to do some serious research on risk modelling for air transport. Ali Mosleh kindly shared his knowledge from the field of nuclear safety engineering and his wisdom in many other disciplines, and this helped me a lot in getting a firm grip on the subject. The CATS people were indispensable for further development of resulting ideas and insights, so thank you Ben Ale, Roger Cooke, Dorota Kurowicka, Pei-Hui Lin, Oswaldo Morales-Napoles, Linda Bellamy, Louis Goossens, Dan Ababei, John Spouge, John Cooper and Rob van der Boom. While being involved in the various projects that were directly or indirectly related to the research topic, I have enjoyed working together with my colleagues at NLR, in particular Peter van der Geest, Gerard van Es, Rombout Wever, Gerben van Baren, Bart Klein Obbink, Hans de Jong, Bas van Doorn, Jelmer Scholte, Hans Post, Jeroen van der Zee, Mariken Everdij, Sybert Stroeve, Lennaert Speijker, Juan Coelho, Harry Smit, Arjen Balk, Joram Verstraeten, Johan Weijts, Ton Nieuwpoort, Arun Karwal, Carolynne Montijn, Margriet Klompstra, Bert Bakker, Koen de Jong, Udo Dees, Tom van Birgelen, Patricia Sijaranamual and Anna Kurlanc. Marijn Giesberts and Job Smeltink deserve special recognition for being such cheerful roommates! Henk Blom conducted a much appreciated review of an early draft of this thesis and Michel Piers and Alex Rutten encouraged me start and also to complete the work.

While the rope group is fighting its way up the mountain slope, the people in base camp actually make it all possible. They provide the necessary supplies and allow the climber to recover from his efforts. I would like to thank the people of my base camp; Rob & Ellen and Jim & Henriëtte for their friendship and Caroline Veugelers for always being interested in the topic of my research. Dear mom, thank you for being the best mother in the world. Bernard and Susana are my base camp heroes. Obviously, the most important member of the base camp is my partner Mijntje Pikaar. I am very grateful for sharing my life with you.

Harrer, H. (1988). Das Buch vom Eiger, Pinguin Verlag, Innsbruck, Austria.
Roelen, B. (1997). TGF-βs and their receptors in early mammalian development, Febodruk B.V., Enschede.
Hunt, J. (1953). The ascent of Everest, Hodder & Stoughton, London.

Table of contents

List of abbreviations

AC	Advisory Circular
ACC	Area Control Centre
AD	Airworthiness Directive
ADREP	Accident/Incident Reporting System
ALARA	As Low As Reasonably Achievable
ALARP	As Low As Reasonably Practicable
ANS	Air Navigation System
ANSP	Air Navigation Service Provider
AOC	Air Operator Certificate
ATC	Air Traffic Control
ATCo	Air Traffic Controller
ATHEANA	A Technique for Human Event Analysis
ATIS	Automatic Terminal Information System
ATL	Aircraft Technical Log
ATM	Air Traffic Management
BBN	Bayesian Belief Net
BFU	Büro für Flugunfalluntersuchungen
CAA	Civil Aviation Authority
CATS	Causal Model for Air Transport System
CFIT	Controlled Flight Into Terrain
CIL	Critical Item List
CREAM	Cognitive Reliability and Analysis Method
CRM	Crew Resource Management
CS	Certification Specification
CTR	Control Zone
CVR	Cockpit Voice Recorder
DME	Distance Measuring Equipment
EASA	European Aviation Safety Agency
EC	European Commission
ECCAIRS	European Co-ordination Centre for Aviation Incident Reporting Systems
EMF	Electric and Magnetic Field
EPC	Error Producing Condition
ESARR	Eurocontrol Safety Regulatory Requirement
ESD	Event Sequence Diagram
EU	European Union
FAA	Federal Aviation Administration
FANOMOS	Flight Track and Aircraft Noise Monitoring System
FAR	Federal Aviation Regulation
FAS	Final Approach Speed
FDR	Flight Data Recorder
FHA	Functional Hazard Assessment
FMECA	Failure Modes Effects and Criticality Analysis
FMS	Flight Management System
FOCA	Federal Office for Civil Aviation
FSF	Flight Safety Foundation
GGR	Gesommeerd Gewogen Risico (summed weighted risk)
GPS	Global Positioning System

GPWS	Ground Proximity Warning System
GR	Group Risk
GTS	Gezamenlijke Tandienst Schiphol (joint fuel service Schiphol)
HEART	Human Error Assessment and Reduction Technique
HEP	Human Error Probability
HIL	Hold Item List
HP	Horse Power
HSE	Health and Safety Executive
IATA	International Air Transport Association
ICAO	International Civil Aviation Organisation
IFR	Instrument Flight Rules
ILS	Instrument Landing System
IOSA	IATA Operational Safety Audit
IR	Individual Risk
JAA	Joint Aviation Authorities
LOSA	Line Operations Safety Audit
MDA	Minimum Descent Altitude
MEL	Minimum Equipment List
MER	Milieu Effect Rapportage (Environmental Impact Assessment)
MRB	Maintenance Review Board
MSAW	Minimum Safe Altitude System
MTOW	Maximum Take-Off Weight
NASA	National Aeronautics and Space Administration
NLL	Nationaal Luchtvaart Laboratorium (National Aviation Laboratory)
NLR	Nationaal Lucht- en Ruimtevaartlaboratorium (National Aerospace Laboratory)
NOTAM	Notice To Airmen
NRC	National Regulatory Commission
NTSB	National Transportation Safety Board
PDP	Piecewise Deterministic Markov Process
PF	Pilot Flying
PNF	Pilot Not Flying
PSA	Probabilistic Safety Assessment
PSF	Performance Shaping Factor
PSSA	Preliminary System Safety Assessment
PSZ	Public Safety Zone
QAR	Quick Access Recorder
QRA	Quantitative Risk Assessment
RSL	Rijks Studiedienst voor de Luchtvaart (State research office for aviation)
RVSM	Reduced Vertical Separation Minima
SADT	Structured Analysis and Design Technique
SAFA	Safety of Foreign Airlines
SAM	Safety Assessment Methodology
SARPs	Standards and Recommended Practices
SASO	Systems Approach to Safety Oversight
SB	Service Bulletin
SES	Single European Sky
SID	Standard Instrument Departure
SMS	Safety Management System
SOP	Standard Operating Procedure
SSS	Stanford Sleepiness Scale

STAR	Standard Arrival Route
STCA	Short Term Conflict Alert System
TCAS	Traffic alert and Collision Avoidance System
TLS	Target Level of Safety
TMA	Terminal Manoeuvring Area
TOPAZ	Traffic Organization and Perturbation AnalyZer
TRG	Totaal Risico Gewicht (total risk weight)
UK	United Kingdom
US	United States
USAF	United States Air Force
UAV	Uninhabited Areal Vehicle
V_1	Take-off decision speed
VACS	Veiligheids Advies Commissie Schiphol (Safety Advisory Committee Schiphol)
VEM	Veiligheid Efficiency Milieu (Safety Efficiency Environment)
VEMER	VEM Effect Report
VFR	Visual Flight Rules
VHF	Very High Frequency
VOR	VHF Omni-Directional Range
V/STOL	Vertical / Short Take-Off and Landing

Chapter 1. Introduction

Affordable and reliable aviation plays a vital role in supporting economic growth and expanding personal options for where individuals can live, work, travel, and conduct business. As the world becomes increasingly interdependent, aviation services will take on even greater importance [JPDO 2004]. In the Netherlands, Amsterdam Airport Schiphol plays an important role as a transport node and in the regional economy [Hakfoort et al 2001]. To maintain its position as main transporter in Europe, the Netherlands must continuously develop its transportation infrastructure. The airport, like many other large international airports, is located in a densely populated area. Further growth of the airport is an economic requirement, but this should not lead to increased burdens in terms of noise, pollution and accident risk.

A main difference between accident risk and other detrimental aspects of aviation, such as noise and pollution, is that risk cannot be directly measured. In particular because the probability of accidents is low with respect to the period that is available for observations, statistical analysis alone is an insufficient basis for risk control. Aviation safety is so well developed that the current accident rate for large commercial aircraft in Europe is approximately one accident in five million flights. Because of this low accident rate, individual organisations such as airlines or airports cannot rely on the number of accidents as a useful indicator of the safety level of their operation. How can an airline that operates 200.000 flights each year and that most likely has not experienced an accident in many years know if their current safety level is better or worse than last year's? Adequate control of risk requires the availability of a method to determine the level of accident risk as a function of changes to the aviation system. This method must be methodologically sound. Determining the level of risk as a function of the status of the aviation system requires insight into cause-effect relations pertaining to accident risk.

The Dutch Ministry of Transport, Public Works and Water Management has recognised this and the Ministry's aviation safety policy plans have proposed the development of system-wide[1] causal risk models. Unfortunately however, they failed to specify or even describe such models and their requirements other than in the most general of terms. Other organisations such as the FAA, NASA and Eurocontrol, have also called for the development of new methods for aviation safety analysis and assessment, albeit without using the term 'causal risk model' [FAA 2004, Luxhøj 2003, Eurocontrol 2004]. Several feasibility studies have been conducted and prototype causal risk models for aviation have been developed in a direct response to the Ministry's policy plans [Roelen et al 2000a, Roelen et al 2002, DNV 2002b, Ale et al 2006, 2007, 2008].

The aviation system is a prime example of a complex multi-actor system. System-wide causal models are currently not used for controlling and managing aircraft accident risk, although more limited accident risk models are, for some specific problems, already well-accepted[2]. The complexity of the subject and the problems that arise in validating the

[1] i.e. describing the complete air transport system.
[2] An example is the ICAO collision risk model, see section 4.3.6.

results of such models are possible reasons for this. Research has primarily been focussed on the technical feasibility of models, without close consideration of the methodological consistency in relation to user requirements. One of the biggest bottlenecks in the past has been the fact that causal model development was stated as a goal in itself, without considering how such a model should be used [VACS 2003b]. From interviews held with expert groups regarding user requirements for a causal risk model for air transport, it was concluded that "It does not appear clear what the goal behind the causal model is, and this is hampering its development as different goals and scopes ask for different choices in developing the model" [De Jong 2006]. A complicating factor is that the Ministry and the Dutch Parliament completely misunderstood what a causal model could do; a causal model was considered to provide a solution for the problem of setting a maximum allowable value for third party risk around airports when this was not a feasible objective of such a model [VACS 2003b] (see Appendix A for details).

The objective of this thesis is to clarify these issues. The approach taken is not to develop yet another prototype model. It is also not the aim to provide a critical analysis and comparison of existing risk models. Existing methodologies such as CATS and TOPAZ are only used to illustrate some of the issues and are not put forward as the primary object of the thesis.

Instead, this thesis will systematically identify and compare user requirements with the performance that can be delivered by various existing modelling techniques. In doing so, the thesis will show what a causal risk model can add to current management approaches and what the criteria are for ensuring it makes a contribution. For practical reasons the thesis will be limited to existing and well-known modelling techniques, lesser known techniques and advanced approaches that are under development are considered to be out of scope. For the purpose of this thesis, a causal risk model will be defined as *a mathematical object that provides an interpretation and computation of causal queries about accident risk. A causal model can be associated with a directed graph*[3] *in which each node corresponds to a variable of the model and the arrows point from the variables that have direct influence to each of the other variables that they influence.*

Economic impact of Schiphol
In 2006 Schiphol Airport facilitated 440.153 aircraft movements which resulted in 40 million passenger movements and the shipment of 1.5 million tonnes of cargo, ranking it the 4th largest European Airport in terms of passenger movements. Schiphol Airport's aviation business had a revenue of 631 million €, the consumer business (airport shopping, car parks, etc) had a revenue of 231 million €, and the real estate's business totalled 109 million €. There were 61.691 people, including temporary staff, working in the airport region. [Schipol Group 2007]. The multiplier of direct employment on Amsterdam Schiphol Airport is approximately 2: one job on the airport leads to approximately one job in indirect and induced employment in the greater Amsterdam region [Hakfoort et al 2001].

[3] Examples of directed graphs are fault trees, event trees, Petri nets, Bayesian belief nets.

1.1. Research question

The main research question is the following:

> *What does causal risk modelling add to current safety management approaches, and what are the criteria for ensuring it makes a successful contribution?*

To answer this question we must first make the term causal risk model more specific. This is done by asking the following sub questions:

> *How can risk be made tangible? What is a proper way to express risk?*
>
> *What is a causal relation and which characteristics of causal relations are important for causal risk model development?*
>
> *What is a causal risk model?*

We must then widen the scope a bit and consider the context to search for clues to help answer the main question. This will be done with the following sub questions:

> *What are the needs of users?*
>
> *What are currently the drivers for aviation safety improvement and what could be the role of a causal risk model in the process of safety improvement?*
>
> *What are shortcomings of the current methods for aviation safety analysis?*
>
> *What can be learned from risk modelling in other industries?*

Having looked at the 'why' we will then focus on the mechanisms for risk modelling to analyse how quantified models could support the users:

> *Which modelling techniques are most appropriate?*
>
> *How should a causal risk model be quantified, what numerical accuracy is required and how can that be obtained?*

As the aviation system is a complex multi-actor distributed system we may expect that some characteristics are more difficult to represent in a model than others.

> *What are, from a modelling point of view, the biggest bottlenecks?*

Finally, the use of a model cannot be justified when there is not some sort of proof of the validity. The last question addresses this issue:

> *How can we demonstrate the validity of a causal risk model?*

1.2. Scope

The scope of this study must be described along three dimensions. The first is that of the aviation processes that are the subject of the study. Is it the straightforward gate to gate processes of an aircraft, or are the subsidiary processes like design and certification, maintenance, etc. involved as well? If so, what are the boundaries of the subsidiary

processes? The second dimension is that of the hierarchy of the organisation. Is only the operational level considered, or are higher managerial levels (procedures level, policy) and even those of the regulators also of interest? The third scoping dimension is that of the accident 'processes', including the causal factors. How far back in time must the causes of the event be traced? Are consequences of the accident also within the scope? For each scoping dimension we need to indicate whether the boundary in scoping is the result of practical or theoretical considerations. These scoping questions are basically part of the user needs and will be addressed in the associated chapter of this thesis.

Apart from the scoping of the causal modelling that will be determined by the user requirements, the scope of the thesis will have to be limited for practical reasons. The study focuses on commercial air transport. Leisure flights, military aviation and aerial work (e.g. crop dusting) are considered out of scope. In practice this limits the study to fixed wing aircraft as commercial air transport by helicopter and lighter-than-air vehicles is negligible in volume and number of flights compared to fixed wing aircraft. The situation in the Netherlands is taken as example to answer the research questions, with a focus on the safety issues related to the growth of Amsterdam Airport Schiphol. Risk for people on-board as well as on the ground (both inside and outside the airport perimeter) will be taken into consideration. Environmental impact effects are excluded. The actors involved include the airlines, the airport and the air traffic control provider, but also the Ministry of Transport, the Ministry of Spatial Planning and the Environment, local municipalities and commissions and advisory groups such as the Dutch Expertgroup on Aviation Safety (DEGAS) and its predecessor the Safety Advisory Committee Schiphol (VACS) and the Commissie MER. Because of the general desire to harmonise regulation within Europe, because of the desire for a 'level playing field' in Europe, and because of the international character of aviation in general, the problem should also be addressed in a European context.

For pragmatic reasons the scope of the thesis will be limited to direct aircraft crash risk. Post crash events, such as the development of post-traumatic stress disorders in people directly involved, are not considered. From a technical point of view it is perhaps perfectly feasible to extend a causal risk model to post-crash events, but such a model would require different subject matter expertise and is therefore considered to be outside the scope.

Aircraft crash risks as a result of unlawful acts (terrorism, revengeful employees, unruly passengers), or military intervention (either on purpose or accidental) are considered out of scope because the information on associated causal influences is considered to be confidential and not suitable for general dissemination.

Primary (flying from A to B) as well as subsidiary processes (air traffic control, aircraft design and maintenance, etc) are relevant for flight safety. While each accident always involves the primary process, the causal chain of events for an accident sequence nearly always involves the subsidiary processes as well. Therefore a causal risk model must encompass the primary and the subsidiary processes. The model should encompass those subsidiary processes that are directly linked to the primary process. They include flight crew training, aircraft design, certification and maintenance, air traffic management and airport processes. A description of these processes is provided in Appendix B.

The primary and subsidiary processes are embedded in national and international policies and regulation. Control of the processes' products, including safety, actually involves a socio technical system of several hierarchical levels. Rasmussen [1997] identifies the

following 6 levels: Government, Regulator, Company, Management, Staff and Work. The scope of the causal risk model should encompass the primary and subsidiary processes across all hierarchical levels from government down to work.

1.3. Directions for the reader

This thesis starts by explaining very briefly and superficially, the basics that are required before the research question can be really addressed. A proper discussion on causal modelling of aviation risk requires first a definition of the concepts of risk and safety in Chapter 2 and in Chapter 3 an explanation of 'causality' in itself, including a description of what constitutes a 'causal model'. User needs are introduced in Chapter 4. The focus then narrows to current practice beginning in Chapter 5 with a look at the way in which today safety assessments for air transport are typically conducted. Chapter 6 follows with an overview of other risk bearing industries with particular attention for the way in which risk models are used in those sectors for managing and controlling risk. Comparison of current practice with user needs is decisive in selecting modelling techniques to be used in causal risk models. Different techniques and their characteristics are described in Chapter 7. Quantification is often mentioned as one of the main problems and will be the topic of Chapter 8. Three traditionally difficult subjects in risk modelling are described in Chapter 9: modelling human operators, modelling safety management and dealing with interdependencies between various parts of the model, while the fourth, validation, is dealt with in Chapter 10. All the ingredients are then available to come to a conclusion and to answer the main research question, which is done in Chapter 11. Additional background information is provided in the Appendices: Appendix A gives an overview of the history of third party risk regulation at Schiphol airport. This information is relevant to appreciate the context from which a call for causal risk models was explicitly generated. Appendix B describes the main and subsidiary process in air transport. Appendix C gives information of the CATS project. CATS is used throughout this thesis as an illustration of several issues.

Chapter 2. Fundamentals of risk

Before being able to say something about a causal risk model for air transport, it must be clear what is meant by 'risk' and if and how this risk can be 'measured'. Policies to control risks are often based on some sort of quantification of the risk that is allowed to exist. There are different views on how to achieve a certain 'amount' of safety and these are discussed. Such information is required to determine what and how to model and what output metrics are needed. It will shed some light on what can and cannot be done with a causal risk model. This chapter also describes the most relevant theories on accident causation and contains a brief historical overview of risk modelling. This information is relevant because further development of current practice is more likely to be accepted by projected users of a causal risk model than a radically different approach.

2.1. Definition of safety

The word safety can be used to indicate freedom from harm or freedom from risk, where risk is a combination of the probability of harm and its severity. The latter use of the word safety, which allows a distinction between degrees of safety instead of just a Boolean safe / unsafe distinction, is more suitable for safety management, because to manage safety it is necessary to know whether the situation is getting better or worse in order to be able to take measures before the harm can occur. A formal, generic definition of safety was given by the International Standards Organisation (ISO) and the International Electrotechnical Commission (IEC) [ISO 1999] and is adopted in this thesis:

> *Safety* is freedom from unacceptable risk.
>
> *Risk* is a combination of the probability of occurrence of harm and the severity of the harm.
>
> *Harm* is physical injury or damage to the health of people either directly or indirectly as a result of damage to property or the environment[4].

The harm in this definition is limited to injury and health effects. Other types of harm, e.g. economic or financial, are excluded and consequently other types of risks, such as financial risk, economic risk, etc. are also excluded. Combining the probability of occurrence of harm and the severity of the harm to obtain risk is similar to, but more general than, the often used definition of risk as 'probability times consequences'. The former definition allows risk to have a more complex dependence on probability and severity. Although not common, the square of severity is sometimes used to obtain a measure of risk for situations where people consider high consequence, low probability events to be worse than low consequence, high probability events [Hubert et al 1990, Joyce et al 2001]. In this thesis the scope will be limited to direct effects of aviation accidents on the health of people. Direct effects are deaths, physical injuries and physical damage that are immediately apparent following an accident. Indirect effects of accidents will not be considered. Indirect effects

[4] For the purpose of this study we explicitly exclude environmental impact. We do not want to get into deliberations on, say, global warming and aviation fuel taxation.

are those that are not immediately apparent but manifest themselves some time after the accident. Examples of indirect effects are health problems of rescue workers caused by exposure to toxic substances, or the development of post traumatic stress disorders among people that have witnessed or were involved in aircraft accidents. The manifestation and degree of indirect effects is largely governed by social, psychological and epidemiological factors and requires different knowledge and expertise than estimating direct aircraft crash effects.

2.2. Risk perception

The definition of safety in the previous section introduces *acceptability* of risk. Acceptability of risk is strongly influenced by risk perception. The level of perceived risk has been found in several studies to be dependent on the degree to which people believe that risk can be controlled and by whom, trusted or not [Slovic et al 1976, Hale & Glendon 1987]. Voluntary and involuntary exposure to risk is also a main driver for risk acceptability (Figure 2). Another factor that is related to risk perception and the level of accepted risk is the chance of multiple fatality accidents [O'Banion 1980, ETSC 1997]. A fourth factor that plays a role is the time passed since a similar event took place. The more retrievable the event, the greater its intuited probability. News media's extensive coverage of aircraft accidents make them particularly 'retrievable'. For instance in a 1996 Associated Press survey, U.S. newspaper editors and television news directors said that the Trans World Airlines Flight 800 accident[5] was the 'biggest' news story of 1996 [Barnett & Wang 2000]. A 1990 study of page-one newspaper articles regarding fatalities in the United States said that coverage of air carrier accidents in *The New York Times* was 60 times greater than its coverage of AIDS, 1,500 times greater than coverage of automobile-related hazards and 6,000 times greater than coverage of cancer [Barnett & Wang 2000]. Events that result in 'identifiable' victims have more impact than events resulting in anonymous or 'statistical' victims. An accident, such as an aircraft crash, therefore has more impact on public perception than for instance exposure of large and hence anonymous populations to toxic substances due to general environmental pollution [Health Council of the Netherlands 1999].

As risk implies possible loss, risk perception is also a direct reflection of perceptions of *value* [Brauer 2004]. As such, there are differences in risk perception across cultures and across time. Accidental death is likely to be more acceptable (to society as a whole) in regions where the life expectancy is lower than the world average due to 'natural' causes of death such as infectious diseases and famine. A comparison of aircraft accident rates and gross domestic product across world regions indeed show a strong inverse relation [Roelen et al 2000c, Visser 1997]. Likewise risk acceptability depends on the benefit associated with risk exposure. The greater the reward, the higher the risk we are willing to take.

[5] On 17 July 1996, Trans World Airlines Flight 800 crashed minutes after take of from John F. Kennedy International Airport, New York. The cause of the accident was an explosion of the centre wing fuel tank. The source of ignition energy for the explosion could not be determined with certainty, but, of the sources evaluated by the investigation, the most likely was a short circuit outside of the centre wing tank that allowed excessive voltage to enter it through electrical wiring associated with the fuel quantity indication system [NTSB 2000].

Figure 1: Voluntary exposure to risk is more acceptable than involuntary exposure.

The causal risk model must be restricted to 'objective' risk because of the lack of uniformity in subjective risk between groups, players, times and places. It might be tempting to require the causal risk model to (also) have perceived risk as output parameter but if the model were to include perceived risk it would have to include too many factors that are most likely very difficult to quantify and to represent in the model. The history of noise level calculations for airports provides a clear example of the difficulty of expressing 'perceived levels of burden'. See for example Everdijk [2006]. For the purpose of a causal risk model, the level of risk must be expressed in objectively quantifiable units. This is needed because the results of the model should be utilizable to make comparisons, for instance on the level of aviation risk before and after the introduction of a technological or managerial change.

2.3. Risk metrics

The objective of this section is to decide on an appropriate metric for a causal risk model. There is no single common metric for risk. Even when the scope is restricted to mortality risk, there are various ways to express this, such as number of accidents with fatalities per aircraft hour, fatalities per aircraft hour, fatalities per passenger mile, fatal accidents per passenger mile and fatal accidents per flight. Expressing mortality risk per flight hour or per passenger mile has the advantage that it makes comparison with other modes of transport possible in a realistic way. But because most accidents occur during the take-off and approach and landing phase of flight [CAA-NL 2007] this metric is also somewhat misleading; the risk per passenger mile is lower for a long flight than for a short flight but

the risk per trip is similar for a long flight and a short flight. Using ‘fatal accidents’ as the numerator for mortality risk has the drawback that the term fatal accident includes all accidents that cause at least one death and does not distinguish between an accident that kills one passenger among 400 and an accident that kills all on-board. According to Barnett & Wang [2000], ‘death risk per flight’ is the most appropriate metric of aviation risk. It is based on the following question: If a passenger chooses a (nonstop) flight completely at random, what is the probability that he or she will be killed during the flight?

However, this metric ignores that aircraft accidents are not only hazardous for those on-board, but also for people on the ground. Risk for those on the ground is called third party risk. Third party risk became a major concern as a result of some major disasters in the chemical industry such as the Bhopal disaster in 1984 when more than 3000 people were killed as a result of the accidental release of some 40 tons of methyl-isocyanate [Ale & Piers 2000]. For stationary installations, the term third party risk is used to indicate the risk for those people that are not employees of the concerned company, i.e. those that are ‘outside the gate’. This definition was also used by the Dutch Ministry of Transport when developing policies for aviation third party risk (see Appendix A). Only risks for those people outside the airport boundary are taken into account, ignoring thousands of people that can be present inside the airport boundaries. Also often only those living and working in the area are included, excluding those temporarily there such as motorway travellers. The decisions to take the airport boundary and temporality of the stay as criteria are policy decisions and are not inherent to the definition of third party risk. There are different metrics for third party risk, the most well-known are individual risk and group risk.

Individual risk (IR) is the probability – per year – that an imaginary person that is permanently residing at the same location (in the vicinity of the airport) dies as a direct result of an aircraft accident. Individual risk is a metric for the personal safety of (imaginary) persons that reside permanently in the vicinity of a risk-bearing activity. Individual risk is location dependent and is present regardless of the presence of persons. In general, IR-values become less when the distance to the risk source increases. Individual risk values are often represented as iso-IR contours on a topographical map.

Group risk (GR) is defined as the probability - per year- that a group of more than a particular number of persons (denoted with ‘N’), other than occupants of the aircraft, dies as a result of a single aircraft accident. GR is often represented in a graph that plots the probability of having an accident with N or more fatalities (the so-called FN curve). The notion of group risk is used to indicate ‘calamities with potentially many victims under the population’. Group risk is sometimes popularly referred to as the ‘disaster probability’. Because group risk not only concerns the probability of a calamity, but also its size, group risk to a certain extent is a metric for the probability of societal disruption. Group risk applies to an entire area and is not location specific within that area. However the ‘area’ can be defined as quite small, so that it approximates to a location specific metric. Whereas individual risk concerns imaginary persons (there is no relation with actually present people), group risk does concern the actual population. When there are no residents in the vicinity of the risk-bearing activity, the group risk is zero.

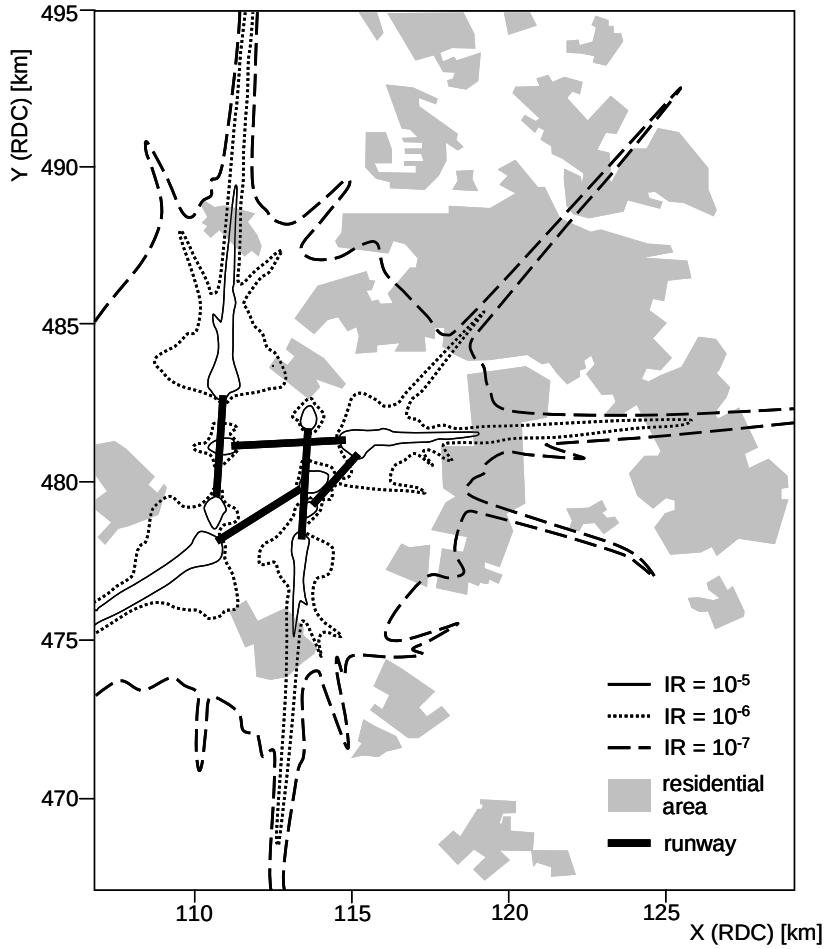


Figure 2: Result of a calculation of individual risk around an airport.

Third party risk is a consequence of aircraft accident risk. Calculating third party risk usually requires combining the aircraft crash probability with the likely location of the accident and the consequences for those on the ground[6]. The aircraft accident probability is hence one of the main inputs for calculating third party aircraft risk, albeit that not every aircraft accident has third party risk potential; an occurrence where an aircraft hits severe turbulence causing some of the passengers to be thrown around in the cabin and sustaining injuries is an accident according to the definition of the International Civil Aviation Organization (ICAO), but it does not contribute to third party risk. In this study, we will focus the attention on a causal risk model with aircraft accident probability as output; the scope is restricted to modelling accident causes, modelling of accident consequences is not considered. Nevertheless, because the aircraft accident probability is an input for third party risk calculations the causal risk model could be used rather straightforwardly in third party risk analysis as well.

[6] Consequences are usually expressed as a combination of crash area and lethality within the crash area.

Fatal accidents are rare events. When using accident data (i.e. the realization of probabilities) to quantify aviation risk, this must be taken into account. Because fatal accidents are so rare even large observed differences need not attain significance. If a coin is tossed only four times, then no possible outcome - even four heads or four tails- would provide statistically significant evidence that the coin is not fair [Czerwinski & Barnett 2004]. The number of aviation accidents with fatalities has become so low that it is problematic to use those as the only indicator of air transport risk, there is too much randomness in the data. Therefore non-fatal accidents and incidents should be included in the considerations on the construction of a causal risk model. But the question then arises which occurrences should be considered. Incidents and (fatal) accidents are not necessarily causally related. Trips and falls for instance result in many injuries but are rarely causally related to catastrophic accidents. Therefore only those occurrences, associated with the operation of an aircraft, which affect or could affect the safety of operation[7] should be included.

2.4. Risk criteria

Unfortunately, safety is not self-sustainable [SAE 2003]. Some sort of safety management is required to improve or even maintain the current level of safety. Policies to control major risks have been in development from the 1960s onwards. Many of these policies are based on some sort of quantification of the risk that could be allowed to continue. This section explains how causal risk models can be used in relation to such quantitative risk criteria and risk control policies.

A Target Level of Safety (TLS) is the 'amount' of safety that is aimed for. The concept of a TLS appears intuitively obvious [Joyce et al 2001]. The prime user of the TLS concept in aviation has been ICAO. Over the years, ICAO has developed TLS concepts in various safety critical areas of the industry, using international groups of experts. Such areas have included the North Atlantic System Planning Group (1992), the All Weather Operations Panel (1994), the Obstacle Clearance Panel (1980) and the Review of the General Concept of Separation Panel (1995). The latter work panel has offered the following definition:
"*A Target of Safety (TLS) specifies an acceptable value of risk which can be used as a yardstick against which the risks associated with a system or procedures can be evaluated. The concept of a TLS is particularly useful when planning changes in safety critical operations such as air traffic control*."
[RGCSP 1995].

The definition for the TLS concept that is offered will vary in accordance with its particular application and its intended use. For example, the UK Civil Aviation Authority (an organisation which has also been instrumental in the development of TLS concepts), defines the TLS concept for controlled airspace as:
"…a fundamental concept in any mathematical / statistical approach to systems planning when questions of safety are involved…The target level of safety is the level of safety which the system is designed to achieve. Put the other way round, the system is designed to an assured level of safety. By this specification it is possible to define planning objectives which fit in with the safety constraints, and also provide a safety yardstick against which potential changes can be assessed and objectives pursued." [Brooker & Ingham 1977].

A Target Level of Safety is a level of safety that either must be achieved in order to carry out some activity (i.e. a mandatory target) or must be aimed for but need not necessarily be

[7] This is ICAO's definition of an 'incident'.

achieved (i.e. an aspirational target). Most TLS used to date in aviation are mandatory targets, set by the safety regulator. They specify a minimum level of safety (or maximum permissible risk) that must be demonstrated before some equipment or system can become operational, or remain in operation. The safety regulator or the industry may also use a TLS with the purpose of allocating risk space to different players and system elements. The main use of aspirational targets is motivational, like the 'zero accidents' action plan of the Federal Aviation Administration (FAA) in 1995 [FAA 1995a]. These are promotional carrots, or sticks to beat those lower in the hierarchy.

The concept of a level of safety has a number of implications. Because safety cannot be measured directly, an alternative approach to quantifying safety is necessary to be able to demonstrate that a certain target has been or will be met. A causal risk model can be used for this purpose. The results from the model should then be reproducible: a person determining a level of safety from the same data on different occasions should obtain the same result. The results should also be objective: different persons running the model should obtain the same result.

Setting a target level of safety and calculating whether a target has been met are two separate steps in a decision making process. Failure to make this distinction may lead to confusion. Setting safety targets is not very useful if the processes that govern safety are not known. It is in this sense that a causal risk model is valuable for it can show the effect of certain decisions on the level of safety. Results of a model can be used to determine that a certain TLS has been met, but can also be helpful if a target has not been met. A causal risk model can then be a tool to determine (effective or efficient) measures to meet safety targets.

Conceptually different from a TLS is the ALARA or ALARP approach; As Low As Reasonably Achievable / Practicable. Here, the main focus is that there should be an improvement and rather less what the end point should be. Whereas a TLS defines the risk space to be utilized and is static, ALARA or ALARP is aimed at continuous improvement and is dynamic. In this approach, a computation must be made in which a quantum of risk is placed on one side and the sacrifice (in money, time, trouble, etc) involved in the measures necessary to avert the risk is placed on the other. If the sacrifice is considered grossly disproportional it is considered unreasonable. Because of the gross disproportionality criterion, it is not required that the cost and quantum of risk are estimated very accurately [Ale 2005]. What is 'reasonably achievable or practicable' is usually established from engineering judgement. A causal risk model can be used in an ALARA/ALARP approach to calculate the quantum of risk. The required accuracy of a model for such application is less than for a TLS application.

A causal risk model has the potential to provide an understanding of the inherent risks of the air transport system over a wide range of conditions. A causal risk model which takes an integrated look at the air transport system, uses realistic criteria for performance and tries to quantify the uncertainties is a basis for 'risk informed' decision making[8]. The increased insight is expected to improve safety beyond the level that can be reached by a 'risk based' approach in which safety requirements are translated in pass/fail rules that are straightforward to implement and to verify compliance [NRC 1998]. It is therefore clear

[8] Risk informed decision making represents a philosophy whereby risk insights are considered together with other factors to establish decisions on design and operational issues commensurate with their importance to safety.

that, while a causal risk model can be used in a TLS approach, the model can fully come to justice in an ALARP approach which strives for continuous improvement.

2.5. Theories about accident causation

Risk from air transport is almost exclusively related to accidents[9]. In a desire to understand and prevent the occurrence of accidents, many researchers have tried to develop a concept or framework for describing accident causation. Some of these theories have been very influential on the way we think about accidents and the representation of accidents in mathematical models. For that reason this section provides a brief overview of some of the most important accident causation theories and concludes with what this means for risk model requirements.

For the purpose of accident investigation, it is often assumed that accidents involve the occurrence of a set of successive events that produce unintentional harm. The start of this sequence is a deviation or perturbation that disturbs the existing equilibrium or state of homeostasis. Benner [1975] pointed out that events during accidents occur both in series and in parallel and proposed flow charting methods to represent this. Accident investigation tools like Event and Causal Factors Analysis (ECFA) [Buys & Clark 1995] and Management Oversight and Risk Tree (MORT) [Johnson 1975] use this concept to provide a structure for integrating and subsequently communicating investigation findings. Already in 1935, Heinrich developed a theory that introduces an additional dimension to such accident chain model. He compared the occurrence of an accident to a set of lined-up dominoes [Heinrich et al 1980]. The five dominoes were a) ancestry and social environment, b) worker fault, c) unsafe act or unsafe condition, d) accident and e) damage or injury. Central to Heinrich's original statement of the model is the assertion that the immediate causes of accidents are of two different types; unsafe acts and unsafe conditions. Heinrich's domino model was also useful to explain how by removing one of the intermediate dominoes, the remaining ones would not fall and the injury would not occur. Reason [1990] took Heinrich's unsafe acts and unsafe conditions a step further by refining the distinction between different types of failures that line up to create an accident. Building upon work by Rasmussen [1983], Reason describes an accident as a situation in which latent failures, arising mainly in the managerial and organisational spheres, combine adversely with local trigger events (weather, location etc) and with active failures of individuals at the operational level. Latent failures are failures that are created a long time before the accident, but lie dormant until an active failure triggers their operation. Their defining feature is that they were present within the system well before the onset of an accident sequence. Like many other high-hazard, low-risk systems, the aviation system has developed such a high degree of technical and procedural protection that it is largely proof against single failures, either human or mechanical. The aviation system is more likely to suffer 'organizational accidents' [Reason 1990]. That is, a situation in which latent failures, arising mainly at the managerial and organizational level, combine adversely with local triggering events and with the active failures of individuals at the execution level [Reason 1997]. This concept is often graphically illustrated as slices of holed cheese, each slice representing a barrier at a different organisational level. The holes in the cheese are barrier failures and an accident occurs when the holes line up. The 'Swiss cheese' model has been useful to underline the importance of organisational factors in accidents. Perrow [1984] also describes accidents as occurrences where apparently trivial events cascade through the system to cause a large event with severe consequences. He uses the term 'normal

[9] Examples of non-accident risk in air transport are deep vein thrombosis and possible health effects of exposure to higher levels of electromagnetic radiation.

accidents' because these accidents start with an event that seems ordinary and happens frequently, almost always without causing great harm. Perrow also uses the term 'system accidents'; accidents are the result of interaction between components of a system rather than failures of individual components. Leveson is another advocate of a systems approach. Leveson [2004a] regards safety as an emergent property of a system. Emergent properties are properties observed at one level of abstraction which cannot be predicted (maybe even not explained) at another level of abstraction. According to Leveson's conception of safety, accidents occur when external disturbances, component failures, or dysfunctional interactions among system components are not adequately handled by the control system. They result from inadequate control or enforcement of safety-related constraints on the development, design and operation of the system. Safety is then viewed as a control problem [Leveson 2004b]. This view on safety is shared by Rasmussen [1997], who described safety management as keeping the organisation within a safe region, away from the boundaries of controllability. The organisation may be constantly pushed towards these boundaries by competitive pressures and financial restrictions, or may drift towards them through complacency. Risk control measures are then defined in terms of measures to detect, diagnose and act on close encounters with the boundary of the safe envelope [Hale et al 2007]. Hollnagel proposes to use non-linear system models. His functional resonance accident model uses the principle of stochastic resonance to describe accidents. Stochastic resonance is a phenomenon in which the combination of a stochastic (random noise) signal and a periodic modulated signal results in resonance, i.e. a relatively large selective response of the system to which the signals are fed [Hollnagel et al 2006]. Latent factors are not necessarily factors that are 'wrong' or outside the norm. Particular combinations of factors that are within the boundaries that one would expect can create a 'resonance' that leads to an accident. Dekker [2005] also emphasises the role of 'normal' factors in accident causation and calls it the 'drift to danger' in reference to Rasmussen. He claims that new, 'organic, co-evolutionary' concepts of accident causation are required. Hollnagel and Dekker were not the first to try to capture the seemingly 'stochastic' and 'organic' behaviour of accidents. Some researchers have made a comparison with epidemiological models to try to account for this 'randomness'. The threshold theorem, which states that the density of susceptible individuals must exceed a certain critical value in order for an epidemic to occur [Bailey 1975] seems to describe similar phenomena to latent failures and functional resonance. This proposition was already established in 1927 [Anderson 1991], even before Heinrich published his domino theory.

What do these theories tell us about (requirements for) causal risk models? The call for more systemic views on accident causation and the underlining of the role of 'normal' occurrences in accident causation emphasises the need for risk models that go beyond traditional 'hard wired' approaches like fault trees and events trees. The causal risk model should be able to represent the seemingly stochastic and evolving nature of the system in relation to its hazards.

2.6. Risk analysis and risk modelling

This section gives a brief historical overview of risk analysis and risk modelling. This is relevant as it is assumed that further development of current practice is more likely to be accepted by projected users of a causal risk model than a radically different approach.

Risk modelling has its roots in business forecasting and reliability engineering. Business forecasting developed in the late 17th century when entrepreneurs began collecting raw data for estimating the probabilities of uncertain outcomes. Lloyds list was launched in a London coffee house in 1696 and was filled with information on arrivals and departures of

ships and intelligence on conditions abroad and at sea [Bernstein 1996]. That information was used by ships' captains as well as insurance underwriters to estimate risks for the many sea routes that were being travelled. In the same period, the brilliant works of people like Jacob Bernouilli (1654-1705), Abraham de Moivre (1667-1754) and Thomas Bayes (1701-1761) provided the basic mathematical theories for risk analysis. Application of those theories was primarily in the field of financial risk analysis. The industrial revolution and increased mechanisation of society led to the application of the same mathematical tools for reliability analysis. Hardware designers however needed more than the ability to predict when their product would fail; they needed formal techniques to help improve the reliability of their products. For the US military it was important that hardware failures should not lead to failure of the mission. Therefore, Failure Modes, Effects and Criticality Analysis (FMECA) was formally introduced by the US military in 1949 (MIL-P-1629). This is regarded as one of the first systematic techniques for failure analysis. FMECA is a qualitative, systematic analysis of all conceivable failures and their effects.

The probabilistic approach to risk analysis and systematic analysis of failure effects were combined through representing whole sequences of events that could produce an accident; the 'causal chain', and calculating the probability of occurrence of those sequences. In risk assessments for technical systems this approach is often referred to as Probabilistic Safety Assessment (PSA)[10]. The representation of the 'causal accident chain' in a PSA is what we call a 'risk model'. The causal chains in PSAs were very limited and almost exclusively hardware related. Even when human factors came were introduced by Swain & Guttmann [1983] it was as a hardware analogy in terms of failure modes. The first attempts to put in the errors of commission and the underlying management factors came in the 1980s and are still ongoing. PSA was first applied for the US manned spaceflight program in the 1960s. NASA had posed a requirement for 0.999 probability of crew safety and 0.99 probability of mission success for each Apollo missions [Kelly 2001]. PSA was developed further in the nuclear power industry in the US. The first full-scale application of these methods to a commercial powerplant was undertaken in the Reactor Safety Study WASH-1400 published by the American Nuclear Regulatory Commission NRC in 1975 [NRC 1975]. Major accidents in chemical factories resulted in the introduction of similar techniques in those industries. After the Flixborough disaster[11] it was recommended that some method should be developed and used to estimate realistic accident consequences and to associate these estimates with a probability [Ale 2005]. Similar developments took place in other countries and probabilistic risk analysis is now applied in most safety critical industries. In several countries or industries there is some degree of standardisation in methodology in order to make the use of quantitative risk analysis viable in policy making and execution.

2.7. Conclusions for this section

Risk is a combination of the probability of occurrence of harm and the severity of the harm. Safety is freedom from unacceptable risk. The current research will be limited in scope to direct effects on the health of people. The acceptability of risk is influenced by risk perception. The level of perceived risk is influenced by many factors, including the level of control, the chance of multiple fatalities, the time passed since a similar event took place, whether the exposure to risk is voluntary, etc. For the purpose of this study the scope will

[10] Probabilistic Risk Assessment (PRA) and Quantitative Risk Assessment (QRA) are also commonly used but the meaning is the same.

[11] This accident happened on June 1, 1974 in a chemical plant in Flixborough, UK. A crack in a temporary bypass pipe caused the accidental release of a vapour cloud of cyclohexane. The vapour cloud exploded, as a result of which 28 people were killed.

be restricted to ‘objective’ risk. The focus of the study are causal risk models that have aircraft crash risk as an output. The research is not extended to third party risk, although the results of such models can easily be used as input for third party risk calculations.

Policies to control major risks have been in development from the 1960s onwards. Many of these policies are based on some sort of quantification of the risk that could be allowed to continue. The concept of a level of safety has certain implications. Because safety cannot be measured directly, an alternative approach to quantifying safety is necessary to be able to demonstrate that a certain target has been or will be met. In most safety critical industries, Probabilistic Safety Assessment is used to quantify risk. A causal model can be a tool for conducting probabilistic safety assessments. It is a probabilistic representation of the ‘causal chain’, the sequence of events that could produce an accident. Such a model can be used in both a TLS and an ALARP approach to safety. For both approaches it is necessary that the model produces output in objectively quantifiable units and that the results are reproducible, but the required accuracy of the causal risk model output is less for application in an ALARP approach than for a TLS application. Inclusion of human and organisational factors in the models is essential because they have a strong effect on safety and are influenced by managerial decisions. The call for more systemic views on accident causation and the underlining of the role of ‘normal’ occurrences in accident causation emphasise the need for risk models that go beyond traditional ‘hard wired’ approaches like fault trees and events trees. The causal risk model should be able to represent the seemingly stochastic and evolving nature of the system in relation to its hazards.

The following requirements were derived in this chapter:

- The model output should be ‘objective’ risk.
- The model output should be expressed in objectively quantifiable units.
- The model results should be reproducible.
- Human and organisational influences on safety should be represented in the model.

Chapter 3. Fundamentals of causation and probability

The subject of this research is causal risk modelling. A causal risk model describes, probabilistically, the causal relations between certain parameters and risk. To better understand these models and how they can be used for risk reduction and safety management it is necessary to describe 'causation' and some of its characteristics. To do that, we must also briefly touch some aspects of probability theory. Their consequences for the feasibility of building causal risk models will be summarised as a set of requirements and issues to be resolved.

3.1. What is causation?

Intuitively, we think of a cause as "something that can bring about an effect or result". Cause and effect are inseparable. Without some kind of effect the word cause has no meaning. However, the causes of an effect may not always be evident. This can be the result of our lack of knowledge, but there is also a second, more subtle characteristic that can make discussions about cause and effect rather complicated. Consider the example of a forest fire, which can be caused by, say, lightning or a match. In both cases however, there would be no forest fire without the oxygen that is in the air. Is it therefore correct to say that the oxygen is a cause of the fire? [Halpern & Pearl 2001]. Any fire requires three elements: Combustible material, oxidizer and heat. From the viewpoint of fire physics, all three elements are equally important: remove any one of these three elements and the fire will not start, or will die-out (Figure 3).

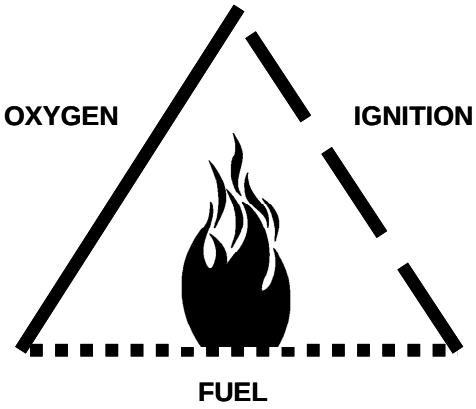


Figure 3: Fire triangle.

In the example of the forest fire, the lightning strike or the match is the final element that completes the fire triangle and as such is the initiator of the event. In legal terms the dried leaves and wood of the forest and the oxygen each are a *causa sine qua non* and the match is a *causa causans*. Most people think of 'cause' in terms of causa causans. This is not something new. In ancient Greek the word αιτια not only means cause, but also guilt and accusation [Muller & Thiel 1986]. When the objective is to assign blame for the forest fire, the difference between causa sine qua non and causa causans is considered relevant, the responsible is the causa causans. When the objective is to extinguish the fire, the difference

is irrelevant. The fire-fighter will remove any one of the three elements of the fire triangle. Which causal factor is removed depends on practical issues. For a forest fire it is more convenient to remove the heat by applying water, for a fire in a confined space, such as a frying pan on fire, it may be more practical to remove the oxygen by covering the fire with a blanket.

Causation versus association

Our knowledge of the chemistry of a fire allows us to determine the causes of the forest fire. In other cases our knowledge of underlying processes is limited, but we may observe that the occurrence of an event and a certain factor is often correlated. Such statistical associations are often used, for instance to demonstrate the beneficial effect of a new kind of medicine. Yet *association* is not necessarily *causation*. In Macedonia there is a strong correlation (in location) between the number of children being born and the number of nesting storks but storks do not cause children to be born. Instead there is a common cause, the villages, which also provide a good habitat for the storks [Van den Brink & Koele 1985]. If one compares sales figures for, say, ice-cream and sunscreen lotion, it will be discovered that these are statistically associated. When figures for ice-cream sales go up, so do figures for sunscreen sales. When sales figures for ice-cream go down, sunscreen sales also drop. Ice-cream sales and sunscreen sales are statistically associated but the one does not cause the other. They are both caused by a third common factor; hot sunny weather. Information on a statistical association between parameters A and B can be very useful when it is difficult to measure parameter A. Data on parameter B can be used to estimate parameter A values. It is not necessary to know the origin of the association to provide this estimate as long as the correlations hold across all conditions covered, but it is vital to distinguish these 'proxy' parameters from the causal ones, in case the model moves outside the region where the correlation holds. When the objective is to control parameter A, for instance increase its average value, a statistical association alone is quite useless. The owner of an ice-cream parlour, knowing the statistical association between ice-cream sales and sunscreen sales can put up bill boards for sunscreen lotion, but this will not increase his sales figures for ice-cream. To be able to control, it is necessary to have knowledge on causation. Because the goal of a causal risk model of aviation is to manage and improve safety (i.e. to control), these models must capture causation.

It is generally agreed that where an association is found it is, in the absence of a known or proven mechanism, ultimately a question of judgement whether the evidence is sufficient to establish a causal relationship [Court of Session 2005]. Causal claims cannot be substantiated from associations alone. Behind every causal conclusion there must lie some causal assumptions or known mechanisms [Pearl 2003]. Inferring causal relations requires subject-specific background knowledge. This means that in order to build a causal risk model of a particular system, we must *understand* that system. We must know the 'mechanisms' within the system. In the example of ice-cream and sunscreen, the causal relationship may seem obvious and even trivial, but in many (complex) systems causal relations are far from obvious. Extensive knowledge of the system mechanisms is required to correctly identify causal relations and thus to be able to construct a causal risk model.

This may seem trivial, but apparently it is not. According to Stephen Jay Gould [1981], the invalid assumption that correlation implies cause is probably among the two or three most serious and common errors of human reasoning. A few examples of the difficulties and controversies with statistical associations and causation are given below.

Confusion between causation and statistical associations

Post war diseases

War syndromes have been associated with armed conflicts at least since the U.S. Civil War (1861-1865) but research efforts to date have been unable to conclusively show causality [Hyams et al 1996, Jones et al 2002]. Explanatory causes that were proposed have varied from the heavy marching packs compressing the chest (U.S. Civil War 1861-1865, Boer War 1899-1902) to concussion from modern weapons (World War I 1914-1918), the use of Agent Orange (Vietnam War 1957-1975) and the use of depleted uranium in armour penetrating ammunition (Persian Gulf War 1991).

Health effects of electromagnetic fields

Concern about possible adverse health effects from exposure to extremely low-frequency electric and magnetic fields (EMF) emanating from the generation, transmission and use of electricity was first brought about by an epidemiologic study concerning a relation between risk of childhood leukaemia and exposure to EMF [Wertheimer & Leeper 1979]. But a more recent review of the epidemiologic literature on EMF and health concludes that in the absence of experimental evidence and given the methodological uncertainties in the epidemiologic literature, there is no chronic disease for which an explanatory relation to EMF can be regarded as established [Ahlbohm et al 2001].

These examples demonstrate why inferring causal relations requires *subject-specific background knowledge*. It is therefore essential for developers of a causal model of aviation safety to have substantial knowledge on all relevant aspects of aviation, including technology, operations, regulation and procedures for the complete lifecycle. Absence of subject matter knowledge will lead to causal risk models that produce misleading results. Going back to the example of post war diseases, it is easy to imagine how ineffective or even counterproductive treatments will be prescribed if any one of the above mentioned cause effect relations are reproduced alone in, say, a diagnostic model.

Indeed, statistical associations can be hopelessly confusing, as is illustrated by *Simpson's paradox*. Simpson's paradox refers to the phenomenon whereby an event C increases the probability of E in a given population p and, at the same time, decreases the probability of E in every subpopulation of p. For example, in Table 1, if we associate C (connoting cause) with taking a drug, E (connoting effect) with recovery, then the drug seems to have no beneficial effect on both males and females compared to no drug and yet is beneficial to the population as a whole.

Table 1: Simpson's paradox.

Combined	Effect	No effect	Recovery rate
Drug	20	20	50%
No drug	16	24	40%
Males	Effect	No effect	Recovery rate
Drug	18	12	60%
No drug	7	3	70%
Females	Effect	No effect	Recovery rate
Drug	2	8	20%
No drug	9	21	30%

In this example, the drug appears beneficial overall because the males, who recover (regardless of the drug) more often then the females, are also more likely then the females to use the drug. Being a male is a factor that influences both C and E [Pearl 2000a].

3.2. Conditional independence

The notion of conditional independence can be used to further illustrate the difference between causality and statistical association. Consider again the events IC that describes ice-cream sales and SC that describes sunscreen sales. We may observe a high correlation between IC and SC. Whenever we observe high values of SC, we observe high values of IC as well. We might well infer that SC causes IC. However, someone points out that IC and SC may have a common cause. Sunshine (SU) may cause IC and SC. In other words, there is no real causal link between SC and IC, but the constant association is caused by some other factor, SU, which causes both. We capture this by saying that given SU, SC tells us nothing about IC. In other words, SC and IC are independent given SU. SC does not cause IC.

Independent events; definition
Events A and B are independent if $P(A \cap B) = P(A)p(B)$

Conditionally independent events; definition
Events A and B are conditionally independent given the event C, written $A \perp B \mid C$, if $P(A \cap B \mid C) = P(A \mid C)\, P(B \mid C)$

This is a symmetric relation between A and B: $A \perp B \mid C$ implies $B \perp A \mid C$. It also implies $A \perp B^c \mid C$. However, it neither implies nor is implied by $A \perp B \mid C^c$. To assume that $A \perp B \mid C$ means that A and B are independent if C occurs and does not say anything of the relation between A and B if C does not occur.

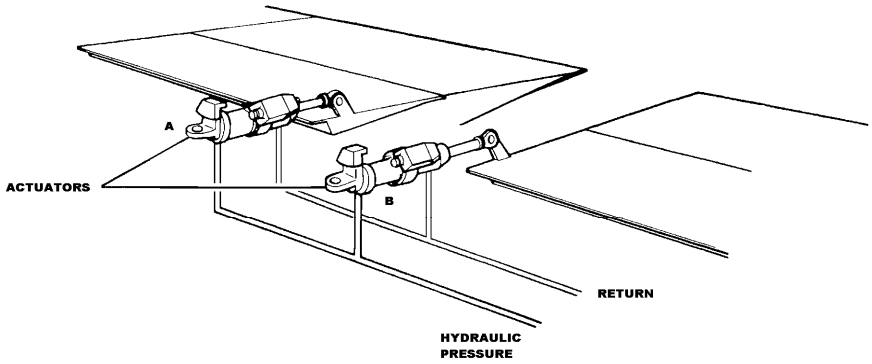


Figure 4: Flight control system.

As an example, consider the flight control system of Figure 4. Let A and B denote the events that actuators A and B fail to operate. Let C be a common mode failure that can cause an operational failure of both actuators, for instance loss of hydraulic pressure. If C occurs, then if B fails to operate, A will fail as well. If C does not occur, then A and B may still fail to operate, but then they would fail independently. Hence A and B are independent given that C does not occur, but given that C does occur, they are highly dependent.

Notice also that P(A|C) does not necessarily imply that A is the effect and C is the cause. Take the example A = car does not start, C = empty fuel tank. Both P(A|C) and P(C|A) exist. In order to be able to differentiate between cause and effect there must be additional information on the *direction* of the relation between C and A. In a causal model this direction is indicated by a directed edge (see section 3.7).

3.3. Causation to predict the future

One of the main reasons why we are interested in causation is because it allows us to predict system behaviour if we assume that the past and present determine the future. Among the first scientists to seriously address this issue were the German philosopher Immanuel Kant (1724-1804) and the French mathematician Pierre Simon Laplace (1749-1827). Kant stated that every effect has a cause which is a prior event and that this cause effect relation is 'fixed'. Therefore, if we have observed, in the past, that certain causes have certain effects we can assume that the same causes will have the same effects in the future [Kant 1781]. Laplace went even further when he published his theory of scientific determinism. He wrote [Laplace 1814]:

> "*We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at any given moment knew all of the forces that animate nature and the mutual positions of the beings that compose it, if this intellect were vast enough to submit the data to analysis, could condense into a single formula the movement of the greatest bodies of the universe and that of the lightest atom; for such an intellect nothing could be uncertain and the future just like the past would be present before its eyes.*"

According to Laplace's conception of nature, all laws in nature are deterministic, and randomness surfaces merely due to our ignorance of the underlying boundary conditions [Pearl 2000b]. Werner Heisenberg's (1901-1976) uncertainty principle shows that Laplace's vision, of a complete prediction of the future, cannot be realised (although Albert Einstein (1879-1955) did not agree with Heisenberg, because, according to Einstein, "God does not play dice with the universe"). According to Heisenberg's principle, effects do not follow causes in a rigorous chain of events [Feynman 1965]. Nevertheless, we assume that to a very good approximation the laws of science behave deterministically. The problem then becomes one of complexity, there are simply too many variables (or in Laplace's words, too many forces and beings that compose nature) to encompass. Because of this, all theories and models are approximations and we use probabilities to express the resulting uncertainty[12]. Apparent stochastic behaviour is then introduced in our deterministic view of the world as a result of our approximations. Each approximation introduces an error term with respect to the 'true' value. The error propagates when numerical operations with other uncertain quantities are being conducted. Under some conditions the error term may grow disproportionately fast. Prigogine [1977] introduced the concept of bifurcation points. According to this theory a system which has bifurcations will imply both deterministic and probabilistic elements. In between two bifurcation points the system behaves deterministically, but in the neighbourhood of the bifurcation points fluctuations play an essential role and determine the 'branch' that the system will follow. The outcome of the model will then be increasingly uncertain when more cause effect relations are called upon, typically when predictions further into the future are made. Rapid error accumulation may

[12] De Finetti (1906-1985) stated that probability is an expression of the observer's view of the world and has no existence of its own – probability does not exist (quoted by Lindley [1986]).

render the model outcome too uncertain to be of practical use. This is the reason why long-term weather forecasts are unreliable. The behaviour of the system can also be extremely sensitive to the values of the input parameters such that 'the flap of a butterfly wing in Brazil may cause a tornado in Texas' [Lorenz 1993]. Small variations in the input parameters may lead to large differences in the model results. For these reasons *uncertainty analysis* and *sensitivity analysis* are an essential element of model development.

But as long as extreme conditions (such as aircraft approaching the speed of light) are avoided and the model is not used to look far into the future (i.e. we do not look beyond the next bifurcation point), there is no fundamental reason why a causal model would not be able to predict the future, albeit with some uncertainty. How far ahead the model is able to predict and with which accuracy depends on the level of detail of the required output, the accuracy of the model input, and the validity of the model.

The challenge is to "know all of the forces that animate nature and the mutual positions of the beings that compose it". As mentioned before, this task is too big and we must think of approximations. For the purpose of this study we are not interested in a complete prediction of the future but only a part of it. We must therefore define a boundary of what we need to represent in the model. The 'rest of the world' can then be ignored, except for changes (e.g. in technology or organisation) outside the defined boundary that affect changes in the elements and structure of the model. Instead of determining 'the forces that animate nature', we could choose to look for causation at a higher level of aggregation. Another solution would be to leave out parts with a (predictable) minor effect.

3.4. Singular and generic causal relations

Confusion between singular and generic causal relations is a potential source for misunderstanding which can undermine one's confidence in model results. Therefore the difference is addressed in this section.

Singular causal relationships are those between concrete singular occurrences of events. An example of a singular causal relationship is "Cigarette smoking is the cause of mrs. McTear's lung cancer". A generic causal relationship would be that "smoking cigarettes causes lung cancer". A causal risk model is most likely a set of generic causal relations; it does not make much sense to construct a model that can only represent a concrete singular occurrence of events. In most cases, we arrive at generic causal relations by generalising from individual cases of occurrence and then apply this general knowledge to other individual occurrences [Hesslow 1988]. Under current law, evidence of a generic causal influence is not always strong enough to enforce liability claims in court. There must be proof that a given phenomenon was the actual singular cause of the observed effect. The following statement by a judge when delivering judgement in the case of an individual against a tobacco company illustrates this:
"*Given that there are possible causes of lung cancer other than cigarette smoking, and given that lung cancer can occur in a non-smoker, it is not possible to determine in any individual case whether but for an individual's cigarette smoking he probably would not have contracted lung cancer. Epidemiology cannot be used to establish causation in any individual case, and the use of statistics applicable to the general population to determine the likelihood of causation in an individual is fallacious* [Court of Session 2005]. [13]

[13] This is one judgment but there are others, especially on asbestos, placing the onus of proof on the company to show that a generic cause was not valid in a specific case.

The difference between singular and generic causal relations is particularly important when causation is used to determine responsibility or guilt, but that is not the purpose of the causal risk models that are subject of this thesis.

The other way around, the absence of a singular cause effect relation is not sufficient to negate a generic causal relation. A 95 year old heavy smoker without health problems is not a proof that smoking is not harmful. Yet such singular examples may be compelling and people's faith in causal risk models may very well be eroded if they happen to know a few of such singular examples that are seemingly contradictory to the generic model relations. A risk model that not only explicitly represents failure scenarios but also success scenarios would be helpful in this respect.

3.5. Strong and weak causal relations

Another potential source of confusion is the difference between weak and strong relationships. In some cases, cause effect relations can be considered strong, in the sense that the occurrence of the cause will always result in the effect. An example is the cause 'empty fuel tank' and the effect 'car does not start'. Other cause-effect relations may be weaker. An example is the cause 'cigarette smoking' and the effect 'lung cancer'. The fact that someone has a habit of smoking cigarettes does not *always* results in the person developing lung cancer, although the probability for lung cancer is greatly increased when someone smokes. The 'weakness' of the cause-effect relation depends on the conditional probability P (effect | cause). If this conditional probability is 1 or close to 1 (as in the case of the car example) the cause-effect is strong. If the conditional probability is close to 0, the relation is weak. A weak statistical association should however not be taken as indicating a non-causal relationship. Weak statistical associations may be seen for instance if an effect is common and there are numerous possible causal pathways towards that effect.

A weak cause-effect relationship does not require high probabilities to be acceptable. What is important is the statistical relevance of the relationship. When a certain drug increases the recovery rate from 5% to 10% it is the statistical significance of the difference between the treatment group and the control group that determines whether we accept that there is proof that the drug is working.

3.6. The beginning and the end of causation

An obvious question to ask when the cause of an effect has been determined is "What causes the cause?" This question is not only interesting from a philosophical point of view but is also relevant for accident prevention. If the cause of an accident is determined to be pilot fatigue this is interesting to know but in order to prevent similar occurrences from happening we usually need to know what caused the fatigue, i.e. what caused the cause[14]. Only then are we able to take measures to safeguard against similar events. If we are sufficiently clever and persevering we can continue determining the causes of causes right up to the big bang that created the universe. In a similar fashion we can think of the effects of effects. In the event of an aircraft crash an immediate effect may for instance be the release of toxic substances into the atmosphere if the aircraft was carrying hazardous materials as cargo. This may then cause, perhaps even years later, health problems for people exposed to those materials. Causal chains never really end, but for practical purposes the model must.

[14] Unless we can detect the presence of this proximal cause and take preventive measures on it, i.e. in this case detect fatigued pilots and prevent them from flying.

If we adopt Leveson's [2004a] system approach to accidents, where safety is seen as an emergent property of a system, we are still faced with a similar problem: the operation of the processes at the lower level of abstraction cause the emergent properties at the higher level of complexity. The question 'how far back or forward the causal chain' is replaced by 'how many levels down or up the scale of abstraction'.

So how do we decide how far back or up along the causal chains to model? In the past the answer to this question has often been data driven. We know how often component X fails, so we do not have to model why. This approach is satisfactory if the purpose of the model is the measurement of risk but if the purpose is to help determine where and how to intervene to control risk we need to use different criteria to determine how far to extend the causal model. We will need to go back along the causal chain until we find a place convenient for intervention. This then needs defining. According to Reason et al [2006], attempts to track down possible errors' and accidents' contributions further back in the causal chain, identifying factors that are widely separated in both time and place from the events themselves, have gone too far. In practice a useful criterion for the extent of the causal risk model is that it will have to include decisions and actions up to the highest managerial level of the actors that are directly involved.

3.7. What is a causal model?

The previous sections addressed causality in general and now and then referred to a causal risk model. But what exactly is a causal risk model? According to a formal definition, a causal model is *a mathematical object that provides an interpretation and computation of causal queries about the domain [Galles & Pearl 1998]. A causal model can be associated with a directed graph in which each node corresponds to a variable of the model and the directed edges (arrows) point from the variables that have direct influence to each of the other variables that they influence.* This formal definition will be adopted in this thesis.

Such a causal model can be used for *backward* reasoning and *forward* reasoning.

Backward reasoning.

Consider a situation where a patient visits a doctor because of, say, unremitting pain in the abdomen. The doctor will investigate the patient using a series of tests and use the results of the tests to come up with a diagnosis and start treatment. The doctor uses the symptoms, i.e. the effects, to determine the most likely cause of the problem. This is called backward reasoning or causal inference: determining the most likely cause, given some observed effects.

Forward reasoning.

Consider a situation where a meteorologist observes particular phenomena, such as the location of a region of low pressure above the North Sea. The meteorologist will investigate the observed phenomena to come up with a weather prediction. The meteorologist uses the causes to determine the most likely effects. This is forward reasoning: determining the most likely effect, given a certain cause.

Causal models for backward reasoning are specifically applied in medical sciences, but some applications in other fields have also been developed, for instance in social sciences. Examples are found in Onisko et al [1998] and Strotz & Wold [1960]. For air transport safety, a causal model for backward reasoning could be applied during accident investigation. The model would then help to determine the most likely cause of the accident, given the evidence found by the investigators. However, given the thoroughness

with which aviation accidents are currently investigated, it seems highly unlikely that a causal model would be able to add anything useful here.[15] However, it could order and organize the results across many accidents and that is essentially useful in support of the forward reasoning.

Causal models for forward reasoning are applied in a wide range of sciences, including meteorology, finance and engineering. Models for *risk assessment* are examples of causal models for forward reasoning. As will be shown in the next sections, there is a need for such causal risk models for air transport safety. Therefore in the remainder of this thesis the scope will be limited to causal models for forward reasoning. Most importantly, there is need for a special case of forward reasoning; predicting the effect of changes. This has added complications because it may require the potential to change the model, thus setting requirements for the dynamics of the model.

3.8. Conclusions for this section

For the purpose of this thesis, the following definition of a causal model was adopted:
A causal model is a mathematical object that provides an interpretation and computation of causal queries about the domain [Galles & Pearl 1998]. A causal model can be associated with a directed graph in which each node corresponds to a variable of the model and the directed edges (arrows) point from the variables that have direct influence to each of the other variables that they influence.

Usability requirements of the model determine whether statistical associations alone can be used to construct the model. If the purpose of the model is to predict, then a model based on statistical associations can be useful. But in our case the objective of the model is to intervene and statistical associations alone are not sufficient, there has to be an underlying assumption, a causal assumption. This causal assumption should be based on deeper knowledge on cause and effect through theoretical studies and accumulated knowledge. Statistical associations should be used carefully, as demonstrated by Simpson's paradox and the examples of controversies related to interpreting statistical associations. It is therefore essential for developers of a causal model of aviation safety to have substantial knowledge on relevant aspects of aviation, including technology, operations, regulation and procedures for the complete lifecycle.

What is actually the cause or causes of an event is not always unambiguous. To a certain extent, it is an assumption that has to be agreed upon by those who work with it. The agreement is necessary because the assumption has consequences. It is therefore necessary that the underlying assumptions are transparent. Causal chains never really end, but for practical purposes the model must. In practice a useful criterion for the extent of the causal risk model is that it will have to include decisions and actions up to the highest managerial level of the actors that are directly involved.

Part of the complexity of causal relations is their apparent probabilistic nature. We often arrive at generic causal relations by generalising from individual cases of occurrence and then apply this general knowledge to other individual occurrences. But evidence of a generic causal relationship is not sufficient to prove a singular causal relationship, and the absence of a singular cause effect relation is not sufficient to negate a generic causal relation. This can lead to confusion and can seriously erode the confidence in causal risk

[15] Aircraft accident investigation is a process that takes approximately 2 years to complete and involves dozens of people, many of them working almost full time on a single accident.

models. A strategy for avoiding this mistrust is explicitly representing failures scenarios as well as success scenarios in the model.

The following requirements were derived in this chapter:

- Model developers should have substantial knowledge on all relevant aspects of aviation, including technology, operations, regulations and procedures for the complete lifecycle.
- The risk model should explicitly represent failure scenarios as well as success scenarios.
- The model will have to include decisions and actions up to the highest managerial level of the actors that are directly involved.

The following issue needs to be resolved:

- A boundary must be defined of what needs to be represented in the model.
- The highest managerial level of the actors that are directly involved needs to be defined.

Chapter 4. User needs

This section will describe user needs for a causal risk model and model design specifications. When explicitly asked, most potential users have difficulty in listing their requirements because they do not have a clear picture of the possible performance of a causal risk model. Therefore user requirements must be derived in a different way. This is done by first describing how flight safety has evolved over the years and identifying the obstacles for further safety improvement. These obstacles are basically the problem that a causal risk model must help solving and are therefore the basis for the requirements. We then ask the question of who the expected users will be. A review of existing quantitative safety regulation in aviation, including a description of how the industry is expected to demonstrate compliance to those regulations, will provide insight into the type of regulatory requirements that can be expected for a causal risk model. Being one of the possible areas of application of a causal risk model, the use of quantitative risk assessment is discussed to be able to comprehend the resulting methodological requirements. Emphasis is placed on the position of the test pilot, because this explains the fundamentally different views on the ability to make quantitative assessments of accident risk, specifically when it comes to modelling human actions and decisions, between aircraft design and air traffic management. As a matter of fact, all stakeholders in the aviation system hold a different view on aviation safety and consequently have different needs with respect to aviation safety modelling. Nevertheless, some common requirements can be identified.

4.1. A brief history of aviation safety

Aviation has realised tremendous improvements in flight safety since the first commercial airlines started operating around 1920. While flying in those pioneering years was still an adventure, and pilots like Jean Mermoz [Mermoz 1937] and Charles Lindbergh [Lindbergh 1953] were true death defying heroes, air transport today is fundamentally safe and the aircraft has become a means of mass transportation. More than 2100 million passengers were carried on scheduled flights in 2006, and 39 million tonnes of freight were transported (ICAO statistics, December 2006). Continuous technological and operational improvements have led to today's exemplary safety record. This section describes the history of aviation safety improvements and explains why further improvements cannot be expected unless something different is done to find possible ways for improvement. The 'something different' includes the development and use of causal risk models.

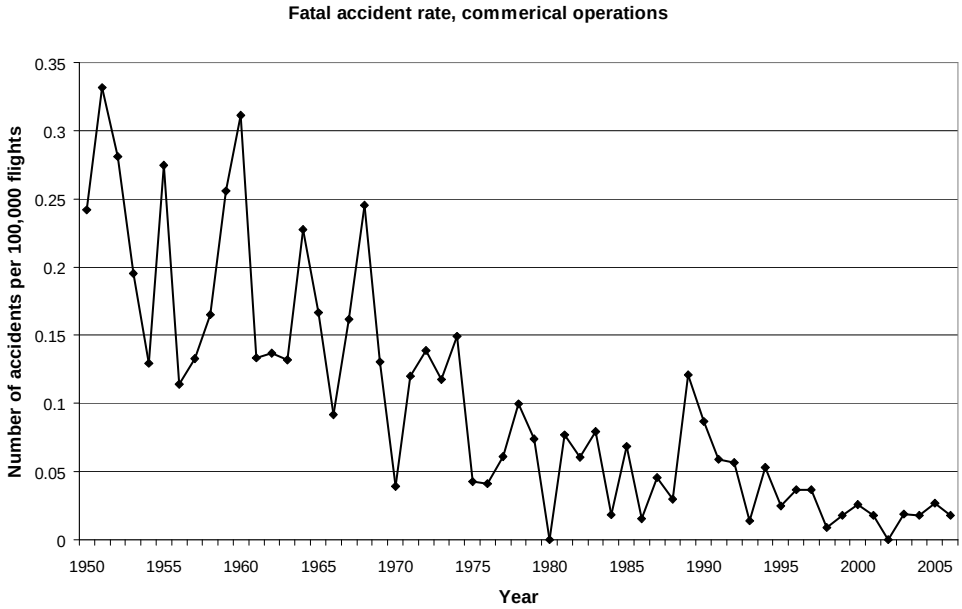


Figure 5: Accident rate history Source: NLR Air Safety Database.

In the early days of aviation, from 1903 to 1914, the man (they were all men) who built the aircraft was often also the first person to fly the aircraft, and he used what he felt and saw during the test flights to further improve the vehicle. After World War I (1914-1918) commercial aviation rapidly developed and so did regulation. Most states established some sort of aeronautics branch that was made responsible for defining and ensuring the airworthiness of aircraft, licensing of airmen and development and enforcement of air traffic rules [Hansen et al 2005]. A Certificate of Airworthiness became a mandatory requirement for each aircraft model. In the Netherlands, the Certificate was only obtained if the engineers of the 'Rijks Studiedienst voor Luchtvaart'[16] (R.S.L.) had provided a positive verdict to the 'Commissie voor Keuring van Luchtvaartmaterieel'[17]. In addition, the State appointed pilot of the R.S.L. had to make a test flight. The description of the certification flight tests of the Lambach HL II provides a vivid example of how such a flight was conducted in 1937 in the Netherlands.

Description of the certification test flights of the Lambach HL II.
The Lambach HL II was a single seat aerobatics aircraft, designed by Dutch engineer Hugo Lambach in 1937. It was powered by a 130 H.P. Gipsy Major engine and specifically developed to compete in aerobatics contests that were very popular in that time. Certification test flights took place in May 1937 from the airfield "Ypenburg", near The Hague. Three test flights were made, but only after the weight of the aircraft had been determined under close supervision of ir. A.G. Von Baumhauer of the 'Commissie voor Keuring van Luchtvaartmaterieel'.

[16] Government Service for Aeronautical Studies
[17] Commission for Approval of Aviation Equipment

The first flight was made by dr. ir. H.J. van der Maas according to a previously established program. The objective of this flight was to determine the flight characteristics of the aircraft. As part of the flight the longitudinal stability was determined, a yaw test was executed, stick forces and stick displacements were measured, etc. The effects of stopping and starting the engine at different flight states were also determined. Next it was the turn of the company pilot, Hein Schmidt Crans, who performed some aerobatics. He showed that the aircraft could perform six consecutive spinning turns, which was a requirement for an approval certificate for unlimited aerobatics. Thereupon he flew two left rolls and one right roll and a half roll followed by inverted flight. During the inverted flight the engine was stopped and again started. This was followed by a half outward loop upwards, a half loop and a half roll. Dr. Ir. Van der Maas then made a third test flight during which also some aerobatics were undertaken. He performed a loop, spin, roll, snap roll and a vertical dive such that the end speed was obtained. Van der Maas performed no less than eight consecutive spin turns, after which he neatly noted in the flight test report "After exiting the spin, which probably took two turns, several seconds passed before the pilot had regained a focused view of the clouds". In addition the test report noted that the engine caused some nuisance because occasionally it stalled briefly. Sometimes even a jolt was felt in the throttle, which indicated problems with the carburettor. Nevertheless, the overall conclusion was that the aircraft conformed to the "requirements for flight characteristics in Order 471 II" [Nijenhuis et al 1996].

Figure 6: Lambach HL II.

The fact that the certification test flight for this small aircraft was conducted by dr. ir. Van der Maas[18] is exemplary of the importance that was (and still is) attached to the role of the certification test pilot, and the (scientific) skills and knowledge that were needed to perform this task. The test pilot combined the functions of pilot, observer and scientist.

[18] Van der Maas had graduated as a maritime engineer from Delft University in 1923, after which he was employed at the R.S.L. He became the first professor in Aeronautical Engineering in the Netherlands in 1940 and played a prominent role in setting up the Faculty of Aeronautical Engineering at Delft University of Technology. In 1950 he was appointed as chairman of the Foundation N.L.L, which was renamed NLR in 1961. Van der Maas died in 1987.

After World War II aviation became a means of mass transportation. In the 1950s accidents during (test) flights were frequent [Wolfe 1979], but the knowledge gained during the flights and as a result of accident investigation contributed greatly to the development of aviation as a reliable and safe means of transport. Safety improved tremendously from 1950 to 2000, the worldwide fatal accident rate dropped from 3 accidents per million flights to less than 0.5 (see Figure 5). Most of the technological developments that contributed to improved safety were driven by a desire to increase the economical viability, with safety as a 'byproduct'. Aircraft make money when they are in the air, not when they are grounded, and thus much effort was put into improving aircraft reliability and the ability to fly in adverse weather conditions. Technological advances that contributed significantly to aviation safety include the following:

- Jet engines, which have a lower failure rate than piston engines due to their inherently simpler design with less accelerating parts [Brown 1981].
- Radio navigation systems have improved the flight crew's ability to navigate and have virtually eliminated occurrences where aircraft get lost and run out of fuel before finding a suitable landing place [Leary 1992, Abbink 1996].
- Precision approach systems and automatic landing systems (precision approaches are five times more safe than non-precision approaches) because they help to keep the aircraft on the correct approach path, even in conditions of reduced visibility, like fog [Enders et al 1996, Abbink 1996].
- On-board weather radar has helped in avoiding hazardous weather.
- Glass cockpits[19] provided an enormous increase in the possibility to directly interpret all aspects of navigation and the avoidance of bad weather and greatly increased the situational awareness of the pilots by integration of all information [Abbink 1996]. They also provide better support to the flight crew in the case of technical failures, reducing the likelihood of flight crew errors in such abnormal or emergency conditions [Roelen & Wever 2005b]. Paradoxically, they have also contributed to accidents due to poor crew understanding of their software logic [Sarter & Woods 1995].

None of these technological advances were safety driven. The reason to develop the jet engine was the need for improved performance in terms of aircraft speed and engine power. Radio navigation allowed the reduction of the flight crew to two pilots and a flight engineer, automatic landing systems increased productivity because aircraft were far less dependent on local weather conditions. Precision navigation by means of GPS became available thanks to the United States Department of Defence that needed accurate targeting and guidance for their weapon systems. The step from electromechanical cockpit instruments to a glass cockpit was made to reduce maintenance costs and weight. It also allowed the elimination of the flight engineer from the flight deck, and thus provided a significant reduction in direct operating costs for the airlines. In addition to technological leaps, overall system reliability has gradually improved. Today an engine failure is such a rare event that most pilots will never experience it in their career[20]. Technological

[19] In a glass cockpit the traditional electromechanical instruments are replaced by displays that are able to present integrated information.

[20] The in-flight engine shutdown rate of a modern engine like the GE90-115B is 0.005 per 1000 flight hours [Horibe et al 2004], i.e. one shutdown every 200,000 hours. For a twin-engine aircraft like the Boeing 777 this means one shutdown per 100,000 hours of flight. An airline pilot will have approximately 20,000 flying hours at the moment of his

advancement in ATM on the other hand has been relatively slow because ATM has traditionally been the responsibility of national authorities and therefore has, until recently[21], not experienced the effects of competition. A result is the use of technology that basically dates from WW II; VHF voice communication and radar.

Regulatory requirements matured. Strength requirements for aircraft structures were specified in terms of limit loads (the maximum loads to be expected in service) and ultimate loads (limit loads multiplied by prescribed factors of safety of 1.5). Aircraft manufacturers had to show compliance to these requirements by analysis supported by strength testing of sub-components, full scale components or full scale tests of assembled components (such as a nearly complete airframe) up to limit loads and ultimate loads. The structure must be able to support ultimate loads without failure for at least 3 seconds. When tests are used to show compliance, an additional safety factor of 1.15 is typically applied to account for variability in material properties [EASA 2003a].

Regulatory requirements regarding aircraft system safety also evolved. When automatic landing systems were designed in the 1950's, aircraft and equipment manufacturers asked the airworthiness authorities what requirements or special conditions would be applied to such systems. The authorities did not consider that they had the background and experience to write detailed requirements. Instead a target level of safety was agreed with the manufacturers as a base for certification. Applicants were required to make a case for their individual systems by assessing them against the declared objective. Subsequently detailed methods of establishing compliance were evolved. With the further development of technology and complex systems, in particular related to supersonic transport aircraft, the authorities from the USA, UK and France produced requirements that included the general principle that *an inverse relationship should exist between the probability of occurrence and the degree of hazard inherent in its effect on the capability of the aircraft*. In addition, they specified a level of safety to be achieved, in qualitative and quantitative terms, and required that a safety assessment should be made. The concept of requiring an inverse relationship between probability of failure and the severity of the failure effect had in practice been established from the early days of aviation and resulted in, for instance, twin magnetos and twin spark plugs on engines, dual wing rigging wires and dual control cables. What was new in the requirements was the introduction of quantitative and numerical methods for addressing risk instead of the intuitive methods that had been used before [Lloyd 1980, Lloyd & Tye 1982]. In assessing the acceptability of a system design it was recognised that rational probability values would have to be established. Historical evidence indicated that the risk of a serious accident due to operational and airframe-related causes was approximately 1 per million hours of flight. Furthermore, about 10 percent of the total could be (albeit arbitrarily) attributed to failure conditions caused by the aircraft's systems problems. It seemed reasonable that serious accidents caused by systems should not be allowed a higher probability than this in new aircraft designs. It was thereby possible to require for new designs that the probability of a serious accident from all such system failure conditions be no greater than 1 per ten million flight hours, or 1×10^{-7} per flight hour. As it is not possible to say whether the target has been met until all the systems on the aircraft are collectively analysed numerically, it was assumed, arbitrarily, that there are

---

retirement. The probability that at some point in his entire career he will have had to shut-down an engine in-flight is approximately 1 in 5.

[21] ANSPs in Europe are becoming more independent from the national authorities; NATS in the UK for example is fully privatised and LVNL in the Netherlands is a 'Zelfstandig bestuursorgaan'.

about 100 potential system failure conditions in an aircraft, which would prevent continued safe flight and landing. The target allowable risk of 1×10^{-7} per flight hour was thus apportioned equally among these conditions, resulting in a risk allocation of not greater than 1×10^{-9} to each. The upper-risk limit for failure conditions, which would prevent continued safe flight and landing would be 1×10^{-9} for each hour of flight, which establishes an approximate probability value for the term 'extremely improbable'. Failure conditions having less severe effects on the capability of the aircraft could be allowed to be relatively more likely to occur.

Accidents involving (many) fatalities have also been important drivers for safety improvements. Each accident is thoroughly investigated and measures are taken to prevent a similar occurrence. Simply stated this is the 'fix and fly' approach. In some instances accidents may reveal weaknesses that hitherto were unknown, such as in the case of the Comet disasters (see infobox Comet disaster), in other cases the sequence of events that leads up to the accident is not unique and has been seen in the past in either accidents or incidents (see infobox the DC-10 story). Usually it is the accident however, i.e. the realisation of a low probability event, that triggers remedial action, rather than the knowledge that such an event *could* happen. In the case of Concorde for example, although there had been five tyre bursts on Concorde that had caused structural damage to fuel tanks, it was not considered necessary to reinforce those fuel tanks. Only after the Concorde disaster was Kevlar lining fitted to the fuel tanks as an extra protection against damage by burst tyres (see info box Concorde accident). Remedial action may include short-term regulatory activities such as issuing Airworthiness Directives, but may also be more fundamental changes to processes or activities. The aircraft maintenance program development process for instance was changed significantly following the Aloha Airlines Boeing 737 accident in 1988[22] and again after the TWA Boeing 747 accident in 1996[23]. In the Netherlands, the El Al Boeing 747 accident in 1992 resulted in fierce debate, a parliamentary enquiry and legislation on third party risk around airports (see also Appendix A: The history of third party risk regulation at Schiphol airport). Sometimes accidents are a driver for technological improvements. Advanced warning systems such as TCAS, GPWS and windshear warning systems were developed as a direct response to (a series of) accidents.

The Comet disasters

The de Havilland DH.106 Comet was the world's first operational passenger jet aircraft. With a cruising speed of more than 800 km/h at an altitude of 40.000 ft it provided unprecedented levels of comfort to the passenger [Goold 2000]. But within two years of entering service, a series of accidents initiated the end of the Comet and of British leadership in aeronautical engineering. First there were take-off accidents in 1952 and 1953

[22] On April 18, 1988, a Boeing 737-200 of Aloha Airlines experienced an explosive decompression and structural failure at 24,000 ft due to fatigue cracking. Approximately 18 feet of fuselage skin and structure separated from the aircraft [NTSB 1989].

[23] On 17 July 1996, Trans World Airlines Flight 800 crashed minutes after take of from John F. Kennedy International Airport, New York. The cause of the accident was an explosion of the center wing fuel tank. The source of ignition energy for the explosion could not be determined with certainty, but, of the sources evaluated by the investigation, the most likely was a short circuit outside of the center wing tank that allowed excessive voltage to enter it through electrical wiring associated with the fuel quantity indication system [NTSB 2000]

caused by overrotation and subsequent wing stall. These problems were solved by a modified wing leading edge. Regulation was also changed as a result of these accidents, requiring demonstration of a minimum unstick speed with maximum angle of attack. [EASA 2003a]. But then in 1954 a Comet disintegrated while climbing to its cruising altitude, and the British Overseas Airways Corporation BOAC grounded its fleet of jet aircraft. However, no defects were found and the aircraft were put back into service. Sixteen days later, another Comet broke-up in mid-air, and the aircraft's Certificate of Airworthiness was suspended. After extensive testing, it was discovered that the accidents had been caused by fatigue cracking of the fuselage skin. This was a new phenomenon, first encountered by the Comet because its high cruise altitude resulted in relatively large stress cycles to the fuselage (pressure cabin) skin. Production of the Comet was halted. A modified version, the Comet 4, flew in 1958 but by that time competing aircraft were on the market and the Comet 4 was not a commercial success.

As a result of these developments, air travel today is astonishingly safe. For air carriers in Europe or the US, the fatal accident rate per million flights is approximately 0.2 [CAA-NL 2007]. If a person were to take a flight each day, he would, on average, experience a fatal accident after 13,700 years. Even then, the likelihood of being killed in that accident is less than the probability of survival; an analysis of fatal accidents by US domestic airlines in 1987-1996 showed that, on average, a fatal accident kills 44% of the occupants of the aircraft [Barnett & Wang 2000]. A fatal accident is defined as an event in which at least one of the occupants is killed, so even in the case of a fatal accident the majority of the occupants may survive. Does this mean that a sufficient level of safety has been reached? Certainly not from the viewpoint of the regulator. Aircraft accidents cost lives and are a cause of emotional trauma to victims' relatives and friends, to people working within the industry, to communities. Crashes consume the attention of the media in a unique way; any fatal airline crash will always be a headline story. As a result, it is feared, the general public may loose confidence in the air transport system. The extent to which 'air crashes' are reported in the media has always been of concern to the industry and has been seen as a possible limiting factor to growth for at least 60 years. In 1943, a confidential 'report' produced by the Curtis Wright Corporation on behalf of the US Government stated that if air safety did not improve from the level achieved during the 1930s, the rapid expansion in air travel expected after the war would result in an unacceptable level of air crashes and would limit growth. These views have been repeated many times over the years, for instance by the US Secretary of Transport at the Aviation Safety Conference in Washington D.C. in 1995 [FAA 1995b] and are still regularly voiced today. Perceived public demand for safe air travel is counterbalanced by a demand from within the industry to remain profitable. As early as 1959, Walter Tye (UK ARB – Air Registration Board) suggested that an optimum balance may have been achieved between airline flight safety and costs. He argued that any further marked increase in safety by (then) existing methods would cost far more than passengers or the industry would be willing or able to pay (as quoted in [Bland 1962]). This view was repeated, at the 1961 FSF Seminar, by M.G. Beard, American Airlines, who said he believed that the major airlines had already reached a position of 'diminishing returns' in that they had (in 1961) reached a very high degree of safety and further improvements would cost more and more. However, the costs associated with an accident can also be enormous, and while much of this is absorbed by insurance, the effect of an accident on the airline's stock price and ticket sales is uninsured. Some airlines have actually gone bankrupt after an accident, examples are Birgenair[24] and Helios

[24] On February 6, 1996 a Birgenair Boeing 757 crashed shortly after take-off from Puerto Plata in the Dominican Republic, killing all 189 occupants. Birgenair was a Turkish holiday

Airways[25], although this is exceptional. Accident prevention and safety management therefore are as important as ever, despite the high level of safety that has been achieved.

The very success of the aviation system in terms of safety performance makes it difficult to determine the linkage between actions and decisions and safety outcomes. Because accidents are so rare it is impossible to determine directly whether actions or decisions had any effect on safety. Hence there is no proper feedback loop from day-to-day operations to safety decision-making [Amalberti 2001]. In other words, the fix and fly approach is no longer effective. Yet new hazards are constantly lurking due to such factors as insertion of new technologies, trends in the economy and change in the regulatory environment. To maintain and improve the current level of safety, there is a need for an aviation safety performance evaluation methodology that is not based on analysis of fatal accidents and hull losses alone. Such a methodology should include a safety evaluation construct, safety hazard identification and risk analysis, safety performance measures and risk indicators and safety diagnostic procedures [FAA 2004]. A causal risk model can play a central role in such methodology.

In conclusion, the main drivers for aviation safety improvement have been technological development and the industry's practice of thorough accident investigation and subsequent remedial actions. Because of the high level of safety, the latter approach is no longer effective and there is need for a new safety performance evaluation methodology, of which a causal risk model could be a central part. We have also seen how in the past the regulatory authorities shifted from prescribing detailed requirements to a quantitative target level of safety, and to requiring the industry to demonstrate numerically that the target is obtained. For parts of the aviation system this is already done, and existing methods for analysis may suffice, but there is a need now to also analyse the aviation system as a whole. This is another potential application of a causal risk model, which requires the model results to be quantitative and reproducible.

The causal risk model should therefore:

- Include risk indicators,
- Provide reproducible results,
- Represent the aviation system as a whole, and
- Have quantitative model output.

The Concorde disaster

The BAC/Sud-Aviation Concorde aircraft was the embodiment of aeronautical progress. With a cruising speed of more than twice the speed of sound it was twice as fast as all other commercial aircraft. To attain these speeds, the shape of the aircraft was special as well, with an elegantly pointed nose and beautifully curved ogival delta wing platform. The shape of the wing was in fact a compromise between the requirement to be able to fly at supersonic speeds and still provide satisfactory characteristics at low speed for take-off and

---

charter operator, most people on-board were German tourists. The accident was initiated by incorrect airspeed indications due to a blocked pitot tube [JIAA 1996].

[25] On 14 August 2005, Helios Airways flight 522, a Boeing 737-300 crashed near Athens in Greece on a flight from Larnaca, Cyprus. All 121 occupants were killed. The immediate cause of the accident was flight crew incapacitation due to hypoxia. Helios Airways was a low fares carrier, established in 1999. At the time of the accident, Helios was operating four aircraft and conducted flights out of Larnaca and Paphos, Cyprus [AAIASB 2006].

landing. Nevertheless, Concorde's typical take-off speed of 200 kts is approximately 35% higher than that of other commercial jet aircraft. These high take-off speeds are more demanding to the tyres, and the rate of tyre bursts for Concorde has always been relatively high at one occurrence per 1,500 cycles rather than the one occurrence in 100,000 cycles which is the rate for modern airliners such as the Airbus A340 [BEA 2001]. Modifications such as strengthened tyres led to a reduction in the tyre failure rate, but it still remained relatively high. This was not considered to be a major safety hazard until 25 July 2000, when during take-off from runway 26R at Roissy Charles de Gaulle Airport, shortly before rotation, the front right tyre of the left landing gear of Air France Concorde F-BTSC ran over a strip of metal which had fallen from another aircraft. The tyre exploded and debris was thrown against the underside of the wing, leading to a rupture of one of the wing's integral fuel tanks. A major fire, fuelled by the leak, broke out immediately under the left wing. Ingestion of tyre debris and hot gases caused malfunctions to the engines on the left wing, and the fire destroyed control surfaces at the wing's trailing edge. The flight crew had no possibility to recover from the catastrophic situation, and the aircraft crashed less than two minutes after take-off. All 109 passengers and crew members, as well as 4 people on the ground, were killed.

Previous to this accident, there had been five tyre burst events on Concorde which had caused structural damage to fuel tanks. Reinforcement of the lower wing was considered after the first incident, but it was considered unnecessary [BEA 2001]. After the accident, Air France immediately grounded its Concorde fleet, and although British Airways (the sole other operator of Concorde) continued operating the aircraft for a few weeks, the CAA ordered the British fleet to be grounded as well. An extensive modification program including the fitment of Kevlar lining to key fuel tanks and installation of strengthened tyres was executed, and the Certificate of Airworthiness was re-issued 14 months after the accident. For commercial reasons, Concorde was retired in October 2003.

Figure 7: Concorde.

Douglas DC-10; an aircraft with a troubled start

The DC-10 was the first widebody aircraft of McDonnell-Douglas, built to meet an airline requirement for a three-engine jetliner for medium and long-range routes. Production started in January 1968, it first flew on August 29, 1970 and first deliveries were in 1971.

On 3 March 1974, a Turkish Airlines DC-10 crashed minutes after take-off from Paris' Orly Airport. All 346 people on-board were killed in what was then the worst accident in aviation history. The accident was caused by a cargo door that was not properly locked. During climb, the door blew out and the resulting damage rendered the aircraft uncontrollable. A similar occurrence had happened almost 2 years earlier, but the crew of that flight managed to land the aircraft safely. The resulting accident investigation showed deficiencies on the cargo door design. Retrofits were mandated but the DC-10's reputation had received a dent. Five years later, on 25 May 1979, the left-hand engine and wing pylon of a DC-10 that was taking off from Chicago O'Hare separated during rotation. The left wing was damaged resulting in an asymmetric stall. The aircraft rolled through 90 degrees and crashed just outside the airport perimeter. All 271 occupants were killed. When inspections of other DC-10s revealed more cases of cracks in engine pylons, the FAA suspended the DC-10 type certificate on June 6, 1979. This effectively grounded the DC-10 fleet. Further investigation showed that the DC-10 did in fact meet FAA's certification criteria[26], and the order of suspension was terminated on July 13, 1979 [NTSB 1979]. The Italian national airline Alitalia cancelled six DC-10 orders on August 23 and Egyptair cancelled 4 shortly afterwards. The total sales revenue from these orders would have been $400 million. Alitalia said that its decision was unrelated to safety, but this is predictable given that it already had a fleet of eight DC-10s. Still in 1979, a DC-10 crashed during a sightseeing trip over Antarctica, killing the 257 occupants. While this accident was not related to the airworthiness of the aircraft (it was a controlled flight into terrain type of accident) the DC-10's reputation was now seriously damaged. McDonnell-Douglas' share price fell significantly because of lower anticipated sales [Chalk 1986]. The accident did not however have an effect on the market shares of routes that were flown by the DC-10 [Barnett & LoFaso 1983].

The next highly publicised accident occurred on 19 July 1989. The centre engine of a DC-10 suffered an uncontained failure, damaging the hydraulic system and depriving the crew of the use of the aircraft's control surfaces. The flight crew, in a display of remarkable airmanship, managed to nurse the aircraft to an airport by clever manipulation of the throttles. The aircraft made a crash landing in Sioux City, and while there were 112 fatalities, 184 people survived [NTSB 1990]. After the crash at Sioux City new bookings on the DC-10 plummeted. The impact on consumer behaviour recovered very quickly however and passenger avoidance was down to about 10 % of the pre-crash behaviour 8 weeks after the accident [Barnett et al 1992], see Figure 8.

The year 1989 also saw the end of the DC-10 production run. 386 commercial DC-10s were delivered, plus 60 KC-10 tanker/cargo models built for the U.S. Air Force. The aircraft still flies today, and despite its initial reputation has an accident rate (overall) that is not different from other aircraft of the same generation.

[26] The cracks were caused by inappropriately executed maintenance procedures.

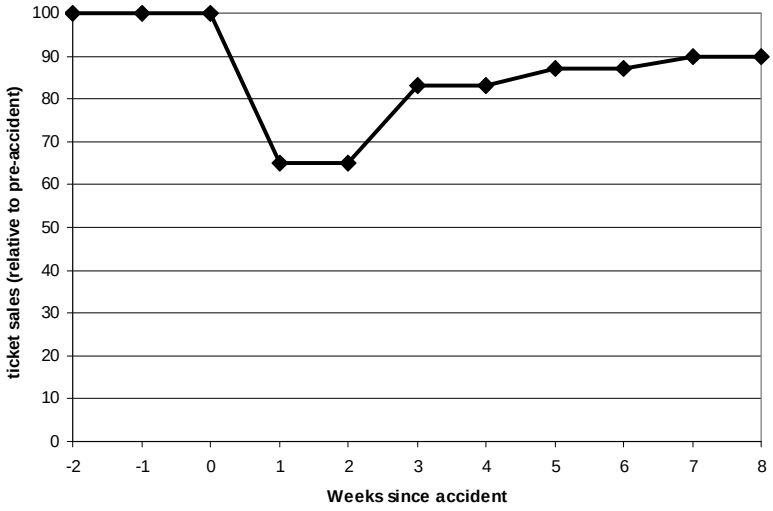


Figure 8: Estimated changes in new DC-10 bookings on competitive routes after 1989 Sioux City DC-10 crash (relative to pre-crash levels). Source: Barnett et al [1992].

4.2. Who are the users?

According to the Dutch Ministry of Transport, the potential users of a causal risk model are airlines, air navigation service providers, airport organisations, maintenance and repair organisations, aircraft manufactures, the central government and the aviation inspectorate. The projected use is not limited to the Netherlands, so the 'central government' could also mean other national governments, the European Commission, or ICAO. However, when some of these organisations were asked directly, many doubted whether a causal risk model could be useful for them. Especially the industry had difficulty to imagine how a causal risk model could support them in taking daily decisions. The model was nevertheless not rejected a priori as useless, the opinion was rather that 'when the causal risk model is ready we will see what we can do with it' [De Jong 2006]. In this thesis, we will thus consider each of the above mentioned organisations as a potential user of a causal risk model.

4.3. Perspectives on aviation safety

This section describes the different perspectives that the stakeholders in the aviation system may have with respect to aviation safety and how this influences their opinion on the need to manage and improve safety. A description of the possible role of a causal risk model is given for every stakeholder based on their needs.

4.3.1. Airlines

For airline management, meeting the minimum safety standards set by the aviation authorities is a necessary requirement for obtaining an Air Operator Certificate (AOC). In addition to that, airline management will be aware of the fact that suffering a major accident may result in significant costs due to decreased ticket sales and may even cause an airline to go bankrupt, even though most airlines have adequate insurance to meet virtually all the direct costs arising from an accident (see infobox on the role of insurance). Especially very small airlines in the holiday charter business are vulnerable in this respect, because they are often not able to survive the immediate loss of income that is the result of one of their aircraft being no longer available and of the effect of the crash on public perception. However, accidents can impact the business of larger airlines as well. Examples

of small airlines that went bankrupt after an accident are Birgenair (Boeing 757 accident in 1996 [JIAA 1996]), Valujet (DC-9 accident in 1996 [NTSB 1997], although Valujet continued operations under a different name; Airtran), Helios Airways (Boeing 737, crashed near Athens in 2005 following flight crew incapacitation due to hypoxia [AAIASB 2006]). But for most airlines the probability of a catastrophic accident is so remote that fear of an accident is not directly a driver for safety improvements. The current catastrophic accident rate for Western operators is in the order of 0.2 accidents per million flights [CAA-NL 2007]. A large European carrier typically conducts 120.000 flights per year[27], which means that on average a large carrier will suffer 1 catastrophic accident in every 42 years. Most managers will thus never be confronted with a serious accident in their entire career. Yet they still invest in safety and want to know about value for money. They also make changes for other reasons and want to know their effect on safety. They may, for example, want to cut costs by stopping certain activities, or reducing staff, without compromising safety.

The role of insurance

Probably because of the potentially catastrophic nature of air accidents, the large amounts of money at risk, and the need to ensure that victims or their relatives get adequately compensated, even if the airline cannot pay, insurance plays a large part in commercial aviation, removing most of the possible direct loss resulting from such accidents. It may be assumed that, as far as airlines are concerned, almost all will have adequate insurance to meet virtually all the direct costs arising from an aircraft accident. Exceptions may arise occasionally with small, perhaps less well established operators in third world countries, although even there it is thought to be rare [Van Dijk et al 2001].

The airline's insurance will respond first to a loss, with other parties becoming involved through subrogation or contribution. Separately, personal insurance, e.g. life insurance, may also be expected to respond [Van Dijk et al 2001].

Aviation accidents usually involve multiple defendants. The most commonly involved defendants are the airline and the aircraft manufacturer but anyone else who may have been involved or contributed in some way to the accident, such as the manufacturer of the engine or any other component or system, the maintenance organisation, the air traffic service provider, etc., will also be included. Unlike some tort litigation with multiple defendants, the defence in air accident litigation is usually highly coordinated and unified in dealing with plaintiffs. The direct insurer of the airline takes the lead in negotiating with the plaintiffs on behalf of all defendants. Defendants usually agree amongst themselves, quite early, on how to share liability, although this may be adjusted later after settlement has been made. Where there is a dispute as to liability, the airline's insurer will often make settlement to plaintiffs in the expectation of gaining partial reimbursement from the other defendants later or on the basis of an interim contract between defendants with the understanding that the amount of each defendant's contribution would be adjusted once liability issues have been resolved.

Insurance companies could perhaps use a causal risk model, e.g. to determine insurance premiums. A requirement is that the model is able to show how changes to the aviation system, either from 'outside', like global economical developments, or from 'inside', like airline policy decisions, influence aircraft crash risk. However, pricing of the airline

[27] KLM data for 2004.

insurance market is strongly influenced by 'spillover' from the general insurance market, the financial markets and overall economy [Swiss Re 1996]. This may have a much bigger impact than the (expected) crash risk and would thus limit the usebility of a causal risk model.

Airlines monitor safety performance of the fleet by routinely analysing in-flight recorded data and checking parameter values for crossings of pre-defined threshold values. Individual threshold crossings are analysed to determine exactly what happened (see also section 8.5.4). In addition to this flight data program, most airlines also have a system in place for the flight crew to report abnormal situations. This information is usually in a free-text format. The combination of flight data and crew reports provide the airline with a good picture of individual incidents, but as of yet a tool that allows analysis on a more generic level does not exist. At best airlines do some sort of trend analysis. A causal risk model would be very valuable to airlines if it could support providing an overall safety picture and, most importantly, show airline management the influence of their decisions on the level of safety. It is therefore required that a causal risk model links-up with the primary safety data information systems: flight data and safety reports.

A challenge for airlines is to maintain current (safety) standards in a changing world. When such changes are local and threaten to influence the 'level playing field' (see section 4.3.6) or their competitiveness, airline management will be keen on taking proper countermeasures to ensure profitability of the company. When locally imposed safety measures threaten the level playing field, the reaction will be to resist them until and unless all others are required to implement them also. Competitiveness is the driving factor. A causal risk model could play a role in analysing future changes, estimating their effect on safety and providing information that allows selection of proper countermeasures. For airlines therefore the ability to perform cost benefit analysis of proposed future changes would be a desirable feature for a causal risk model. This implies also the need for a model in which the way that future changes in how functions are operationalised can be incorporated. Because of the low accident frequency, a causal risk model for airlines should be based on safety indicators other than the accident rate.

4.3.2. Repair stations

Maintenance technicians consider aircraft safety to be the result of their professional skills. They feel responsible for safety. In the process of releasing an aircraft back into service, some maintenance technicians even think that they sign off for the safety of the aircraft rather than (as required by regulation) for having followed prescribed procedures [Biemans et al 1998]. These procedures are specified in the operator's maintenance program, which is normally based on instructions for continued airworthiness prepared by the manufacturer. The operator may rewrite the structure and format of these maintenance recommendations to better suit his operation. Once aircraft enter service, the initial maintenance program is subject to continuous development and update as modifications, product improvements and operational feedback are incorporated. To evaluate the effectiveness of the maintenance program and to update it, operators develop a reliability program. The actions resulting from the reliability program may be to escalate, de-escalate, omit or add maintenance tasks. By proving to the authority for instance that increasing a servicing or inspection interval for a particular component does not adversely affect safety, the operator could save money in maintenance expenditures. The Alaska Airlines MD-83 accident in 2000 (see infobox) is an example of a case where inspection intervals were extended without the operator (Alaska Airlines) or the authority (FAA) being sufficiently aware of the safety implications.

Throughout the service life of an aircraft type, the authorities continue to monitor the safety of the design through a service difficulty reporting and analysis system. If at any time during the service life of an aircraft, the airworthiness authorities determine that an unsafe condition exists in a product (aircraft or system) and that the unsafe condition is likely to exist or develop in other products of the same type design, they may issue an Airworthiness Directive (AD). The purpose of an AD is to correct or prevent an unsafe condition. An AD gives a description of required corrective action or operating limitations for a particular aircraft or system. Compliance to ADs is compulsory.

Service Bulletins (SBs) are issued by the aircraft, component, or engine manufacturers to update operators with information relevant to a particular aircraft type. Maintenance service bulletins may advise to inspect, repair, rework or modify the aircraft (see infobox Boeing Service Bulletin 737-56-1019). While some service bulletins address safety or airworthiness related problems, other service bulletins relate solely to operational or economic matters. A service bulletin may be prompted by problems identified through in-service experience or production, and may introduce product improvements or changes to operational requirements. It is up to the operator to apply or ignore the SB. This requires a management process in the company to monitor these, to anticipate them and to respond with good decisions on them. A causal risk model would be useful if it could calculate the expected safety implications of each SB.

Alaska Air accident

Alaska Airlines began operating McDonnell-Douglas MD-80 aircraft in 1984. The MD-80 is a derivative of the DC-9 series aircraft. The aircraft is equipped with a 'T-tail'; the horizontal stabilizer is mounted on top of the vertical stabilizer, they are connected by two hinges at the aft spar of the horizontal stabilizer and with a single jackscrew assembly at the front spar of the stabilizer. Initially, that jackscrew was lubricated every other B check (every 700 flights hours), but in the course of the years this interval had been increased by Alaska Airlines to 2,550 flight hours (see Figure 9).

On January 31, 2000, Alaska Airlines Flight 261, a McDonnell Douglas MD-83 crashed into the Pacific Ocean. The 2 pilots, 3 cabin crewmembers and 83 passengers on board were killed and the aircraft was destroyed by impact forces. The accident was caused by a loss of aircraft pitch control resulting from a failure of the horizontal stabilizer jackscrew assembly. The failure was caused by Alaska Airline's insufficient lubrication of the jackscrew assembly. Contributing to the Alaska Air MD-83 accident in 2000 was the FAA's approval of Alaska Airlines' extended interval for lubrication and check interval of the horizontal stabiliser trim system jackscrew assembly [NTSB 2002a].

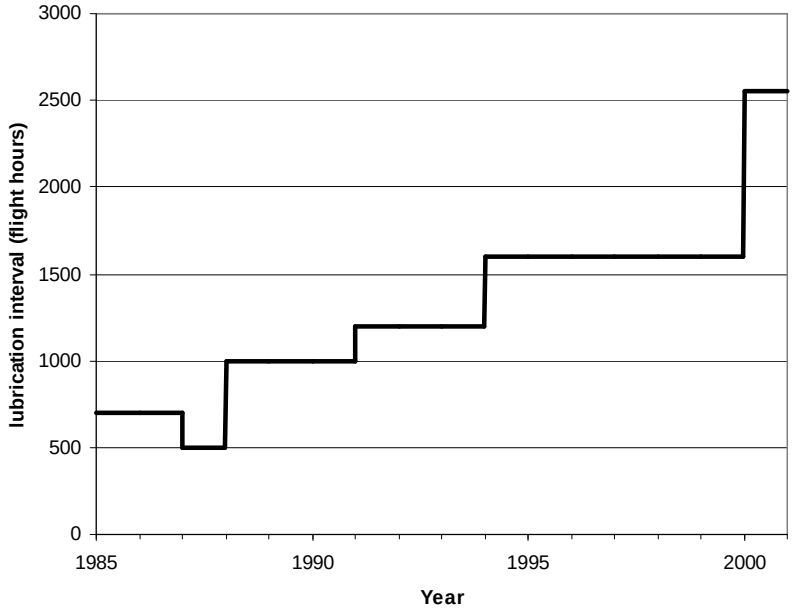


Figure 9: Alaska Airlines horizontal stabilizer jackscrew lubrication interval escalation.

The maintenance organisation will be constantly looking for more cost-effectiveness in the reliability program. The reliability program requires approval from the authority and therefore one of the main drivers for safety is the need to demonstrate that the proposed changes will not adversely affect safety. The Alaska Air accident revealed that the process by which airlines make changes to the maintenance program and demonstrate to the authority that these changes will not present any potential hazards requires improvement [NTSB 2002a]. A causal risk model could help improve this process. This application would require a causal risk model that captures very specific (specific to aircraft type or even tail number) and detailed (up to aircraft component level) information.

Boeing Service Bulletin 737-56-1019
Boeing originally designed the Boeing 737 jet aircraft with cockpit eyebrow windows in order for the flight crew to maintain a sufficient and unobstructed view during turns. According to Boeing, today's advanced navigation and air traffic control systems have made these windows obsolete and from 2005 onwards the eyebrow windows were deleted on newly produced 737 aircraft. This design change has reduced aircraft weight by 20 pounds and eliminates approximately 300 hours of periodic inspections per aircraft. Service Bulletin 737-36-1019 allows replacing the eyebrow windows of the Boeing 737-300 and -400 aircraft (manufactured before 2005) with an aluminium plug [KLM 2007].

4.3.3. Aircraft manufacturer

For aircraft manufacturers, meeting the minimum safety requirements as specified in FAA FAR 25 and EASA CS 25 is necessary to obtain a type approval for the aircraft. FAR 25.571 and CS 25.571 for instance specify requirements for damage tolerance and fatigue evaluation of the structure, FAR 25.603 and CS 25.603 specify requirements for materials. In particular, FAR 25.1309 and CS 25.1309 specify requirements for aircraft system design and analysis. These requirements for aircraft systems are based on the principle stating that

an inverse relation should exist between the probability of malfunctions and the degree of hazard to the aircraft and its occupants. Depending on the criticality of the system and the failure, the applicant must demonstrate a failure probability of less than 10^{-5}, 10^{-7} or 10^{-9} per flight hour. A detailed Failure Modes and Effects Analysis (FMEA) and Fault Tree Analysis are often necessary to demonstrate compliance with these regulations.

Fokker F-10; end of the wooden wing

The F-10 was an aircraft designed and built by de Fokker Aircraft Company in the US which entered service in 1928. The aircraft was of typical Fokker construction, combining a wooden wing and fabric covered steel-tube fuselage. Initially the aircraft was a commercial success, despite it being of rather outdated design; all-metal aircraft like the Douglas DC-2 were showing to be faster and more efficient, having 20% less operating costs. But on March 31, 1931 a Fokker F-10 crashed during a thunderstorm, killing all on-board including Knute Rockne who was the celebrated coach of the popular 'Notre Dame' football team. The accident was headline news for a week. Initial investigation immediately revealed a broken wing and the emerging Bureau of Air Commerce issued a directive that called for frequent and expensive inspections of all Fokker wings, despite furious attempts by Anthony Fokker to prove that bad weather had caused the accident. Fokker aircraft now had a reputation of being inherently dangerous. In combination with the Great Depression and Fokker stubbornly maintaining the wooden wing concept this was the end of Fokker in America. On July 10, 1931, Anthony Fokker was forced to resign from the company that bore his name, and he soon returned to the Netherlands [De Leeuw 1994, Dierikx 1997, Schatzberg 1994].

The manufacturers are in a position to directly influence aviation safety by introducing new technology. As we have seen in the previous sections, technological improvements have been fundamental in improving aviation safety. Any new technology can only be introduced if the manufacturers can demonstrate to the certifying authority that it is sufficiently safe. Currently aircraft manufacturers use analysis and physical testing for this. In the case of aircraft systems, the analysis is conducted with the aid of fault tree analysis and comparable methods. Aircraft manufacturers could use a causal risk model during the type certification process in a similar way as they currently use fault tree analysis and comparable methods. The objective is to demonstrate in the system safety assessment that all aircraft systems comply with relevant regulation. Increased integration of aircraft systems and integration of airborne and ground-based systems lead to a need for methods that can properly represent these integrated systems, including human operators, in a safety assessment. Fault tree methods are no longer sufficient. For this application the causal risk model must be able to properly represent the role of the human operator and be able to capture dependencies between different system components. Using the model as part of the certification process requires that the model be validated and accepted by the regulators. This also makes it necessary that the model produce similar results if applied by different users.

History provides some examples of aircraft that allegedly suffered from a bad safety reputation; see for instance the story of the Fokker F-10 and the Douglas DC-10. However, aircraft purchase decisions are complex, political and confidential, and therefore do not easily allow a clear statement as to whether public mistrust had an impact on a company's purchase decision. From that perspective, there is no direct driver for the aircraft manufacturer to develop aircraft that are safer than others.

Aircraft manufacturers are also working on safety because of product liability. Liability claims, in some cases even extended to personal accusations of manslaughter (see infobox: Aircraft design and manslaughter), can be a reason for manufacturers to equip their aircraft with the best available technology. Contrary to the situation at airlines, for which a crash is only a remote probability, aircraft manufacturers are regularly confronted with accidents involving their products. From 1986 (the year of the first flight of the Fokker 100) to 2006 for instance there have been at least 108 accidents involving Fokker aircraft - 6 Fokker 100, 1 Fokker 70, 5 Fokker 50, 23 F-28 and 72 Fokker F-27 - of which 51 accidents caused fatalities (Source: NLR Air Safety database). Because of this exposure the safety awareness of manufacturers is high.

Aircraft design and manslaughter

An Airbus executive went on trial in May 2006 for manslaughter over the 1992 crash of an Airbus A320 of France's domestic airline Air Inter in eastern France that killed 87 people. Investigators considered the argument that the cockpit was ergonomically flawed, and that the pilots were confused between the controls for the aircraft's descent. Airbus' then engineering chief Bernard Ziegler was indicted over the ergonomic design of the A320 cockpit, as well as the possibility that the aircraft's navigation system was faulty [Flight International 2006a, 2006b]. The French authorities also launched a manslaughter case against the former head of the Concorde programme in connection with the Air France Concorde accident in 2000 near Charles de Gaulle airport that killed 113 people. Henri Perrier, chief engineer on Concorde's first test flight in 1969, was placed under investigation for manslaughter and involuntary injury as part of an investigation into the crash [Bentley 2005].

An aircraft manufacturer could use a causal risk model to anticipate and respond with suitable priority to potential accident scenarios. This would require a model that is able to represent current as well as possible future accident scenarios.

4.3.4. Air navigation service provider

The air navigation service provider (ANSP) is responsible for ensuring that air traffic proceeds safely and efficiently while minimising the burden to the environment [LVNL 2001a]. Decisions taken by the ANSPs can have direct consequences for safety, efficiency and the environment, and accordingly affect many actors. Often, decisions that positively affect one area will have a negative effect on another; for example accepting higher crosswinds (and consequently higher risks) during the landing approach so that a runway can be used that is preferred for reasons of noise abatement. Therefore all actors demand transparency of the decision making process at the ANSP. ANSPs are in need of a way to estimate the effects of proposed changes (human, procedural, hardware and software elements of the ATM System as well as its environment of operations) in such a way that allows comparison of those effects. Currently, effects on capacity and the environment can be predicted reasonably well, but safety still is a problem. As a result, decisions are sometimes made which, with the benefit of hindsight, were unwise. See for instance the case of the Crossair BAE 146 accident at Zurich, where the aircraft was forced to fly a more risky non-precision approach instead of a precision approach because of noise abatement procedures.

Safety versus the environment: the Crossair BAe 146 accident at Zurich [AAIB Switzerland 2004].

On 24 November 2001, a Crossair BAe 146-RJ100 aircraft took-off from Berlin-Tegel airport for a scheduled flight to Zurich. After an uneventful flight, the aircraft received the clearance for the approach to runway 28 at Zurich airport. The weather was close to the minimum for this runway. When the aircraft reached the minimum descent altitude (MDA) of 2,390 ft, the captain mentioned to the first officer that he had visual ground contact and continued the descent. Shortly afterwards the aircraft crashed into the ground. Twenty-one passengers and three crew members died from their injuries at the site of the accident; seven passengers and two crew members survived the accident.

Zurich airport is characterised by a system of three runways. Runways 16 and 14 are equipped with a Category CAT III instrument landing system (ILS) and are therefore suitable for precision approaches. Runway 28 only allows non-precision VOR/DME approaches. The approach sectors of runways 14 and 16 are equipped with a minimum safe altitude warning system (MSAW). This system triggers a visual and acoustic alarm in air traffic control if aircraft infringe defined minimum altitudes. No MSAW is installed in the runway 28 approach sector. At the time of the accident the noise abatement procedures in force for Zurich airport played a decisive role in determining take-off and landing runways. Approaches from 22:00 local time (LT) to 06:00 LT had to take place on Runway 28.

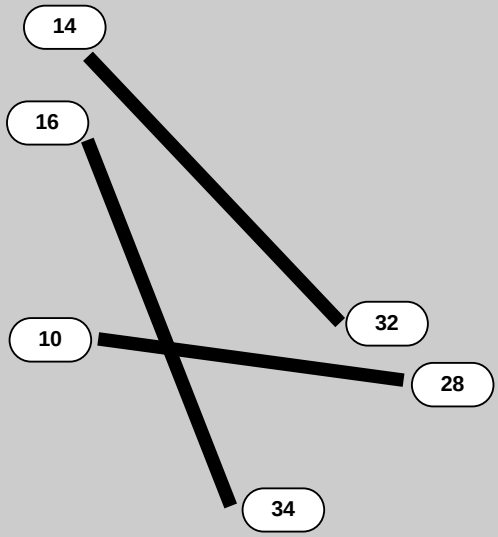


Figure 10: Zurich airport runway layout.

Because of noise abatement procedures, the aircraft had to carry out a non-precision approach, rather than the precision approach that is the norm when noise abatement is not in place. The accident risk during non-precision approaches is on average 5 times higher than the risk during precision approaches [Enders et al 1996]. The absence of the MSAW for this approach was another missing safety barrier. In this particular case, apparently it was accepted that, for the interest of noise abatement, aircraft had to fly a non-precision approach in marginal weather conditions, even though a much safer precision approach was available. In this case, the non-precision approach and the lack of MSAW were not the *causa causans* of the accident, but rather the *causa sine qua non*.

Safety assessments for ANSPs are often driven by the need for capacity increase on one hand whereas it is not clear how to achieve this without jeopardizing safety criteria (such as for instance ICAO's TLS for mid-air collision risk). A quantitative risk analysis when introducing changes to the ATM system is also required by Eurocontrol under ESARR 4. A causal risk model could fulfil Eurocontrol's requirement for a quantitative risk-based approach in Air Traffic Management when introducing and/or planning changes to the ATM system. If the model were to properly represent the influence of factors like the contribution of human performance or safety management systems and their interactions, some of the current ambiguity in ESARR 4 would be resolved. A causal risk model representing the total aviation system, i.e. not limited to the ATM component, would also remove a weak point of Eurocontrol's current approach by fully taking account of integration of the different elements in the aviation system. This requires careful representation of dependencies in the model. Note that the requirement comes from the regulatory role of Eurocontrol and not from Eurocontrol as an air traffic service provider.

Eurocontrol is very well aware of the potential benefits of a causal risk model, and when the High Level European Action Group for ATM Safety (AGAS) identified priority actions to improve safety in European Airspace, it included research to develop an 'integrated risk picture' for ATM in Europe. The purpose of such an integrated risk picture is to show the contribution of ATM in both causing and preventing aircraft accidents, and the safety impacts of changes in ATM as it develops in the future [Perrin et al 2006, Kirwan 2007]. The integrated risk picture is the output of a risk model which uses separate 'causal models' for each accident category, constructed as fault trees that represent subsequent barrier failures.

From these considerations we can derive the following requirements from ANSPs for a causal risk model:

- Able to represent safety effects of changes in human, procedural, hardware and software elements of the ATM System as well as its environment of operations.
- Able to represent the contribution of ATM in both causation and prevention of accidents.
- Representation of the total aviation system, including dependencies between system elements.
- Provide quantitative output.
- Able to show the safety impacts of ATM as it develops in the future.

4.3.5. Airports

The airport facilitates the primary aviation processes, and the direct influence of the airport on aviation safety is via the airside facilities that it provides: the runways, taxiways and aprons, including lighting and marking, de-icing facilities and emergency response. To be able to control risk, Amsterdam Airport Schiphol has expressed the need to obtain insight into safety risks that are associated with the airside processes, in particular, the following is required [AAS 2001]:

- Insight into cause and effect of safety risks,
- Insight into which risks are already sufficiently covered by safety measures, and those risks for which adequate safety measures are lacking,
- Insight into the effectiveness of safety measures for the management of risk,
- Insight into the probability and the effect of safety risks.

A causal risk model could support the airport and its safety management if the model is able to meet these requirements.

4.3.6. Policy makers and regulatory bodies

The Netherlands

The starting point for policy with respect to aviation safety transport in the Netherlands is the notion that ease of access to the Netherlands by air is an important factor for the national economy [RLD 1996, DGTL 2005]. A safe aviation system is a prerequisite to good air connections. The national government supports the accessibility of the Netherlands by setting certain boundary conditions. Simultaneously, there is an objective from the government to be more efficient, with a less complex internal structure and mode of operation. This means that industry and Dutch citizens themselves are expected to do more. They should also be provided with the means to do so. Government should only provide safeguards when other actors are insufficiently capable to take care of public interest. The Aviation Inspectorate (IVW) is tasked to (re)activate the industry's responsibility by means of its safety oversight activities. The increased role of citizens requires from the government as well as from the industry an open attitude and a duty to inform the citizen on relevant aspects of aviation. It also means that the Ministry has decided to use safety perception[28] by Dutch citizens as a performance indicator of the aviation safety policy in the Netherlands [DGTL 2005].

The vision of the Ministry of Transport and Water Management concerning aviation safety consists of four pillars [DGTL 2005]:

1. Continuous improvement of safety
2. Explicit and transparent decision making
3. Preparedness for inevitable risks
4. Development and maintenance of safety management and safety culture within the Ministry.

The Ministry understands that aviation and aviation safety is largely governed by international agreements and cooperation. An important starting point for the Netherlands is to be able to operate on a level playing field (see also under that heading in this section).

The development of a causal model of aviation safety is one of the primary topics of the Ministry's Safety Policy agenda. According to the policy implementation plan, a causal risk model should be used to identify weak spots in the aviation processes and assess their effect on overall safety (see also Appendix C).

The use of a causal risk model as an enforcement tool was also mentioned in the policy documents. This would be a similar use as the way in which the statistical models for third party risk around airports and noise calculations are being used. This intended use has caused some panic amongst the industry, primarily because they see the model as potentially restrictive. There is quite some suspicion, or even paranoia, of the industry towards the authorities in this respect because they already feel severely restricted by all kinds of local regulations. The fact that the authorities initially failed to explain precisely what a causal risk model is and how it could be used contributed to this situation (see also Appendix A on the history of third party risk regulation for Schiphol airport). Regulation on third party risk at Schiphol is a complex and sometimes controversial issue. A seemingly

[28] Safety perception is considered to be outside the scope of the causal risk models of this thesis, see section 2.2.

unavoidable gap between (legal) safety standards and safety perception by the public further complicates the issue. This is particularly true for third party aviation risk, where safety standards are based on a combination of tradition, professional risk estimates, estimates of accident probabilities, expected accident locations and accident consequences. These are combined into a single artificial number that represents 'risk' for a single year. This artificial number (e.g. individual risk or group risk) does not allow straightforward comparison with other expressions for fatality risk. A causal risk model has been proposed as a panacea for those issues, but this initially created even more difficulties and misunderstandings. These have been resolved, and the Ministry of Transport and Water Management now argues that causal risk models can be used to quantify improvements of safety on the accident probability [Tweede Kamer 2003b]. It is then possible to use this accident probability for the calculation of third party safety levels and compare those with the target. In doing so, the causal risk model provides directions to the sector for the improvement of safety; it becomes a tool for risk informed decision making and can help the industry to capitalize on measures they have already taken. A requirement for the causal risk model therefore is the ability to estimate the effect of local, airport specific (safety) measures and local airport characteristics (like e.g. weather and surrounding terrain, type of operations, etc.) on aircraft accident probability.

What the Ministry should also be wanting is a model that helps to develop and implement a national safety policy and the associated safety oversight activities. Setting and implementing safety policies requires the ability to make a risk inventory based on projected trends and changes, and to monitor safety from safety performance indicators. It also requires a way to estimate the effect of proposed safety measures on the level of safety.

In a 'vision document' on safety oversight [IVW 2007], the Dutch Transport and Water Management Inspectorate laid down the framework for future safety oversight of the 'system Schiphol'. The system Schiphol is described as 'all actors who directly or indirectly influence safety, economy or the environment within the airport processes, including interactions between actors'. Key points of this framework are:

- Safety oversight of the system Schiphol should be integrated,
- Special attention should be given to interfaces between actors,
- Priority will be given to those areas where the risks are high and deficiencies in compliance are numerous.

To facilitate safety oversight as described in the vision document, an analysis tool is required that identifies and prioritises risks on the basis of accident and incident information, (future) threats and compliance information for the entire system Schiphol. A causal risk model could fulfil this need if it meets the following requirements:

- System wide representation, including interactions.
- Able to utilise compliance information. This requires a link to safety regulation.
- Able to use accident/incident information, specifically ECCAIRS as this is the main incident database system used by the European aviation authorities.
- Able to rank risks (according to a combination of probability and severity of the occurrence).
- Able to represent emerging threats.

United States

In the United States, the Federal Aviation Administration (FAA) is responsible for certifying air carriers, commercial operators, air agencies and airmen in order to ensure

adequate levels of safety in the civil aviation system. These activities are conducted by the Flight Standards Services (AFS) organization. Within AFS, the Certification and Surveillance Division (AFS-900) is responsible for safety oversight activities, including certification, surveillance, and investigation. The activities are collectively termed the oversight system. Like other regulatory agencies, FAA and AFS face significant challenges in fulfilling their safety oversight mission. They must contend with a disparity between the magnitude of activities they are mandated to oversee and the limited resources available for oversight. Additionally, the activities involved are highly specialized and the relationships between individual activities and system safety are very complex. The very success of the aviation system in terms of safety performance makes it difficult to determine the linkage between oversight activities and safety outcomes. The FAA recognized that a system safety approach is required to guide the use of limited resources in order to increase the efficiency and effectiveness of the oversight activities and therefore initiated the Systems Approach to Safety Oversight (SASO) initiative. Such a system safety approach requires a comprehensive aviation safety performance evaluation methodology including a safety evaluation construct, safety hazard identification and risk analysis, safety performance measures and risk indicators, safety diagnostic procedures, risk management decision support tools and oversight evaluation methodology [FAA 2004]. According to SASO's research requirements, the safety hazard identification and risk analysis requires a methodology to provide a foundation for conducting risk analysis. That methodology should include basic definitions such as cause, hazard, failure, their relationships and types, and the causal pathway. At a minimum, the method must allow a probabilistic evaluation of risk associated with hazards in the area of operational functions and procedures, but, if possible, hazards and risk in other areas of the aviation system such as aircraft and equipment malfunctions, hazards due to organisational set-up, human errors, environmental hazards (wind, turbulence, terrain) and contributory hazards (regulatory, economy) should be included. A specific requirement is to have a top-level representation as the user interface for performing risk analysis [FAA 2004]. Validation of the methods is proposed by empirical testing using historical data and secondly by a panel of subject matter experts from the FAA and industry.

When comparing these FAA requirements to those from the Dutch Ministry of Transport for a causal risk model it is apparent that FAA's requirements are more detailed and the role a causal model in the regulatory process is better described. So for the Dutch Ministry of Transport the model developers will have to describe the requirements for them.

Europe
The European Commission's vision on European aeronautics for 2020 asks for a five fold reduction in the average accident rate of global operators (relative to 2001). According to the same vision, safety research challenges to achieve that goal include the following [Group of Personalities 2001]:

- Flight hazard protection
- Advanced avionics
- Probability and risk analysis
- Computational methods
- Human error checking systems

The document, high level as it is, fails to provide useful details on what is exactly meant under the heading probability and risk analysis, but at least it shows that it is recognised by

the European Commission as a key research area. This is further demonstrated by the fact that the EC has funded several research initiatives on causal aviation risk modelling, most notably the projects DESIRE, ASTER and ASICBA[29] in the fourth, fifth and sixth framework programme respectively [Piers et al 2006].

The EC funded project 'ASTER' also attracted the attention of the JAA because they were at that time looking into ways to perform regulatory impact assessments. JAA required that "All safety impacts of the option(s) being evaluated should be identified and wherever possible and necessary be quantified e.g. in terms of incident/accident probability and severity. This should, if appropriate, include Human Factors and Operational aspects" [Morier 2002]. With the transition of tasks from JAA to EASA this has now become an EASA requirement. For a causal risk model to be able to assist in a regulatory impact assessment it needs to be able to assess the effect of regulatory changes on accident risk.

Eurocontrol and the European Commission launched the Single European Sky (SES) ATM program in 2006. SESAR, the technological part of SES, aims to achieve a high-performance European air traffic control infrastructure by the year 2020 which will enable the safe, environmental friendly and sustainable development of air transport. Part of the safety approach in SESAR is a top-down approach to show compliance with a TLS by apportioning the overall safety targets for individual components. According to Eurocontrol, to apportion the TLS in a systematic way, a fundamental requirement is a quantitative model of the causes of accident risks. This risk model must quantify the causal relationship between the safety measure used in the TLS and the system elements that require apportioned targets [Eurocontrol 2007].

Level playing field

In air transport, when locally new rules are announced or implemented, response from industry is often that those local rules distort the 'level playing field'. The notion of a 'level playing field' is used to indicate that an organization (be it an airline, airport organization, or other) does not enjoy advantages or suffer disadvantages as a result of locally prevailing regulation compared to rivals under other local regulation. This seems to imply that the existence of a 'level playing field' is fully determined by aspects that are controlled by the authorities; rules and enforcement. However, the competitiveness of an organization is not only influenced by rules and regulations but also by factors such as the geographical and demographical location, environmental conditions such as prevailing weather etc. This raises some fundamental issues. One could imagine that in a country where the environmental conditions are unfavourable for a particular activity, the authorities provide compensation by subsidizing, which then can lead to a certain controversy. A neighbouring country could claim that the subsidies bring about a distortion of competition, while the home country could claim that the subsidies just do the opposite and result in a balanced competition.

Assessing whether a level playing field exists in aviation is an extremely complex endeavour, not only because aviation itself is a complex and highly distributed activity, but

[29] The objectives of these projects was to develop a method to set and optimise safety targets to achieve the optimum level of safety for the aviation system as a whole by distributing the safety burden more evenly among the different stakeholders, and the development of a method for cost benefit analysis to assess safety benefits of any change, including changes in legislation and rulemaking, in relation to the costs of implementing those changes.

also because of the large amount of- and differences in rules and regulations and the way in which they are applied in different countries. A good example of the way in which different rules regarding norms for airport noise levels influence airport operations and associated costs is provided by the Netherlands Environmental Assessment Agency [MNP 2005]. However this information is still insufficient to show whether there exists a level playing with respect to this subject. A causal risk model could possibly provide a substantiated answer to questions or alleged claims regarding the level playing field, but this would require the model to have some sort of representation of current national and international regulation and also allow for national differences in climate and geography.

Quantitative safety requirements in aviation safety regulation

The majority of aviation regulations are qualitative descriptions of the conditions that have to be satisfied to obtain a license or certificate to build, fly, operate and maintain aircraft, provide air navigation services and design, operate and maintain infrastructures such as airports. Aviation safety regulation contains sometimes also quantitative safety requirements. By considering these requirements, the targets that are set, the methods used and the way in which compliance can be demonstrated, we obtain some idea of the user requirements that one might expect for a causal risk model.

The most influential are ICAO's collision risk targets, FAA/JAA/EASA's targets for aircraft system failures, and Eurocontrol's target for 'ATM directly contributing to aircraft accidents'. Targets for aircraft system failures have already been described in section 4.3.3, the others will be discussed below. Compliance with mandatory targets is commonly shown by analysis, laboratory tests or flight testing. FAA/JAA/EASA and Eurocontrol publish advisory material on acceptable means of compliance. For aircraft system safety assessments, the advisory materials on FAR 25.1309 and EASA CS 25.1309 list various methods for assessing the causes, severity, and probability of failure conditions. The list includes Functional Hazard Assessment, Failure Modes and Effects Analysis, Fault Tree Analysis, Markov Analysis, and Common Cause Analysis. Eurocontrol's guidance material on risk assessment and mitigation in ATM [Eurocontrol 2004] is intended to present the acceptable means of compliance recognised by Eurocontrol's Safety Regulation Commission as possible harmonised ways to meet ESARR 4 provisions. The document presents 3 different methods, but, interestingly enough, also states that none of those methods is considered to be fully compliant.

ICAO collision risk requirement

ICAO prescribes a TLS of 5×10^{-9} fatal accidents due to collisions per flight hour per dimension i.e, lateral, longitudinal and vertical [ICAO 2002c]. The basis for assessing ICAO's mid-air collision safety targets is a collision risk model developed by Reich in 1966 [Reich 1966a, 1966b, 1966c]. It is partially based on real and simulated data of observed error rates. According to Reich, the collected data shows that 'flying errors'[30] do not obey the Gaussian distribution and that no simple theoretical distribution adequately describes all the observed data. Four decades later, the need for real data rather than 'estimates' when it comes to probabilities of events that are dominated by human performance continues to be relevant [Brooker 2004].

[30] Imperfections in navigation and piloting.

For ILS approaches[31], the ICAO Obstacle Clearance Panel developed a collision risk model in the form of a computer programme that calculates the probability of collision with obstacles by an aircraft on an ILS approach and possible subsequent missed approach. The collision risk model was developed as a result of an extensive data collection programme followed by detailed mathematical analysis [ICAO 1980]. The data collection was necessary to be able to predict probabilities of events that are dominated by human performance, such as the spread of aircraft about the nominal path during an ILS approach.

In conclusion, the ICAO collision risk requirements are expressed as a target of fatal accidents per flight hour. ICAO prescribes the way in which the calculation must be conducted to demonstrate that the target is met. ICAO is indeed an appropriate body to standardise and prescribe a risk model for a particular application. It would therefore be useful to inform and involve ICAO in the development of a causal risk model.

Eurocontrol requirement for ATM risk

Eurocontrol Safety Regulatory Requirement (ESARR) 4 'Risk Assessment and Mitigation in ATM' [Eurocontrol 2001b] concerns the use of a quantitative risk-based approach in Air Traffic Management when introducing and/or planning changes to the ATM system. This requirement covers the human, procedural and equipment (hardware, software) elements of the ATM system as well as its environment of operations. The requirement is consistent with amendment 40 to ICAO Annex 11, mandating the use of safety assessment of significant changes in ATS. According to ESARR 4, the maximum tolerable probability of ATM directly contributing to an accident is 1.55×10^{-8} per flight hour, or 2.31×10^{-8} accidents per flight. Professional judgement was used to determine the maximum acceptable ATM direct contribution to accidents; and 20 years of historic data were selected to confirm the credibility of this target. The explanatory material on ESARR 4 requirements states [Eurocontrol 2003]:

> "The feasibility of setting quantitative objectives / targets for specific parts of the ATM system was discussed, especially when it comes to their allocation to human contributions, procedures and software. It is recognised that demonstration of compliance won't always be quantitatively – based, as it does not seem feasible to demonstrate a priori and in a quantified manner that a good working process, such as training, Safety Management System, or software codes of practices, enable specific quantitative objectives to be met. This will only be based on professional judgement and potentially verified over time."

This is an illustration of the dilemma between the objective, which implies the need of some sort of causal risk model, and the belief in the possibility of constructing one to meet the objective.

The increasing integration, automation and complexity of the ATM system requires a systematic and structured approach to risk assessment and mitigation, including hazard identification, as well as the use of predictive and monitoring techniques to assist in these processes [Eurocontrol 2003]. ESARR explanatory material recognises that a combination of quantitative (e.g. mathematical model, statistical analysis) and qualitative (e.g. good working processes, professional judgement) arguments may be used to provide a good enough level of assurance that all identified safety objectives and requirements have been

[31] An ILS approach is an approach to landing that is conducted with the aid of a ground navigation system ILS: the Instrument Landing System. ILS allows aircraft to land under weather conditions where the pilots cannot see the runway, like low-hanging clouds or fog.

met. The explanatory material emphasises the development of safety monitoring and data collection mechanisms. The rationale is that any model (e.g. Collision Risk Model) used in risk assessment must be validated, and real data could contribute significantly to this validation process.

Eurocontrol provides the following rationale for a quantified risk based approach in Air Traffic Management when introducing and/or planning changes to the ATM system [Eurocontrol 2003]:

- A risk based approach is considered as meeting the objectives of Eurocontrol's Safety Regulation Committee (SRC) in assessing and controlling risk related to changes in ATM.
- A risk based approach is the most commonly used in aviation as well as other safety critical industries such as chemical and nuclear.
- There is an added value in using quantitative objectives, as it avoids diverging understanding by states on the range of frequencies of occurrences included in a risk classification scheme,
- Quantitative criteria provide a clear target and, when derived and applied to lower level events, allow manufacturers to design equipment without having to analyse the entire ATM system.
- The safety performance measurement of actual safety occurrences (with ATM contribution) may enable the verification a posteriori of whether or not quantitative objectives allocated to specific failures in a 'fault tree' for example are being met. Indeed, quantitative objectives could be derived from ATM safety minima and allocated to specific failures in a 'fault tree', hence providing for a range of probabilities which can be reasonably assessed over a short or medium period of time.
- In addition to the measurement of accidents and serious incidents with associated ATM contributions, such a feed back from experience on events of a lower level would enable the measurement of the effectiveness of the qualitative processes selected for software, human, or procedures and the identification of related appropriate "good working practices" to be used in ATM as acceptable measures to show compliance with quantified objectives.

Eurocontrol has developed an ANS Safety Assessment Methodology (SAM) to support the demonstration that safety is being managed within safety levels which as a minimum meet those approved by the designated authority [Eurocontrol 2006a]. The SAM methodology is based upon the method used for certification of civil aircraft systems and equipment and follows three main steps: Functional Hazard Assessment, Preliminary System Safety Assessment and System Safety Assessment. SAM suggests the use of event trees and fault trees to create a bow-tie with the hazards, as identified in the functional hazard assessment, in the centre. The SAM considers the three types of system elements; people, equipment and procedures and their interactions (within the system and with the environment) in a specific environment of operation, but does not address organisational and management aspects. The ability to represent managerial and organisational factors therefore is a requirement for a causal risk model in order to help resolving this omission in the Eurocontrol SAM.

4.3.7. Passengers

For passengers safety is a very personal thing; safety is *their* safety. From their perspective, the aviation system is the airline. People are in general only superficially aware of the role of the other actors in the system, and demand that the airline will take care of their safety.

Passengers, in particular those that travel for leisure, have some, but limited, control over two variables that may influence their safety: the destination and the airline. Yet very few people will consider flight safety when they select their holiday destination. Some studies have tried to determine whether aircraft accidents really influence people's willingness to travel by air, but significant, long lasting effects have not been demonstrated, see also Figure 8 [Barnett & LoFaso 1983, Barnett et al 1992]. This is not to say that passengers are ignorant of aviation safety. The case of Onur Air (see infobox) illustrates that passengers indeed would like to see an 'airline safety rating' similar to for instance 'safety stars' that are awarded to cars in EuroNCAP crash tests. These safety ratings are even used by car manufacturers for sales promotion.

Onur Air
On 12 May 2005, the Turkish airline Onur air was banned from Dutch airspace for a period of two weeks after it had failed on a number of SAFA inspections. Initially the ban was intended to last for a period of four weeks, but when Onur Air had shown improvements, and after considerable political pressure from the Turkish government, the ban was lifted after two weeks. In the meantime, the Dutch Inspectorate received numerous calls (up to 800 per day) from worried passengers who had questions about the safety of air travel in general, and airlines from Turkey in particular [Tweede Kamer 2005].

In reaction to the accident in Puerto Plata (Dominican Republic) in 1996 which mainly involved German tourists on board a Turkish charter aircraft [JIAA 1996], and following a string of accidents in the summer of 2005, the European Commission drafted a regulation on air carrier identity. The regulation introduces two innovations: a blacklist of airlines which have been banned for safety reasons, and passenger information about the identity of the air carrier which is operating their flight. Common safety criteria, which are listed in the annex to the regulation, were drawn up on the basis of work performed by a committee of national experts in aviation safety and relate essentially to the findings of SAFA inspections carried out at European airports [EC 2005]. A report on safety communication that was produced for the Safety Advisory Committee Schiphol (VACS) concludes that the publication of a blacklist should be accompanied by the publication of a 'gray list' of airlines that are not banned but are considered more risky than others. Safety communication via black lists alone could lead to a false picture of safety [Stallen & Hudson 2006]. The VACS report also recommends providing additional safety information to the general public, including a simple explanation of complex and counterintuitive facts. This would result in a better appreciation of the ability to control risk, and this is an important factor in risk perception. The FAA is also in favour of informing the public better: "Because a large component of the public perception of aviation risk may not be easily assuaged by quantifiable risk ratios and accident rates, any communication system intended to inform and reassure the public about safety probably has to address more than the likelihoods of various outcomes and events" [FAA 1997]. Informing the public on safety policy and safety efforts is one of the user objectives of a causal risk model for aviation, as listed in the final report of the project group[32] causal modelling [VACS 2003a]. From a series of surveys in the Netherlands [De Gier 2005] it was concluded that the general public is greatly interested in what they themselves can do to improve aviation safety. In the chemical industry this had been recognised much earlier and obligations for government and industry with respect to public information on risk was laid down in

[32] This group consisted of representatives from the Dutch Ministry of Transport, Schiphol Airport, KLM and Air Traffic Control the Netherlands.

Article 8 of the Seveso Directive (EC Directive 82/501/EC) and a subsequent addendum, the so-called 'post Seveso Directive' (EC Directive 88/610/EC) which specifies the items the people should be informed about and requested the active provision of information to the public [Eijndhoven et al 1994]. A causal risk model would thus be useful for passengers if it can indeed be used for safety communication. The government could then use it to explain complex and counterintuitive matters, while the passengers could use it to assess what they themselves can do to improve safety and to assess the 'safety quality' of airlines. This requires a causal risk model that is not a black box, but provides insight into how various components interact and influence aviation safety. Requirements for a causal risk model to be used by passengers or for safety communication are the following:

- The model representation should be transparent.
- The model should use laymen terminology in the parts accessible to the user.
- The model should be able to represent the influence on safety of those parameters that can be directly influenced by passengers:
    - airline
    - aircraft type
    - airport of departure
    - airport of destination.

4.3.8. People living or working in the vicinity of airports

For people living or working in the vicinity of the airport, safety is a personal thing, safety is *their* safety. They do not have any direct control over safety insofar as aviation risk is concerned, other than the decision to live in the vicinity of an airport. As formulated by the chairman of the Provincial Council of Limburg: *"If a civilian lives in the vicinity of for example an airfield or another business that produces noise nuisance it is up to the civilian to accept the nuisance or not. In granting an environmental permit the business receives from the authority permission to cause a certain nuisance that is considered acceptable. The business is held to observe this permit. The authorities must oversee compliance by the business. How the authorities should oversee compliance of businesses to regulations and permits is prescribed in the provincial enforcement policy. If a civilian does not accept the 'permitted nuisance' he should not, but could, move. This is a personal consideration and he has this freedom of choice* [GS 2003]. People living or working in the vicinity of the airport expect the government to take care of air accident risk. They do not have an objective instrument to measure the level of risk, but they sometimes perceive low flying or noisy aircraft as 'risky'. "Whenever an aircraft flies low over the neighbourhood, I crawl under the staircase from sheer fright" [GVA 2005]. "On Saturday evening at approximately half past seven an aircraft with deployed landing gear flew so extremely low and with such thunderous noise over the neighbourhood that we thought he was making an emergency landing" [SP 1999]. Indirectly, citizens can influence aviation safety or third party safety by reporting safety concerns to the authority and by taking part in resident's lobby groups. In 2005, there were at least 20 active groups of residents that were lobbying against Schiphol, see Table 2. The primary focus of most of these groups is on noise and the environment. Only the Werkgroep Vliegverkeer Bijlmermeer[33] also explicitly concentrates on safety.

[33] This group was established immediately after the El-Al Boeing 747 disaster in 'the Bijlmermeer', an Amsterdam suburb, in October 1992 [Netherlands Aviation Safety Board 1994].

Table 2: List of resident's lobby groups. Source: Milieudefensie

1. Platform Vliegoverlast Amsterdam
2. Werkgroep Vliegverkeer Bijlmermeer
3. SWAB Amstelveen/Buitenveldert
4. Platform Leefmilieuregio Schiphol
5. GEUS-Aalsmeer
6. De Baanbrekers
7. Dorpsraad Vijfhuizen
8. Dorpsraad Spaarnwoude-Halfweg
9. Bewonersbelangen Isolatie Assendelft
10. Platform Vliegoverlast Assendelft
11. S.O.S. Spaarndam
12. Milieudefensie Haarlem
13. Platform Velsen Overlast Schiphol
14. Werkgroep Luchtruim IJmond-Noord
15. Milieudefensie Heemskerk
16. Platform Vlieghinder Regio Castricum
17. Zaans Natuur en Milieu Kontakt, Platform Vliegtuigoverlast Zaanstad
18. Milieu Contact Heiloo
19. Vliegtuig Overlast Sassenheim
20. Platform Overlast Schiphol Uithoorn

Research studies indicate that aviation risk is not a major concern for people living in the vicinity of airports. Effects such as noise and environmental effects are considered to be more important than risk. People are aware of the fact that the airport creates additional risk, but feelings of unsafety in general are driven by car risk and crime risk and fear for toxic substances in food, but not by aviation risk [Van het Loo et al 1999]. In comparison to discussions about effects of air transport on pollution and noise, there is little societal pressure to improve the safety performance [K+V 2005].

National and local governments are keenly aware of the importance of citizen's opinions. Citizen participation and involvement is stimulated by local meetings where people can voice their opinion. Communication is an important section in the government's plan for the evaluation of Schiphol policy [DGL 2004]. In their pursuit of citizen participation and involvement, it is understandable that the Dutch government is looking for instruments that can help them communicate about safety. Similar to the use of a causal risk model to inform passengers about the safety of various airlines, this would require a model that provides insight into how various components in the air transport system interact and influence flight safety.

4.4. Summary of user requirements and discussion on consistency

The user requirements derived in the previous section can be summarised as follows:

Airlines

1. Ability to perform cost benefit analysis of proposed future changes.
2. Future changes in how functions are operationalised can be incorporated.
3. Link with in-flight recorded data and crew safety reports.
4. Based on safety indicators other than the accident rate.

Repair stations

5. Calculate the expected safety implications of each Service Bulletin.
6. Capture very specific (specific to aircraft type or even tail number) and detailed (up to aircraft component level) information.

Manufacturers

7. Properly represent the role of the human operator and be able to capture dependencies between different system components.
8. Represent current as well as possible future accident scenarios.
9. Model is validated and accepted by the regulators.
10. Model should produce similar results if applied by different users.

Air navigation service providers

11. Ability to represent safety effects of changes in human, procedural, hardware and software elements of the ATM System as well as its environment of operations.
12. Ability to represent the contribution of ATM in both causation and prevention of accidents.
13. Representation of the total aviation system, including dependencies between system elements.
14. Provide quantitative output.
15. Ability to show the safety impacts of ATM as it develops in the future.

Airports

16. Provide insight into cause and effect of safety risks.
17. Provide insight into which risks are already sufficiently covered by safety measures, and those risks for which adequate safety measures are lacking.
18. Provide insight into the effectiveness of safety measures for the management of risk.
19. Provide insight into the probability and the effect of safety risks.

Policy makers and regulators

20. Ability to estimate the effect of local, airport specific (safety) measures and local airport characteristics (like e.g. weather and surrounding terrain, type of operations, etc.) on aircraft accident probability.
21. Ability to make a risk inventory based on projected trends and changes, and to monitor safety from safety performance indicators.
22. A way to estimate the effect of proposed safety measures on the level of safety.
23. System wide representation, including interactions.
24. Able to utilise compliance information. This requires a link to safety regulation.
25. Able to use accident/incident information, specifically ECCAIRS as this is the main incident database system used by the European aviation authorities.
26. Able to rank risks (according to a combination of probability and severity of the occurrence).
27. Able to represent emerging threats.
28. Provide a foundation for conducting risk analysis.
29. Include basic definitions such as cause, hazard, failure, their relationships and types, and the causal pathway.
30. Allow a probabilistic evaluation of risk associated with hazards in the area of operational functions and procedures.
31. If possible, hazards and risk in other areas of the aviation system such as aircraft and equipment malfunctions, hazards due to organisational set-up, human errors, environmental hazards (wind, turbulence, terrain) and contributory hazards (regulatory, economy) should be included.
32. Have a top-level representation as the user interface.
33. Ability to assess the effect of regulatory changes on accident risk.
34. Representation of current national and international regulation.

Passengers

35. The model representation should be transparent.
36. The model should use laymen terminology in the parts accessible to the user.
37. The model should be able to represent the influence on safety of those parameters that can be directly influenced by passengers:
    - airline
    - aircraft type
    - airport of departure
    - airport of destination.

People living near airports

38. Provide insight into how various components in the air transport system interact and influence flight safety.

The requirement (29) to include basic definitions such as cause, hazard, failure, their relationships and types, and the causal pathway is assumed to be inherently met in a causal risk model.

Grouped by type of requirements, the remaining user requirements can be placed under the following high level headings:

- Integrated, to represent the complexity of the complete aviation system and its many interdependencies, including the human operators and organisations (3, 4, 5, 6, 7, 11, 12, 13, 17, 18, 20, 21, 22, 23, 24, 25, 28, 30, 31, 33, 34, 37)
- Quantitative; to allow comparison with quantitative safety targets (1, 14, 19, 21, 22, 26, 28, 30)
- Transparent and clear, to provide insight and for communication purposes (16, 19, 32, 35, 36, 38)
- Model is validated and turns out reproducible results, required if the results are being used in a certification process (9, 10)
- Able to represent current and future accident scenarios (2, 8, 15, 21, 22, 27, 28).

The summary of user requirements also shows that some requirements are incompatible or cannot be achieved. The following requirements are problematic:

- The requirement for transparency
- The requirement to use laymen terminology.
- The requirement to capture very specific (specific to aircraft type or even tail number) and detailed (up to aircraft component level) information.
- Representation of current national and international regulation and the ability to utilise compliance information.
- The requirement to properly represent the role of the human operator.

As will be explained below, the requirement for transparency requires a choice to be made. The requirement to use laymen terminology can be met if the model representation and accessibility is adapted to the user. This will require a different user interface for each user group. The requirement to capture very specific information and the requirement to represent regulation cannot be met at present. The requirement to represent the human operator, while problematic, should be maintained.

Transparency

In this context, a model is considered transparent if a simple set of rules and symbols applies and if it is intuitive. Transparency and model completeness and correctness are in conflict. A model is an abstraction of reality which in itself is not transparent. If more aspects of reality are represented in the model it will become less transparent. Transparency also depends on the user. For a mathematician a set of mathematical equations will be more transparent than for an engineer or a safety manager, which means that transparency can also be increased by educating the users on the rules and symbols that are being used. In general there are three alternative ways to obtain a transparent model:

- Keep the model simple
- Educate the users
- Present a simplified representation of the model via a user interface.

Which way (or combination of ways) is selected is a choice to be made and will depend on things such as the number of different users that are expected, the type of users, whether the model will be used only once or repeatedly, the available resources, etc.

Use of laymen terminology

The requirement to use laymen terminology in the parts accessible to the user is incompatible with the general requirement to represent the total aviation system, including

dependencies between system elements, and to allow a probabilistic evaluation of risk associated with hazards in the area of operational functions and procedures. Because the level of safety in aviation is already very high, safety improvements must be obtained in the details. A causal risk model will therefore need to be able to represent these details and hence be a complex model. For communication with a general public on the other hand a simple model is required. The core of a causal risk model consists of mathematical equations, but these are counterproductive for communication with lay people. For communication it is often best to avoid mathematics but to use for instance metaphors to explain concepts. Reason's 'Swiss cheese model' is a good example of a model that is fit for communication purposes but for calculations that 'model' is rather useless, unless the concept is translated into mathematical equations. On the other hand, graphic representation of the causal flows of accidents have proven to be very effective for communication of assumptions and findings *within* a risk analysis team [Svedung & Rasmussen 2002]. A solution for this dilemma is to develop different model representations for different users. The model representation for the general public would then have to be simple, using laymen terminology and allowing only limited access to the model, the model representation for the safety specialist can be complex and allow access to a much larger part of the model.

Capturing specific information

The requirement to capture very specific (specific to aircraft type or even tail number) and detailed (up to aircraft component level) information is incompatible with the ability to show the safety impacts of the aviation system as it develops in the future and to represent current as well as possible future accident scenarios. From a practical point of view it is also incompatible with the requirement to represent the total aviation system. For instance, a causal risk model to estimate the safety implications of Service Bulletins[34] requires detailed technical representation of the aircraft down to component level. This is no easy task; it is estimated that a Boeing 747 is made up out of six million individual parts [Sutter & Spencer 2006]. On the other hand, organisational and cultural influences need not be represented. In the case of assessing a new ATM safety concept for instance, the situation is different. The difficulty for a model in this case is to be sufficiently broad to represent technical, procedural and managerial changes in all relevant fields. Detail is of less importance because the input parameters, describing the new ATM safety concept, are high level.

A balance must be found between broadness and detail. Users are not able to define precisely what they need or want, and there are some differences between the requirements from the various potential users. But the level of detail and broadness of the model are also determined by our knowledge of causal influences and the availability of data. Human performance and safety management are subject areas for which both our knowledge and the availability of data is lagging behind our understanding and data accessibility on performance of technical components for both hardware and (to a lesser extent) software. A more elaborate consideration of these aspects follows in Chapter 9, but it is already clear that our knowledge and available data are insufficient to develop a causal risk model of the total aviation system that represents very specific (specific to aircraft type or even tail number) and detailed (up to aircraft component level) information.

[34] A Service Bulletin is information from the manufacturer describing proposed technical alterations to the equipment. Implementation of these Service Bulletins is not mandatory; it is up to the airline to decide whether the updates are incorporated.

Representation of regulation

Regulation can be seen as making mandatory some risk control measure or some aspect of their management. We propose to represent management of risk control measures explicitly in the model (see section 9.2), so representation of regulation would 'only' require adding a box or node to every measure to represent whether or not it is already mandatory and the related steps of inspection by the regulator for compliance with the regulation and issuing sanctions against its absence or inadequacy. Although this is a formidable task it is not undoable. This would then enable identification of the quantitative relation between regulation and aviation safety performance, a relation which is currently not known [Hansen et al 2005]. However, representation of managerial influences in itself is already one of the bottlenecks (see section 9.2) and although we propose solutions to those bottlenecks it is recommended to focus the efforts first on making these solutions work without running the risk of being bogged down in a simultaneous effort to attach the effects of regulation on those managerial influences.

Representation of human operators

The dilemma

The requirement to represent the role of the human operator is problematic because of different views on how to deal with human operators in quantitative safety assessments. Although certification regulation on aircraft systems contain quantitative failure probability requirements and says that regardless of the type of assessments used, the analysis should always be accomplished with consideration of all relevant factors, including the flight crew, the current regulatory advisory material on aircraft system design and analysis (FAA Advisory Circular AC 25.1309-1A) states that quantitative assessments of the probabilities of crew errors are not considered feasible [FAA 1988]. Demonstrating compliance with this part of the regulations is therefore to a large extent done on qualitative criteria. Ultimately the certification test pilot will then judge on the acceptability of a particular system design. So while quantitative safety assessments are well established in aircraft system design and development, as soon as a human operator (such as a pilot) plays a role, the judgment of acceptability of a design switches from quantitative to qualitative. Tasks are qualitatively evaluated to determine if the crew can realistically be anticipated to perform them. The advisory material allows full credit for correct pilot action in response to a system failure, based on the following guidance:

> "When assessing the ability of the flight crew to cope with a failure condition, the warning information and the complexity of the required action should be considered (Paragraph 8g(5) of FAA Advisory Circular AC 25.1309-1A). If the evaluation indicates that a potential failure condition can be alleviated or overcome during the time available without jeopardising other safety-related flight crew tasks and without requiring exceptional pilot skill or strength, credit may be taken for correct and appropriate corrective action, for both qualitative and quantitative assessments."

So while for aircraft systems a quantitative safety assessment is required, this is not considered feasible for assessing the characteristics of the combination of man and machine. This relies to a large extent on a qualitative assessment by the certification test pilot to determine, during a series of prescribed test flights, whether the behaviour of the aircraft and the man-machine interface are acceptable. Requirements for acceptability are described qualitatively in the certification requirements. Aircraft handling characteristics are rated on a qualitative scale such as the Cooper - Harper rating scale, see textbox [Cooper & Harper 1967]. The importance of the test pilot is essential, not only for performing the test flight but also in simulation, in the design or evaluation of the flight

control system, and test-flight preparation [Thompson 2000]. It is significant that the field of industry that was one of the key drivers for the development of probabilistic risk assessment and quantified safety targets (i.e. aircraft certification) has always held the opinion that estimating human error probabilities is not possible even though the nuclear power industry, which was another key driver, accepts human error probability estimates as part of the certification of nuclear power plants. One of the reasons for the reluctance of aircraft developers and manufactures to accept quantitative human error techniques is that they have always had a good alternative in the qualitative assessments by the test pilot.

The Cooper-Harper rating scale

A well-known quantitative method for rating aircraft handling characteristics is the Cooper Harper scale [Cooper & Harper 1967]. This utilizes a decision-tree process to guide the user through a series of questions. The answers lead the user to a set of three sub alternatives which ultimately result in a numerical rating from 1 – 10. Cooper and Harper used four broad categories within which to describe aircraft handling qualities:

- Uncontrollable: Unsuitable for any task.
- Unacceptable: Not suitable for the task but aircraft still controllable
- Unsatisfactory but tolerable: adequate for the task but improvement desirable.
- Satisfactory: No improvement required.

The following three questions help the user placing the system into one of the four categories:

- Is the vehicle controllable?
- Is adequate performance attainable?
- Is the system quality satisfactory without improvement?

By separating the main categories into subcategories, a ten-point scale is achieved. The worst rating is a 10 "Major deficiencies -control will be lost during some portion of required operation", the best possible score is a 1 "Excellent, highly desirable - pilot compensation is not a factor for desired performance". The Cooper-Harper rating scale has become the standard way of measuring flying qualities.

Flight testing

The certification test pilot provides an expert opinion on the acceptability of a particular design. He or she will have been given dedicated training, for instance at a test pilot school, to be able to provide this expert judgement. Rather than being gung-ho heroes who risk their lives for the advancement of technology, as they are portrayed in Tom Wolfe's 'The right stuff' and Chuck Yeager's autobiography [Wolfe 1979, Yeager & Janos 1985], a test pilot is trained (among other things) to be able to make assessments of the implications on safety of particular characteristics or features of the aircraft, based on a limited sample during a series of test flights [Duke 1953, Lithgow 1956, Johnston & Barton 1991]. This requires knowledge on the aircraft's design and systems, aerodynamics, flight mechanics, stability & control, and the ability to interpret the effect of certain characteristics and or features as displayed during the circumstances of the test flight and to assess their consequences for other potential circumstances. Many test pilots have a degree in (Aeronautical) Engineering[35] so that they are able to translate a deficiency encountered in the air into precise engineering terminology [Crossfield & Blair 1960]. These

[35] A degree in Engineering is one of the requirements to get accepted at the USAF Test Pilot School.

characteristics make test pilots very suitable for providing quantified judgements on causal risk model parameters for which other data is lacking, see also section 8.5.5.

During flight testing the flight envelope of the aircraft is explored. Inevitably, when operating close to the boundaries of the flight envelope while these boundaries have not yet been precisely defined, there is an increased risk of accidents. This is evident when examining statistics of flight test accidents. In Table 3, fatal flight test accidents of commercial air transport aircraft are listed. We assume that on average each aircraft type will require approximately 1000 test flights for certification [Kingsley-Jones 1997, 2005, Norris 1988], and that there have been approximately 180 different types of large commercial aircraft (100 jets, 80 turboprops) since 1960. These aircraft will have accumulated a total of 180.000 test flights while according to the table there have been 22 fatal flight test accidents. The corresponding accident rate is 1.2×10^{-4} per flight, which is approximately 100 times higher compared to the accident rate for commercial flights in the same time period. Table 3 demonstrates one of the reasons why the approach of, say, the nuclear power industry towards certification of nuclear installations has to differ from the approach of the aviation industry towards certification of aircraft. A crash of a prototype aircraft during a test flight would seem to be acceptable to society, a serious accident with a nuclear powerplant during its certification process clearly is not. The crash of the aircraft will almost certainly only affect those on-board[36] and for them the risk is voluntary, while a nuclear disaster may also affect a great number of people in the vicinity of the powerplant and for them the risk is not voluntary. The nuclear power industry has to rely - more than aircraft certification- on safety and performance calculation rather than tests, and risk assessment methodologies were developed for this purpose. Whereas aircraft certification considers quantitative assessments of human error probabilities not feasible and partially relies on the test pilot's judgement on assessing the ability of the flight crew to cope with certain conditions, the nuclear power industry started development of human reliability methods to be used within a probabilistic risk assessment to calculate human error probabilities. The nuclear power industry does not have the equivalent of a test pilot; in the nuclear power industry there are no 'test-operators', trained, qualified and experienced in taking the plant to the edge of the performance envelope to judge whether the plant's behaviour is acceptably safe. They cannot do this 'edge testing' as the possible consequence (a core meltdown) is unacceptable. As a matter of fact, trained 'test-operators' do not exist for any type of control room, including ATM control rooms. The requirements for certification of ATM systems indeed explicitly refer to human reliability methods. The current approach in risk assessment and mitigation for Air Traffic Control systems in Europe, as described in Eurocontrol's ESARR 4, is based on a quantitative target level of safety. This quantitative safety requirement is somewhat similar to the quantitative requirements for system safety in aircraft certification. A difference is that the ATM requirement is not limited to systems, but should address 'the three different types of ATM elements (human, procedures and equipment), the interactions between these elements and the interactions between the constituent part under consideration and the remainder of the ATM System' [Eurocontrol 2001b]. Human error models are being proposed that are a further development of the basic concepts and models that have primarily been developed in the nuclear industry (e.g. HERA - predict) [Isaac et al 2004].

[36] Predominantly, test flights are deliberately conducted over uninhabited regions.

Test pilots

The two decades following World War II (1939-1945) was a period of enormous technological advances. Aircraft were being designed to fly into realms that were unexplored and not yet well understood [Rotundo 1994]. Almost every aircraft could fly faster, further and higher than its predecessor. There was hardly any knowledge to be able to predict the behaviour of the aircraft, and again the role of the test pilot was invaluable. He was tasked to fly the aircraft into the unknown flight regimes, observe the aircraft's behaviour, and communicate that information back to the aircraft design engineers [Bridgeman & Hazard 1955, Crossfield & Blair 1960, Beaumont 1994].

Figure 11: Test pilot A. Scott Crossfield in the cockpit of the Douglas D-588-II after the first Mach 2 flight (twice the speed of sound) on 20 November 1953 (source: NASA Dryden Flight Research Center).

Requirements for test pilots matured, test pilot schools were established in quick succession and began providing dedicated training. Typically a test pilot school provides a year-long curriculum of classroom academic, simulator training and flight exercises, with main topics of performance, flying qualities, systems, and test management. Test pilots became engineers.

Table 3: List of fatal flight test accidents, 1960-2003, commercial air transport aircraft (jets and turboprops) involving the aircraft manufacturer. Source: NLR Air Safety Database.

DATE	AIRCRAFT TYPE	LOCATION	NUMBER OF FATALITIES	ACCIDENT
22/10/1963	BAC 1-11	United Kingdom	7	Entered deep stall during stall testing.
25/02/1965	Ilyushin IL-62	Zhukovsky, Russia	6	Overran runway after abandoning the take-off.
12/05/1965	HFB HansaJet	Torrejon, Spain	1	Deep stall followed by flat spin.
03/06/1965	Hawker- Siddeley HS-121 Trident	United Kingdom	4	Deep stall.
14/01/1966	Tupolev Tu-134	Russia	8	Loss of control during high speed maximum rudder deflection.
12/05/1970	Aero Spaceline 377 Mini Guppy	Edwards AFB, USA	4	Lost control following simulated engine failure during take-off.
19/11/1970	IAI Arava	Jordan	3	Loss of control during flutter test.
23/05/1971	Aérospatiale SN600 Corvette	Istres, France	3	Failure of horizontal stabilizer.
01/02/1972	VFW-614	Bremen, Germany	1	Tab flutter, aircraft entered vertical dive.
31/10/1972	Dassault Falcon 10	France	2	Aft fuselage separated during rudder trim test.
06/07/1977	Let 410	Czechoslovakia	4	Structural failure after rudder deployment at high speed.
23/05/1978	Tupolev Tu-144	Russia	2	Crash landed after in-flight fire.
03/04/1980	Canadair CL-600 Challenger	Mojave, California, USA	1	Stall and loss of control during stall testing.
26/03/1982	Dornier Do-228	Germany	3	Entered steep dive during trim test.
13/10/1992	Antonov 124	Ukraine	8	Loss off control when nose cargo door opened at high speed test.
03/02/1993	Lockheed L-100-20 Hercules	Marietta, Georgia, USA	7	During V_{mcg} testing inadvertently lifted off below V_{mca} and spun in.
05/07/1993	Ilyushin IL-114	Ramenskoye, Russia	5	Rapid pitch-up and stall immediately after take-off.
26/07/1993	Canadair RJ100	USA	3	Entered deep stall during low speed steady-heading sideslip.
30/06/1994	Airbus A-330	Toulouse, France	7	Pitch up and stall during autopilot certification test.
10/02/1995	Antonov 70	Russia	7	Collided with chase aircraft.
10/10/ 2000	Canadair CL-604 Challenger	Wichita, Kansas, USA	2	Stalled immediately after take-off with aft centre of gravity.
26/04/ 2003	Sino Swearingen SJ-30-2	USA	1	Aircraft became unstable during flutter test.

The need for a change of views

The distinction between which element is part of the ATM system and which element is part of the aircraft is becoming less clear than it was in the past, and this development can be expected to continue in the future. According to Eurocontrol [2001b], the ATM system includes both ground-based and airborne components. This is consistent with a growing awareness of the need to apply a total systems approach in safety assessments, considering the complete aviation system rather than its individual components. This will have implications for the methodologies that have been and are being used for the assessment of the safety of the components of the aviation system. The principle of decomposition (which is one of the cornerstones of reliability theory) according to which the reliability of a system is the result of the reliability of its individual components does not necessarily hold for the integrated system. Accidents that were the results of failures in the interfaces of the individual elements of the system rather than failures of the system themselves, like the mid-air collision above Überlingen, illustrate the importance of interfaces [FAA 2002, Roelen et al 2003, Roelen 2004]. It can therefore be expected that regulatory requirements will be developed for the integrated aviation system rather than the individual components. The complexity of the matter will discourage the authorities from specifying detailed requirements. Instead, a target level of safety will probably be specified. Quantitative safety assessment methods such as causal risk models will have to be applied to demonstrate to the authorities that these targets have been met. Because the human operator (pilot, air traffic controller, etc) still plays a decisive role in the aviation system these quantitative risk assessments methods must be able to represent the role of the human operator in a way that is satisfactory for all parties involved.

Inappropriate interface: the Überlingen mid-air collision

On July 1, 2002, a Tupolev TU-154M, en-route from Moscow to Barcelona, collided in mid-air with a Boeing 757-200 which was en-route from Bergamo to Brussels. The collision took place at an altitude of 10,800 m above the German city of Überlingen. Both aircraft crashed after the collision and all people on board (2 in the Boeing and 69 in the Tupolev) were killed. At the moment of the collision ACC Zurich was providing air traffic control to both aircraft. Due to maintenance activities to the ATC system, the Short Term Conflict Alert System (STCA) of ACC Zurich was not operational. The Air Traffic Controller initially failed to observe that the aircraft were on collision course. The Traffic Alert and Collision Avoidance System (TCAS) in both aircraft operated as designed; the TCAS in the Boeing alerted the crew to ‘descend’ and the TCAS in the Tupolev alerted the crew to ‘climb’. At the same time, the air traffic controller noticed the conflict and instructed the Tupolev to descend. Subsequently the Boeing’s TCAS provided an ‘increase descent’ alert while the Tupolev’s TCAS provided an ‘increase climb’ alert. The crew of the Boeing followed their TCAS alerts and started to descend. The crew of the Tupolev, confronted with conflicting instructions, decided to comply with the Air Traffic Controller’s instruction to descend rather than following the TCAS instruction to climb. The collision occurred 50 seconds after the first TCAS advisory. In the subsequent accident investigation, the German air accident investigation board BFU concluded that one of the ‘systemic’ causes of the accident was the insufficient integration of TCAS in the aviation system [BFU 2004]. This allowed a situation where a flight crew was confronted with contradictory instructions.

4.5. User expectations: lessons from CATS

This section addresses some issues specifically arising from CATS (Appendix C). These are used here to illustrate the problems of defining user needs and of trying to reconcile incompatible needs.

On a general level, three expectations regarding the function of the causal risk model being developed for the CATS project have been indicated [De Jong 2006]:

- The model should provide understanding how safety depends on different elements of the current operation. With this understanding, it should be able to indicate the big risks, strong and weak areas and provide strategic directions for safety improvements and safety research.
- The model should be used as a monitoring/ evaluation tool.
- The model should be able to determine the safety effects of future changes in the aviation system or subsystem, thereby supporting decision making and policy development.

During a series of interviews with potential users of a causal risk model [De Jong 2006], it was recognised that changes in one part of the aviation system may have effects at a very different place in the system. A causal risk model should be able to find such effects. The model should not only tell what happens with safety when changes are made to the operation itself, but also to its development, management and environment. The risk results of a causal risk model should to some extent be quantified, for instance by means of expected accident probabilities, although it is recognised that quantification may not always be feasible, and qualitative answers may be sufficient. Such results can be used to verify conformance to target levels of safety, for the overall system or parts of it. Reliability of the results of the model is crucial; indications of the reliability should be given. The results need to be understandable to the users – the model should not be a black box. This is not the same as merely confirming what people thought they knew already. Reliable and understandable results are necessary for confidence in the model. A causal risk model should identify where the big risks are and what are causes and contributes to the risk. It should indicate possible solution areas to help the right organisations (the risk owners, also to be identified by the model) to solve the problem. A causal risk model was perceived by the interviewees as a tool for strategic changes for which time is usually not critical; there are usually at least months available. In summary, the users were of the opinion that the model should (also) be able to study the effects of changes behind the direct operation, such as development, management and environment. Risk results should be quantified (to some extent), reliable (or have well-defined reliability) and understandable (the model should not be a black box) in order to have confidence in the model. The results of the model should help to identify possible solutions areas.

A second series of interviews with potential users of a causal risk model [De Jong 2007] was felt to be necessary because the results of the first round did not yet provide a clear and sufficiently detailed set of requirements. However, the second round revealed even more how difficult it was for many users to define how and when a causal risk model should be used, other than in the most general terms. The expression ‘causal risk model’ is too generic and open to interpretation, depending on a person’s own terminology and understanding of the subject. Furthermore, because a comprehensive causal risk model for aviation had not been developed at the time of the interviews, it was difficult for users to envision how such a model should actually be used to help them in making decisions related to safety, particularly since ‘safety’ itself is such an intangible concept. Under these conditions, it are the model developers who should provide guidance on possible model

performance. Several alternative uses should be presented, including the consequences of selecting a particular choice. Model developers should not shy away from a conclusion that some requirements may be incompatible, but should be explicit about this.

The required level of detail of a causal risk model depends on the type of user. Local users, such as individual airlines (e.g. KLM) or airport authorities (e.g. Heathrow airport), need models that accurately describe their specific local situation, both in the causal structure as well as in the quantification. General users, such as international authorities (e.g. EASA) and trade organisations (e.g. IATA) need models that accurate describe the general situation. The models for local users should be relatively narrow in scope and deep in detail, while models for general users should be relatively broad in scope and shallow in detail. There is a difference in dynamics and consequently in required accuracy as well. Local users have operational and tactical questions, their decisions have immediate effect, and long term for an airline means 2-5 years. General users have tactical and strategic questions, the effect of decisions may take decades to materialise. Developing a model which can help answer local operational questions as well as general strategic questions is practically unfeasible. It would require a 'broad and deep' model and consequently a very substantial development effort. If the purpose of the model is not clear at the beginning of the development effort, progress will continue to be frustrated by the ambition to represent detail as well as broadness. Eurocontrol's Integrated Risk Picture (IRP) [Perrin et al 2006, Kirwan 2007] is an example of a model which was from the outset aimed at providing higher level strategic insights. For the CATS model of the Dutch Ministry of Transport and Water Management (see Appendix C) the objective was not so clearly defined but was considered to be part of the research question. It was decided that the model should give a global picture of the risks involved in commercial fixed wing aviation from gate to gate, including aerodrome operations, ATS operations, aircraft operations in all flight phases for all aircraft involved. The model should also reflect the risks deriving from aircraft design and maintenance. Simultaneously, the model should be 'as detailed as possible' [Ale 2007]. Interviews with projected users resulted in the requirement to support the authority and the sector parties and to be sufficiently simple to be understood by the general public [De Jong 2007]. It was even suggested at the CATS website (www.verkeerenwaterstaat.nl/causaal model) to use the model to determine the causes of individual incidents. The CATS model was developed top-down (i.e. from the generic to the specific), and because users were not really aware of what to expect they were sometimes frustrated that their urgent operational problems (like an increase in overspeed warnings on the Boeing 777 fleet of a particular airline, fluctuations in bird strike incidents at Schiphol airport) were not immediately recognisable in the model while this was being developed. Such frustrations are potentially harmful as they can cause members of the aviation industry to suspend their support for model development. Support from the industry is essential for the provision of data and to provide insight, to the model developers, into operational aviation processes.

The user of the model should be able to relate daily operational practice and decision making to the elements in the model or the model representation. The model developers should therefore have sufficient knowledge of- or insight into those operational practises and decisions. For a successful and workable model, the developers of the model themselves should have sufficient knowledge of the aviation system and its processes and at the same time ascertain the scientific and mathematical correctness.

Feedback from the modellers to the potential users is required to get a better grip on user requirements. The modellers should indicate possibilities and limitations of the model, and should come-up with real-life examples of cases in which a causal risk model would be a

helpful tool. This will have to be an iterative process. The following possibilities and limitations are apparent from experience in CATS and during the development of a causal risk model for the FAA:

- One single model cannot be used to answer each question of every possible user. Depending of the problem at hand, some parts of the model may have to be developed further in order to answer the question.
- Modelling involves numerous assumptions. In case the assumptions do not hold, the model cannot be used without taking the assumptions into account.
- A causal risk model should not be used as the sole input for assessing compliance with regulation unless the model and its use are agreed upon by the relevant stakeholders[37].
- A causal risk model is intended for use by professionals within the aviation industry. It is not suitable as an instrument for safety communication from the government to the general public.
- A system-wide causal risk model is suitable to support decision making at the strategic level. It is not useful at the level of day-to-day operations because this almost invariably involves very specific situations which cannot be represented correctly in a causal risk model. This is not a fundamental limitation of a causal risk model but a practical limitation, depending on time and money available for model development.
- A strength of a causal risk model is that it assembles information from different disciplines. It is therefore specifically suitable to support decision making for situations involving multiple actors or disciplines.
- A strength of a causal risk model is that it assembles current knowledge.
- The case for which a causal risk model is used should be defined at the proper level of detail. The model cannot be estimated to assess for example the safety impact of the introduction of uninhabited areal vehicles (UAVs) in commercial airspace unless we define where and when this introduction will take place, and the operating procedures, performance characteristics, on-board equipment and other relevant information of those UAVs.

Obtaining valuable user requirements demands that the users are not distrustful towards the model. Due to the fierce competition in aviation, industry parties are paranoid about everything that even has a remote chance of disturbing the 'level playing field'. Historically, many risk analysis have been performed to demonstrate that an installation, aircraft or spacecraft conforms to requirements [Bedford & Cooke 2001]. The industry's natural reaction to proposals that may potentially be restricting is to resist. For that reason they tend to resist changes of the instruments of the authorities for regulatory compliance assessment. Modellers should therefore emphasise the possibility of using a causal risk model as input to the decision making process instead of using the model for strictly assessing compliance with regulation[38].

4.6. Conclusions for this section

User requirements for a causal risk model have been derived in various ways, starting from a description how aviation has realised tremendous safety improvements since the first commercial airlines started operations, improvements that were the result of technological

[37] This is a policy choice, not an inherent limitation of a causal risk model.

[38] The use of the model is not a decision that the modellers must make, but what they can do is warn against possible consequences.

advances and the systematic analysis of accidents. Because of the very success of aviation safety, the fix and fly approach is no longer effective on its own. To maintain and improve the current level of safety, there is need for an aviation safety performance evaluation methodology that is not based on fatal accidents and hull losses alone. While accidents are rare events, the consequences of an accident can be enormous for the airline, the aircraft manufacturer, the air navigation service provider and society. Accident prevention and safety management is therefore as important as ever, but due to the rarity of accidents it becomes more and more difficult to understand the 'current state of affairs' and to determine what the effects of changes are on the level of safety. There are simply too few 'data points'. Aviation is a competitive environment, and the investments are often huge. Any proposed safety improvement measure must be well-reasoned. Therefore an integrated safety assessment methodology is needed. For the industry, an important characteristic of such method will be the ability to compare projected costs to estimated benefits, in terms of increased safety, of particular decisions. For the regulator, reproducibility of the results is crucial if the method is destined to be part of regulation, and transparency is an important feature for communication regarding the model and model results. For both the regulator and the industry it is important to be able to feed the model with data that is currently being gathered in various occurrence reporting systems.

The increased integration, automation and complexity of the aviation system, including integration of ground-based and airborne systems, will lead to new regulatory requirements. The complexity of the matter will discourage authorities from specifying detailed requirements. Instead, quantitative targets will be specified. There is a growing awareness of the need to apply a total systems approach in safety assessments, considering the complete aviation system rather than the individual components. Quantitative safety assessment methods such as causal risk models will have to be applied to demonstrate to the authorities that the requirements have been met. Because the human operator (pilot, air traffic controller) plays a decisive role in the aviation system, these quantitative safety assessment methods must be able to satisfactory represent the human operator. Up to now, the opinion in regulatory requirements for aircraft system safety assessments has always been that assessing the probability of human error is not feasible. This opinion is contradicted by the use of 2nd and 3rd generation human reliability methods in the nuclear power industry (see also section 9.1). Historically, there has never been a real necessity for such estimation because aircraft system safety assessments relied on the opinion of the certification test pilot, who is specifically trained to provide such judgements. This opinion could be a major obstacle in the development and implementation of causal risk models. As long as this view is maintained it is rather unlikely that a truly integrated quantitative aviation safety assessment can be conducted. Given the experience with human reliability methods in the nuclear power industry and the need for integrated assessments it becomes unlikely that aircraft certification can afford to maintain its view on the lack of feasibility of estimating human error probabilities. Representation of human operators in quantitative models is further discussed in section 9.1.

This need for quantitative safety assessment methods is explicitly voiced by safety regulators. FAA, EASA, Eurocontrol and the Dutch Ministry of Transport have all expressed the need for a causal risk model. There are differences though. The FAA calls for the development of a method for a probabilistic evaluation of risk associated with hazards in the area of operational functions and procedures, if possible extended with hazards and risk in other areas of the aviation system such as aircraft and equipment malfunctions, hazards due to organisational set-up, human errors, environmental hazards (wind, turbulence, terrain) and contributory hazards (regulatory, economy). A specific requirement

is to have a top-level representation as the user interface for performing risk analysis. The European regulator EASA is specifically looking for a methodology to quantify the safety effects of proposed regulatory changes in terms of accident probability and severity. The Dutch Ministry of Transport and Water Management also sees a role for causal risk models in the regulation of third party risk. Additionally, the Ministry believes that a causal risk model can be used for communication with citizens, to provide insight in how various components interact and influence aviation safety. Eurocontrol has mandated the use of quantitative safety assessments of significant changes in the ATM system to demonstrate that the probability of ATM directly contributing to accidents is less than 1.55×10^{-8} per flight hour. Eurocontrol also emphasises the development of safety monitoring and data collection mechanisms. The rationale is that any model used in risk assessment must be validated, and real data could contribute significantly to this validation process. With the exception of the requirement to use the model for communication with citizens, the different viewpoints of the regulators are not necessarily incompatible, but the modellers should be aware of them and should make choices, because it will be difficult to develop a model that meets all needs simultaneously.

The industry (aircraft manufacturers, airlines, ANSPs, airports) has not explicitly expressed a need for a causal risk model. Perhaps there is a link with a fear that such a model might upset the level playing field. That does not mean that such a model could not be useful for them. Modellers should emphasise the possibility of using a causal risk model as input to the decision making process instead of strictly using it for assessing compliance with regulations. A feature that would make such models attractive for them is the ability to perform cost benefit analysis. Some of potential industry applications may require very detailed models however, that describe the aircraft down to component level.

A detailed set of requirements for a causal risk model was defined (see section 4.4). Some of these requirements proved to be problematic because they are incompatible with other requirements and could not be maintained. The remaining requirements can be grouped under the following high level headings:

- Integrated, to represent the complexity of the aviation system and its many interdependencies, including the human operators and organisations.
- Quantitative; to demonstrate to authorities that quantitative requirements have been met.
- Transparent and clear, to provide insight and for communication purposes. What is transparent depends on the user.
- Model is validated and turns out reproducible results, required if the results are being used in a certification process.
- Able to represent current and future accident scenarios.

The requirement for transparency requires a choice to be made between educating the user and keeping the model or model representation simple. The detailed requirement to use laymen terminology can be met if the model representation and accessibility is adapted to the user. This will require a different user interface for each user group. The requirement to capture very specific information and the requirement to represent regulation cannot be met at present. The requirement to represent the human operator, while problematic, should be maintained.

User requirements demand a quantitative and transparent causal risk model that should be able to represent environmental and managerial influences. The results of the model should

be reliable, and the model should be useable to indicate possible solution areas for further safety improvement. The variety of users and diversity in the users' questions make it impossible to develop a causal risk model that is capable of answering all needs simultaneously. The requirements call for a system-wide representation of air transport. Such a model is not suitable to answer very specific questions. This limitation should be explained to the aviation community to avoid the creation of false expectations. This is particularly relevant in view of the fact that most of the expected users only have a vague picture of what a causal risk model is and how it could or should be used. It should also be verified, for each kind of user, that the causal risk model definition as adopted in this thesis, is suitable for them.

Due to the diversity of possible users and the limited view they currently have on the use of causal risks models, significant works remains to be done on the systematic identification of user needs. Feedback from the modellers to the potential users is required to get a better grip on user requirements. The modellers should indicate possibilities and limitations of the model, and should come-up with real-life examples of cases in which a causal risk model would be a helpful tool. This will have to be an iterative process.

Support from the industry is essential for the provision of data and to provide insight, to the model developers, into operational aviation processes.

Chapter 5. Examples of aviation safety analyses

There is no binding reason why safety assessment or safety analyses should involve quantitative causal risk models. There are other constructs that can be used for assessing or analysing safety as well and these methods can be seen as alternatives to causal risk models. This section will discuss some examples of methods that are currently used for aviation safety assessment and management, and compare those to the requirements and needs identified in the previous chapter to determine whether the use of such alternative methods is sufficient to bring about the required safety improvements or fulfil future needs. The examples were selected on the basis that they represent questions pertaining to actual safety issues for which a stakeholder needed an answer. They are considered typical for the type of approaches being used. The aim here is to provide a representative, albeit not exhaustive, overview of methods to determine if current approaches to safety assessment and analysis meet the criteria (transparency, reproducibility, etc) and answer the practical questions of the stakeholders. Because the methods described in this chapter can be seen as alternatives to a causal risk model, the examples described here will then help to further specify the requirements for a causal risk model by indicating what they do not have which such a causal model should have.

5.1. Safety of mixed VFR/IFR air traffic at Geneva Airport

Background

Geneva International Airport operates two parallel runways, one long concrete runway and a shorter grass strip. The concrete runway is used predominantly by heavy aircraft flying under Instrument Flight Rules (IFR), and the grass strip is restricted to small single-engine aircraft operating under Visual Flight Rules (VFR). The lateral spacing between the two runways is 250 metres, qualifying them, according to ICAO [2004a], as closely spaced parallel runways with regards to wake turbulence separation minima. Mixed streams of heavy and light aircraft during simultaneous closely spaced arrivals and departures, in particular at a busy international airport such as Geneva, require specific attention to assure the safety of such operations.

The research question

Because of growing traffic volumes, the question arose whether the existing mix of light aircraft and commercial traffic could still be regarded as compatible and safe, now and in the future. It was beyond the scope to aspire to reach the completeness of a full safety assessment.

The research approach

To answer the research question three lines of activity were carried out. The first was an assessment of whether the existing mixed VFR/IFR operations were in compliance with the applicable rules and regulations. The second concerned an evaluation of safety management practices and organisation at the airport. The third line of approach comprised a hazard assessment of the mixed VFR/IFR operation at Geneva airport. These activities included structured interviews with representatives of the airport, local Air Traffic Control unit and representatives of people living near the airport. A structured brainstorm was also held, involving a panel of eight operational experts and three risk management specialists. This

information was completed by supporting data and analysis, including an analysis of aircraft trajectories and meteorological conditions at Geneva airport [Van der Geest et al 2005].

Discussion
In this study, the question as to whether the system was sufficiently safe was not answered by deriving the level of safety and comparing it to pre-defined criteria. Instead, the system was analysed by comparing the existing situation to best practice and common sense. Such an approach is only realistically feasible when conducted by individuals who themselves are subject matter experts and cover the full scope that is relevant. The study looked at separate system elements and their interconnections. Although the nature of the issue itself (mix of VFR and IFR traffic) is an integration issue, there was no systematic way of addressing dependencies between system elements. All results of the study were qualitative. This was sufficient to answer the research question and not necessarily a characteristic of the research approach. No specific data gathering was required for the study, although some radar track data were analysed to obtain a detailed picture of the situation at the airport. This data had been collected by the airport to answer a question not directly related to this study, but came in handy. All conclusions and recommendations were based on argumentation. As a result the report may appear transparent, but the results were partially based on the outcome of a brainstorm session and because of that the reproducibility is questionable and the method is not fully transparent.

5.2. Safety assessment of parallel approaches at Helsinki-Vantaa Airport

Background
To increase airport capacity, Helsinki Vantaa airport considered a switch from dependent parallel approaches to using independent parallel approaches on the two parallel runways 04L/22R and 04R/22L. The new parallel approach procedures would enable aircraft to land simultaneously on both runways.

The research question
According to Eurocontrol requirements, particularly ESARR 4, a safety assessment of the proposed procedures had to be carried out before implementation of the procedures could be approved.

The research approach
The safety assessment was executed according to the Eurocontrol Safety Assessment Methodology or SAM [Eurocontrol 2006a], which prescribes the following three steps: Functional Hazard Assessment (FHA), Preliminary System Safety Assessment (PSSA) and System Safety Assessment. The results of the first two steps for the independent parallel approaches at Helsinki airport are described in Eurocontrol's safety assessment report on independent approaches to parallel instrument runways at Helsinki-Vantaa airport [Eurocontrol 2006b].

Discussion
Hazards were identified in this study by means of brainstorming and a functional analysis. The objective of the PSSA was then to derive a set of requirements that would mitigate the effects of the hazards identified in the FHA. The likelihood of the hazards was not estimated, but there was a qualitative estimate of the possible effect of the hazard in terms of a severity class. Because the analysis results were partially based on the outcome of a brainstorming session, the reproducibility of the results is questionable and the method is not fully transparent. There was no explicit attention to possible interdependencies and

therefore the approach cannot be considered integrated. As a matter of fact, because the results were largely based on a functional analysis, there is a significant risk that 'non functional' hazards, like violations, were overlooked [De Jong 2004]. To develop mitigation measures, fault trees were created to identify causes of hazards, and event trees were developed to identify possible consequences of the hazards. These fault trees and event trees were developed specifically for this study and were not based on existing models.

5.3. Safety assessment of offset steep approaches at Lugano Airport

Background

The airport of Lugano, Switzerland, is situated in mountainous terrain. In 2006 one of the instrument approach possibilities used a glide path angle of 6.65 degrees. It was expected that future use of this approach would not be accepted by the Swiss regulatory authority unless aircraft were to be certified for such a steep approach. This would severely limit the number of aircraft types that can use Lugano Airport commercially. To counter the effect of this measure the airport developed an alternative approach with a glide angle of 5.5 degrees. This approach procedure was submitted to the regulatory authority (the Federal Office for Civil Aviation - FOCA).

The research question

FOCA responded by indicating a number of concerns regarding the procedure and requested Lugano Airport to conduct an Aeronautical Study to analyse and evaluate the risks of the new procedure and to propose, if necessary, possible mitigating measures or alternative solutions. The National Aerospace Laboratory NLR was contracted by the airport to perform such an Aeronautical Study.

Research approach

The study included an assessment of the flight operational and flight safety risks related to the introduction of the draft procedure. This was a qualitative safety assessment based on inputs from operational experts, databases and other sources for estimates of severities and frequencies rather than mathematical modelling and simulation. Operational safety experts were NLR safety experts and operational experts from Lugano airport, ATC and pilots with flying experience at Lugano. The safety assessment method was scenario-based; safety issues were analysed in the context of accident scenarios rather than in the context of individual hazards. These scenarios are bundles of event sequences, each centred on a central hazardous situation, including the possible causes and the possible consequences of that situation. Severity and likelihood of event sequences were estimated by experts, using qualitative rating scales: *hazardous*, *major*, *minor* and *no significant safety effect* for severity; *probable*, *remote*, *extremely remote* and *extremely improbable* for likelihood [Wever et al 2006].

Discussion

The study addressed several different accident scenarios and as such was not focussed on a single system element. The approach was not fully integrated because interdependencies between system elements were not explicitly addressed. Some interdependencies will have been implicitly accounted for by the experts during the analysis. The results of the study were qualitative. Probabilities of occurrence of events were subjectively estimated by experts and expressed semi-quantitatively, using terms like 'remote'. No other data analysis was conducted and hence there was no need to collect data as part of the study. A classic risk matrix, combining likelihood and severity of the event, was used to illustrate acceptability of the risk. Because the results of the study were primarily a product of

brainstorm sessions, the overall result of the study depended on the subjective opinion of the experts involved. The process was not fully transparent, because it was not completely traceable what the expert's opinions were based on. Reproducibility of the results is questionable, as different experts may have had different opinions and arrived at different conclusions.

5.4. Reduced vertical separation minimum in Europe

Background

On 24 January 2002 the introduction of a Reduced Vertical Separation Minimum (RVSM) in European airspace created six additional flight levels between 29,000 ft and 41,000 ft. This was achieved by reducing the vertical separation minimum between aircraft from 2,000 ft to 1,000 ft. RVSM was implemented simultaneously in the airspace of 42 European and North African countries and was the biggest change in Europe's airspace in 50 years [Tiemeyer 2003].

Research question

It had to be demonstrated to the international aviation community that the Target Level of Safety set out by ICAO for the vertical collision risk would not be exceeded in the European RVSM airspace. During the initiation of the RVSM program, it was felt by Eurocontrol that to demonstrate the achievement of ICAO's collision risk criterion would be viewed as necessary but not sufficient. For instance, there was the question whether the 'switch over' to RVSM would cause additional hazards to the ongoing air traffic[39] [Tiemeyer 2003]. Therefore a RVSM safety policy was developed, and a pre-implementation safety case was conducted to assure that (i) all identified hazards and risks were managed and mitigated, (ii) the collision risk met the ICAO Target Level of Safety and (iii) States showed that they would safely implement RVSM through the development of national safety documentation [Eurocontrol 2001a]. Consequently, the three main deliverables of the RVSM Safety activities were a functional hazard assessment, a collision risk assessment and national safety plans.

Research approach

Only the functional hazard assessment and the collision risk assessment will be discussed here. The functional hazard assessment was conducted as a series of brainstorming sessions to identify and classify all potential hazards. Operational experts participating in the sessions were pilots and air traffic controllers. In addition to the operational experts, representatives from the RVSM Program were present at the sessions. Each listed hazard was then classified according to severity and probability. Both severity and probability were subjectively estimated by the experts.

For collision risk, two specific safety targets have been defined by ICAO, a Target Level of Safety (TLS) for technical vertical risk, i.e. risk due to height keeping performance of the autopilot and the flight crew, of 2.5×10^{-9} fatal accidents per flight hour, and a TLS for total vertical risk, i.e. including the risk resulting from blunders like mishearing the assigned altitude, of 5×10^{-9} fatal accidents per flight hour. ICAO prescribes the use of the Reich collision risk model [Reich 1966a, 1966b, 1966c] for assessing vertical risk:

[39] For instance, if the introduction of RVSM would result in a significantly larger number of flight level changes for each flight, fuel consumption would be higher which could result in an increased likelihood that an aircraft runs out of fuel. Another hazard could be any interference of the RVSM equipment with on-board instrumentation.

$$N_{az} = 2P_z(S_z)P_y(0)\left[\begin{array}{l} n_z(same)\left\{1+\frac{2\lambda_x}{2\lambda_y}\frac{\overline{|\dot{y}|}}{\overline{|\Delta V|}}+\frac{2\lambda_x}{2\lambda_z}\frac{\overline{|\dot{z}|}}{\overline{|\Delta V|}}\right\}+ \\ n_z(opp)\left\{1+\frac{2\lambda_x}{2\lambda_y}\frac{\overline{|\dot{y}|}}{2\overline{V}}+\frac{2\lambda_x}{2\lambda_z}\frac{\overline{|\dot{z}|}}{2\overline{V}}\right\} \end{array}\right]$$

Parameter	Definition
N_{az}	The expected number of fatal aircraft accidents per flight hour due to the loss of vertical separation
S_z	The vertical separation minimum
$P_z(S_z)$	The probability of vertical overlap for aircraft nominally flying on adjacent flight levels
$P_y(0)$	The probability of lateral overlap for aircraft nominally flying the same route
$n_z(same)$	The frequency with which same direction aircraft on adjacent flight levels of the same route are in longitudinal overlap
$n_z(opp)$	The frequency with which opposite direction aircraft on adjacent flight levels of the same route are in longitudinal overlap
$\overline{\lvert\Delta V\rvert}$	The average of the absolute value of the relative along-track speed between two same direction aircraft flying at adjacent flight levels of the same route
$\overline{V}$	The average ground speed of a typical aircraft
$\overline{\lvert\dot{y}\rvert}$	The average of the absolute value of the relative cross-track speed between two typical aircraft flying at adjacent flight levels of the same route
$\overline{\lvert\dot{z}\rvert}$	The average of the absolute value of the relative vertical speed between two typical aircraft which have lost S_z feet of vertical separation
λ_x	The average length of a typical aircraft
λ_y	The average width of a typical aircraft
λ_z	The average height of a typical aircraft

The equation expresses accident probability as a function of the probability of vertical and lateral overlap for aircraft flying at adjacent flight levels and on the same route respectively, together with kinematic factors (relative speeds and aircraft dimensions). The probability of overlap is calculated from observed probability distributions. Data was collected by dedicated data collection infrastructure comprising ground based and portable airborne units. In total more than 350,000 measurements were made [Eurocontrol 2001a].

Discussion

It is interesting to note that for this problem the available standard approach (the ICAO collision risk model) was not considered sufficient because of the possible existence of new, previously not existing hazards. The ability to handle ‘new’ hazards has been defined as a desirable characteristic of a causal risk model. Both the functional hazards assessment and the collision risk assessment were strictly focussing on collision risk. Even within this limited scope there was no explicit attention for system interdependencies and the approach therefore cannot be considered ‘integrated’. It was for instance assumed that the risk of

collision in an RVSM environment with TCAS[40] would be lower than the risk of collision without TCAS, thereby ignoring the possibility that interactions of TCAS with other system elements can introduce additional risk[41]. Some system interdependencies will have been implicitly considered by the experts in the functional hazard assessment. The results of the functional hazard assessment were qualitative. Probabilities of occurrence of events were subjectively estimated by experts and were expressed semi-quantitatively, using terms like 'remote', but the results of the collision risk assessment were qualitative. This part of the study required an extensive data gathering campaign. The collision risk assessment is not transparent to non mathematicians because the outcome is the result of a mathematical equation that requires probability distributions as input parameters. For operational managers and safety experts in the air transport system this may be difficult to comprehend. The functional hazard assessment is not fully transparent, because it is not completely traceable what the expert's opinions were based on. Reproducibility of the results is questionable, as different experts may have had different opinions and arrived at different conclusions.

5.5. VEMER ATM System increment 2002

Background

The method of VEM[42] Effect Reports (VEMERs) has been developed by ATC the Netherlands in order to support developments with respect to the ATM system with appropriate management information. Every proposed change in the ATM system is evaluated for its consequences on safety, efficiency of the operations, and environmental aspects. The results are published in a VEMER. A distinction is made between increment VEMERs, which focus on the entire operation, and in-depth VEMERs, which focus on envisaged partial operations in detail. Increment VEMERS are conducted at a higher level of abstraction than in-depth VEMERS.

The research question

The objective of the VEMER 'ATM system increment 2002' [LVNL 2002] was to provide information on the envisaged introduction of the 5th runway at Schiphol in November 2002. The purpose of the study was to support decision making within ATC the Netherlands and to support external arrangements with the stakeholders (like KLM, the Dutch government and the airport authority). The VEMER ATM system increment 2002 was not aimed at providing a full safety assessment. This was done later on when VEMERs of specific operations were made.

Research approach

Each VEMER starts by defining the objective of the project in terms of safety, efficiency and environment -the VEM targets- followed by describing specific sets of options for changing the ATM system with which it is envisaged that the VEM targets will be met. VEM effects are determined for each set of options. Designing options and deriving VEM effects is an iterative process; more detail is added when it becomes available in the course of the project. The process stops as soon as the results of the VEMER are considered

[40] TCAS stands for Traffic Collision Avoidance System. It is an airborne system that warns the flight crew of imminent collision risk and provides instructions to avoid collisions.
[41] That such interactions can occur with catastrophic results was demonstrated with the Überlingen accident. Two aircraft collided when the instructions provided by TCAS were conflicting with instructions provided by the air traffic controller [BFU 2004]. See also infobox 'Inappropriate interface: the Überlingen mid-air collision'.
[42] VEM stands for Veiligheid, Efficiency, Milieu (Safety, Efficiency and Environment).

acceptable. The question, scope and level of detail are determined before each iteration cycle. Evaluation of the VEM effects focuses on those elements of the ATM concept that have substantial influence. For increment VEMERs the effects are classified by operational experts as Large, Medium and Small. For in-depth VEMERS the corresponding categories are Unacceptable, Tolerable and Small. The VEM effects are then quantified on the basis of expert judgement or studies. The 'ATM system increment 2002' VEMER was quantified on the basis of expert judgement. Information was derived from structured interviews with experts representing different operational functions (tower/approach controller, ground controller, clearance delivery, air traffic control assistant). For other VEMERs interviews have also been conducted with airline pilots.

Discussion

The VEMER approach has the potential to address the total aviation system including interdependencies between different system elements. However, these interdependencies are not actively pursued. The focus of the analysis (in terms of both the changes and the effects of those changes) is on risk that can be influenced by the ATM part of the aviation system. The results of the analysis are quantitative, where quantification is done by expert opinion. The process of expert judgement is well developed. Strong points include the traceability of the process, confidentiality and the existing infrastructure; both analysts and experts are conveniently familiar with the process of expert judgement elicitation because it is used quite frequently. Weak points are the traceability of the combination of experts' opinions and the reproducibility of the process [Roelen et al 2004a, Van der Plas & Van Luijk 2005]. Because of these weak points there is a lack of transparency of the overall project result, which is unfortunate as one of the objectives of the method is to support external arrangements with the stakeholders. Other than the need to perform structured interviews with experts there is no additional need to collect data.

5.6. Conclusions for this section

In this section we examined different methods that are currently being used to assess or analyse the safety of (parts of) the aviation system. Specific attention was paid to whether the method looks at the integrated system or only an element of the system in isolation, to data requirements, reproducibility of the results, and transparency.

There is no standardized or prescribed procedure for conducting safety analyses for air transport. Selection of the safety analysis methodology for a specific problem depends on the scope of the safety question, the type of operation, the level of involvement of the customer, the domain knowledge of the customer and the usage of the results. This was also concluded by Van Doorn [2006] in an analysis of several ATM safety studies. Although the different methods often follow similar steps, they differ in the details. One of the consequences is that quite often safety analysis starts with an activity that has already been performed earlier, although slightly differently. Activities such as hazard identification and accident scenario development are repeatedly done, often without taking full advantage of the results and lessons learned from previous studies. Data collection is also an activity that sometimes requires a significant effort. Often data are not available or there are not sufficient resources to collect data. Most safety analyses described here therefore relied heavily on the experts, or more precisely, on the black boxes inside the expert's heads, for hazard identification, accident scenario development and quantification. Each expert has his own 'model of the world' which he uses to come to certain conclusions. These models were not made explicit in the studies reviewed here. When consensus was achieved among a group of experts, this consensus was only concerning the outcome, and not concerning the way these outcomes had been derived. The background for reaching a certain conclusion

can vary among experts. Consensus can also be influenced by group dynamics. Extrapolating on the basis of the consensus decisions is risky, because nothing is known about how each expert has given thought to possible scenarios or consequences. Criteria for experts did not exist in the studies mentioned. In aviation safety analyses, selected experts are often people with (ample) operational experience, such as pilots and air traffic controllers. Analytical ability and ability to provide judgement on future operations are only rarely taken into consideration. Experts are not 'calibrated'. These aspects have a negative effect on the transparency and reproducibility of analysis results.

It is no coincidence that all the examples presented here address issues of air traffic management or airports. These two subjects are, in comparison with aircraft design, operation and maintenance, far less regulated. Air navigation service providers and airport authorities are relatively free in developing infrastructure, procedures etc., that fulfil their specific needs. However, they still need to demonstrate that all applicable target levels of safety are met. As a result, there is a growing need for support in conducting risk assessments in the field of ATM and airports. In aircraft design, operation and maintenance, there is less freedom. It is assumed that compliance with existing regulation will ensure that (minimum) safety targets are met. This difference between ATM and airports on the one side and aircraft design, operation and maintenance on the other has developed historically (see also section 4.4). As explained in section 4.4, it can be expected that regulatory requirements will in the future be developed for the integrated system rather than the individual components, but the complexity of the matter will discourage the authorities from specifying detailed requirements. It can therefore be expected that also for aircraft design, operation and maintenance, a need for support in conducting risk assessments will emerge.

None of the methods described in this chapter fulfils all requirements that have been derived in the previous chapter: ability to analyse the integrated system including dependencies, reproducibility of results, transparency, ability to provide quantitative results and the ability to represent current as well as future accident scenarios. The customers for these studies would have had better (and cheaper) answers to their questions if there had been an existing causal risk model. This chapter also showed that the ability to handle 'new' hazards would be a desirable characteristic of a causal risk model. An effort to develop a method that meets these criteria therefore seems fully justified.

Chapter 6. Risk models in other industries

There are various ways in which risk modelling approaches can be compared. For instance one can look across various scientific disciplines, or across different countries, or across different industries, etc. In this thesis the comparison is restricted to a comparison across industries. This chapter investigates the role of causal risk models in the nuclear power industry, manned spaceflight, the offshore industry, the process industry, rail transport and health care. Characteristics that seem to be vital for the success or failure of causal risk models are identified. The final section of the chapter concludes on the relevancy of it all for the air transport industry.

6.1. Nuclear power

Initially, the nuclear power industry relied on the use of multiple barriers and large design margins to assure the safety of nuclear reactors. During the 1960s the general public became more worried about the hazards of radiation as a result of expert concern and accidents such as Windscale[43], and there was a desire from the public to know whether nuclear power plants were sufficiently safe. In roughly the same time frame, the theoretical basis was provided for applications of reliability methods in the risk assessment of complex systems. The desire to meet risk targets and to quantify and to evaluate the effects of design improvements of nuclear power plants led to the introduction of probabilistic risk models in the nuclear industry. The methods for probabilistic risk analysis that were used originated from the aerospace industry (fault trees) and decision theory (event trees). These techniques are very suitable for the representation of hardware failures [Ericson 1999]. The first full-scale application of these methods to a commercial power plant was undertaken in the Reactor Safety Study WASH-1400 published by the American Nuclear Regulatory Commission NRC in 1975 [NRC 1975]. This landmark study was controversial and an independent evaluation of the study [Lewis et al 1979] was requested. The evaluation report concluded that the WASH-1400 study had important shortcomings but provided the most complete single picture of accident probabilities associated with nuclear reactors. The use of fault trees and event trees coupled with an adequate database was considered to be the best available tool to quantify these probabilities. The lack of scrutability of the calculation process and lack of data on which to base the component reliability estimates were mentioned as shortcomings of the study. The evaluation also revealed shortcomings in relation to the representation of human factors and the role of the organisation. The Three Mile Island accident in 1979[44] and the subsequent investigation of it underlined the value of probabilistic risk models, because some of the characteristics of the accident had been highlighted by the WASH 1400 study [Kemeny 1979]. When the PRA procedures guide

[43] On October 10, 1957, the core of a British nuclear reactor at Windscale caught fire. As a result, substantial amounts of radioactive contamination were released into the surrounding area.

[44] The accident at the Three Mile Island nuclear power plant near Middletown, Pennsylvania, on March 28, 1979, was the most serious in U.S. commercial nuclear power plant operating history, even though it led to no deaths or injuries to plant workers or members of the nearby community. The accident led to a partial meltdown of the reactor core but only very small off-site releases of radioactivity [NRC 2007a].

[NRC 1983] was published by the NRC in 1983 it became a main reference. Probabilistic risk models became even more accepted when in 1995 the NRC issued a revised policy statement on the use of probabilistic risk assessment which stated that "The use of PRA technology should be increased in all regulatory matters" [NRC 1995]. The results of probabilistic risk assessments are now used together with traditional deterministic analysis for regulatory decision making, the so-called risk-informed regulation [Keller & Modarres 2005]. Traditionally, PRAs of nuclear powerplants take account of technical systems and human operators but do not include an explicit representation of the possible impact of plant organisation and management on the safety performance of equipment and plant personnel. As a result it is acknowledged that these PRAs may not have accounted for the contributions of organisational facts to total risk [Mosleh & Goldfeiz 1994]. A first attempt to resolve this issue is the work process analysis model (WPAM) which represents the dependencies among the parameters that are introduced by organisational factors by calculating new (organizationally dependent) minimum cut-sets for major accident sequences [Davoudian et al 1994]. There is general recognition that organisational factors need to be evaluated further for their contribution to plant safety performance [NEA/CSNI 1999] and much of the current research work is still focussing on this topic.

6.2. Manned spaceflight

When US President Kennedy in 1961 set his ambitious goal[45] of going to the moon, the total US experience in manned space flight consisted only of Alan Shepard's 15 minute suborbital flight in a Mercury capsule [Chaikin 1994]. The lunar program was tremendously bold and ambitious and NASA management was aware that identification of potential failures and their risks was essential to a successful design and thus to a successful mission. Quantitative numerical goals were set for the Apollo missions: a risk of 1 out of 100 per launch was considered acceptable for mission non-completion and 1 out of 1000 per launch was set for loss of the crew. Quantitative risk models were developed for the different components of Apollo (the command module, the moon lander and the Saturn V rocket) to determine if these goals could be met [Kelly 2001]. As described in the previous section, the theoretical basis for applications of reliability methods in risk assessment of complex systems had by then (the 1960's) just been developed. However, data to populate quantitative models was simply non-existent. Because Apollo was still very much in the design stage, and due to the time pressure to meet Kennedy's goal (and also to beat the Russians in the race to the moon) NASA fell back on familiar qualitative risk methods, in particular FMEA. Individual components were analysed bottom-up, and those that could put the mission at risk were put on a Critical Item List (CIL). Abundant personnel and test facility resources were available to ensure that each critical item was properly addressed. The success of Apollo and the subsequent (initial) success of the Space Shuttle made NASA fully comfortable with the FMEA/CIL approach. It was only after the accident with the Space Shuttle Challenger in 1986 [Shayler 2000] and the subsequent accident investigation report [Rogers 1986] that probabilistic risk assessment methods were (re)-introduced at NASA. The methods used were similar to those in the nuclear industry, i.e. combinations of fault trees and event trees. Yet there was still much resistance. It was stated that the money spent in quantitative safety analysis could be better invested in improving the system directly. Because of the many assumptions that had to be made in the models, the results of the analyses were considered to be too subjective to be used for decision

[45] "I believe that this nation should commit itself to achieving the goal, before this decade is out, of landing a man on the Moon and returning him safely to the Earth.", John F. Kennedy, 25 May 1961, "Urgent National Needs" speech to a Joint Session of Congress in Washington DC.

support [Paté-Cornell & Dillon 2001]. Endorsement grew when probabilistic risk models had been used for safety assessments of Shuttle payloads (such as the Galileo and Ulysses spacecraft[46]) and had produced results that were favourable to the program (the consequence of the results was that the launches could proceed on schedule). Following the accident with the Space Shuttle Columbia in 2003, the accident investigation board concluded that the Space Shuttle program hazard analysis was performed on components and elements, but was not required on the Shuttle as a whole. Since it was designed for bottom-up analysis, it could not effectively support the kind of top-down hazard analysis that was needed to inform managers on risk trends and to identify potentially harmful interactions between systems. The Board therefore recommended to conduct *integrated* hazard analysis, i.e. not only anticipating failure modes of individual components but also considering the components integrated into a total system and working in concert [Columbia Accident Investigation Board 2003]. Probabilistic safety analysis has now been adopted as one of the decision support instruments for the management of the Space Shuttle, the International Space Station and some unmanned missions. The most important issues that require further attention are the need for an overarching model, consistency in the choice of methods, analytical depth and the treatment of data [Paté-Cornell & Dillon 2001, Fragola 1996].

6.3. Offshore industry

The world's first formal requirement for probabilistic safety analysis in the offshore industry was by the Norwegian Petroleum Directorate in their 'Guidelines for a safety evaluation of the platform concept' of 1981. The analyses initially focused on the availability of safety functions such as escape routes and shelter areas, but were later extended into total risk analysis. From 1990 onwards the Norwegian Petroleum Directorate required operators to manage safety systematically, using probabilistic risk assessment methods as a tool and defining their own safety targets and risk criteria [Brandsaeter 2002].

Developments in the UK were roughly similar. Probabilistic risk assessments methods were applied from the 1980s onward, but only to specific aspects of the design of offshore installations. Like in the nuclear power industry and in manned spaceflight, it was a major accident that led to a more prominent role for quantitative risk analysis: the Piper Alpha disaster[47]. Lord Cullen's Report on the Public Inquiry into the Piper Alpha disaster, issued in 1990, recommended that the operator or owner of every offshore installation should be required to prepare a Safety Case, which should include identification and assessment of the risks of major accidents, and submit it for acceptance by the HSE [DNV 2002a]. The use of risk assessment in the offshore industry was firmly established in the 1990s, particularly in the UK where many aspects became subject to full risk assessment, notably

[46] The objective of Galileo was to explore the planet Jupiter. It was launched on October 18, 1989, from the Space Shuttle Atlantis [Meltzer 2007]. Primary mission of the unmanned Ulysses spacecraft is to investigate properties of the heliosphere as a function of solar latitude. The craft was launched on October 6, 1990, from the Space Shuttle Discovery [Forsyth et al 2002]. Both spacecraft used a small amount of plutonium-238 as fuel, and the possibility of accidental release of plutonium into the atmosphere following a launch mishap was a cause of great concern.

[47] On July 6, 1988, a fire and subsequent explosions completely destroyed the Piper Alpha oil production platform. The fire, resulting from an oil condensate leak, affected and melted the gas pipes, which then caused the explosions. A total of 167 people died making it the world's worst offshore oil disaster.

the Temporary Refuge[48] assessment which was mandated to be analysed using Quantitative Risk Assessment. The specific tool of QRA was the subject of criticism in a HSE commissioned survey as to the effectiveness of current offshore regulations. The technique was considered too mathematical and there was insufficient agreement within the industry and the HSE on how to use the results of QRA. Nevertheless, HSE's Guide to the Offshore Installations (Safety Case) Regulations 1992[49] states [HSE 1998]: "The evaluation of risk should involve both a qualitative and quantitative approach. Where relevant good or best practice is clear, the balance should be in favour of qualitative arguments to show that the risks have been properly controlled. Where relevant good or best practice is less clear, appropriate support from quantitative arguments will be necessary". Similarly, the UK oil and gas industry has developed a framework to assist risk-related decision making (Figure 12), which helps decision makers choose an appropriate basis for their decisions. The framework takes the form of a spectrum of decision bases, ranging from those decisions dominated by pure engineering concerns to those where company and societal values are the most relevant factors. The concept is not unlike the risk informed regulation in the nuclear power industry. This approach shows that quantitative risk assessment has a major input to the types of decisions that involve some uncertainty, deviation from standard practice, risk trade-offs, etc.

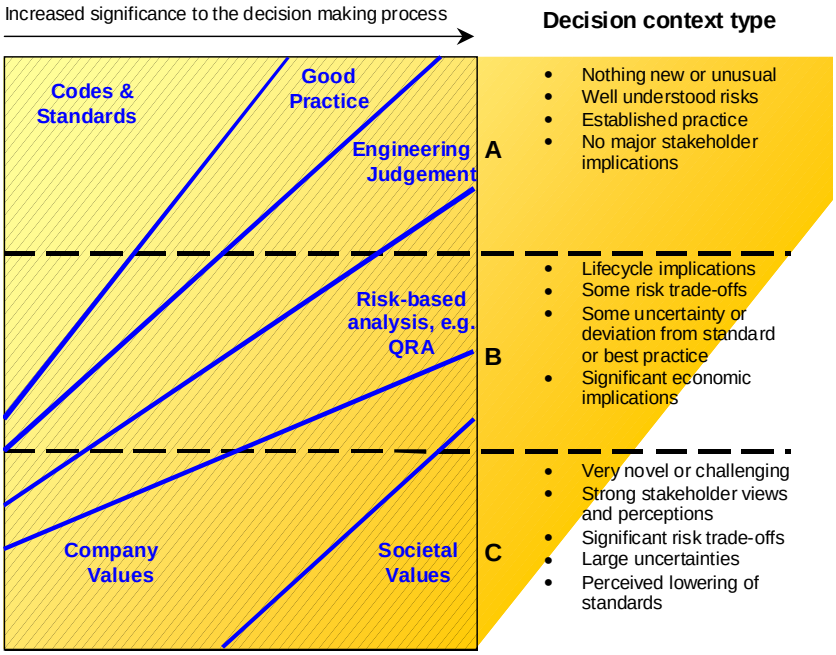


Figure 12: Risk-related decision support framework. Source: [UKOOA 1999].

[48] A Temporary Refuge is a specially designated area on board every offshore oil and gas production platform which is designed to ensure that personnel can muster in safety for a period of protection while an incident is being assessed and a decision taken on whether or not to abandon the installation.

[49] The Regulations implement the central recommendation of Lord Cullen's Report that the operator of every offshore installation should prepare a Safety Case.

Techniques that have been used in the offshore industry for estimating the frequency of occurrence of failure causes include historical accident frequency data, fault tree analysis, event tree analysis, simulation, human reliability analysis, expert judgement and Bayesian analysis [DNV 2002a]. Historical accident frequency data is the basis for most quantitative safety analysis in the offshore industry.

6.4. Process industry

On June 1, 1974, in a chemical plant in Flixborough, UK, a crack in a temporary bypass pipe caused the accidental release of a vapour cloud of cyclohexane. The vapour cloud exploded, as a result of which 28 people were killed. After this disaster it was recommended that some method should be developed and used to estimate realistic accident consequences and to associate these estimates with a probability [Ale 2005]. On November 17, 1975, an explosion occurred at the DSM works in Beek, the Netherlands, due to the accidental release of propane gas. There were 14 fatalities and many injuries [Ale 2005]. Then in 1976 a vapour cloud containing dioxin was accidentally released from a chemical plant near the town of Seveso in Italy. More than 600 people had to be evacuated and as many as 2000 were treated for dioxin poisoning. These accidents motivated the development of Council Directive 82/501/EEC [EC 1982] on the major-accident hazards of certain industrial activities, better known as the Seveso Directive. For the prevention of major accident hazards involving dangerous substances, the European Commission issued the so-called Seveso I and later Seveso II Directive. The Directive prescribes that plant operators need to establish a Safety Management System that must address identification and evaluation of major hazards, adaptation and implementation of procedures for systematically identifying major hazards arising from normal and abnormal operation and the assessment of their likelihood and severity [EC 1996].

Because it was known that the results of risk analysis may vary considerably depending on the methodology used and the assumptions underlying the analysis, national and international guidelines to standardise the approaches have been developed, like for instance the Guidelines for Chemical Process Quantitative Risk Analysis [CCPS 1989]. The Dutch government made an attempt to standardise the methodology for risk analysis of sites involving flammable or toxic substances. Methods to calculate probabilities were published in the ‘coloured books’; methods for calculating probabilities in the Red book, methods for calculating consequences in the Yellow book, methods to calculate damage in the Green book and guidelines for performing risk analysis in the Purple book. But results of a benchmark study [Ale et al 2001] showed that, depending on the interpretation and the method used, differences of up to one order of magnitude were still being obtained. The Dutch government therefore decided on a prescribed and unified calculation method. Safeti-NL, developed by DNV in London, was selected as the package for determining whether the operator complies with criteria for third party risk.

6.5. Rail transport

Historically, safety in railway companies was managed from a technical and rule-based point of view. The main focus was on reliability of infrastructure and rolling stock and technological solutions for the prevention of ‘human error’ and a system of rules and procedures that eliminated (in theory) the need for independent decision making by operators in the field [Hale 2000]. The privatisation of railway companies that occurred in several countries in Europe in the mid-nineties and the European Commission’s policy to encourage liberalisation of the railway industry across all EU member states resulted in significant changes. Railway infrastructure, traffic control, maintenance, etc., were separated and became the responsibility of individual companies. Competition, which

previously did not exist, started to play a role. Performance requirements, including requirements for safety performance, were introduced in European and national regulation[50]. Unlike the nuclear power or offshore industry or manned spaceflight there was no single major accident that forced regulatory requirements, but accidents are also in the railway industry a driver for change, for instance the recommendations for changes to the regulatory requirements for railway safety set out by Lord Cullen [2001] in the inquiry on the accident at Ladbroke Grove[51]. These changes also led to a different approach towards safety management. Prorail, a provider of rail infrastructure in the Netherlands, has developed a safety management system [Heimplaetzer & Busch 2006] that is based on the quality control cycle [Deming 1990]. Safety performance is measured with the use of a monthly updated database of accidents, incidents and other safety related parameters such as audit results [Wright 2006]. Safety critical processes were also defined. (Causal) modelling is used to identify likely system factors or elements which could be considered as precursors to future accidents. According to the Dutch policy for rail safety (Tweede Kadernota voor de veiligheid van het railvervoer in Nederland), all design plans and resolutions on rail transport should explicitly involve safety. Preferably, this should be done by means of an Integral Safety Plan or a Safety Case. On a European level, the European Railway Safety Directive [EC 2004] is aimed at ensuring the development and improvement of safety on the European Community's railways and improved access to the market for rail transport services. According to the Directive, common safety methods should be gradually introduced to ensure that a high level of safety is maintained and, when and where necessary and reasonably practicable, improved. 'Common safety methods' are the methods to describe how safety levels and achievement of safety targets and compliance with other safety requirements are assessed by elaborating and defining the following:

- Risk evaluation and assessment methods,
- Methods for assessing conformity with requirements in safety certificates and safety authorisation,
- Methods to check that the structural subsystems of the rail systems are operated and maintained in accordance with the relevant essential requirements.

In line with the tradition in the rail industry, the EU directive still puts much emphasis on compliance with regulations, but the gradual introduction of risk assessment methods is otherwise similar to what is happening in other industries as described previously.

6.6. Health care

The health care industry is interesting to compare with aviation and rail because these industries are all responsible for managing the safety of their 'customers', i.e. patients and passengers respectively. This is a marked difference with e.g. the off-shore industry where safety is to a large extent synonymous to occupational safety and third party risk. As in air transport, customer safety is the main driver for safety improvement in the health care industry, rather than occupational safety. But there is also a different risk dynamic. The choice for a patient is often certain death now or possible life later, which is the reverse that

[50] For example, the 'Tweede Kadernota voor de veiligheid van het railvervoer in Nederland' of November 2004 states a goal of less than 1.5 passenger fatalities per 10 billion passenger kilometres.

[51] On 5 October 1999 at Ladbroke Grove junction, about two miles west of Paddington Station, London, there was a head on crash at high speed between trains operated by Thames Trains and First Great Western. This caused the death of 31 persons including both train drivers, and inflicted injuries, some of them critical, on over 400 other persons.

airline passengers face. A complicating factor in considerations on health care safety is the fact that it is even less tangible than air transport safety or nuclear safety. We use the words 'health care risks' for injuries resulting from a medical intervention, but not for those due to the underlying condition of the patient. In transportation and other industries, safety is usually 'measured' by the numbers of dead and injured, but measuring the safety of the health care industry is not so straightforward. Patients in hospitals are already 'severely injured' and some patients will die even when they receive 'safe' health care. Whether or not such deaths are *preventable* is often difficult to determine. If a patient has surgery and dies from pneumonia he or she got postoperatively it could be the result of poor hand washing by the surgeon (preventable). But the patient could also have carried the infection already before the operation where the surgery and resulting weakening of the immune system allowed full development of the pneumonia (not or at least less preventable). This is one of the reasons why data on health care safety are difficult to obtain, and are at best estimates. This may also be one of the reasons why probabilistic risk assessment is not commonly performed in the health care industry. Yet causal models are routinely and successfully used, not for safety assessment but for diagnostic analysis. Bayesian Belief Nets (see section 7.2.2) are very suitable for diagnostic analysis, as the models take advantage of the ability to update the model when observations have been made. The representation of the models as influence diagrams is considered to be helpful when communication among experts of probabilistic relationships is important [Burnside et al 2006, Owens et al 1997, Onisko et al 1999]. The latter observation is a relevant lesson for causal modelling of air transport safety. A causal model can indeed be a tool for communication, not from the government to the general public, as the Dutch Ministry of Transport initially envisioned (see Chapter 4), but for communication amongst experts.

Health care safety: Going to a hospital is as risky as going to the Moon.
The estimates that we have paint a sombre picture of the safety level of health care. According to Corrigan et al [2000], the probability of death due to preventable injuries resulting from medical intervention in U.S. hospitals in 1997 was approximately 1 in 1,000 hospitalisations. This death rate per exposure is equivalent to the safety objective set by NASA in the 1960's for crew loss in spaceflights to the moon and three orders of magnitude worse than that in commercial aviation or the nuclear power industry. According to Amalberti et al [2005] the failure of the health care industry to achieve a level of safety that is comparable with commercial aviation and the nuclear power industry is caused by the culture of safety. The health care industry is characterised by chronic shortage of staff, systematic overtime working, overloaded work schedules and excessive fatigue on the job [Van der Heijden 2006]. Surgeons have the status and self-image of a craftsman rather than adopting a position that values equivalence among their ranks. In the early days of aviation, pilots also had such a 'craftsman' aura. This image was maintained by 'heroes' such as the crew of the KLM Douglas DC-2 'Uiver' that won the London-Melbourne race in 1934 [Nater 1983], but disappeared when air transport became a mass product and passengers are now accustomed to the notion that all pilots are equivalent to one another in their skills. In modern surgery, such 'craftmanship' still exist. Historical and cultural precedents and beliefs on performance and autonomy have resulted in an attitude to obtain a specified high level of production, no matter what it takes. The low level of safety does not arise from incompetence but from the experts who challenge the boundaries of their own maximum performance [Amalberti et al 2005]. The failure to accept that there are limits to maximum performance is one of the barriers that prevents the health care industry from achieving 'ultrasafety'.

6.7. Conclusions for this section

The purpose of most risk models in other industries is primarily to be able to conduct probabilistic risk analysis of the system including technical elements as well as human operators and management systems. Inclusion of human and organisational factors in the models is considered essential because they have a strong effect on safety and are, to some extent, easily influenced by managerial decisions.

The way in which probabilistic risk models were introduced in the nuclear power industry, manned spaceflight, and the oil and gas industry followed a similar pattern. Traditionally, risk assessment methods were deterministic and qualitative. Probabilistic risk analysis methods were occasionally applied but were not seen by everyone as an improvement over the traditional methods. Lack of data to populate the models, lack of transparency and failure to adequately represent human and organisational performance were mentioned as weak points. However, almost within a single decade several major accidents[52] resulted in regulatory changes that forced the introduction of probabilistic risk models for system-wide analysis. Even then it was a slow process, and it took a long time to develop methods and models that are considered acceptable. It is recognised that these methods should not be used as a substitute for the traditional approaches, but as a complement. Traditional approaches are fail safe design and 'defence in depth', detailed regulatory requirements to assure safety and applying competent judgement to determine that a risk is acceptably low. A probabilistic, model based approach allows representation and analysis on dependencies and interactions among system components, including human operators. A probabilistic model based approach can identify possible flaws in the fail safe design or defence in depth concept, (e.g. generated by common cause failures), it can be used in cases where detailed regulatory requirements are absent but instead a target safety level is posed, and it is informative when the complexity practically prevents the application of 'competent judgement' for safety analysis. The combination of qualitative, deterministic and quantitative probabilistic methods allows risk informed regulation and decision-making. Yet there is still sometimes resistance, when people believe that the money spent in quantitative safety analysis could be better invested in improving the system directly. Because of the many assumptions that have to be made in the models, the results of the analyses are at times considered to be too subjective to be used for decision support. It is only when people realise that there are no alternatives, that current decision making is even more subjective, and that complex integrated systems require systematic and consistent analysis, that people come to appreciate the value of causal risk models. Traditionally, probabilistic risk models take account of technical systems and human operators but do not include an explicit representation of the possible impact of organisation and management on safety performance. It is generally recognised that organisational factors should be included, but analysing and modelling the complexities of a human organisation has proven to be a difficult task (see section 9.2 for a more elaborate discussion on this topic). To get rid of the variation in results of risk analysis, the method to be applied is sometimes prescribed in regulation.

Based on these experiences it could be expected that introduction of causal risk models in air transport will be similarly jerkily and it can take decades before the industry has fully accepted the use of quantitative risk models, even though risk models are, for some specific

[52] In the process industry is was the Seveso calamity (1976), in the nuclear power industry it was the Three Mile Island accident (1979), in manned spaceflight the Challenger explosion (1986), and in the oil and gas industry the Piper Alpha disaster (1988).

problems, already well accepted in air transport (for instance the ICAO collision risk model, see section 4.3.6). Data to populate the model, transparency of the models and the representation of human and organisational influences are important to promote and support the development. It is also essential to state from the outset that the models should be used as a complement rather than as a replacement of current practice. The use of diagnostic models in the health care industry demonstrates that such models are indeed suitable instruments for communication among experts on probabilistic relationships and to provide a coherent representation of domain knowledge under uncertainty.

Chapter 7. Modelling

A causal risk model is a mathematical object. It can be regarded as being composed of a representation scheme or conceptual model and a computational engine. Several methods exist that can be used for model representation, and these all have different mathematical properties and limitations. This chapter addresses the characteristics of the most important and most frequently used modelling techniques. It starts with a definition of a causal model as adopted in chapter 3. The relevancy of the model representation is discussed. The relevant modelling methods are described, including a brief overview of strengths and weaknesses. The information is then used to draw a conclusion on which modelling technique or set of techniques is most suitable for aviation causal risk models by comparing their (mathematical) properties with the causal model requirements.

7.1. Model representation

According to the definition adopted in section 3.7, a causal model is *a mathematical object that provides an interpretation and computation of causal queries about the domain* [Galles & Pearl 1998]. What does this mean? First, a causal model is a *mathematical* object. This means that building causal models will inevitably require some mathematics. But it does not necessarily imply that the output of the model is quantitative. Second, the *domain* must be specified. In our case the domain is aviation safety, but obviously this domain, or the scope, must be defined more strictly[53]. Third, the model provides an *interpretation*. It is just one way of looking at reality; there may be others that are different without being necessarily wrong.

The causal model definition of section 3.7 also implies that a causal risk model can be regarded as being composed of a representation scheme or conceptual model and a computational engine [Van Waveren et al 1999, Labeau et al 2000]. The representation scheme is the interface between the computational engine and the user of the model. The representation scheme is the primary means of communicating the model to the outside world [Van Waveren et al 1999]. Many users will regard the representation scheme as 'the model'. The representation scheme must therefore be comprehensive and, most importantly, transparent without being an oversimplification. Representation schemes can be useful for verification of the model's internal consistency, its applicability to the actual system being modelled and as an aid to the modelling process itself [Labeau et al 2000]. In a classical probabilistic risk assessment the representation scheme consists of accident scenarios in the form of of Event Sequence Diagrams and fault trees. This representation was selected in the WASH 1400 study (see the section on nuclear power, 6.1) and also in NASA's software tool for quantitative risk assessment [Groen et al 2006]. The accompanying computational engine is mathematically simple, combining only Boolean algebra and basic probability theory. The price that has to be paid for this are the simplifications and compromises which need to be made to represent reality as a static series of Boolean expressions. To overcome this limitation other methods have been developed that are more complex than fault trees and event sequences and are

[53] E.g. military or civil aviation, commercial or non-commercial, primary and/or subsidiary processes (see also Appendix B).

computationally more intensive. A drawback of these advanced methods is that the corresponding representation schemes do not always provide an intuitive understanding of the scenarios (see for instance the sections on Bayesian belief nets and Petri nets in this chapter). The classical fault tree / event sequence diagram approach on the other hand is self-explanatory[54].

The representation schemes of models used for safety assessments of nuclear powerplants are, at the highest level of abstraction, relatively simple. This is achieved by the methodology (event sequence diagrams) and by a functional or physical decomposition. The top level of the model represents only top level functions or physical components. Compared to the aviation system, a nuclear powerplant is a simpler system. It has only a few inherent classes of design features that will keep it working, and a few that will cause problems. The fundamentals involved (in accident scenarios) do not exhibit many feedback loops or multiple paths for recovery. In comparison with a nuclear powerplant, the aviation system has more design features that keep it working and that can cause problems, and there are more feedback loops. The greater role of the human operator (pilots, ATC controllers, maintenance technicians) and their interactions is one of the main causes of the greater complexity. Because of this greater complexity, there may be a need for a second high-level decomposition of the aviation system. In the probabilistic risk assessment for the Space Shuttle, a second decomposition is made by considering the mission time-line and distinguishing the various operational phases of the shuttle (ascent, orbit, descent) [Groen et al 2006]. A similar approach in air transport would be a break down into flight phases: taxi out, take-off, climb, cruise, descent, approach, landing, taxi in and standing. But the operational phases of the shuttle and the phases of flight of an aircraft are not independent. Hazards that originate in one phase may create a risk in a later phase. An example is the accident with the Space Shuttle Columbia; a failure during launch resulted in a catastrophe during re-entry 17 days later (see infobox Space Shuttle Columbia accident). Whenever a decomposition is made, a mechanism must be in place that keeps track of all possible dependencies between components. A modelling method that allows easy handling of such interdependencies between decomposition elements is desirable.

Space Shuttle Columbia accident

The Space Shuttle

The Space Shuttle consists of an Orbiter, an External Tank and two Solid Rocket Boosters. The External Tank and Solid Rocket Boosters are only used during launch of the vehicle. The External Tank accommodates 543 m^3 of liquid oxygen and 1458 m^3 of liquid hydrogen. In order to keep the super-cold propellants from boiling and to prevent ice from forming on the outside of the tank while it is sitting on the launch pad, the External Tank is covered with a one-inch-thick coating of insulating foam.

The accident

Mission STS-107 was the 113th in the Space Shuttle program and the 28th trip into space of the orbiter Columbia. Launch occurred at the Kennedy Space Center, launch pad 39A, on January 16, 2003. Approximately 82 seconds after launch, pieces of foam came off the External Tank and damaged the left wing of Columbia. The orbiter was at an altitude of 65,860 feet, travelling at Mach 2.46 at the time of impact. The impact had no immediate consequences, the launch was otherwise uneventful and Columbia reached its planned orbit.

[54] While a fault tree is self explanatory, this does not necessarily mean that it is also correct. Correctness should be tested as part of validation (Chapter 10).

While Columbia was on-orbit, there was no indication of damage from the ascent foam impact. After a successful 17 day scientific mission, the crew of Columbia prepared the spaceship for deorbit, reentry and landing. Columbia successfully completed the deorbit burn over the Indian Ocean but during reentry the damage in the wing leading edge allowed hot gas intrusion onto the wing structure leading to extreme heating and eventual deformation of the left wing. While travelling at 200,700 feet and Mach 18, complete vehicle break-up occurred and all 7 crewmembers were killed [Columbia Accident Investigation Board 2003].

The representation scheme is basically only a picture. The computational engine is the translation of this picture into mathematical equations and operations, allowing the user of the model to perform analyses like sensitivity analysis or cut set identification. Because the computational engine is a (large) set of mathematical expressions it is not transparent to most engineers and safety managers. A software tool that combines the function of the representation scheme and the computational engine is therefore essential. Such a tool will also make configuration control[55], very important for complex models, a little less difficult.

7.2. Modelling techniques

This section provides a brief overview of techniques that can be used for causal risk modelling. The basic characteristics of the methods are described and compared with requirements for use of a causal risk model as derived in previous sections. The intention of this section is not to provide an exhaustive overview of available methodologies, but rather to describe the relevant and valuable ones that are currently being applied in aviation risk modelling. These methods are used or proposed in the CATS model [Ale et al 2006, 2007, 2008], Eurocontrol's IRP [Perrin et al 2006], the FAA causal model [Mosleh et al 2004, Roelen et al 2008], TOPAZ [Blom et al 2001, 2006, 2008] and NASA's Aviation System Risk Model (ASRM) [Luxhøj 2004]. A more exhaustive description of techniques is found in Labeau et al [2000], Bedford & Cooke [2001], Rouvroye & Van den Bliek [2002] and Everdij [2004].

7.2.1 Boolean Trees

Fault trees

A fault tree is a graphic model of the various parallel and sequential combinations of faults that will result in the undesired event. The undesired event constitutes the top event in the fault tree diagram. In fault tree analysis an undesired state of the system is specified and the system is then analysed in the context of its environment and operation to find all credible ways in which the undesired event can occur [Vesely et al 1981]. A fault tree is composed of a number of symbols, see Figure 13.

Fault trees are used to identify combinations of component failures or human errors that can lead to an undesired event (top event), which can be a system failure, a component failure (if regarded as a subsystem), a loss (degradation) of function, a human error, etc. The technique is preferred when combinations of failures are expected. The logic in fault trees is binary. Events (faults) either occur or not and the mathematical operations are in essence very simple.

[55] Configuration control is the process for the proposal, review, acceptance, implementation and documentation of changes to the model.

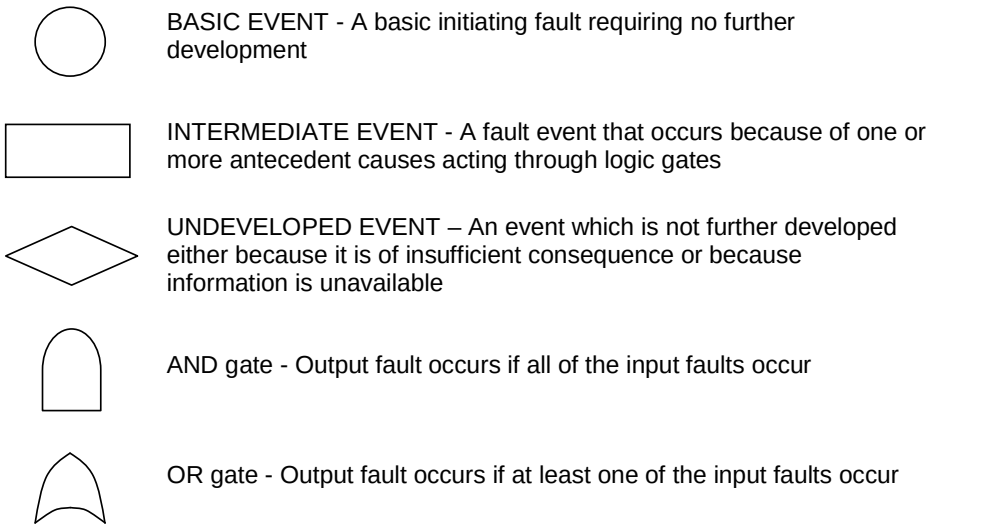


Figure 13: Basic Fault Tree symbols. Source: Vesely et al [1981].

A fault tree is quantified by establishing the probability of occurrence of the base events. Once those are defined the probability of the top event can be calculated. The great advantage of the method is that a very simple set of rules and symbols provides the mechanism for analyzing very complex systems [Ericson 1999]. One of the characteristics of a fault tree is that safety influencing factors that occur much before the occurrence of the top event (sometimes referred to as latent factors and often management/organisational in nature) are located deep in the tree structure. In practice, this means that in order to capture these factors in the fault tree, the tree must be expanded enormously. Beyond about 5 or 6 levels of a fault tree however the events are influenced very strongly by many to many influences of common mode factors such as competence, procedures, maintenance, etc. This leads to a combinatorial explosion of the trees. Computer power can also not solve this problem, since the assumption in fault tree logic is that branches are independent events (i.e. not linked by common cause). Currently there is no definitive solution for the handling of common modes in a fault tree. Hence the breaking down of the fault tree must stop at the level where common mode influences start. The elements at that level must be linked to the most important common modes through an interface with another sort of model. Another drawback of Fault Trees is that they are particularly limited for portraying dynamic aspects and iterative loops.

Event trees and Event Sequence Diagrams

An event tree (Figure 14) represents the possible consequence sequences of a hazardous situation or event, called an initiating event. The construction of an event tree begins with the specification of an initiating event. The following levels of the tree describe the working of safety systems that are able to counteract, recover or mitigate the effects of the initiating event. For each safety system, there are two branches in the tree corresponding to success or failure of the safety system. An event tree can be particularly helpful for developing counter measures to reduce the consequences of the initiating event.

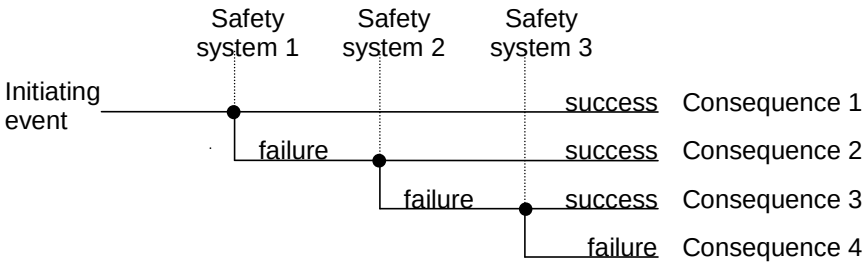


Figure 14: Event tree

To quantify event trees, the likelihood of occurrence of the risk initiating event as well as the failure and success probabilities of the safety systems are needed. Because fault trees are often used to provide the failure probabilities of systems, and because they both use Boolean logic, fault trees and event trees are regularly combined.

An Event Sequence Diagram (ESD) (Figure 15) is a representation of an event tree which distinguishes different types of events. The event sequence starts with an initiating event such as a perturbation that requires some kind of response from operators or one or more systems [Stamatelatos 2002]. Along each path, pivotal events are identified as either occurring or not occurring with paths leading to different end states. Each path through the flowchart is a scenario.

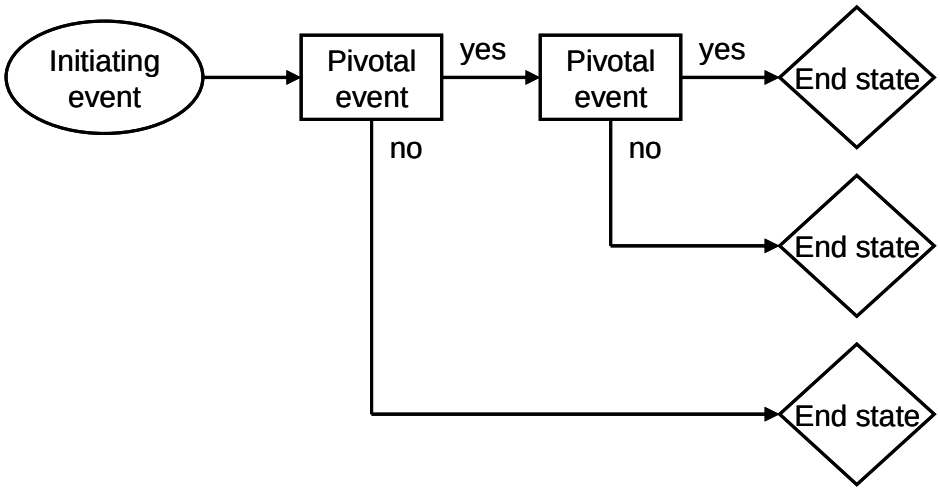
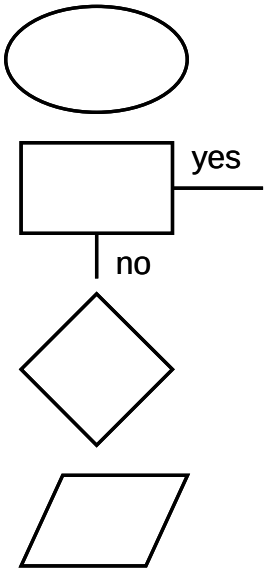


Figure 15: Event Sequence Diagram

Conditional operators can be included to represent different outcomes depending on whether the condition is met or not. Figure 16 shows types of events and condition in an ESD and their iconic representation. Intentionally, the building blocks of the scenarios are kept broad and generic to cover many ‘similar’ situations. The detailed specific or possible causes or contributing factors of these events are not directly of interest at the scenario level. Event Sequence Diagrams are often combined with fault trees. In practice, Event Sequence Diagrams are typically used to portray the progression of events over time, while

fault trees best represent the logic corresponding to the failure of complex systems [Stamatelatos 2002]. Fault trees are then used to model initiating and pivotal events in Event Sequence Diagrams in sufficient detail. The initiating and pivotal events in the Event Sequence Diagram are the top events in the fault trees. Initiating events can then be regarded as the centre of a bow-tie diagram that is sometimes used to represent accident sequences, see for instance Roelen et al [2000a] and Figure 17. The left hand side of the bowtie represents the causes of the initiating event, the right hand side the effects of the initiating event. This makes the bowtie representation valuable for analysis of risk prevention measures, which will be represented in the left hand side, and risk alleviation or mitigation measures, which will be represented in the right-hand side of the bowtie.



Initiating event: The first event in an ESD which initiates a sequence of events terminating in an end state.

Pivotal event: An event which has two outcomes, typically ‘yes’ and ‘no’, corresponding to event occurrence and non-occurrence.

End state: It is the terminating point of an ESD scenario. An ESD can have multiple end states.

Comment box: Used for providing information regarding the development of the accident sequence.

Figure 16: Types of events in an ESD and their iconic representation.

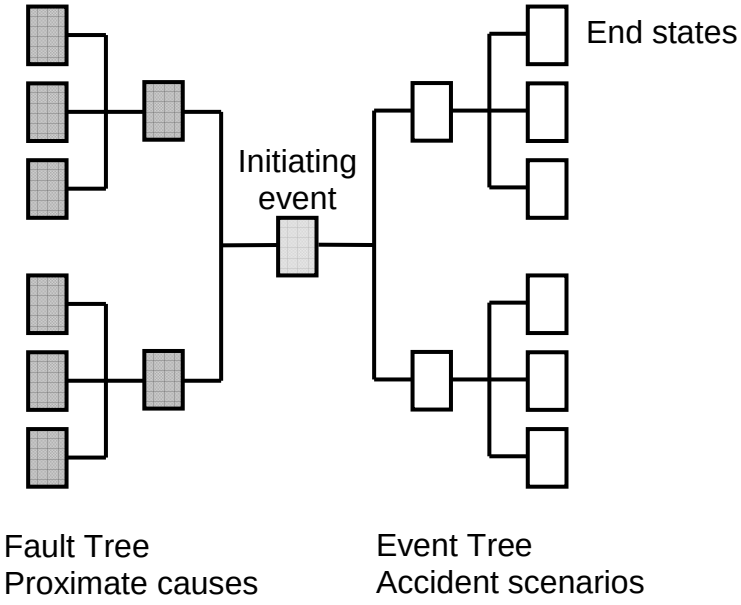


Figure 17: Bow-tie schematic

7.2.2 Bayesian Belief Nets

A Bayesian Belief Net (BBN) (Figure 18) is an acyclic directed graph in which nodes represent variables and connecting arcs represent conditional dependence. Acyclic means that it contains no path from a variable back to itself. Conditional dependence is the mathematical equivalent of 'causality (see also section 3.2). It is expressed as conditional probabilities, which thus quantify the strength of the causality in the model. A high conditional probability refers to strong causality, a low conditional probability indicates weak causality. 'Bayesian' refers to Bayesian[56] probability theory which is used to explore causal relations.

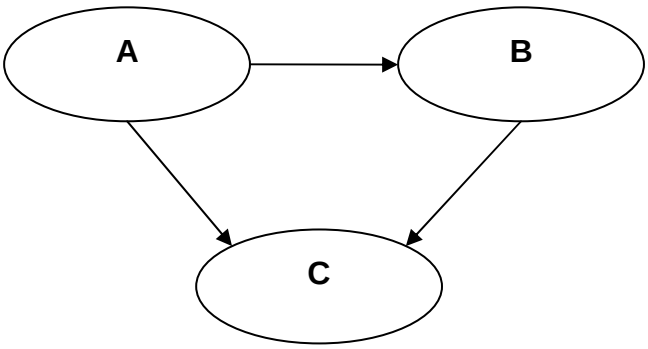


Figure 18: Bayesian Belief Net schematic

[56] Thomas Bayes (1702-1761) was a British mathematician who developed a solution to the problem of 'inverse probability'. It was published posthumously and is now known as Bayes' Theorem.

Bayesian Belief Nets are convenient tools for making inferences about uncertain states when limited information is available, and are therefore frequently used for diagnosis, with for instance applications in medical science [Jensen 1995]. But the use of BBNs is not necessarily restricted to diagnostic models; they can also be used instead of fault trees and event trees in a quantitative risk assessment. The advantages of a BBN over a fault tree is that it allows multi valued variables instead of binary events as in a fault tree, local dependencies among components instead of the classical assumption on fault trees that events are statistically independent and soft interactions among component behaviour instead of deterministic AND/OR interactions as in a fault tree. Experience has also shown that BBNs are an attractive modelling tool in which the user readily recognizes his problem [Roelen et al 2004b]. The graphical problem representation is also the user interface with which the user can do 'what if' analysis. BBNs come in various colours and shapes. For the purpose of this research, discrete BBNs and distribution free continuous BBNs are described in more detail because both have been proposed for use in causal risk models for air transport.

Discrete BBN

In discrete BBNs, the nodes represent discrete variables, e.g., 'true' and 'false' or 'bad', 'medium', 'good', or '1', '2', '3', '4', and '5'. Conditional probability tables specify the dependencies between nodes. The main drawback of discrete BBNs is the excessive assessment and maintenance burden. The number of probabilities that must be assessed and maintained for a child node is exponential to the number of parents. If a given node X has K parent 'influences' where each influence originates from a chance node with M possible outcomes, then the conditional distribution of X must be assessed for each of the M^K input influences. This exorbitant burden can only be reduced by grossly coarse-graining the outputs from nodes and /or introducing simplifying assumptions for the compounding of influences. In practice, chance nodes are often restricted to two possible values, just like in fault trees [Kurowicka et al 2005]. Discrete BBNs are not very flexible with respect to changes in modelling. If we add one parent node, then we must do again all previous quantification for the children of this node. In a fluid modelling environment this is a serious drawback. We would much prefer to be able to add a new node by adding one number for each child node, indicating influence, without re-doing the previous quantification, particularly in cases where data are sparse and expert judgement is needed to quantify the influences. This stresses some of the weakest features of discrete BBN methodologies. To overcome these difficulties, distribution free continuous BBNs have been developed.

Distribution free continuous BBN

Kurowicka & Cooke [2004] introduced an approach to continuous BBNs using vines[57] together with copulas[58] that have the zero independence property. In this approach, nodes are associated with arbitrary continuous distributions and arcs with conditional rank correlations. For non-mathematicians this whole concept is difficult to comprehend, but it comes with a significant advantage; the assessment burden for this type of BBN is limited to a one dimensional distribution for each node and for each arc a conditional rank correlation. Complexity is linear with the number of parents, rather than exponential. Elicitation procedures for conditional rank correlations are described in Morales et al

[57] A vine is a graphical model for dependent random variables [Bedford & Cooke 2002].
[58] The copula of two continuous random variables X and Y is the joint distribution of $(F_X(X), F_Y(Y))$ [Bedford & Cooke 2002].

[2008]. While distribution free continuous BBNs as described in Kurowicka & Cooke [2004] and Morales et al [2008] have significant advantages, these come at a price. One disadvantage of this method is that it fails to capture non-monotonic dependencies. Dependencies that show large jumps for specific threshold parameter values may also not be properly captured. Perhaps the biggest disadvantage is the lack of transparency. Probability distributions are more complicated than simple failure/success probabilities and a rank correlation is not intuitive for non-mathematicians. Unlike fault trees, calculations of even simple models are difficult to reproduce with pencil and paper. The advantage of course is the added functionality of these types of models, but we should be very careful only to apply them to those problems in which such functionality is indeed necessary. This technique was applied in the CATS model [Ale et al 2007] for the Dutch Ministry of Transport and Water Management.

Using BBNs; lessons learned from CATS

The CATS model is unique in the sense that the aviation risk model is (mathematically) represented as a single (large) BBN. The network representation is very suitable for describing the interdependencies between system elements. This combination of system-wide representation and representing the system as a network makes the CATS model a unique and potentially powerful instrument. But in the CATS model, the computational engine is mathematically complex relative to the model structure and the data. This is the result of the request by the Dutch Ministry of Transport to take proper account of interdependencies and uncertainties. This led to using a single BBN for the model, introducing variables with continuous distributions to avoid a computational explosion. Dependencies between model elements are represented by conditional and unconditional rank correlations and many of these correlations are retrieved by expert judgement using a conditional quantile approach.

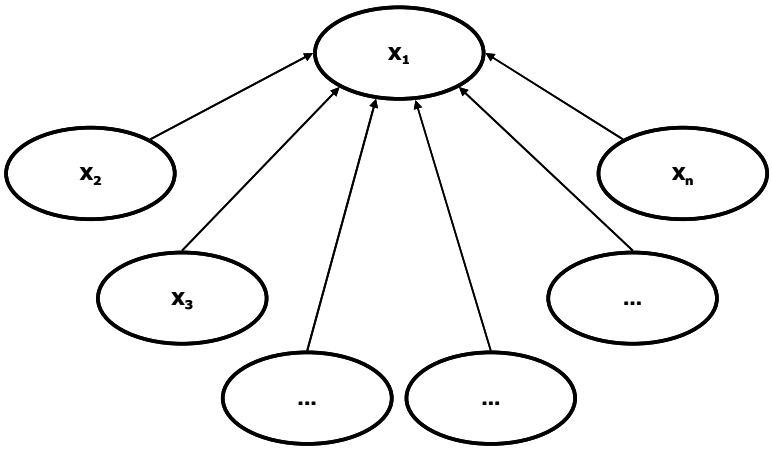


Figure 20: Dependence between variables represented in an influence diagram.

To elicit the unconditional rank correlation $r_{1,2}$, the expert is asked the following:
Suppose that the variable X_2 was observed above its q^{th} quantile. What is the probability that also X_1 will be observed above its q^{th} quantile?

The approach for CATS was with q = 0.5 (the median) and assuming the minimum information copula realizing the rank correlations in the joint distribution. To assess the conditional rank correlation $r_{1,3|2}$ the expert is asked the following question:
Suppose that not only variable X_2 but also X_3 were observed above their median. What is the probability that also X_1 will be observed above its median value?
The possible answers to this question are constrained by the answer to the first question. For instance if the expert believes that X_2 and X_1 are fully correlated this means that X_1 would be completely explained by X_2. In this situation X_3 cannot have additional influence on X_1 and the answer to the second question cannot be anything else but 0. In general the upper and lower band of possible answers are not self-evident or easy to calculate. They have to be calculated on-line during the elicitation to help experts avoiding inconsistencies in their answers [Morales et al 2008]. All additional required conditional rank correlations are obtained by asking increasingly 'nested' questions, the final question becoming
Suppose that X_2 and X_3 ....and X_n were observed above their medians. What is the probability that also X_1 will be observed above its median value?
The nested constraints on successive conditional probabilities make it even more complex and requires the experts to imagine combinations of events which are very difficult to grapple with, like for instance the following example from Roelen et al [2007]:

Suppose that you select 1,600,000 flights at random. Suppose that out of the 1,600,000 you select 800,000 for which crew suitability is at least equal to its median value and out of those 800,000 you select 400,000 for which aircraft generation is also at least equal to its median value. Additionally suppose that out of these 400,000 you select 200,000 for which weather is also at least equal to its median value and out of the last 200,000 you select 100,000 for which abnormal situation is also at least equal to its median value. What is your probability that in this (not randomly chosen) pool, the median value of flight crew errors will be more than your median estimate?

Answering this question requires mental skills which are, most probably, unrelated to what it takes to be an operational expert. The apparent necessity to help the expert in answering the questions also justifies some real concern regarding this method. If the expert is not able to provide consistent answers without help, how can we be confident that the answers are reliable estimates of the probabilities we want to obtain? Another considerable problem is the fact that, as the upper and lower bounds of possible values are not self-evident, there is hardly any possibility for peer review. The values themselves are difficult to interpret as they do not provide a direct overview of the relative strength of the dependencies. Mathematically it is all correct, but practically this approach introduces a lot of uncertainty, the exact size of which is unknown because there are insufficient ways to determine if the expert's answer makes sense or not.

A different approach is to assume that the dependence of, say, X_1 and X_2, is independent of the value of, say, X_3. Under this assumption we ask the experts to estimate the unconditional rank correlation $r_{1,2}$ (similar to the way described above) but the other dependences are obtained by asking the expert to estimate them as a portion of the dependence of X_1 and X_2. In this case, a single simplifying assumption leads to less complicated questions. The simplifying assumption in itself creates additional uncertainty[59], but because the questions for the experts become less complicated, the approach is much

[59] Cases in which the dependence between X_1 and X_2 depends on the value of X_3 cannot be accurately represented. An example would be X_1 alertness, X_2 drug use (e.g. dexchlorpheniramine, an antihistamine) and X_3 the use of alcohol [Starmer et al 1994].

more credible to the experts and their answers will allow peer review. This simpler approach was adopted at a later stage in the development of the CATS model.

This example shows how modelling solutions for one problem may create difficulties somewhere else. This should be closely guarded by the development team and requires good coordination among the different team members.

7.2.3 Petri nets

A Petri[60] net is a graphical and mathematical modelling tool for describing and analysing distributed systems. A directed graph structure is used to represent a Petri Net (Figure 19). It is composed of a set of places and transitions, connected by arcs, and an initial marking representing the number of tokens in each place at time t = 0. Arcs are either from a place to a transition or from a transition to a place. In modelling, using the concept of conditions and events, places represent conditions and transitions represent events. In graphical representation, places are drawn as circles, transitions as bars or boxes. The execution of the Petri net is by means of movements of tokens among places. The presence of a token in a place is interpreted as indicating the truth of the condition associated with that place. Events in one box trigger state changes in the boxes to which it is connected by directional arrows. The possibility of changing the systems characteristics over time gives a Petri net its dynamic character. When more complicated systems are to be modelled with Petri nets, extensions to the original concept have to be made, thus creating high level Petri nets. The three most important extensions are the possibility to model multiple in- and outputs, the addition of types of influence (illustrated by means of colour), and the addition of time by means of time stamps.

Dynamically Coloured Petri Nets can be used to represent Piecewise Deterministic Markov Processes (PDPs) [Everdij & Blom 2005]. PDPs are useful to describe complex continuous time stochastic processes. A PDP consists of a discrete valued component and a continuous valued component. The discrete valued component models the mode process. At discrete times, the discrete state may switch to another mode value which is selected according to some probabilistic relation. The continuous valued component models the drift process as a solution of a differential equation that depends on the discrete state. At discrete moments in time, the continuous state may jump according to some relation, which makes it only piecewise continuous [Everdij & Blom 2005].

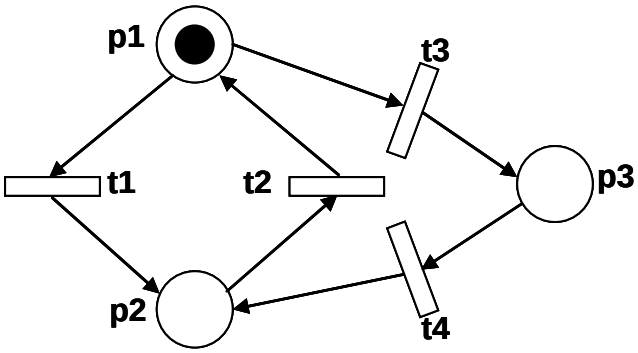


Figure 19: Petri Net

[60] Named after the German mathematician Carl Adam Petri.

PDPs can be used to construct models in which both discrete and continuous variables evolve over continuous time, possibly affected by probabilistic disturbances [Blom et al 2001]. Quantitative analysis of the Petri net can be performed by Monte Carlo simulation. An advantage of Petri nets is that the influence of time and process dynamics on scenarios is captured. This is an added value when describing systems with many interactions or feedback loops. Air transport is typically a system with many interactions, e.g. communication between flight crews and air traffic controllers. Examples of application of Petri nets for aviation risk assessment have demonstrated the power of Petri nets in combination with Monte Carlo simulations for handling interactions [Blom et al 2008]. Petri Nets in general do not provide an easy understanding of scenarios, because the representation is less intuitive than for instance an event tree [Murata 1989].

7.3. Size, depth, complexity and uncertainty

For practice, but not necessarily scientific reasons, a limit must be set on the size of the causal model and the depth of causes in the model and the attachment of common mode influences and adaptive mechanisms. The degree of detail and sophistication of any given part of the model should be decided upon based on some sort of sensitivity analysis of how much that aspect could contribute to the overall risk figure. Only for major influences is further detailed modelling worthwhile. Another criterion is that a user has to see how a particular influence feeds to the 'top event'. The required level of detail also depends on the purpose of the model, ranging from global risk assessment to detailed risk management. During development, a situation may easily arise in which every *possible* refinement of the model is considered *necessary* by the modellers, thereby losing track of the model's intended use [LVNL 2001b]. In a peer review of a causal risk model feasibility study, a weakness of the causal risk model was described as "*The quickly escalating complexities as one attempts to improve the fidelity of the model(s). This attribute can potentially soon overwhelm the output of the model or the value of the output. As a research tool, this may not be so critical, inasmuch as skilled analytical researchers are working with it; as it moves toward the potential end user (i.e. the Director of Flight Operations, the Safety Manager, etc.) the output must be sure and simple*" [EAI 2002]. Developers should be keenly aware of this and should ensure that the complexity of the model is in balance with its intended use. A practical solution is to model first at a more aggregate level and add further detail in a next step, taking account of those elements to which the model appears to be sensitive or areas of specific interest (e.g. because one wants to analyse the effect of certain proposed safety measures or other measures which might compromise safety). Essentially, stop rules for causal factor analysis are anchored to the applicable span-of-control of the problem owner and the incident severity [Koornneef 2000]. Another natural point to stop detailing is at the point where our knowledge of the system becomes inadequate to properly describe the cause effect relations involved, or where the available data (through measurements, observations, or expert judgement) are insufficient for quantification. The size and depth of a causal risk model for air transport is not significantly limited by mathematical restrictions. Experience in the CATS model demonstrates that even if a causal risk model is represented as a single BBN with hundreds of nodes with continuous variables, this presents no insurmountable mathematical difficulties [Ale et al 2007] and requires no more computational power than an off the shelf personal computer [Ale et al 2008]. Even though the development of a causal risk model involves mathematical challenges such as the correct representation of dependencies and dynamic interactions, developing a causal risk model is not primarily a mathematical problem. Developing a causal risk model is much more a problem of assembling all available knowledge and data on a very large and complex system and unravelling this complexity

into cause-effect relations that can be assembled in a model and presenting this information in a useful way.

A distinct feature of the CATS model is its truly system-wide representation of aviation. Other models, such as Eurocontrol's IRP and FAA's model for SASO (see sections 4.3.4 and 4.3.6), only represent a portion of the system (ATM and airline operations, respectively). Although these models are useful in their own right, a system-wide representation in a single model is an excellent way to capture possible interactions and most importantly to represent safety hazards that are created by interface problems between different aviation system elements.

Any model inherently contains uncertainty and noise. An estimate of the degree of uncertainty for intermediate model results is required to correctly interpret overall model results and also to identify those parts of the model for which further development is justified. It makes no sense to put a lot of effort into improving the accuracy of one part of the model if this does not contribute to improvement of the overall result. The accuracy of the model results are determined by the model structure, the mathematical operations being performed 'inside' the model and the data used to populate and feed the model. A correct balance between those three ingredients is desirable; it makes no sense to build, for example, a complex and detailed model structure when there is no data to populate the model. User requirements are important for determining what constitutes a 'correct balance'.

The functionality of techniques for causal modelling is inversely proportional to their transparency; greater functionality is paired with less transparency. Because of the importance of transparency, the model developers should choose a technique (or combination of techniques) that fits the functional requirements, instead of selecting a technique with more functionality than is actually required.

7.4. Time dependency

The ability to represent time as an independent variable does not follow directly from the user requirements. Yet the question whether or not time should be represented as an independent variable is important because of the consequences for the selection of the modelling method. The main question is whether dynamic behaviour is a significant factor.

A system can be termed static with respect to risk when the behaviour of the system does not change rapidly with time. Strictly speaking, the aviation system, and in particular an aircraft in flight, cannot be considered static for the purpose of risk analysis. The probability of some events changes over time. For example engine failure is more probable when high engine power is selected (such as during take-off). The available time for recovery may also depend strongly on the flight phase. Interactions between human operators such as air traffic controllers and pilots are also time dependent. Think of for example the interactions between a controller and a pilot when an aircraft is crossing an active runway. Another example of a time dependency is a wear related component failure. There are also slower changes over time, such as influences from management, maintenance and regulation.

There are two possible approaches in a risk analysis for a system whose behaviour can change with time. The first is a discretisation of the dynamic character of the system, for example by considering short sections of an aircraft's flight. Failure probabilities may differ for each section but can be considered static within a period. In this manner the risk

analysis is repeated for each section of the flight. A second and more elegant approach is the application of dynamic modelling techniques.

Fault trees and BBNs are particularly limited for portraying dynamic aspects and feedback loops. ESDs can capture some dynamic characteristics of accident scenarios implicitly and heuristically by the model developer's knowledge of the accident evolution [Labeau et al 2000, Swaminathan & Smidts 1999]. Two of the most feasible computational approaches for dynamic modelling of large scale problems are discrete dynamic event trees and Monte Carlo simulation [Labeau et al 2000]. Monte Carlo simulation using Petri Nets is also proposed by Blom et al [2006] for accident risk assessments in air transport. Comparison with a static fault/event tree approach indeed showed differences in the result.

Dynamic methods are more abstract and mathematically more complex. Many engineering disciplines are content with applying static techniques because they are easy to understand, representation schemes are simple and computation is simple. However, explicit incorporation of dynamic aspects into scenarios will lead to more correctness of modelling [Swaminathan & Smidts 1999].

Representing all rapid and slow changes over time within the total air transport system in a model would be a gigantic task. Existing examples of aviation accident risk assessment with support of Monte Carlo simulation, such as the example described in Blom et al [2006], were limited to a sub part of the air transport system. The degree to which dynamic modelling is necessary will therefore require careful consideration and will depend on the type of user and type of application of the model. For a system-wide causal risk model, which is used for decision making at a strategic level, a static model is for practical reasons more appropriate. For local users who have operational and tactical questions, a dynamic model may be required and dynamic modelling techniques such as Monte Carlo simulation should be considered. The possibility of linking local dynamic models with an overarching static model could be a topic for future research.

7.5. Conclusions for this section

A causal risk model can be regarded as being composed of a representation scheme and a computational engine. The representation scheme is the interface between the computational engine and the user of the model. The representation scheme is the primary means of communicating the model to the outside world. Many people will regard the representation scheme as 'the model'. The representation scheme must therefore be comprehensive, and most importantly, transparent. A scheme is transparent when a simple set of rules and symbols applies and when it is intuitive. Simplicity of the representation scheme can also be achieved by a (physical, functional, or other) decomposition. Whenever a decomposition is made a mechanism must be in place to keep track of possible dependencies between components.

Modelling methods with a very simple set of rules and symbols are tree-like structures such as fault trees and event sequence diagrams. They are therefore suited to provide the top layer of the representation scheme, the level that is used most directly by the user.

However, beyond a few levels of the trees the events are strongly influenced by many to many influences of common mode factors. Since the assumption in fault tree logic is that all events are independent events, the model must be linked at this level with another type of model. Bayesian Belief Nets are a suitable candidate as they can capture dependencies and soft interactions among component behaviour. The graphical representation of a

Bayesian Belief Net is still intuitive, although the rules are more complex than those of a fault tree. A BBN is therefore less transparent. When dynamic interactions between components are important it becomes necessary to shift to another type of modelling technique such as Petri Nets in combination with Monte Carlo simulation. The draw back then is that the model representation is not intuitive.

The degree of detail and sophistication of the model representation must be decided upon based on the purpose of the model and the performance that is required by the computational engine. If the computational engine must be able to handle interdependencies and dynamic behaviour, a simplified representation of the model via a user interface is required to obtain transparency.

Even though a causal risk model is a mathematical object, developing a causal risk model is not primarily a mathematical problem. Examples show that proper mathematical techniques are available and the required computer power is not excessive. Developing a causal risk model is much more a problem of assembling all available knowledge and data on a very large and complex system and unravelling this complexity into cause-effect relations that can be assembled in a model and representing this information in a useful way.

The development of a causal risk model for aviation should be guided by questions and functional requirements from the users of the model. The model developers should use techniques, develop (mathematical) solutions, and go to the level of detail that fits the user requirements. The degree to which dynamic modelling is necessary will require careful consideration and will depend on the type of user and type of application of the model.

Chapter 8. Quantification

The causal risk model describes the output parameter, aircraft accident risk, as a function of the input parameters. Because risk is a combination of severity and probability, which both imply rank ordering, it must be expressed as a unit that can at least be rank ordered. This has far reaching consequences for both the development and use of the model. To understand these consequences it is necessary to look more closely at quantification, and particularly units and values. Especially when trying to quantify 'soft' causal influences, finding appropriate units is not an easy task. A discussion on quantification is not complete without addressing the topic of numerical accuracy and uncertainty. The credibility of the causal model depends on the quality of the model structure and on the sources of the numerical data to be used and the confidence that one can have in the figures. This section describes the most relevant sources of data for quantification of an aviation causal risk model.

8.1. Measurements, quantities, units and values.

The causal risk model is primarily[61] a quantitative model which expresses the level of safety as the value of one or more physical quantities. The model consists of variables and conditional probabilities or rank correlations where the variables describe causes and effects and the conditional probabilities describe the strength of the causal relationship. A conditional probability describes the probability of occurrence of a certain value of a variable given a particular condition, e.g. the occurrence of a breaching of the dykes in Zeeland given a high sea level. The condition can be described qualitatively ('high sea level') or quantitatively (e.g. sea level at Flushing (Vlissingen) of +4.55 meter NAP[62]). Quantification of the probability of occurrence of variables and the conditional probabilities is done by using existing data or by expert opinion. Often different sources of data will need to be combined, for instance recorded sea level measurements are used to determine the probability of occurrence of a high sea level, and expert judgement is used to determine the probability of breaching of the dykes given the high sea level. In these cases it is important that the description of 'high sea level' in the database is the same as what the experts consider to be 'high sea level'. When different sources of data are combined it is always extremely important to ascertain that the same variable descriptions and units of measurement are used. The following definitions are relevant [Taylor. 1995]:

> A **quantity** is a property ascribed to phenomena, bodies, or substances that can be quantified for, or assigned to, a particular phenomenon, body, or substance. Examples are mass and electric charge. A physical quantity is a quantity that can be used in the mathematical equations of science and technology.
>
> A **unit** is a particular physical quantity, defined and adopted by convention, with which other particular quantities of the same kind are compared to express their value.

[61] The heart is quantitative, but semi quantitative or qualitative could be considered.
[62] This was the sea level at Flushing (Vlissingen) on 1 February 1953 when the Netherlands' flood disaster occurred.

> The **value** of a quantity is its magnitude expressed as the product of a number and a unit, and the number multiplying the unit is the numerical value of the quantity expressed in that unit.

An example of a physical quantity is 'sea level' which has the unit 'meter', where a meter is defined as '*The meter is the length of the path travelled by light in vacuum during a time interval of* 1/299,792,458 *of a second*' [Taylor 2002]. In the Netherlands, the (arbitrary) zero-point of the scale is NAP.

Measurement has been defined as 'the assignment of numerals to things so as to represent facts and conventions about them' [Stevens 1946]. Often we measure because we want to apply some statistical manipulation, and which manipulations can be applied then depends on the type of scale against which the measurements are ordered. In a *nominal* scale the numerical values just name the attribute uniquely. No ordering of the cases is implied, and words or letters would serve equally well. With an *ordinal* scale the attributes are rank-ordered but distances between attributes do not have any meaning. Strictly speaking, statistics involving means and standard deviations cannot be used with ordinal scales, because these statistics imply knowledge of something more than the relative rank order. With an *interval scale* the distance between attributes has a meaning. Almost all statistical measures are meaningful, except those that require the existence of a true zero point. The zero point in an interval scale is a matter of convention, and as a result ratios are meaningless. A *ratio scale* consists not only of equidistant points but also there is a true zero that is meaningful (not arbitrary) and a fraction (or ratio) can be constructed. All types of statistical manipulations are applicable to ratio scales, including logarithmic transformations, such as are involved in the use of decibels [Stevens 1946]. Using the same units of measurement is vital when information from different sources is combined, see the example of the Mars Climate Orbiter and Figure 21.

Mars Climate Orbiter
On September 23, 1993, the Mars Climate Orbiter was scheduled to enter into an orbit around the planet Mars. The vehicle had been launched 9 months earlier by a Delta rocket from Cape Canaveral Air Station, Florida, but the Mars orbit insertion manoeuvre failed, resulting in a lower than planned spacecraft trajectory. As a consequence the obiter was lost. The exact fate of the spacecraft is unknown; it was either destroyed in the atmosphere or re-entered heliocentric space after leaving Mars' atmosphere. The Mars Climate Orbiter accident investigation report concluded that the cause of the accident was the failure to use metric units in a segment of ground-based navigation related software. That software required input in metric units of Newton seconds, but instead the data was delivered in English units of pound seconds [Stephenson 1999].

In everyday language, we often ignore the difference between quantities, units and values. Human beings are able to interpret descriptions of the values of quantities and rank different descriptions of the value of the same variable. If the quantity is for instance 'sea level', we are able to interpret what 'high sea level' means and rank it against 'low sea level'. In doing so, we are able to communicate complex issues in a rather effective way. A prerequisite is that sender and receiver share a common context or frame of reference, otherwise miscommunication will occur. Two pieces of information can only be combined when they share a common frame of reference. Whether a 'small bolt' will fit into a 'small

hole' is impossible to tell unless the diameter of bolt and hole are expressed in a similar unit of length. This is precisely the reason why units have been defined.

Figure 21: Without units things do not make much sense.

The use of standard units offers the greatest flexibility to express the condition of variables and is therefore preferred. But sometimes it is impossible or impractical to apply a standard quantitative unit of measurement. For example if we want to know the distribution of 'sleepiness' among flight crews during different stages of the flight we could resort to measurement of brain activity, but this is quite impractical. It is much easier to ask pilots whether they feel sleepy. We then use a descriptive rating scale. When developing a descriptive rating scale it is important to provide sufficient *anchor points*; qualitative descriptions that allow little room for interpretation. Key words and phrases must be easily understood and yet sufficiently definitive so that each rating will be clearly separated from every other rating. Particularly when expert judgement is used it is important to remove as much ambiguity as possible. Simply providing a scale that runs from 'none' via 'a little' and 'much' to 'very much' will not do. An example of a scale for sleepiness with proper anchor points is the Stanford Sleepiness Scale [Hoddes et al 1973]:

Stanford Sleepiness Scale

1. Feeling active and vital; alert; wide awake.
2. Functioning at a high level, but not at peak; able to concentrate.
3. Relaxed; awake; not at full alertness; responsive.
4. A little foggy; not at peak; let down.
5. Foggy; beginning to lose interest in remaining awake; slowed down.
6 Sleepy; prefer to be lying down; fighting sleep; woozy
7. Almost in reverie; sleep onset soon; losing struggle to remain awake.

A potential problem with such a scale is the suggestion that the difference between 1 and 2 is of the same size as the difference between 2 and 3. In reality this may not be the case; because it is an ordinal scale the values only represent a rank order.

A different approach for quantifying variable descriptions is to select a proxy variable that is strongly associated with the variable of interest and which can be quantified. An example is the variable 'weather'. This actually combines a whole set of atmospheric conditions including wind speed, precipitation, cloud coverage, temperature, etc. For a pilot it is important to be aware of the actual weather conditions on his planned route because 'bad weather' is not only uncomfortable for passengers but can also be dangerous. An exact unit for the variable 'weather' does not exist but in practice the intensity of rainfall comes pretty close. Often intense rainfall is associated with strong and variable winds and possibly lightning, which all can be aviation hazards. Therefore rainfall can be used as a proxy variable for weather. Rainfall intensity can be objectively measured in mm/day using standardised gauge. This example also shows that a proxy has boundary conditions for its appropriateness; if we are interested in clear air turbulence[63] as part of the variable weather, rainfall rate is definitely not a good proxy. Therefore, whenever proxies are introduced, the boundary conditions within which they work need to be defined.

From the potential users of a causal risk model comes a requirement to represent the influence of composite factors such as safety management and safety culture [De Jong 2006], see section 4.4. One of the main difficulties lies in the need to express such composite factors in objectively quantifiable units. What are proper units for complex issues such as the quality of procedures, commitment of personnel, non-technical skills of pilots or safety culture? Can they be constructed so that experts can consistently use them? Proxy variables are also difficult to find as these often reflect only one of the 'dimensions' of the variable of interest. The difficulty of developing clear anchor points for a qualitative scale or finding relevant proxy variables is one of the main obstacles associated with quantifying the influences of management on accident risk [Bellamy et al 1999].

We could strictly apply the requirements for quantification and only allow elements in a causal risk model that can be described using well defined physical quantities and units of measurement or for which decent proxy variables can be found. However, many (allegedly) influencing factors would almost certainly disappear from the model, and some of the user requirements would not be met. Alternatively we could incorporate such factors using less well defined quantities and units of measurement or rating scales at the risk of introducing additional variability in the model and increased uncertainty in the model results. The question is whether the advantages of having the factors in the model outweigh the disadvantage of additional variability. To answer this question it is necessary to estimate the degree of variability that can be expected. The variability is introduced because, due to the lack of properly defined units, different people will use different yardsticks for measurement and ranking. These yardsticks are partly based on individual experiences and beliefs. When a particular groups shares beliefs and experiences, the variability within that group is much less than the variability between members of different groups. This can be observed in different cultures. Within a cultural group, even ideas on complex properties such as, say, beauty and taste are quite consistently shared. It is very difficult, if not impossible, to describe the beauty of the human face or the overall quality of a restaurant in objectively quantifiable units. Yet there is much consistency among people on what is considered to be an attractive face [Zaidel & Cohen 2005, Braun et al 2001], or what is an excellent restaurant.

[63] Clear air turbulence is characterised by the absence of clouds. It typically occurs at cruise altitudes.

An example of a complex issue for which a broadly accepted rating scale has been devised concerns the quality of restaurants. The Guide Michelin contains the most influential gastronomic rating in the world, based on rather loose definitions:

- Three stars: Exceptional cuisine, worth a special journey.
- Two stars: Excellent cooking, worth a detour.
- One star: A very good restaurant in its category[64].

The *quantity* considered is the quality of the restaurant, the *unit* is the Michelin Star, and the *value* is 0, 1, 2 or 3. Among the elements taken into account are the quality of the ingredients, the technical ability shown in their preparation, the combination of flavours, the creativity and the consistency of the cooking and the value for money it offers. Despite its apparent weak definition and lack of transparency, the system of Michelin Stars is well established and broadly accepted. It has created a yardstick for a complex multidimensional issue. The yardstick can be used and applied with consistency even without detailing all the dimensions associated with it.

A similar rating system for, say, safety culture or flight crew non-technical skills, would allow an acceptable representation of it in a quantitative causal risk model. Several attempts have been made for developing rating scales for safety culture and non-technical skills, see for instance Hudson [2003], Von Thaden et al [2003], Kho et al [2005], Van Avermaete [1998] and Davoudian et al [1994]. But a 'Guide Michelin' in which stars are awarded to companies that have an excellent safety culture does not yet exist , although the airline 'blacklist' that has been published by the EU [EC 2005] is an attempt, albeit very crude. And in spite of all this research the current regulation on flight crew licensing does not contain an evaluation scheme of non-technical skills. One of the difficulties is that there are hardly any data on the relation between non-technical skills assessment and the actual flight crew performance on non-technical skills [Van Avermaete 1998]. This is because these skills are insufficiently specified to be clearly measurable. They need to be tied to tasks in order to operationalise them. In the short term, a broadly accepted and applied rating system for non technical skills or safety culture cannot be expected. But if we do not make a start, we will never get anywhere, not even to know if we have consensus now or not and if so, built in what group(s). The only way to find out if proper units for these complex issues can be constructed is to try (harder) to develop and validate them.

8.2. The need for ratio scales

For some applications of a causal risk model an outcome using an ordinal scale may be sufficient, for instance when comparing two possible alternatives it may be sufficient if we know which of those is the 'better'. This does not mean however that it is therefore sufficient to express the model parameters with ordinal scales. Consider the following example from Hesslow [1976]: The use of birth control pills is a risk factor for thrombosis. At the same time, birth control pills are an effective preventer of pregnancy, which is in turn a risk factor for thrombosis. The use of birth control pills may thus affect one's chances of suffering from thrombosis directly and indirectly via the effect of the pills on the probability of becoming pregnant (Figure 22). Whether birth control pills raise or lower the overall probability of thrombosis will depend on the relative strengths of these two causal routes. The example illustrates that if the outcome of the model is required to be ordinal, this does not mean that it is sufficient to express each individual element of the model as an

[64] This is the customer rating, the judges from Michelin have a more detailed rating scale set and definitions.

ordinal value. The direct and indirect effects of birth control pills must be expressed in meaningful numbers (using an interval scale or ratio scale) in order to determine the overall effect. Similarly, it is a misconception to think that the required numerical accuracy of the model is the same as the required numerical accuracy of the model elements. This misconception is clearly present in a recommendation made in 2003 by the Safety Advisory Committee Schiphol [VACS 2003b]. According to the recommendation, "a causal model can be a proper tool to make well-justified weightings in decisions with safety consequences or when choices between different safety measures need to be taken". According to the same recommendation, "this does not necessarily require a high numerical accuracy, but rather a high degree of consistency". While it is true that a high numerical accuracy may not necessarily be needed for the final model outcome, the numerical accuracy of the underlying causal relations is important. Errors in the input and errors due to model uncertainty propagate throughout the model and eventually accumulate in the end-result. Exactly how the individual error terms propagate depends on the model structure and the type of dependencies, but generally speaking this means that the error in the model input en the elements of the model should be *smaller* than the acceptable error in the model output. Requirements for individual model elements are thus more strict than the overall model requirement.

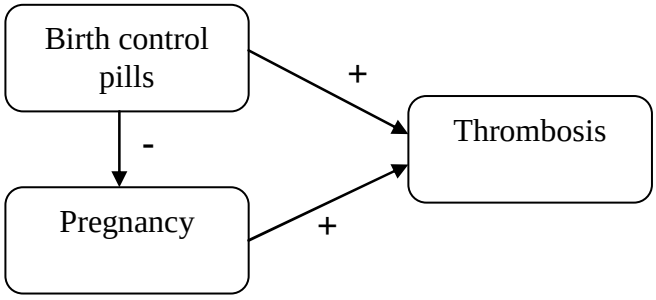


Figure 22: Dependence between birth control pills, pregnancy and thrombosis.

8.3. Uncertainty

A result of a quantitative risk assessment that only gives point estimates of accident probabilities provides a very loose basis for decision making. Levels of uncertainty of the model results are as equally important as point estimates. In many fields of science and technology, the parameters necessary for modelling physical, chemical or biological behaviour are not known with certainty. Different values may be reported in literature, and scientists sometimes disagree as to which values should be used in practical applications. The amount of uncertainty is often represented as a confidence interval. A confidence interval gives an estimated range of values which is likely to include the true parameter value. The width of the confidence interval gives information on the uncertainty surrounding the estimate; a larger interval indicates more uncertainty.

Sometimes a distinction is made between aleatory uncertainty and epistemic uncertainty. Aleatory uncertainty is the result of randomness in the input parameters, epistemic uncertainty is the result of lack of knowledge. Aleatory uncertainty is 'objective', epistemic uncertainty is 'subjective'. For the final end result there is no need to distinguish between different types of uncertainty, as there is ultimately one aggregated uncertainty measure. From a practical point of view the difference is relevant. In experiments that involve the drawing of coloured marbles from a jar a 95% confidence interval can be uniquely

determined, given the results of the experiment. The 95% confidence interval will converge if the experiment involves more sampling. Uncertainty is strictly 'objective' and methods for estimating aleatory uncertainty are well-established, see for instance Everdij et al [2006]. Estimating uncertainty of results of a causal risk model is less straightforward. Consider for instance input parameter values which are quantified by expert judgement. Experts can estimate 95% confidence intervals, but these are not unique and undisputable as in the case of uncertainty intervals for 'objective' uncertainty. The underlying probability distribution is not known, but is assumed. In a strict sense it is incorrect to refer to such intervals as 95% confidence intervals, because the 95% has no real meaning other than as some indication of the degree of uncertainty (of the expert). When the judgements of more experts are elicited the 95% confidence interval does not necessarily converge. The 95% does not refer to numerical probabilities but to a qualitative 'scale of likelihood'. While calculations can be made with confidence intervals provided by experts, it is incorrect to treat them in the same way as confidence intervals that are the result of controlled experiments. Lack of information cannot be compensated by mathematics. On the other hand, sometimes very elegant mathematical solutions to estimate epistemic uncertainty can be found, like Wechseler's Range Ration.

Wechseler's Range Ratio

The human component in a system is much more variable than the equipment component. For this reason estimates of uncertainty of failure of people in a system will usually be larger than the estimates of uncertainty assigned to equipment in that system. It is therefore particularly important that human error probabilities are not expressed as a point estimate but that they are accompanied by appropriate uncertainty estimates. Because of lack of real data on human error probability distributions, often some kind of probability distribution is assumed with the single point estimate as the median of the distribution. On the presumption that for skilled operators most errors will fall near the low end of the distribution a lognormal distribution was proposed in [Swain & Guttmann 1983]. It was further observed that despite variability among people, selection and other managerial influences tend to restrict variability. If a person consistently performs far below average for his group, he will usually be retrained or reassigned. This phenomenon is expressed as Wechseler's Range Ratio: "Given a population that is homogeneous with respect to some ability that can be measured with a ratio scale, the highest score in the distribution will rarely be more than five times the smallest score and usually not more than three times the smallest score". By assuming a 4:1 range ratio between the 95^{th} and 5^{th} percentiles of error probabilities for tasks performed under routine conditions it was hypothesized that a lognormal probability density function with a standard deviation of 0.42 provides a reasonable fit [Swain & Guttmann 1983].

8.4. Model assumptions

Epistemic uncertainty arises from abstracting the endless complexity of reality into the limited complexity of the model. Abstraction is obtained by making a (large) number of assumptions. By making these assumptions explicit, and by determining the effect of those assumptions on the model results, we are able to get a feeling for the epistemic uncertainty. A weakness of a diagrammatic format of a model representation scheme, like a fault tree or an influence diagram, is that it discourages model developers from explicitly stating the assumptions and conditional probabilities for the model elements and their interconnections. An arrow between two model nodes is easily drawn but it is equally easy

to 'forget' that by drawing the arrow a number of assumptions are adopted[65]. To get a proper picture on the uncertainty of the results of a causal risk model it is necessary to conduct a thorough identification and analysis of all assumptions adopted, including assumptions on parameter values used, and to assess the individual and combined effect of these assumptions on the model results. The analysis should also include sensitivity analysis, which analyses how modifications in the values of the input parameters affect the value of the output parameters. If the value of an output parameter is sensitive to the value of an input parameter, it is also sensitive to aleatory and epistemic uncertainties in the input parameter. Therefore the sensitivity analysis helps to identify those input parameters for which it is most important to determine the level of uncertainty and to estimate the effect of underlying assumptions [Everdij et al 2006].

8.5. Data sources

The credibility of the whole causal risk model depends, apart from the quality of the model structure, on the sources of the numerical data (to be) used and the confidence that one can have in the figures. The core of the model needs to be populated with accident data. The accident investigation report is without doubt the single most complete and accurate source of information regarding a particular accident. In addition to accident and incident data, data on normal operations are required to determine the frequency of occurrence of factors that can contribute to accidents. This exposure data is often harder to find than accident data because 'normal events' are simply not recorded. Ideally, parameters of the model should be estimated based on operational data of the (sub) system under consideration. Often however the analysis has to rely on a number of sources and types of information. This section describes the most relevant sources of data for quantification of an aviation causal risk model and their qualities and limitations and looks at what could potentially be used.

8.5.1 Accident or incident data?

Aircraft accidents are rare events. In 2005, worldwide commercial operated aircraft with a take off weight of 5,700 kg or heavier suffered only 30 fatal accidents while performing 35 million flights [CAA-NL 2007]. The number of possible accidental pathways is orders of magnitude higher, which means that accident data alone is insufficient to populate an aviation causal risk model, even if data over many years are aggregated. Aggregation over (too) many years also runs the risk that the type of aircraft, rules and regulations, instrumentation, airfields, etc., have changed in the meantime. Obvious alternative sources of data are incidents.

Accident causation theories such as Heinrich's domino model [Heinrich et al 1980] and Reasons 'Swiss cheese' model [Reason 1990] suggest that accidents and incidents within one scenario follow similar causal pathways, at least partially. Indeed Heinrich's accident pyramid, or better, the way in which the pyramid concept is used by subsequent researchers, suggests a fixed ratio between the number of accidents and the number of incidents. Some authors challenge this. O'Hare [2006] for instance claims that errors observed in accidents and incidents are not the same, stating that 'failures at higher cognitive levels (e.g. goal or planning failures) are more lethal than failures at lower cognitive level (e.g. in detection or action)' and he suggests that the difference between accidents and incident is 'more than just the outcome'. But O'Hare fails to provide the definitions used for accidents and incidents, which makes it difficult to interpret the results.

[65] These assumptions depend on the type of model representation.

Indeed it can be misleading to think that actions to control minor incidents will also control major hazards. Some large chemical companies have used lost time injuries as a general performance indicator of their safety management, despite the fact that analysis of their accidents shows that these occur mainly on peripheral tasks such as falls from stairs or scaffolding, or tripping injuries moving around the plant. A classical example is that of BP which before 2005 was considered one of the leading companies as regards safety management. However, the organisation failed to notice that safety management efforts were too heavily focussed on reducing personal injuries and achieving significant improvements in personal safety performance rather than focussing on preventing major accidents. BP mistakenly interpreted improving personal injury rates as acceptable process safety performance at its refineries [Baker 2007]. This only became apparent after the BP Texas City refinery blew up on March 23, 2005, resulting in 15 deaths and more than 170 injuries. Minor incidents like tripping injuries only indicate the quality of the risk control performance on peripheral tasks, not on the operations and maintenance tasks in which the scenarios occur which can lead to major accidents.

Hale [2001] explains that in answering the question about the relation between major and minor accidents, we should ask specifically whether we can distinguish from early on in the sequences those that will lead ultimately to serious accidents from those that will only lead to minor ones and those which will be detected and recovered before they reach the damage step. In other words, was the sequence leading to a minor incident also a sequence that could lead to a much more serious event? This requires clearly articulated and understood accident scenarios. The advantage of using (complex) risk models involving major accident scenarios is that their careful application will allow performance indicators to be clearly defined, linked with the important hazards scenarios. In the BP case, the causal pathways resulting in personal injuries were different from the causal pathways resulting in catastrophic accidents and therefore the incident data could not be used to predict major accidents.

The conclusion is that, once we have developed accident scenarios and represented them in a causal model, it becomes a helpful tool for focussing on actions that might control major hazards. But the question remains whether incident data can be used to *construct* accident scenarios. The definition of accidents and incidents is important here. By properly defining what we classify as accidents and incidents, confusion such as in the BP example can be avoided. ICAO's definitions [ICAO 2001] of accidents and incidents, which are widely used throughout the aviation industry, clearly link incidents to major accidents:

> An *incident* is an occurrence, other than an accident, associated with the operation of an aircraft which affects or could affect the safety of the operation.
>
> A *serious incident* is an incident involving circumstances indicating than an accident nearly occurred.

By this definition accidents and incidents share similar occurrence pathways and only differ in the outcome. This means that, if this definition is consistently applied, incident data can be used to gain insight in causal accident pathways and can be used for quantification of accident models.

8.5.2 Accident investigation

International standards for accident investigation are defined by ICAO in Annex 13 according to which each accident and serious incident must be investigated and the results

of the investigation must be assembled in an investigation report which has to be provided to ICAO. The objective of accident investigation is to establish the causes of the accident sufficiently to be able to devise and implement remedial and preventive action. Responsibility for an investigation belongs to the state in which the accident or incident occurred, but representatives of the state of registry of the aircraft involved, the operator and the manufacturer usually also take part in the investigation. The investigation process includes the gathering, recording and analysis of all relevant information, the determination of the causes, formulating appropriate safety recommendations and the completion of the accident investigation [ICAO 2001]. Most countries have set up an independent accident investigation board to conduct the investigation without conflict of interest; examples are the National Transportation Safety Board (NTSB) in the USA, the Air Accidents Investigation Branch (AAIB) in the UK and the Bureau d'Enquêtes et d'Analyses (BEA) in France. Although the quality of the investigation may differ between countries, the standard is usually very high in Europe, North America, Australia and some parts of Asia. The investigation report will often also include information about underlying causes such as organisational factors that contributed to the accident. Because accident investigation reports are usually rather extensive and very detailed, retrieving information from those reports requires quite some knowledge and understanding of basically all aviation disciplines. Because there is no consistent, recognisable causal model behind the investigations, there is a translation step necessary to be able to populate the causal risk model with accident information. Due to the low frequency of major accidents and incidents, the investigation reports are not sufficient to populate all parts of the causal risk model. Other sources of information are required to complete quantification of the model.

8.5.3 Incident reporting

Investigating an aviation accident is usually a lengthy and costly process. The actual level of effort will normally reflect the severity or significance of the accident. A 'disaster' or some other significant accident will generate very considerable investigative effort, typically in the order of €2.5 million. [Deazle et al 1999]. Relatively minor accidents and incidents are often investigated less extensively and the reports are scantier. To facilitate collection of information on actual or potential safety deficiencies, ICAO requires that states establish a mandatory incident reporting system. All EU countries for instance submit mandatory occurrence reports to the European Co-ordination Centre for Aviation Incident Reporting Systems (ECCAIRS). A reportable occurrence is any incident which endangers or which, if not corrected, would endanger the aircraft, its occupants or any other person. The requirement applies to pilots, but also to individuals involved in aircraft manufacturing, aircraft maintenance, air traffic control and ground handling of aircraft. This includes managerial functions as well. The regulation on occurrence reporting expects employers to refrain from disciplinary or punitive action which might inhibit their staff from duly reporting incidents of which they may have knowledge [CAA 2005].

ICAO ADREP and ECCAIRS

ICAO operates a computerized database known as the Accident/Incident Data Reporting (ADREP) system. In the ADREP system, the occurrence (accident or incident) is described by listing the 'events'. Examples of events are 'runway incursion', 'electrical system failure', 'hard landing', 'fuel starvation', etc. The expression 'phase' is used to indicate in which phase of flight a certain event occurred and is always paired with an event. To describe the events, up to five 'descriptive factors' can be entered for each event. To explain the events, up to three 'explanatory factors' can be entered for each descriptive factor. Descriptive factors describe in detail what happened during an event by listing all phenomena present. Explanatory factors explain why the event happened. They are used to

determine what preventive action may be required. If possible, the descriptive and explanatory factors are coded in chronological order. If not possible, background information on terrain and weather is coded first and descriptions of what people did are coded last, and the factors must be coded such that subsequent factors explain the preceding ones. If more than three explanatory factors are required, the three most important ones must be coded and the remainder has to be mentioned in the narrative.

Since 2004, ICAO ADREP is using the European ECCAIRS system to store and analyse its data. Development of this system originated from a 1989 study in the field of accident investigation and incident reporting systems on behalf of the Commission of the European Community. One of the conclusions of the study suggested to bring together the knowledge derived from the collection of incident reports existing in the aviation authorities of various member states. As the existing systems were not compatible, the study recommended the setting up of a European co-ordination centre for mandatory incident reporting systems. In 1993 the Joint Research Centre at the request of Directorate General VII (Transport) of the European Commission started a feasibility study. The main objective of the study was the pilot implementation of an automated incident reporting system able to collect information from various existing, incompatible sources. The project was called ECCAIRS. In 1995 the feasibility of the ECCAIRS approach was demonstrated by integrating on an experimental basis more than 40,000 occurrences originating from Scandinavia, the United Kingdom and Germany. From 1995 until 1999 the first production release of ECCAIRS was set up. Current efforts focus on the actual implementation of the ECCAIRS network and the usage (analysis) of the collected data. According to EU Directive 2003/42/EC [EC 2003], EU member states must require that occurrences which endanger or which, if not corrected, would endanger an aircraft, its occupants or any other person, are reported to the competent authorities. The authorities then classify the occurrence and store the information according to the ECCARS system.

The information content for each individual incident in these systems is rather low, but the importance comes from the fact that these occurrences are much more numerous than the accidents and incidents that are fully investigated. Therefore these reporting systems are suitable for statistical analysis and to populate a larger part of a causal risk model, provided that the information is collected and stored consistently and unambiguously and provided that the incident-accident relationship is well established and known. Coding systems have been developed to facilitate the process of information storage and retrieval, (see text box ICAO ADREP and ECCAIRS); ICAO's ADREP was the first in 1976, and others have followed, using more or less the same format. ECCAIRS for instance uses exactly the same format as ADREP. The strength of these systems is their ability to capture vast amounts of data with relatively little effort. The weakness lies in the coding process during which information may get lost or is incorrectly classified. Since its first conception in 1976, ADREP has evolved into a rather elaborate system. The coding is generally performed by individuals who have not had any formal training in accident investigation and the likelihood of misinterpretation or misclassification of the information during the coding process is considerable. When first developed, the system was only intended for accidents and serious incidents, but with the introduction of EU Directive 2003/42/EC [EC 2003] the system is also used for classification of 'occurrences'. An occurrence is defined as an operational interruption, defect fault or other irregular circumstance that has or may have influenced flight safety and has not resulted in an accident or serious incident'. This means that the number of data entries has increased by a significant amount; for instance, in 2007 there were 7,881 reported occurrences in the Netherlands [IVW 2008]. Only 11 of those were possible severe incidents that required reporting before the introduction of the

Directive. After a year of experience, the Dutch Transport and Water Management Inspectorate concluded in 2008 that ECCAIRS at some points does not meet the requirements [IVW 2008]. Another problem is that of underreporting. Although the reporting of incidents may be mandatory, 'minor' incidents tend to be underreported even when reporting is mandatory. Figure 23 for instance shows a comparison of the frequency of occurrence of aborted take-offs[66] for different speed regimes according to data from mandatory occurrence reports and automatically recorded data from the aircraft's quick access recorder. This chart shows the underreporting of low speed aborted take-offs, but it also shows that for serious incidents (like high speed rejected take-offs) the mandatory reporting systems are a reliable source of quantitative information.

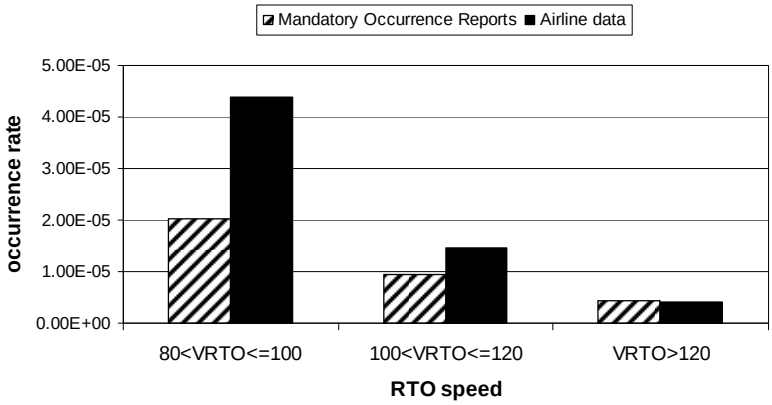


Figure 23: Comparison of the occurrence rate of aborted take-offs (aborted take-offs / take-off attempts) for different speed regimes: Operator data versus Mandatory Occurrence data. Source: Roelen & Wever [2004].

In addition to mandatory reporting, ICAO recommends that states establish a voluntary incident reporting system to facilitate the collection of information that may not be captured by a mandatory incident reporting system. The voluntary system should be non-punitive and afford protection to the sources of the information. Voluntary incident reporting programs can be very useful to gain insight into the types of incidents that have occurred, but there are some known biases in voluntary incident data that render it less useful for statistical analysis [Chappell 1994]. Only those that are familiar with the program, who have access to the reporting system and are motivated will report when they experience an incident. Furthermore individuals tend to report errors made by other individuals but when they make a similar mistake themselves they will take responsibility for ensuring that the error is not repeated, but not necessarily report it. Usually there are only limited ways to verify the contents of the report. Analysis is also complicated because most of the information is provided in the form of free text. The reporting forms are designed for the convenience of the reporter rather than for the convenience of the database encoder and therefore allow the reporter to provide free format textual descriptions of the occurrence. The level of detail and accuracy is consequently not very consistent. All these factors make voluntary incident reporting systems not very well suited to conduct statistical analysis or

[66] An aborted take-off is a reportable occurrence [CAA 2005].

to provide quantitative data to the causal risk model. They are more useful in helping to gain insight into potential accident scenarios and the causal pathways.

8.5.4 In-flight recorded data

For the purpose of accident investigation, aircraft have been equipped with 'black boxes' since the 1960s. The 'black box' is actually orange to facilitate quick location in the aftermath of a crash and consists of two separate units: the Flight Data Recorder (FDR) and the Cockpit Voice Recorder (CVR). Regulatory requirements such as FAR § 121.344 prescribe the parameters to be recorded and the sample rates. These include airspeed, altitude, attitude, control inputs, control surface positions, ILS deviations, FMS selections and warnings [FAA 2003], see Figure 24.

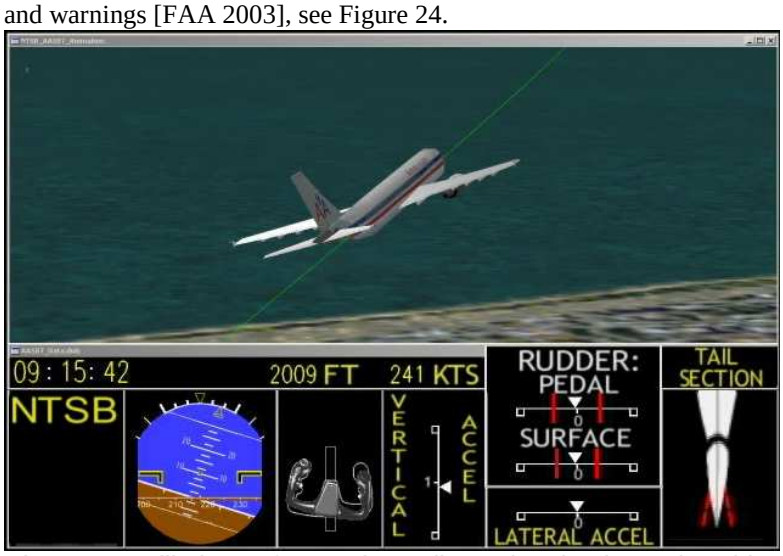


Figure 24: Still picture from a three-dimensional animated accident reconstruction of the flight for American Airlines flight 587, which crashed shortly after take-off from JFK International airport on November 12, 2001. The lower portion of the picture depicts a set of instruments and indicators, which display selected parameters from the flight data recorder. The investigation was performed by the National Transportation Safety Board NTSB.

It has always been common practice in the aviation industry to keep records of failures of components in service operation with a view to optimising availability of the aircraft. When flight data recorder systems developed from analogue tape recorders into digital data storage devices, many airlines quickly realised that flight data analysis could also be used to improve the overall efficiency of the operation by providing a better overview of the reliability of aircraft systems and thus allowing the airline to anticipate on aircraft maintenance actions. By closely monitoring engine vibration levels for instance it can be observed that an engine failure is imminent and the engine can be replaced and repaired before the very costly failure actually occurs. Quick Access Recorders (QARs) were introduced for this purpose. QARs register similar data as the Flight Data Recorder but are not as crash resistant. They allow routine flight data analysis because extracting the data from the aircraft is much less cumbersome than in the case of the Flight Data Recorder. Almost immediately airlines expanded the use of in-flight recorded data from reliability analysis to safety management. Regulatory authorities have also understood the safety potential of routinely analyzing in-flight recorded data and have required that all operators of aircraft of a maximum take-off mass in excess of 27,000 kg must have implemented a

flight data analysis programme as part of its accident prevention and flight safety programme, in accordance with ICAO Annex 6 part 1 [ICAO 2002a]. Hundreds of parameters can be recorded with a typical sample rate of 1 Hz. For each of the parameters recorded with the QAR, upper and lower threshold limits can be defined, after which stored data can be screened for threshold exceedances. Even when no thresholds are exceeded, the data provides valuable information that can be used to quantify some of the risk model elements. Van Es & Van der Geest [2006] provide an example of the use of data from quick access recorders to analyse the operational landing field performance of two aircraft types, see Figure 25.

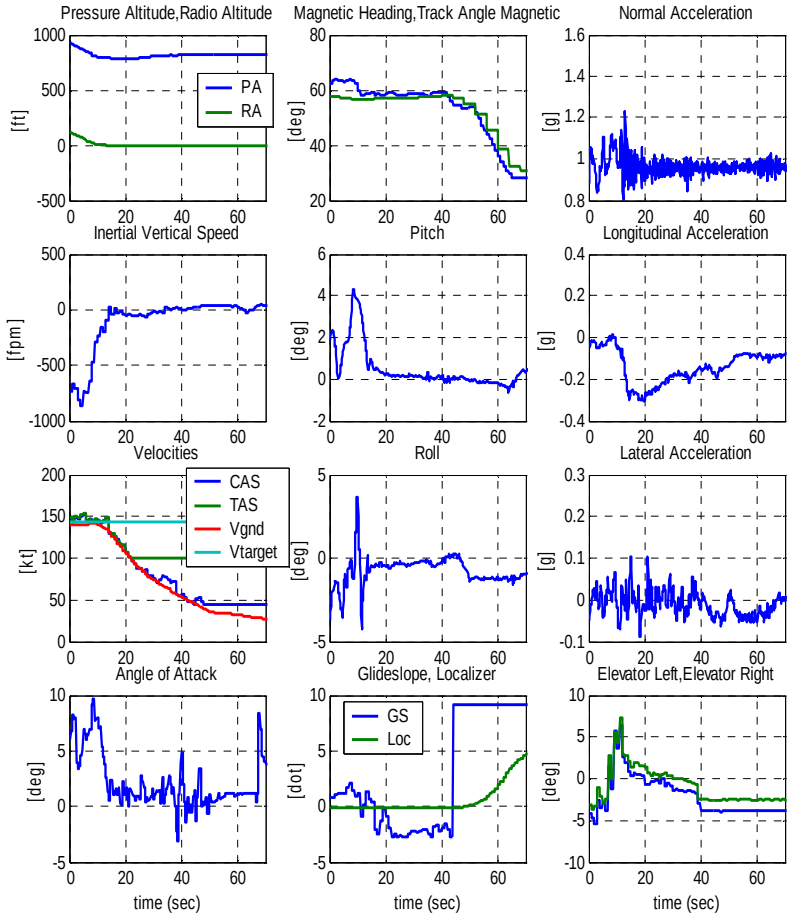


Figure 25: Example QAR time series Boeing 737-400. Source Van Es & Van der Geest [2006].

For airlines, reliability and accuracy of such data is essential. Keeping track of all abnormal occurrences is not only important for flight safety, but first and foremost for their operational management. Abnormal occurrences are usually associated with additional costs, e.g. because of delays or because of extra unscheduled maintenance, and to record these occurrences is therefore vital for the airlines' profitability. This, together with the impressive data capturing capabilities of modern Quick Access Recorders, makes airline

data the most accurate and detailed data, suited to quantify the causal risk model. Unfortunately for model developers this data is often company confidential because of the competition and liability issues. Nevertheless, experience has shown that airlines are sometimes quite willing to share data, provided that certain conditions are met. For quantification of the CATS model for instance, KLM has provided in-flight recorded data. However, a prerequisite for using this data is that the descriptions and definitions used in the flight data program are sufficiently similar to the model elements. In the CATS model, fault tree events were defined without considering the definitions used in flight data recording programs. A result was that very few (less than 10%) fault tree events could be quantified directly from in-flight recorded data [Roelen 2008].

8.5.5 Expert judgement

Experts may have valuable knowledge about parameters for problems in their field of expertise. Quantification and aggregation of experts' knowledge is frequently used for assessing variables for which other adequate data are lacking, either because values for quantified variables are not registered or because objective units of measure simple do not exist, like in the case of safety culture. The experts' knowledge is not certain, but comes with an implicit level of subjective confidence, or degree of belief. Often information from experts is obtained and combined in an informal and ad hoc way, but formal procedures for eliciting and processing expert judgement are increasingly applied, particularly when the quality and transparency of the results are important. The overall goal of these formal methods is to achieve rational consensus in resulting assessments. This requires that the process itself optimises performance, as measured by valid expert judgement performance criteria. Performance criteria are based on 'calibration', that is, assessments of uncertain quantities, closely resembling the variables of interest, for which true values (e.g., from experiments) are known beforehand. Criteria for analysing 'calibration' assessments are closely related to standard statistical methods, and are applied both to expert assessments, and to the combinations of expert assessments. The underlying assumption is that experts' performance on calibration variables predicts their performance on the variables of interest. If systematically obtained from experts who are both knowledgeable and well-calibrated for the relevant area of expertise, expert judgement can offer quite acceptable accuracy of quantification [Cooke & Goossens 2000]. The methods have proven to be mature and provide a scientific tool for achieving additional data that would otherwise remain unavailable [Goossens et al 2008]. The experts should preferably combine analytical skills and the ability to provide judgement on unknown (e.g. future) operations. As was explained in section 4.4, test pilots are specifically trained to provide quantitative estimates on variables for which other data is lacking and therefore test pilots are particularly suitable to provide expert judgements on flight operational topics and should be preferred above line pilots in this respect. In the CATS project, expert judgement was extensively used to quantify sub-models of flight crew and air traffic controller performance [Roelen et al 2007]. The drawbacks of the use of expert judgement are primarily practical issues: finding proper experts is not always easy and expert elicitation can be a time consuming activity.

8.5.6 Empirical studies

Empirical studies can sometimes also provide valuable information for quantification of causal model elements. Human factors have been an area of quite extensive study and some of the research results are available and usable to quantify causal model parameters. A good example is flight crew fatigue. Commissioned by the Netherlands Civil Aviation Authority, the Aviation Medicine Group of TNO Human Factors has conducted a number of field and laboratory studies on the different determinants of aircrew fatigue and their effects on alertness and performance of aircrew. The studies, employing subjective and objective

measures of in-flight performance and alertness, concerned quality and duration of sleep, effects of early reporting times, effects of night flying, effects of alcohol and medication, and the effects of countermeasures, such as onboard sleep (augmented crew) and pre-planned napping in the cockpit seat [Simons & Valk 1993, 1997, 1998, Simons et al 1994, Valk & Simons 1996, 1998].

The results of these studies provide an extensive database on factors causing aircrew fatigue and impaired performance and alertness in flight. This database is further complemented with data collected by fellow participants in the European Committee on Aircrew Scheduling and Safety. Based on the results of the aircrew fatigue studies, a causal model for flight crew fatigue/alertness has been developed [Roelen et al 2002], see Figure 26. Flight crew fatigue is determined by the fitness of the crew before the flight (pre-flight fitness), whether it is a day or a night flight, and in-flight operational factors (operational loads). Factors that influence pre-flight fitness are recent workload and the quality of preflight sleep. Operational factors are the flight duty period, and the quality of in-flight rest facilities. In-flight rest facilities are only used on long haul flights. This means that there is a *conditional dependence* between flight duty period and rest facility, hence the arrow linking the two nodes.

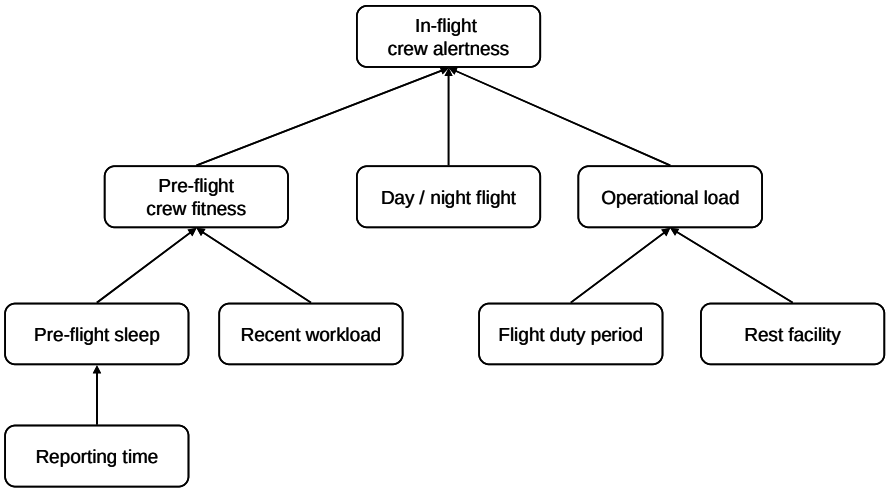


Figure 26: Causal risk model of in-flight crew alertness. Source: [Roelen et al 2002].

The effects of psychosocial factors and the use of alcohol and drugs are not included in the model, although it is known that these factors do influence flight crew fatigue/alertness [Valk et al 1999, Valk & Simons 1994, Valk et al 1997]. Psychosocial factors have not been explicitly modelled because of the difficulty of defining them and the variability of the effects they have on different persons. Alcohol and drug use have not been included because of a lack of good quality data and the difficulty in separating out the effects of alcohol on pre-flight sleep. However, these effects are implicitly included in the model because they will have influenced the performance of the flight crew that have contributed to the dataset that has been used for quantification. This is an important point to keep in mind generally and underlines the need to be explicit about data sources and what they include.

It takes years to set up and conduct experiments such as those described above for flight crew fatigue, and it seems unlikely that such experiments will be conducted for the sole purpose of quantifying causal risk model parameters. But many experiments are already being conducted for other purposes, and many of those are in the field of human performance. This is exactly an area where incident reporting systems and especially in-flight recorded data are usually rather weak, so the empirical studies can provide very valuable complementary data which are usable for model development and quantification. Particularly relevant for pilot performance and air traffic controller performance are studies involving flight simulators and air traffic control simulators. The simulators provide a controlled environment for analysing influences on human performance. An example is the use of the Boeing 747-400 level D[67] qualified full fight simulator at NASA Ames in an experimental study of the balked landing and missed approach manoeuvres to provide data for pilot control behaviour for different aircraft configurations and weather conditions [Hörmann et al 2005, Hosman et al 2005, De Leege et al 2006]. Another simulator at NASA Ames, the Advanced Concepts Flight Simulator, was used to conduct human in the loop simulations to understand the factors that contribute to taxiway navigation errors [Hooey et al 2000]. Another study at NASA Ames involved human in the loop simulations with air traffic controllers and included measurements of controller workload [Smith et al 2004], similar experiments have been conducted at NLR [Hilburn et al 1998]. Applying such data may require some flexibility and creativity from the causal model developers because the empirical studies may provide data on parameters that are defined slightly differently than those in the model. Unfortunately, simulation facilities to study organisational influences do not exist.

8.5.7 Safety audits

Audits are becoming an interesting source of data as safety audits are beginning to be widely used and, most importantly, standardised. A great advantage of safety audits is their ability to capture information on the quality of safety management and safety culture. Two audit programs are especially important in this respect; the Line Operations Safety Audit (LOSA) and the IATA Operational Safety Audit (IOSA).

In a Line Operations Safety Audit (LOSA), trained observers take in-flight notes from the cockpit jump seat to evaluate several aspects of crew performance. LOSA originated at the request of Continental Airlines as an instrument to check whether crew resource management (CRM) concepts taught in training actually transferred to the line [Helmreich et al 1994]. Later, threat and error management performance were added to the analysis. At the core of the LOSA process is a model of threat and error management, which provides a framework for data collection. In-flight observers record the various threats encountered by aircrew, the types of errors committed, and they record how flight crews operationally deal with these situations to maintain safety. A large LOSA dataset, airline de-identified, is maintained by the University of Texas [Klinect et al 2003, Klinect 2005]. ICAO acts as an enabling partner and promoter in the LOSA program [ICAO 2002b] and the FAA encourages airlines to voluntarily conduct LOSA programs [FAA 2006]. De-identified LOSA results (all references to the operator are removed) are available from the University of Texas, although experience in the CATS project has shown that it is difficult to access the data and use it [Lin et al 2008]. A potential weakness of the data is that the (causal) link between threats and errors is subjectively made by the in-flight observers. For example, during a flight the following observation was made:

[67] Flight simulator fidelity is indicated by levels A-D, with D the highest level of fidelity.

"*While in descent the crew was advised by ATC that their runway was to be 26R. After entering downwind at 6000 ft, the runway was changed to 27. The Captain overshot the runway 27 centerline slightly while trying to brief the new approach*".

The observer classified this event as follows:

threat code	ATC runway change
threat outcome	linked to flight crew error.

The observer in this case subjectively assumes the centreline overshoot is caused by the late runway change and not by for instance poor training or strong crosswinds. The data are therefore only valuable when observers are very familiar with an operational flight deck environment and very familiar with the LOSA threat and error management framework. Typically retired pilots are used who have had 5 days of training on the LOSA methodology [Klinect et al 2003], which makes it plausible that indeed the data is sufficiently valid. Unfortunately, this type of operational observations is only good for capturing active failures, latent failures are likely to remain hidden. LOSA data could be used to populate those parts of the model that represent flight crew performance, including the most direct managerial influences on this performance.

The name 'audit' given to this data source is an unfortunate one, since in general safety management literature, the term is confined to check on the safety management system and its functioning. LOSA, in contrast, is a behavioural observation technique which confines itself to the actions and reactions of flight crew and has no direct data about organisational factors influencing those.

The IATA[68] Operational Safety Audit (IOSA) conforms much more to the traditional safety management audit and is more suitable for identifying latent failures. IOSA is designed to assess the operational management and control systems of an airline. Table 4 gives an overview of the areas that are covered during the audit [IATA 2008a]. During an audit, an operator is assessed against the IOSA Standards and Recommended Practices contained in the IOSA Standards Manual [IATA 2008b]. The standards and recommended practices cover issues such as training and qualification, provision of resources, emergency response, procedures, communication and coordination. To determine conformity with any standard or recommended practice, the IOSA Auditor will assess the degree to which specifications are documented and implemented by the operator. Registration is the final objective of the IOSA audit process. An airline that has been audited under the IOSA program and demonstrated full conformity with the IOSA standards will be entered in the IATA registry as an IOSA Operator. The audits are conducted by organisations that are accredited by IATA [Bisignani 2006]. The main disadvantage of this program is that the data, common to virtually all data that is collected by IATA, is difficult to get hold of. The audited airline is the sole owner of its audit report. An interested party desiring access to an audit report must submit an application to IATA, which retains electronic versions of audit reports in a custodial capacity. Any and all requests for access to an audit report must be expressly authorised by the audited airline before IATA will grant access.

[68] International Air Transport Association (IATA) is the air transport industry global trade organisation. Its members comprise some 250 airlines.

Another problem with the use of audit results is the danger of lack of focus. If the goal of the audit is to assess an organisation's effectiveness for a focussed objective such as major hazard control, the auditors (and those that use the results of the audit) must check that the links to the major accident scenarios are in place, not just the links to other parts of the company processes [Hale 2001]. The only auditors who can penetrate the heart of the matter are ones with sufficient knowledge of the technology or activity to know what the crucial aspects of control are. For both the LOSA and the IOSA audits this seems to be the case, so indeed results from these audits are useful and relevant.

Data from LOSA and IOSA will provide exposure data for the model, as they will show how often particular deviations (flight crew errors in the case of LOSA, managerial and control system failures in the case of IOSA) occur. This can then be compared to the occurrence of the same deviations in accidents to estimate the relative contribution of the errors to the accident.

Table 4: IOSA standards

Discipline	**Activities and Processes**
Organisation and Management System	
Flight Operations	Line flight – flight deck operations
	Simulator session
	Training flight
Operational Control and Flight Dispatch	Flight planning
	Flight monitoring
Aircraft Engineering and Maintenance	AD/ASB process
	Maintenance activities
	Maintenance processes
Cabin and Cargo Compartment Operations	Line flight – passenger cabin operations only
Ground Handling	Weight and balance calculation
	Ground handling activities
Cargo Operations	Aircraft loading or unloading
Operational Security	Baggage reconciliation
	Pre-Board/hold room screening
	Aircraft access control
	Preflight crew security briefing

8.6. Denominator data

Quantification of the occurrence rate of certain events requires not only a counting of the number of occurrences of the events, but also requires counting the associated number of attempts.

Occurrence rate = number of occurrences of event / number of attempts

The 'number of attempts' is what we call denominator data. Event occurrences are routinely reported and registered, but the non-events, the number of attempts, quite often remain 'unnoticed'. Efforts focus on event data capture, but quantification of rates is impossible without denominator data. When the objective is quantification, non-event data is just as important as event data. Due to the large number of variables and combinations of variables, collection and analysis of denominator data is no easy task. Consider for example the issue of landing overrun accidents. A landing overrun accident occurs if during landing the aircraft is not able to stop before the end of the runway. The hypothesis is that one of

the causal factors of such accidents is a wet runway due to rain. To test this causal claim we must know the number of overrun accidents, the number of overrun accidents that involved wet runways, the total number of landings regardless of the runway condition and the number of landings on wet runways. The accident data is easy to find in accident and incident databases, the total number of landings can be found in air traffic statistics such as airline timetables [Vos 1996], but nobody records the number of landings on wet runways. To circumvent this difficulty we must combine weather information with traffic data. This approach was followed by Van Es et al [1998] to show a five-fold increase in risk for passenger jet aircraft landing on a wet runway. In general, denominator data can be obtained by combining different databases. Databases that are typically required as a minimum for a causal risk model are those that capture data on aircraft movements (origin-destination), aircraft utilisation (cycles for each aircraft), airports (runways and navigation aids), weather, aircraft (physical characteristics, equipment and performance), airlines (state of registration, fleet size, aircraft types, destinations, financial situation), etc.

In-flight recorded data is potentially a significant source for denominator data, but currently the focus there is primarily on parameter threshold crossings. Data on parameters that stay within their normal operational boundaries is usually discarded, although there are some initiatives to conduct flight data analysis of full datasets including normal parameter values as well as parameter threshold crossings [Roelen et al 1998].

Denominator data for human (flight crew) performance can potentially be obtained from LOSA (see also previous section). Although LOSA was developed for individual airlines to improve their safety performance, we are able to get a much broader picture of flight crew behaviour in normal operations by combining the results of several airlines [Klinect 2005]. The potential of using LOSA data for this purpose is also recognised by Castano & Graeber [2002], but it can only be fulfilled if agreements can be made with the airlines and pilot's associations on collection and distribution of this information, a difficult but certainly not impossible task.

8.7. Using the data

Existing databases can only be used for quantification of a causal risk model if it is possible to match the data set to model elements. In the case of databases that store occurrences in a free text format, the mapping of occurrences to model elements requires analysis and interpretation by the model developer and hence introduces subjectivity. It will also involve a significant manpower. Automated keyword searches can be used for pre selection, but after that each individual incident description will have to be read to determine if it really meets all criteria.

If data is stored according to a classification system it all depends on the way in which the definitions and criteria of the classification system match with the definitions and criteria used for the model elements. Ideally, there is a one to one match from the database to the model elements. Reality is often different. An airline's flight safety reporting system can for instance contain the event descriptor 'ATM confusing instructions'. Mapping this event descriptor to model elements requires information on the content of the confusion (is it not clear to which aircraft the instruction is intended or is the phraseology non-standard, etc.) and the context (e.g. is it in the air or on the ground). Additional information must be obtained, for instance from the free-text description of the occurrence, to get a one to one match. This is time-consuming and the process cannot be automated.

Ambiguity in the classification system is also problematic. The classification system of ECCAIRS is an example of a system that contains ambiguity. The same occurrence can be coded in different ways and criteria and definitions are not sufficiently clear. For instance the event type with code 2060200 'aircraft damage caused by foreign object' is defined as 'an event involving foreign object damage to the aircraft'. The event type with code 2060800 'damage caused by object' is defined as 'an event involving damage to the aircraft when the aircraft was struck by an object'. The difference between an 'object' and a 'foreign object' is not clear. Similarly ambiguous are events 2061000 'damage caused by turbulence' and event 2061200 'damage caused by wind'; the difference between 'turbulence' and 'wind' is unclear.

The problem is amplified when different databases need to be combined to quantify a model element. Take for instance the event 'unstabilized runway approach[69]'. Airlines record unstable approaches automatically with their flight data monitoring program, so most airlines have accurate data on where and when unstable approaches have occurred. But if airline A defines 'correct speed' as between Vref and Vref +15 kts, and airline B considers an approach speed between Vref and Vref+20 as 'correct', the data from airline A and B cannot be straightforwardly combined to estimate the unstabilized approach probability to be used in the model. Even for seemingly unambiguous parameters such as, say, crosswind, there are differences in how speed measurements are averaged over time, if and how gusts are included, at what height and location the wind is measured, etc. For parameters which cannot be measured directly and are interpreted and classified before they get stored in the data system it is even more complex. When complex data systems are used, like ICAO's ADREP, there is the possibility of differences in interpretation. Consider for instance the two following 'events' in ADREP:

> 2010101 Altitude bust
> An event related to the aircraft not obtaining/maintaining the assigned altitude
>
> 2020517 Deviation from clearance – assigned flight level
> An event related to a deviation from an air traffic control assigned flight level

Considering that a flight level indicates an altitude, and that altitudes can only be assigned by Air Traffic Control, it is difficult, if not impossible, to discern the difference between these two events.

A way to minimise these difficulties is by using definitions and criteria from existing data classification systems in defining the model elements. Despite the inherent problems of ECCAIRS it is prudent to use this as the main reference. ECCAIRS' classification system is a European and world standard (as it is similar to ADREP) and feeds the largest European database on aviation accidents and incidents. A significant amount of resources was spent to develop ECCAIRS and associated legislation and it took a lot of effort to have it accepted by most of the European aviation authorities[70]. Access to the ECCAIRS database is not restricted and the associated software is freely available. Development of a causal risk model that uses the same definitions as the database is beneficial to both the model and the data base. For the database the advantage is the possibility to use the data for

[69] An approach is considered stabilized when the aircraft is in the correct landing configuration and with the correct speed on the correct glide path, and only small changes in heading and pitch are required to maintain the correct glide path.
[70] Some authorities, like the British CAA, refuse to accept ECCAIRS.

analysis instead of just storing it, and for the model the advantage is the continuous and direct feeding of the model with the most recent information from an accepted system, increasing the likelihood that the model will be accepted as well.

In the CATS model the breakdown of events and causal factors into individual model elements was guided by the principle to consider the potential causes of accidents as barrier failures. This has the advantage of being very structured, but also results in base events which cannot always be mapped onto existing data classification systems. An event such as a 'disorientating manoeuvre', which is part of CATS' fault tree related to the spatial disorientating type accident cannot be found directly in ECCAIRS. Ideally, the causal risk model is fed directly and continuously with data from an occurrence reporting system. This will strengthen the credibility of the risk model and simultaneously enhance the usability of the data system. ECCAIRS is a potential candidate, but it is currently not yet entirely suitable due to the ambiguity in the classification system and the errors that are being made during data entry.

Whatever data source is used, verification of the data analysis process must be possible, even after model development has been completed. Therefore accurate bookkeeping is essential if data is used to generate numbers for the causal risk model. The process for obtaining and processing data must be transparent and results must be traceable for third parties. A complicating factor is the confidentiality of some of the data. Good quality data starts with a good quality database. Database control procedures must ensure that each database is uniquely identified by a name and a version number. A procedure must be in place to record, for each data base, the following characteristics:

- Source organisation (e.g. ICAO, Airclaims, etc);
- Description of the type of data (e.g. accident, incident, etc.);
- Applicable time period;
- Scope. This should be defined as specifically as possible. Think of e.g. type of operation, geographical region, etc;
- Data base change log.

Procedures for data selection and analysis must ensure that for each result (probability estimate) the following information is registered:

- The data source (uniquely identified by the database name and version number) that was used;
- Data query;
- Description of criteria used for 'manual' selection of data;
- A copy of the resulting dataset.

A positive feature of the CATS model is that it uses 'real' and existing data for quantification of occurrence frequencies and exposure, and the procedures for data selection and analysis as described above were strictly followed. The NLR Air Safety Database and ECCAIRS were used as main sources of data. Because the NLR Air Safety Database combines information from many different sources the sample size is in general larger than when using a single source of data. As a result, generic as well as specific situations can be represented, like for instance a (generic) accident rate for jet aircraft and a (specific) landing accident rate for Boeing 737 aircraft. If existing data were not available for model quantification, expert judgement was used, applying the techniques developed by Cooke & Goossens [2000]. The values of the variables used in the CATS model are all expressed in objectively quantifiable units. The model and its results are therefore less

vulnerable to differences in interpretation. This all contributes to minimising the potential ambiguity in the model and uncertainty in the model results.

8.8. Conclusions for this section

Quantification of the probability of occurrence of variables and the conditional probabilities in the causal risk model is done by using existing data or by expert opinion. Often different sources of data will need to be combined. It is therefore extremely important to ascertain that the same variable descriptions and units of measurement are used. The use of standard units offers the greatest flexibility to express the condition of variables and is therefore preferred. But sometimes it is impossible or impractical to apply a standard quantitative unit of measurement. We can then use a qualitative rating scale, or select a proxy variable that is closely related to the variable of interest and which can be quantified. One of the main difficulties in developing a causal risk model for aviation lies in the need to express 'soft' factors in objectively quantifiable units. Proper units for the quality of procedures, commitment of personnel, safety culture or non-technical skills of pilots do not yet exist. Proxy variables are also difficult to find, as many of these soft factors are multidimensional. We will have to use less well defined quantities and units of measurement at the risk of adding variability to the model and increased uncertainty of the model results. Simultaneously we must continue trying to find proper units for these complex issues.

For some applications of a causal risk model an outcome using an ordinal scale may be sufficient, but this does not mean that it is sufficient to express each individual element of the model as an ordinal value. A high numerical accuracy may not necessarily be needed for the final model outcome but the numerical accuracy of the underlying causal relations is important. Levels of uncertainty of the models are equally important as point estimates. Confidence intervals for expert judgement data require special attention. It is incorrect to treat confidence intervals provided by experts in the same way as confidence intervals that are the result of controlled experiments. To get a proper picture on the uncertainty of the results of a causal risk model it is necessary to conduct a thorough identification and analysis of all assumptions adopted, including assumptions on parameter values used, and assess the individual and combined effect of these assumptions on the model results.

The credibility of the whole causal risk model depends on the quality of the model structure and on the sources of the numerical data to be used and the confidence that one can have in the figures. The accident investigation report is without doubt the single most complete and accurate source of information regarding an accident. Due to the low frequency of major accidents the investigation reports are not sufficient to populate all parts of the causal risk model. Other sources of information are required to complete quantification of the model. Incident reporting systems capture vast amounts of data with relatively little effort, but during the coding process information may get lost or is incorrectly classified. Voluntary incident reporting systems are not very well suited to conduct a statistical analysis or to provide quantitative data to the causal risk model. In-flight recorded data are the most accurate and detailed data, suited to quantify the causal risk model. Unfortunately for model developers these data are often company confidential because of the competition and liability issues. Nevertheless, experience has shown that airlines are sometimes quite willing to share data, provided that certain conditions are met.

Incident reporting systems and in-flight recorded data are usually rather weak in capturing data on human performance. Empirical data can sometimes be used to fill this gap. Applying the data may require some flexibility and creativity from the causal model developers because the empirical studies may provide data on parameters that are defined

slightly differently than those in the model. Studies involving flight simulators and air traffic control simulators are useful for data on performance of the pilot and air traffic controller performance. Unfortunately there are no simulators to obtain data on organisational performance. The only alternative then is expert judgement. The experts should preferably combine analytical skills and the ability to provide judgement on unknown (e.g. future) operations. If systematically elicited from experts who are both knowledgeable and well-calibrated for the relevant area of expertise, expert judgement can offer quite acceptable accuracy of quantification. The drawbacks of the use of expert judgement are primarily practical issues: finding proper experts is not always easy and expert elicitation can be a time consuming activity.

Quantification of the occurrence rate of events requires denominator data. Due to the large number of variables and combinations of variables, collecting and analysis of denominator data is more than simply counting the number of flights but requires combining the information from several different databases. In-flight recorded data is potentially a significant source for denominator data, but currently the focus there is primarily on parameter threshold crossings. Non-event data is usually discarded, although there are some initiatives to conduct flight data analysis of full datasets including normal as well as threshold crossing data. LOSA results are a potentially valuable source of denominator data for flight crew performance, IOSA results are a source of denominator data for managerial risk control failures within an airline. FDR data and LOSA and IOSA results are considered confidential information by the airlines, so proper agreements with the airlines and pilot associations are required in order to use this kind of data to quantify causal risk models.

To avoid problems due to differences in definitions between model elements and data classifications, a causal risk model should conform to definitions used in an existing database, preferably ECCAIRS because it is widely used and is freely accessible. This will also strengthen the credibility of the risk model and simultaneously enhance the usability of ECCAIRS. But simultaneously ECCAIRS should also be improved. The ambiguity in the classification system should be resolved and the system should be able to provide the data and data quality that is required by the model.

Verification of the data analysis process must be possible, even after model development has been completed. Therefore accurate bookkeeping is essential if data are used to generate numbers for the causal risk model.

Chapter 9. Modelling challenges

In the previous sections we identified three user requirements (section 4) that are difficult from a modelling point of view:

- Representing the role of the human operator (problematic because of different views on how to deal with human operators in quantitative safety assessments, see section 4.4).
- Including an explicit representation of the possible impact of organisation and management on safety performance (Section 6.1 shows that this is still a weakness in the nuclear power industry and other industries).
- Representation of dependencies between model elements. Assembling all available knowledge and data on a very large and complex system and unravelling this complexity into individual cause-effect relations that each can be represented separately (Section 7).

This chapter describes those aspects. It is explained why they are so difficult and what are possible solutions to the problems encountered. The concept of accident archetypes is introduced as part of the proposed solution for getting to grips with the complexity of accident scenarios and the role of human performance in those accidents.

9.1. Modelling human operators

Safety in the aviation system is influenced by human actions and decisions at every level of management and throughout all disciplines. The strongest influence is at the operations level, where the actions of pilots and air traffic controllers play a much greater role in risk control than in such systems as nuclear power and chemical industries, which have been the subject of the majority of human performance modelling in the past. Causal aviation risk models necessarily require causal sequences and probabilities to be estimated for (failure) events involving human actions and decisions. They need to describe the cause of the human actions as well as the consequences of those actions, particularly when they are 'erroneous'. The human is extremely interactive and adaptive and does not function in the same way as hardware components. There is greater variability and interdependence in human performance and people rely much more on error detection and correction as strategies for error reduction and risk control, compared to hardware that is designed to function correctly the first time. To provide the necessary representation of human actions, specific human performance models must be developed. In particular, these models must be able to quantify human error probabilities. These models, in essence, fill in the gap between a defined task in a risk control measure and the delivery systems for competence, commitment, etc. (see also the section on the modelling of safety management). Ideally, the models should include feedback loops to represent the adaptive nature of human 'systems' [Wewerinke 1989].

Techniques for human error prediction analysis can be either qualitative, like task analysis to identify opportunity for error, or quantitative, like identifying probabilities associated with particular error types. There are vast amounts written on human error (types of error and why the errors occur), and on human reliability assessment techniques [Swain & Guttmann 1983, Wiegman & Shappell 2001, Kirwan 1994, Hollnagel 1998, Rasmussen

1983]. Much of this work has been stimulated by the nuclear industry. Human reliability assessment involves the identification of opportunities for error, generally in an operating system, and then calculation of the probability of error. This calculation can be either through structured expert judgement techniques or selection of data from a database [Roelen et al 2000b]. Human error modelling started in earnest in the 1970s. The vast majority of the early development of techniques for hazard identification and hazard representation was based on modelling hardware failures. Since human behaviour and failure is different from hardware performance and failure, new techniques had to be developed to incorporate it. In particular, human behaviour is strongly influenced by recovery actions as well as initial failures and by use of mental 'models' to anticipate things to come.

The classical approach in human error quantification is applied to errors of omission and is to use basic error probabilities that are modified to account for specific circumstances or contexts. Human error probabilities for general types of tasks are adjusted for the influence of possible circumstances or contexts by the application of performance shaping factors [Swain & Guttmann 1983]:

$$P(error) = HEP * \sum_{i=1}^{N} PSF_i * W_i \qquad (1)$$

HEP is the human error probability for a certain generic task, PSFi is the ith performance shaping factor and Wi is the weight of PSFi. Examples of generic tasks are selecting a control, operating a control, etc. These are very small steps in tasks and therefore lose the context of the whole task, its objectives and variability, which all has to be catered for in PSFs. Examples of performance shaping factors are workload or time since last recurrent training. Performance shaping factors are not normally seen to be causal in a strict sense, but are generally thought of as factors that contribute to how erroneous actions manifest themselves [Hollnagel 1998]. A strict methodology for deriving performance shaping factors does not exist. They are known, through experience, to have consequences for how the task is carried out and how frequently errors occur. Two of the most often used techniques that follow this approach are HEART (Human Error Assessment and Reduction Technique) [Kirwan 1994] and THERP (Technique for Human Error Rate Prediction) [Swain & Guttmann 1983]. Application of these and similar first generation methods usually require a task analysis in combination with engineering techniques such as fault and event tree logic to identify where errors are important. These latter methods are better at dealing with omissions/procedures and are not good at supporting the identification of errors of commission[71]. Errors of commission are particularly difficult to identify with techniques that only look at the logical operation and failures of the system according to design and planned procedures. Identifying such errors requires special support that is not readily available in the traditional way that fault trees and event trees were built to include human error [Hale et al 1999]. In particular the construction of fault trees from planned/intended system performance, by systematically incorporating the omission or failure of each intended step, will incorporate many slips and lapses, but miss many mistakes and some violations. Accident and incident analysis has shown that human failures that contribute to severe consequences, and that involve highly trained staff using

[71] An error of omission is a human failure event resulting from a failure to take a required action. An error of commission is a human failure event resulting from an overt unsafe action [NRC 2007b].

considerable procedure guidance, do not usually occur randomly or as a result of simple inadvertent behaviour such as missing a procedure step or failing to notice certain indications. Instead, such failures occur when the operators are placed in an unfamiliar situation where their training and procedures are inadequate or do not apply, or when some other unusual set of circumstances exists [NRC 2007b].

A second drawback of the way in which first generation techniques combine human error probabilities and performance shaping factors is the assumption that the performance shaping factors are mutually independent. This is not a very reasonable assumption. As an example, stress cannot be considered independent of workload. Second generation methods have been developed that try to cope with these issues. Examples of second generation methods are ATHEANA (A Technique for Human Event Analysis) [NRC 2007b] and CREAM (Cognitive Reliability and Analysis Method). A key-issue in ATHEANA is the identification of plausible error-likely situations and potential *error forcing contexts.* ATHEANA estimates the probability of making an error in such situations. According to this approach, equation 1 is only valid for situations in which a strong error forcing context is not likely. The CREAM method starts with identifying the context and common performance conditions [Hollnagel 1998]. It is assumed that the most important factor to estimating human performance is the degree of control that human operators have over the situation or context. CREAM suggests a division of the degree of control into four categories: scrambled, opportunistic, tactical and strategic. CREAM provides a method for determining the control mode (and associated error probability) from factors (called Common Performance Conditions) that describe the context.

In a causal model representation, some of the limitations of human reliability methods like equation 1 can be overcome by using influence diagrams such as Bayesian Belief Nets to represent the influence of the context or performance shaping factors on error probability. In such a BBN the overall error probability is calculated as a conditional probability of a particular set of performance shaping factors rather than a sum of separate and independent performance shaping factors. Dependencies between performance shaping factors are represented in the influence diagram by an arc, in Figure 27 for example the likelihood of PSF 1 depends on the value of PSF 2.

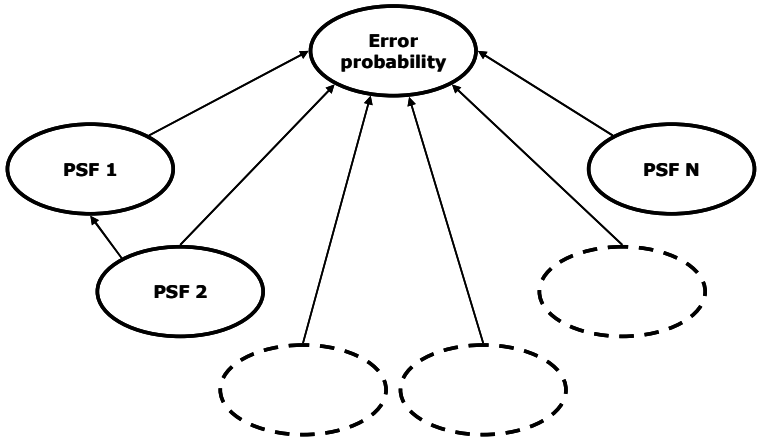


Figure 27: Influence of performance shaping factors on error probability represented in an influence diagram

Such an approach is adopted in the CATS model [Ale et al 2007] where the actual flight crew error follows from the part of the model which the human error model ties into. Mapping of accident and incident data on those fault trees will ensure that both errors of omission as well as errors of commission will be captured in the model. The context is thus defined by the model structure and the associated event in the fault tree is used to estimate the basic human error probability from accident and incident data instead of deriving this from a task analysis as is the case in first generation human performance methods, see Figure 28. The human performance model attached to the fault tree base event that defines the error (of commission or omission) describes the influence of human performance factors on the error probability. Whether this approach is indeed suitable for analysis of errors of commission is still open and needs solving, but is considered to be beyond the scope of this thesis.

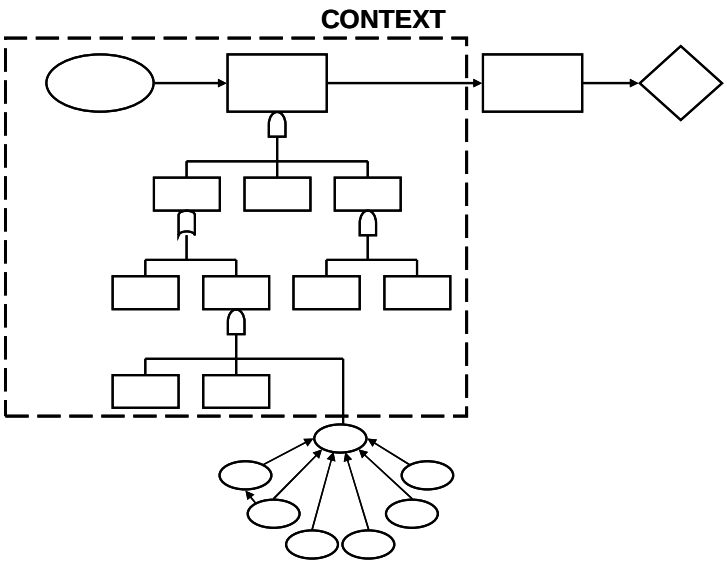


Figure 28: Schematic of the way in which a human performance influence diagram is linked to an accident scenario in CATS.

But the problem still exists of how to select and assess the performance shaping factors or common performance conditions. It is also important to keep in mind how this could be linked to management of the human error probability. The 'levers' for management to influence the human error probabilities are primarily the performance shaping factors since the nature of the tasks are often not subject to change[72]. Safety management can be considered as the process of delivering the necessary resources and criteria for the front-line workforce at the task execution level to operate safely (see next section). Generic Delivery Systems were defined to describe the process of providing the resources and criteria [Bellamy et al 1999]. This means that how 'levers' for management (the performance shaping factors) can be changed can be described as delivery systems, see Figure 29.

[72] Changing the nature of the task is often more drastic and long term (requires redesign, automation, etc) than changing the performance shaping factors.

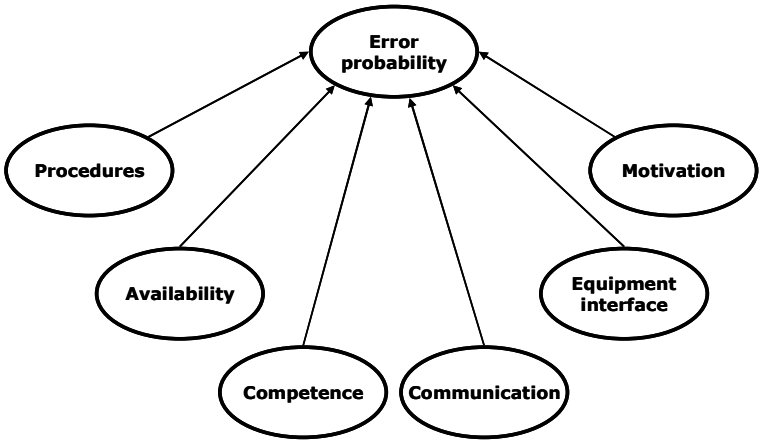


Figure 29: Representation of delivery systems as influences on human error.

They then provide a convenient framework for selecting performance shaping factors and providing the link between management and human error probability. This approach was followed in the CATS model, see Appendix C. It is of course necessary to keep in mind the requirement from the previous chapter to express the performance shaping factors in objectively quantifiable units. Performance shaping factors and delivery systems are also not fully compatible without some work to fit them together, but this can be done as proposed in Table 5 which links error-producing conditions from HEART with the delivery systems. This table is not intended as a definitive comparison, but merely as an illustration of how performance shaping factors could be linked to delivery systems.

Irrespective of whether first generation human reliability methods with performance shaping factors or more advanced methods with control modes and common performance conditions are preferred, a remaining difficulty is the great variability of contexts for a flight crew. Traffic density, aircraft type, weather conditions, route structures, aircraft system status, flight phase, flight delay status, fatigue, crew composition, etc., are all part of the context. The possible combinations of factors are almost unlimited, and establishing the most appropriate performance shaping factors for each combination will be very difficult indeed. A solution to this problem is presented in section 9.3, where the concept of accident archetypes is presented.

Table 5: Proposed linkage of HEART error producing conditions and delivery systems. EPCs from Kirwan [1994].

HEART error producing condition	**Delivery system**
Unfamiliarity with a situation which is potentially important but which only occurs infrequently, or which is novel.	Competence; Procedures
A shortage of time available for error detection and correction.	Availability
A low signal-to-noise ratio.	Interface; Communication.
A means of suppressing or overriding information or features which is too easily accessible.	Interface.
No means of conveying spatial and functional information to operators in a form which they can readily assimilate.	Interface.
A mismatch between an operator's model of the world and that imagined by a designer.	Interface; Competence.
No obvious means of reversing an unintended action.	Interface.
A channel capacity overload, particularly one caused by simultaneous presentation of non-redundant information.	Interface; Communication.
A need to unlearn a technique and apply one which requires the application of an opposing philosophy.	Competence.
The need to transfer specific knowledge from task to task without loss.	Communication.
Ambiguity in the required performance standards.	Procedures.
A mismatch between perceived and real risk.	Commitment; Competence.
Poor, ambiguous or ill-matched system feedback.	Interface.
No clear, direct and timely confirmation of an intended action from the portion of the system over which control is exerted.	Interface; Communication.
Operator inexperience.	Competence.
An impoverished quality of information conveyed by procedures and person-person interaction.	Procedures; Communication.
Little or no independent checking or testing of output.	Procedures.
A conflict between immediate and long-term objectives.	Commitment.
No diversity of information input for veracity checks.	Procedures; Interface.
A mismatch between the educational-achievement level of an individual and the requirement of the task.	Competence.
An incentive to use other more dangerous procedures.	Commitment.
Little opportunity to exercise mind and body outside the immediate confines of a job.	Commitment.
Unreliable instrumentation.	Interface.
A need for absolute judgements which are beyond the capabilities of an operator.	Competence.
Unclear allocation of function and responsibility.	Procedures; Communication.
No obvious way to keep track of progress during an activity.	Procedures; Interface.

9.2. Modelling safety management

ICAO Standards provided in Annexes 6, 11 and 14 oblige States to establish Safety Management Programmes requiring air traffic service providers, aerodrome and aircraft operators as well as maintenance organizations to implement Safety Management Systems (SMS). ICAO defines an SMS as an organized approach to managing safety, including the necessary organizational structures, accountabilities, policies and procedures [ICAO 2006]. The formal introduction of Safety Management Systems in air transport emphasises the fundamental need to be able to predict the effect on safety of changes which are made by decisions of (safety) managers. Regulatory authorities are attempting to implement system safety approaches in regulatory oversight of airline operations [FAA 2001]. Traditional oversight programs have focussed upon observations of operations, facilities, equipment and resources, and individual activities of employees in airlines. Most data that are collected are reflective of assessments of these individual operations. Most measurements of performance currently in use by the regulators are, therefore, tallies, ratios, or other manipulations of the number of times that unsatisfactory observations were recorded, or broad generalisations based upon a small number of observations. The aviation system is, however, composed of organisations of humans who use physical resources as tools to accomplish organisational goals. The performance of the aviation system should be measured more in terms of how the system's processes achieve these organisational goals, as well as individual instances of failure [FAA 2004]. Technical and human performance are interwoven with managerial decisions and this leads to a desire for the causal risk model to be able to properly represent the effect of managerial influences on the safety level, but this is not a straightforward task; organisational factors are enigmatic [Luxhøj et al 2001] in the sense that their effects may be hard to predict and subject to much interaction and managerial decisions are often common mode influences on technical and human performance. The purpose of modelling management is to understand the influence of the factors which can be manipulated. To describe the effect of management on safety we must therefore find a way to link the description of the factors that can be influenced in the managerial system (the 'management model') to the description of the failure events and probabilities on the operational level. The interface between a safety management model and quantitative risk analysis models requires a meeting point of two models which are philosophically different. An important difference is one of mechanism versus holism. A quantified risk analysis looks for causal chains, failure pathways and failure combinations. This mechanistic philosophy would be pure determinism if it was not for the necessary probabilistic and stochastic nature of failures and hence of risk analysis itself. Safety management however is considered a control function aimed at maintaining a particular operational process within the boundaries of safe operation [Rasmussen 1997]. Management is about maintaining functions, technical models deal with failures, but not the process of failure..

A literature review in 1997 concluded that there had been few attempts to produce coherent and comprehensive models of a Safety Management System [Hale et al 1997]. That study also identified the need for an explicit model of safety management as a starting point to assess the completeness of audits to determine whether a safety management system is adequate and how it can be improved. In Hale et al [1997] the Structured Analysis and Design Technique (SADT) is proposed as a modelling approach because of its ability to represent information flows which make up the processes of the management system and the controls which prevent failure. Again, like in Rasmussen [1997], safety management is represented as a control problem. SADT also allows packing and unpacking at different levels of aggregation and thus easily represents different levels of decision making like the classical split between policy, planning and control, and execution. But the SADT model by

Hale et al [1997] merely describes how a safety management system should be structured, it does not allow analysis of what its influence is on failure probability nor how the system behaves following a disturbance (at least not without adding more complication). The desire to analyse system behaviour as a function of disturbances, combined with the view to regard safety management as a control problem, logically results in attempts to apply control theory as a modelling approach. Indeed Rasmussen [1997] concludes that a closed loop feedback point of view is required for modelling safety management.

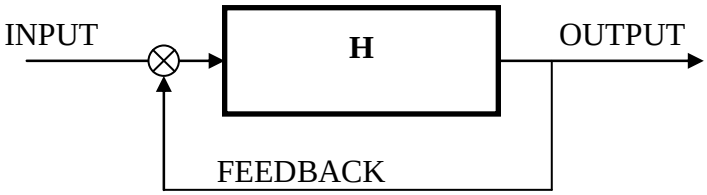


Figure 30: Simple closed loop feedback system

The behaviour of the closed loop feedback system in Figure 30 as a function of the input is completely described by the transfer function H. It is used to describe the characteristics of a physical system e.g. consisting of a mass, a spring and a damper, or an electrical system e.g. consisting of a resistance and a capacity, but also a human or a system in a control task, like a safety management system. If needed, the model can be made more complex by adding feedback loops or even feedforward loops (e.g. to represent proactive elements of safety management, like anticipating on the results of a risk analysis) but this does not really make the analysis more difficult. Simply speaking, if we were able to determine the safety management system's 'transfer function', modelling safety management would be an easy task. But unfortunately things are not that simple. First of all, Figure 30 suggests a distinct 'input' and 'output' signal, but what should be considered the input of safety management, and what is the output? Risk is generated (or prevented) at the operational level, and managerial decisions sometimes influence the hazards (like deciding where to fly) but more often they influence the risk control measures. If input and output could be clearly defined, risk control could be seen as a process of detecting and recovering from deviations from some 'ideal state'. This is essentially a control engineer's view, strongly influenced by the ideas of quality control. It works well for known risks in stable and mature technologies, where one right, safe way of working can be defined, based on long experience, for a single organisation. However, in other situations where risks are incompletely known and technology and work methods are rapidly or constantly changing, or where processes are managed by a complex arrangement of organisations, there is no one "ideal state" which can be usefully defined. Unexpected combinations of essentially normal system factors can produce effects which propagate through complex organisations and find loopholes in the risk defences. Rasmussen [1997] has characterised safety management in such cases as keeping the organisation within a more broadly defined "safe" region, away from the boundaries of uncontrollability. The company is constantly being pushed towards these boundaries by competitive pressures and financial restrictions, or may drift towards them through complacency. Hollnagel's description of accidents as occurrences of 'resonance' [Hollnagel 2004] also seems to fit well within the control theory approach because resonance as a physical phenomenon is typically a control problem, but Hollnagel uses the term resonance merely as a metaphor which is far removed from the physical

phenomenon[73]. As a matter of fact he proposes a modified SADT model to represent his Functional Resonance Accident Model. For the purpose of causal risk modelling, this approach suffers the same limitation as the SADT model proposed in Hale et al [1997]; it might be sufficient as a descriptive model, but it is inadequate for performing calculations. Rasmussen's safe envelope also resembles the approach of Leveson [2004] in which accidents are viewed as resulting from interactions between components that violate system safety constrains. Dekker [2005] echoes Leveson's and Hollnagel's call for control models (see section 2.5) but also fails to describe just how such a model should be constructed other than in the vaguest of terms. So the conclusion is that despite an apparent need for models based on control theory to describe safety management, no such approach has yet been developed that enables quantification of managerial influences on accident risk. The main difficulty seems to be the formal description of the input and output of safety management, the description of the 'ideal state', and identification of the transfer function, including the level at which this should be done; the whole SMS or individual risk control functions?

However there have been efforts in the past to quantitatively link management to risk models. The simplest way to incorporate the effect of management in a causal risk model is by multiplying the probability of occurrence of the top event of the risk model by a factor which expresses the overall quality of management. Initial research efforts therefore attempted to address safety management within a quantitative risk analysis by linking directly from the top event of a number of accident scenarios to the influences in the management system [Bellamy et al 1993]. The work process analysis model (WPAM), developed for the nuclear power industry, represents the dependencies among the parameters that are introduced by organisational factors by calculating new (organizationally dependent) minimum cut-sets for major accident sequences [Davoudian et al 1994]. A simple parameterization of the impact of organisational factors on component performance, as proposed by Mosleh & Goldfeiz [1994] divides the component failure rates into two contributors, the rate of inherent failures and the rate of failures due to adverse organisational influences:

$$\lambda_T = \lambda_I + \lambda_o$$

Here λ_I is the inherent failure rate and λ_o is the rate of failure due to organisational factors. The inherent portion of the failure rate represents failure mechanisms which are 'beyond the control of the organisation' like hardware failures linked to design[74]. By introducing the parameter ω defined as:

$$\omega = \lambda_o / \lambda_I$$

component reliability can be expressed as a function of this ω-factor:

$$\lambda_T = (\omega + 1)\, \lambda_I$$

The ω-factor represents the quantitative influence of organisational factors on component reliability. Such a simplified approach gives little insight into the way in which the

[73] In physics, resonance is used to describe a situation in which the natural frequency of an object coincides with the excitation frequency.
[74] Organisations can influence this during selection of the hardware, but once that choice is made it is beyond their control.

management factors influence the risk numbers, and determining the value of ω is a problem by itself. It is a reasonable approach when 'management' has a single common mode effect, whose management quality and its effects can be assessed (e.g. from audits) and which is fairly stable over time. The aviation system, however, is very different to this; it has up to a hundred independent organisations managing the different causal factors and events. An easy solution is the simple split into different organisations, each with its own ω factor. But again it will be difficult to determine the value of ω if the mechanisms of how safety management influences risk events are not made explicit. A more sophisticated and for the aviation system a more appropriate approach breaks down the quality of management into different common mode factors, where these factors can individually influence risk control measures that are represented as (the probability of occurrence of) the base events of the causal risk model. In Bellamy et al [1999] these common mode factors are derived by considering safety management as providing the necessary resources and criteria for safety critical tasks to be correctly executed. Generic delivery systems are defined to describe those resources and criteria, each of these delivery systems consists of a number of tasks which indicate the process of delivery meeting the needs at the front lines of performance and incorporates a learning loop [Bellamy et al 1999]. The following delivery systems related to human performance were defined:

Competence: The knowledge, skills and abilities in the form of first-line and/or back-up personnel who have been selected and trained for the safe execution of the critical tasks and activities in the organisation. This system covers the selection and training function of the company, which delivers sufficient competent staff for overall manpower planning. Competence should be seen as not only cognitive, but also physiological, i.e. it includes factors such as health and physiology (e.g. vision for pilots).

Availability: Allocating the necessary time (or numbers) of competent people to the tasks which have to be carried out. This factor emphasises time-criticality, i.e. people available at the moment (or within the time frame) when the tasks must be carried out. This delivery system is the responsibility of manpower planning. A critical aspect is planning for peak demands, particularly in emergency situations or other times when deviations from normal or planned operations occur.

Commitment: The incentives and motivation which personnel have, in order to carry out their tasks and activities with suitable care and alertness, and according to the appropriate safety criteria and procedures specified by the organisation or by the workforce themselves for unexpected situations. This delivery system deals with the incentives of individuals carrying out the primary business activities not to choose other criteria above safety, such as ease of working, time saving, social approval, etc. The delivery system for this is often diffuse within companies, but includes many of the activities of supervision, social control, staff appraisal and incentive schemes.

Interface: This covers the ergonomics of the interfaces which are used/operated by operations, inspection or maintenance. Included are both the appropriateness of the interface for the activity and the user-friendliness needed to carry out the activities.

Communication: Communication refers to on-line communication necessary for risk control. It occurs implicitly or explicitly within any task activity when it involves more than one person. Proper communication ensures that the tasks are co-ordinated and everyone knows who is doing what. Communication and co-ordination is particularly critical within the cockpit (captain and first officer), between the aircraft and ATC and at maintenance

shift changeovers or when an aircraft is handed over from operations to maintenance and back again.

Procedures: Rules and procedures are specific performance criteria which specify in detail, often in written form, a formalised 'normative' behaviour or method for carrying out an activity. They may also cover informal 'good practice'. They represent the 'design' of the human tasks.

In addition to these human delivery systems, there are also technology delivery systems:

Design
This delivery system deals with the process for ensuring that the hardware/ software risk control measures and risk control measure elements which have been specified are acquired or designed, either by purchase from outside, or by construction on site, are put in place and adjusted and that the spare parts or replacements purchased and stored for the maintenance phase of their life cycle are the correct ones and are in good condition when used.

Maintain
This factor deals with the management processes for ensuring that the hardware/software risk control measures and risk control measure elements are kept in an effective state as specified by design or as modified to take account of improvements.

These delivery systems are subordinate to a higher level cycle of risk analysis, selection of risk control measures and monitoring and implementation of these at an integrated level above the more specific learning loops incorporated in each delivery system.

This list of delivery systems corresponds well with a list of 'organisational factors' regarded as important to safety as listed by NEA/CSNI [1999]. Table 6 lists those factors, in arbitrary order, as well as a proposed mapping onto the delivery systems.

In I-risk the 'delivery systems' approach made the link between management factors and risk at a more detailed level of the base events in the risk model [Papazoglou & Aneziris 1999]. The output of the management model was therefore a set of weighting factors, one for each of the specified influences on each of the technical parameters determining each of the types of failure events. A richer insight into the causal links between management factors and technical failures was provided, but of course this method required much more effort to develop. Even more insight is provided by linking safety management to operational *risk controls* instead of plain technical failures. Later research efforts therefore made an attempt to represent the influence of safety management on the life cycle of *safety barriers*. A safety barrier is a component or procedure that is installed to prevent or mitigate hazards. Their life cycle includes design, installation, use, maintenance and improvement activities. The basic concept of this approach is that various events in an accident sequence can be prevented or mitigated if the safety barriers corresponding to each function are put into place and if the performance of the safety barriers is assured. Design and maintenance of technology is ultimately a human task, and the interface with which humans have to work needs to be designed and maintained. The risk control steps can therefore be repeated at a different level. This process can be repeated again and again. The iteration of organisational influences, which influences other organisational influences, has been referred to as the Russian doll problem. This is a reality of organisational and societal factors, and not an artefact of the representation method.

Table 6: Comparison of 'organisational factors' with delivery systems

NEA/CSNI organisation factor	Corresponding delivery system(s)
External influences (from outside the boundary of an organisation)	Depends on what the influence is[75].
Goals and strategies	Procedures
Management functions and overview	This factor maps on all delivery systems
Resource allocation	Availability and Design
Human resources management	Availability and Competence
Training	Competence
Co-ordination of work	Communication
Organisational knowledge	Competence
Proceduralization	Procedures
Organisational culture	Commitment
Organisational learning	This factor maps on the learning loops within the delivery systems and at the higher, strategic level
Communication	Communication

Such iterations of the analysis lead inevitably, at some point, outside the initial organisation's responsibility for the risk control measure and its functioning, and into other organisations or parts of the broader society. From a practical point of view, these iterations have to be cut off in the model. In the ARAMIS project [Hourtolou & Salvi 2003, Hale & Guldenmund 2004] the delivery systems were directly linked to barriers. This produces a simpler formulation and explanation of the risk control measures. But the concept of safety barriers introduces additional difficulties as well. In the case of multiple hazards the risk control measure for one hazard can become a factor in increasing the risk of another hazards (see infobox Liberty Bell). In the model proposed for the CATS project [Ale et al 2006] the link to the technical model is represented as risk control measures to stay within a safe envelope. This step was taken to cope with the much greater use of behavioural controls in aviation as opposed to physical ones in chemical processes for which the model was originally developed and to represent the view mentioned in the beginning of this section that safety management is actually a control function, with a monitoring and improvement cycle as its basis [Hale et al 2007]. So the safe envelope should not only be described at an abstract level of the organisation, as proposed by Rasmussen [1997], but also at the operational level as proposed in Hale et al [2007].

A safe envelope is defined as a multi-dimensional space in which activities take place without accident. If the boundary of the safe envelope is breached, damage is unavoidable, although the extent of the damage may still be subject to influence. Inside the safe envelope there will exist a 'defined operational boundary' within which the organization or system wishes to keep the activity so that there is a margin of safety. Risk control measures ensure that the organisation or system stays within the safe envelope and preferably within the defined operational boundary. In order to do so, there must be ways for detecting, diagnosis / decision making and acting on information that the boundaries are being approached and

[75] Some influences, like regulation or market influences, impinge on individual delivery systems, e.g. by making procedures mandatory or influencing commitment. Others affect the strategic risk analysis and learning loop by making new risk controls available, making them mandatory, etc.

will be crossed if no action is taken. The detection-decision-action sequence acts as a risk control measure. Additionally risk control measures serve to locate the 'position' of the system within the boundary and to keep the functions operating which define the boundary and keep it where we expect it to be.

The combination of delivery systems and the safe envelope solves some of the problems identified earlier in representing safety management as a closed loop control system, i.e. that of identification of input/output and description of the ideal state: the input and output signals are defined by the delivery systems, and the ideal state is defined by the safe envelope. All the ingredients are thus available to develop an appropriate control model of safety management. But, in the context of the research question, developing a management model is not an objective by itself. The aim is to develop a causal risk model which combines the description of failure events at the operational level with a description of the managerial system. A suitable approach for representing the failure events at the operational level of a system-wide model seems to be a 'static' model using techniques such as event trees, fault trees and influence diagrams. By definition fault trees, event trees and Bayesian belief nets cannot contain feedback loops. But we want to combine this with a model of a safety management system, which is best represented by a control model in which feedback loops are essential. These two kinds of model are not directly compatible. Either the 'technical model' must allow feedback loops, or the feedback loops in the 'management model' must be removed. Which solution would be the most appropriate? To answer that question we must consider the consequences of either solution.

According to Rasmussen [1997] one of the reasons why a control model is needed is the dynamic interaction between different parts of the (aviation) system that is the result of a 'very fast pace of change of technology'. But what rate of change necessitates the shift from a static to a dynamic model? Is the technology in aviation really changing that fast? The first flight of the Boeing 747 took place in 1969 [Sutter & Spenser 2006] and four decades later the aircraft is still in production. Compared to the dynamics of an actual flight, technology moves at a snail's pace. So at the abstract, strategic level of the organisation there is no need for a dynamic model. At the operational level however, dynamics are more important and a control model seems appropriate.

Another argument for control models is the observation that safety management systems incorporate control loops. The task of piloting an aircraft also is a control task, and indeed sophisticated control models have been developed to describe pilot behaviour in a variety of circumstances. See McRuer et al [1965] for early work, Wewerinke [1989] for an all-round theoretical overview and Hosman et al [2005] for an example of a more recent pilot control model. A risk assessment model in which variables evolve over continuous time, affected by probabilistic disturbances and including ('cognitive') models for human operators involved is TOPAZ as described in Blom et al [2001], which also has its roots in control theory. Human operators are represented in TOPAZ as interacting agents who each have a certain situation awareness. The situation awareness is a dynamic state, and accidents are considered as emergent phenomena from the variability of the situation awareness updating process [Stroeve et al 2007]. Application of this approach requires a high level of expertise in stochastic analysis however. The greater simplicity is the primary reason to represent the flight crew in risk assessment models by simple static techniques like the model described in section 9.1 rather than the more sophisticated dynamic model. In any case, when it is being questioned whether control loops should be introduced explicitly it should be considered if static modelling techniques could be sufficiently adequate. It is here perhaps more than anywhere else necessary to find a proper balance between the desire to

adequately represent the complexity of reality and the need to produce transparent and quantifiable models. Mosleh et al [2004], Perrin et al [2006] and Ale et al [2006] all propose to use simple static techniques like fault trees, event trees and BBNs to model flight crew behaviour. For the same reason simple static models using representation techniques such as fault trees, event trees and Bayesian belief nets can be used for representation of the 'technical' parts well as the managerial influences in the causal risk model. Indeed Bayesian belief Nets are particularly suited to represent the influence of managerial decisions [Roelen et al 2004b, Luxhøj et al 2001].

Section 9.1 has already described a convenient way to provide a link between managerial influences and human performance by describing human performance shaping factors as the levers for managerial influence and representing their influence using a Bayesian belief Net. In the CATS model, the influence of human performance on aviation accident risk is described using three generic human operator models for flight crew, air traffic controllers and maintenance mechanics. Each human operator's influence on the accident sequence (or accident avoidance sequence) is represented by an instantiation of one of these three models. The human operator models have been developed such that the delivery systems can be 'hooked' directly to the performance shaping factors. With this approach, two of the main problems that frustrated previous efforts to model managerial influences have been solved:

- Both the 'management model' and the 'technical model' are represented using static modelling techniques; event trees, fault trees and Bayesian belief nets diagrams, solving the problem of static versus dynamic.
- Managerial influences on human task performance are described by only six delivery systems, connecting to three generic human operator models. This limits the model development burden to acceptable levels.

In a similar fashion, the managerial levers on the hardware (i.e. design and maintain) can be represented in a Bayesian belief Net as separate influences on hardware failure probabilities.

Despite its simplifications this approach in essence is still capable of representing the safe envelope concept. The safe envelope is defined by the accident avoidance pathways of the event sequence diagrams.

Liberty Bell, safety barriers or hazards?

An example of a safety barrier that became a risk cause is provided by the experience of the second American in space, Gus Grissom, immediately following landing after his space flight on July 21, 1961. The spacecraft, Liberty Bell 7, was the first Mercury spacecraft to include a newly designed explosive hatch. Although the hatch had not been tested previously, it was considered to be superior in design to the older model, enabling the pilot to make a quicker and easier egress from the capsule. After splashdown, Grissom began final preparations for egress. "I opened up the faceplate on my helmet, disconnected the oxygen hose from the helmet, unfastened the helmet from my suit, released the chest strap, the lap belt, the shoulder harness, knee straps and medical sensors. And I rolled up the neck dam of my suit." Grissom was lying in his couch, waiting to receive final confirmation that it was time for him to blow the hatch and exit the spacecraft "when suddenly, the hatch blew off with a dull thud". Water flooded the cabin. Grissom automatically threw off his helmet, grabbed the sill of the hatch, hauled himself out of the sinking capsule and swam furiously to get away from the spacecraft. While helicopters tried to hoist-up the sinking

capsule, Grissom realized that he was having a hard time just keeping his head above the water. "Then it dawned on me that in the rush to get out before I sank I had not closed the air inlet port in the belly of my suit, where the oxygen tube fits into the capsule. Although this hole was not letting much water in, it was letting air seep out, and I needed that air to help me stay afloat." His suit was quickly losing buoyancy. The space suit that was designed as a safety barrier against the risk of capsule decompression had become a cause of drowning risk. Unaware of the difficulty Grissom was having in staying afloat, none of the helicopters surrounding him were dropping him a life line. Their rotor blades were churning up the surface of the water, making it necessary for Grissom to swim even harder to keep from going under. As exhaustion set in, he thought, "Well, you've gone through the whole flight, and now you're going to sink right here in front of all these people." Finally, a helicopter approached and dropped Grissom a horse collar. He managed to loop it over his neck and arms, albeit backwards, and was hoisted up [Carpenter et al 1962].

Figure 31: Marine helicopter has astronaut Virgil I. Grissom in harness and is bringing him up out of the water. The Liberty Bell 7 spacecraft has just sunk below the water.

9.3. Complexity, completeness and dependencies

A model that describes causal pathways of accidents in the aviation system will consist of many elements, irrespective of the modelling technique that is used. This is merely a reflection of the complexity and diversity of the air transport system. The model elements are logically linked and quite likely there will be several hierarchical levels. Think of for instance a fault tree, where elements are linked using ‘AND’ and ‘OR’ connectors and the

failure events become more generic when moving upwards in the tree. When two model elements are not directly connected they are by definition independent[76]. The base events in a fault tree are independent and each base event is linked to only one event on the next higher level. This simplicity has drawbacks as well; common cause factors are notoriously difficult to represent in simple fault trees and there may also be cases in which base events are not independent. In the thrombosis example of section 8.2, pregnancy and the use of birth control pills are not independent. A Bayesian belief net is a much better vehicle to represent dependencies and conditional dependencies. From a modelling perspective we would like to keep the number of interconnections as small as possible. More interconnections means increased complexity and hence more development effort and less transparency. But managerial influences are typically common causes so if we want to represent those influences in the model we introduce a great many common cause factors. And when we zoom in on specific areas of the model we will identify all sorts of possible dependencies. Consider for instance the simple flight crew error model that has been developed in the CATS project, in Figure 32.

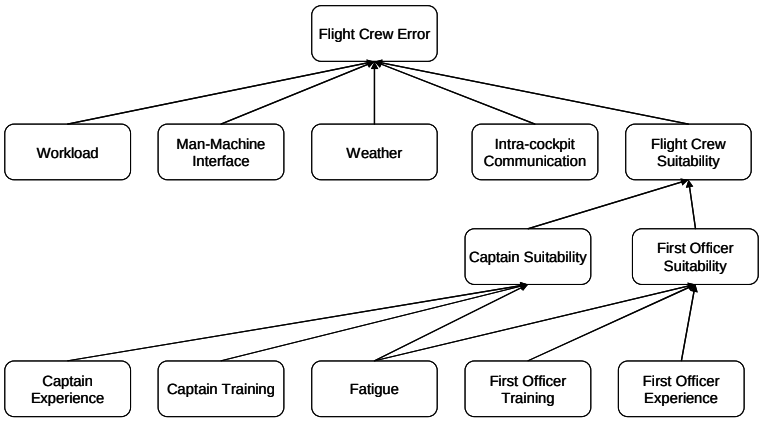


Figure 32: CATS flight crew performance model.

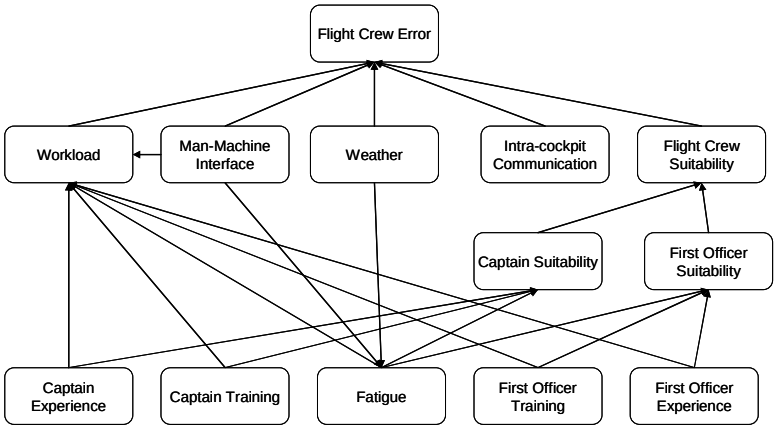


Figure 33: CATS flight crew performance model expanded.

[76] Dependence and conditional dependence are formally defined in the chapter on causality.

According to this model, ‘fatigue’ for instance is independent of ‘weather’, but we know that prolonged flight in bad weather conditions can be quite exhausting, so strictly speaking there should be an arc between weather and fatigue. A good man machine interface design can reduce the task load, so there should be an arc between man machine interface and workload. High levels of fatigue decrease operator capabilities and will, for the same task demand, result in higher workload, so there also should be an arc from fatigue to workload. Likewise, experience and training can improve operator capabilities and hence reduce workload. Good ergonomics design can influence fatigue as well, so there should also be an arc from man machine interface to fatigue. This would then result in the model of Figure 33.

The number of arcs in this model has increased from 13 to 21, and the node ‘workload’ now has 6 parent nodes. When each of the parent nodes only has 2 possible states we need to assess 64 conditional probabilities to quantify workload. The transparency of the model is also affected. This is a result of the increased number of arcs and also because arcs now not only run from the lower to the upper levels, but also the other way around, e.g. from weather to fatigue. While Figure 32 still looks neat and orderly, Figure 33 is becoming more obscure and beginning to look like a ‘giant hairball’. And this is just a simple example; imagine what will happen when an arc is drawn for every possible interaction in a model consisting of, say, 5000 nodes. The assessment burden will then become excessive and it may be simply impossible to complete model development. It is necessary to find a proper balance between model completeness and practical feasibility. The proposed way of doing this is by developing the model from the generic to the specific, starting at the highest level of abstraction and introducing additional detail step-by-step. This can be done systematically if the model is hierarchical, with separate levels of abstraction that are formally defined. The first instance of the model only considers dependencies at the highest level of abstraction and the model is then incrementally detailed by moving to the next lower level at each development step [Mosleh et al 2004].

This top-down approach obviously raises the question whether it will ever be possible to reach a level that is sufficiently detailed for our purposes. Every accident is a unique occurrence. Each accident involves different causes, different components, different locations, environmental circumstances, organisations and people. Representing every single detail of every possible accident is practically impossible. Accident risk modelling requires generalisation and discretisation of such individual occurrences. A model is an abstraction of reality. Individual uniqueness must be generalised such that the essential parts are not lost while a balance must be found between uniqueness and generality.

There are two possible approaches to this; inductive and deductive. The inductive approach is to reason from the individual cases to a generic construct. A deductive approach involves reasoning from the generic to the specific. An example of deductive approach to aviation risk modelling would be to start with the main functions that are required to prevent an accident, e.g. lift, thrust, control, structural integrity and separation between aircraft. Each next step in the modelling process would then make a further functional decomposition. This kind of reasoning is similar to that used to construct Fault Trees. The approach is straightforward and works well when only hardware is involved, but is more complicated when humans, organizations and management play an important role. Task analysis and functional analysis of for instance barriers and delivery systems are sometimes helpful, but there are many risk control failures that are difficult to identify by means of a functional decomposition [De Jong 2004]. Errors of commission may not be found at all. The

inductive approach starts with the individual accidents and then tries to develop a generic construct. This is only useful when certain patterns exist at the higher level of abstraction but is useless if accidents are just random combinations of random events. A review of aircraft accidents indeed shows that often event sequences are very similar, even for cases where human error plays an important role in the accident sequence. An example of such a recurring accident type is an aircraft stall and loss of control following an attempt to take off while the aircraft's wing is contaminated with snow or ice. See infobox 'Take-off stall in icing condition'. Crash due to stall and loss of control following an attempt to take off with a contaminated wing in icing conditions can be considered an *accident archetype*. Another example of an accident archetype is a runway overrun following landing long and fast on a wet runway, possibly in combination with cross- or tailwind. A review of a large set of aircraft accidents identified 33 of such accident archetypes [Roelen & Wever 2005a]. Later in this thesis we will also show that the set of 33 accident scenarios is sufficiently complete to provide a suitable backbone for a causal risk model.

In Roelen & Wever [2005a], each accident archetype is described as a chain of discrete events. This is a useful concept because it is easy to understand; the structure is identical to the way we tell stories. The accident can be prevented by removing any one of the links in the accident chain. Each link represents an opportunity for accident prevention. If the link is removed the sequence of events enters a different pathway that does not result in an accident. This can be represented in an event sequence diagram, where each of the links of the accident chain is an initiating event or a pivotal event.

The advantage of the use of accident archetypes is that this is a data driven approach which results in realistic accident scenarios. It easily captures the most frequent accidents which means that the model quickly focuses on those events that are really relevant for flight safety. The total number of accident archetypes (33 in Roelen & Wever [2005a]) is easily surveyable. There are a very large number of consequences that are the possible result of a set of causes, but there are a much smaller number of consequences that are probable. The accident archetypes capture the most probable cause - consequence relations. The accident archetype / event sequences therefore provide a suitable interface between (more detailed layers of) the model and the users of the model. In Roelen & Wever [2005a] the accident archetypes are described as a series of active failures. This has the advantage that the archetypes are mostly independent. Dependencies are only introduced at the detailed level where also latent failures are described. This is convenient because the model remains more transparent and the assessment burden for quantification is reduced. A disadvantage of the accident archetypes is that the sequence delineation is more or less analyst dependent [Labeau et al 2000]. Peer review will therefore be an essential part of the scenario validation process (see Chapter 10).

Archetype accident example 1: Take-off stall in icing conditions

On March 5, 1993, Palair Flight PMK301, a Fokker 100, crashed shortly after take-off from Skopje Airport, Republic of Macedonia, for a scheduled passenger flight to Zürich, Switzerland. Seventy-nine passengers and four crewmembers were fatally injured and thirteen passengers and one cabin crewmember survived. The accident was caused by degraded aerodynamic characteristics due to contamination of the wings with snow and ice. The situation resulted from an omission to de-ice the aircraft with de-icing or anti-icing fluid in weather conditions conducive to icing. At the time of the accident there was moderate snowfall and the temperature was 0 degrees Celsius [Netherlands Aviation Safety Board 1996]. This accident was not unique, in the decade before the Palair accident there had been at least 7 major accidents of Western built jet aircraft during take-off that were caused by degraded aerodynamic characteristics due to contamination of the wings with snow and ice, see table 7. The dangers of taking off with a contaminated wing were well known to the aviation community. The NTSB's first investigation of an air transport category aircraft accident caused by an attempt to take off with a contaminated wing involved a DC-9 aircraft [NTSB 1993]. That accident occurred on December 27, 1968 at the Sioux City Airport at Sioux City, Iowa. The Safety Board's finding of probable cause in that accident was "*a stall [..] with subsequent loss of control as a result of the aerodynamic [..] penalties of airfoil icing. The flight crew failed to have the airfoil ice removed prior to the attempted take-off from Sioux City*" [NTSB 1970]. This type of accident continues to occur today: on January 25, 2007, a Régional Fokker 100 crashed shortly after takeoff from the airport of Pau, France. Control was lost because the aircraft wing was contaminated with snow [BEA 2007].

The lift which is developed by a wing depends on the angle of attack (the relative angle of the impinging air to the wing chord), and the airspeed. The higher the angle of attack and the higher the speed, the greater the amount of lift developed so long as the airflow over the wing is smooth and adheres to its contour surface. When the airflow separates from the surface, the lift produced by the wing diminishes. The airflow starts to separate from any wing when its angle of attack reaches a critical value, typically 15 - 18 degrees. As the angle of attack is increased further, the lift will decay rapidly. Even a small amount of snow or ice on the wing surface has an influence on the smooth flow of air over the surface contour. Changes in the contour shape and roughness of the surface will cause the airflow to begin to separate from the wing at a lower angle of attack than normal and cause a reduction in the lift. The extent and way that performance will be affected depends on the position of the contaminant on the wing as well as the nature of the contaminant. Generally, contamination of the forward leading edge of the wing will be the most degrading to the lift-producing efficiency of the wing. For this reason, aircraft without leading edge slats or flaps are more susceptible to wing contamination than aircraft with leading edge slats. Note that all but one of the accidents in Table 7 involves an aircraft without leading edge slats (the Boeing 737 is the exception). Due to the influence of the Reynolds number on airflow characteristics, contaminated wings are more of a problem for smaller aircraft than for very large aircraft.

Aircraft that have snow or ice on the wings are de-iced before take-off to prevent these types of accidents. De-icing is conducted by spraying a de-icing fluid, typically a mixture of glycol and water, on the aircraft's surfaces. The de-icing fluid removes snow or ice and also prevents the build-up of additional layers of snow or ice for a predefined time period of time called the hold-over time. Typical values for the hold over time are 30 minutes.

Table 7: Accidents involving attempted take-off with contaminated wing, 1982-1993, Western built commercial jet aircraft.

Date	Aircraft type	Location	Weather	Ref.
13/01/82	Boeing 737-200	Washington, D.C. USA	-4 deg. C. Heavy snowfall.	NTSB 1982
05/02/85	Douglas DC-9-15	Philadelphia, Pennsylvania, USA	-2 deg. C. Ice pellets, snow.	MacIntosh 1993
15/10/87	Douglas DC-9-14	Denver, Colorado, USA	-2 deg. C. Moderate snow.	NTSB 1988
10/03/89	Fokker F-28	Dryden, Ontario, Canada	+ 2 deg. C. Locally heavy snow.	Moshansky 1992
17/02/91	Douglas DC-9-15	Cleveland, Ohio, USA	-5 deg. C. Light snow.	NTSB 1991
22/03/92	Fokker F-28	Flushing, New York, USA	0 deg. C. Drifting snow.	NTSB 1993
05/03/93	Fokker 100	Skopje, Macedonia.	0 deg. C. Moderate snowfall.	Netherlands Aviation Safety Board 1996

Accident archetype 2: Landing long and fast on a wet or contaminated runway

On June 1, 1999, American Airlines flight 1420, a McDonnell Douglas MD-82, crashed after it overran the end of runway 4R during landing at Little Rock National Airport in Little Rock, Arkansas. The captain and 10 passengers were killed; the aircraft was destroyed by impact forces and a postcrash fire. Analysis of the accident revealed that the aircraft had touched down at a speed of 160 kts, which was 29 kts too fast, at 2000 ft past the runway threshold. A normal touchdown point is 1000 - 1500 ft beyond the runway threshold. Because the runway was wet, the braking performance of the aircraft was also negatively affected [NTSB 2001]. Less than a year later, an accident occurred under similar circumstances. On March 5, 2000, Southwest Airlines flight 1455, a Boeing 737-300, overran the departure end of runway 8 after landing at Burbank-Glendale-Pasadena Airport in Burbank, California. The aircraft came to rest on a city street near a gas station off of the airport property. Of the 142 persons on board, 2 passengers sustained serious injuries; the aircraft sustained extensive exterior damage. In this accident the aircraft touched down at a speed of 182 kts, which was 44 kts too fast at 2150 ft past the runway threshold. Again the runway was wet [NTSB 2002b]. Both instances are examples of landing overrun accidents that are the result of landing long and fast on a wet or contaminated runway, possibly in combination with tailwind. Additional examples of this type of accident are presented in Table 8.

Table 8: Runway overrun accidents after landing long and fast on a wet or contaminated runway.

Date	Aircraft type	Location	Touchdown point/ total RWY length (ft)	RWY condition	Wind	Ref.
21/02/1986	McDonnell-Douglas DC9	Erie, PA, USA	2000/6000	snow/wet	10 kts tail	NTSB 1987
14/09/1993	Airbus A-320	Warsaw, Poland	2500/9100	flooded	18 kts tail	
01/06/1999	McDonnell-Douglas MD-82	Little Rock, AR, USA	2000/7500	wet	5 kts tail, 20-25 kts cross	NTSB 2001
22/09/1999	Boeing 747-400	Bangkok, Thailand	3300/10300	flooded	8 kts cross	ATSB 2001
05/03/2000	Boeing 737-300	Burbank, CA, USA	2150/6000	wet	6 kts tail	NTSB 2002b
02/11/2002	Fokker F-27	Sligo, Ireland	1500/3900	wet	3 kts cross	AAIU 2005
02/08/2005	Airbus A-340	Toronto, Canada	4000/9000	flooded	5 kts tail	TSB 2007
08/12/2005	Boeing 737-700	Chicago, IL, USA	2000/6500	snow	9 kts tail	NTSB 2007

How do we know if the set of accident archetypes is sufficiently complete? The accident archetypes should be mutually exclusive categories and the total set of archetypes must represent all probable accident occurrences, so that their results can be added together and should give complete coverage of all accident risks. Based on ICAO's definition, an accident[77] can be divided into subcategories 'personal injury' (including fatality), 'aircraft destroyed' and 'aircraft damaged'. On a high level of abstraction there are four ways for a person to get fatally or seriously injured, three of which result in the aircraft being destroyed or sustaining major damage:

- Personal injury (without the aircraft being damaged or destroyed)
- Collision of aircraft with the ground
- Collision of aircraft with an object
- General disintegration of aircraft.

There can be occurrences in which the aircraft receives only little damage, but the occupants are seriously injured or even killed. Other than security related events, this can happen when there is an abrupt manoeuvre, e.g. a sudden and unexpected turbulence encounter resulting in passengers thrown around in the cabin, or in case of an event involving the cabin environment like a lack of oxygen. A collision with the ground is called 'controlled' when the aircraft has no malfunctions and is under control of the flight crew, albeit that the crew is not aware of the fact that they are flying towards terrain, i.e. controlled flight into terrain. A collision with the ground can also be uncontrolled, when the flight crew has lost control of the aircraft, which can be induced by the crew, an aircraft system/mechanical malfunction or the environment. Thirdly, collision with the ground can also be a forced landing, where the crew makes an unscheduled landing off a runway

[77] An accident is an occurrence where a person has been fatally or seriously injured, or the aircraft sustains major damage or structural failure [ICAO 2001].

because they are unable to continue flight. This happens when for instance the aircraft runs out of fuel and all engines stop producing power. A collision with an object can be an in-flight collision with another aircraft, or it can be a collision on the ground with another aircraft, a vehicle, or other object. A 'general disintegration' is an occurrence where the aircraft's structure is overloaded and damaged, including in-flight disintegration, either due to a structural failure or due to fire or an explosion (e.g. fuel tank explosion or the detonation of a bomb) .

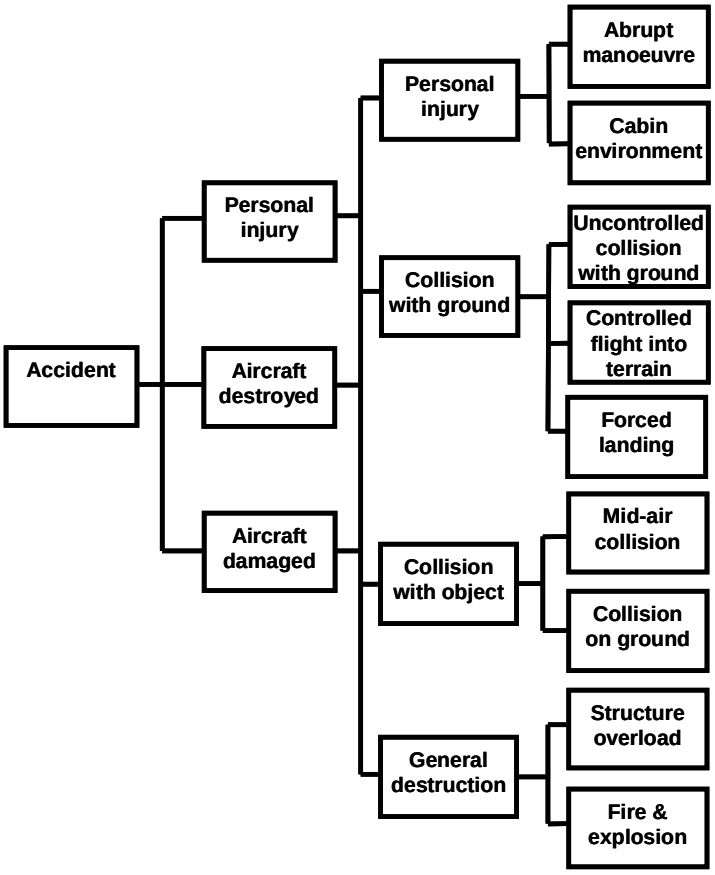


Figure 34 Accident breakdown into subtypes.

The resulting set of accident types is shown in Figure 34. This way of categorizing accidents is generally in agreement with the tradition in aircraft accident statistics (see for instance brochures published by the CAA-UK [1998, 2000] and CAA-NL [2007]), although there is no strict convention. It provides a convenient framework to map the accident archetypes and allows a straightforward check if the set of archetypes is sufficiently complete. A next step is to map past accidents on the set of accident archetypes. In CATS this was done for 8969 accidents without obvious omissions being identified.

By using generic accident scenarios as the backbone of a causal risk model, the model structure is explicitly based on observations in practice, rather than a theoretical approach such as structural decomposition or functional abstraction. Aggregation in the accident scenarios was performed heuristically by the analysts' knowledge of the accident

evolution[78]. An advantage of this approach is the ability to simultaneously capture technical malfunctions, human behaviour and managerial influences without being confined by a theoretical modelling construct. On the down side, this practical approach is predominantly based on observations from the past and 'new' accidents and accident types where risk control measures work so well that they do not happen are not necessarily captured. A failure to capture 'new' accidents is often used as an argument against models that are based on past accidents, but as a matter of fact new accidents are rare indeed [Levy 2001]. Nevertheless, this restriction should be made clear to the user and it should be verified that this is indeed acceptable.

The concept of accident archetypes is also useful in identifying the relevant human performance shaping factors, a problem identified in section 9.1. Because accidents are not random event sequences, not all errors are equally hazardous. In many cases there will be sufficient time for the flight crew to detect an error and recover from the situation, without flight safety being compromised. Instead of starting human performance assessment with the human operator and determining how he or she can fail, it is therefore proposed to start with the accidents and the errors that have contributed to those accidents. The accident archetypes define the types of errors and associated contexts that have been shown to be relevant.

Consider Controlled Flight Into Terrain (CFIT) accidents as an example. CFIT accidents are those in which an aircraft, under the control of the crew, is flown (unintentionally) into terrain, obstacles or water with no or insufficient prior awareness on the part of the crew of the impending disaster [Khatwa & Roelen 1996]. It is a class of accidents that received special attention in the 1990s because it was seen as the deadliest 'killer' in aviation [FSF 1999]. In these accidents, situational awareness errors[79] and tactical decision making errors[80] are dominant. Situation awareness errors were identified in almost 100% of CFIT accidents in the approach phase of flight [Khatwa & Roelen 1996]. Non-precision approaches are particularly risky [Enders et al 1996]. During a precision approach the flight crew receives detailed information on the position of the aircraft relative to the desired approach path, both in the horizontal plane (deviation from the localiser) as well as in the vertical plane (deviation from the glideslope). When such information is missing, i.e. when the approach is non-precision, it is easier to become 'lost'. If during a non-precision approach the flight crew reaches the minimum descent altitude[81] and they do not have the runway in sight the approach should be abandoned. Continuation of the approach is a tactical decision error. This error, under these circumstances, in combination with a situation awareness error, is especially dangerous.

This example of CFIT accidents shows how certain types of human error can be linked to certain types of aircraft accidents. If we are also be able to link certain delivery systems to particular error types, the connection can be made from organisational influences all the way to accident occurrence. Research by Lin [2007, 2008] is aimed at identifying the links between managerial influences on flight crew error with the aim of directly connecting

[78] Peer review is required to validate this aggregation (see section 10).
[79] Situational awareness error was defined as controlling the aircraft to wrong parameters.
[80] Tactical decision making error was defined as failing to revise action in response to a signal to do so or failing to heed warnings or alerts that suggest a revision of action.
[81] The minimum descent altitude is the lowest altitude to which descent is authorized on final approach in execution of a non-precision approach procedure. Continuation of the approach is then only permitted when the runway or runway lights can be seen.

management models (section 9.2) to human performance models (section 9.1). Figure 35 shows the relative contribution of delivery systems to five types of flight crew error. This figure is based on a sample of accident and incident data from the ADREP database. According to this information, competence is the dominant delivery system for all five types of flight crew errors being identified, but the research also shows significant differences for the next most important delivery system across error types. For perception/judgement and aircraft handling the 2nd most influential delivery system is technology interface design, for decisions it is manpower planning and execution, for action in respect to procedures it is communication and coordination, and for operation of equipment there is a large part not specific. Notice that the five error types are described on a generic level rather than being very specific. Again the approach in the analysis is to start at a generic level to identify the most important patterns and trends, and only add more detail if this is required by the user and allowed by the available data.

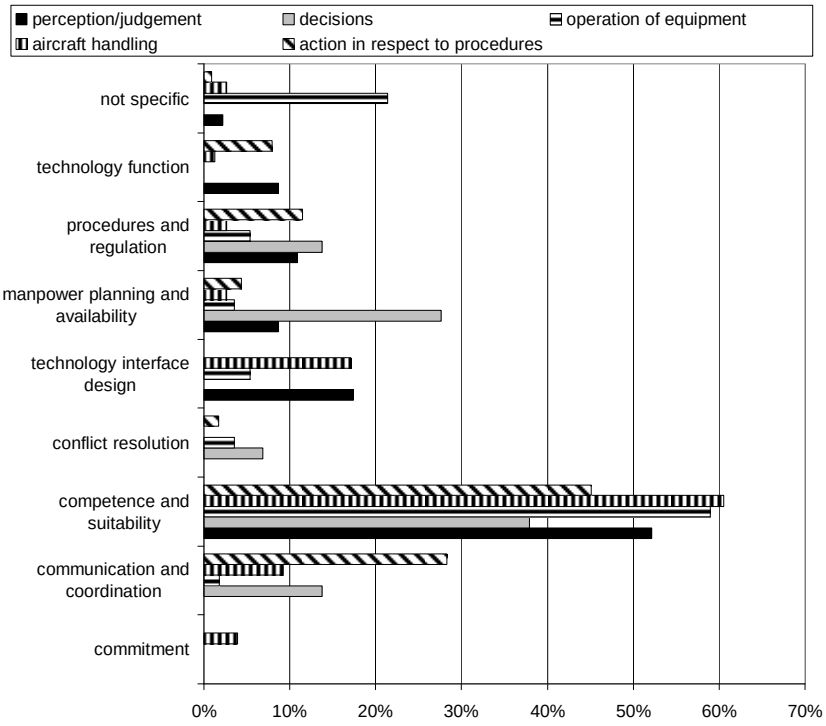


Figure 35: Relative contribution of delivery systems to flight crew error types.

9.4. Conclusions for this section

Causal aviation risk models necessarily require causal sequences and probabilities to be estimated for (failure) events involving human actions and decisions. Because there is greater variability and interdependence in human performance than in hardware performance, specific human performance models must be developed. The classical approach in human error quantification is to use basic error probabilities for specific tasks and to modify those to account for specific circumstances or contexts. This approach is not good at supporting the identification of errors of commission and assumes that the performance shaping factors are independent, which is not a very reasonable assumption. Some of these limitations can be overcome by using influence diagrams like Bayesian Belief Nets to represent the influence of the context or performance shaping factors on error probability. The performance shaping factors can also be described as the output of delivery

systems. They then provide a convenient link between safety management and human error probability. It is also proposed to use the associated event in e.g. the fault tree (or any other type of model representation that is used), which the human performance model ties into, to estimate the basic human error probability instead of deriving this from a task analysis as is the case in first generation human performance methods. Mapping of accident and incident data on those fault trees will ensure that both errors of omission as well as errors of commission will be captured in the model. Whether this approach is indeed suitable for analysis of errors of commission is still open however and needs solving.

There is also a requirement for the causal risk model to be able to represent the effect of managerial influences on safety, but this is not a straightforward task. The difficulty is that the influence of management is always via the failure at the operational level. To describe the effect of management on safety the description of the managerial system must be linked to the description of the failure events at the operational level. This requires a meeting point of two model types that are inherently different. The failure events at the operational level are described as causal chains and failure pathways, but safety management is considered a control function aimed at maintaining a particular operational process within the boundaries of safe operation. This way of formulating safety management as a control function logically results in attempts to apply control theory as a modelling approach and to use dynamic models with control loops. Strictly technically this seems feasible, but it can be problematic to link this to ‘simple’ static models using representation techniques like fault trees and Bayesian belief nets. If static models are used to represent the failure events at the operational level, the management influences must be described in a different way, one that allows the link with these static models. As is proposed in this chapter, this can be done by describing managerial influences on human task performance by six delivery systems, connecting to generic human operator models. The delivery systems are represented as parent nodes of Bayesian Belief Net representations of human operator performance.

There are potentially many interactions between individual model elements and if these must all be represented we may end up with an excessive development and assessment burden and a model that looks like a giant hairball which is not very transparent. It is necessary to find a proper balance between model completeness and practical feasibility. The proposed way of doing this is by developing the model from the generic to the specific, starting at the highest level of abstraction and introducing additional detail step-by-step. This can be done systematically if the model is hierarchical, with separate levels of abstraction that are formally defined and if certain patterns exist at higher levels of abstraction rather than accidents being random combinations of random events. A review of accidents shows that indeed event sequences are often very similar even in cases where human error plays an important part in the accident sequence. A review of a large set of accidents identified 33 of such accident archetypes. These provide a suitable backbone for a causal risk model. The advantage of the use of accident archetypes is that it is a data driven approach which results in realistic scenarios. It easily captures the most frequent accidents which means that it quickly focuses on those elements which are most relevant for flight safety. The accident archetypes create a top level of the model that is intuitively transparent and reduce the assessment burden because dependencies are limited. The concept of accident archetypes is also useful to help in identifying the relevant performance shaping factors for human reliability assessment.

Chapter 10. Model validation

10.1. Introduction

Model validation is usually described as a process of checking if the model provides, for the objective for which it is applied, a correct description of the reality that it represents. An alternative definition is that model validation is a process of checking *the degree to which* the model provides, for the objective for which it is applied, a correct description of the reality that it represents [AIAA 1998]. The fact that a model is by definition unequal to reality is the basis for a validation approach developed by Everdij et al [2006]. This shifts the validation activities from proving that the model is correct to evaluating what the difference between model and reality means in terms of the model output. In this thesis the first definition of validation is adopted.

The reason we need validation is because the model is only a representation of reality and during the process of modelling numerous assumptions and simplifications have been made. Typically, validation involves comparison of the behaviour of the model with the behaviour of the part of the 'real' world that the model is supposed to represent. In this case the represented reality is the safety behaviour of the air transport system and in the model this is expressed as accident probability (or probability distribution) as a function of the probability of occurrence of certain base parameters. Validation of a causal risk model of air transport is problematic because of the complexity of the system, the small number of accidents now occurring and the intangibility of risk. Full validation of a complete causal risk model of air transport is not possible because model results cannot be compared with a continuous observation of an unaltered system; the time period involved in observing reality until sufficient events have accumulated makes real life observations impractical and creating these events on purpose is unethical. Full validation would also involve comparison of model results with reality for every possible set of conditions. For the sort of models that are considered here the number of possible conditions is so large that this is impossible. Moreover a causal risk model of air transport exhausts all accident data for the development of the model and checking the model results with independent accident data is not possible. Therefore alternative methods for demonstrating the 'validity' of the model must be sought.

To circumvent the problems of complexity and intangibility of risk, individual segments of the model with directly observable outputs will have to be validated, where 'validation' is restricted to a few specific conditions. Proof must be provided that the overall process for combining these segments into a single model is sufficiently correct for the objective(s) under consideration and that this not only holds for the specific condition but also for other possible conditions. The proposed hierarchical structure of a causal risk model is advantageous in this respect. The hierarchical structure is composed of accident scenarios (e.g. represented as ESDs) at the highest level of aggregation, direct causal factors of scenario events (e.g. represented as fault trees) at the next highest level, and common mode human and managerial influences (e.g. represented as BBNs) at the most detailed level. The modelling techniques used in each of the 'layers' of the model necessitate specific assumptions and simplifications and by demonstrating the validity of a small section of each layer it is plausible that the results are equally valid for the rest of the layer, provided

that the rest contains no 'new' general assumptions or simplifications. For instance, where the BBNs at the most detailed level of a causal risk model are quantified using expert judgement, by demonstrating the validity of this quantification process for a small part, the validity of this process for the rest of the model components is made plausible as well, assuming the experts there can be considered to be equally knowledgeable. On top of this, there is the issue whether experts have the impression that the model looks right (for instance with respect to the sequence delineation in the scenarios) and behaves as experts expect. This is called face validity and peer review.

This section provides examples of the way in which validation of a causal aviation risk model could be conducted. It simultaneously provides 'proof' for some of the statements that have been made in previous sections. Four aspects will be discussed:

- Validity of the process of decomposition (e.g. breakdown into different accident scenarios) and aggregating the results
- Case validity: comparison of behaviour of (a section of) the model with real world behaviour
- Face validity and peer review
- Assumption analysis

Because of the problems identified earlier (comparison of model results with a continuous observation of the unaltered system is not possible, the development of the model exhausts all accident data, the set of possible conditions is infinitely large) this is all that can (and should) be done.

10.2. Validation of the generic accident scenarios

10.2.1. Validation of take-off and landing overrun probability estimates

This section gives an example of checking the validity of the process of breakdown into different accident scenarios and aggregating the results. Such a validity check was done for a causal risk model developed for the FAA. That model represents risks of all flight phases, but the validity check was limited to take-off overrun accidents and landing overrun accidents. The FAA model is a hybrid causal model combining three hierarchical levels. The highest level is identical to that used in the Dutch CATS model and consists of 33 Event Sequence Diagrams that each represent a generic accident scenario [Roelen et al 2006b]. Take-off overrun accidents are represented as possible end-states in 10 different scenarios, landing overrun accidents are possible end states in 8 scenarios. The model was quantified with worldwide accident data for large Western-built aircraft (MTOW > 5,700 kg) in commercial operations from 1990 - 2003. Quantified model results were compared with three other research reports that contain estimates of the probability of occurrence of take-off overrun accidents or landing overrun accidents. These studies had similar data inclusion criteria, although there are some differences in the time frame. Results of the comparison are presented in table 9.

Table 9: Comparison of model results with study reports

Accident type	FAA Causal Model 1990-2004	NLR study[82] 1980-2004	Boeing study[83] 1990-1999	NLR study[84] 1970-2004
Take-off overrun	1.27×10^{-7}	1.36×10^{-7}	1.4×10^{-7}	
Landing overrun	4.17×10^{-7}	3.87×10^{-7}		5.0×10^{-7}

The match between the model results and the other results are good. All results are based on more or less the same data source and are in that sense not independent, but the comparison is important because it shows that indeed the process of breaking accidents down into generic accident scenarios and then aggregating the results is valid or at least acceptable. The model reproduces the data that was used to construct it so there are no overrun accident scenarios missing from the FAA model.

10.2.2. Completeness of the accident scenarios

The concept of generic accidents scenarios as the top level of the model hinges on the assumption of repeatable accidents. The validity of this assumption was discussed in section 9.3. Aggregation in the accident scenarios was performed heuristically by the analysts' knowledge of the accident evolution. Peer review is required to validate this aggregation (see also section 10.4).

10.3. Validation of a model for missed approaches: case validity

This section provides an example of the validation process for a small segment of a causal risk model. The output of this sub-model is a directly observable variable -the average missed approach rate- which makes validation more straightforward than in the case of intangible variables such as 'risk'. Like every other sub-model of the causal risk model, the missed approach model consists of input variables, a computational engine and an output variable. The model is considered valid if the values of the input variables correspond with reality *and* the value of the output variable as a function of the values of the input variables corresponds with reality. In the example the observed missed approach rate at Schiphol airport is compared with model results. The model includes an aspect of human performance and is therefore representative of the most detailed level of a causal risk model.

10.3.1. Qualitative description of the model

For the purpose of demonstrating causal modelling techniques for safety analysis, a Bayesian Belief Net describing the case of a missed approach was developed [Roelen et al 2002]. A 'missed approach' should be initiated by the pilots when a situation arises that would make the continuation of the approach and landing unsafe. Generally speaking, during a missed approach the flight crew advances the throttle to go-around power, the flap setting is reduced (typically to 20 degrees) and the aircraft is rotated to 15 degrees pitch attitude. The aircraft climbs to a predefined altitude from where a new approach is initiated, or the aircraft diverts to an alternate airport. The purpose of the missed approach procedure is to reject flying into unsafe conditions or under unsafe circumstances and to enable the crew to carry out a new approach and landing under safe circumstances. The missed approach phase is a dynamic and complex phase of flight, requiring decision making, and

[82] [Cheung & Post 2005].
[83] [MacKinnon 2000].
[84] [Van Es 2005].

quick handling by the flight crew and air traffic controller. It is characterised by relatively large changes of the aircraft configuration (from landing to climb out configuration). There can be various reasons why an approach is discontinued and a missed approach procedure is executed. The failure to recognise the need for and to execute a missed approach when appropriate is a major cause of approach and landing accidents. According to a study published by the Flight Safety Foundation in 1998, "the most common primary causal factor [of approach and landing accidents] was judged to be omission of action/inappropriate action, which most often referred to the crew continuing the descent below the Decision Height or Minimum Descent Altitude without visual reference or when visual cues were lost" [FSF 1998]. Although it has been recognised that the execution of the missed approach manoeuvre itself can, if conducted incorrectly, play a role in the sequence of events that results in an accident, this issue is not described in the model validated here[85]. The model is, therefore, strictly focussing on 'failure to execute a missed approach'. This failure to execute a missed approach is a pivotal event in several approach and landing accident scenario's.

A missed approach is ultimately initiated by the flight crew, based on their mental representation of the current situation, a sort of cognitive envelope which the flight crew uses to assess dynamically the safety state of the approach and landing phase. A potentially unsafe situation exists when there is a mismatch between the flight crew's mental representation of the situation and the 'actual' situation. The main factors that are important in the missed approach decision making are considered to be the following:

- Visibility
- Cross wind
- Longitudinal separation from the preceding aircraft in the approach path
- Speed deviation from the reference approach speed at 500 ft altitude
- Crew alertness
- Fuel weight

The corresponding influence diagram is presented in Figure 36. In the missed approach model, the following assumptions/techniques are applied:

- Expert judgement for the quantification of some causal influences,
- Existing databases for the quantification of other causal influences,
- Combination of expert judgement results with other quantified data,
- Bayesian Belief Net to represent human reliability.

Quantification of the model requires quantification of the parent nodes and of the influence of the parent nodes on the child node. This is further described in the next two sections.

[85] In a later development of the model for CATS, the incorrect missed approach execution is indeed represented in the model as potentially leading to an accident.

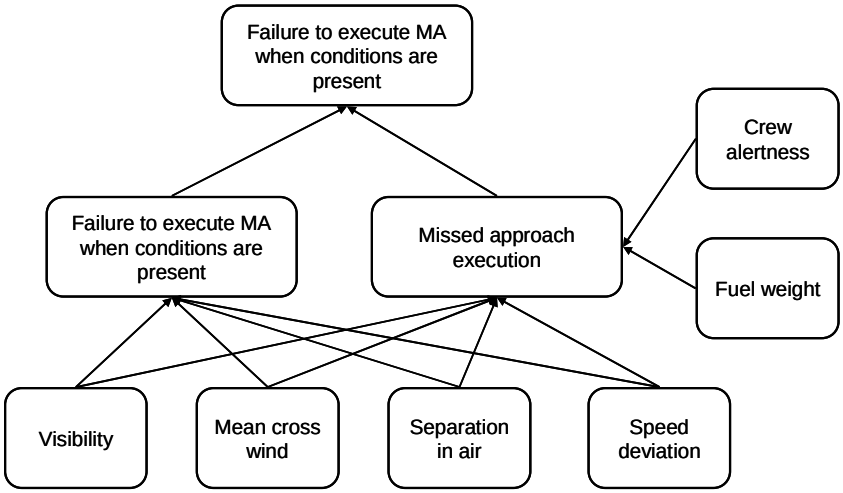


Figure 36: Missed approach model influence diagram

10.3.2. Quantification of the model variables (the parent nodes)

The initial model described in Roelen et al [2002] is a discrete BBN in which each of the parameters can only take two values; 'OK' or 'not OK'. Because of the inherent limitations of a discrete BBN the model was further developed into a continuous BBN with variables modelled as continuous quantities [Morales et al 2008]. The probability distributions of the parent nodes were quantified using readily available data from the NLR Air Safety Database. Because the validation centres around comparison of model results with observed values for Schiphol airport, the aim of the quantification was to have probability distributions that are representative for the situation at Schiphol.

Visibility

Described in metres, visibility is based on 27 million observations across Europe. The cumulative probability distribution is presented in Figure 37, the unit of measure for visibility is meters.

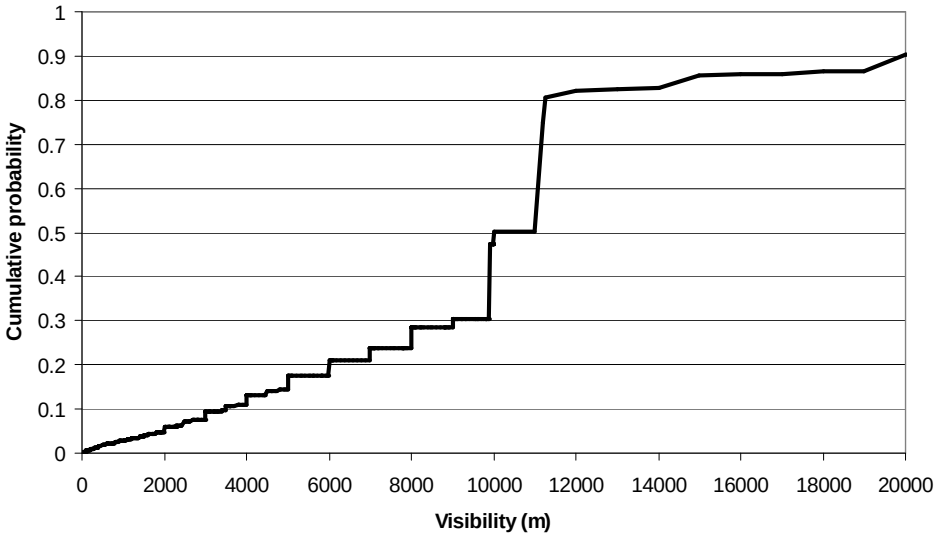


Figure 37: Probability distribution for visibility

Mean cross wind

The mean cross wind is the steady state crosswind measured at landing. The probability estimate is based on world-wide in service operations of UK airlines (sample size about 2000) as presented as reference in EASA's aircraft certification specifications for all weather operations [EASA 2003b]. Crosswinds from the left and right are be equally likely. In Figure 38 the cumulative probability of the crosswind is shown, the unit of measure is knots.

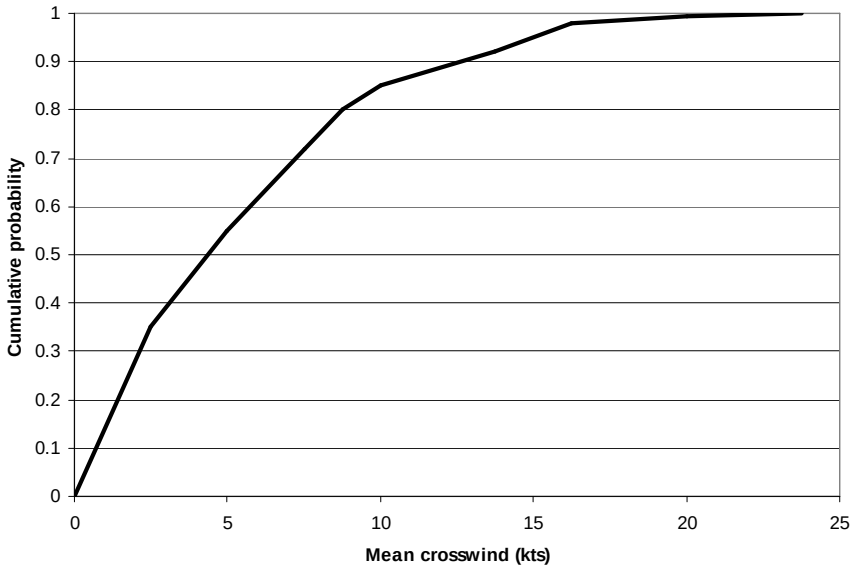


Figure 38: Cumulative probability for mean cross wind

Traffic separation in the air

Traffic separation in the air refers to the longitudinal separation distance between two successive aircraft on the approach to the same runway. The minimum required separation distance depends on the type of approach and the weight category of the aircraft. The probability distribution was determined by analyzing a sample of FANOMOS data. FANOMOS stands for “Flight Track and Aircraft Noise Monitoring System”. The system is controlled by the Dutch Aviation Inspectorate (IVW). The FANOMOS system calculates actual flight paths of individual aircraft by combining radar track data and flight plan data. These data are received automatically from ATC the Netherlands. Figure 39 shows a probability density that is based on a sample of 2382 approaches on Schiphol airport runway 18R between 1 and 8 March 1999. Separation distance is measured in nautical miles. Although the sample size is relatively small, and despite the fact that the sample only represents one approach for one particular airport, the data is considered sufficiently representative because the rules on separation distances are largely governed by international standards. Correctness of this assumption is provided by comparison with a distribution of separation distances that was found at the Dallas Fort Worth International Airport [Ballin & Erzberger 1996].

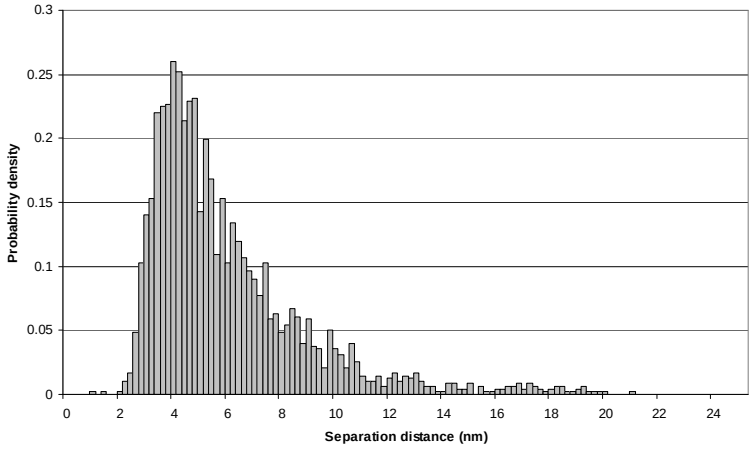


Figure 39: Probability distribution for traffic separation

Speed deviation

During the approach, the aircraft must be on the correct lateral and vertical flight path (based on radio navigation guidance or visual reference), the aircraft must be in the desired landing configuration and at the desired speed and the landing checklist must have been accomplished. For precision approaches, deviations from the lateral and vertical glide path are hardly an issue. Speed is a different matter. Before the approach, the crew will have calculated the approach reference speed for their particular aircraft. This reference speed mainly depends on the weight of the aircraft. A correction will be added to this reference speed to cater for wind conditions or other external factors. The reference speed plus the correction determines the ‘bug-speed’, which is the desired speed during the approach. When at 500 ft altitude the aircraft’s speed deviates too much from the bug speed (e.g, the speed drops below bug speed minus 5 knots or exceeds bug speed plus 10 knots), a missed approach is required. Figure 40 presents the distribution of aircraft speed relative to the bug

speed[86] for the 500 ft altitude gate. This data represents a total sample of 13,753 approaches flown by a large European airline. Speed is indicated airspeed in knots.

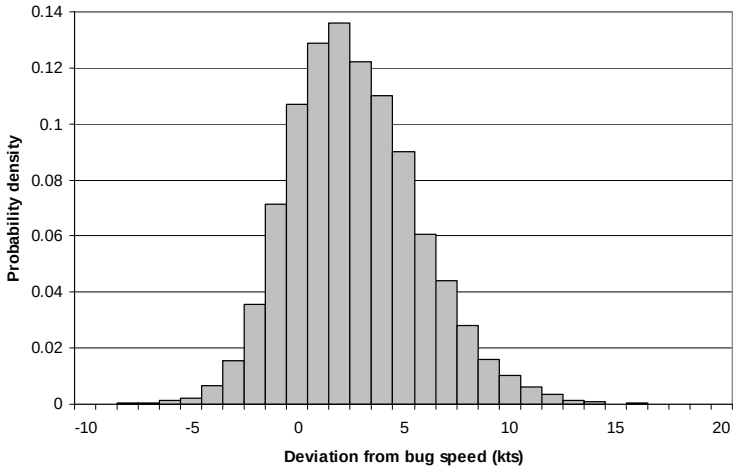


Figure 40: Probability distribution for speed deviation at 500 ft altitude.

Flight crew alertness

Alertness (or its opposite, flight crew fatigue) is an important issue in flight crew performance. Fatigue was reported as a contributory factor in 21% of the incident reports to the NASA Aviation Safety Reporting System [Lyman & Orlady 1980]. Since 1990, fatigue has been a major issue in national and international transportation research programs. The concept of fatigue is very complex; it can be defined in terms of performance decrement, subjective feelings of fatigue, or physiological changes. For practical purposes, excessive fatigue can best be considered as a general description of the multi-causal condition of feeling unfit. The influence of alertness on flight crew performance justifies its representation in the model. To quantify the probability of occurrence of the degree of alertness of the flight crew during the approach phase, results from field and laboratory studies on the different determinants of aircrew fatigue and their effects on alertness and performance of aircrew are used. These studies were conducted by the Aviation Medicine Group of TNO Human Factors as also described in section 8.5.6. The studies, employing subjective and objective measures of in-flight performance and alertness, concerned quality and duration of sleep, effects of early reporting times, effects of night flying, effects of alcohol and medication, and the effects of countermeasures, such as onboard sleep (augmented crew) and pre-planned napping in the cockpit seat. The results of these studies provide an extensive database on factors causing aircrew fatigue and impaired performance and alertness in flight. It has been demonstrated that pre-flight levels of sleepiness and vigilance are good predictors of the level of in-flight crew alertness [Valk & Simons 1998]. Therefore, sleepiness is used as input for the model. Pre-flight and in-flight sleepiness were measured by means of the Stanford Sleepiness Scale (SSS). The result of the SSS is a score with increasing sleepiness from 1 to 7, where 1 signifies “feeling active and vital; wide awake” and 7 stands for “almost in reverie; sleep onset soon; losing struggle to remain

[86] The bug speed is the target speed for the approach. ‘Bug’ refers to a movable pointer on the airspeed indicator that can be positioned by the pilot at the target speed to act as a visual aid for maintaining the correct speed during the approach.

awake", see also section 8.1. Studies have indicated a high correlation between SSS measures and flying performance. Field studies by the Aviation Medicine Group of TNO Human Factors were used to determine probability densities for crew alertness. Sleepiness was measured before the flight and at the top of descent. Sample size was 807 for 'before flight' and 12,965 for 'top of descent'. The results are presented in Figure 41.

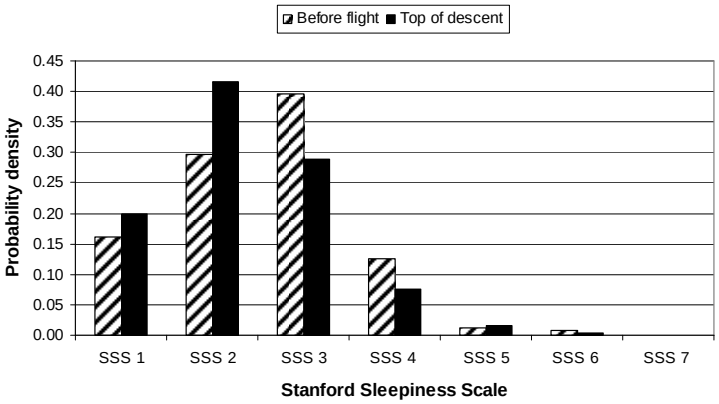


Figure 41: Probability distribution for flight crew alertness

Fuel weight

The amount of fuel on-board the aircraft may affect a crew's decision to fly a missed approach. If the amount of fuel is critical, they may decide to continue the approach in spite of unfavourable conditions. Airline fuel policies are delicate issues. On the one hand, an airline will want to have its aircraft land with the minimum amount of fuel, because any surplus amount means either a lower payload or reduced range. On the other hand, (international) regulation requires that on arriving at the destination airport, the aircraft has sufficient fuel to fly to the alternative airport and then to fly for a further 30 minutes at holding speed at 1500 ft. Actual fuel data is difficult to acquire because, as aviation is such a competitive environment, airlines are reluctant to share fuel information. Figure 42 shows a limited sample of fuel data for one particular aircraft type of a large European airline. The total sample size is 172 flights. The most appropriate unit of measure is remaining fuel in kilograms after landing as this is the way it is represented to the pilots and hence is used in the flight crew's decision making process.

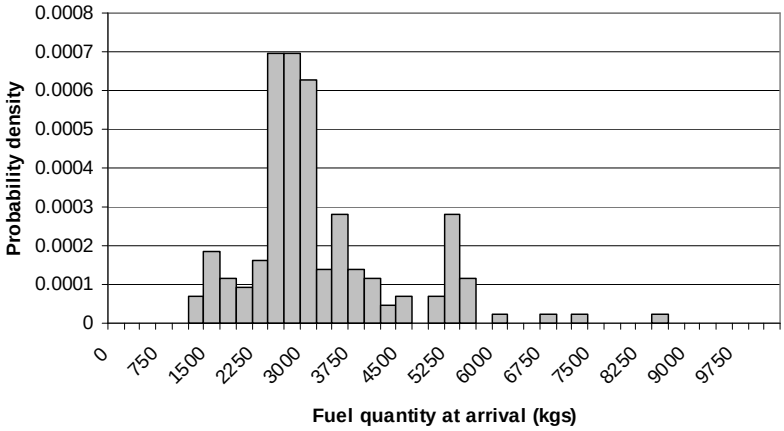


Figure 42: Aircraft fuel weight distribution.

10.3.3. Dependencies

The dependencies between the nodes are described as unconditional and conditional rank correlations that were estimated on the basis of expert judgement utilizing the procedures described in Morales et al [2008]. A single expert, a pilot for a large international airline, answered a total of 7 questions. For the marginal distribution of the missed approach rate the expert was asked: *'Consider 100,000 randomly chosen flights at Schiphol airport , on how many of these flights will a missed approach be executed?'*. Next the dependence information was obtained starting with the rank correlation of missed approach execution and separation as follows: '*Suppose the variable separation was observed and its value was found to lie above its median value. What is your probability, that in this situation, the number of missed approaches will be above its median value?*' The conditional rank correlation of missed approach execution and mean cross wind given separation was then determined by asking '*Suppose the variables separation and mean crosswind were observed and their values were was found to lie above their median values. What is your probability, that in this situation, the number of missed approaches will be above its median value?'* The rest of the rank correlations were elicited in a similar way by sequentially adding information about the variables entering the conditional set.

The results of the elicitation for the 6 arcs in the BBN for missed approach are summarized in Table 10. A negative rank correlation indicates a decreasing missed approach probability if the variable increases. In this case for instance the missed approach probability decreases if the separation distance increases. The value of the rank correlation is an indication of the strength of the dependence, the higher the number the stronger the dependence. A functioning model was obtained using the copula vine approach and the UniNet software application.

Table 10: Conditional rank correlations

$r_{7,6}$	-0.88
$r_{7,5 \mid 6}$	0.20
$r_{7,4 \mid 6,5}$	0.12
$r_{7,3 \mid 6,5,4}$	0.23
$r_{7,2 \mid 6,5,4,3}$	-0.11
$r_{7,1 \mid 6,5,4,3,2}$	0.11

7 = missed approach execution
6 = separation in air
5 = mean cross wind
4 = speed deviation at 500 ft
3 = crew alertness
2 = visibility
1 = fuel weight

10.3.4. Comparison of model results with observations in practice

Flight track data was analysed to validate this missed approach model. The objective was to obtain a realistic estimate of the probability of a missed approach in correlation with the separation distance and to compare this with model estimates. The flight track data was derived from FANOMOS [Kreijkamp & Veerbeek 1997]. The flight tracks in FANOMOS are collected through measurements with the secondary surveillance radar (SSR) located at Amsterdam Airport Schiphol. This raw data is updated every 4 seconds and is used by FANOMOS to reconstruct the actual flight track of the aircraft. FANOMOS computes the position of the aircraft (in the "Rijksdriehoek" co-ordinate system (RDC)), the ground speed, and the total recorded length of the track. All reconstructed flight tracks are stored in a database. For this study, flight tracks of aircraft landing on any runway in the period from November 2004 until October 2005 were selected from the database, resulting in a total of 207,478 approaches. For each approach the flight track was inspected to verify if a missed approach had occurred. In addition, for each approach the longitudinal separation with the previous approach on the same runway was determined. Since a missed approach is by definition executed in the final approach segment, it was sufficient for this study to determine the separation in the segment from final approach fix (FAF) to runway threshold. The total number of missed approaches found in the dataset is 277. This means that approximately 1.3 of 1000 approaches is aborted. An example of a missed approach flight track is shown in Figure 43.

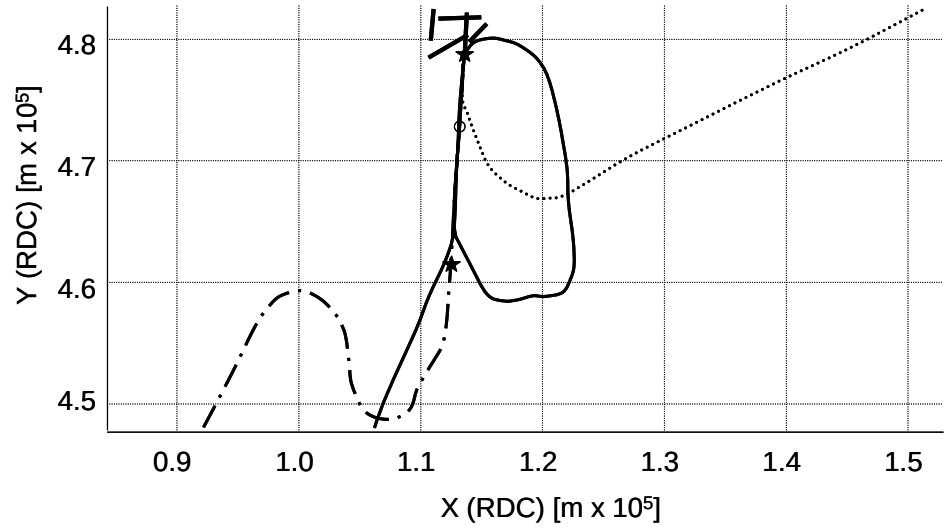


Figure 43: Flight Track Analysis

This example concerns an approach from the south to runway 18R that includes a missed approach. The figure shows the plan view of the situation with the airport at the top center. The flight track of the approach under consideration is depicted in as a solid line. The tracks of two other aircraft are also included. The dotted line corresponds to the leading aircraft before the missed approach is executed, the dash-dot-dash track corresponds to the leading aircraft after the missed approach. In this example the order of occurrences is as follows: First the aircraft represented by the dotted line conducts the approach, then the aircraft represented by the solid line an approach but this is aborted and a missed approach manoeuvre is conducted and a right hand circuit pattern is flown. While the aircraft turns to get lined-up again, the aircraft represented by the dash-dot-dash line conducts the approach and lands. Finally, the solid line aircraft conducts its second approach and lands. The position of the aircraft at the moment the missed approach is initiated is indicated by the 'O' marker. The FAF and runway threshold are denoted by stars.

Analysis of all flight tracks resulted in the probability of a missed approach for a given separation distance. The missed approach probability as a function of separation distance that follows from the FANOMOS data is compared in Figure 44 with results from the missed approach model. The model calculations were made in UniNet, a continuous and discrete non parametric Bayesian belief net application, functioning as a module of UniCorn, a stand-alone uncertainty analysis software package. UnitNet was developed by Delft University of Technology to support the CATS project [Cooke et al 2007]. The order of magnitude of the probabilities and the shape of the curve of the model results corresponds reasonably well with observed results, albeit that the model underestimates the missed approach probability for very small separation distances. Whether the model results are an acceptably close approximation of the real world will depend on the use of the model.

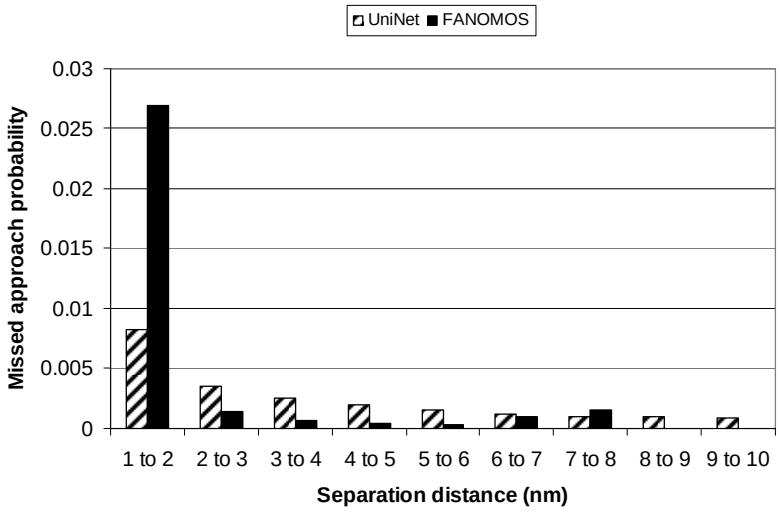


Figure 44: Comparison of FANOMOS data with model results (UniNet).

10.4. Face validity and peer review

Face validity is important to check to quality of the model structure. Face validity is obtained by presenting the model and model results to experts and obtaining their feedback. Obtaining face validity is essential for the model to be able to be accepted for use. If an expert considers the model not valid, it is difficult to challenge that opinion, because there is no alternative way for validation. The only possibility (apart from rejecting the model of course) then would be to find an expert with more authority to contradict the first expert, which is undesirable because of the risk of getting involved in a fruitless discussion on who is the most authoritative expert.

During the development of the CATS model, face validity was obtained during regular meetings with an 'expert group'. Importantly, this expert group consisted of operational experts (e.g. pilots) as well as experts in risk modelling and human behaviour.

The effect of presenting the model to operational experts is that they will always and almost immediately spot where the model is incorrect, which will lead to the question whether the model should be rejected or adapted. This is aggravated when the experts are considered specialists in a particular field. Without proper introduction of the purpose of the model and its intended use, the experts will invariably call the model 'too simplistic'. Experts should be asked to review the model *in relation to its intended use*. For CATS, the intended use of the model was initially not specified, which made it difficult for the experts to provide comments and for the developers to determine what to do with those comments.

For the FAA causal risk model [Roelen et al 2008], face validity was obtained by having the 33 accident scenarios checked by former employees of the National Transportation Safety Board (NTSB). Principal Inspectors and Operational Research Analysts from the FAA who are the proposed users of the model, were invited during progress meetings and this provided further face validity of the model. The advantage of involving future users is that they are in a much better position to appreciate the assumptions and simplifications in relation to the intended use of the model.

In summary, face validity is essential. It should be obtained from experts with various backgrounds, including potential future users of the model. The intended use of the model should be defined and explained to the experts at the start of the process.

10.5. Assumption analysis

During the development and quantification of a model, several assumptions are always adopted. An assumption analysis identifies and analyses all assumptions adopted and assesses the individual and combined effect of these assumptions on the model-based risk result. Assumptions are made due to the method selected, due to scoping, or they are adopted during the modelling, e.g. due to unavailability of data or lack of time [Everdij & Blom 2006]. The validity of each and every assumption of the causal risk models needs to be assessed separately. This requires strict discipline of the model developers in meticulously keeping records of all the assumptions, including parameter value assumptions, numerical approximation assumptions, model structure assumptions. During development of the CATS model for instance, parameter value and numerical approximation assumptions were painstakingly stored in a database called CATSPAWS (CATS PArameters With Sources). It is imperative to do this during development because it will be almost impossible to infer assumptions from a completed model. Without an assumption analysis it is impossible to define the boundaries within which the model is considered valid for use.

10.6. Conclusions for this section

This section discussed four aspects of validation:

- Validity of the process of decomposition (e.g. breakdown into different accident scenarios) and aggregating the results
- Case validity: comparison of behaviour of (a section of) the model with real world behaviour
- Face validity and peer review
- Assumption analysis

When the risk model is constructed the developers should keep record of all assumptions, including parameter value assumptions, numerical approximation assumptions and model structure assumptions. Because of the expected size and complexity of a causal risk model for aviation the four types of validation cannot be obtained for the model as a whole. Instead, the validity of parts of the model should be checked. This chapter showed some examples of how this could be done. The example on missed approaches showed how, for specific cases and a specific part of the model, comparisons of the model results with real world data are possible. This type of validation provides valuable insight but is also labour intensive. The choice of cases should be done carefully, such as to represent the full spectrum of intended model use. Case validity can only be obtained for situations where data on real world behaviour is available.

Nevertheless, full validation of the complete model cannot be obtained. Therefore the results of a causal risk model should always be used carefully. For every use it should be established how the model assumptions could impact the result. The model results should never be used as the sole source of information for decision making. Whenever possible, alternative sources of information should be considered as well.

The added value of using the alternative definition of validation as described in the introduction (i.e. validation is a process of checking the degree to which the model provides, for the objective for which it is applied, a correct description of the reality that it represents [AIAA 1998]) could be a topic for future research.

Chapter 11. Summary, discussion and conclusions

Affordable and reliable aviation plays a vital role in supporting economic growth and expanding personal options for where individuals can live, work, travel, and conduct business. Adequate control of aviation risk is a necessity, and this requires the availability of a method to determine the level of accident risk as a function of changes to the aviation system. This method must be methodologically sound. Causal risk models have been proposed for this purpose. Aviation is possibly well-suited for the application of causal risk models, but research has primarily been focussed on the technical feasibility, without close consideration of the methodological consistency in relation to user requirements. This thesis has investigated the usefulness of causal risk models for air transport by analysing user needs, appropriateness of modelling techniques, model structure, quantification methods and validation.

The main research question of this thesis was "*What does causal risk modelling add to current safety management approaches, and what are the criteria for ensuring it makes a successful contribution?*". Several sub-questions were defined to help answer the main research question. The preceding sections have provided the information to answer the sub-questions and subsequently the main question.

Sub-question 1: What is a proper way to express risk?
A decisive characteristic of 'safety', with important consequences for (the application of) a causal risk model is its inherent intangibility. Safety is defined as freedom from unacceptable risk, where risk is a combination of the probability of occurrence of harm and the severity of the harm across a defined range of scenarios. The acceptability of risk is influenced by risk perception, which is shaped by many factors, including the level of control, the chance of multiple fatalities, the time passed since a similar event took place, whether the exposure to risk is voluntary, etc. Policies to control major risks have been in development from the 1960s onwards. Many of these policies are based on some sort of quantification of the risk that could be allowed to continue. Most safety critical industries apply probabilistic risk assessment to quantify the risk in their business. Correctly applied, probabilistic risk assessment can be the basis for risk informed decision making. Probabilistic risk assessment has only sporadically been applied in the aviation industry, but it has been suggested that a causal model for aviation safety should be developed and applied to assess risk and to evaluate the effect of operational and managerial changes (decisions, measures, etc.) on the level of safety. The need for quantitative safety assessment methods is explicitly voiced by safety regulators like the FAA, EASA, Eurocontrol and the Dutch Ministry of Transport. The purpose of such a model is to reduce aviation accident risk by improving safety management, for instance by identifying those areas where intervention in the future would be most efficacious

A proper way to express aircraft crash risk for this type of objective is accident probability per flight, using ICAO's definition of an accident. This is an objective metric, not influenced by the perception of risk. If deemed necessary it can be used as input for third party risk calculations and it can be objectively compared with a target level of safety. The ICAO terminology defines incidents as precursors to accidents. They can be seen as

accident chains in which the ultimate link(s) towards an accident did not occur, a formulation that is compatible with the accident causation theories of Heinrich, Reason and Rasmussen. The way in which an incident is defined ensures that incident data can be used to estimate accident probability; this is a necessity for aviation where the accidents are so rare that even large observed differences need not attain significance. Using incident data will decrease the uncertainty boundaries of the model results.

Sub-question 2: What is a causal relation and which characteristics of causal relations are important for causal risk model development?

A causal relation is a formal description of the link between cause and effect. Statistical associations alone are not sufficient to construct a causal relation. There also has to be an underlying assumption, a causal assumption, which should be based on deeper scientific knowledge on cause and effect through theoretical studies and accumulated knowledge. Inferring causal relations requires *subject-specific background knowledge.* It is therefore essential for developers of a causal model of aviation safety to have substantial knowledge on relevant aspects of aviation, including technology, operations, regulation and procedures for the complete lifecycle. Even then, what is actually the cause of an event can be ambiguous. To a certain extent, it is an assumption that has to be agreed upon by those who work with it. The agreement is necessary because the assumption has consequences for the model output. It is therefore required that the underlying assumptions are transparent. Part of the challenges of causal relations is also their apparent probabilistic nature. The causal relations in a risk model are generic. We often arrive at generic causal relations by generalising from individual cases of occurrence and then apply this general knowledge to other individual occurrences. But evidence of a generic causal relationship is not sufficient to prove a singular causal relationship, and the absence of a singular cause effect relation is not sufficient to negate a generic causal relation. Whereas a causal model can be used for both forward and backward reasoning, a causal risk model for air transport is most useful for forward reasoning; predicting the most likely effect, given a cause or a change in the set of causes. The call for more systemic views on accident causation and the underlining of the role of ‘normal’ occurrences in accident causation emphasise the need for risk models that go beyond traditional ‘hard wired’ approaches like fault trees and events trees. The causal risk model should be able to represent the seemingly stochastic and evolving nature of the system in relation to its hazards.

Sub question 3: What is a causal risk model?

A causal model is a mathematical object that provides an interpretation and computation of causal queries about the domain. A causal model can be associated with a directed graph in which each node corresponds to a variable in the model and the arrows point from the variables that have direct influence to each of the other variables that they influence. The causal risk models addressed in this thesis are models for forward reasoning and are used to predict the effect of changes in the aviation system on aircraft accident risk.

Sub question 4: What are the needs of users?

The main drivers for aviation safety improvement have been technological development and the industry’s practice of thorough accident investigation and subsequent remedial actions. Because of the high level of safety, the latter approach is no longer effective and there is a need for a new safety performance evaluation methodology, of which a causal risk model could be a central part. The increased integration, automation and complexity of the aviation system, including integration of ground-based and airborne systems, will lead to new regulatory requirements for safety. The complexity of the matter will restrain authorities from specifying detailed requirements. Instead, quantitative target levels of

safety will be specified. Quantitative safety assessment methods such as causal risk models will have to be applied to demonstrate to the authorities that the requirements have been met. There is a growing awareness of the need to apply a total systems approach in safety assessments, considering the complete aviation system rather than the individual components. Because the human operators (pilot, air traffic controller) play a decisive role in the aviation system, these quantitative safety assessment methods must be able to satisfactorily represent the human operators.

The potential users of a causal risk model are airlines, air navigation service providers, airport organisations, maintenance and repair organisations, aircraft manufactures, the central government and the aviation inspectorate.

This need for quantitative safety assessment methods is explicitly voiced by safety regulators. FAA, EASA, Eurocontrol and the Dutch Ministry of Transport have all expressed the need for a risk assessment method. There are differences though. The FAA calls for the development of a method for a probabilistic evaluation of risk associated with hazards in the area of operational functions and procedures, if possible extended with hazards and risk in other areas of the aviation system such as aircraft and equipment malfunctions, hazards due to organisational set-up, human errors, environmental hazards (wind, turbulence, terrain) and contributory hazards (regulatory, economy). A specific requirement is to have a top-level representation as the user interface for performing risk analysis. The European regulator EASA is specifically looking for a methodology to quantify the safety effects of proposed regulatory changes in terms of accident probability and severity. The Dutch Ministry of Transport and Water Management also sees a role for causal risk models in the regulation of third party risk. Furthermore, the Ministry believes that a causal risk model can be used for communication with citizens, to provide insight in how various components interact and influence aviation safety. Eurocontrol has mandated the use of quantitative safety assessments of significant changes in the ATM system to demonstrate that the probability of ATM directly contributing to accidents is less than 1.55×10^{-8} per flight hour. Eurocontrol also emphasises the development of safety monitoring and data collection mechanisms. The rationale is that any model used in risk assessment must be validated, and real data could contribute significantly to this validation process. The different views by the regulators on the use of a causal risk model are not fully compatible. Particularly the requirement to use a causal risk model for communication with citizens is incompatible with the other requirements.

The industry (aircraft manufacturers, airlines, ANSPs, airports) has not explicitly expressed a need for a causal risk model. That does not mean that such a model could not be useful for them. A feature that would make such models attractive for them is the ability to perform cost benefit analysis. Some of potential industry applications may require very detailed models however, for instance that describe the aircraft down to component level.

Detailed user requirements were defined (see chapter 4) which can be grouped under the following high level headings: The model must be *integrated*, meaning that it represents the complexity of the complete aviation system and its many interdependencies and explicitly includes human operators and represents managerial influences. It should produce *quantitative* results to demonstrate to authorities that quantitative requirements have been met. The model structure representation and the model development process should be *transparent and clear* to provide insight and for communication purposes. The model should be *validated* and turn out *reproducible* results, this is particularly required if the

results are being used in a certification process. Finally, the model should be able to represent *current and future* accident scenarios.

Requirements that cannot be met at present are the requirement to capture very specific information and the requirement to represent the influence of regulations.

User requirements demand a quantitative and transparent causal risk model that should be able to represent environmental and managerial influences. The results of the model should be reliable, and the model should be useable to indicate possible solution areas for further safety improvement. The variety of users and diversity in the users' questions make it impossible to develop a causal risk model that is capable of answering all needs simultaneously. The requirements call for a system-wide representation of air transport. Such a model is not suitable to answer very specific questions. This limitation should be explained to the aviation community to avoid the creation of false expectations.

Due to the diversity of possible users and the limited view they currently have on the use of causal risks models, significant work remains to be done on the systematic identification of user needs. Feedback from the modellers to the potential users is required to get a better grip on user requirements. It should also be verified, for each kind of user, that the causal risk model definition as adopted in this thesis, is suitable for them. Because users are not good at imagining what a causal model could do they are not good at articulating uses and needs for it. Therefore, the modellers should indicate possibilities and limitations of the model, and should come-up with real-life examples of cases in which a causal risk model would be a helpful tool. This will have to be an iterative process. Support from the industry is essential for the provision of data and to provide insight, to the model developers, into operational aviation processes.

Sub-question 5: What are currently drivers for aviation safety improvement and what could be the role of a causal risk model in the process of aviation safety improvement?
Aviation has realised tremendous safety improvements since the first commercial airlines started operations. The drivers for safety improvements have been primarily catastrophic accidents, regulation, technological advances and liability. Systematic analysis of accidents and subsequent remedial action ('fix and fly') is the natural evolutionary way of improvement and leads to spectacular advances when technology is still immature. However, when the technology matures the rate of improvement reduces until a level plateau has been reached; this phenomenon is well known in reliability engineering and is often represented as the first part of the bathtub curve, also referred to as infant mortality. The maturity of aviation causes this driver to become more and more ineffective. Technological advances have been focussing on improving reliability of aircraft as a necessity to be more profitable and this has naturally improved safety as well. However, currently the biggest challenge is not aircraft reliability but efficiency, and technological development is aimed at improving that. But what is good for efficiency is not always good for safety. Regulation traditionally has a role of safeguarding minimum safety levels but it follows safety developments rather than initiating them and regulation is aimed at specific disciplines rather than the integrated system. Liability issues are perhaps currently the strongest drivers for aviation safety improvement, especially since some highly visible cases have involved personal accusations of manslaughter (e.g. the Überlingen midair collision and the Milan Malpensa runway collision). But this can also be very counterproductive, when known safety problems are hidden and incident reporting systems stop working because of fear of liability claims.

The air transport industry is often driven by the need of capacity increase on one hand whereas it is not clear how to do so without jeopardizing safety criteria. To maintain and improve the current level of safety, there is need for an aviation safety performance evaluation methodology that is not based on fatal accidents and hull losses alone. While accidents are rare events, the consequences of an accident can be enormous for the airline, the aircraft manufacturer, the air navigation service provider and society. Society is becoming less tolerant towards accidents. Accident prevention is therefore as important as ever, but due to the rarity of accidents it becomes more and more difficult to understand the 'current state of affairs'; there are simply too few 'data points'. Feedback of lessons learned is frustrated by a tendency towards criminal prosecution of those involved in accidents and incidents. As a complicating factor, the aviation system is characterised by a multifarious arrangement of organisations and activities. It is a distributed system with many different players. Each discipline in the aviation industry has its own process for managing safety and regulation is aimed at specific disciplines rather than targeting the integrated system. The lack of an integrated approach is worrisome as the breakdown of the communication paths between the members of the aviation industry is often causal or contributory to accidents and incidents. The aviation industry cannot afford to resign itself to the current situation as environmental and economical pressures will then slowly but surely force the level of safety down. Aviation is a competitive environment, and the investments are often huge. Any proposed safety improvement measure must be well-reasoned. Therefore an integrated safety assessment methodology is needed. A causal risk model could potentially provide an accurate estimate of the actual accident probability and simultaneously indicate the relative contribution to risk of each of the different actors, processes and risk control measures within the aviation system. This will then allow the choice of adjustments to the system which are most effective in improving safety. It will also provide solid arguments to safety managers for standing up against decisions to change the system which are motivated by environmental and economic grounds but are counter-productive for safety.

Sub question 6: What are shortcomings of the current methods for aviation safety analysis?
Although causal risk models for aviation that represent the complexity of the complete aviation system and its interdependencies including human operators and organisations have not yet been developed and applied, safety analyses are frequently conducted in the aviation industry. Generally speaking there is no standardized or prescribed procedure for conducting safety analyses for air transport. Selection of the safety analysis methodology for a specific problem depends on the scope of the safety question, the type of operation, the level of involvement of the customer, the domain knowledge of the customer and the usage of the results. While the currently used methods often follow similar steps, they differ in the details. One of the consequences is that quite frequently safety analyses start with an activity that has already been performed earlier, albeit slightly differently. Activities such as hazard identification and accident scenario development are repeatedly done, without taking full advantage of the results and lessons learned from previous studies. Data collection is also an activity that sometimes requires a significant effort. Often data is not available or there are insufficient resources for a dedicated data collection campaign. Many safety analyses therefore rely heavily on experts, or more precisely, on the black boxes inside the expert's heads, for hazard identification, accident scenario development and quantification. Each expert will have his own 'model of the world' which is used to come to certain conclusions. These models have not been made explicit, and criteria for experts have not existed. In aviation safety analyses, selected experts are often people with (ample) operational experience, such as pilots and air traffic controllers. Analytical skills and ability to provide judgement on future operations are only rarely taken into consideration. Experts

are not ‘calibrated’. These aspects have a negative effect on the transparency and reproducibility of model results.

Air traffic management and airports are, in comparison with aircraft design, operation and maintenance, far less regulated in the form of detailed requirements for specific risk control means. Air navigation service providers and airport authorities are relatively free in developing infrastructure, procedures, etc., that fulfil their specific needs. However, they still need to demonstrate that all applicable target levels of safety are met. As a result, there is a growing need for support in conducting risk assessments in the field of ATM and airports. In aircraft design, operation and maintenance, there is less freedom due to more detailed regulations.

Current methods for safety analysis are not unified, resulting in unnecessary duplication of work and possibly diverse outcomes, undermining the credibility of the analyses. None of the currently available methods for safety analysis and assessment fulfils the main requirements: ability to analyse the whole integrated system including dependencies, reproducibility of results, transparency, ability to provide quantitative results and ability to represent current as well as future accident scenarios. An effort to develop a method that will meet these criteria therefore seems fully justified.

Sub question 7: What can be learned from other industries?
The way in which probabilistic risk models were introduced in the nuclear power industry, manned spaceflight, and the oil and gas industry followed a similar pattern. Traditionally, risk assessment methods were deterministic and qualitative. Probabilistic risk analysis methods were occasionally applied but were not seen by everyone as an improvement over the traditional methods. Lack of data to populate the models, lack of transparency and failure to adequately represent human and organisational performance were mentioned as weak points. However, almost within a single decade several major accidents[87] resulted in regulatory changes that forced the introduction of probabilistic risk models for system-wide analysis. Even then it was a slow process, and it took a long time to develop methods and models that are considered acceptable. It is recognised that these methods should not be used as a substitute for the traditional approaches, but as a complement. The combination of qualitative, deterministic and quantitative probabilistic methods allows risk-informed regulation and decision-making. Yet there is still sometimes resistance, when people believe that the money spent in quantitative safety analysis could be better invested in improving the system directly. Because of the many assumptions that have to be made in the models, the results of the analyses are at times considered to be too subjective to be used for decision support. It is only when people realise that there are no alternatives, that current decision making is even more subjective, and that complex integrated systems require systematic and consistent analysis, that people appreciate the value of causal risk models. Prescription of a unified method for risk assessment, as is the case in the Netherlands for sites involving flammable or toxic substances, leads to less variability in the outcomes and better acceptance of model outcomes. Based on these experiences it can be expected that introduction of causal risk models in air transport will be similarly jerkily and it could take decades before the industry has fully accepted the use of quantitative risk models. Data to populate the model, transparency of the models and the representation of

[87] The Seveso calamity (1976) in the process industry, the Three Mile Island accident (1979) in the nuclear power industry, the Challenger explosion (1986) in manned spaceflight and the Piper Alpha disaster (1988) in the oil and gas industry.

human and organisational influences are important to promote and support the development. It is also essential to state from the outset that the models should be used as a complement rather than as a replacement of current practice. The use of diagnostic models in the health care industry demonstrates that such models are indeed suitable instruments for communication on probabilistic relationships and to provide a coherent representation of domain knowledge under uncertainty.

Sub-question 8: Which modelling techniques are most appropriate?
A causal risk model can be regarded as being composed of a representation scheme and a computational engine. The representation scheme is the interface between the computational engine and the user of the model and is the primary means of communicating the model to the outside world. Many people will regard the representation scheme as 'the model'. The representation scheme must therefore be comprehensive, and most importantly, transparent. A scheme is transparent when a simple set of rules and symbols applies and when it is intuitive. Simplicity of the representation scheme can also be achieved by a (physical, functional, or other) decomposition. Whenever a decomposition is made, a mechanism must be in place to keep track of possible dependencies between components.

Modelling methods with a very simple set of rules and symbols are tree-like structures such as fault trees and event trees. They are therefore ideally suited to provide the top layer of the representation scheme, the level that is used most directly by the user. However, beyond a few levels of the trees, the events are strongly influenced by many to many influences of common mode factors, particularly related to management. Since the assumption in fault tree logic is that all events are independent events, the model must be linked at this level with another type of model. Bayesian Belief Nets are a suitable candidate as they can capture dependencies and soft interactions among component behaviour. The graphical representation of a Bayesian Belief Net is still intuitive, although the rules are more complex than those of a fault tree. A BBN is therefore less transparent. When dynamic interactions between components are important it becomes necessary to shift to another type of modelling technique such as Petri Nets in combination with Monte Carlo simulation.

The degree of detail and sophistication of the model representation must be decided upon based on the purpose of the model and the performance that is required by the computational engine. If the computational engine must be able to handle interdependencies and dynamic behaviour, a simplified representation of the model via a user interface is required to obtain transparency.

Even though a causal risk model is a mathematical object, developing a causal risk model is not primarily a mathematical problem. Examples show that proper mathematical techniques are available and the required computer power is not excessive. Developing a causal risk model of the air transport system is much more a problem of assembling all available knowledge and data on a very large and complex system and unravelling this complexity into cause-effect relations that can be assembled in a model and representing this information in a useful way.

The development of a causal risk model for aviation should be guided by questions and functional requirements from the users of the model. The model developers should use techniques, develop (mathematical) solutions, and go to the level of detail that fits the user

requirements. The degree to which dynamic modelling is necessary will require careful consideration and will depend on the type of user and type of application of the model.

Sub question 9: How should a causal model be quantified, what numerical accuracy is required and how can that be obtained?

Quantification of the probability of occurrence of variables and the conditional probabilities in the causal risk model is done by using existing data or by expert opinion. Often different sources of data will need to be combined. It is therefore extremely important to ascertain that the same variable descriptions and units of measurement are used. The use of standard units is preferred, but sometimes it is impossible or impractical to apply a standard quantitative unit of measurement and a qualitative rating scale must be used, or a proxy variable that is closely related to the variable of interest and which can be quantified must be found. One of the main difficulties in developing a causal risk model for aviation lies in the need to express 'soft' factors in objectively quantifiable units. Proper units for e.g. the quality of procedures, commitment of personnel, safety culture or non-technical skills of pilots do not yet exist. Proxy variables are also difficult to find as many of these soft factors are multidimensional. Less well defined quantities and units of measurement must be used at the risk of adding variability to the model and increased uncertainty of the model results.

The credibility of the whole causal risk model depends on the quality of the model structure and on the sources of the numerical data to be used and the confidence that one can have in the figures. The accident investigation report is without doubt the single most complete and accurate source of information regarding an accident. Due to the low frequency of major accidents the investigation reports are not sufficient to populate all parts of the causal risk model. Other sources of information are required to complete quantification of the model. Incident reporting systems capture vast amounts of data with relatively little effort, but during the coding process information may get lost or be incorrectly classified. Voluntary incident reporting systems are not very well suited to conduct a statistical analysis or to provide quantitative data to the causal risk model. In-flight recorded data is the most accurate and detailed data, suited to quantify large pieces of the causal risk model. Unfortunately for model developers these data are often company confidential because of the competition and liability issues. Nevertheless, experience has shown that airlines are sometimes quite willing to share data, provided that certain conditions are met. But incident reporting systems and in-flight recorded data are usually rather weak in capturing data on human performance and managerial influences. Empirical data can sometimes be used to fill this gap. Applying the data may require some flexibility and creativity from the causal model developers because the empirical studies may provide data on parameters that are defined slightly differently than those in the model. Unfortunately there are hardly any usable empirical studies on organisational influences. The only alternative is expert judgement. The experts should preferably combine analytical skills and the ability to provide judgement on unknown (e.g. future) operations. If systematically elicited from experts who are both knowledgeable and well-calibrated for the relevant area of expertise, expert judgement can offer quite acceptable accuracy of quantification. The drawbacks of the use of expert judgement are primarily practical issues: finding proper experts is not always easy and expert elicitation can be a time consuming activity.

Quantification of the occurrence rate of events also requires denominator data. Due to the large number of variables and combinations of variables, collection and analysis of denominator data is more than simply counting the number of flights but requires combining the information from several different data sources. In-flight recorded data is potentially a significant source for denominator data, but currently the focus there is

primarily on parameter threshold crossings. Non-event data is usually discarded, although there are some initiatives to conduct flight data analysis of full datasets including normal as well as parameter threshold crossing data.

For all quantitative data in the model the levels of uncertainty of the models are as equally important as point estimates of parameter values. Confidence intervals for expert judgement results require special attention. They should be regarded as 'levels of belief or confidence' instead of 95% confidence levels in the strict mathematical sense. To get a proper picture of the uncertainty of the results of a causal risk model it is necessary to conduct a thorough identification and analysis of all assumptions adopted, including assumptions on parameter values used, and assess the individual and combined effect of these assumptions on the model results.

Sub question 10: What are, from a modelling point of view, the biggest bottlenecks?
Because there is greater variability and interdependence in human performance than in hardware performance, specific human performance models must be developed. Representation of human reliability in risk models has always been one of the major hurdles. The classical approach in human error quantification is to use basic error probabilities for specific tasks and to modify those to account for specific circumstances or contexts. This approach is not good at supporting the identification of errors of commission and assumes that the performance shaping factors are independent, which is not a very reasonable assumption. Some of these limitations can be overcome by representing the influence of the context in an influence diagram. It is also proposed to use the associated event in e.g. the fault tree (or any other type of model representation that is used), into which the human performance model ties, to estimate the basic human error probability instead of deriving this from a task analysis as is the case in first generation human performance methods. Mapping of accident and incident data on those fault trees will ensure that both errors of omission as well as errors of commission will be captured in the model. Whether this approach is indeed suitable for analysis of errors of commission is still open however and needs solving. If the influence of management is described by means of 'delivery systems', these provide a convenient framework for selecting appropriate performance shaping factors and linking management influences and human error probability.

A second bottleneck is representation of managerial influences. There is a requirement for the causal risk model to be able to represent the effect of managerial influences on safety, but this is not a straightforward task because the influence of management is always via the failure events at the operational level. Therefore a link must be found between the managerial system and the description of the failure events at the operational level. A possible solution is that of combining the generic delivery systems, to describe the provision by management of the resources and criteria for the frontline workforce to operate safely, with the concept of a safety envelope. Risk control measures ensure that the organisation stays within that safe envelope. Safety management is then considered a control function aimed at maintaining a particular operational process within the boundaries of safe operation. This requires a meeting point of two model types that are inherently different; failure events at the operational level are traditionally described as specific accident scenarios, while management is seen as a control function, which logically results in attempts to apply control theory as a modelling approach for safety management. Linking the generic control models, with their dynamic feedback loops, to the static accident scenarios has proven to be difficult. Therefore it is, at least for the moment, better to use 'simple' static representation techniques for modelling managerial influences, even if

this may seem to be an oversimplification. In fact, the same simplification is used by representing the actions of the front line human operators (pilots, air traffic controller, maintenance technicians) by linear static models rather than by dynamic control models incorporating feedback loops.

The third bottleneck is the dependence between model elements. There are potentially many to many interactions between individual model elements and if we were to represent them all we would end up with an excessive development and assessment burden and a model that looks like a giant hairball which is not very transparent. It is necessary to find a proper balance between model completeness and practical feasibility. The proposed way of doing this is by developing the model from the generic to the specific, starting at the highest level of abstraction and introducing additional detail step-by-step. This can be done systematically if the model is hierarchical, with separate and formally defined levels of abstraction and if accidents are not completely random combinations of random events. A review of aircraft accidents shows that indeed accident event sequences are often very similar, even in cases where human error plays an important part in the accident sequence. As a matter of fact, almost all accidents of the last 30 years or so can be mapped onto a set of 33 *archetype accidents*. These archetypes provide a suitable backbone for a causal risk model. The advantage of the use of accident archetypes is that it is a data driven approach which results in realistic scenarios. It easily captures the most frequent accidents which means that it quickly focuses on those elements which are most relevant for flight safety. The accident archetypes create a top level model that is intuitively transparent. By using generic accident scenarios as the backbone of a causal risk model the model structure is explicitly based on observations in practice, rather than a theoretical approach such as structural decomposition or functional abstraction. An advantage of this approach is the ability to simultaneously capture technical malfunctions, human behaviour and managerial influences without being confined by a theoretical modelling construct. On the down side, this practical approach is predominantly based on observations from the past and 'new' accidents are not necessarily captured.

Sub question 11: How should the validity of the model be demonstrated?
Full validation of a complete causal risk model of air transport is not possible because model results cannot be compared with a continuous observation of an unaltered system; the time period involved in observing reality until sufficient events have accumulated makes real life observations impractical and creating these events on purpose is unethical. Therefore alternative methods for demonstrating the 'validity' of the model must be pursued. Individual segments of the model with directly observable outputs will have to be validated, and proof must be provided that the overall process for combining these segments into a single model, including the possible interactions between various segments, is sufficiently correct for the objective(s) under consideration. When the risk model is constructed the developers should keep record of all assumptions, including parameter value assumptions, numerical approximation assumptions and model structure assumptions. It is imperative to do this during development because it will be almost impossible to infer assumptions from a completed model. Assumption analysis should be part of the validation process, and face validity should also be obtained. Nevertheless, full validation of the complete model cannot be obtained. Therefore the results of a causal risk model should always be used carefully. For every use it should be established how the model assumptions could impact the result. The model results should never be used as the sole source of information for decision making. Whenever possible, alternative sources of information should be considered as well. The added value of using a different approach to validation,

which emphasizes on evaluating what the difference between model and reality means in terms of the model output, could be a topic for future research.

Main research question: What does causal risk modelling add to current safety management approaches, and what are the criteria for ensuring it makes a successful contribution?
Current drivers for safety improvement fall short to bring about further safety improvements. Because accidents are very rare, the industry's practice of thorough accident investigation and subsequent remedial action is no longer sufficient. A causal risk model can be used to assemble available knowledge and data of the air transport system to identify weaknesses in the system before they result in accidents or incidents. This systemic and pro-active approach would be an addition to current safety management approaches which are ad-hoc and reactive.

A method for safety assessment and safety analysis is required that is not solely based on accident frequencies, provides a quantified accident probability and is able to properly represent the complex arrangements within the air transport system, including human and organisational influences. The method should be transparent and yield reproducible and demonstrable valid results. Using a causal risk model as described in this thesis is able to meet those requirements and therefore is a method to identify possibilities for safety improvements that cannot be obtained with a reactive response to accidents.

Each and every day decisions are taken that potentially have an influence on air transport safety. Often these decisions involve a consideration of safety against economic or environmental aspects. Safety analyses are conducted to help in making those decisions but the current methods used in those analyses fail to meet important requirements of transparency, reproducibility, quantification and the ability to consider the integrated air transport system rather than a part of it. A causal risk model can meet those requirements. Developing such a causal risk model is not a problem of mathematics; the tools for calculation and representation of the model are mathematically relatively simple and well-developed. The classical problems of risk modelling, i.e. representation of human operators and managerial influences present significant challenges but they are not insurmountable. The real problem is not so much describing those influences but more their quantification. A third dilemma that is often mentioned, the lack of data, is not as problematic in aviation as in other industries. The aviation industry has always been strong in data collection. This has not been limited to system reliability data but has been expanded to data on human performance and managerial influences. There is an industry-wide standard for accident and incident data collection (ADREP/ECCAIRS) which, while not perfect, provides a solid basis for quantification. In combination with FDR data and LOSA results (both of which are also standardised) this will help the proper representation of human and managerial influences. By using accident archetypes as the backbone of the model and top level representation and further development of the model from the generic to the specific with separate and formally defined levels of abstraction the seeming infinity of interdependencies can be taken care of. Even though 100% representation of all dependencies cannot be obtained, this approach will result in a practicable and useable model and helps avoiding being bogged down in development difficulties. In summary, developing a causal risk model for aviation is not overly difficult; it is just a lot of work. And the majority of this work will have to be conducted by people with sufficient background knowledge of air transport, because behind each causal conclusion there must be a known mechanism. Inferring causal relations requires subject specific background knowledge. Without a doubt, representing the complexity of the air transport system in a causal risk model will require numerous assumptions, each of which can be challenged. The causal risk model can only be a success if these assumptions are agreed upon by all parties involved, and the air transport regulator is the appropriate lead in this process. Preferably

this should be an international regulator like EASA or ICAO. They should initiate and lead the process to come to an industry standard model and to agree on how and when to apply this model. Examples from other industries, most notably the nuclear power industry and the process industry demonstrate the benefits of such a unified model.

For further development of causal risk models of air transport, it is essential to link such models directly to a data system, preferably ECCAIRS as this is a European and world standard and data accessibility is almost unrestricted. Feedback from the modellers to the potential users and vice versa is required to get a better grip on user requirements. It should also be verified, for each kind of user, that the causal risk model definition as adopted in this thesis, is suitable for them. The modellers should indicate possibilities and limitations of the model, and should come-up with real-life examples of cases in which a causal risk model would be a helpful tool. This will have to be an iterative process. Current experience shows that causal risk models are suitable to support strategic decision making by aviation professionals. They are specifically useful when the decision involves multiple disciplines or stakeholders. They should be used as input to the decision making process instead of using them for strictly assessing compliance with regulation. Support from the industry is essential for the provision of data and to provide insight, to the model developers, into operational aviation processes.

Final conclusion

This thesis did not describe the development of some new methodological risk modelling approach. Instead it considered risk modelling approaches with application examples from literature and investigated if and under which conditions the needs from the users can be met. The list of user needs that was developed is in itself an important result that had not been obtained before. The thesis also showed how difficult it will be to meet the needs of the user and how different scientific disciplines need to be integrated to obtain the desired result. The model developers should be aware of these difficulties and should know what must be made explicit to allow such integration.

Model development is very much comparable to aircraft design. The difficulty of aircraft design is not in aerodynamics, nor in structures. It is not in propulsion, and not in avionics. It is not in hydraulics and also not in aircraft performance. The difficulty lies in integrating all these disciplines into a single design, in making compromises such the overall result meets the objectives. Compromises are even needed in the objectives, because they can be conflicting as well. In safety model development, the disciplines that need to be integrated are various shades of engineering, mathematics, psychology and a touch of philosophy and economics. The users of the model are the industry, the regulator, and the society. The added value of this thesis is that it brought together the various disciplines and articulated the challenges associated with integrating them into a model that meets the needs of the users.

References

AAIASB. (2006). Accident Investigation report 11 / 2006, Accident of the a/c 5B-DBY of Helios Airways, Flight HCY522 on August 14, 2005, in the area of Grammatiko, Attikis, 33 km Northwest of Athens International Airport, Hellenic Republic, Ministry of Transport and Communications, Air Accident Investigation Board, Athens, Greece.

AAIB Switzerland. (2004). Investigation Report of the Aircraft Accident Investigation Bureau on the accident to aircraft AVRO 146-RJ100, HB-IXM, operated by Crossair under flight number CRX 3597 on 24 November 2001 near Bassersdorf/ZH, Nr. u1793, Federal Department of the Environment, Transport, Energy and Communications, Bern, Switzerland.

AAIU. (2005). Formal report 2005-015, Air Accident Investigation Unit, Dublin, Ireland.

AAS. (2001). Airside Aerodrome Manual, Version 1, Amsterdam Airport Schiphol, Business Unit Airlines.

Abbink, F.J. (1996). Integrated free-flight and 4-D gate-to-gate air traffic management, possibilities, promises and problems, NLR TP 96239U, National Aerospace Laboratory NLR, Amsterdam.

Ahlbohm, A., Cardis, E., Green, A., Linet, M., Savitz, D., Swerdlow, A. (2001). Review of the epidemiologic literature on EMF and health, Environmental Health Perspectives, Volume 109, Supplement 6, p. 911-933.

AIAA. (1998). Guide for the verification and validation of computational fluid dynamics simulations, G-077-1998, American Institute of Aeronautics and Astronautics, Reston, VA, USA.

Ale, B.J.M., Post, J.G., Bellamy, L.J. (1998). The interface between the technical and the management model for use in quantified risk analysis. In: Mosleh, A., Bari, R.A. (Eds.), Probabilistic Safety Assessment and Management, Vol. 3. Springer, p. 2087–2092.

Ale, B.J.M., Piers, M.A. (2000). Policy options for dealing with societal risk around Schiphol, NLR-CR-2000-084, RIVM nr. 610066011, National Aerospace Laboratory NLR, Amsterdam.

Ale, B.J.M., Golbach, G.A.M., Goos, D., Ham, K., Janssen, L.A.M., Shield, S.R. (2001). Benchmark risk analysis models, RIVM Report 610066015/2001, RIVM, Bilthoven, the Netherlands.

Ale, B.J.M. (2005). Tolerable or acceptable: A comparison of risk regulation in the United Kingdom and in the Netherlands, Risk Analysis, Vol. 25, No. 2, p. 231-241.

Ale, B.J.M., Bellamy, L.J., Cooke, R.M., Goossens, L.H.J., Hale, A.R., Roelen, A.L.C., Smith, E. (2006). Towards a causal model for air transport safety - an ongoing research project, Safety Science 44, p. 657 - 673.

Ale, B.J.M. (2007). Causal Model for Air Transport Safety, modified 7th interim report, Risk Centre, Delft University of Technology, Delft.

Ale, B.J.M., Bellamy, L.J., Van der Boom, R., Cooper, J., Cooke, R.M., Goossens, L.H.J., Hale, A.R., Kurowicka, D., Morales, O., Roelen, A.L.C., Spouge, J. (2007). Further development of a Causal model for Air Transport Safety (CATS): building the mathematical heart. Aven & Vinnem (eds), Risk, Reliability and Societal Safety, p 1431-1439, Taylor & Francis Group, London, UK.

Ale, B.J.M., Bellamy, L.J., Van der Boom, R., Cooper, J., Cooke, R.M., Lin, P.H., Morales, O., Roelen, A.L.C., Spouge, J. (2008). Using a Causal model for Air Transport Safety (CATS) for the evaluation of alternatives, ESREL2008, Valencia, Spain.

Amalberti, R. (2001). The paradoxes of almost totally safe transportation systems, Safety Science, 37, p. 109-126.

Amalberti, R., Auroy, Y., Berwick, D., Barach, P. (2005). Five system barriers to achieving ultrasafe health care, Annals of Internal Medicine, Volume 142, Number 9, p. 756-764.

Anderson, R.M. (1991). Discussion: The Kermack-McKendrick epidemic threshold theorem, Bulletin of Mathematical Biology, Vol. 53, Numbers 1 – 2, p. 1–32.

ATSB. (2001). Boeing 747-438, VH-OJH, Bangkok, Thailand, 23 September 1999, Australian Transport Safety Bureau, Canberra, Australia.

Avermaete, J.A.G. van. (1998). NOTECHS: Non-technical skill evaluation in JAR-FCL, NLR-TP-98518, National Aerospace Laboratory NLR, Amsterdam.

Bailey, N.T.J. (1975). The mathematical theory of infectious diseases and its applications, second edition, Oxford University Press, New York.

Baker, J.A. (chairman). (2007). The report of the BP US refineries independent safety review panel.

Ballin, M.G., Erzberger, H. (1996). An analysis of landing rates and separations at the Dallas / Fort Worth International Airport, NASA Technical Memorandum 110397.

Barnett, A., LoFaso, A.J. (1983). After the crash: the passenger response to the DC-10 disaster, Management Science, Vol. 29, issue 11, p. 1225-1236.

Barnett, A., Menhigetti, J., Prete, M. (1992). The Market Response to the Sioux City DC-10 Crash, Risk Analysis, March 1992, p. 12.

Barnett, A., Wang, A. (2000). Passenger-mortality risk estimates provide perspectives about airline safety, Flight Safety Digest, Vol. 19, No. 4, p. 1-12.

BEA. (2001). Accident on 25 July 2000 at La Patte d'Oie in Gonesse (95) to the Concorde registered F-BTSC operated by Air France, report number F-SC000725A, Bureau d'Enquêtes et d'Analyses pour la Sécurité de l'Aviation civile, le Bourget, France.

BEA. (2007). Rapport préliminaire. Accident survenu le 25 janvier 2007 sur l'aérodrome de Pau Pyrénées (64) au Fokker 28-100 immatriculé F-GMPG exploité par Régional, Bureau d'Enquêtes et d'Analyses pour la Sécurité de l'Aviation civile, le Bourget, France.

Beaumont, R. (1994). Flying to the limit: reminiscences of air combat, test flying and the aircraft industry, Haynes Publishing, Somerset, UK.

Bedford, T., Cooke, R. (2001). Probabilistic Risk Analysis: foundations and methods, Cambridge University Press, New York.

Bedford, T.J., Cooke, R.M. (2002). Vines - a new graphical model for dependent Random Variables. Annals of Statistics, Vol. 30, No. 4, p. 1031-1068.

Bellamy, L.J., Wright, M.S., Hurst, N.W. (1993). History and Development of a Safety Management System Audit for Incorporating into Quantitative Risk Assessment, International Process Safety Management Workshop, San Francisco, USA.

Bellamy, L.J., Papazoglou, I.A., Hale, A.R., Aneziris, O.N., Ale, B.J.M., Morris, M.I., Oh, J.I.H. (1999). Development of an integrated technical and management risk control and monitoring methodology for managing and quantifying on-site and off-site risks, I-Risk main report, May 1999, EU Contract number ENVA CT96-0243.

Benner, L. (1975). Accident investigations: Multilinear events sequencing methods, Journal of Safety Research, Vol. 7, Nr. 2, p. 67-73.

Bentley, R. (2005). Former Concorde programme head under investigation, Flight International, 27 September 2005.

Bernstein, L.N. (1996). Against the Gods, John Wiley & Sons, New York.

BFU. (2004). Untersuchungsbericht AX001-1-2/02, Bundesstelle für Flugunfallenuntersuchung, Braunschweig, Germany.

Biemans, M.C.M., Avermaete, J.A.G. van, Roelen, A.L.C., Meulen, G.J. van der. (1998). ADAMS Aircraft maintenance operational systems, identification of bottlenecks in the existing maintenance and dispatch system, Technical Report ADAMS –WP2B-TR01, Issue 1.0, NLR-TR-98120, National Aerospace Laboratory NLR, Amsterdam.

Bisignani, G. (2006). We need to do even better. Aviation Safety World, August 2006 issue, p. 11-12, published by the Flight Safety Foundation, Alexandria, VA, USA.

Bland, L.M. (1962). The problem of air transport flight safety, paper presented at the 15th Annual International Air Safety Seminar, Flight Safety Foundation.

Blom, H.A.P., Bakker, G.J., Blanker, P.J.G., Daams, J., Everdij, M.H.C., Klompstra, M.B. (2001). Accident risk assessment for advanced air traffic management, NLR-TP-2001-642, National Aerospace Laboratory NLR, Amsterdam.

Blom, H.A.P., Stroeve, S.H., Jong, H.H. de. (2006). Safety risk assessment by Monte Carlo simulation of complex safety critical operations, Proceedings of the 14th Safety Critical Systems Symposium, Bristol, UK.

Blom, H.A.P., Stroeve, S.H., Scholte, J.J., Jong, H.H. de. (2008). Accident risk analysis benchmarking Monte Carlo simulation versus event sequences, 3rd international conference on research in air transportation (ICRAT), Fairfax, VA, USA.

Brandsaeter, A. (2002). Risk assessment in the offshore industry, Safety Science, Vol. 40. p. 231-269.

Brauer, C. (2004). The risk landscape of the future. Swiss Reinsurance Company, Zurich, Switzerland.

Braun, C., Gründl, M, Marberger, C, Scherber, C. (2001). Beautycheck, Ursache und Folgen von Attraktivität, Universität Regensburg.

Bridgeman, W., Hazard, J. (1955). The lonely sky, Henri Holt and Company, New York, USA.

Brink, W.P. van den, Koele, P. (1985). Statistiek, deel 1, Datareductie, Boom, Amsterdam.

Brooker, P., Ingham T. (1977). Target Levels of Safety for Controlled Airspace, CAA Paper 77002, Civil Aviation Authority , London, UK.

Brooker, P. (2004). P-RNAV, Safety Targets, Blunders and Parallel Route Spacing, The Journal of Navigation, Vol. 57, p. 371-394.

Brown, E. (1981). The helicopter in civil operations. Granada Publishing, London, UK.

Burnside, E.S., Rubin, D.L., Fine, J.P., Shachter, R.D., Sisney, G.A., Leung, W.K. (2006). Bayesian network to predict breast cancer risk of mammographic microcalcifications and reduce number of benign biopsy results: initial experience. Radiology: 240: 3, p. 666-673.

Buys, J.R., Clark, J.L. (1995). Events and Causal Factors Analysis, SCIE-DOE-01-TRAC-14-95, Technical Research and Analysis Center, SCIENTECH Inc, Idaho Falls, USA.

CAA. (1998). Global fatal Accident Review 1980-1996, CAP 681, Safety Regulation Group, Civil Aviation Authority, London, UK.

CAA. (2000). Aviation Safety Review 1990-1999, CAP 701, Safety Regulation Group, Civil Aviation Authority, London, UK.

CAA. (2005). The mandatory occurrence reporting scheme, CAP 382, Safety Regulation Group, Civil Aviation Authority, London, UK.

CAA-NL. (2007). Civil aviation safety data, 1991-2005, Civil Aviation Authority Netherlands, The Hague.

Carpenter, M.S., Cooper, L.G., Glenn, J.H., Grissom, V.I., Schirra, W.M., Shepard, A.B., Slayton, DK. (1962). We Seven, Simon and Schuster, New York, USA.

Castano, D.J., Graeber, R.C. (2002). Aircraft manufacturer eager to learn from reliability data on normal flight operations, ICAO Journal, Vol. 57, Nr. 4, p. 10-11.

CCPS. (1989). Guidelines for Chemical Process Quantitative Risk Analysis, Centre for Chemical Process Safety (CCPS), American Institute of Chemical Engineers, New York, USA.

Chaikin, A. (1994). A man on the Moon, Michel Joseph, UK.

Chalk, A. (1986). Market forces and aircraft safety: The case of the DC-10, Economic Enquiry, Vol. 24, issue 1, p. 43-60.

Chappell, S.C. (1994). Using voluntary incident reports for human factors evaluations, in McDonald, N., Johnston, N, Fuller, R. (eds), Aviation Psychology in Practice, Avebury Technical, Aldershot, UK.

Cheung, Y.S., Post, J.A. (2005). Revised accident rates of third generation aircraft for NLR IMU model 2004 (RANI-2004), NLR-CR-2005-656, National Aerospace Laboratory NLR, Amsterdam.

Columbia Accident Investigation Board. (2003). Report Volume 1.

Cooke, R.M., Goossens, L.H.J. (2000). Procedures guide for structured expert judgement, EURATOM document EUR 18820EN, European Communities.

Cooke, R.M., Kurowicka, D., Hanea, A.M., Morales, O., Ababei, D.A., Ale, B, Roelen, A. (2007). Continuous/Discrete Non Parametric Bayesian Belief Nets with UNICORN and UNINET, Proceedings of Mathematical Methods in Reliability MMR 2007, Glasgow, UK.

Cooper, G.E., Harper, R.P. (1967). The use of pilot rating in the evaluation of aircraft handling qualities, AGARD report 567, France.

Corrigan, J., Kohn, L.T., Donaldson, M.S. (eds). (2000). To err is human. Building a safer health system. National Academy Press, Washington D.C., USA.

Court of Session. (2005). Statement by the Right Hon. Lord Nimmo Smith when delivering judgment in the cause mrs. Margaret McTear against Imperial Tobacco ltd at the Court of Session, Edinburgh on 31 May 2005, Edinburgh, UK.

Crossfield, A.S., Blair, C. (1960). Always another dawn: the story of a rocket test pilot, The World Publishing Company, Cleveland, Ohio, USA.

Cullen, W.C. (2001). The Ladbroke Grove Rail Inquiry, HSE books, Sudbury, Suffolk, UK.

Czerwinski, D., Barnett, A. (2004). Airlines as baseball players: An alternative approach to evaluating air carrier safety record, Draft report prepared for FAA Tech Center, Atlantic City, NJ. USA.

Daso, D.A. (2003). Doolittle, aerospace visionary, Brassey's, Dulles, VA, USA.

Davoudian, K., Wu. J-S., Apostolakis, G. (1994). The work process analysis model (WPAM) Reliability Engineering and System Safety, Vol. 45, p. 107-125.

Deacon, N., Rochard, B. (2000). Fifty years of airfield grass management in the UK, In J. van Nugteren (ed), Proceedings of the 25th meeting of the International Bird Strike Committee, IBSC, Amsterdam, p. 71-79.

Deazle, A., Hayes, P, Schmidlin, M., Tress, G. (1999). The development of a taxonomy of cost factors, DESIRE technical report WP-3-TR3B.

Dekker, S.W.A. (2005). Why we need new accident models, Technical report 2005-02, Lund University of Aviation, Sweden.

Deming, W.E. (1990). Out of crisis: quality, productivity and competitive position. Cambridge University Press, Cambridge, UK.

DGL. (2002a). Luchthavenindelingbesluit Schiphol, Ministerie van Verkeer en Waterstaat, Directoraat Generaal Luchtvaart, The Hague.

DGL. (2002b). Luchthavenverkeersbesluit, Ministerie van Verkeer en Waterstaat, Directoraat Generaal Luchtvaart, The Hague.

DGL. (2004). Evaluatie Schipholbeleid, Plan van Aanpak, Ministerie van Verkeer en Waterstaat, Directoraat Generaal Luchtvaart, The Hague.

DGTL. (2005). Beleidsagenda Luchtvaartveiligheid, Ministerie van Verkeer en Waterstaat, The Hague.

Dierikx, M. (1997). Dwarswind, een biografie van Anthony Fokker, Sdu publishers, The Hague.

Dijk, W. van, Molemaker, R.J., Hayes, P., Roelen, A.L.C. (2001). The cost of unsafety, ASTER WP3, NLR-CR-2001-154, National Aerospace Laboratory NLR, Amsterdam.

DNV. (2002a). Marine Risk Assessment, Offshore Technology Report 2001/063, Published by the Health and Safety Executive, UK.

DNV. (2002b). Causal modelling of air safety, CM-DNV-018, DNV London, UK.

Doolittle, J.H. 1952. The Airport and Its Neighbors, President's Airport Commission.

Doorn, B.A. van. (2006). Analysis of a selection of safety assessments with respect to ATSF strategic objectives. NLR Memorandum ATSF-2006-068, National Aerospace Laboratory NLR, Amsterdam.

Duke, N. (1953). Test pilot, Allan Wingate, London, UK.

EAI. (2002). Review of Causal Modeling of Air Safety for the Directorate General of Civil Aviation - the Netherlands, Report of Peer Reviews, Enders Associates International, Bethesda, MD, USA.

EASA. (2003a). Certification Specification for Large Aeroplane, CS-25, European Aviation Safety Agency, Brussels, Belgium.

EASA. (2003b). Decision No. 2003/6/RM of the Executive Director of the Agency of 17 October 2003 on certification specifications, including airworthiness codes and acceptable means of compliance, for all weather operations (« CS-AWO »), European Aviation Safety Agency, Brussels, Belgium.

EC. (1982). Council Directive 82/501/EEC 1982 on the major-accident hazards of certain industrial activities, Council of the European Communities, Luxemburg.

EC. (1996). Council Directive 96/82/EC on the control of major-accident hazards involving dangerous substances, Council of the European Union, Brussels, Belgium.

EC. (2003). Council Directive 2003/42/EC on occurrence reporting in civil aviation, Council of the European Union, Brussels, Belgium.

EC. (2004). Corrigendum to Directive 2004/49/EC of the European Parliament and of the Council of 29 April 2004 on safety on the Community's railways and amending Council Directive 95/18/EC on the licensing of railway undertakings and Directive 2001/14/EC on the allocation of railway infrastructure capacity and the levying of charges for the use of railway infrastructure and safety certification (Railway Safety Directive), Official Journal of the European Union L 220 of 21 July 2004.

EC. (2005). Rapid agreement reached on airline blacklist, Press release IP/05/1429, European Commission, Brussels, Belgium.

Eijndhoven, J.C.M. van, Weterings, R.A.P.M., Worrell, C.W., Boer, J. de, Pligt, J. van der, Stallen, P.J.M. (1994). Risk communication in the Netherlands: the monitored introduction of the EC 'post-Seveso' Directive, Risk Analysis, Vol. 14, no. 1, p. 87-95.

Enders, J.H., Dodd, R., Tarrel, R., Khatwa, R., Roelen, A.L.C., Karwal, A.K. (1996). Airport safety: A study of accidents and available approach-and-landing aids, Flight Safety Digest, Vol. 15, No. 3, Flight Safety Foundation, Alexandria, VA, USA.

Ericson, C.A. (1999). Fault tree analysis - a history, 17th International System Safety Conference, Orlando, Florida, USA.

Es, G.W.H. van, Roelen, A.L.C., Kruijsen, E.A.C., Giesberts, M.K.H. (1998). Safety aspects of aircraft performance on wet and contaminated runways, paper presented at the 10th European Aviation Safety Seminar, Amsterdam.

Es, G.W.H. van. (2005). Running out of runway, analysis of 35 years of landing overrun accidents, NLR-TP-2005-498, National Aerospace Laboratory NLR, Amsterdam.

Es, G.W.H. van, Geest, P.J. van der. (2006). A study of normal operational landing performance on subsonic civil narrow body jet aircraft during ILS approaches, NLR-CR-2006-049, National Aerospace Laboratory NLR, Amsterdam.

ETSC. (1997). Transport accident costs and the value of safety, European Transport Safety Council, Brussels, Belgium.

Eurocontrol. (2001a). The EUR RVSM pre-implementation safety case, RVSM 691, version 2.0, Eurocontrol, Brussels, Belgium.

Eurocontrol. (2001b). Eurocontrol Safety Regulatory Requirement (ESARR) 4, Risk assessment and mitigation in ATM, Edition 1.0, Eurocontrol Safety Regulation Commission, Brussels, Belgium.

Eurocontrol. (2003). Explanatory Material on ESARR 4 Requirements, Edition 1.0, Eurocontrol Safety Regulation Commission, Brussels, Belgium.

Eurocontrol. (2004). EAM 4/ AMC Acceptable means of compliance with ESARR 4, Edition 3.0, Eurocontrol Safety Regulation Commission, Brussels, Belgium

Eurocontrol. (2006a). Safety Assessment Methodology (SAM), SAF.ET1.ST03.1000-MAN-01, version 2.1, Eurocontrol, Brussels, Belgium.

Eurocontrol. (2006b). Safety Assessment report on independent approaches to parallel instrument runways at Helsinki-Vantaa Airport, final version 1.2, Eurocontrol, Brussels, Belgium.

Eurocontrol. (2007). EEC Mid-Term Validation Safety Plan, version 1.0, Eurocontrol, Brussels, Belgium.

Evans, A.W., Foot, P., Mason, S.M., Parker, I.G., Slater, K. (1997). Third party risk near airports and public safety zone policy, R&D Report 9636, National Air Traffic Services, London, UK.

Everdij, M.H.C. (2004). Review of techniques to support the EATMP safety assessment methodology, Volume I, EEC Note No. 01/04, Project SRD-3-E1.

Everdij, M.H.C., Blom, H.A.P. (2005). Piecewise Deterministic Markov Processes represented by Dynamically Coloured Petri Nets, Stochastics, Vol. 77, p. 1-29.

Everdij, M.H.C., Blom, H.A.P., Stroeve, S.H. (2006). Structured assessment of bias and uncertainty in Monte Carlo simulated accident risk, paper presented at the 8th International Conference on Probabilistic Safety Assessment and Management (PSAM 8), New Orleans, Louisiana, USA.

Everdijk, H. (2006). Luid maar duidelijk. Commissie deskundigen vliegtuiggeluid, The Hague.

FAA. (1988). Advisory Circular AC 25.1309-1A, System Design and Analysis, 21 June 1988, Federal Aviation Administration, Washington D.C., USA.

FAA. (1995a). Aviation safety action plan. Zero accidents, a shared responsibility, Federal Aviation Administration, Washington D.C., USA.

FAA. (1995b). Proceedings of the Aviation Safety Conference held in Washington DC on January 9 and 10, 1995, Federal Aviation Administration, Washington D.C., USA.

FAA. (1997). Aviation Safety Data Accessibility Study Index, a report on issues related to public interest in aviation safety data, Office of System Safety, Federal Aviation Administration, Washington D.C., USA.

FAA. (2001). System Approach for Safety Oversight, Mission Need Statement, Federal Aviation Administration, Washington D.C., USA.

FAA. (2002). The report on the FAA Associate Administrator for Regulation and Certification's Study on the Commercial Airplane Certification Process, Federal Aviation Administration, Washington D.C., USA.

FAA. (2003). FAR 121.344, Digital flight data recorders for transport category airplanes, Doc. No. 28109, 62 FR 38378, July 17, 1997; 62 FR 48135, Sept. 12, 1997, as amended by Amdt. 121–300, 68 FR 42936, July 18, 2003; 68 FR 50069, Aug. 20, 2003, Federal Aviation Administration, Washington D.C. USA.

FAA. (2004). System Approach for Safety Oversight, Commercial Aviation, RE&D Requirements, FY2004-2005, version 7.0, Federal Aviation Administration, Washington D.C., USA.

FAA. (2006). Advisory Circular 120-90, Line Operations Safety Audit. Federal; Aviation Administration, Washington D.C., USA.

Feynman, R. (1965). The character of physical law, The British Broadcasting Corporation. London, UK.

Flight International. (2006a). Editorial, The danger of hindsight in criminal air accident cases, 21 March 2006.

Flight International. (2006b). France prosecutes six over 1992 Air Inter Crash, 21 March 2006.

Forsyth, R.J., Balogh, A., Smith, E.J. (2002). Latitudinal variation of the underlying heliospheric magnetic field direction: Comparison of the Ulysses first and second orbits, Space Science Reviews, 97, p. 161-164.

Fragola, J.R. (1996). Risk management in US manned spacecraft, from Apollo to Alpha and beyond, Proceedings of the ESA Product Assurance Symposium and Software Product Assurance Workshop, ESA SP 377, p. 83-92.

FSF. (1998). A Study of Fatal Approach-and-Landing Accidents Worldwide, 1980-1996, Flight Safety Digest, Vol. 17, No. 2/3, Flight Safety Foundation, Alexandria, VA, USA.

FSF. (1999). Killers in Aviation: FSF Task Force Presents Facts About Approach-and-landing and Controlled-flight-into-terrain Accidents, Flight Safety Digest, Vol. 17, No. 11-12, Vol. 18, No. 1-2, Flight Safety Foundation, Alexandria, VA, USA.

Galles, D., Pearl, J. (1998). An axiomatic characterization of causal counterfacts, Foundations of Science, Volume 3, Issue 1, p. 151-182.

Geest, P.J. van der, Kruijsen, E.A.C., Ramsay, C.G., Slater, D.H. (2005). Safety and compatibility of mixed VFR/IFR air traffic at Geneva Airport, NLR-CR-2005-102, National Aerospace laboratory NLR, Amsterdam.

Gier, M, de. (2005). Luchtvaartveiligheid, de beleving van de Nederlandse bevolking, TNS-NIPO consult, report E1537, TNS-NIPO Amsterdam.

Goold, I. (2000). The Modern Jet Airliner - the trailblazers, in Ph. Jarret (series editor), Modern Air Transport, worldwide air transport from 1945 to the present, Putnam Aeronautical Books, London, UK.

Goossens, L.J.J., Cooke, R.M., Hale, A.R., Rodíc-Wiersma, Lj. (2008). Fifteen years of expert judgement at TU Delft, Safety Science, Vol. 46, Issue 2, p. 234-244.

Gould, S.J. (1981). The mismeasure of man, W.W. Norton & Company, New York.

Groen, F.J., Smidts, C., Mosleh, A. (2006). QRAS - the quantitative risk assessment system, Reliability Engineering and System Safety, Volume 91, Issue 3, p. 292-304.

Group of Personalities. (2001). European Aeronautics: a vision for 2020. Meeting society's needs and winning global leadership, Office for Official Publications of the European Communities.

GS. (2003). Antwoord van Gedeputeerde Staten op schriftelijke vragen van het lid mevrouw E. Tans inzake Uitlatingen van Gedeputeerde Vestjens in het programma Confrontatie van de RVU over geluidsoverlast rond MAA, 28 oktober 2003.

GVA. (2005). Gazet van Antwerpen, www.gva.be, Dossier Enschede.

Hakfoort, J., Poot, T., Rietveld, P. (2001). The regional economic impact of an airport: the case of Amsterdam Schiphol Airport, Regional Studies, 35:7, p. 595-604.

Hale, A.R., Glendon, A.I. (1987). Individual behaviour in the control of danger. Elsevier, Amsterdam.

Hale, A.R., Heming, B.H.J., Carthey, J., Kirwan, B. (1997). Modelling of safety management systems, Safety Science, Vol 26, No. 1/2, p. 121-140.

Hale, A.R., Goossens, L.H.J., Bellamy, L.J. (1999). Modelling errors of commission: applicability of ATHEANA in the chemical industry. In: L.H.J. Goossens (eds.); Proceedings Rotterdam '99. Risk Analysis: Facing the New Millennium (Rotterdam, 10/10/99), Delft University Press, Delft, p. 243-246.

Hale, A.R. (2000). Railway safety management: the challenges of the new millennium, Safety Science monitor, issue 1, vol. 4.

Hale, A.R. (2001). Conditions of occurrence of major and minor accidents, Journal of the Institution of Occupational Safety and Health, Vol. 5, No. 1, p. 7-21.

Hale, A.R. Guldenmund, F. (2004). ARAMIS audit manual, version 1.3Safety Science Group, Delft University of Technology, Delft, the Netherlands.

Hale, A.R., Ale, B.J.M., Goossens, L.H.J., Heijer, T., Bellamy, L.J., Mud, M.L., Roelen, A., Baksteen, H., Post, J., Papazoglou, I.A., Bloemhoff, A., Oh, J.I.H. (2007). Modeling accidents for prioritizing prevention, Reliability Engineering and System Safety, Vol. 92, No. 12, p. 1701-1715.

Halpern, J.Y., Pearl, J. (2001). Causes and Explanations: A Structured-Model approach, Part I: Causes, in Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, San Francisco, CA, Morgan Kauffmann, p. 194-202.

Hansen, M., McAndrews, C., Berkeley, E. (2005). History of aviation safety oversight in the United States, Research Report NR-2005-001, Institute of Transportation Studies, University of California at Berkeley, USA.

Haubrich, E.J.A. (1984). The certification of a new type of transport category aircraft, NLR TR 84001, National Aerospace laboratory NLR, Amsterdam.

Health Council of the Netherlands. (1999). Public health impact of large airports. Health Council of the Netherlands 1999/14E, Committee on the Health Impact of Large Airports, The Hague, Netherlands.

Heijden, F.M.M.A., van der. (2006). Toegewijd, maar oververmoeid, Medisch Contact, Nr. 45, p. 1792-1995.

Heimplaetzer, P., Busch, C. (2006). Safety management in rail infrastructure, paper presented at the 3rd International Conference 'Working on Safety', 12-15 September 2006, Zeewolde, the Netherlands.

Heinrich, H., Petersen, D., Roos, N. (1980). Industrial accident prevention, McGraw-Hill, New York (original edition, H. Heinrich, 1931).

Helmreich, R.L., Butler, R.E., Taggart, W.R.,Wilhelm, J.A. (1994). The NASA / University of Texas / FAA Line / LOS Checklist: A behavioral marker based checklist for CRM skills assessment. NASA/UT/FAA Technical Report 94-02, revised 12/8/95, The University of Texas Austin, TX, USA.

Hesslow, G. (1976). Discussion: two notes on the probabilistic approach to causality, Philosophy and Science, 43, p. 290-292.

Hesslow, G. (1988). The problem of causal selection, in Hilton, D.J. (ed.) Contemporary science and natural explanation, Common sense conceptions of causality, Harvester Press, Sussex, UK.

Hilburn, B.G., Bakker, M.W.P., Pekela, W.D. (1998). Free flight and the air traffic controller: an exploratory study of human factors issues, NLR-TP-98237, National Aerospace Laboratory NLR Amsterdam.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., Dement, W.C. (1973). Quantification of sleepiness: A new approach. Psychophysiology, 10, p. 431-436.

Hollnagel, E. (1998). Cognitive reliability and error analysis method, Elsevier Science, Oxford, UK.

Hollnagel, E. (2004). Barriers and accident prevention, Ashgate Publishing, Aldershot, UK.

Hollnagel, E., Woods D.D., Leveson, N.G. (eds). (2006). Resilience engineering: Concepts and precepts, Ashgate Publishing, Aldershot, UK.

Hooey, B.L., Foyle, D.C., Andre, A.D. (2000). Integration of cockpit displays for surface operations: The final stage of a human-centered design approach. SAE Transactions: Journal of Aerospace, 109, p. 1053-1065.

Horibe, K., Kawahari, K., Sakai, J., Sakaki, J. (2004). Development of GE90-115B turbofan engine, IHI Engineering Review, Vol. 37, No. 1.

Hörmann, H.J., Berg, P. van den, Peixoto, J.L., Robinson, J., Rager, T., Belyavin, A., Hosman, R. (2005). Analysis of pilot control behaviour during balked landing manoeuvres, AIAA 2005-5881, AIAA Modelling and Simulation Technologies Conference and Exhibit, San Francisco, California, USA.

Hosman, R., Schuring, J., Geest, P.J. van der. (2005). Pilot model development for the balked landing manoeuvre, AIAA 2005-5884, AIAA Modelling and Simulation Technologies Conference and Exhibit, San Francisco, California, USA.

Hourtolou, D., Salvi, O. (2003). ARAMIS project: development of an integrated accidental risk assessment methodology for industries in the framework of SEVESO II directive, in Bedford T., Van Gelder P.H.A.J.M (eds), Safety and Reliability - ESREL 2003, p. 575 - 581.

HSE. (1998). A Guide to the Offshore Installations (Safety Case) Regulations 1992, Health & Safety Executive, HSE Books, Sudbury, UK.

Hubert, Ph., Barni, M.H., Moatti, J.P. (1990). Elicitation of criteria for management of major hazards, 2sd SRA conference, Laxenburg, Austria.

Hudson, P. (2003). Applying the lessons of high risk industries to health care, Quality and Safety in Health Care, 12, p. 7-12.

Hyams, K.C., Wignall, F.S., Roswell, R. (1996). War syndromes and their evaluation - From the U.S. Civil War to the Persian Gulf War, Annals of Internal Medicine, Volume 125, Issue 5, p. 398-405.

IATA. (2008a). Operational Safety Audits, Programme Manual, 3rd edition, International Air Transport Association, Montreal, Canada.

IATA. (2008b). IOSA Standards Manual, 2nd edition, Revision 1, International Air Transport Association, Montreal, Canada.

ICAO. (1980). Manual on the Use of the Collision Risk Model (CRM) for ILS Operations, Doc 9274 AN/904, International Civil Aviation Organization, Montreal, Canada.

ICAO. (2001). Aircraft accident and incident investigation, Annex 13 to the Convention on International Civil Aviation, ninth edition, International Civil Aviation Organization, Montreal, Canada.

ICAO. (2002a). Operation of Aircraft, Part 1 International Commercial Air Transport – Aeroplanes, Annex 6 to the Convention on International Civil Aviation, eighth edition, International Civil Aviation Organization, Montreal, Canada.

ICAO. (2002b). Line Operations Safety Audit (LOSA). Doc 9803 AN/761, International Civil Aviation Organization, Montreal, Canada.

ICAO. (2002c). Manual on airspace planning methodology for the determination of separation minima, Doc 9689 AN/953, Amendment No. 1, International Civil Aviation Organization, Montreal, Canada.

ICAO. (2004a). Manual on simultaneous operations on parallel or near-parallel instrument runways (SOIR), Doc 9643 AN/941, International Civil Aviation Organization, Montreal, Canada.

ICAO. (2004b). Aerodromes, Volume 1, Aerodrome design and operations, Annex 14 to the Convention on International Civil Aviation, fourth edition, International Civil Aviation Organization, Montreal, Canada.

ICAO. (2006). Safety Management Manual, Doc 9859, International Civil Aviation Organization, Montreal, Canada.

Isaac, A., Straeter, O., Damme, D, van. (2004). A method for predicting human error in ATM (HERA-PREDICT). Eurocontrol report HRS/HSP-002-REP-07, EURONTROL, Brussels, Belgium.

ISO. (1999). Safety aspects – guidelines for their inclusion in standards, ISO/IEC guide 51:1999, International Organisation for Standardisation, Geneva, Switzerland.

IVW. (2003). Handhavingsbeleid: Wet Luchtvaart; De Luchthaven Schiphol, Inspectie Verkeer en Waterstaat, Divisie Luchtvaart, Hoofddorp.

IVW. (2007). Visiedocument toezicht in beweging, Regiemodel toezicht Schiphol, Inspectie Verkeer en Waterstaat, Toezichteenheid Luchthavens en Luchtruim, Hoofddorp, the Netherlands.

IVW. (2008). Implementatie van de EU richtlijn (2003/42/EG), rapportage over het eerste jaar van meldingen van voorvallen in de Burgerluchtvaart, Inspectie Verkeer en Waterstaat, Hoofddorp, the Netherlands.

Jensen, F.V. (1995). Introduction to Bayesian networks, University College London Press, UK.

JIAA. (1996). Final Aviation Accident Report, Birgenair flight ALW-301, Puerto Plata, Dominican Republic, February 6, 1996; Original report published by the Junta Investigadora de Accidentes Aéreos (JIAA), Director General of Civil Aeronautics, Dominican Republic; Translated copy distributed by the Air Line Pilots Association.

Johnson, W.G. (1975). MORT: The management oversight and risk tree, Journal of Safety Research, Vol. 7. No. 1, p. 4-15.

Johnston, A.M., Barton, C. (1991). Tex Johnston, jet-age test pilot, Smithsonian Institution Press, Washington D.C., USA.

Jones, E., Hodgins-Vermaas, R., McCartney, H., Everitt, B., Beech, C., Poynter, D., Palmer, I., Hyams, K., Wessely, S. (2002). Post-combat syndromes from the Boer war to the Gulf War: a cluster analysis of their nature and attribution, BMJ, Volume 324.

Jong, H.H. de. (2004). Guidelines for the identification of hazards, how to make unimaginable hazards imaginable? NLR-CR-2004-094, National Aerospace Laboratory NLR, Amsterdam.

Jong, H.H. de. (2006). Identification of user requirements for CATS, NLR Memorandum ATSF-2006-096, National Aerospace Laboratory NLR, Amsterdam.

Jong, H.H. de. (2007). Second round of interviews to identify CATS user requirements, NLR Memorandum ATSI-2007-085, National Aerospace Laboratory NLR Amsterdam.

Joyce, T., Graham, G., Kinnersly, S., Van Eenige, M.J.A., Roelen, A.L.C. (2001). A study into Target Levels of Safety (TLS) within the aviation industry, including comparative analyses with the rail and nuclear power sectors, ASTER WP 1, NLR-CR-2001-145, National Aerospace Laboratory NLR, Amsterdam.

JPDO. (2004). Evaluation working group of the Joint Planning and Development Office, Overall evaluation process and first year test plans, DRAFT 1407 EWG Plan 1-1, version 2.

Kampen, A. van. (1960). Plesman. De Boer publishing company, Hilversum.

Kant, I. (1781). Kritik der reinen vernunft.

Keller, W., Modarres. M. (2005). A historical overview of probabilistic risk assessment development and its use in the nuclear power industry, a tribute to the late Professor Norman Carl Rasmussen, Reliability Engineering and System Safety, 89, p. 271-285.

Kelly, T. J. (2001). Moon Lander, Smithsonian Institution Press, Washington D.C., USA.

Kemeny, J., (1979). Report of the President’s Commission on the accident at Three Mile Island, Washington D.C., USA.

Khatwa, R., Roelen, A.L.C. (1996). An analysis of controlled-flight-into-terrain (CFIT) accidents of commercial operators, 1988 through 1994, Flight Safety Foundation, Flight Safety Digest, Vol. 15, No. 4/5, p. 1-45.

Kho, M.E., Carbone, J.M., Luicas, J., Cook, D.J. (2005). Safety Climate Survey: reliability of results from a multicenter ICU survey, Quality and Safety in Health Care, 14, p. 273-288.

Kingsley-Jones, M. (1997). A330-200 flight testing programme takes off, Flight International, 20 August 1997.

Kingsley-Jones, M. (2005). A380 powers on through flight-test, Flight International, 20 December 2005.

Kirwan, B. (1994). A practical guide to human reliability assessment, Taylor and Francis, London, UK.

Kirwan, B. (2007). Safety informing design, Safety Science 45, p. 155-197.

Klinect, J.R., Murray, P., Merritt, A., Helmreich, R. (2003). Line Operations Safety Audit (LOSA): Definition and operating characteristics, Proceedings of the 12th International Symposium on Aviation Psychology, Dayton, OH, USA, p. 663-668.

Klinect, J.R. (2005). Line operations safety audit: a cockpit observation methodology for monitoring airline safety performance, PhD dissertation, The University of Texas, Austin, TX, USA.

KLM. (2007). Continuation Training Bulletin 07/11, p. 4.

Koornneef, F. (2000). Organised learning from small scale incidents, Delft University Press, Delft, the Netherlands.

Kreijkamp, H.A., Veerbeek, H.W. (1997). Mid-life Update FANOMOS RLD: Reference manual version 1.0, Report NLR-CR-97386 L, National Aerospace laboratory NLR, Amsterdam

Kurowicka, D., Cooke, R.M., Charitos, T., Speijker, L.J.P. (2005). ATC- wake risk assessment model, Annex 1, Continuous Bayesian Belief Networks, ATC Wake, EC contract number IST-2001-34729, Deliverable d3_5b, Annex 1.

Kurowicka, D., Cooke, R.M. (2004). Distribution - free continuous Bayesian Belief Nets, Proceedings of the Conference on Mathematical Methods in Reliability, Santa Fe, New Mexico, USA.

K+V. (2005). Eindrapport, Veiligheidsadviescommissie Schiphol, Veiligheidsonderzoek Schiphol 2005.

Labeau, P.E., Smidts, C., Swaminathan, S. (2000). Dynamic reliability: towards an integrated platform for probabilistic risk assessment, Reliability Engineering and System Safety 68, p. 219-254.

P.S. Laplace, P.S. (1814). Essai Philosophique sure les Probabilites. Courcier, New York, English translation by F.W. Truscott and F.L. Emory, Wiley, NY, 1902.

Lawson, T.W., Considine, R. (1944). Thirty seconds over Tokyo, Blue Ribbon Books, Garden City, NY, USA.

Leary, W.M. (1992). Safety in the air: the impact of instrument flying and radio navigation on U.S. commercial air operations between the wars, Leary, W.M. (ed), The history of civil and commercial aviation, Vol. 1 Infrastructure and environment, Smithsonian Institution Press, Washington D.C., USA.

Leege, A.P.M. de, Hörmann, H.J. (2006). Balked landing study, modelling of flare initiation altitude and determination of decrab initiation altitude, Boeing Research & Technology Europe, Madrid, Spain.

Le Maitre A.S. (Chairman). (1957). Report of Departmental Committee on Safeguarding Policy, Ministry of Transport and Civil Aviation, London, UK.

Leeuw, R. de (ed). (1994). Fokker commercial aircraft, Public Relations department of the N.V. Koninklijke Nederlandse Vliegtuigenfabriek Fokker.

Leveson, N. (2004a). A systems-theoretic approach to safety in software-intensive systems, IEEE Transactions on Dependable and Secure computing, Vol. 1, Issue 1, p. 66-86.

Leveson, N. (2004b). A new accident model for engineering safer systems, Safety Science, Vol. 42, No. 4, p. 237-270.

Levy, R.A. (2001). There are no new accidents - most aircraft accidents preventable, Flying safety, July 2001.

Lewis, H.W., Budnitz, R.J., Kouts, H.J., Lowenstein, W.B., Rowe, W.D., Von Hippel, F., Zachariasen, F. (1979). Risk assessment review group report to the U.S. Nuclear Regulatory Commission, NUREG/CR-0400, U.S. Nuclear Regulatory Commission, Washington D.C., USA.

Lin, P.H. (2007). Causal model for air transport safety: findings from the accident and incident data, in P.H. Lin and J. Klungers (eds), The research agenda of risk and design anno 2007, PhD student symposium, Faculty of Technology, Policy and Management, Delft University of Technology, Delft, the Netherlands.

Lin, P.H., Hale, A.R., Gulijk, C. van, Ale, B.J.M., Roelen, A.L.C., Bellamy, L.J. (2008). Testing a Safety Management System in aviation, paper presented at the Ninth International Probabilistic Safety Assessment and Management Conference, Hong Kong.

Lindbergh, C.A. (1953). The Spirit of St. Louis, Charles Scribner's Sons, USA.

Lindley, D.V. (1986). Obituary: Bruno de Finetti, 1906-1985, Journal of the Royal Statistical Society, Series A149, p. 252.

Lithgow, M. (1956). Vapour Trails, Allan Wingate, London, UK.

Lloyd, E. (1980). The development of requirements for safety assessment, Cranfield College of Aeronautics, UK.

Lloyd, E., Tye, W. (1982). Systematic Safety, Civil Aviation Authority, London, UK.

Loo, M. van het, Bosman, S., Vader, J., Frinking, E., Kahan, J. (1999). Berichtgeving en percepties ten aanzien van de externe veiligheid van de luchtvaart, RE-99.013, Rand Europe.

Lorenz, E.N. (1993). The Essence of Chaos, Jessie and John Danz lecture, University of Washington Press, Seattle, USA.

Luxhøj, J.T., Choopavang, A., Arendt, D.N. (2001). Risk assessment of organizational factors in aviation systems, Air Traffic Quarterly, Vol. 9, p. 135-137.

Luxhøj, J.T. (2003). Probabilistic causal analysis for system safety risk assessments in commercial air transport, paper presented at the Second Workshop on the Investigation and Reporting of Accidents and Incidents, IRIA 2003, Williamsburg, Virginia, USA.

Luxhøj, J.T. (2004). Building a safety risk management system: a proof of concept prototype, FAA/NASA Risk Analysis Workshop, Arlington, VA, USA, August 19, 2004.

LVNL. (2001a). Veiligheid is niet het doel, Air Traffic Control the Netherlands.

LVNL. (2001b). Evaluatie VEM rapportage UDP, version 1.0, Air Traffic Control the Netherlands.

LVNL. (2002). VEM Effect Report ATM System Increment 2002, Volume 1, main document, version 1.1, Department Research & Development / Performance Analysis, Air Traffic Control the Netherlands.

Lyman, E.G., Orlady, H.W. (1980). Fatigue and associated performance decrements in air transport operations, NASA Contract NAS2-100060, Batelle Memorial Laboratories, Aviation Safety Reporting System ASRS, Mountain View, CA, USA.

MacIntosh, R.M. (1993). Accidents show need for comprehensive ground deicing programs, Airport Operations, Vol. 19, No. 6, Flight Safety Foundation, Arlington, VA, USA.

MacKinnon, R.A. (2000). Rejected Take-off Studies, Aero magazine 11, published quarterly by Boeing Commercial Airplane Group.

McRuer, D.T., Graham, D., Krendel, E.S., Reisener, W. (1965). Human pilot dynamics in compensatory system. theory, models and experiments with controlled element and forcing function variations. AFFDL-TR-65-15. Wright Patterson Air Force Base, Ohio, USA.

Meltzer, M. (2007). Mission to Jupiter, a history of the Galileo project, NASA SP-2007-4231, NASA, Washington D.C., USA.

Mermoz, J. (1937). Mes vols, Flammarion, France.

MNP. (2005). Het milieu rond Schiphol, 1990 - 2010, feiten en cijfers, Milieu- en NatuurPlanbureau, Bilthoven.

Molemaker, R.J., Piers, R., Adamidis, P., Cloos, B., Geest, P. van der. (2005). Impact Assessment on the extension of EASA competences to ANS, ATM and Airports, Final report, ECORYS, Rotterdam.

Morales, O., Kurowicka, D., Roelen, A. (2008). Eliciting conditional and unconditional rank correlations from conditional probabilities, Reliability Engineering and System Safety, Vol. 93, Issue 5, p. 699-710.

Morier, Y. (2002). NPA 11-2; Regulatory Impact Assessment Guidance, letter to NPA Distribution List, Sectorial Teams and Working Parties, reference 07/03-2-2 02-L197, Joint Aviation Authorities, Hoofddorp, the Netherlands.

Moshansky, V.P. (1992). Commission of Inquiry into the Air Ontario crash at Dryden, Ontario, Final report (Vol 1-4), Published by the Ministry of Supply and Services, Canada.

Mosleh, A., Goldfeiz, E.B. (1994). An approach for assessing the impact of organisational factors on risk, Unites States Nuclear Regulatory Commission, Office of Nuclear Research, Washington D.C., USA.

Mosleh, A., Dias, A., Eghbali, G., Fazen, K. (2004). An integrated framework for identification, classification and assessment of aviation system hazards. Probabilistic safety assessment and management: PSAM 7 - ESREL '04: proceedings of the 7th International Conference on Probabilistic Safety Assessment and Management, Berlin, Germany.

Muller, F., Thiel, J.H. (1986). Beknopt Grieks-Nederlands Woordenboek, elfde druk, Wolters-Noordhoff, Groningen.

Murata, T. (1989). Petri Nets: Properties, Analysis and Applications, Proceedings of the IEEE, Vol 77, No. 4, p. 541-580.

Nater, J.P. (1983). De Uiver, glorie en tragiek in de Melbourne race, hoogtepunten uit de luchtvaarthistorie in de jaren dertig. Donker, Rotterdam.

NEA/CSNI. (1999). Identification and assessment of organisational factors related to the safety of NPPs, State-of-the-art report, Volume 1, NEA/CSNI/R(99)21/VOL1, Nuclear Energy Agency, Committee on the Safety of Nuclear Installations, Le Seine St-Germain, France.

Netherlands Aviation Safety Board. (1994). Aircraft Accident Report 92-11, El Al Flight 1862, Boeing 747-258F, 4X-AXG, Bijlmermeer, Amsterdam, October 4, 1992, SDU Uitgeverij

Netherlands Aviation Safety Board. (1996). Report of the views of the Netherlands Aviation Safety Board on the accident to Palair Flight PMK301, Fokker 100, PH-KXL, Skopje, Republic of Macedonia, March 5, 1993, Aircraft Accident Report 93-01.

Nijenhuis, W.A.S., Spek, F., Moeleker, P. (1996). De Lambach HL II. De geschiedenis van het origineel en de bouw van de replica, Barjesteh van Waalwijk van Doorn & Co's Uitgeversmaatschappij, Rotterdam, the Netherlands.

Norris, G. (1998). 757-300 prepared for flight test, Flight International, 10 June 1998.

NRC. (1975). Reactor Safety Study, WASH-1400, NUREG –751014, United States Nuclear Regulatory Commission, Washington D.C., USA.

NRC. (1983). PRA Procedures guide, a guide to the performance of probabilistic risk assessments for nuclear power plants, Final Report, Vol. 1-2, NUREG/CR-2300, United States Nuclear Regulatory Commission, Washington D.C., USA.

NRC. (1995). Use of probabilistic risk assessment methods in nuclear regulatory activities, final policy statement, Federal register, Vol. 60, No. 158.

NRC. (1998). White Paper on Risk-Informed and Performance-Based Regulation, SECY-98-144, United States Nuclear Regulatory Commission, Washington D.C., USA.

NRC. (2007a). Fact Sheet, Three Mile Island accident, United States Nuclear Regulatory Commission, Washington, D.C., USA, http://www.nrc.gov/reading-rm/doc-collections/fact-sheets/3mile-isle.html,

NRC. (2007b). ATHEANA user's guide, NUREG-1880, United States Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, Washington, D.C., USA.

NTSB. (1970). Aircraft Accident Report: Ozark Airlines, Douglas DC-9-15, N974Z, Sioux City Airport, Sioux City, Iowa, December 27, 1968, AAR-70-20, National Transportation Safety Board, Washington D.C., USA.

NTSB. (1979). Aircraft Accident Report: American Airlines Inc, DC-10-10, N110AA, Chicago-O'Hare International Airport, Chicago, Illinois, May 25, 1979, AAR-79-17, National Transportation Safety Board, Washington D.C., USA.

NTSB. (1982). Aircraft Accident Report: Air Florida, Inc., Boeing 737-222, N62AF, Collision with 14th Street Bridge, near Washington Nat'l Airport, Washington, DC, January 13, 1982. AAR-82/08, National Transportation Safety Board, Washington. D.C., USA.

NTSB. (1987). Aircraft Accident/Incident Summary Report, NTSB/AAR-87/02/SUM, National Transportation Safety Board, Washington D.C., USA.

NTSB. (1988). Aircraft Accident Report: Continental Airlines, Inc., Flight 1713, McDonnell Douglas DC-9-14, N626TX, Stapleton International Airport, Denver, Colorado, November 15, 1987, AAR-88/09, National Transportation Safety Board, Washington D.C., USA.

NTSB. (1989). Aircraft Accident Report: Aloha Airlines Flight 243, Boeing 737-200, N73711, near Maui, Hawaii, April 28, 1988, AAR-89/03, National Transportation Safety Board, Washington D.C., USA.

NTSB. (1990). Aircraft Accident Report: United Airlines Flight 232, McDonnell Douglas DC-10-40, Sioux Gateway Airport, Sioux City, Iowa, July 19, 1989, AAR-90/06, National Transportation Safety Board, Washington D.C., USA.

NTSB. (1991). Aircraft Accident Report: Ryan International Airlines DC-9-15, N565PC, Loss of control on takeoff, Cleveland-Hopkins International Airport, Cleveland, Ohio, February 17, 1991, NTSB/AAR-91/09, National Transportation Safety Board, Washington D.C., USA.

NTSB. (1993). Aircraft Accident Report: Take-off stall in icing conditions, US Air Flight 405, Fokker F-28, N485US, LaGuardia Airport, Flushing, New York, March 22, 1992, AAR-93/02, National Transportation Safety Board, Washington D.C., USA.

NTSB. (1997). Aircraft Accident Report: In-Flight Fire and Impact with Terrain, Valujet Airlines Flight 592, DC-9-32, N904VJ, Everglades, Near Miami, Florida, May 11, 1996. AAR-97/06, National Transportation Safety Board, Washington D.Cc, USA.

NTSB. (2000). Aircraft Accident Report: In-flight breakup over the Atlantic Ocean Trans World Airlines Flight 800, Boeing 747-141, N93119, near East Moriches, New York July 17, 1996, AAR-00/03, National Transportation Safety Board, Washington D.C., USA.

NTSB. (2001). Aircraft Accident Report: Runway overrun during landing, American Airlines Flight 1420, McDonnell Douglas MD-82, N215AA, Little Rock, Arkansas, June 1, 1999, AAR-01/02, National Transportation Safety Board, Washington D.C., USA.

NTSB. (2002a). Aircraft Accident Report: Loss of control and impact with Pacific Ocean, Alaska Airlines flight 261, McDonnell-Douglas MD-83, N963AS, about 2.7 miles north of Anacapa Island, California, January 31, 2000, AAR-02/01, National Transportation Safety Board, Washington D.C., USA.

NTSB. (2002b). Aircraft Accident Brief: Southwest Airlines Flight 1455, Boeing 737-300, N668SW, Burbank, California, March 5, 2000, AAB-02/04, National Transportation Safety Board, Washington D.C., USA.

NTSB. (2007). Aircraft Accident Report: Runway overrun and collision, Southwest Airlines Flight 1248, Boeing 737-7H4, N471WN, Chicago Midway International Airport, Chicago, Illinois, December 8, 2005, AAR-07/06, National Transportation Safety Board, Washington D.C., USA.

O'Banion, K. (1980). Public reaction to imposed risk, paper presented at the US Environmental Protection Agency Workshop on Environmental Risk Assessment of Synfuels, Alexandria, Virginia, USA.

O'Conner, S., Reynolds, T. (1996). A review of aircraft maintenance and dispatch, ADAMS working document ADAMS-DRA-WD.01.

O'Hare, D. (2006). Cognitive functions and performance shaping factors in aviation accidents and incidents, The International Journal of Aviation Psychology, Vol. 16, nr. 2, p 145-156.

Onisko, A., Druzdzel M.J., Wasyluk, H. (1998). A probabilistic causal model for diagnosis of liver disorders. In Proceedings of the Seventh Symposium on Intelligent Information Systems (IIS-98), p. 379-387, Malbork, Poland.

Onisko, A., Druzdel, M.J., Wasyluk, H. (1999). A Bayesian network model for diagnosis of liver disorders, Research Report CBMI-99-27, Center for Biomedical Informatics, University of Pittsburgh.

Owens, D.K., Shachter, R.D., Nease, R.F. (1997). Representation and analysis of medical decision problems with influence diagrams, Journal of Medical Decision Making, Vol 17, No. 3, p. 241-262.

Papazoglou I.A., Aneziris O.N. (1999). On the quantification of the effects of organisational and management factors in chemical installations. Reliability Engineering and System Safety 63, p. 33-45.

Papazoglou, I.A., Bellamy, L.J., Hale, A.R., Aneziris, O.N., Ale, B.J.M., Post, J.G., Oh, J.I.H. (2003). I-Risk, development of an integrated technical and management risk methodology for chemical installations, Journal of Loss Prevention in the Process Industries, Vol. 16, p. 575-591.

Paté-Cornell, E., Dillon, R. (2001). Probabilistic risk analysis for the NASA Space Shuttle: a brief history and current work, Reliability Engineering and System Safety, Volume 74, p. 345-352.

Pearl. J. (2000a). Causality, Cambridge University Press, New York.

Pearl. J. (2000b). The logic of counterfactuals in causal inference, Journal of the American Statistical Association, Vol. 95, No. 450, p. 428-435.

Pearl, J. (2003). Statistics and causal inference: A review, Sociedad de Estadística e Investigatión Operativa, Test, Vol. 12, p. 281-345.

Perrin, E., Kirwan, B., Stroup, R. (2006). A systematic model of ATM safety: the Integrated Risk Picture, paper presented at the Conference on Risk Analysis and Safety Performance in Aviation, September 19-21, 2006, Atlantic City, New Jersey.

Perrow, C. (1984). Normal accidents: Living with high-risk technology, Basic Books Inc, New York.

Piers, M.A., Ale, B.J.M. (2000). Policy Options for dealing with societal risk around Schiphol, NLR-CR-2000-084, RIVM nr. 610066011, NLR Amsterdam.

Piers, R., Lebouille, R., Roelen, A., Smeltink, J. (2006). Safety has value! An approach for the assessment of the costs and benefits of safety measures, Second International Conference on research in Air Transport (ICRAT), Belgrade, June 24-28 2006, ICRAT Conference proceedings p. 351-355.

Plas, M, van der, Luijk, C.M. van. (2005). Performance Luchtverkeersleiding schiphol, Beoordeling van het VEM Raamwerk, report 6200030001/2005, RIVM, Bilthoven, The Netherlands.

Poot, M.J.M., Lensink, R., Brandjes, J., Dirksen, S., Buurma, L.S. (2000). Spatial patterns of bird movements on and around an airport, a case study on Eindhoven Airport 1998-99. In J. van Nugteren (ed), Proceedings of the 25th meeting of the International Bird Strike Committee, IBSC, Amsterdam, p. 175-186.

Prigogine, A. (1977). Time, structure and fluctuations, Nobel lecture, 8 December 1977.

Rasmussen, J. (1983). Skills, rules, and knowledge: Signals, signs, and symbols, and other distinctions in human performance model, IEEE Transactions on Systems, Man, and Cybernetics, SMC-13(3), p. 257-266.

Rasmussen, J. (1997). Risk management in a dynamic society, a modelling problem, Safety Science Vol. 27, No. 2/3, p. 183-213.

Reason, J. (1990). Human Error, Cambridge University Press, New York.

Reason, J. (1997). Maintenance related errors: The biggest threat to aviation safety after gravity?, in H. Soekkha (ed.) 'Aviation Safety , Proceedings of the IASC-97 International Aviation Safety Conference VSP publishing, Utrecht, The Netherlands.

Reason, J., Hollnagel, E., Paries, J. (2006). Revisiting the 'Swiss cheese' model of accidents, EEC note no. 13/06, Eurocontrol Experimental Centre, Brétigny-sur-Orge, France.

Reich, P.G. (1966a). Analysis of long range air traffic systems, separation standards - I, Journal of the Institute of Navigation, Vol. 19, Nr. 1, p. 88-89.

Reich, P.G.(1966b). Analysis of long range air traffic systems, separation standards - II, Journal of the Institute of Navigation, Vol. 19, Nr. 2, p. 169-186.

Reich, P.G. (1966c). Analysis of long range air traffic systems, separation standards - III, Journal of the Institute of Navigation, Vol. 19, Nr. 3, p. 331-347.

RGCSP. (1995). A Review of Work on Deriving a Target Level of Safety (TLS) for En-route Collision Risk, Review of the General Concept on Separation Panel Working Group A, Brussels, 1st – 12th May, 1995 RGCSP-WG/A-WP/8.

RLD. (1996). Safety Policy Report Civil Aviation, Ministry of Transport, Public Works and Watermanagement, Directorate General of Civil Aviation, The Hague.

RMI. (2003). Regeling Milieu-informatie Luchthaven Schiphol, Staatscourant 18 februari 2003, nr. 34, page 15.

Roelen, A.L.C., Lynch, R.E., Statler, I.C. (1998). KLM Royal Dutch Airlines user needs study for the Aviation Performance Measurement System, NLR CR 97545, National Aerospace Laboratory NLR, Amsterdam.

Roelen, A.L.C., Bellamy, L.J., Hale, A.R., Molemaker, R.J., Paassen, M.M. van. (2000a). Feasibility of the development of a causal model for the assessment of third party risk

around airports, Part 1: Main report, NLR-CR-2000-189-PT-1, National Aerospace Laboratory NLR, Amsterdam.

Roelen, A.L.C., Bellamy, L.J., Hale, A.R., Molemaker, R.J., Paassen, M.M. van. (2000b). Feasibility of the development of a causal model for the assessment of third party risk around airports, Part 3: Modelling methodologies/techniques, NLR-CR-2000-189-PT-3, National Aerospace Laboratory NLR, Amsterdam.

Roelen, A.L.C., Pikaar, A.J., Ovaa, W. (2000c). An analysis of the safety performance of air cargo operators, NLR-TP-2000-210, National Aerospace Laboratory NLR, Amsterdam.

Roelen, A.L.C., Wever, R., Cooke, R.M., Lopuhaä, R., Hale, A.R., Goossens, L.H.J., Simons, M., Valk, P.J.L. (2002). Causal modelling of air safety, Demonstration model, NLR-CR-2002-662, National Aerospace Laboratory NLR, Amsterdam.

Roelen, A.L.C., Wever, R., Verbeek, M.J. (2003). Improving aviation safety by better understanding and handling of interfaces, A pilot study, NLR-CR-2004-025, National Aerospace Laboratory NLR, Amsterdam.

Roelen, A.L.C. (2004). A description of main safety affecting interface problems between different aviation disciplines, NLR-CR-2004-398, National Aerospace Laboratory NLR, Amsterdam.

Roelen, A.L.C., Wever, R. (2004). A causal model of a rejected take-off, NLR-CR-2004-039, National Aerospace Laboratory NLR, Amsterdam.

Roelen, A.L.C., Cooke, R.M., Goossens L.H.J. (2004a). An assessment of the validity of expert judgement techniques and their application at Air Traffic Control the Netherlands, NLR-CR-2004-360, National Aerospace Laboratory NLR, Amsterdam.

Roelen, A.L.C., Wever, R., Cooke, R.M., Lopuhaä, R., Hale, A.R., Goossens, L.H.J., Simons, M., Valk, P.J.L. 2004b. Causal modeling using Bayesian belief nets for integrated safety at airports, Risk Decision and Policy, Vol 9, number 3, p. 207-222.

Roelen, A.L.C., Wever, R. (2005a). Accident scenarios for an integrated aviation safety model, NLR-CR-2005-560, National Aerospace Laboratory NLR, Amsterdam.

Roelen, A.L.C., Wever, R. (2005b). An analysis of flight crew response to system failures, paper presented at the 2005 seminar of the International Society of Air Safety Investigators, ISASI 2005, September 12-17, Fort Worth, Texas, USA.

Roelen, A.L.C., Bos, T.J.J., Modiri, B. (2006a). Aircraft maintenance program development needs improvement; An analysis of the maintenance program development process and its effect on safety for large transport aircraft in commercial service, NLR-CR-2006-351, National Aerospace Laboratory NLR, Amsterdam.

Roelen, A.L.C., Van Doorn, B.A., Smeltink, J.W., Verbeek, M.J., Wever, R. (2006b). Quantification of Event Sequence Diagrams for a causal risk model of commercial air transport, NLR-CR-2006-520, National Aerospace Laboratory NLR, Amsterdam.

Roelen, A.L.C., Van Baren, G.B., Smeltink, J.W., Lin, P.H., Morales, O. (2007). A generic flight crew performance model for application in a causal model of air transport, NLR-CR-2007-562, NLR Amsterdam.

Roelen, A.L.C. (2008). Presentation for CATS group of experts, 14 January 2008, Victoria Hotel, Amsterdam.

Roelen, A.L.C., Wever, R, Mosleh, A, Growth, K. (2008). Development and validation of a comprehensive hybrid causal model for safety assessment and management of aviation systems, ninth International Probabilistic Safety Assessment and Management Conference, PSAM 9, Hong Kong, China.

Rogers, W.P. (1986). Report of the Presidential Commission on the Space Shuttle Challenger Accident.

Rotundo, L.C. (1994). Into the unknown: the X-1 Story, Smithsonian Institution Press, Washington D.C., USA.

Rouvroye, J.L., Bliek, E.G. van den. (2002). Comparing safety analysis techniques, Reliability Engineering and System Safety, Vol. 75, No. 3, p. 289-294

SAE. (2003). Safety Assessment of Transport Airplanes in Commercial Service, SAE ARP 5150, proposed draft 2003-03-04, Society of Automotive Engineers, Aerospace Standards Development and Research Division, Warrendale, PA, USA.

Sarter, N.B., Woods, D.D. (1995). “How the world did we ever get into that mode?” Mode error and awareness in supervisory control, Human Factors, 37 (1), p. 5-19.

Schatzberg, E. (1994). Ideology and technical choice: the decline of the wooden airplane in the United States, 1920-1945. Technology and Culture, Vol. 35, No. 1, p. 34-69.

Schiphol Group. (2007). Facts and Figures 2006, published by Schiphol Group Corporate Communications, Schiphol, the Netherlands.

Shayler, D.J. (2000). Disasters and accidents in manned spaceflight, Springer-Praxis, Chichester, UK.

Simons, M., Valk, P.J.L. (1993). Review of human factors problems related to long distance and long endurance operation of aircraft. NATO-AGARD CP-547: Recent Advances in Long Range and Long Endurance Operation of Aircraft. Neuilly sur Seine: NATO-AGARD. p. 15/1-15/9.

Simons, M., Valk, P.J.L., de Ree, J.J.D., Veldhuijzen van Zanten, O.B.A., D'Huyvetter, K. (1994). Quantity and quality of onboard and layover sleep: effects on crew performance and alertness. Report RD-31-94. Netherlands Aerospace Medical Centre, Soesterberg.

Simons, M., Valk, P.J.L. (1997). Effects of a Controlled Rest on the Flight Deck on Crew Performance and Alertness. Report: NLRGC 1997-B3. Netherlands Aerospace Medical Centre, Soesterberg.

Simons, M., Valk, P.J.L. (1998). Early starts: effects on sleep, alertness and vigilance. AGARD-CP-599; NATO-AGARD, Neuilly-sur-Seine, France. p. 6/1-6/5.

Sinha, N.K. (2002). Environmental and runway surface conditions during friction tests at North Bay Airport: January - February 2002, TP 14158E, Transport Canada, Montreal, Canada.

Slovic, P., Fischhoff, B., Lichtenstein, S. (1976). Cognitive processes and societal risk taking. In J. S. Carroll and J. W. Payne (Eds.), Cognition and social behavior, p. 165-184, Erlbaum, Potomac, MD, USA.

Smith, N.M., Lee, P.U., Prevot, T., Mercer, J., Palmer, E.A., Battiste, V., Johnson, W. (2004). A Human-in-the-loop Evaluation of Air-Ground Trajectory Negotiation. AIAA 4th Aviation Technology, Integration, and Operations Forum, Chicago, Illinois, USA.

SP. (1999). SP wil opheldering over laagvliegend toestel boven woonwijk, Nieuwsberichten 09-08-1999, www.sp.nl.

Stallen, P.J., Hudson, P. (2006). Publieke communicatie over luchtvaartveiligheid, VACS/06.01.04.

Stamatelatos, M. (2002). Probabilistic risk assessment procedures guide for NASA managers and practitioners, Version 1.1.

Starmer, G.A., Mascord, D.J., Tattam, B., Vine, J.H., Watson, T.R. (1994). The effects of single acute therapeutic doses of Dexchlorpheniramine, alone and in combination with alcohol, on human performance in driving related tests: Exploration of the relationships between performance impairment and blood concentrations of Dexchlorpheniramine and Alcohol, Consultant Report CR 1/94, Roads and Traffic Authority, Road Safety Bureau, Australia.

State of California. (2002). California Airport Land Use Planning Handbook, 2002. State of California, Department of Transportation, Division of Aeronautics.

Stephenson A.G. (chairman), (1999). Mars Climate Orbiter Mishap Investigation Board Phase 1 Report. National Aeronautics and Space Administration, Washington D.C., USA.

Stevens, S.S. (1946). On the theory of scales of measurement, Science, Vol. 103, No. 2684, p. 677-680.

Stroeve, S.H., Bakker, G.J., Blom, H.A.P. (2007). Safety risk analysis of runway incursion alert systems in the tower and cockpit by multi-agent systemic accident modelling, ATM 2007, 7th USA/Europe ATM R&D seminar, Barcelona, Spain.

Strotz, R.H., Wold, H.O.A. (1960). Causal models in the social sciences. Econometrica, 28, p 417-427.

Sutter, J., Spenser, J. (2006). 747: Creating the World's first Jumbo Jet and other adventures from a life in aviation. Smithsonian Books, Washington D.C., USA.

Svedung, I., Rasmussen, J. (2002). Graphic representation of accident scenarios: mapping system structure and the causation of accidents. Safety Science, Vol. 40, p. 397-417.

Swain, A.D., Guttmann, H.E. (1983). Handbook of human reliability analysis with emphasis on nuclear power plant applications, Final report, NUREG/CR-1278-F, SAND80-0200, Sandia National Laboratories, Albuquerque, NM, USA.

Swaminathan, S., Smidts, C. (1999). The event sequence diagram framework for dynamic probabilistic risk assessment, Reliability Engineering and System Safety, 63, p. 73-90.

Swiss Re. (1996). High volatility in aviation insurance: are premium rates due for a nosedive?, Sigma, No. 1/1996, Swiss Reinsurance Company, Zurich, Switzerland.

Taylor, B.N. (Ed.). (1995). Guide for the Use of the International System of Units (SI), National Institute of Standards and Technology (NIST) Special Publication 811, 1995 Edition, U.S. Government Printing Office, Washington D.C., USA.

Taylor. B.N. (Ed.). (2002). The International System of Units (SI), National Institute of Standards and Technology (NIST) Special Publication 330, 2001 Edition, U.S. Government Printing Office, Washington D.C., USA.

Thompson, M.O. (2000). Flight research: problems encountered and what they should teach us. Monographs in Aerospace History, number 22, NASA SP-2000-4522, NASA, Washington D.C., USA.

Tiemeyer, B. (2003). Reduced vertical separation minimum (RVSM): pre- and post-implementation safety cases, in Bedford, T and Van Gelder, P.H.A.J.M (eds), Safety and Reliability, A.A. Balkema publishers, Lisse, the Netherlands.

Transport Canada. (2004). Overview of the Joint Winter Runway Friction Measurement Program, TP 13361, Transport Canada, Montreal, Canada.

TSB. (2007). Aviation Investigation Report A05H0002, Runway Overrun And Fire, Air France Airbus A340-313 F-GLZQ, Toronto/Lester B. Pearson International,Airport, Ontario, 02 August 2005 Transportation Safety Board of Canada, Gatineau, Quebec, Canada.

Tweede Kamer. (1983). Integrale Nota LPG, Tweede Kamer der Staten General, vergaderjaar 1983-1984, 18233 nrs 1-2, SDU, Den Haag, the Netherlands.

Tweede Kamer. (2001). Planologische Kernbeslissing Vijfde Nota ruimtelijke ordening, deel 3: kabinetsstandpunt, Tweede Kamer der Staten-Generaal, vergaderjaar 2001-2002, 27 578, nr 5, SDU, Den Haag, the Netherlands.

Tweede Kamer. (2003a). Vaststelling van de begrotingsstaat van het Ministerie van Verkeer en Waterstaat (XII) voor het jaar 2004, Verslag houdende een lijst van vragen en antwoorden, Tweede Kamer, vergaderjaar 2003-2004, 29 200 XII, nr. 10, SDU, Den Haag, the Netherlands.

Tweede Kamer. (2003b). Toekomst van de nationale luchthaven, lijst van vragen en antwoorden, Tweede Kamer der Staten-Generaal, vergaderjaar 2003-2004, 26959, nr. 54, SDU, Den Haag, the Netherlands.

Tweede Kamer. (2003c). Motie van de leden Samsom en Duyvendak, voorgesteld 25 november 2003, Tweede Kamer der Staten-Generaal, vergaderjaar 2003-2004, 26959, nr. 55, SDU, Den Haag, the Netherlands.

Tweede Kamer. (2005). Vragen van de leden Koopmans en Haverkamp (beiden CDA) aan de staatssecretaris van Verkeer en Waterstaat over een incident met een vliegtuig van Onur Air. (Ingezonden 12 mei 2005); Antwoord. Aanhangsel van de handelingen, Tweede Kamer der Staten-Generaal, Vergaderjaar 2004-2005, 23125, nr. 1749, SDU, Den Haag, the Netherlands.

UKOOA. (1999). A framework for risk-related decision support. UK Offshore Operators Association, London, UK.

USAir. (1997). Fokker 100 Pilot's Handbook. US Airways, Flight Publications Department Pittsburgh, PA, USA.

VACS. (2003a). Inhoudelijke rapportage projectgroep, VACS/03.04.21a.

VACS. (2003b). Advies Causale modellering en groepsrisico, VACS/03.08.33.

Valk, P.J.L., Simons, M. (1994). Aircrew and Hypnotics: Residual effects of temazepam and brotizolam on performance. Report NLRGC 1994-K8. Netherlands Aerospace Medical Centre, Soesterberg.

Valk, P.J.L., Simons, M. (1996). Effects of early reporting times and irregular work schedules on sleep, alertness, and performance of pilots engaged in short-haul operations Report: NLRGC 1996-B2. Netherlands Aerospace Medical Centre, Soesterberg.

Valk, P.J.L., Simons, M., Struyvenberg, P.A.A., Kruit, J., Van Berge Henegouwen, M. (1997). Effects of a single dose of loratadine on flying ability under conditions of simulated cabin pressure. American Journal of Rhinology, 11(1), p. 27-33.

Valk, P.J.L., Simons, M. (1998). Pros and cons of strategic napping on long haul flights. AGARD-CP-599; NATO-AGARD, Neuilly-sur-Seine, France, p. 5/1-5/5.

Valk, P.J.L., Roon, D.B. van, Simons, M. (1999). Effects of an alcohol hangover on performance and alertness. Report 1999-B2. Netherlands Aeromedical Institute, Soesterberg, The Netherlands.

Vesely, W.E., Goldberg, F.F., Roberts, N.H., Haasl, D.F. (1981). Fault Tree Handbook, NUREG-0492, U.S. Nuclear Regulatory Commission, Washington D.C., USA.

Visser, H.C. (1997). If I were a rich man ... my accident record wouldn't be so bad!, in H. Soekkha (ed.), 'Aviation Safety , Proceedings of the IASC-97 International Aviation Safety Conference VSP publishing, Utrecht, the Netherlands.

Von Thaden, T.L., Wiegmann, D.A., Mitchell, A.A., Sharma, G., Zhang, H. (2003). Safety culture in a regional airline, results from an aviation safety survey, 12th international symposium on Aviation Psychology, Dayton, Ohio, USA.

Vos, P.R. (1996). The development of a relational database, based on airline timetables, to solve the denominator problem for aviation safety, Memorandum M-793, Faculty of Aerospace Engineering, Delft University of Technology, Delft.

Waveren, R.H. van, Groot, S., Scholten, H., Geer, F.H. van, Wösten, J.H.M., Kroeze, R.D., Noort, J.J. (1999). Good modelling practice handbook, STOWA report 99-05, Dutch Department of Public Works, Institute for inland water management and waste water treatment, Lelystad, the Netherlands.

Wertheimer, N., Leeper, E. (1979). Electrical wiring configurations and childhood cancer, American Journal of Epidemiology, nr. 109, p. 273-284.

Wever, R., Karwal, A.K., Smit, H.H. (2006). The Lugano Aeronautical Study, Safety assessment of the offset steep approach procedures, NLR-CR-2006-038, National Aerospace Laboratory NLR, Amsterdam.

Wewerinke, P.H. (1989). Models of the human observer and controller of a dynamic system, Ph.D. Thesis, Delft University of Technology.

Wiegman, D., Shappell, S. (2001). A human error analysis of commercial aviation accidents using the Human Factors Analysis and Classification System (HFACS). DOT/FAA/AM-01/3, Federal Aviation Administration, Washington D.C., USA.

Wolfe, T. (1979). The right stuff. Farrar, Straus and Giroux, New York, USA.

Wright, L.B. (2006). A PROMISE for improvement, the ProRail management information for safety and environment database, paper presented at the 3rd International Conference 'Working on Safety', 12-15 September 2006, Zeewolde, the Netherlands.

Yeager, C.E., Janos, L. (1985). Yeager, Bantam Books, New York, USA.

Zaidel, D.W., Cohen., J.A. (2005). The face, beauty and symmetry: perceiving asymmetry in beautiful faces International Journal of Neuroscience, Nr 155, p. 1165-1173.

Summary

Aviation safety is so well developed that individual organisations cannot rely on the number of accidents as useful indicators of the safety level of their operation. Adequate control of risks requires the availability of a method to determine the level of safety as a function of the current status and of proposed or expected changes to the aviation system. Aviation safety policy plans have therefore proposed the development of causal risk models. Unfortunately however, they failed to specify or even describe such models other than in the most general of terms. Causal model development was stated as a goal in itself, without consideration of how such a model should be used. The objective of this thesis is to clarify these issues by comparing user requirements with the performance that can be delivered by various modelling techniques. The thesis answers the question what causal risk modelling adds to current safety management approaches and what the criteria are for ensuring it makes a successful contribution to safety. Experience gained in several causal model development projects (particularly for the Federal Aviation Administration and the Dutch Ministry of Transport and Water Management) are used to illustrate how a causal model should and should not be developed and used.

Chapter 2 describes what we mean with the term ‘safety’ and in what ways it can be expressed. The most relevant theories on accident causation are described, emphasizing the role of ‘normal’ occurrences in the accident sequence. These normal occurrences should therefore be correctly represented in a causal model. The influence of human behaviour and of organisations is also important and should be represented adequately.

Chapter 3 delves deeper into causality and the difference between causality and association. From this it follows that cause-effect relations can only be substantiated from specific knowledge on the mechanisms of the system. This means that substantial knowledge of the aviation system is a prerequisite for the development of a causal risk model for air transport. To avoid confusion between singular and generic causal relations, a causal risk model should represent success scenarios as well as failure scenarios. Causal chains never really end, in practice a useful criterion for the extent of a causal risk model is that it will have to include decisions and actions up to the highest managerial level of the actors that are directly involved.

Requirements from potential users of a causal risk model for air transport are the subject of Chapter 4. A short overview of the history of aviation safety shows that the current level of safety is so high that the mechanisms that previously resulted in continuous safety improvement do not work anymore. Because each of the potential users of a causal risk model (airlines, air navigation service providers, airports organisations, maintenance and repair stations, aircraft manufacturers and aviation authorities) can influence safety in a different way and each has different drivers for improving safety, some of the user requirements are not compatible. The most important general requirements are that the model should represent to complete air transport system including dependencies, that the model should properly represent the influence of the human operator and of organisations, that the model is validated and produces quantitative and reproducible results, that the

model representation is transparent and that the model is able to represent current as well as future accident scenarios.

Chapter 5 demonstrates with some practical examples how safety analyses are currently conducted. The approach followed in each example is compared to the user requirements that were derived in Chapter 4. This provides an even better insight into the performance that is required for a causal model in order to have added benefits compared to current approaches. From this chapter it follows that currently there is no standardised procedure for conducting safety analyses, each time it is performed differently. Available results from previous analyses are not fully utilised and results rely heavily on expert judgement. None of the examples fulfils all the user requirements that have been derived in Chapter 4. An effort to develop a method that meets these criteria therefore seems fully justified.

Causal risk modelling is applied for management of safety in industries other than air transport. Chapter 6 gives an overview of how causal risk models are being applied in the nuclear power industry, manned spaceflight, offshore industry, process industry, rail transport and health care and it discusses the lessons than can be learned with respect to development and application of similar models in air transport. Initially there was resistance against the application of risk models in the nuclear power industry, manned spaceflight, offshore industry and process industry, but catastrophic accidents led to regulatory changes that forced the industry to use probabilistic risk methods. Development of models and methods that were considered acceptable required a lot of time. Based on these experiences it can be expected that the introduction of causal risk model in air transport will be similarly jerkily and it could take decades before the industry has fully accepted the use of quantitative risk models. It is essential to state from the outset that the models should be used as a complement rather than as a replacement of current practice.

Several techniques are available for model representation and calculations. The most commonly applied techniques (fault trees, event trees, Bayesian belief nets and Petri-nets) and their characteristics are described in Chapter 7. Fault trees and event trees have the important advantage of simplicity. This makes them transparent but the areas of application are limited. Bayesian belief nets are better suited for the representation of ‘soft’ influences, but as a drawback are less transparent. Petri-nets can also represent dynamic situations but are even less transparent. The conclusion of this chapter is that a combination of techniques should be applied, e.g. with fault trees and event trees for the representation of the uppermost level of the model (the part which is directly visible to the user), while Bayesian belief nets and perhaps also Petri-nets should be applied, where necessary, for the correct representation of details.

The ability to provide quantitative results is one of the requirements of a causal risk model. Quantification is the subject of Chapter 8. One of the problems in modelling is finding the right units, specifically if the influence of complex factors like safety culture, safety management or non-technical pilot skills must be expressed. The credibility of the model depends on the model structure and the sources used for quantification. Important sources of information are accident investigation reports and incident reporting systems of airlines. Results from empirical studies can sometimes also be used. If correctly elicited, expert opinion can offer a quite acceptable accuracy of quantification. In addition to information on accidents and incidents, information on normal operations is required to determine the relative influence of causal factors. Flight data recorders can be an important source of this information. To use the information from existing data bases as efficiently as possible, it is important that definitions and descriptions of model elements match with those used in

existing data bases. Preferably the model should conform to the definitions used in ECCAIRS, the European standard for incident reporting, but this also requires that some of the current problems of ECCAIRS are resolved.

The biggest difficulties are to be expected in representing the influences of the human operator, management and the organisation, and the many dependencies between model elements. These problems and possible solutions are discussed in Chapter 9. Solutions are the application of Bayesian belief nets for representing 'soft' influences and dependencies and also the definition of a limited number (33) of archetype accidents. These archetype accidents can be the backbone of a causal risk model. The human operator can be represented in the model by means of a Bayesian belief net, while the influence of management on human performance can be represented by means of generic 'delivery systems'.

Chapter 10 addressed the possibilities for validating a causal risk model. Full validation is not possible due to the complexity of the air transport system, but limited validation of model components is feasible. Some examples are provided in the chapter.

Conclusions of this thesis, presented in Chapter 11, are that current drivers for safety improvement fall short to bring about further safety improvements. A method for safety assessment and safety analysis is required that is not solely based on accident frequencies, provides a quantified accident probability and is able to properly represent the complex arrangements within the air transport system, including human and organisational influences. The method should be transparent and yield reproducible and demonstrable valid results. Using a causal risk model as described in this thesis is able to meet those requirements and therefore is a method to identify safety improvement opportunities that will not be obtained with the current methods. Developing a causal risk model for aviation is not overly difficult; it is just a lot of work. And the majority of this work will have to be conducted by people with sufficient background knowledge of air transport. Representing the complexity of the air transport system in a causal risk model requires numerous assumptions, each of which can be challenged. The causal risk model can only be a success if these assumptions are agreed upon by all parties involved, and the air transport regulator is the appropriate lead in this process. Preferably this should be an international regulator like EASA or ICAO. They should initiate and lead the process to come to an industry standard model and to agree on how and when to apply this model. For further development of a causal risk model of air transport, it is essential to link the model directly to an accident and incident data system, preferably ECCAIRS. Feedback from the modellers to the potential users and vice versa is required to get a better grip on user requirements. The modellers should indicate possibilities and limitations of the model, and should come-up with real-life examples of cases in which a causal risk model would be a helpful tool. This will have to be an iterative process. Support from the industry is essential to provide data and insight, to the model developers, into operational aviation processes.

Samenvatting

Luchtvaartveiligheid is zo goed ontwikkeld dat voor individuele organisaties het aantal ongevallen geen goede indicator is voor de veiligheid. Voor adequate risicobeheersing is een methode vereist waarmee het veiligheidsniveau als functie van de huidige status en van voorgestelde of te verwachten veranderingen van het luchtvaartsysteem kan worden vastgesteld. In plannen voor luchtvaartveiligheidsbeleid is daarom de ontwikkeling van causale risicomodellen voorgesteld. Helaas worden deze modellen in de beleidsplannen niet verder gespecificeerd of beschreven anders dan in zeer algemene termen. Ontwikkeling van een causaal model is als een doel op zich gesteld, zonder aandacht voor het gebruik van zulke modellen. Het doel van dit proefschrift is hier helderheid in te verschaffen door gebruikerseisen te vergelijken met de prestaties die met verschillende modelleertechnieken kunnen worden bereikt. Het proefschrift beantwoordt de vraag wat causale risicomodellen kunnen toevoegen aan de huidige benadering voor veiligheidsmanagement and wat de criteria zijn om te verzekeren dat de bijdrage van causale modellen aan veiligheid succesvol is. Ervaringen opgedaan in verschillende projecten voor de ontwikkeling van een causaal risicomodel (in het bijzonder voor de Federal Aviation Administration en het Nederlandse Ministerie van Verkeer en Waterstaat) worden gebruikt ter illustratie van hoe een causaal model wel of niet ontwikkeld en toegepast dient te worden.

Hoofdstuk 2 beschrijft wat we verstaan onder het begrip ‘veiligheid’ en op welke manier dit kan worden uitgedrukt. Er wordt ingegaan op gangbare theorieën over de manier waarop risico tot stand komt. Hierin wordt de rol van ‘normale’ gebeurtenissen bij het tot stand komen van ongevallen benadrukt, en dat moet derhalve correct kunnen worden gerepresenteerd in een causaal model. Ook de invloed van menselijk gedrag en de rol van organisaties is belangrijk en moet correct worden gerepresenteerd.

In hoofdstuk 3 wordt dieper ingegaan op het begrip causaliteit en het verschil tussen causaliteit en associatie. Hieruit blijkt dat oorzaak-gevolg relaties alleen kunnen worden opgesteld op basis van specifieke kennis aangaande de werking van het systeem. Dit betekent dat substantiële kennis van het luchtvaartsysteem een vereiste is voor de ontwikkeling van een causaal risicomodel voor luchttransport. Om mogelijke verwarring tussen singuliere en generieke oorzaak-gevolg relaties te voorkomen is het belangrijk dat het model in staat is zowel scenarios van falen als van herstel en succes te representeren. Het begin en einde van oorzaak-gevolg relaties is arbitrair, een praktisch criterium voor een causaal model is dat het model de effecten van beslissingen en handelingen van betrokkenen op het hoogste management niveau moet kunnen representeren.

Eisen die mogelijke gebruikers stellen aan een causaal risicomodel voor luchttransport worden behandeld in hoofdstuk 4. Een kort overzicht van de geschiedenis van luchtvaartveiligheid laat zien dat de veiligheid op dit moment zo groot is dat de mechanismen die voorheen tot continue veiligheidsverbeteringen hebben geleid nu niet meer werken. Omdat elk van de mogelijke gebruikers van een causaal risicomodel (luchtvaartmaatschappijen, luchtverkeersleidingsorganisaties, luchthaven organisaties, onderhoudsorganisaties, vliegtuigfabrikanten en luchtvaartautoriteiten) elk op een andere manier invloed hebben op de veiligheid en elk verschillende drijfveren hebben voor het

verder verbeteren daarvan, zijn er verschillende gebruikerseisen die soms niet compatibel zijn. De belangrijkste eisen zijn, op hoofdlijnen, de eis dat het hele luchtvaartsysteem inclusief afhankelijkheden wordt gerepresenteerd, dat het model de invloed van de mens en van organisaties representeert, dat het model is gevalideerd en kwantitatieve en reproduceerbare resultaten levert, dat de model representatie transparant is en dat het model huidige en toekomstige ongevallenscenarios kan representeren.

In hoofdstuk 5 wordt aan de hand van enkele praktijkvoorbeelden beschreven op welke wijze op dit moment veiligheidsanalyses worden uitgevoerd in de luchtvaart. De in die voorbeelden gekozen aanpak wordt vergelijken met de gebruikerseisen die in hoofdstuk 4 zijn afgeleid. Hierdoor wordt een nog scherper inzicht verkregen in wat een causaal risicomodel moet kunnen om een meerwaarde te hebben ten opzichte van huidige methodes en technieken. Uit het hoofdstuk blijkt dat er op dit moment geen standaard is voor het uitvoeren van veiligheidsanalyses. Dit leidt telkens opnieuw tot een andere aanpak. Beschikbare resultaten van eerdere analyses worden vaak onvoldoende benut en de resultaten zijn sterk afhankelijk van expert mening. Geen van de besproken voorbeelden voldoet volledig aan de in hoofdstuk 4 genoemde gebruikerseisen. Ontwikkeling van een methode die wel aan die eisen voldoet lijkt daarom gerechtvaardigd.

In andere industrieën dan de luchtvaart worden causale risicomodellen toegepast voor de beheersing van veiligheid. In hoofdstuk 6 wordt besproken op welke wijze causale modellen worden gebruikt in de nucleaire industrie, bemande ruimtevaart, offshore industrie, proces industrie, rail transport en de medische wereld en welke lessen daaruit kunnen worden getrokken voor de ontwikkeling en gebruik van soortgelijke modellen in de luchtvaart. Voor de nucleaire industrie, bemande ruimtevaart, offshore en de proces industrie geldt dat aanvankelijk weerstand bestond tegen de toepassing van risicomodellen. Enkele catastrofale ongevallen leidden echter tot veranderingen in regelgeving die de partijen er toe dwongen om probabilistische methoden te gaan gebruiken. Ontwikkeling van modellen en methoden die als acceptabel konden worden beschouwd vergde veel tijd. Op basis van deze ervaringen kan worden verwacht dat ontwikkeling en invoering van causale risicomodellen in de luchtvaart eveneens met horten en stoten zal geschieden and dat het tientallen jaren kan duren voordat het gebruik ervan volledig is geaccepteerd. Het is essentieel om vanaf het begin duidelijk te maken dat deze modellen bedoeld zijn om bestaande methoden aan te vullen en niet om ze te vervangen.

Voor modelrepresentatie en berekeningen zijn verschillende technieken beschikbaar. De meest gangbare methoden (foutenbomen, gebeurtenisbomen, Bayesian belief nets en Petri-nets) en hun belangrijkste eigenschappen worden beschreven in hoofdstuk 7. Foutenbomen en gebeurtenisbomen hebben als belangrijk voordeel dat ze eenvoudig zijn. Hierdoor zijn ze transparant maar de toepassingmogelijkheden zijn beperkt. Bayesian belief nets zijn beter geschikt voor de representatie van ‘zachte’ invloeden, maar hebben als nadeel dat ze minder transparant zijn. Met Petri-nets kunnen ook dynamische situaties worden gerepresenteerd, maar hiervan is de transparantie nog minder. De conclusie in dit hoofdstuk is dan ook dat een combinatie van technieken moet worden toegepast, bijvoorbeeld met gebeurtenisbomen en foutenbomen voor de representatie van de bovenliggende laag van het model (het deel wat direct door de gebruiker wordt gezien), terwijl invloedsdiagrammen en mogelijk Petri-nets gebruikt kunnen worden, daar waar nodig, voor correcte representatie van dieper liggende details.

Een van de eisen aan een causaal risicomodel is dat het kwantitatieve resultaten moet leveren. Hoofdstuk 8 is geheel gewijd aan kwantificering. Een van de problemen bij het

modelleren is het vinden van de juiste eenheden, in het bijzonder als de invloed van samengestelde factoren zoals veiligheidscultuur, veiligheidsmanagement of niet-technische vaardigheden van piloten moet worden uitgedrukt. De geloofwaardigheid van het model hangt af van de kwaliteit van de modelstructuur en van de gegevensbronnen die worden benut voor kwantificering. Belangrijke bronnen zijn ongevalsonderzoeksrapporten en incident rapportage systemen van luchtvaartmaatschappijen. Soms kunnen resultaten van empirische studies worden gebruikt. Expert mening kan, mits op de juiste wijze verkregen, resulteren in gegevens van voldoende kwaliteit en nauwkeurigheid. Naast gegevens over ongevallen en incidenten zijn voor het vaststellen van de relatieve invloed van factoren ook gegevens nodig over normale operaties. Flight data recorders kunnen hiervoor een belangrijke bron zijn. Om de gegevens uit bestaande databestanden zo efficiënt mogelijk te benutten is het een vereiste dat definities en beschrijvingen van gebeurtenissen in het model overeenkomen met die zoals gebruikt in de bestaande databestanden. Bij voorkeur moet het model zich conformeren aan de definities van het Europese systeem voor incidentenrapportage ECCAIRS. Een voorwaarde daarbij is wel dat een aantal van de huidige problemen van ECCAIRS worden opgelost.

Naar verwachting zullen de grootste moeilijkheden optreden bij de representatie van de mens, de representatie van de invloed van organisaties en management en het representeren van de vele onderlinge afhankelijkheden tussen delen van het model. In hoofdstuk 9 worden deze problemen besproken en worden oplossingen aangedragen. Enerzijds worden die gevonden in het toepassen van Bayesian belief nets voor representatie van ‘zachte’ invloeden en afhankelijkheden, anderzijds door definitie van een beperkt aantal (33) ‘archetype’ ongevallen. Deze archetype ongevallen kunnen de ruggengraat vormen van een causaal risicomodel. De menselijke operator kan in het model gerepresenteerd worden door middel van een Bayesian belief net, waarbij de invloed van het management op het presteren van de mens wordt gerepresenteerd door middel van generieke ‘delivery systemen’.

In hoofdstuk 10 wordt ingegaan op mogelijkheden tot validatie van een causaal risicomodel. Volledige validatie van een model is niet mogelijk vanwege de complexiteit van het luchtvaartsysteem, maar validatie van deelcomponenten is wel beperkt mogelijk. Hiervan worden enkele voorbeelden gegeven.

De conclusies van dit proefschrift worden vermeld in hoofdstuk 11. De huidige mechanismen voor veiligheidsverbetering zijn onvoldoende om voortdurende verbetering van de veiligheid te bewerkstelligen. Er is een methode voor veiligheidsanalyse nodig die op niet alleen is gebaseerd op ongevalfrequenties. Met de methode moet de kwantitatieve kans op een ongeval kunnen worden bepaald, en de methode moet op de juiste wijze de complexe inrichting van het luchttransportsysteem beschrijven, inclusief de invloed van mensen en organisaties. De methode moet transparant zijn en aantoonbaar valide resultaten opleveren. Gebruik van een causaal risicomodel zoals beschreven in dit proefschrift voldoet aan deze eisen en is derhalve een manier om mogelijke veiligheidsverbeteringen te identificeren die met gangbare methoden niet mogelijk zijn. De ontwikkeling van een causaal risicomodel voor luchttransport is niet uitzonderlijk moeilijk, het is alleen veel werk. Het grootste deel van dit werk zal moeten worden uitgevoerd door mensen met voldoende achtergrondkennis van luchttransport. Om de complexiteit van het luchttransport te kunnen representeren in een causaal model zijn vele aannames nodig, en elk van deze aannames is opponeerbaar. Een causaal risicomodel kan alleen succesvol zijn indien overeenstemming aangaande deze aannames wordt bereikt met alle betrokken partijen. Dit proces zou geleid moeten worden door een regelgevende instantie, bij voorkeur een

internationale regelgever zoals EASA of ICAO. Zij zouden een proces moeten initiëren en leiden om te komen tot een internationaal standaard model en om overeenstemming te bereiken over hoe en wanneer het model dient te worden toegepast. Voor de verdere ontwikkeling van causale risicomodellen voor luchttransport is het essentieel dat deze modellen direct gekoppeld kunnen worden aan een gegevensbestand van ongevallen en incidenten, bij voorkeur het ECCAIRS systeem. Terugkoppeling van modelontwikkelaars naar potentiële gebruikers en vice versa is vereist om de gebruikerseisen beter te definiëren. De modelontwikkelaars moeten de mogelijkheden en beperkingen van het model aangeven, en moeten praktijkvoorbeelden aandragen van gevallen waarbij een causaal risicomodel een bruikbaar instrument zou zijn. Dit is een iteratief proces. De steun van de industrie is nodig voor het leveren van gegevens en om de modelontwikkelaars inzicht te verschaffen in operationele luchtvaart processen.

Appendix A: The history of third party risk regulation at Schiphol

Background
The involuntary exposure to the risk of aircraft accidents of the people living in the vicinity of an airport is referred to as third party risk. It is usually expressed as individual risk and group risk.

In 1952, the President's Airport Commission in the USA, best known as the Doolittle Commission in honor of its chairman, James Doolittle[88], conducted one of the first comprehensive studies of the noise and safety relationships between airports and surrounding communities. Among other things, the commission plotted the location of over 30 off-airport commercial and military aircraft crashes which caused death or injury to persons on the ground (there is no indication in the report that any data was gathered regarding non-injury accidents). Despite the rather limited database, the commission's report led to the establishment of what became known as clear zones and are now called runway protection zones at the ends of airport runways [Doolittle 1952, State of California 2002].

The importance of risk around airports was recognised in the UK in the 1950s. Public Safety Zones (PSZs) were introduced in 1958 following the recommendations of the Committee of Safeguarding Policy [Le Maitre 1957]. A PSZ was defined as an area adjacent to the end of a runway in which development of land is restricted if it would be likely to increase significantly the numbers of persons "residing, working or congregating there". The suggestion of the 1957 Committee of Safeguarding Policy was for a longitudinal limit of 4,500 feet (1,372 m) for the PSZ. This was based on the subjective choice that 65% of landing and take-off crashes should be contained within the PSZ area [Evans et al 1997].

A string of high profile accidents in the mid seventies (Flixborough, UK, chemical plant explosion 1974, explosion and fire at DSM in Beek, the Netherlands, 1975, the accidental release of dioxin near the Italian town of Seveso,1976) started a debate in Europe on the extent to which governments should control hazardous industry. New regulations were developed such as CIMAH (Control of Industrial Major Accident Hazard) in the UK. Regulation on risk management in the Netherlands was mainly shaped by the regulation on

[88] James 'Jimmy' Doolittle (1896 - 1993) was one of the most famous pilots during the inter-war period. He conducted flight testing of artificial horizons and directional gyroscopes which resulted in the first take-off and landing conducted on instruments alone, i.e. with no outside view. In 1932, Doolittle set the world's high speed record for land planes in the notorious Gee Bee R-1 racer with a speed averaging 252 miles per hour [Daso 2003]. In 1942, he led the daring one-way attack of sixteen B-25 bombers from the aircraft carrier USS Hornet, with targets in Japan [Lawson & Considine 1944].

LPG [Tweede Kamer 1983], which set quantitative limits to individual risk and societal risk.

Stand still for Schiphol risk

Third party risk around Schiphol Airport was formally addressed for the first time in the 'Plan van Aanpak Schiphol en Omgeving' (PASO) at the end of the 1980's. In the framework of the expected doubling of the number of movements at Schiphol, an Environmental Impact Assessment (Milieu Effect Rapportage, MER) was conducted within which third party risk was also investigated. Starting point was the concept that the growth of traffic at Schiphol should not lead to an increase in risk, the so-called *stand still* principle. The discussion on third party risk was intensified on October 4, 1992, when a Boeing 747-200 cargo aircraft lost two engines immediately after take-off from Schiphol and subsequently crashed into an apartment building in a suburb of Amsterdam during an attempt to return to the airport. All 4 crew members and 39 people on the ground were killed.

At that point, the stand still principle for third party risk was not yet translated into measurable norms. As a result of the first Schiphol MER, the MER Commission advised that the stand still principle should be interpreted as no increase of the number of houses within the 10^{-5}, 10^{-6}, 10^{-7} and 10^{-8} individual risk contours and no increase of group risk', thereby accepting risk criteria (individual risk and group risk) that had been used for third party risk of stationary installations.

When it became clear that stand-still of third party risk was not feasible with these criteria, new policy was developed in the Planologische Kernbeslissing Schiphol en Omgeving (PKB) using a new risk measure: the 'Gesommeerd Gewogen Risico' (GGR) (summed weighted risk) [Tweede Kamer 2001]. GGR is the sum of all houses within a particular area multiplied by the individual risk at each house. To some extent, the GGR takes aspects of spatial planning into consideration, and therefore it is a risk measure that falls between individual risk and group risk. The GGR only considers houses and not objects such as office buildings, schools, hospitals etc. The stand-still policy was now formulated as: The GGR shall not increase within the 10^{-5} and 10^{-6} individual risk contours.

New law for Schiphol

With the change in 2001 of the aviation law with respect to Schiphol, the GGR was dropped. External safety is considered in both the luchthavenindelingbesluit (airport layout decree) [DGL 2002a] and the luchthavenverkeersbesluit (airport traffic decree) [DGL 2002b] of the new Chapter 8 of the Aviation Law. In the luchthavenindelingbesluit restrictions regarding permissible buildings are described, taking into consideration third party risk. Four types of areas are indicated upon which different building restrictions act. These areas are defined by noise contours and contours of individual risk. For most locations the noise contours are larger than the individual risk contours and hence determining. In the luchthavenverkeersbesluit the Total Risk Weight (TRG) was presented as a measure for third party risk. The TRG is defined as the product of the average accident probability per aircraft movement per year and the summed maximum take-off weight of all aircraft movements in that year. The TRG was an attempt to express the risk to which the whole neighbourhood of the airport is subjected as a single number. This TRG does not take into account (changes in) use and location of routes, runways and changes in buildings and population densities in the surroundings of the airport. The maximum allowable limit value for the TRG was determined at 9.724 tons per year. The rules of the luchthavenverkeersbesluit encourage that the limit value for third party risk is not exceeded

but they do not guarantee this. The explanation of the luchthavenverkeersbesluit states “To optimally conduct the luchthavenverkeersbesluit within the limit values requires tuning and co-operation between sector parties. Article 8.18 of the law requires sector parties to take such measures, as can reasonable be demanded, individually and in joint co-operation, in order to accomplish that the limit values are not exceeded”. The way in which the norms for third party risk with respect to Schiphol airport must be upheld is described in the safety oversight policy for Schiphol [IVW 2003]. This document refers among other things to the Regeling Milieu Informatie Schiphol (Ruling on Environmental Information for Schiphol) [RMI 2003]. When limit values are exceeded, the government can take sanctions, such as issuing a fine. In 2005 the number of aircraft movements was substantially more limited by noise contours than by the TRG norm.

In the first instance the luchthavenverkeersbesluit [DGL 2002b] did not contain a limit value for (a measure for) group risk. Up to then it had not proven to be feasible to establish a method suitable for aviation to ‘capture’ group risk in a number and to compare this number with a fitting limit value. The failure to find a feasible metric for group risk that was suitable for aviation led, in the debate between the government and the parliament about the proposal for an amendment, to a change in the original bill, wherein it was laid down that a statistical-causal safety model should be ready by 2005. The model should also serve as foundation for the implementation of a measure to determine group risk. This group risk should not be higher than the group risk of 1990, calculated in the same manner.

Group risk, particularly when presented in the form of FN-curves, is a concept that is difficult to grasp. It does however provide insight into the probability and size of a calamity or disaster. Additional to this is the problem of a limit value for group risk. For static installations this was perfectly possible, be defining an unacceptable area under a line on the FN graph. The fact that in the last few years no norm for group risk at Schiphol was established has led to a situation where concrete limit values for group risk have not been included in the Aviation Law. This clearly illustrates the difference between calculating a level of safety and setting a *target* level of safety. With group risk the calculation is not a problem. The problem is the limit value, since all limits proved unattainable almost as soon as suggested. This was not a problem of setting a limit, but of predicting in advance if it could be met and what would be the (political) consequences of keeping to it [Piers & Ale 2000]. These would have been the closure of the airport to any more traffic, thus damaging its role as driver of the Dutch economy.

By then the linkage between a statistical causal model and a norm for group risk had also been abandoned. The main issue was that the causal model would not solve the problem because the Ministry and Chamber completely misunderstood what it could do. In answer to questions from parliament it was also explained that the sector parties, given the results of the demonstration phase of the project causal modelling, had insufficient faith in the feasibility of a complete causal model. The loss of faith however was more general and linked to the anticipated use of the model by the Ministry, i.e. to use the model for enforcement. In any case, the sector parties stopped their contribution to the development of a statistical causal model. In consequence of this, it was no longer possible that, in accordance with the amended aviation law, a causal model Schiphol would be operational before 1 January 2005 [Tweede Kamer 2003a, 2003b, 2003c].

In almost all relevant studies that have been conducted since 1990 it was concluded that the intended stand still for group risk is not feasible due to the increase in the number of movements at Schiphol and the increase of buildings in the vicinity of the airport. Because

of this a situation has been created in which requirements for group risk are an obstruction for the growth of the airport. Suitable criteria for group risk have proven to be politically difficult to find. This has complicated and delayed the discussion regarding external safety.

Causal model as a solution?
In conclusion it can be stated that the government, in its policy with respect to third party risk in the vicinity of Schiphol Airport, has mainly been focussing on individual risk. Criteria for group risk are too complicated to develop a policy that fulfils the requirements of the government (transparent, maintainable, measurable, controllable, technically and operationally feasible) and there is no political will to put limits on the use of Schiphol. The stand-still norm is, given the risk measures that were selected in the past (in particular with respect to group risk), restricting for the growth of Schiphol. This has complicated and delayed the discussion regarding third party risk. The government is also looking for a policy to control the 'disaster potential'. This desire is to a large extent, but not completely, a result of the Bijlmer disaster (see for example the motion of Samson and Duyvendak of 25 November 2003) [Tweede Kamer 2003c].

The industry is confronted with an individual risk limit value similar to limit values for aircraft noise and air pollution. These limit values are in the first place regarded by the industry as possible restrictions for the activities at and around the airport and not as limits to prevent the neighbourhood from being subjected to a risk that is too high. The present limit value for third party risk (TRG of 9.724 tons per year) is calculated from an accident probability, the number of aircraft movements and the maximum take-off weight. The sector can influence the TRG by the types of aircraft that are being allowed to use Schiphol and the number of movements. With respect to the types of aircraft, already 95% (percentage from 2002) of all movements at Schiphol are being conducted by third generation aircraft, the most modern and hence safest class of aircraft. Practicably, a reduction of the number of flights is the only option that is available for the sector when the limit value for TRG is in danger of being exceeded. Article 8.18 of the Aviation Law obliges the sector parties to take such measures as can reasonably be demanded, individually and in co-operation, to accomplish that the limit values are not exceeded. All in all, the sector has little choice regarding those 'measures'. In any case this policy does not provide impetus for the sector to start focussed development and to take measures for the improvement of internal and external safety.

The Ministry of Transport and Water Management has recognised this and is looking for ways to come to a solution. Causal risk models are being considered as one possible solution. The Ministry argues that causal risk models can be used to quantify improvements in safety in the accident probability [Tweede Kamer 2003b]. It is then possible to use this accident probability for the calculation of safety levels and compare those with the target. In doing so, the causal risk model provides directions to the sector for the improvement of safety. It becomes a tool for risk informed decision making and can help the industry to capitalize on measures they have already taken. Only then can we expect support from the industry in the development on such a model.

Appendix B: The aviation system

The aviation system is characterised by a complex arrangement of organisations and activities that has virtually no geographical boundaries. Air transport has always been an international affair. Because of its relatively high speed the aircraft is an attractive means of transportation for long distance travel. In Europe this is synonymous with international travel; KLM's very first destination in 1920 was Amsterdam on a flight from London [Van Kampen 1960]. Aviation today is an international business in almost every aspect. Airlines have become international organisations (SAS, Air France/KLM) that cooperate in code sharing alliances such as One World, STAR and Skyteam. The aircraft that are being operated are manufactured by international consortia like Airbus. Air traffic control, particularly in Europe, requires coordination and integration of national traffic control centres in the pursuit of a 'single European sky. 'Rules of the air' are developed by a special body of the United Nations that is called the International Civil Aviation Organisation (ICAO). Aviation is also a business with many different actors (airlines, aircraft manufacturers, air traffic control service providers, airport organisations, maintenance providers, security etc) and it is a distributed system in the sense that the different actors are not physically located in a single spot. At minimum there are two locations; the location of departure and the destination.

Central to all aviation system processes is the primary process of flying 'from A to B' as schematically represented in Figure B1. The primary- and subsidiary aviation processes are described in the following sections. They are the current processes in developed countries. In some underdeveloped countries the aircraft and infrastructure may be lacking some technologies and procedures may also be less well developed. These processes are also in a continual state of change due to technological advances and demands from society.

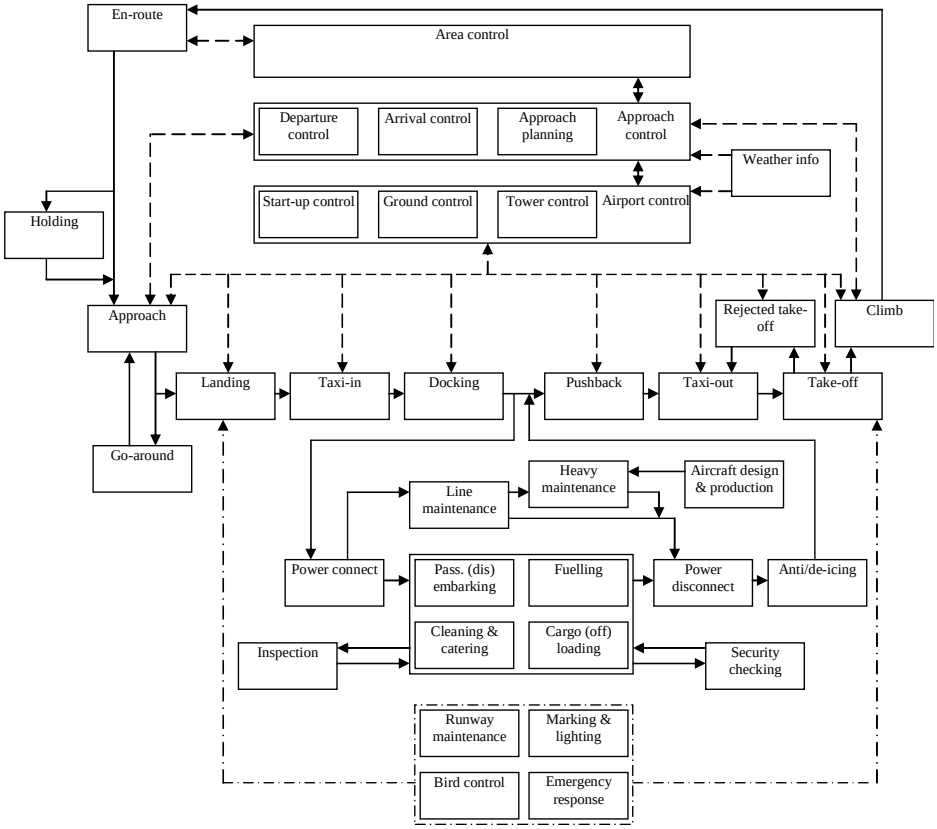


Figure B1: Air transport processes

A typical flight

Pre-flight, crew centre

The cockpit crew normally consists of a Captain and a First Officer. On long flights an additional First Officer or Cruise Relief Pilot (CRP) may be scheduled for the flight. The cockpit crew reports for duty at the Crew Centre at the airport typically 90 minutes before scheduled departure time. For larger airlines much of the preparation has been done by then by the Dispatcher. A briefing package has been prepared with anticipated loading figures and weights, special loads (if any), aircraft deficiencies relevant for flight planning (if any), flight plan route, including speed and altitude schedule, a copy of the ATC filed flight plan, departure slot-time (if any), actual and forecasted weather and NOTAMs[89] for departure airport, destination, alternate and en-route airports, temperature and wind charts and significant weather charts for en-route and a minimum block fuel for the planned flight. Based on the information available the cockpit crew may choose to take extra fuel. When the flight plan is accepted, the final fuel figure is then forwarded to the fuelling department by the Dispatcher. The cockpit crew then proceeds through airport customs and security checks to the aircraft and normally arrives at the aircraft approximately 40 minutes prior to departure, just before passenger boarding commences. If not already met in the crew centre, the cockpit crew meets the cabin crew on-board the aircraft and discusses any relevant

[89] NOTAM stands for Notice to Airmen. NOTAMs are messages to alert the pilots of hazards or abnormalities anywhere along the planned route.

issues with the Senior Flight Attendant, such as anticipated delays, flight time, forecasted turbulence and security procedures.

Pre-flight, aircraft at the gate

Any deficiencies to the aircraft, cockpit or cabin are stated in the Hold-Item-List (HIL) in the Aircraft Technical Log (ATL) for aircraft and cockpit matters and Cabin Technical Log (CTL) for cabin matters. The release of the aircraft by the maintenance department is done after the pre-flight inspection by the ground engineer and signed in the ATL. The cockpit crew verifies that all open items recorded on the previous flight are either repaired or added to the HIL and verifies release of the aircraft. The aircraft may be released to service with aircraft or cockpit system deficiencies if these deficiencies are stated in the Minimum Equipment List (MEL). Any contingency procedure for unserviceable equipment is mentioned in the MEL as well. Acceptance of any cabin deficiencies is done by the cockpit crew in concert with the Senior Flight Attendant.

During the boarding of the passengers, loading of the cargo and fuelling of the aircraft, the cockpit crew prepares the aircraft for departure. On the flight deck duties are normally divided between Pilot Flying (PF) and Pilot Not Flying (PNF) but some actions are specifically assigned to the Captain or First Officer irrespective of which of the two is PF on the stretch. The decision who is going to be PF on the stretch is taken either at flight planning or when arriving in the cockpit.

The time that the aircraft is parked at the gate is a busy time for the cockpit crew. One pilot (normally the PF but this also may be assigned to the First Officer) enters the flight plan data in the Flight Management System (FMS) and prepares all other aircraft systems for departure; the other pilot communicates with the gate agent, the airline maintenance department, the cabin crew, airline hub control and the Dispatcher to keep track of latest developments. When fuelling is complete the fuelling department will bring the fuelling ticket to the cockpit, the flight crew then checks the actual fuel loading with the flight plan fuel.

The latest weather observation for the departure airport is recorded from the Automatic Terminal Information System (ATIS). Approximately 20 minutes prior to departure ATC Departure Clearance Control is contacted to obtain a departure clearance for the departure runway, Standard Instrument Departure (SID) route, transponder code and ATC slot-time (if any), this is done either by voice or via datalink. The ATIS weather observation, the latest estimate of the take-off weight and the expected departure runway are used to calculate take-off performance figures (weight limitation, engine duration, bleed-air demand and take-off speeds).

When all systems settings and take-off performance calculations have been examined by both pilots, these are verified by reading the pre-departure checklist. The PF briefs the PNF on the taxi- and departure route, special weather conditions, applicable NOTAMS and other relevant issues such as discussing high terrain in the departure route, related safe altitudes, and the engine-failure procedure and highlighting some Standard Operating Procedures (SOPs)[90].

[90] All actions of the flight crew, including those in abnormal or emergency situations, are prescribed in procedures. These are described in the Basic Operating Manual, which contains the general procedures of the airline, and the Aircraft Operating Manual, which contains procedures specific for the type of aircraft.

Approximately 5 minutes prior to departure the gate agent (or alternatively the Dispatcher through datalink) hands over the cargo manifest for special cargo, passenger information list, load sheet and any last-minute changes to this data, all based on the actual loading figures. The latest data are entered in the FMS and checked against the take-off performance data that was based on pre-flight estimates.

Push-back and engine-start

When all passenger and cargo doors are closed and the push-back truck has arrived, start-up and pushback clearance is requested from ATC Start-Up Control and Ground Control successively. The direction of the push-back and possible subsequent pull-out is prescribed per gate or overruled by ATC. During the push-back or after the pull-out the engines are started. During the push-back the PF normally remains in contact with the push-back driver, the PNF remains in contact with ATC Ground Control.

Taxi-out

The push-back truck will disconnect after completion of the push-back manoeuvre. When the engines have successfully been started and relevant cockpit systems have been set for taxiing, the taxi clearance is obtained from ATC Ground Control. The taxi clearance indicates the cleared taxi route from the gate to the departure runway. This may be a standard route if this route is published in the airport information (AIP) or else a detailed instruction on which route to take. Runway signs and markings as well as a map of the airport layout are used by the cockpit crew to navigate on the airport. During taxi-out aircraft systems and configuration are set and verified for departure by reading the taxi-out checklist. The cabin crew reports that the cabin is secured and ready for departure. The flight crew will report their position and ready for departure when the aircraft has reached the runway, where ATC Ground Control will hand the aircraft over to ATC Tower Control. Tower must then check the position of the aircraft either visually or by using the ground radar system, after which a line up approval will be given when the preceding aircraft has initiated the take-off roll.

Take-off

ATC Tower Control will issue a take-off clearance when the preceding aircraft has lifted off and no further airspace restrictions apply. When the before take-off checklist is completed, the take-off roll is commenced by the PF. All actions and call-outs, performed for the purpose of cross-checking, during take-off and initial climb are prescribed and performed according to SOPs. As the aircraft accelerates down the runway, the PNF calls out airspeed, typically at 80 kts, V_1 (take-off decision speed) and V_R (rotation speed). If a problem occurs before V_1, the take-off will be aborted. After passing V_1 the aircraft is committed to take-off. Immediately after lift-off the gear is retracted, followed later by thrust reduction, flap retraction and acceleration to climb speed. Also the autopilot will be engaged.

Climb

During flight, duties are assigned according to PF/PNF roles. The PF will monitor the flight path and adjust where required by selecting autopilot modes or changing data in the FMS. The PNF will assist the PF where required, will do the communication with ATC and perform administrative duties. The departure is flown according to the applicable SID, but sometimes a flight may have to deviate from the planned departure route because of weather or by ATC request. As the aircraft climbs to the assigned cruising altitude, and during cruise flight, the flight crew will have to switch to different ATC centres. The flight

proceeds to its destination by following assigned airways. Navigation is predominantly done using the FMS that obtains its information from a range of different sources (air data measurements, fuel calculation, GPS and radio position information, aircraft system status). An on-board weather radar provides information on actual cumulative weather conditions ahead of the aircraft. The useable range of the weather radar is approximately 100 NM or 15 minutes flying time ahead.

Cruise/descent

During the flight updated actual weather observations and forecasts of key airports (destination, en-route alternates) can be obtained directly through datalink, via dedicated radio transmissions or as a request to ATC. Also the flight crew can obtain any information as required from the Dispatcher who may be reached via datalink, radio (when in reach) or satellite phone (if installed on the aircraft). With the information available the flight crew can continuously update contingency planning for an en-route diversion or emergency.

The flight is normally flown along the airways that were planned in the pre-flight phase, but ATC may issue short-cuts where appropriate. The FMS will continuously calculate important parameters such as distance to, time until over and fuel remaining at the waypoints that make up the flight plan and will calculate optimum speed and altitude. The cockpit crew will negotiate the flight profile with ATC so that an optimal altitude and speed schedule will be flown, subject to traffic restrictions.

The flight crew will start preliminary landing preparations approximately 45 minutes prior to landing. When in range the destination airport ATIS is copied. Landing calculations are made using the latest weather data, expected landing runway, estimated landing weight and planned landing configuration. Navigation systems (FMS programming, radio beacon pre-tuning) and flight deck preparation for descent, approach and landing are performed well in time and correct settings verified according to the descent checklist. The company handling agent may be contacted on a dedicated radio frequency to obtain the gate assignment and discuss items such as special handling requests for passengers or cargo and discussion of possible maintenance actions upon arrival. The PF will give the crew briefing on expected arrival route and approach type, cockpit settings, aircraft landing configuration, weather conditions, applicable NOTAMS, expected taxi-route to the gate after landing, instructions for a possible missed approach and other relevant issues such as high terrain during arrival, related safe altitudes and highlighting applicable SOPs.

The descent and arrival may be conducted according to a Standard Arrival Route (STAR) until radar vectoring to final approach (on the extended centreline of the landing runway) is applied by ATC Arrival or Approach Control. During descent and arrival speed is reduced from cruise speed to final approach speed according to the optimal schedule, as prescribed by the STAR or by ATC. Approximately 10 minutes before landing the cabin crew is alerted to prepare the cabin for landing and the approach checklist is read.

Approach/landing

Weather permitting, non-precision or CAT I[91] ILS approaches can be flown manually following the Flight Director (FD) or with the autopilot (AP) engaged. The PF is either hand-flying the aircraft and following the FD or is monitoring autopilot operation, being ready to take over if required. Meanwhile, the PNF scans alternately inside and outside the

[91] Three main categories of automatic approach and landing are distinguished: CAT I, CAT II and CAT III. CAT III indicating an automatic landing under practically zero visibility.

cockpit announcing flight parameter deviations and standard call-outs for cross-checking purposes according to SOPs. CAT II automatic approaches can be followed by a manual landing, although usually SOPs will prescribe an automatic landing with the autopilot engaged. In CAT III weather conditions an automatic landing is mandatory.

When established on final approach, ATC Approach Control will hand over to ATC Tower Control who will provide the latest weather and runways conditions update. When the preceding aircraft has vacated the landing runway, Tower Control will issue a landing clearance. Sometimes this is just seconds prior to anticipated touchdown.

On final approach the aircraft is configured for landing: flaps are extended in steps, gear is extended and speed is reduced to Final Approach Speed (FAS), which is a function of aircraft weight, aircraft configuration and weather conditions. Before reaching the applicable stabilisation height (usually 1000 or 500 ft, depending on weather conditions), the aircraft must be on the correct lateral and vertical flight path (based on radio navigation guidance or visual reference), the aircraft must be in the desired landing configuration and at the desired speed (FAS) and the landing checklist must have been accomplished. Otherwise a missed approach must be initiated.

During the approach, at any time the cockpit crew may abandon the approach if the visibility is below the required minimums, criteria for stabilised approach are not achieved, or when the flight crew is not convinced that safe landing is possible for whatever reason. When a missed approach is initiated, the thrust is advanced, gear and flaps are retracted. The missed approach path is followed using FMS or radio navigation or additional instructions may be obtained from ATC. The aircraft then goes around for a second approach of diverts to an alternate airport.

After touchdown, the following braking devices are used to decelerate the aircraft and bring it to a complete stop: ground spoilers, wheel brakes (including anti-skid and autobrake systems) and the thrust reverser system. Typically at 80 kts the thrust reverse levers are returned to the reverse idle position and then to the stow position when reaching taxi speed. ATC Ground Control will provide a taxi clearance from the runway to the gate or parking spot on the apron. When the engines are shut down the doors may be opened and cargo and passengers are off-loaded. During disembarkation the cockpit crew shuts down the different aircraft systems and completes administrative duties. If on an outstation and the same crew will perform the return flight, initial preparations may be started for the return flight.

Abnormal and emergency procedures

During a flight, events may occur that differ from normal operational practise. For instance a technical malfunction may occur on-board the aircraft or weather conditions at the airport of destination may prevent the aircraft from landing. Like almost all aviation processes, the course of actions in such abnormal conditions is for the greater part described in procedures. For technical failures on-board the aircraft, the associated procedures are described in the form of checklists in the abnormal and emergency procedures sections of the aircraft operations manual. While every attempt is made by the aircraft manufacturer to establish necessary non-normal checklists, it is not possible to develop checklists for all conceivable situations, especially those involving multiple unrelated failures. Pilots should follow these procedures as long as they fit the emergency or abnormal situation. If at any time they are not adequate or do not apply, the Captain’s best judgement should prevail. Only the flight crew operating the aircraft at the time the emergency occurs can evaluate the situation sufficiently to make the proper decision [USAir 1997]. The aircraft operations

manual also contains procedures for events such as missed approaches and encounters with windshear or severe turbulence. Procedures that are not aircraft type specific are described in the airline's Basic Operations Manual.

Because normal as well as abnormal situations and procedures can be causal or contributing to accidents and incidents it is important for the scope of the causal risk model to encompass all.

Subsidiary processes

Flight crew training
An average commercial air transport pilot's career typically starts at a flying school where the pilot receives training for, successively, a private pilot license, a commercial pilot license, an IFR multi engine rating and a multi crew coordination course. Some pilots will have had their initial training in the military services. After this the pilot may spend some time at a smaller operator (e.g. business aviation), which initially requires type qualification training, and of course the periodic recurrent training, proficiency checks and line checks. When the pilot moves to a large airline, (s)he will initially again receive type qualification training, and during the operational career the periodic recurrent training, proficiency checks and line checks. Command qualification training will be required when the progress step is made from First Officer to Captain. New type qualification training will be required when the switch is made from one aircraft type to another. Each training step may involve ground school, simulator training, training flights and line flight under supervision. Training flights are only required in case the simulators are not qualified for zero flight time training or in case of the first type qualification entering an airline.

Air Traffic Control
The captain of a flight is always responsible for the safety of the flight. In those areas that are under Air Traffic Control (ATC), the air traffic controller is responsible for the separation between aircraft. Sufficient separation is required to prevent collisions in the air and to prevent an aircraft from entering the wake of a preceding aircraft. This wake consists of a pair of counter rotating and slowly dissipating vortices. When an aircraft enters a wake that has not yet dissipated, it can suddenly and unexpectedly start to roll, which can be dangerous if close to the ground. To meet this responsibility, the air traffic controller issues instructions that need to be obeyed by the pilot, but when the pilot is of the opinion that following the controller's instruction compromises flight safety, the pilot is allowed to disregard the controller's instruction. The airspace is divided into different blocks (see Figure B2) and generally speaking each ATCo has responsibility for one such block. Horizontally the airspace is divided into different sectors with each normally having a separate area controller. Within these sectors, the aircraft fly along ATS routes. In the vicinity of the airport, the control zone (CTR) and terminal area (TMA) are additional blocks of airspace to help control the flows of arriving and departing aircraft. The area controllers handle aircraft entering their sector on the way to the airport and also crossing traffic overflying the area, either military or civil. Aircraft can be led to a stack if they have to wait for arrival clearance. The area controller hands over to the arrivals controller either directly on the way to the airport, if there is no delay, or when leaving the stack, if there is delay. The arrivals controller takes the aircraft from there until they are lined up for the runway and ILS is established. The tower controller then brings the aircraft in to land. Once it has cleared the runway it is handed over to the ground controller who supervises the taxiing up to the gate. For departure there are usually two additional controllers; one who gives the en route clearance and a second who gives clearance for start-up and push-back.

The ground controller then takes over for the taxiing to the holding point just before the take-off runway. The tower controller then takes over for the take-off up to 2000ft, when the departure controller takes over up to the point where the aircraft is free of all conflict with other traffic at and around the airport. At this point the area controller takes over for the transit to the following sector/area.

In controlled airspace, flight crew and Air Traffic Control communicate by means of VHF radio. Communication involves instructions according to a protocol that is laid down in ICAO Annex 10, part II, Chapter 5: R/T. Air-ground radiotelephony communications must be conducted in the language normally used by the station on the ground or, preferably, in the English language.

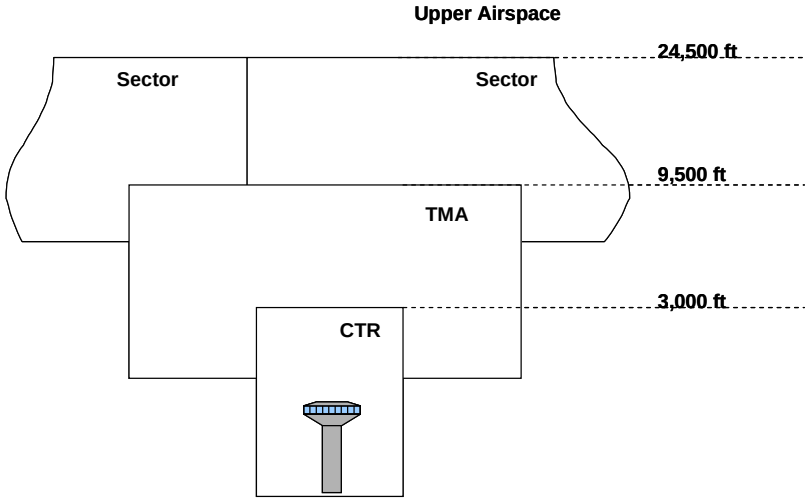


Figure B2: Airspace schematic.

Aircraft design and certification

The Aircraft Type Certification process is officially initiated by the submission of an "application for a type certificate" by the aircraft manufacturer. The regulator establishes the Type Certification Basis; the applicable rules and any additional requirements to which the manufacturer must show (and the regulator must find) compliance in order to be granted a Type Certificate. The regulator will now create a Type Certification Board, a management level group charged with confirming that all applicable regulations have been complied with prior to final approval of the particular design under evaluation. The regulator identifies significant certification items which will require special attention during the certification program. These items are described in "Issue Papers". The manufacturer then submits data on how it is intended to demonstrate compliance with the applicable regulation in the form of a Certification Plan. After approval of the Certification Plan the actual certification process will start. Detailed design submittals begin, including drawings, reports, analyses, test plans and test reports. These data are reviewed and any revisions or additions that are found necessary must be incorporated before final acceptance by the authorities [Haubrich 1984].

An important aspect of the aircraft certification process is the flight test program. Flight test programs are divided into an evaluation flight test program and a certification flight-test

program. The evaluation flight tests are executed by the manufacturer, usually in close co-operation but without direct involvement of the certifying authority in order to obtain a first overview of the handling and performance characteristics of the aircraft. The certification flight tests are executed with the participation of representatives of the authority and partially flown by the authority's test pilots.

A Compliance Check List is composed which provides information concerning the substantiation information which has been provided to show compliance with applicable regulation. When the regulator determines that the manufacturer has demonstrated compliance with all applicable airworthiness standards, the required inspections have been satisfactory completed, and the required documentation is approved, the Type Certificate is issued. A Type Certificate Data Sheet is issued as part of the Type Certificate. It describes the conditions and limitations under which the aircraft meets the applicable airworthiness requirements. Within 90 days after issuing the Type Certificate, a Type Inspection Report has to be issued which includes the final documentation of the ground inspection and flight test phase of the certification program. In addition to the Type Certification, a number of provisions must cater for the airworthiness of each individual aircraft. Before the manufacturer can produce aircraft intended for sale it must obtain a Production Certificate. For each individual aircraft coming off the production line a Certificate of Airworthiness has to be issued. After the Type Certification is completed and aircraft of the type are introduced into service, the authorities continue to monitor the safety of the design through a service difficulty reporting and analysis system.

Aircraft maintenance

Maintenance refers to 'all activities necessary to keep the aircraft in, or restore it to, a specified condition'. Aircraft maintenance is a complicated and costly business and is characterized by large amounts of regulations, procedures and documentation. This section describes generally the way in which a maintenance program is created and how it is implemented and executed.

To ensure that the necessary regulations, guidelines and standards are applied and adhered to, and that the correct tasks and inspections are conducted at the correct time, every aircraft type will have an approved aircraft maintenance program. This lays down the mandatory minimum maintenance program and is produced by the Maintenance Review Board (MRB) consisting of manufacturers and aviation authorities, in consultation with the airlines, to ensure the continuous maintenance needs of the aircraft are met. It includes details relating to the minimum maintenance tasks and inspections which have to be carried out on each aircraft at pre-determined times, depending upon either the number of flying hours, number of flights or calendar time. Tasks and inspections above the absolute minimum should normally accompany those which are required and advice is included in the manufacturer's documentation. Due to an airline's fleet size, route structure, aircraft utilization etc and from years of operational experience, the maintenance program can be customized for the individual airline.

The actual implementation of the aircraft maintenance schedule is a detailed and complex process. Most airlines have departments which are dedicated to the detailed planning, scheduling and control of aircraft maintenance tasks and inspections. Aircraft maintenance is usually scheduled in 'checks' of varying proportions, ranging from walk-around inspections to heavy maintenance 'D' checks. The Production Planning / Engineering Department controls the contents of any maintenance check, in accordance with the aircraft maintenance program and the manufacturer's manual and decides when and where the

aircraft input will occur. They will therefore plan which particular aircraft is due for which check and prepare and produce all job cards and documentation necessary to complete the check. They also ensure the correct spares, manpower, tools, equipment, hangar space and aircraft are organised for a scheduled check in order that everything should be ready when the aircraft comes in. This planning process will take about 4-5 weeks for a large commercial aircraft like a Boeing 767.

In the case of heavy maintenance, once the work has been planned, production control teams take over about 4 weeks before the aircraft arrives. Sometimes a pre-input survey of the aircraft will be carried out by qualified personnel, who will establish the condition of the aircraft and highlight any areas where extra work will be needed, e.g. repair of unforeseen minor damage and long standing defects. In the 4 weeks prior to the aircraft coming in, the proposed work schedule will be thoroughly reviewed and all work scheduled to teams of engineers. Milestones are created to enable the progress to be monitored. The same process would apply for light maintenance except time scales would be significantly reduced since the length of the input and amount of work would be less. Some airlines may try to schedule major work for periods of reduced activity (typically the winter months from September to March), to save on revenue which would otherwise be lost in down time.

Once the actual contents of a check are known, resourcing of hangar space, people, materials, tools and equipment also takes place in this department to ensure that the work to be carried out can be achieved. This process is not as straightforward as it might seem, since for many inputs only about 60% of the work will be predictable beforehand; the rest of the work is raised only as the results of inspections or entries made in the aircraft technical log.

A few days before the aircraft is scheduled to arrive in the hangar, the completed work package for that aircraft will be forwarded to the appropriate hangar control centre. A work package is literally the package of work which defines and sets out all the activities, i.e. tasks and inspections which must be conducted for a scheduled check on a particular aircraft. It typically contains a list of contents, job cards, a job card deviation list (to notify planning of jobs not done and therefore carried as Acceptable Deferred Defect (ADD)), Additional Work Requirements (i.e. non-critical extra tasks such as changing the cabin boarding tape), any relevant Service Bulletins or Airworthiness Directives, a list of parts and equipment needed and ordered, the dispatch reliability list and Certificate of Maintenance release. The work package will be organised and checked by the shift manager who is responsible for monitoring the work progress and output of the work. Shift managers will locate tools, equipment, parts and documentation necessary for the job. They are also responsible for ensuring that they have enough staff on the shift and that such staff has received adequate training. The shift manager will liaise with Planning to advise on the man hours necessary for particular tasks in order to help plan the work.

Once the aircraft has come into the hangar the shift manager will allocate staff to a particular work area. Typically the resourcing of tasks and work allocation is written up on a control board in the control centre of each hangar, so that each technician will know when and where they will be working.

Tasks are scheduled to technicians via ‘job cards’ which detail information in relation to an individual task, or a number of tasks, depending on the practices of the maintenance organisation. A major check may have as many as 5,000 - 10,000 job cards. Each card will

contain a description of the task and contain information pertaining to the type of aircraft, the identity of the aircraft, the original work package the card is from and the zone of the aircraft where the task is to be conducted. They will also indicate how long the task is expected to take and what license requirements are needed to be able to sign the job off as completed. Although all job cards will contain the same basic information, the overall content and format of these cards will differ between maintenance organisations. All tasks will also have a reference to the relevant section of the Maintenance Manual. This Maintenance Manual contains a description of how to carry out every conceivable task on the aircraft, and as such runs to thousands of pages. It includes detailed diagrams where necessary to aid in the location of components and structures. As previously noted, detailed planning can only be done for those tasks which were scheduled prior to input. The remaining unscheduled defect items will have been raised as defect cards from the original scheduled inspections, i.e. any defects which are discovered have job cards raised as they are discovered. During the pre-planning of a check, all job cards are arranged into suitable functional groups ready for the input. The job cards are distributed to the technicians by placing them on racks in designated areas. For example job cards appear on a number of separate display boards depending on the nature of the tasks. The technicians will collect their assigned job cards from these racks to perform the tasks. On completion of the tasks the technician must sign or stamp in the designated place depending on what type of tasks they are licensed to complete. Safety critical tasks will require duplicate inspections. Once signed the job cards are returned to the aircraft's technical records where they are stored for a minimum of two years.

For ramp maintenance, task resourcing and work allocation is a bit more hectic. For typical European airlines from about 6:00 -22:30 there are some 100 to 200 narrow body departures each day and it is the Traffic Coordinators job to assign technicians to particular aircraft. The traffic coordinator will have a 'time-line' of daily departures with the technicians name, radio call and aircraft registration shown against each time window. If an aircraft is delayed the coordinator will try and reassign the work for an aircraft to another technician. In this way they attempt to keep a technician or team of technicians assigned to an aircraft for the duration of its stay rather than swapping and changing the schedule [O'Conner & Reynolds 1996, Roelen et al 2006a].

Airport processes

The airport operator is responsible for providing and maintaining the required infrastructure such as ground equipment, runways, buildings etc and provides services supporting the users of the airport in ground operations. The airport operator shares a safety responsibility and therefore sets requirements regarding the knowledge and the capabilities of its personnel, especially for those that work on 'airside'. The following processes under responsibility of the airport organisation are most relevant with respect to aviation safety:

- Design, construction and maintenance of infrastructure
- Runway friction control
- Prevention of foreign object damage
- Prevention and mitigation of bird risks
- De-icing aircraft
- Control of obstacle clearances
- Fuelling

Infrastructure

Basic principles for the design, construction and maintenance of runways and taxiways are prescribed by ICAO. Heavy maintenance of the runways is scheduled according to a long-range plan with the aim to accommodate the maintenance activities within the operational activities.

Runway friction control

Safe operation of aircraft on the ground requires runways, taxiways and aprons to have sufficient friction. The slipperiness of an existing section of a runway can vary significantly, even within a matter of minutes, as a result of rubber deposits, (heavy) rain, snow and ice [Transport Canada 2004, Sinha 2002]. Especially under winter conditions, additional activities like sweeping and application of grit are required to prevent runways from becoming too slippery.

Foreign object damage

Any loose item that is lying around in the vicinity of the apron, taxiway or runway can cause serious damage to the aircraft when the objects are ingested in the engine or strike vulnerable parts of the aircraft during take-off and landing. The airport tries to minimise the risk of this so-called 'foreign object damage' by a combination of instructions and procedures for all personnel working on airside and by regular visible checks and sweeping activities.

Birds

Birds are a special case of 'foreign objects. Collisions between aircraft and birds can be catastrophic if the bird strikes a vital part of the aircraft like the engine. The majority of bird strikes take place during landing or take-off, because the bird density is higher at the lower altitudes [Poot et al 2000]. The bird strike probability is influenced by the number of local birds and by migration. The local bird population depends on the layout of the airport and the airport vicinity. Local measures by the airport, such as habitat management and bird control (scaring, culling) can significantly reduce the risk [Deacon & Rochard 2000]. Bird migration is more difficult to control for the airport, because this is largely unaffected by the airport itself.

De-icing aircraft

Snow or ice on the wings of aircraft can be a safety hazard, especially during take-off, as the contaminants can disrupt the airflow over the wing and cause the aircraft to stall out of control. Aircraft that have snow or ice on the wings are de-iced before take-off to prevent these types of accidents. De-icing is performed by mobile de-icing trucks or in a stationary de-icing stand at the airport and involves spraying a de-icing fluid, typically a mixture of glycol and water, on the aircraft's surfaces. The de-icing fluid removes snow or ice and also prevents the build-up of additional layers of snow or ice for a predefined time period of time called the hold-over time. Typical values for the hold over time are 30 minutes.

Obstacle free zones

The airport must be permanently aware of any construction plans in the vicinity of the airport to prevent the obstacle free zones for departing and arriving aircraft from becoming obstructed. If necessary the airport must take action at the level of the competent authorities.

Fuelling

This section describes the fuelling process at Amsterdam Airport Schiphol as performed by the Gezamenlijke Tankdienst Schiphol BV (GTS). The fuelling process at other airports may be different.

Aircraft fuel is directly transported to the airport by a pipeline from the storage facilities of the oil companies in Amsterdam and Rotterdam. At the airport fuel is stored in the fuel depot. The aircraft fuel suppliers (the major oil companies) have to guarantee the quality of the aircraft fuel. This quality is checked several times in different stages of the fuel production process. Upon arrival at the airport, the fuel will be left in the storage for 24 hours to allow water and sediments to settle down and be drained.

There are two ways of tanking at Schiphol: A fuel truck can be filled with fuel from the storage facility and this fuel truck can then fuel the aircraft directly from its tank, or the fuel can be distributed using the hydrant system; a network of pipelines underneath the apron. Using the hydrant system one worker with one fuel dispenser, a light vehicle, can perform refuelling.

The airline may have requested a certain amount of fuel before the fuel truck or dispenser leaves GTS for refuelling a certain aircraft. Alternatively the fuel operator may be informed by the airline representative on the ramp about the fuel quantity that has to be delivered when he arrives at the aircraft. The airline can requests an initial amount of fuel, which will be filled up later with more fuel when accurate data on cargo and passengers has become available.

The fuel operator connects the fuel hoses to the fuel panel on the aircraft and the hydrant system, or directly from the tanker to aircraft. The parking brake of the fuel truck engages automatically until the fuel hose is fully stowed. The required fuel quantity is entered either in the fuel panel underneath the wing by the airline representative or by the flight crew from the cockpit. The aircraft's fuel system will automatically divide the fuel over the tanks and will shut off the valves when the requested amount has been reached. A fuel filter monitors the fuel flow and shuts off the flow when it detects water in the fuel. The fuel operator monitors the process: he keeps an eye on the indicators, his equipment and ramp safety. There are four 'safety tools' to prevent ramp accidents: the dead mans handle (must be activated each 2 minutes, otherwise the fuel flow will stop automatically), a lanyard (when pulled, the valves shut), a motor stop (when the pump fails, the fuel flow stops) and, finally, the hydrant emergency stop (which stops the fuel flow from the hydrant system).

After tanking has been completed, the representative or ground-handling agent receives a fuel receipt (with fuel quantity, truck number etc.), that will be used for the preparation of the loadsheet. That sheet and the data on the delivered fuel quantity are provided to the cockpit crew. Finally, the fuel operator will disconnect and stow away the fuel hoses, remove the fuel tanker or dispenser and drive to the next aircraft.

Safety regulation and oversight

Until the Second World War, aviation regulation had primarily been the responsibility of national governmental bodies like for instance the Civil Aeronautics Authority in the US [Hansen et al 2005]. Towards the end of the war however the need for international

regulation became evident and the International Civil Aviation Organisation ICAO, a specialised agency of the United Nations, was founded in 1944. ICAO's main task is rulemaking. Certification and supervision are still national tasks [Molemaker et al 2005].

ICAO is responsible for developing international rules governing all areas of civil aviation. In the Convention of Chicago in 1944 the ground rules were established to ensure a safe and orderly growth of civil aviation throughout the world. All European countries have ratified this treaty. ICAO's Air Navigation Commission develops regulation in the form of international Standards and Recommended Practices (SARPs). SARPs are designated as Annexes to the Convention on International Civil Aviation (the Chicago Convention) and are signed and adopted, in part and sometimes with reservations, by the individual ICAO member states.

Annex 1		Personnel Licensing
Annex 2		Rules of the Air
Annex 3		Meteorological Service for International Air Navigation
Annex 4		Aeronautical Charts
Annex 5		Units of Measurement to be used in Air and Ground Operations
Annex 6		Operation of Aircraft
	Part I	International Commercial Air Transport – Aeroplanes
	Part II	International General Aviation – Aeroplanes
	Part III	International Operations – Helicopters
Annex 7		Aircraft Nationality and Registration Marks
Annex 8		Airworthiness of Aircraft
Annex 9		Facilitation
Annex 10		Aeronautical Telecommunications
	Volume I	(Radio Navigation Aids)
	Volume II	(Communication Procedures including those with PANS status)
	Volume III	(Part I - Digital Data Communication Systems and Part II – Voice Communication Systems)
	Volume IV	(Surveillance Radar and Collision Avoidance Systems)
	Volume V	(Aeronautical Radio Frequency Spectrum Utilization)
Annex 11		Air Traffic Services
Annex 12		Search and Rescue
Annex 13		Aircraft Accident Investigation
Annex 14		Aerodromes
	Volume I	Aerodrome Design and Operations
	Volume II	Heliports
Annex 15		Aeronautical Information Services
Annex 16		Environmental Protection
	Volume I	Aircraft Noise
	Volume II	Aircraft Engine Emissions
Annex 17		Security - Safeguarding International Civil Aviation against Acts of Unlawful Interference.
Annex 18		The Safe Transport of Dangerous Goods by Air

Although all regions of the world follow the same global rules that are established through the regulatory framework of ICAO, this does not ensure that all states achieve similar levels of safety. One of the reasons is that States can circumvent compliance with ICAO standards by filing a formal notice of difference to ICAO[92]. Another reason for this is that ICAO's main task is rulemaking. Certification and supervision are national tasks, and as such prone to local interpretations and to the local ability to provide sufficient oversight [Molemaker et al 2005]. Within Europe, some of the national tasks are being transferred to a European level.

There is not a single European body that is responsible for aviation safety. There are at least six organisations, and not every European state is a member of each organisation (Figure B3).

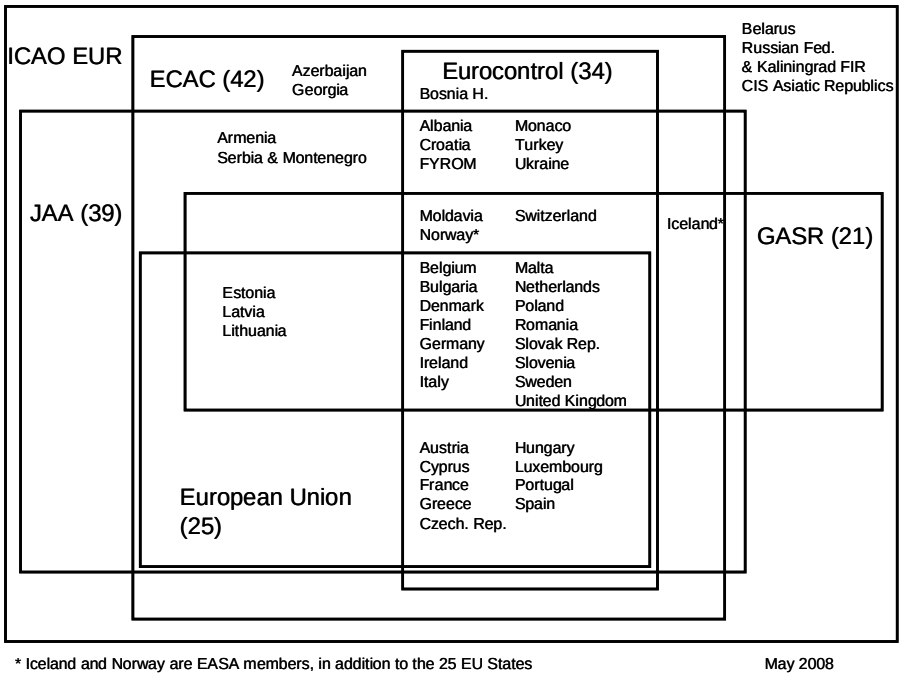


Figure B3: Distribution of European member states across different bodies associated with flight safety.

Almost all European states are a member of the European Civil Aviation Conference (ECAC). Its objective is to promote the continued development of a safe, efficient and sustainable air transport system. The Joint Aviation Authorities (JAA) is an associated body of ECAC representing the civil aviation regulatory authorities of a number of European States who have agreed to co-operate in developing and implementing common safety regulatory standards and procedures. This co-operation was intended to provide high and consistent standards of safety and a "level playing-field" for competition in Europe. Much emphasis is also placed on harmonising the JAA regulations with those of the USA. One of

[92] Many States do this. In the case of Annex 14 volume 1 (Aerodrome design and operations) for instance, 14 countries, including the Netherlands, have filed a notice of difference [ICAO 2004b].

its functions was to develop and adopt Joint Aviation Requirements (JARs) in the fields of aircraft design and manufacture, aircraft operations and maintenance, and the licensing of aviation personnel. Because JAA acted merely as a coordinating body, relying on the good will of national authorities across Europe, there remained differences in the application of JAA rules across the member states. To further improve aviation safety the European Aviation Safety Agency (EASA) was established by the European Parliament and Council Regulation (EC)1592/2002 of 15 July 2002. The aim of EASA is to create a single regulatory framework to promote the highest common standards of safety and environmental protection in civil aviation, to oversee their uniform application across Europe, and to promote them at world level. As a first step, part of JAA's competences (in the domains of certification of aeronautical products, parts and appliances and the approval of organisations and personnel engaged in the construction and maintenance of these products) have been transferred by the European Parliament to EASA. The European Commission later enlarged the competences of EASA with air operations, the licensing of air crew and safety of foreign aircraft (first extension). The intention is to further extend the competences of EASA in the field of regulation (including safety & interoperability) of airports, air traffic management and air navigation services.

Eurocontrol, the European Organisation for the Safety of Air Navigation, was established in the 1960s by six states, including the Netherlands, with the intention of creating a single upper airspace. The current (2008) membership of Eurocontrol numbers 31 states and the primary goal is "One Sky for Europe"; a seamless, pan-European Air Traffic Management (ATM) system. As one of the means to accomplish that goal, Eurocontrol's Regulatory Committee and the accompanying Regulatory Unit draw up regulation with respect to ATM safety. Most prominent of these are so-called Eurocontrol Safety Regulatory Requirements (ESARRs) that impose high level mandatory requirements to ensure harmonised ATM safety regulation. ESARRs are implemented and enforced by Member States through transposition into the national legal order. In December 2003 the European Commission and Eurocontrol signed a memorandum of cooperation in a number of areas, including the implementation of a Single European Sky. Eurocontrol was given mandates for the development of implementing rules for the Single European Sky. The rules are then partially or completely adopted by the European Commission. The fact that Eurocontrol requires conversion of rules into national or Community legislation leads to differences in implementation of ESARRs among member states. Also, Eurocontrol does not have the authority to certify or approve systems and to supervise and enforce the implementation of regulation.

Until EASA will become responsible for regulation of airports, related subjects are left to the responsibility of individual states on the basis of ICAO provisions. In 1995 a number of European states established the Group of Aerodrome Safety Regulators (GASR) to develop, on a voluntary basis, harmonised safety regulation standards for airports and ground aids.

On a national level, states have governmental organisations for rulemaking, certification and oversight. In the United States this is the Federal Aviation Administration FAA. The US has a very strong position in all fields of air transport, including rulemaking. US aviation rules are published as Federal Aviation Regulations (FARs). EASA and FAA work together in harmonisation groups to keep the differences between US rules and European rules limited.

Appendix C: Causal Model for Air Transport Safety (CATS)

In February 1997 a Safety Policy Report on Civil Aviation was accepted by parliament as the policy framework for aviation safety in the Netherlands. An update of policy goals and an accompanying implementation plan for 2006 - 2010 was published in 2005. The development of a causal model of aviation safety is one of the primary topics of the Safety Policy agenda. According to the plan, a causal risk model should be used to identify weak spots in the aviation processes and assess their effect on overall safety. This plan was put into practice: in 2005, the Ministry of Transport and Water Management launched a three-year research effort to develop a causal model for air transport. The project, which received the acronym CATS (Causal model for Air Transport Safety) was conducted by a consortium of Delft University of Technology, NLR, DNV and White Queen [Ale et al 2006].

The original design of the model was based on results of feasibility studies for air transport risk modelling [Roelen at al 2000a, Roelen et al 2002, DNV 2002b] and work done in the area of occupational safety [Ale et al 1998, Bellamy et al 1999, Papazoglou et al 2003]. This design called for the combination of three modelling techniques, Event Sequence Diagrams (ESDs), Fault Trees and Bayesian Belief Nets (BBNs), in a single model.

The backbone of the model consists of 33 generic accident scenarios represented as Event Sequence Diagrams. All ESDs are linked-up and are seen as challenges that have to be faced during a flight, from taxi through the take-off, climb, en-route, approach and landing phase. Each event in the ESD is defined such that there are only two possible outcomes. The outcome is determined by an underlying fault tree. There are separate fault trees for each event in each accident scenario. Human operator models are attached to the fault trees wherever humans are involved in the fault tree events. These human operator models are represented as Bayesian Belief Nets. Separate models were developed for flight crew, air traffic controllers, and maintenance technicians. The whole model was converted into one single integrated BBN. This allows using distributions of values rather than point estimates, dependencies can be handles conveniently and consistently throughout the model, and it removes the need for artificial transfer point between ESDs, fault trees and BBNs. The BBN is of the ‘distribution free continuous’ type, developed by Kurowicka and Cooke [2004]. Nodes in the BBN are associated with continuous distributions and arcs with conditional rank correlations. This means that the complexity is linear with the number of parent nodes rather than exponential as in a discrete BBN. Model calculations are made in UniNet, a continuous and discrete non parametric Bayesian Belief Net application, functioning as a module of UniCorn, a stand-alone uncertainty analysis software package. UniNet was developed specifically for CATS.

Managerial influences on accident probability are represented in the model as the resources and direction which management provides for front line workers (pilots, air traffic controllers, maintenance technicians) to perform their task of controlling risk, this being done through their own actions and through use of hardware and software. Those resources are clustered under the headings of human and technical delivery systems. The ‘quality’ of

the delivery systems determines a management modifier, which is used to modify the distribution of human error probabilities in the human operator model [Lin et al 2008].

Model quantification was conducted from accessible accident and incident data, audit information, data on normal operations and from expert judgement. The method for structured expert judgement developed by Cooke and Goossens [2000] was used to maximise the chance of unbiased estimates. The origin and a characterisation of the quality of each number are stored in a separate database to help users in interpreting the results of the analysis, and also forms a basis for future data requirements.

The model was developed down to a level of detail for which it was considered feasible to derive probability numbers from either data or expert judgement. In many cases however this is not the way in which safety managers actually look at the air transport system. They often use aggregated notions such as the complexity of an airport, the complexity of airspace, runway conditions, etc. These notions translate influences on the probabilities of many of the model elements. Therefore a translation or mapping was made from the notions common in the air transport industry to the base events of the model.

Curriculum Vitae

Alfred Roelen was born on 21 April 1968 in Vught, the Netherlands. In 1986 he received his VWO diploma from the Maurick College in Vught, after which he studied aeronautical engineering at Delft University of Technology. He obtained his MSc degree in aeronautical engineering in 1992, the title of his Master's thesis was 'Design and analysis of a tilt-wing V/STOL aircraft for 19 passengers'. From 1992 to 1994 he joined the Technological Designer programme at Delft University of Technology, Faculty of Aerospace Engineering, obtaining a Master of Technological Design (MTD) degree in 1994. In 1994 he joined the National Aerospace Laboratory NLR, Department of Flight Testing and Safety. He is now a senior scientist at NLR's Air Transport Safety Institute.