Safety Management and Risk Modelling in Aviation

The challenge of quantifying management influences

Pei-Hui Lin

Safety Management and Risk Modelling in Aviation

The challenge of quantifying management influences

Proefschrift

ter verkrijging van de graad van doctor

aan de Technische Universiteit Delft,

op gezag van de Rector Magnificus prof.ir. K.C.A.M. Luyben,

voorzitter van het College voor Promoties,

in het openbaar te verdedigen op dinsdag 21 juni 2011 om 10.00 uur

door

Pei-Hui LIN

Master of Science in Transportation Engineering and Management

geboren te Tainan, Taiwan

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. P.T.W. Hudson	Technische Universiteit Delft
Prof. dr. B. Kirwan	University of Nottingham
Prof. dr. ir. J.A. Mulder	Technische Universiteit Delft
Prof. dr. ir. A.C. Brombacher	Technische Universiteit Eindhoven
Dr. ir. A.L. Roelen	Nationaal Lucht- en Ruimtevaartlaboratorium

ISBN 978-90-79787-32-6

Published and distributed by:
Next Generation Infrastructures Foundation
P.O. Box 5015
2600 GA Delft
The Netherlands
T: +31 15 278 2564
F: +31 15 278 2563
E: info@nginfra.nl
I: www.nginfra.nl

This research was funded by the Next Generation Infrastructures Foundation programme and the Dutch Ministry of Transport, Public Works and Water Management.

Keywords: aviation, safety, risk modelling, safety management system

Cover illustration: © alperium/ 123RF

Copyright © 2011 by Pei-Hui Lin.

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without the prior permission of the author.

Printed by Gildeprint Drukkerijen, Enschede, The Netherlands

To my child,

Rozen

Contents

1 Introduction

Air transport is one of the safest forms of travel (Amalberti, 2001). The average rate of accidents for scheduled commercial aircraft involving passenger fatalities in the region of the 27 EU Member States plus Iceland, Liechtenstein, Norway and Switzerland is approximately 3.6 accidents in 10 million flights (EASA, 2008), as shown in Figure 1-1. Yet as air traffic continues to grow, we continue to witness high-profile failures and loss of lives, such as the Tenerife disaster, Japan Airlines Flight 123, and Singapore Airlines Flight 006. Other countries around the world also continue to struggle with their accident rates. The regions of Africa, Eastern Europe, and West and Central Asia (see Figure 1-1) have the highest rates of fatal accidents in the world. Hence a common initiative is needed at both the state and international level to keep air transport safe and sustainable.

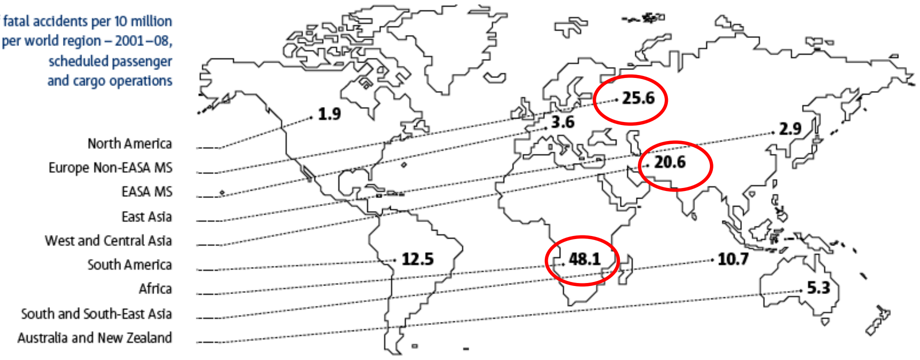

Rate of fatal accidents per 10 million flights per world region – 2001–08, scheduled passenger and cargo operations
North America
Europe Non-EASA MS
EASA MS
East Asia
West and Central Asia
South America
Africa
South and South-East Asia
Australia and New Zealand
1.9
3.6
25.6
2.9
20.6
12.5
48.1
10.7
5.3

Figure 1-1 Rate of fatal accidents per 10 million flights per world region from 2001 to 2008 (EASA, 2008)

For the past ten years, major airlines have started to employ safety management activities such as accident prevention and flight safety programmes, wishing to improve their safety record. Such activities are often overseen by a safety office which monitors overall operating experience and provides independent advice to company management on the action needed to eliminate or avoid identified hazards, or to reduce the associated risk to an acceptable level. The failures and accidents rarely occur from random failures of hardware. They usually arise from a complex combination of factors both in the operational level (pilot, crew, ATC, dispatcher) and in the supporting processes of maintenance, airline management policy, aircraft design, etc.

To eliminate or prevent identified failure in this complex system, the safety program has to go beyond the available local information and make causal inferences stretching much further back in time and up into the higher system levels to understand why events happened. The analyst is able to do so for causal factors closely related in time and space to the event by applying individual knowledge and expertise. But it is extremely difficult to systematically identify latent failures in management activities and monitor management interventions to control risk. Without a systematic model to see the broad picture and show the relative safety priorities, the effectiveness of management remedial action is difficult to monitor. It is therefore not surprising to see that many airline safety improvements are still using a fly-crash-fix-fly approach rather than a proactive approach (Rasmussen and Svedung, 2000). Such a broad model needs to link the on-line risk controls, whose failures represent the proximal causes of accidents, with the underlying management processes which put these on-line risk controls in place and keep them functioning through time. It is only recently that such explicit models have been considered relevant in aviation. Prior to that aviation has operated with very implicit models of how the integrated safety measures work (Roelen, 2008).

There is also a demand for a model which can be used for systematic analysis for safety regulation and inspection, to see the effectiveness of management action on safety. The International Civil Aviation Organization (ICAO), a specialized agency of the United Nations, has mandated its 190 member States to develop and implement safety management system (SMS) programs to achieve an acceptable level of safety in aviation operations (ICAO Annex 6, 2006). According to these ICAO requirements, by 2010 service providers (e.g. airlines and air traffic management) are responsible for establishing an SMS, which is accepted and overseen by their State. The state aviation authorities will have to inspect the SMSs throughout the airline industry. Inspectors will then not only have to make sure that companies comply with detailed practical regulations about on-line risk control, but they may want to encourage them to use the SMS to drive monitoring and improvements in safety in a more efficient and effective way. Such an overarching, explicitly modelled SMS will allow inspectorates to look at and talk to company planning and operations to understand how they identify and address safety hazards before they become manifest in real time. It is also not always possible to cover all aspects of the management processes to be interrogated in one audit. Without a scientifically-based risk management tool to help them systematically identify the priorities of safety management action on safety, it is difficult for authorities to go beyond compliance inspections and to step in when the critical aspects of the company safety management process is poorly managed.

A similar development, from implicit management of risk to an explicit modelling of not only the on-line risk scenarios, but also the safety management processes driving them, has occurred in other industries and activities, such as chemical process (Bellamy et al., 1999; Hourtolou & Salvi, 2003), occupational risk (RIVM, 2008; Ale et al., 2007), nuclear (Davoudian et al., 1994a, b; Mosleh et al., 1997; Yu et al., 2004) and railways (v.d. Top, 2010).

The Dutch Ministry of Transport, Public Works and Water Management recognized the importance of this and, after a long series of feasibility studies starting in 2000 (Roelen et al., 2001), embarked in 2005 on a project called CATS (Causal Model for Air Transport Safety) [(Ale et al., 2006), (Ale et al., 2007), (Ale et al., 2008a), (Ale et al., 2008b), (Ale et al., 2009)] to develop an integrated risk model of air transport for the whole flight cycle from (departure) gate to (arrival) gate. The essential goal of CATS was to model the whole aviation system as it related to risk control and accident prevention. It was intended to link a safety management model with the technical/human factors model and quantify the risk implications of different technical and management changes to prevent accidents. CATS adopted a general structure and management model resulting from a long line of development in the United Kingdom, the Netherlands and France in a range of industries (the I-Risk, ARAMIS and WORM projects cited above). The general structure of the CATS model will be described in the second part of this first chapter (Section 1.2), and its management model and its limitations will be explained and discussed in Chapter 2 of this thesis.

The general structure of the CATS model and its management part had been decided before I was taken on to work on this thesis. I participated in the further development and implementation of that CATS model, but modifications to it within the time scale of the project could only be partial. In particular, the strong emphasis of CATS on quantification led to a number of restrictions in what was eventually modelled in that management section of the model. The main advantage of a quantification approach to modelling is that it offers the numerical probability of error occurrence, and the relative influence of individual elements in the model on it, as an output. But, there is often a strong requirement imposed by this need for quantification to provide a probabilistic risk assessment model. Some influences (especially the organisational and to a lesser extent the human) are extremely complicated in their causal relations and are therefore difficult to represent quantitatively at the current stage of modelling development. There is also a severe limit on the availability of data to quantify the causal relationships which are modelled. The result is that these factors either have to be very limited in their definition in order to quantify them in numerical units or they have to be left out of (or at least unquantified in) the model. As will be shown in Chapters 2 and 3 this was also the case with CATS, so that its model does not, as yet, tell the whole story of influences at that human and management levels, because important influences have been excluded and so lost.

It was therefore decided that this thesis would take a step back from the CATS project, in which I had participated as part of the modelling team, and would re-examine the place and role of the human and management models and their quantifications in a more fundamental way, in order to see what solutions might be proposed in the longer term. The general question for this thesis was, therefore, formulated as:

Is it possible to develop a safety management model which can link with the human and technical factors as modelled in CATS, or compatible with it, in such a way that it lends itself to quantification of the contribution of those management factors to the risk?

One constraint on the work was that any recommendations that it would make should still potentially be able to be incorporated into the CATS general model structure. Radically different approaches were therefore not suitable for exploration and development.

1.1 Aviation as a complex, hierarchical system

Rasmussen (1997) has used multi-levels to describe the complexity of a socio-technical system involved in the control of safety, as shown in Figure 1-2. At the top, society seeks to control safety through the legal system and sets rules for acceptable human and societal conditions and consequences. Every company has to interpret and implement these rules by setting company goals, choosing suitable risk control measures, and deploying resources. Inside the company, management has to put the risk control measures in place and keep them working effectively by using those suitable and sufficient resources and controls for ensuring that the measures work for their full life cycle. Individuals who are at the bottom level of the safety performance chain are required to carry out a series of actions as specified by company goals, procedures and rules, to keep the working process preferably within the defined operational boundary and at least within the safe limits of operation, and prevent any failure events before the accident happen. They also have to cope adequately with any situations occurring that have not been anticipated or planned for in those goals, procedures and rules. Rasmussen’s model also shows that the different levels of this hierarchy have been studied and modelled by different research disciplines and are subject to different dynamics from environmental stressors. This goes a long way to explaining why the models developed at each level are often difficult to link to each other.

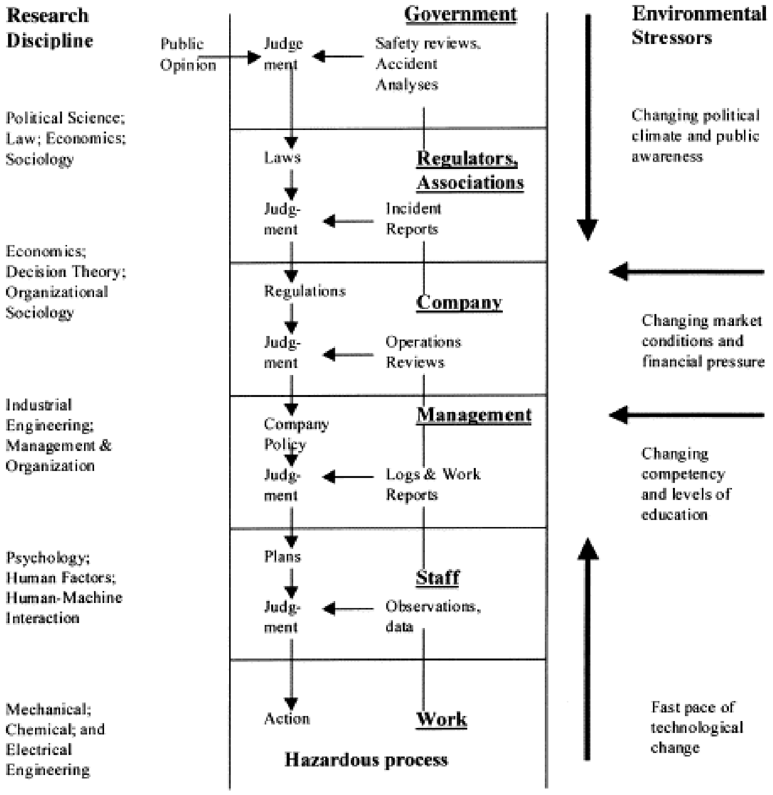

Research Discipline
Political Science; Law; Economics; Sociology
Economics; Decision Theory; Organizational Sociology
Industrial Engineering; Management & Organization
Psychology; Human Factors; Human-Machine Interaction
Mechanical; Chemical; and Electrical Engineering
Public Opinion
Government
Judge ment
Safety reviews, Accident Analyses
Laws
Regulators, Associations
Judg-ment
Incident Reports
Regulations
Company
Judg-ment
Operations Reviews
Company Policy
Management
Judg-ment
Logs & Work Reports
Plans
Staff
Judg-ment
Observations, data
Action
Work
Hazardous process
Environmental Stressors
Changing political climate and public awareness
Changing market conditions and financial pressure
Changing competency and levels of education
Fast pace of technological change

Figure 1-2 The socio-technical system involved in risk management (Rasmussen, 1997)

As it involves multiple stakeholders to provide a variety of services to the primary process of a flight from A to B, the aviation system is a prime example of Rasmussen's hierarchy. It is a complex socio-technical system, as shown in Figure 1-3. The primary process of a flight is executed by the flight crew and aircraft operations providing a particular mix of those that directly control the activities and hence the hazards (the dashed box in Figure 1-3). The supporting processes are provided by airlines, air navigation services providers, aircraft manufacturers and providers, aerodrome operators, maintenance organizations, etc. Airlines are responsible for ensuring that the tasks can be carried out safely by their flight crew and aircraft, and that the hardware and software used in them is in an effective and safe operational state. Air traffic controllers have to support flight crew to maintain sufficient vertical, lateral, and longitudinal separations to manage the risk of collision and provide additional flight information such as weather, navigation information and NOTAMs[1] to assist pilots operating in the airspace. Airports also serve a key safety role in transportation of people and goods in regional, national, and international commerce. To keep the system safe, airports have to maintain a good quality surface of runway and taxiways for aircraft to land and take-off. Big airports also provide reliable navigation and communication aids to the flight crew and direction and information to taxiing aircraft and to airport vehicles by operational guidance signs. A sound aircraft (ATC and airport) maintenance system is necessary to support the continuous airworthiness of these technical systems and maintain them for a safe and operationally efficient flight during its designed life. In addition, aircraft manufacturers also play a key role, by providing the aircraft, which need to be not only

[1] NOTAMs (Notices to Airmen) alert aircraft pilots to any hazards en route or at a specific location, e.g. closed runways, inoperable radio navigational aids, notifications of runway/taxiway/apron status with respect to snow, ice and standing water, etc.

structurally safe, but also to be designed to support the flight crew in all situations, from normal to extreme emergency.

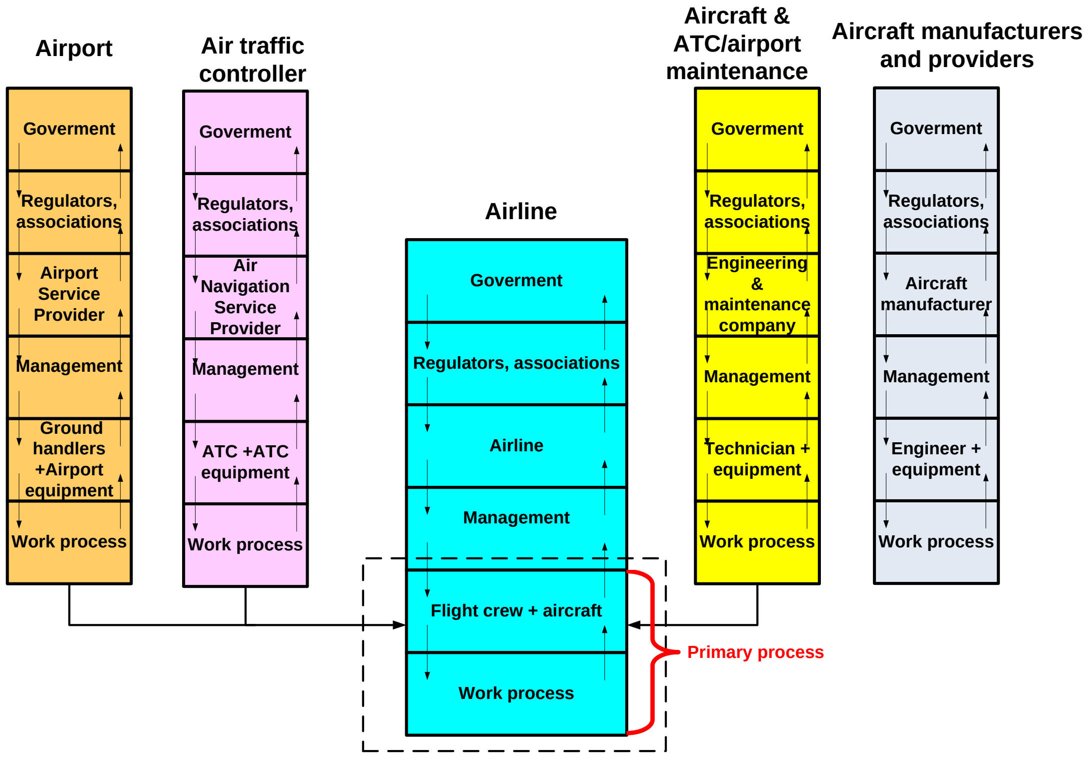


Figure 1-3 The socio-technical system involved in risk management in aviation (adapted from Rasmussen, 1997)

Each stakeholder has a system across hierarchical levels from government down to their individual work processes in the control of safety. The overall safety relies not only on each individual responsibility and contribution to safety but also how those systems interrelate and are incorporated within an overall safety management system and safety culture. Since aircraft and flight crew have the most central roles of execution of the primary process of a flight and of dealing on line with the risks arising in them, it was decided, for reasons of manageability, to limit the research scope of this thesis to the safety management influences related to the performance of flight crew and aircraft in the primary process. The assumption is that, if the modelling of this aspect can be developed satisfactorily, the other management areas from Figure 1-3 can be filled in using the same approach.

1.2 The general structure of the CATS model

The technical model of CATS is based on the combination of three modelling techniques: Event Sequence Diagrams (ESD), Fault Trees (FT) and Bayesian Belief Nets (BBNs) (Figure 1-4). ESDs delineate the possible accident scenarios. FTs describe the events, conditions and causes of the scenarios or barrier failures. Each cause of a barrier failure in an FT is a base event. The base events of the fault trees include events representing technical failures and events representing failures of human reliability, for instance “autopilot incorrectly used by flight crew”. Human operator models are attached to the fault trees, represented by BBNs, wherever humans are involved.

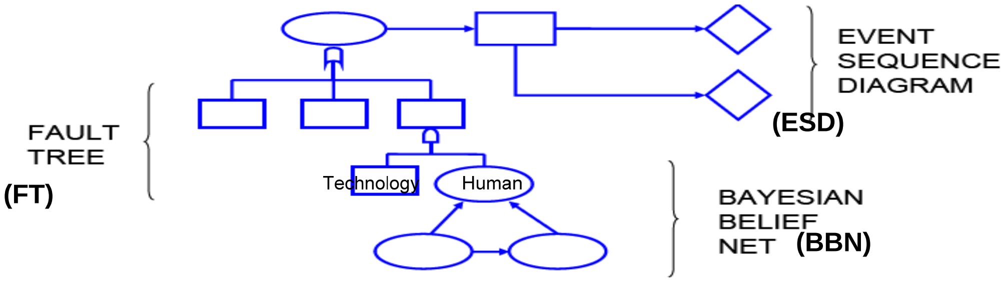


Figure 1-4 The basic constituents of CATS

The CATS model is intended to provide a comprehensive model of all of the relevant levels in Rasmussen's hierarchy, in order to support the identification and implementation of risk control measures at all levels, to reduce aviation risks and improve safety (Ale et al., 2006). It is a bottom-up approach to modelling, the backbone of which consists of 33 generic accident scenarios represented as Event Sequence Diagrams (ESDs), formulated at the level of the primary flight process. Figure 1-5 shows an example of a complete ESD of "uncontrolled collision with ground due to inappropriate landing roll during landing". This ESD describes the scenario in which a touchdown is made with a correct speed and sink rate, but due to an action by the crew during the landing roll, control of the aircraft is (temporarily) lost or maximum braking is not achieved. Along each path, pivotal events are identified as either occurring or not occurring with paths leading to different end states (i.e. runway veer-off, runway overrun, aircraft continues landing roll). Each path through the flowchart can be considered as a scenario.

All the ESDs are linked to each other as a set of challenges that have to be faced during a flight, from taxi through take-off, climb, en-route, approach landing and final taxi. Whether the flight survives depends on how the flight crew and aircraft system copes with the hazards. The event representation in ESDs is usually kept broad and generic to portray the progress of events over time. FTs are developed more elaborately to identify combinations of technical component failures and/or human errors that can lead to an undesired event identified in the ESD. There are separate fault trees for each event in each accident scenario. Figure 1-6 shows the FTs for the event of "failure to achieve maximum braking" in Figure 1-5.

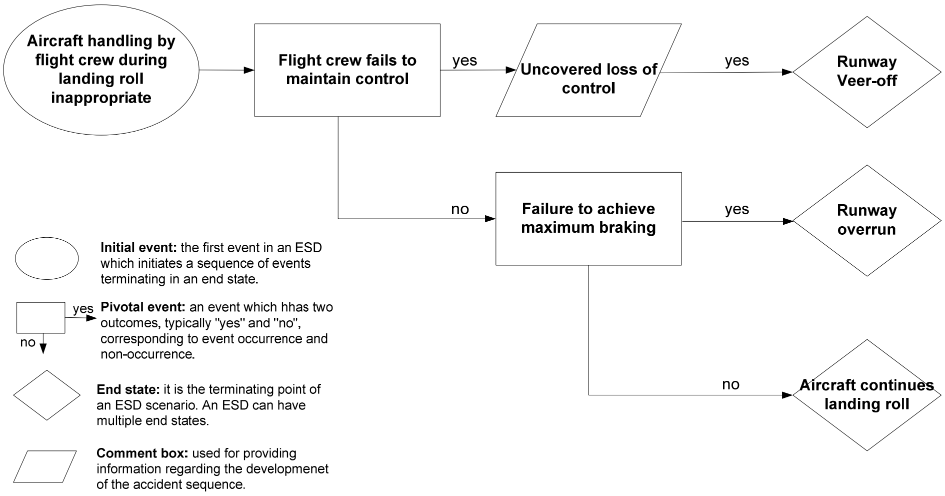

Aircraft handling by flight crew during landing roll inappropriate
Flight crew fails to maintain control
yes
Uncovered loss of control
yes
Runway Veer-off
no
Failure to achieve maximum braking
yes
Runway overrun
no
Aircraft continues landing roll
Initial event: the first event in an ESD which initiates a sequence of events terminating in an end state.
yes
no
Pivotal event: an event which hhas 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 developmenet of the accident sequence.

Figure 1-5 ESD of uncontrolled collision with ground due to inappropriate landing roll during landing

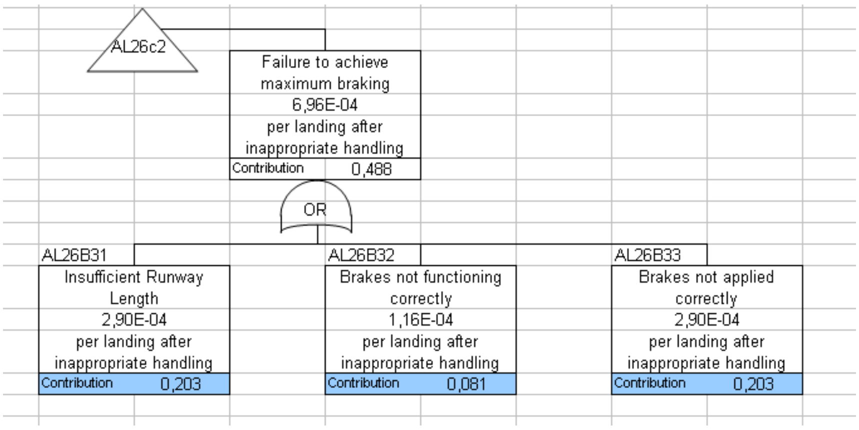

AL26c2
Failure to achieve maximum braking
6,96E-04
per landing after inappropriate handling
Contribution 0,488
OR
AL26B31
Insufficient Runway Length
2,90E-04
per landing after inappropriate handling
Contribution 0,203
AL26B32
Brakes not functioning correctly
1,16E-04
per landing after inappropriate handling
Contribution 0,081
AL26B33
Brakes not applied correctly
2,90E-04
per landing after inappropriate handling
Contribution 0,203

Figure 1-6 FT of uncontrolled collision with ground due to inappropriate landing roll during landing (part of)

FTs are usually constructed from the analysis of accident descriptions. This analysis is performed by dissecting these accident histories one by one to find potential causes of events already in the causal chain towards a pivotal event in the ESD. This continues until no new events (the failure of an identifiable technical system or a human action) can be established from data.

Human operator models are attached to the fault trees wherever humans are involved in the fault tree events. Separate models have been developed for flight crew, air traffic controllers, and maintenance technicians using the concept of the Performance Shaping Factors (PSF), which will be explained further in Chapter 3, to deal with human factors. These PSFs were envisaged as the way in which management processes would influence the probability of failure of the human at each point. In the final version of CATS it was envisaged that the technical failures would also be linked to the relevant processes of management influence, seen as the management processes of the hardware life cycle from design, through installation, use and maintenance to replacement (see Chapter 4), but this stage has not yet been reached due to delays in funding. These PSF models are represented as Bayesian Belief Networks (BBNs). The whole structure of the CATS model as described up to this point is shown in Figure 1-7.

Figure 1-8 shows the list of six PSFs envisaged by CATS for the human factors model. These are intended to represent all of the management processes which deliver the attribute named in each ellipse. In principle each of the six was envisaged as potentially relevant to all human failure events, but as each event was modelled a specific choice of the most relevant could be made. The derivation of the six categories will be discussed in Chapter 2.

This was the conceptualisation of the whole CATS model structure when I joined the modelling team. Work was carried out by that team according to this formulation, but suffice it to say here that the actual practice in CATS did not fill in this generic model anywhere near completely. Only limited modelling of the six PSFs has been made up to now, and this was driven very strongly by which of the influences could be quantified with existing data and the limited use of expert elicitation based on BBNs (see Chapter 5).

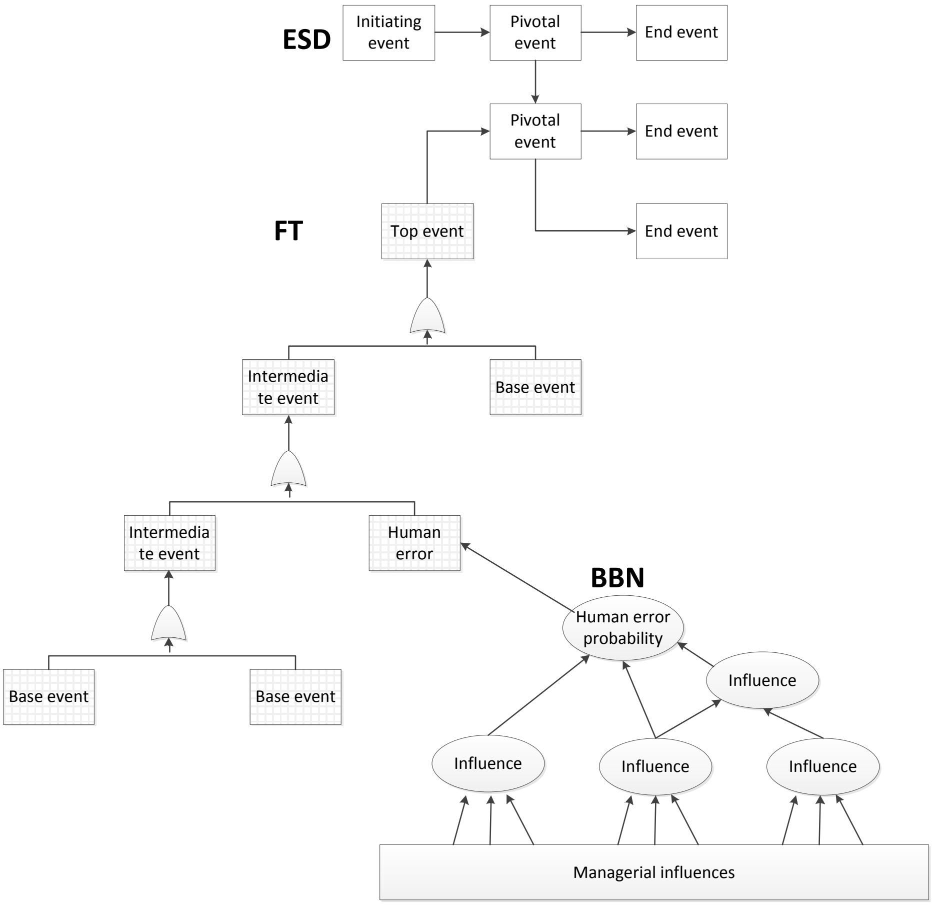


Figure 1-7 The basic constituents of CATS

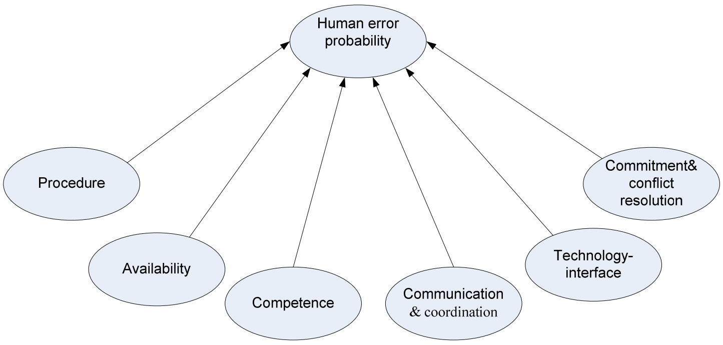


Figure 1-8 Scheme of human performance model in CATS

1.3 Safety management influences on risk

Safety management influences on risk are still generally ill-understood. This particularly holds true for aviation, which is complex in nature and operates in a highly dynamic environment. CATS is one of the first attempts to make management influences in aviation explicit (Ale et al., 2009; Roelen 2008).

In other industries in the past a number of probabilistic risk assessment (PRA) tools have tried to incorporate the management influences and organizational factors in their probabilistic risk models. The amount of similarities between the different frameworks with respect to the organizational models or sets of factors is rather limited. Øien (2001) concludes that “this is no big surprise" since even for pure classifications of organizational factors there are significant differences (Wilpert, 2000). The safety management influences are often treated as simple influences on the failures (PSFs) in the causal chain and therefore modelled in simple failure terms to fit the PRA. This does not meet the full objectives of CATS, which had as goal to show managers (and inspectors) what actions they could take to influence safety. Therefore the decision was made in CATS to follow earlier modelling in the I-Risk, ARAMIS and WORM projects, cited above. These see safety management influences as a set of actions which can be taken by managers (a management process) to influence the problems identified in the human and technical failures by providing resources or controls for the barrier design and operation.

General safety management theories and management standards (such as the ICAO safety management standard (ICAO Annex 6, 2006), or more generic ISO and national SMS standards such as BS18001 (OHSAS 18001, 1999) are usually developed from top-down, general policy being developed into a number of generic management activities, such as risk assessment, monitoring and learning. The ISO standards have tended to be independent of the substance matter context of a given organization, or in other words independent of what you are managing in the operations and technical systems. Rasmussen (1997) criticized the fact that there have been few studies emphasizing the vertical interaction between the different levels in his model (Figure 1-2 and Figure 1-3). This means that there is a problem in incorporating these theoretical management models as a tool for resolving issues related to human performance or technical failure in the accident analysis, namely that they do not connect to the current practice of safety data collection and analysis.

The aim of CATS was to fill these gaps. In doing so it decided early on in its history to draw on a management approach developed in the Netherlands, through a series of projects (I-Risk (Bellamy et al., 1999), ARAMIS (Hourtolou & Salvi, 2003; Hale & Guldenmund, 2004) and WORM (RIVM, 2008; Ale et al., 2007) which have modelled safety management as *a series of functions* which have to be carried out in order to prevent the failure events. This means that the processes of providing e.g. training or procedures to the staff level are modelled in some detail. The implication of doing so is that the management factors included in the model would therefore be able to be seen by managers and regulators as actions they could take and hence they could have the leverage to influence the probabilities of the failure at the operational level.

However, as stated briefly above, and as will be explained in Chapter 2, things did not turn out exactly as envisaged and planned. It was therefore decided, for this PhD to take a step back and examine the assumptions made in arriving at the original intentions for management modelling in CATS and the decisions made in putting them into practice. This involved revisiting the development of the Dutch model used there and critically examining its structure, the assumptions built into that structure, both intrinsically and in relation to other relevant models of safety management appropriate to aviation in order to answer the question whether it has the potential to model all aspects that other frameworks model? This is a step towards an examination of its concurrent validity.

Part of the critical analysis is to look at the issue of the connection between the management model and the human factors and technology failure models at the lower level in Rasmussen's hierarchy. In particular this thesis will concentrate on the links with the human factors models available in aviation, since this has been the emphasis in CATS. It will give less attention to the links to the technology failure, which have not yet been linked in CATS to the safety management system.

As has been indicated, one of the main driving forces of the CATS model has been the requirement to quantify influences. This raises the issue of what data are available within the aviation industry of failures and errors/incidents/accidents classified under human factors models, and do these data models offer a possibility to link with the management failures? To the extent that they do not, we are left with the only other potential source of quantification, namely expert judgement, used in CATS in relation to the BBNs. As described in Chapter 5, this expert elicitation was complex when based on the BBNs and it was decided to experiment with a simpler form of elicitation based on paired comparisons to assess its potential for quantification.

1.4 Scope, approaches and outline of this thesis

The conceptual model for the topics of our research is organized as in Figure 1-9.

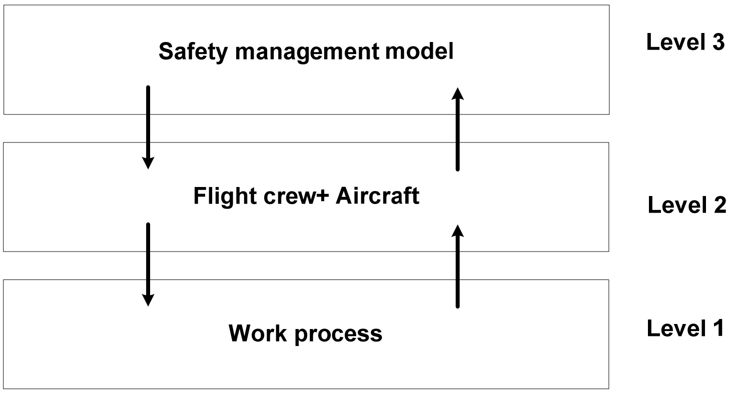


Figure 1-9 Hierarchical model

The overall research question is:

Is it possible to develop a model of safety management influences which can link with the human and technical factors as modelled in CATS, or compatible with it, in such a way that it lends itself to quantification of the contribution of those management factors to the risk?

This PhD, therefore, takes the choice of the Dutch model as a given for this research. It begins by describing in Chapter 2 the development of that Dutch management model and the dilemmas met and resolved in that development process, and by (re)assessing its suitability, by comparing it with the other relevant aviation safety management models, airline practice in safety management and relevant management standards applicable to aviation, in order to see if the Dutch management model does, or can in theory, encompass the insights from these other sources, in other words whether it is as complete and appropriate as possible from a scientific and practical point of view to be applied to aviation. This is level 3 in Figure 1-9. Any areas of improvement will be carried forward to Chapter 7. Since the findings in Chapter 2 were not finalised before the end of the current phase of work on the CATS model, we were unable to modify the CATS management model within the CATS' time frame. The management modelling in CATS is therefore of limited scope.

Next we review in chapters 3 and 4 the human and technical factors in relevant aviation models from bottom up, elaborating in more detail levels 1 & 2 in Figure 1-9. In particular, since the experience is always that human factors are harder to model than technical factors, this examines what are the underlying causes that contribute to human performance deficiencies that affect the risk of aviation. This is dealt with in Chapter 3.

The underlying mechanisms which contribute to aircraft (or other hardware in ATC and airports) deficiencies found in the work process are more straightforward to model, compared to human performance. Aircraft (ATC and airport) functional status is determined by design, manufacture, maintenance, and human operation[2]. Since there are many studies for these aspects in the field, we will only briefly discuss each of the aspects for completeness sake in relation to the management of design and maintenance, but will not go into much detail. This is the subject of Chapter 4. The further development of the safety management model specific to these technical aspects falls outside the scope of this thesis, but it is believed that the approach used could be the same as that proposed here.

In chapters 3 and 4 we will also analyse whether the Dutch management model can support (i.e. link to) the control functions analyzed in the two operational levels. Can level 3 be linked to level 2 and through to level 1 using the human factors and technical models available? Are these models mappable onto the management factors? Where they do not match, additional functions will be proposed and discussed.

We then turn to quantification and present in chapters 5 and 6 a critical discussion of the CATS quantification method linked to the BBNs (Chapter 5) and some experiments in developing a supplementary method to get round the complexity of the expert elicitation that the BBN method involves (Chapter 6).

Based on all of these analyses, Chapter 7 puts together the findings of the preceding chapters (chapters 2 to 6) into a new integrated model, which is proposed for further testing and eventual use in an extension of CATS. In particular the new insights for the Dutch management model found in Chapter 2 will be taken on board. The discussion will also indicate whether these changes affect the proposed method of quantification.

[2] Operating any system beyond the design limits causes irrecoverable damage to the components due to unintended exposure to overload, overheating, corrosion and wear.

2 Safety management model

As the introduction in the previous chapter indicated, before this thesis started a decision had been made within the CATS project to build upon an existing safety management model which had been developed in earlier studies in the Netherlands. This model has a long development history, through projects called I-Risk (Bellamy et al., 1999), ARAMIS (Hourtolou & Salvi, 2003; Hale & Guldenmund, 2004), and WORM (RIVM, 2008; Ale et al., 2007), in which it has been tested a number of times. As management model it therefore has a respectable pedigree and has been tested by peer review and the approval of the organisations of regulators (e.g. the British Health and Safety Executive, and the Social Affairs and Employment Ministry of the Netherlands and INERIS in France) and of participating industrial companies, particularly in the chemical process industry. It had also been used as a teaching model on post-graduate courses for safety professionals in a wide range of industries. The assumption made in the CATS project was that it would also be suitable for the aviation industry. This was a working hypothesis, which was not tested in detail in CATS. It was therefore decided that this thesis would put the model used to such a test, so far as this was possible. This would provide some more validation to the model used, at least at a concurrent and face validity level. The focus of this chapter is therefore to answer the following questions:

- *Is the existing Dutch model of safety management suitable for application in aviation?*
- *Is the Dutch model of safety management complete from a scientific and practical point of view when compared to other formulations of safety management system applied in the aviation field?*

Figure 2-1 shows the structure for reviewing and validating a SMS for this purpose. The chapter (Section 2.1) starts with a review of the history of the development of the Dutch safety management models. This section will identify the crucial assumptions made in that development and how these were related to the purpose for which the model was designed and to the data available to quantify the model. At the end of this section the experience of using the model, as then developed in the CATS project, will be presented in order to indicate the issues still remaining to be resolved.

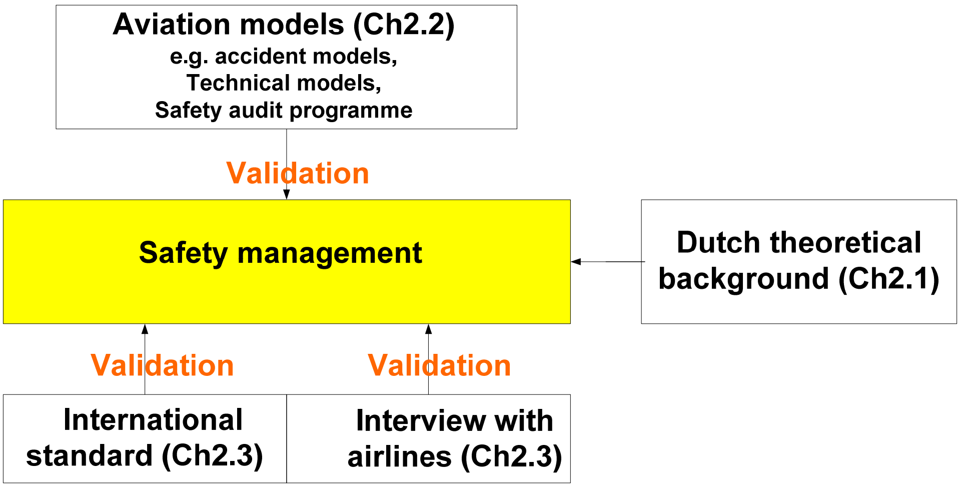


Figure 2-1 Structure of Chapter 2

The second section of this chapter (Section 2.2) is to test and compare the Dutch model as devised with other relevant models in order to validate the model and indicate how it needs to improve. This analysis will concentrate on models which have been developed for, or applied to aviation. Models from other areas and industries will be mentioned in passing where relevant. In this section, the different sources from aviation-related accident models, technical models, and safety audit programmes will be mapped onto the Dutch model to see if all can be accommodated.

Section 2.3 will compare the Dutch management model with international safety management standards, and report the findings of interviews with safety professionals within airlines to check whether what the companies do in practice in the name of safety management fits the model we are using. Together these will form a (partial - concurrent) validation of the model. The final section of the current chapter summarizes the findings from these four perspectives and gives new insights for the current Dutch safety management model, in particular, for a new formulation later in Chapter 7.

2.1 Dutch safety management models

Work began in the Netherlands, at Delft University of Technology, in the early 1990s on the modelling of safety management systems (Hale et al., 1994). This work was linked from 1996 to studies which had started in the UK with the development of Manager (Technica, 1988) and continued in the development of the PRIMA model and audit system (Bellamy et al., 1993; Hurst et al., 1996). This was achieved in a project called I-Risk (Bellamy et al., 1999). The combined model, called in this thesis “the Dutch model”[3], received in I-Risk project its characteristic form, which portrays safety management as the provision of the resources and controls which management provides for front line workers to perform their task of controlling risk, through their own actions and through use of hardware and software. This section describes the main steps which followed, as the model was adapted through different applications, having different objectives and application fields.

2.1.1 I-Risk[4]

I-Risk (Bellamy et al., 1999) was a European-funded collaboration whose main partners were the British Health and Safety Executive and the Dutch Ministry of Social Affairs and Employment. They brought together two strands of modelling, from UK and the Netherlands. The resulting Dutch model was originally designed to link the performance of the SMS to technical failures in chemical installations. The technical model in I-Risk consisted of event trees and fault trees, modelled in the classical QRA form with initiating and base events. These define the safety critical parts of activities, including technical failures, unavailabilities of safety features or human errors in the primary (operations, emergency, inspection, maintenance and modifications) processes. The whole management system linked with the technical model through these critical parts of activities in the primary processes as defined by the technical model and its parameters. The technical model calculated the frequency of

[3] The main architects of this model were Linda Bellamy who had worked in a number of consultancy companies as one of the initiators of Manager and subsequent developer of PRIMA and Andrew Hale and his colleagues from Delft University of Technology who had worked on the SMS modelling there. Whilst during the initial collaboration the former had worked from UK, she subsequently emigrated to the Netherlands and worked on later projects from there, justifying the title of “Dutch model”.

[4] Standing for ‘Integrated Risk’.

occurrence of base events in terms of 10 parameters which were used to quantify the Loss of Containment[5] risks:

(*f*i) frequency of initiating events,
(ls, lo) failure rate of unmonitored or monitored components,
(*T*) time between testing,
(*Q*M1) error in test and repair,
(*Q*M2) failure to detect and recover previous error in test and repair,
(*f*m) frequency of routine maintenance,
(*T*m) duration of routine maintenance,
(*T*R) duration of repair,
(*Q*o1) probability of error in operations or emergency,
(*Q*o2) probability of not detecting and recovering error,

A company exercises control over major hazards by managing a number of "primary business activities" (i.e. operations, emergency operations, inspection and testing, maintenance, and modifications). The SMS is modelled to show how the resources and controls are delivered to the primary business activities, which directly influence the technical risk parameters. These primary business activities were modelled as being controlled proactively by allocating suitable resources to them and by imposing suitable criteria and controls on the way in which they are carried out by the safety management system. Hence, safety management was seen as the provision of those systematic *resources and controls* of the risks which were derived from the risk analysis of the plant as reflected in a technical model.

The supply of these resources and controls is achieved by secondary management processes, which were called *delivery systems* in I-Risk. The delivery systems were originally developed from a combination of the "*management control and monitoring loops"* derived from the PRIMA audit with the frameworks of the systematic logic imposed by the "SADT technique" (Structured Analysis and Design Technique) (Hale et al., 1997).

This SADT technique (shown in Figure 2-2) is a control theory which consists of an activity transformation controlled by three aspects together to produce the outputs (O): the input (I), the resources (R) and the controls/criteria (C) (Hale et al., 1997). The "inputs" in an SMS analysis will be the transformation outputs of earlier steps in the activity. This can be data or consumables that are needed by the activity. "Outputs" are the products of the process. They can be desired products, by-products or unwanted outcomes. Management make plans to control the output to meet defined safety criteria. "Controls/ criteria" include all laws, regulations, and standards and procedures that are used to direct and judge the performance of an activity, to ensure that the outputs meet the objectives of the transformation. "Resources" identify the means, e.g. tools, equipments, and operating individuals, used to accomplish the activity. Resources and control/criteria influence the execution of an activity but are not themselves transformed.

[5] Loss of containment means a discontinuity or loss of the pressure boundary between the hazardous substance and the environment, resulting in a release of hazardous substances

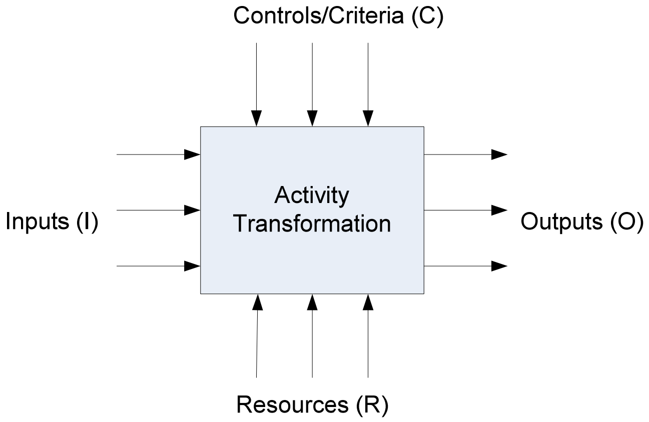


Figure 2-2 Structured Analysis and Design Technique (SADT)

The "resources" and the "control/criteria" were grouped in the I-Risk model under the following **delivery systems** to produce the desired outcomes.

1) **procedures, goals and plans** for safety critical tasks,
2) **availability** of people to carry out these tasks,
3) their **competence** to do so,
4) their **commitment** to, or **motivation** for safe execution of the tasks (= their choice to apply their competence safely rather than pursuing other goals),
5) the necessary **communication and coordination** between several people or groups needing to collaborate for a given safety critical task,
6) **conflict resolution** at an organization level of any conflicts between safety and other goals related to that task (related closely to 4 above),
7) the good functioning of **equipment & spares** which are installed during maintenance which covers both the correctness of the equipment and spares for their use, and the availability of them when and where needed to carry out the activities,
8) the **ergonomics** of all aspects of the plant which are used/operated by operations, inspection or maintenance.

7) and 8) should be closely related and provided by the same design, selection, installation, use and maintenance cycle paying attention to both sets of criteria. In I-Risk there was no delivery system for the original design of the plant or installation being modelled, since the model treated the plant design as a given[6].

The primary delivery systems were supplemented by the inclusion of management control and monitoring loops derived from the PRIMA audit (Bellamy et al., 1993, Hurst et al., 1996). That was where the idea of the delivery systems having two higher level management functions of "**feedback**" and "**learning loops**" came from.

The interface between the 8 primary delivery systems and the 10 parameters of the base events was given by a detailed table (Table 2-1) showing the cross-tabulation of which management influences were relevant for each parameter. The values shown in the table are the proportional importance weightings of the assessments of each of the delivery systems in influencing the ten technical parameters listed along the top. The weightings of all of the delivery systems for any one parameter add to 1.0, which is the total effect of all management

[6] This was because the modelling tool was being developed to assist risk analysis of existing plants and not as support for the design process itself. Any change to the design would have meant changing the basic parameters of the model, which could not be done dynamically in an essentially static model.

factors on that technical parameter. Some deliveries have no effect on a given parameter, e.g. competence on time between tests (*T*) and on frequency of routine maintenance (*fm*), or spares and tools on frequency of initiating events (*fi*). On the other hand some have relatively heavy weightings, like commitment and conflict resolution on time between tests (*T*), etc.

The complexity of applying this table was that, in practice, each of the parameter is affected by many management tasks within delivery systems. In total, there are many hundreds of tasks which need to be assessed to get a full picture of a company's management system and quantified to show their influence on the base events. Although the project did use systematic expert judgment to try to overcome this problem (Hale et al., 1999, 2000), the quantification for management modelling had not been worked out by the time the project ended in 1999, due to its complexity in classifying these influences into manageable links to the assessments of the management system. However, the combination of the two techniques (SADT& management control and monitoring loops) in one modelling technique was felt to be an important development, which forms the basis of the later development of the Dutch SMS approach used in later projects.

Table 2-1 Delivery systems that affect basic event parameters

	Qo1	*Qo2*	*Qm1*	*Qm2*	*fi*	λ	*T*	*fm*	*Tr*	*Tm*	*Total*
Availability	0.06	0.05	0.08	0.05	0.1	0.08	0.05	0.05	0.12	0.12	0.76
Commitment	0.15	0.14	0.19	0.13	0.2	0.12	0.24	0.21	0.07	0.08	1.53
Communication	0.07	0.05	0.06	0.05	0.1	0.12	0.14	0.16	0.21	0.21	1.17
Competence	0.16	0.21	0.14	0.22	0.1	0.08	0	0	0.09	0.08	1.08
Conflict resolution	0.18	0.21	0.14	0.18	0.1	0.08	0.28	0.32	0.1	0.12	1.71
Interface	0.2	0.2	0.08	0.18	0	0.08	0.05	0.05	0.19	0.17	1.2
Procedures	0.18	0.14	0.17	0.15	0.4	0.16	0.19	0.16	0.1	0.08	1.73
Spares & tools	0	0	0.14	0.04	0	0.28	0.05	0.05	0.12	0.14	0.82

2.1.2 ARAMIS

The Delft team collaborated after the end of the I-Risk project on another EU funded project called ARAMIS led by the French government research organisation for environment and safety (INERIS) (Hourtolou & Salvi, 2003; Hale & Guldenmund, 2004). This project aimed to develop models covering not only the preventive side of major hazard control, but also the modelling of the impact on those living around the plant and on the environment. Its relevance to this thesis is in the further developments of the management model from I-Risk.

In the ARAMIS project the decision was made to use bow-tie diagrams (Figure 2-3) as a modelling framework showing scenario development and to assess the potential of the concept of safety barriers (Haddon, 1973) as a way of formulating the link between technical, human, and organizational factors. This replaced the link via the parameters of the base events. The barrier concept presumes a hazard (a dangerous source of energy) and a target (a vulnerable object like humans), which is protected by the barriers. Figure 2-4 shows Haddon's hazard-barrier-target model.

The definition of the barrier developed in ARAMIS is hardware and/or behaviour which detects, diagnoses, and acts (actively or passively) to control the hazard. The classification in ARAMIS was based on whether the barriers were permanent or temporary, whether they work passively once put in place or have to be activated, and whether they have a pure safety (barrier) function or one which also forms part of the normal process control (column 1 in Table 2.2). In the case where the barrier is operated by human being, the distinction is made between whether the behaviour required is at the skill, rule, or knowledge based level (column 4). This distinction resulted in 11 types of categories of the barriers (Table 2.2).

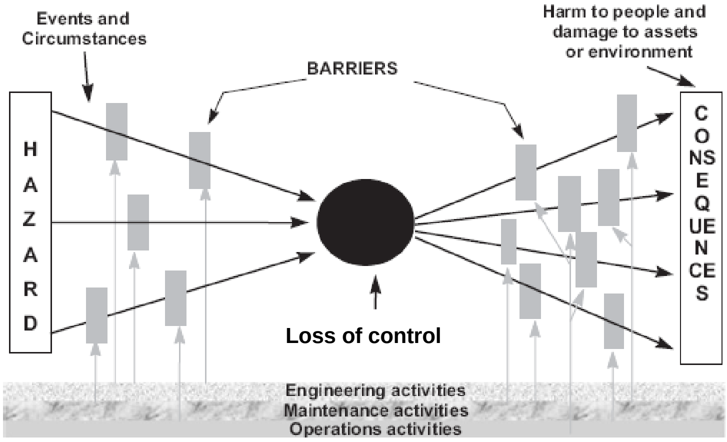


Figure 2-3 Bow-tie model

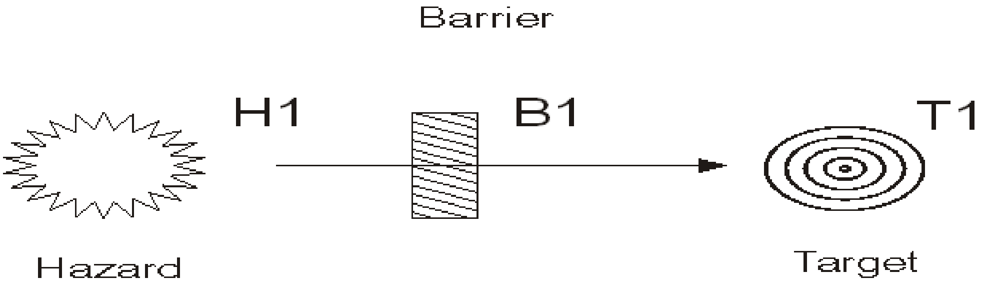


Figure 2-4 Barrier model

In order to fulfil their functions as barriers on a continuing basis, the management system has to manage their life cycle, which is conceptualized as being to design them, put them in place, ensure their use by the operators, maintain or modify them to retain or restore their functioning and replace or improve them where necessary. With the insight into the management processes which deliver controls and resources to the barriers' life cycle to ensure its proper functioning, came the realization that the management safety task could best be seen as managing the life cycle of the barriers (Figure 2-5). In other words, the steps ("design", "install", "use", "maintain", and "improve") in the barrier life cycle are then tasks to be organized and carried out by the management system. This also led to a reformulation of the eight delivery systems in I-Risk into two categories. There were 2 delivery systems related to delivering "hardware" barriers or components of barriers and 5 delivery systems delivering "behaviour" as a barrier or part of a barrier.

Table 2-2 Classification of barriers in the ARAMIS safety management evaluation

	Barrier	Examples	Detect	Diagnose/ Activate	Act
1	*Permanent – passive –control*[7]	Wall of pipe, hose or tank; anti-corrosion paint; tank support; floating tank lid; flange connection; seals; viewing port in vessel	Passive	Passive	Hardware
2	*Permanent – passive –barrier*	Tank bund, dyke, drainage sump, railing, fence, blast wall, lightning conductor,	Passive	Passive	Hardware
3	*Temporary – passive* Put in place (and removed) by person	Barriers round repair work, blind flange over open pipe, helmet/gloves/safety shoes/goggles, inhibitor in mixture	Passive	Passive (human must put them in place)	Hardware
4	*Permanent – active*	Active corrosion protection, heating or cooling system, ventilation, system to maintain inert gas in equipment.	Passive	Built-in (may need activation by operator for certain process phases)	Hardware
5	*Activated – hardware on demand – barrier or control*	Pressure relief valve, interlock with "hard" logic, sprinkler installation, electro-mechanic pressure, temperature or level control	Hardware	Hardware	Hardware
6	*Activated – automated*	Programmable automated device, control system or shutdown system	Hardware	Software	Hardware
7	*Activated – manual* Human action triggered by active hardware detection(s)	Manual shutdown or adjustment in response to instrument reading or alarm, evacuation, donning breathing apparatus or calling fire brigade on alarm, action triggered by remote camera, drain valve, close/open (correct) valve	Hardware	Human (Skill-, Rule- or Knowledge-based)	Human/ remote control

[7] The difference between "control" and "safety barrier" follows from the terminology of the (MORT) methodology, (W.G. Johnson, "MORT - the Management Oversight & Risk Tree", SAN 821-2, February 1973). A control is a component that is necessary to perform the primary process, but which serves also to control hazards (e.g. a pipe wall, a level control), a barrier is a component that is installed solely to prevent or mitigate hazards (a tank bund, a pressure relief valve).

	Barrier	**Examples**	**Detect**	**Diagnose/ Activate**	**Act**
8	*Activated – warned* Human action based on passive warning	Donning personal protection equipment in danger area, refraining from smoking, keeping within white lines, opening labelled pipe, keeping out of prohibited areas	Hardware	Human (Rule-based)	Human
9	*Activated – assisted* Software presents diagnosis to the operator	Using an expert system	Hardware	Software – human (Rule- or Knowledge-based)	Human/ remote control
10	*Activated – procedural* Observation of local conditions not using instruments	(Correctly) follow start up/shutdown/batch process procedure, adjust setting of hardware, warn others to act or evacuate, (un)couple tanker from storage, empty & purge line before opening, drive tanker, lay down water curtain	Human	Human (Skill- or Rule-based)	Human/ remote control
11	*Activated – emergency* Ad-hoc observation of deviation + improvisation of response	Response to unexpected emergency, improvised jury-rig during maintenance, fight fire	Human	Human (Knowledge-based)	Human/ remote control

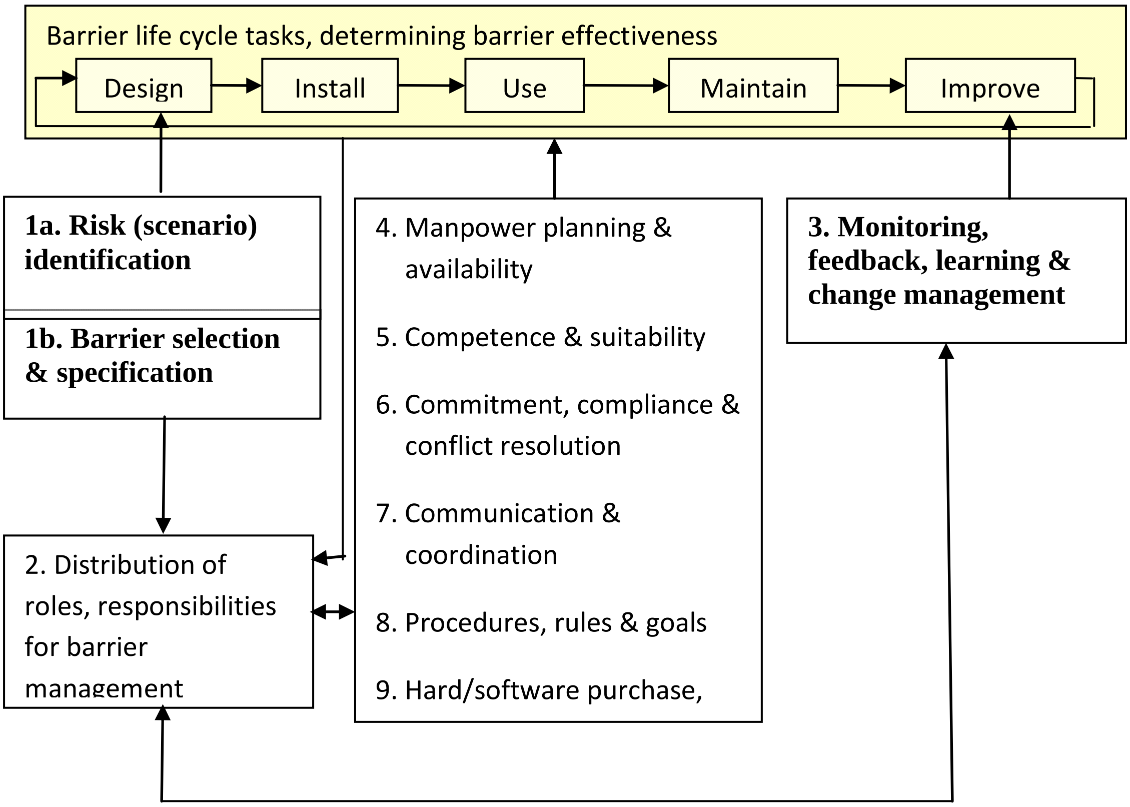

Barrier life cycle tasks, determining barrier effectiveness
Design
Install
Use
Maintain
Improve
1a. Risk (scenario) identification
1b. Barrier selection & specification
4. Manpower planning & availability
5. Competence & suitability
6. Commitment, compliance & conflict resolution
7. Communication & coordination
8. Procedures, rules & goals
9. Hard/software purchase,
3. Monitoring, feedback, learning & change management
2. Distribution of roles, responsibilities for barrier management

Figure 2-5 Model of ARAMIS with barrier life cycle and delivery systems

The two hardware related systems (equipment and spares) in I-Risk were reformulated as the hardware life cycle applying to any sort of hardware, whether that was the original plant design, the original barriers, or the spares and replacements installed during maintenance. This hardware life cycle was envisaged as being required to deliver not only the functionality of the hardware but its ergonomic interface determining its usability (the two original I-Risk conceptualisations of the delivery systems). The life cycle was split into two blocks to emphasise that these tasks are often performed by different department or even organizations:

1) **specification, design, installation and adjustment of the technology** whether operation equipment, spares, or tools (including the layout, labelling and design of the interface where it was to be used by people);
2) the **inspection and maintenance** of the technology to keep it working to specification.

The six behaviour related systems in I-Risk were reduced to five by combining “commitment” and “conflict resolution”, since both related to the choice between alternative behaviours where the requirement was not to choose other criteria above safety, one being at individual level and the other at a higher management level. The re-formulated delivery systems for human behaviour were:

3) **Procedures, plans, rules and goals** to specify the behaviour required of the human barrier element;
4) **Availability, manpower planning** of personnel to operate or be the barrier;
5) **Competence, suitability** of these people;
6) **Commitment, conflict resolution** (choice to deploy competence);
7) **Coordination, communication** during the direct barrier tasks.

The full definitions of what these delivery systems cover are shown in Appendix A. At the time of the ARAMIS project, it was not made very explicit that these five delivery systems were in fact workings out of the life cycle of human behaviour as barrier. That clarification only came later in the WORM project.

Apart from the delivery systems, which are all part of the life cycle of barriers, the safety management system also has to manage the processes of "**risk (scenario) identification, barrier selection and specification**", and the process of "**monitoring, feedback, learning and change management**" from experiences gained in all steps of the life cycle. These functions of management were included in the ARAMIS structure as shown in Figure 2-5 as well.

Each delivery system consists of a number of steps according to the well-known Plan-Do-Check-Adjust (PDCA cycle) of Deming quality cycle (Deming, 1968), which can be pictured as cycles of steps. For instance the delivery system of competence is shown in the block diagram in Figure 2-6. The other delivery systems are shown in the same form in Appendix A. Whether the company takes systematic and effective actions for each of the steps determines the effectiveness of the barrier functions.

It should be noted that all of these cycles (those within each delivery system with their Plan-Do-Check-Adjust cycle, or the barrier life cycle of Design-Install-Use-Maintain-Improve) have a similar structure, but they engage with each other at different hierarchical levels. So, in principle the whole model is full of these cycles. This is one of the things that was not clear to some of the participants in the later projects. The dynamic nature of these cycles also made it difficult for the mathematical modelling because the traditional approaches and those adopted for CATS cannot cope with that dynamic.

An audit technique was developed to assess the quality of each delivery system by scoring these steps using a five-point rating scale. When it was worked out into a model which tried to establish a valid link between the quantitative technical model and various types of barrier and its management control from delivery systems, it provide a clearer model than I-Risk. A successful audit technique (Hale & Guldenmund, 2004) was developed and trialled based on the model. That step fulfilled the main objective for the management model in ARAMIS. However, like in I-Risk the workload to apply this approach quantitatively to all barriers still proved to be large in a complex process and the quantification of the management effects on risk level remained a challenge which was not tackled by the end of the ARAMIS project.

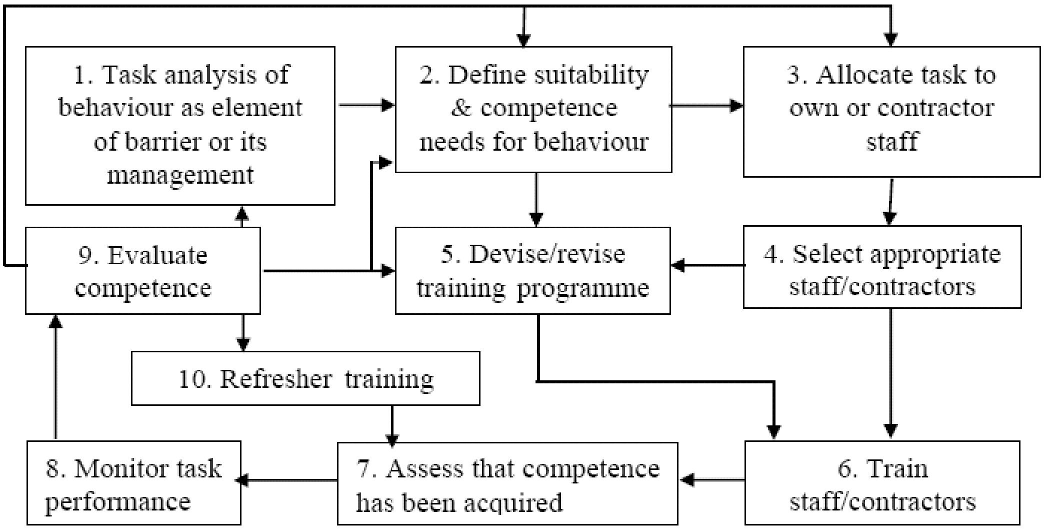

1. Task analysis of behaviour as element of barrier or its management
2. Define suitability & competence needs for behaviour
3. Allocate task to own or contractor staff
9. Evaluate competence
5. Devise/revise training programme
4. Select appropriate staff/contractors
10. Refresher training
8. Monitor task performance
7. Assess that competence has been acquired
6. Train staff/contractors

Figure 2-6 Block diagram of competence delivery

2.1.3 WORM

Essentially the same team as had worked on I-Risk was reconstituted towards the end of the work on ARAMIS, in order to work on a new Dutch programme related to the development of a causal model for the full range of occupational and process accidents reported to and investigated by the Dutch labour inspectorate, encompassing their causation and prevention. The project, called WORM (RIVM, 2008; Ale et al., 2007), adopted the use of barriers as the way of formulating risk controls, as in ARAMIS, and developed the conceptualisation of accidents as in the bow-tie model into an extended modelling of the scenarios of all accidents, not just those related to chemical releases and explosions. This resulted in a typology of accidents relating to different ways of losing control through such accident scenarios, which was captured in a software depiction called "Storybuilder" (Bellamy et al., 2008).

WORM also, therefore, identified failures of barriers as the key elements which could act as the linkage with the management system. Initially, the eight delivery systems from I-Risk were used, rather than their modified form derived from ARAMIS. However, as stated above, confusions arose among the researchers and coders because it was not clear exactly how the delivery systems should relate to the barriers. The concept of the life cycle of the barriers as it should be in Figure 2-5 had not at first been suitably translated into the WORM project, so it was not clear that the formulation used for ARAMIS already incorporated the life cycle in the detailed block diagrams. Eventually, as a compromise the decision was made to define directly four management tasks (to "provide", "use", "maintain" and "monitor" the barriers) in order to represent the barrier life cycle. This echoed the ARAMIS solution of setting "design", "install", "use", "maintain", and "improve" to the hardware barrier's life cycle. Table 2-3 shows the comparison between the steps of the life cycles in the two projects, with "provide" covering the steps from design to installation, "use" and "maintenance" being the same in both, and "monitoring" being equivalent to the modification and improvement step. But the last step did cause confusion for some analysts, who saw "monitoring" as a preliminary to "maintenance" and not to improvement.

Table 2-3 Barriers life cycle in ARAMIS &WORM projects

WORM	ARAMIS
Provide	Design
	Installation
Use	Use
Maintain	Maintain
Monitor	Improve

This formulation worked well for the technical barriers, and hence the decision was made in WORM to formulate all barriers as "technical or physical". Hence, for example, the skill of an operator in keeping his hands out of the way of sharp or moving parts of a machine was formulated as "sufficient distance" being the (physical) barrier which had to be provided and used, and the skill as being the competence which had to be delivered in order to use that barrier. This was a very different formulation to ARAMIS, where the human behaviour had been seen as part or all of the barrier – represented as the detection of the hazardous part of the machine, the diagnosis that the operator needed to keep his hands out of the way of the hazard and the action of keeping the distance. The decision in WORM led to the formulation shown in Figure 2-7, where the defences are purely physical.

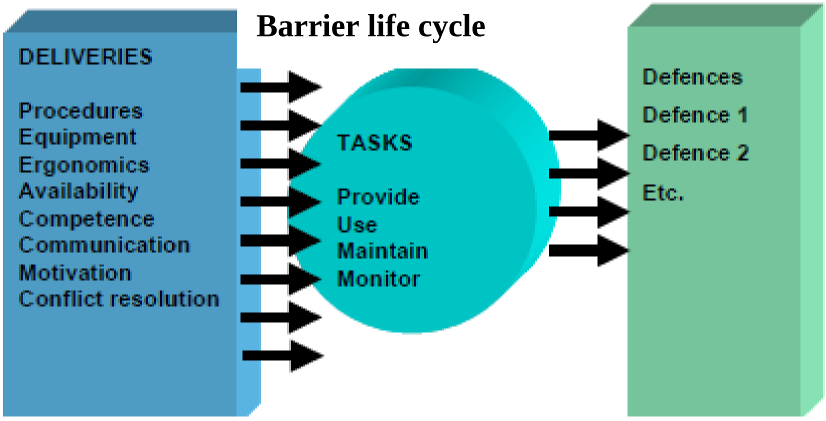


Figure 2-7 Model of WORM

The difference between these two formulations of barriers led to confusion in those past projects which needed to be resolved in any future application. The clarification of the need to be explicit about where the barrier life cycle fitted in the model was a contribution of WORM. Given the separation of the life cycle (as the separate four management tasks) from the delivery system, it was appropriate that WORM returned to the formulation of the two hardware delivery systems as the functioning of the hardware and the ergonomics or user-friendly nature of it. The delivery systems were treated very much as black boxes in WORM and not as the set of management tasks seen when the box was opened to reveal the block diagrams shown in Appendix A. In other words they were treated as a whole rather than as a sum (or some other function) of their constituent steps.

2.1.4 CATS

CATS, which employed much of the same team as had worked for WORM, but now working for another Dutch Ministry, that of Transport and Waterways, was the inheritor of the combined experience of the preceding projects. Now the different elements of the model had to be applied to aviation. This is a very different activity to those which had been the subjects of much of the earlier work, particularly the chemical process industry. In the latter many of the risk control measures or barriers (and almost all of the most crucial ones) are (largely) hardware barriers. In aviation, whilst there is a great deal of sophisticated hardware, the competence of the pilot, air crew and air traffic control is a key element in risk control. The formulation of all barriers as physical manifestations, as in WORM, is much less appropriate, and the formulation of behaviour as a vital and integral part of the barrier itself (in the detection and diagnosis of the hazard and action to control it), as envisaged in ARAMIS, is more appropriate.

In the CATS project, a model describing the effect of safety management on the risks of air traffic was formulated. The management model adopted the concept of *barrier life cycle, deliveries,* and *steps within each delivery (block diagrams)* from I-Risk, ARAMIS and WORM. Barriers were translated into devices for detection, diagnosis and action (DDA) against the developing hazard of going outside the safe envelope of flying. The barriers could have one or the two elements, human or technology or a combination of the two. The delivery systems were seen again as providing the essential resources and controls so that those barrier functions were carried out effectively. The formulation of the behavioural delivery systems defined in CATS was more or less the same as those from ARAMIS, including the combination of commitment with conflict resolution. The deliveries for hardware were modified from I-Risk by renaming “ergonomics” as “technology interface” and “equipment” as “technology function”, and combining the two phases of specification/ design/ installation/ adjustment and inspection/ maintenance into one life cycle covering all those phases. The life cycles of both technology and human behaviour in CATS were represented explicitly in the WORM terms, apart from the fact that “use” was defined as the actions of the barrier elements (see Figure 2-8).

This was the idea proposed in CATS, but some of the ideas were actually not implemented in the structure of the risk management process implemented in CATS (Figure 2-8). The model in Figure 2-8 reads from left to right in “sentences”, e.g. through the use of refresher training, the competence of people (4) to use the technology (3) to detect the approach (2) to the safe envelope boundary (1) in a certain scenario. But for instance, what was finally modelled in CATS did not explicitly incorporate the management cycle of risk assessment and learning in its final conceptual model; neither did CATS explicitly incorporate the block diagrams as the set of management actions in its final modelling.

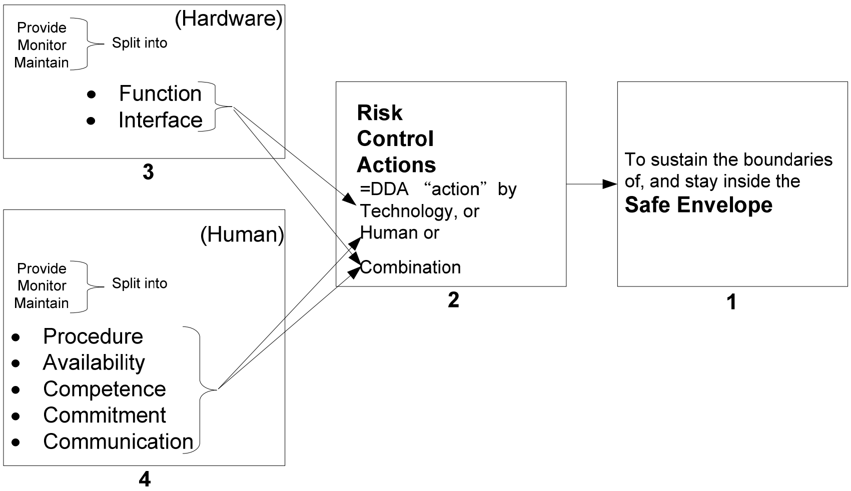


Figure 2-8 Model of CATS

2.1.5 Internal conclusions for this section

Tracing the development of the modelling through earlier projects has shown that the Dutch model has a long development history to arrive at Figure 2-8, in which it has been tested a number of times. Some parts of it have consistent feature that have stood the test of time, whilst the other parts are still criticized. There are a number of consistent features of the Dutch model:

- In the safety management systems in I-Risk, the model linked management to the occurrence and control of errors and failures. In the ARAMIS project and later, this link was explicitly formalised and barriers were designed consisting of the correct functioning of hardware and/or human behaviour to prevent the full range of accident scenarios for the activity. Barriers have to be put and kept in place for their whole life cycle. The company management exercises control over major hazards by allocating suitable *resources and controls* to ensure the continued correct functioning of that barriers. So, safety management was seen as "delivering" those systematic resources and controls to those barriers, in order to have the barriers put and kept in place for their whole life cycle.
- The supply of these resources and controls is achieved by secondary (management) processes. The *delivery systems* were detailed as workings out of the barrier life cycle both for hardware and behaviour. In different projects, this led to slightly different formulations of a limited number of delivery systems related to hardware and behaviour. Apart from these delivery systems, the safety management system also has to manage the processes of "*risk (scenario) identification, barrier selection and specification*", and the process of "*monitoring, feedback, learning and change management*". These two systems are as essential as the others, because they drive the whole safety management system in their own cycle. In the past, a lot of emphasis was put into the technical and human delivery systems and their links with the technical model, but the "risk assessment" and "feedback and learning" deliveries were the ones which got least attention in the Dutch models in those projects. This is probably because they are at another level of the system hierarchy that is not linked directly to barriers, but to the whole cycle of designing and selecting the barriers and improving them.
- Each delivery system consists of a number of *steps* to clearly show what actions should be taken by managers to deliver resources and controls from the delivery systems to the barriers. Whether the company takes systematic and effective actions for each of the steps determines the effectiveness of the barrier functions.

However, a few general modelling problems can be concluded as follows:

- Throughout the development of the Dutch model we still can put a question mark on where precisely these delivery systems should link to human and technical systems, if we are to take a closer look at the human factors or technical failures in the accident analysis. In the past, the management system was defined as to manage barriers' life cycles, which were conceptualized as being to put them in place, ensure their use by the operators, maintain or modify them to retain or restore their functioning and replace or improve them where necessary. This conceptualization worked fine for physical manifestations in I-Risk and WORM and served as a good audit tool in ARAMIS, which was the main objective for the management model in ARAMIS. But when the model had to be applied to CATS with the main goal of resolving human errors which are subject to a great deal of underlying human factors, the Dutch management model was still too conceptual, generic, and difficult to apply. This led to the conclusion that, to create effective management functions, individual process analysis must be performed to understand what “resources and controls” are needed and how they are provided.

- Second, during the implementation or quantification stages, neither I-Risk, ARAMIS, WORM or CATS had successfully implemented the theoretical management model to its full extent. For instance, the formulation of WORM (Figure 2-7) and CATS (Figure 2-8) did not explicitly incorporate the management process “steps” in the full formulation of the delivery systems in their final modelling. In I-Risk and ARAMIS, to establish a valid link between the quantitative technical model and various types of barriers and their management control from delivery systems had proven to be too complex and almost unmanageable. This lack of completion was partly due to the limited time scales available for the project, but also because the conceptual model was not easy to apply and therefore needed some simplification.

Hence, along with the findings from other validations (the rest of this chapter), in Chapter 7 we need to revise the model to resolve some of the confusions in the previous project and propose solutions for them, either in the short term or proposed for the long term.

2.2 Mapping the Dutch model with other aviation management models

In this section, we are looking for the mapping between the Dutch model and the other relevant models. This constructs the first validation of Figure 2-1. The relevant models start with the management parts of accident/incident reporting systems, risk models, and safety audit programmes. A number of different models developed for or applied to aviation were preliminary reviewed in terms of the Dutch model to indicate what management issues are encompassed within their models and how these map to the Dutch model. These were ADREP (ICAO Accident/Incident Data Reporting System, LOSA (Line Operations Safety Audit), HFACS (Human Factors Analysis and Classification System), Eurocontrol IRP (Integrated risk picture for Air traffic management in Europe), and SoTeRiA (Socio-technical risk analysis)

Of the modelling approaches, the ADREP taxonomy mainly emphasizes the operational side of the factors (level 1 and level 2 of Figure 1-9) and LOSA mainly focuses on the observation of the flight crew’s external error mode (level 1). As a result neither of these models have explicit management models or factors, so neither of these models are included for review here but will be dealt with in Chapter 3 where the human factors in accident/incident reporting systems are reviewed.

Therefore, the only comparisons which were considered useful and feasible were with HFACS, Eurocontrol IRP, and SoTeRiA. Although Eurocontrol’s IRP indicated that it considered ATC safety management in its report, it did not specify this in further detail in the contribution. It only identified the elements of safety management as “policy”, “planning”, “achievement”, “assurance”, and “promotion”. This is the Deming circle again and hence the Dutch model at a generic level maps onto that. We will therefore not consider it further in the management review. In the following sections (2.2.1. and 2.2.2) HFACS and SoTeRiA are mapped onto the Dutch management model to see if it could accommodate all of the insights that they contained.

2.2.1 Human Factors Analysis and Classification System (HFACS)

HFACS (Wiegmann & Shallell, 2003) was specifically developed to define the latent and active failures implicated in Reason's Swiss cheese model (Reason, 1990) so that it could be used as an accident investigation and analysis tool. The framework of HFACS is described diagrammatically in Figure 2-9. Level 1 shows the unsafe acts of operators classified into errors and violations. Level 2 depicts the preconditions for unsafe acts, addressing the latent failures within the causal sequence of events as well as the more obvious active failures, i.e. environmental factors, personnel factors, and condition of operators. Reason's Swiss cheese model includes supervisors who influence the condition of pilots and the type of environment they operate in. Therefore HFACS also identifies a third level of unsafe front-line supervision. Level 4 describes the contributions of fallible decisions in upper levels of management that directly affect supervisory practices. Selected examples of level 1 and level 2 of HFACS can be found in Appendix B. For the management model we are concerned with levels 3 and 4 so their selected example will be demonstrated in the following of this section.

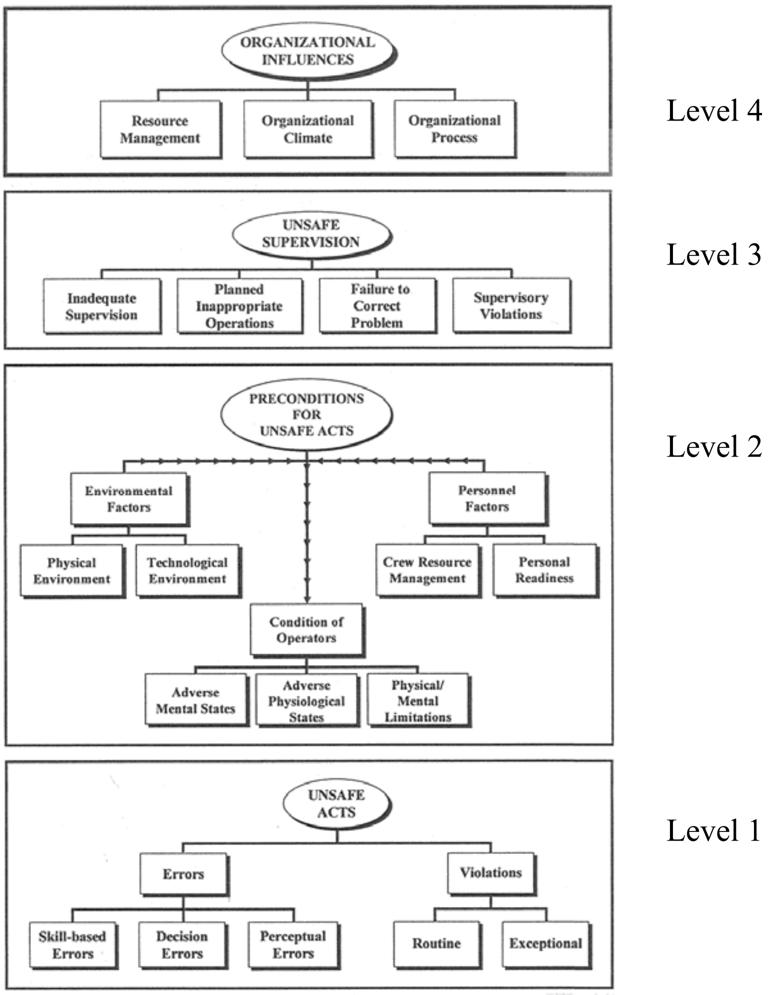


Figure 2-9 Human Factor Analysis and Classification System (HFACS)

Level 3 Unsafe Supervision

Unsafe supervision includes inadequate supervision, planned inappropriate operations, failure to correct a known problem, and supervisory violations. In the Dutch model in its various development stages, we do not have "supervision" as a separate delivery system. It was considered that all of the actions of supervisors could be incorporated under either "competence" (providing back-up knowledge when consulted), "commitment" (making sure rules are followed and the operators are motivated) "communication" (coordination of the work of different operators) or monitoring (checking and correcting or initiating learning). Comparing the factors in the first two categories of supervision in HFACS, namely the "inadequate supervision" and "planned inappropriate operations" in Table 2-4 with the Dutch models, their factors appear to be covered to a great extent by the delivery systems for human behaviour indicated in the Dutch model. On the basis of the mapping of all the factors in Table 2-4 into the Dutch safety management model, we acknowledge that the supervision aspects referred to in these two categories is reflected in the "provide" and "monitor (track)" steps of Deming cycle through our delivery systems. For instance, "fail to provide proper training" and "fail to provide professional guidance" are related to provide the designed measures to the people in the system to be trained and informed about procedures. "Fail to track qualification" and "fail to track performance" are monitoring steps to detect (potential) deviations from specified functioning. In other words, the factors in these two categories can be seen as a representation of how well the company takes these steps in its own management process. This means that these management factors are largely dealt with already, at least in principle, in our safety management model.

Table 2-4 Factors of unsafe supervision (I)

HFACS unsafe supervision categories	Delivery system
Inadequate supervision	
Failed to provide proper training	Competence
Failed to provide professional guidance/oversight	Procedure, competence
Failed to provide current publications/adequate technical data and/or procedures	Procedure
Failed to provide adequate rest period	Availability of manpower
Lack of accountability	Commitment, monitoring
Perceived lack of authority	Commitment
Failed to track qualifications	Competence
Failed to track performance	Generic "monitoring" steps in the delivery systems- e.g. step 8 in Figure 2.6 and its fellows in other systems
Failed to provide operational doctrine	Procedure
Over-task/untrained supervisor	Availability of manpower/ Competence
Planned inappropriate operations	
Poor crew pairing	Availability of manpower
Failed to provide adequate brief time/supervision	Communication & Coordination
Risk outweighs benefit	Commitment and conflict resolution
Failed to provide adequate opportunity for crew rest	Availability of manpower
Excessive tasking/workload	Availability of manpower

In Table 2-5, the third category of "failure to correct a known problem" in HFACS's supervision refers to those instances when deficiencies among individuals, equipment, training or other related safety areas are known to the supervisor, yet are allowed to continue. The fourth category of "supervisory violations" refers to those instances when existing rules and regulations are wilfully disregarded by supervisors. In the Dutch model, commitment is about the decision to choose one from several possible courses of action, all of which are within the competence of the pilot or other operator concerned. But some of the conflicting pressures of pilots are created in the management hierarchy, which places incompatible demands on front-line personnel in their roles as risk control measures. Hence, conflict resolution at a management level can be seen as a management product that influences the functioning of the commitment of the workforce. Therefore, the last two categories identified in HFACS's supervision can be accommodated within commitment and conflict resolution at management level in that delivery system.

Table 2-5 Factors of unsafe supervision (II)

HFACS unsafe supervision categories	Delivery system
Failed to correct a known problem	
Failed to correct inappropriate behaviour/identify risky behaviour	Commitment and conflict resolution
Failed to correct a safety hazard	
Failed to initiate corrective action	
Failed to report unsafe tendencies	
Supervisory violation	
Authorized unqualified crew for flight	Commitment and conflict resolution
Failed to enforce rules and regulations	
Violated procedures	
Authorized unnecessary hazard	
Wilful disregard for authority by supervisors	
Inadequate documentation	
Fraudulent documentation	

To conclude, although the Dutch model seems not to have supervision as a separate delivery system, if we zoom in much more on the factor content and map with the different delivery systems, "supervision" appears to be covered quite well under the various steps in the block diagrams of the five behavioural delivery systems. For this reason, we do not need to add a new delivery system for it.

Level 4 Organizational influences

HFACS's level 4 describes the contributions of fallible decisions in upper levels of management that directly affect supervisory practices, and through them (or directly) the conditions and actions of front-line operators. This level includes three main categories: resources management, organizational process, and organizational culture.

"Resource management" encompasses the realm of corporate-level decision making regarding the allocation of organizational assets, which includes three sub-categories of "human resources", "equipment and facilities", and "monetary/budget resource" (see Table 2-6). In the Dutch model, the delivery systems for "availability of manpower" and "competence" are concerned with allocating the necessary time (or numbers) of competent people to do the safety-critical primary business tasks which have to be carried out. They also provide competence through selection, training and experience management to personnel whose behaviour, alone or using equipment, constitutes the risk control measure. The delivery systems for "technology-function" and "technology-man-machine interface" cover the desired functioning of the technology coupled with the question whether it can be operated easily and correctly when and where needed to carry out the activities. These deliveries fit the first two perspectives of “resources management” in HFACS. The last sub-category in Table 2-6, namely “monetary/budget resource”, links with the higher level of commitment to safety, as corporate decisions about how such resources should be managed are based upon conflicts between the goal of safety and the goal of performance (e.g. punctuality) and cost effective operations.

Table 2-6 Factors of organizational influences (Resource management)

HFACS organizational influences--Resource management	Delivery system
Human resources	
Selection	Competence
Staffing/manning	Availability of manpower
Training	Competence
Background checks	Competence, monitoring
Equipment/Facility resources	
Poor aircraft/aircraft cockpit design	Tech-Man-machine interface
Purchasing of unsuitable equipment	Tech-Function
Failure to correct known design flaws	Tech-Function
Monetary/budget resources	
Excessive cost cutting	Commitment and conflict resolution
Lack of funding	Commitment and conflict resolution

However, the factors in the second and third categories at the organizational level of HFACS, namely “organizational process” and “organizational climate”, are currently not well coded in our delivery systems (see Table 2-7 & Table 2-8). They deal with aspects at a higher management level, which is not well modelled in the Dutch model, although there are some links to delivery systems and the steps of designing the SMS and monitoring and improving it. “Procedure” and “oversight” in Table 2-7 are mapped well with the procedure delivery system and the high level design and learning system loops, but “operations” maps relatively less well and in ill-defined ways compared to the others. It is not clear exactly what HFACS means with these categories, especially how they articulate with the lower levels. We have put on the delivery systems to these factors as best as we can, but with question marks to indicate it is not entirely clear what it is covered. Dissatisfaction with these higher levels of HFACS has also been expressed in another study of management influences using it (Hale et al., 2010). The lack of specificity makes it difficult to connect with the Dutch model.

The issue of safety culture and safety climate is considered important because many users and people working in aviation believe that this construct influences personal safety attitudes to a significant degree and can predict unsafe behaviour and accidents. In HFACS, the author includes this under “organizational climate” (see Table 2-8).

Table 2-7 Factors of organizational influences (Organizational process)

HFACS organizational influences--Organizational process	Delivery system
Operations	
Operational tempo	? Workload pressure
Incentives	Commitment
Quotas	? Commitment
Schedules	? Availability
Procedures	
Performance standards	Procedure
Clearly defined objectives	Procedure
Procedures/instructions about procedures	Procedure
Oversight	
Established safety programs/risk management programs	Overall SMS design
Management's monitoring and checking of resources, climate, and processes to ensure a safe work environment	High level monitoring & learning

Table 2-8 Factors of organizational influences (Organizational climate)

HFACS organizational influences--Organizational climate	Delivery system
Structure	
Chain-of-command	Distribution of roles in fig 2.5
Communication	Communication at management level
Accessibility/visibility of supervisor	Monitoring
Delegation of authority	Competence, commitment
Formal accountability for actions	Monitoring, commitment
Policies	
Promotion	Competence, commitment
Hiring, firing, retention	Competence
Drugs and alcohol	Suitability
Accident investigation	Monitoring, feedback and learning
Culture	
Norms and rules	Procedure
Organizational customs	Procedure, commitment
Values, beliefs, attitudes	Commitment

But the concept of safety culture still remains ill-defined in HFACS, though we can see some mappings to procedure and commitment delivery. This is a problem which is indeed general in the literature (as will be discussed in more detail in Section 7.2.2.8). Whether we need a separate delivery system for “culture” will be discussed in that section. The other aspects of “organizational climate” map to some extent again to the higher level monitoring and learning processes and to the competence and commitment delivery systems.

In general we have to conclude that the Dutch model has more difficulty coping with the level 4 influences of HFACS, but that model is also somewhat vague in its definitions of those higher levels. Since it would only be possible to model at that management level if the links to and through the lower levels are agreed upon and well modelled, we are inclined to put off the resolution of the vagueness at level 4 until such time as we have made progress with levels 1 to 3. We return to this in Chapter 7.

2.2.2 SoTeRiA

In SoTeRiA (Mohaghegh, 2007; Mohaghegh & Mosleh, 2009), the author includes a sector called "organizational structure & practices" which resembles the safety management system in the Dutch model. This safety management system has been developed for application in the aviation industry and applied to aircraft maintenance. In SoTeRiA, organizational safety activities include all organizational practices that support the resources, procedures, and human actions in the "unit process model" (e.g. maintenance unit, operation units) that includes the direct activities that affect safety critical performances. The direct activities are decomposed to their direct resources, procedures, and the involved individuals' performances in the unit process model. All organizational practices that influence the resource, individuals, and procedures in the unit process models are defined as organizational safety practices. SoTeRiA classifies them into four groups including (1) human-related activities, (2) procedure-related activities, (3) resources-related activities, and (4) common activities. All the first three of activities are supported by the fourth one (common practices) (see Figure 2-10).

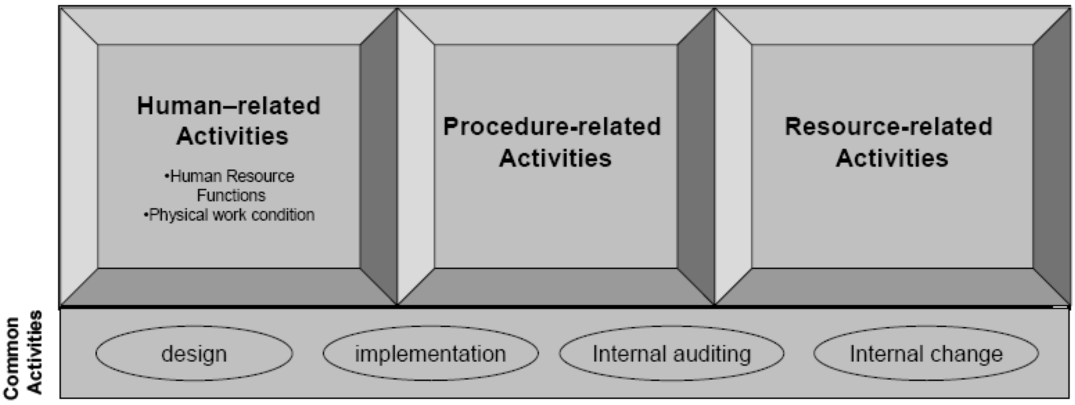


Figure 2-10 Organizational safety practices (maintenance-specific)

"Resource-related" and "procedure-related" activities are the practices that support the resources and procedures of the unit process model. According to Mohaghegh (2007), they are more specific than human–related activities. For example, in the case of the maintenance unit process, one of the resources-related activities is in-house calibration and testing that supports locally produced tools and equipment, and one of the procedure-related activities is the process of "alteration" that supports records and reporting. Since the former aspect of SoTeRiA mainly focuses on supporting technical resource (tools/equipments), it fits into the "technology-function" delivery system of the Dutch model, whilst the procedure-related activities match with the Dutch "procedure" delivery system. However, Mohaghegh then went much deeper into the maintenance unit process and its related organizational safety practices in her case study. She built another layer to detail the causal paths for these two example activities, tracing their resources, procedures and actions. But from there, she identified another layer of resources, procedures, and human actions that are needed to provide those resources, procedures and actions, which brings the modelling into a complex regression (what procedure defines the competence of those who write the procedures to select the people who will write the operational procedures, etc.). We recognize this as a process of repetition or iteration, which was labelled the "Russian doll problem" in the series of Dutch projects (I-Risk, ARAMIS, WORM and CATS) (the iteration of behavioural and organisational influences, which influence other organisational influences). An arbitrary decision has to be made about the point to cease the detailed modelling.

The "human-related" activities in SoTeRiA are those that support individual performance in the unit process model. The factors in human-related activities in SoTeRiA's were extracted from a study by Ostroff (1995). The first column in Table 2-9 shows all the human-related activities in SoTeRiA adapted from Ostroff (1995).

Table 2-9 Human-related activities from SoTeRiA

SoTeRiA	Delivery system
Selectivity in recruiting/hiring	Competence
Internal staffing	Availability of manpower
Contingent workforce (e.g. contractors)	Availability of manpower
Training and employee development	Competence
Appraisal	Monitoring of competence, Commitment and conflict resolution
Compensation and reward system	Commitment and conflict resolution
Job analysis	Risk assessment and risk control measure choice
Job enrichment	Commitment/motivation, Competence
Team system	Commitment/motivation, Competence, Learning
Employee assistance	Suitability, Availability, Commitment
Due process	Commitment
Employee voice/empowerment	Commitment, feedback& learning
Diversity	Commitment, competence, human resource management
Legal compliance	Monitoring, feedback and learning
Safety	Competence, Monitoring and learning
Union relation	Not link very much to safety, but to human resource management

It was hard to link the topics in the first column of Table 2-9 literally to delivery systems because they are quite general and SoTeRiA did not take these aspects and model them in any detail. Therefore, we have to refer to Ostroff's more detailed explanations shown in Appendix C and map these onto the Dutch model. The mapping is shown in the second column. Overall, with the definitions provided by Ostroff some of the human related activities can be mapped into the delivery systems. But there are quite a number of influences (e.g. increase minority representation and diversity in the company, discrimination against members of protected classes, etc.) which are more related to human resource management than to safety management. This list is therefore a rather strange mixture of influences and these remote influences are not relevant to map to our model.

The fourth activity, named common activities, consists of "design", "implementation", "internal auditing", and "internal change system". All bottom layer "procedures" and "resources" in the framework are affected by these design, implementation, internal auditing, and internal change factors. This is identical to the PDCA cycle (plan, do, check, and adjust) rooted in the quality control of the Deming cycle. The Dutch model does not have this separated out as antecedents of management activities, but rather integrates them into each of the delivery systems and into the review process of the safety management system.

In addition to these organizational practices, SoTeRiA also included psychological terms in its scheme (Figure 2-11) such as group safety climate, organizational safety culture, and emergent processes. However, even from a thorough study of the report it was not clear what elements the "organizational structure" and "organizational climate" contain, and why the relationship between "organizational structure" and "organizational climate" is presented as such in the research. Because these points are not clear at the moment, we cannot cross reference this aspect.

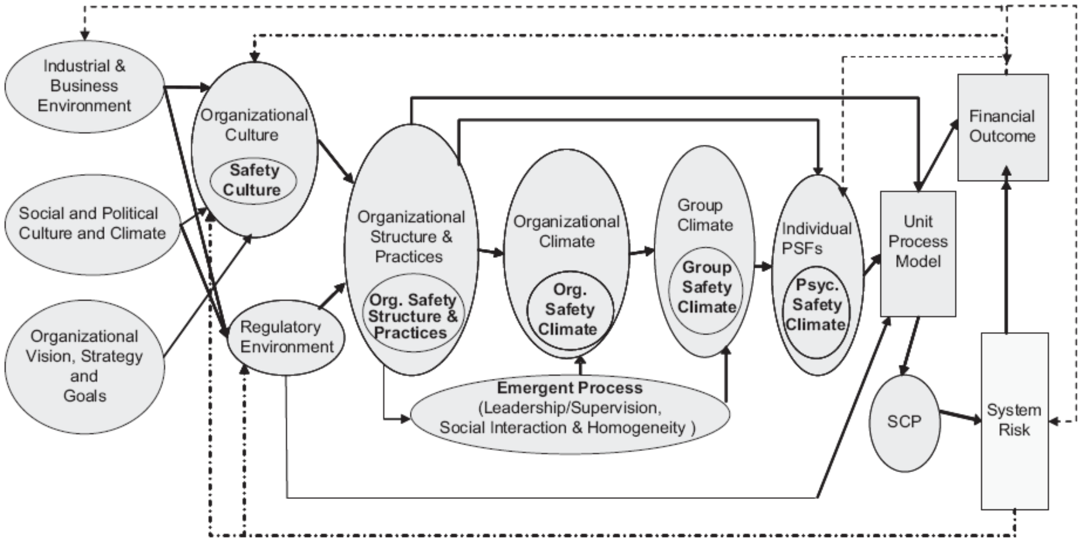

Industrial & Business Environment
Social and Political Culture and Climate
Organizational Vision, Strategy and Goals
Organizational Culture
Safety Culture
Regulatory Environment
Organizational Structure & Practices
Org. Safety Structure & Practices
Organizational Climate
Org. Safety Climate
Group Climate
Group Safety Climate
Individual PSFs
Psyc. Safety Climate
Emergent Process (Leadership/Supervision, Social Interaction & Homogeneity)
Unit Process Model
Financial Outcome
SCP
System Risk

Figure 2-11 Schematic representation of SoTeRiA

As with HFACS, SoTeRiA also presents some problems in understanding the exact meaning of a number of the terms which are used in the model and so of mapping them to the Dutch model. Again this is particularly true for safety climate and culture.

2.3 Safety management system in practice

Section 2.1 has shown the theory of the early Dutch models. To find out how safety management works in practice in aviation, Section 2.3.1 reviews the international SMS standards and requirements in aviation. Section 2.3.2 summaries the interview results with two airlines about their SMSs.

2.3.1 International SMS standards

The International Civil Aviation Organization (ICAO), a specialized agency of the United Nations, has mandated its member States to develop and implement SMS programs to achieve an acceptable level of safety in aviation operations (ICAO Annex 6, 2006). ICAO gives a definition for "safety management system" as *"an organized approach to managing safety,* *including the necessary organizational structure, accountabilities, policies and procedures" (ICAO,2006, 1-2).*

In this regard the European Aviation Safety Agency (EASA) stated its intention to translate the SMS related provisions in ICAO Annex 6 and JAR-OPS 1.037 into their upcoming rulemaking proposals, so that they will be similar for aircraft operators, maintenance organizations, air navigation services providers and aerodrome operators. (EASA TOR No: OPS.001 Subgroup Authority requirements and SMS). EASA has published several best practice materials on safety management systems to help stakeholders comply with ICAO requirements, but up to the time of writing of this thesis they do not yet have any legal status. A proposal for future EASA rules for organizational requirements is currently the subject of public consultation as part of the EASA Notice of Proposed Amendment NPA 2008-22. Draft requirements can be found on the EASA website, published on 31 October 2008. But the final content of legal requirements regarding an SMS is still to be determined. However, all air operator certificate holders in Europe are already being encouraged to have an SMS implementation plan that will provide for a fully functional SMS in two to three years.

In the U.S. the Federal Aviation Administration (FAA) also supports the harmonized implementation of international standards, and is currently working to make U.S. aviation safety regulations consistent with ICAO standards and recommended practices. Similar to Europe, an SMS is not currently required for U.S. certificate holders, but the FAA Advisory Circular No. 120-92 (FAA, 2006) introduces the concept of SMS to aviation service providers.

Key Generic Features of the ICAO SMS

There is no definitive definition attached to the term "SMS". Every organization (e.g. Europe and U.S.) and industry, for that matter, has its own interpretation of what it is. In general the SMS components from the civil aviation perspective may contain broad features assigned to the SMS. According to ICAO SMS manual (ICAO, 2009), the following are the four components which constitute the basic building blocks of an SMS. Each component is subdivided into elements, which encompass the specific sub-processes or tasks to conduct the management of safety. They are

- Safety risk management
  a) hazard identification
  b) risk assessment and mitigation
- Safety assurance
  a) safety performance monitoring and measurement
  b) the management of change
  c) continuous improvement of the SMS
- Safety policy and objectives
  a) management commitment and responsibility
  b) safety accountabilities
  c) appointment of key safety personnel
  d) coordination of emergency response planning
  e) SMS documentation
- Safety promotion
  a) training and education
  b) safety communication.

The two core operational activities of ICAO SMS are safety risk management and safety assurance (ICAO, 2009). Safety risk management must be considered as an early system design activity, aimed at initial identification of hazards in the context in which operations related to the delivery of services will take place. Safety assurance is considered as a continuous, ongoing activity to ensure that the operations that support the delivery of service are properly protected against hazards and that learning takes place and change is responded to. These two activities are implemented through safety policy and objectives and are supported by safety promotion. Without these two components, hazard identification and safety risk management would be impossible or seriously flawed. Therefore, ICAO considers that safety risk management and safety assurance are the actual "doing" of the SMS; they are the operational activities underlying a performing SMS. On the other hand, safety policies and objectives and safety promotion, provide the frame of reference and support that allow the operational activities underlying safety risk management and safety assurance to be effectively conducted.

Mapping what ICAO identifies on to the Dutch model (Table 2-9), the "safety risk management" heading corresponds well to the risk assessment and choice of prevention/risk control measures, whilst the "safety assurance" heading covers the monitoring, and feedback, learning and improvement loop in our model. The "safety policy and objectives" heading encompasses many different aspects of the organisation's safety management system. "Management commitment and responsibility" are equivalent to our commitment and conflict resolution delivery system. "Safety accountabilities" equate to the allocation of tasks overall in the Dutch model and represent actors for functions and not management functions themselves. "Appointment of key safety personnel" relates to the competence delivery system. The "coordination of emergency response planning" covers all of the delivery systems applied to that emergency phase, with particular emphasis on the procedures. "SMS documentation" covers both our "procedures, rules and goals" delivery system for task performance and for the reporting and learning system, as well as overall documentation of the complete SMS manual. "Training and education" underlying "safety promotion" matches well with our competence delivery system that ensures personnel are well trained and competent to perform their safety management duties. The major focus of "safety communication" in the ICAO framework is the appropriate communication of SMS objectives and procedures to all operational personnel to achieve their understanding and commitment. This therefore contains elements of (management) commitment as well as of procedures and some aspects of coordination and communication. Such organizational communication to promote safety can be performed in different forms, e.g. via safety management manual, safety procedures, safety newsletter, bulletins or via website or email, which should also ensure communication flow freely between the safety manager and operational personnel throughout the organization. It even contains some elements of the learning loop, since it proposes that the safety manager should ensure that lessons learned from investigations, both internally and from other organizations, are distributed widely via these channels.

Table 2-10 ICAO safety management system

ICAO SMS		Dutch model
Safety risk management	a) hazard identification b) risk assessment and mitigation	Risk identification, barrier selection and specification
Safety assurance	a) safety performance monitoring and measurement b) the management of change c) continuous improvement of the SMS	Monitoring, feedback, learning and change management
Safety policy and objectives	a) management commitment and responsibility	Commitment and motivation
	b) safety accountabilities	Distribution of tasks
	c) appointment of key safety personnel	Competence & distribution of tasks
	d) coordination of emergency response planning	All delivery systems for this activity
	e) SMS documentation	Procedure
Safety promotion	a) training and education	Competence
	b) safety communication	Commitment and motivation + others

Table 2-10 summarised the mapping, mostly accommodated under the delivery systems of "commitment and motivation", "procedure", "competence", and the two higher level management functions of risk assessment and learning. It should be noted that the international SMS standards and requirements mentioned above are created in a way to emphasize more on "what to do" rather than "how to do it", as is the Dutch model. The reason behind this is to create standards which are set in a way that suits a wide variety of types and sizes of organizations. These standards are designed to allow the companies to integrate the safety management system into their individual operational models. However, based on the mapping with the Dutch model, although we can accommodate all of the ICAO requirements in the Dutch model, there are elements of the Dutch model which have very little if any place in the ICAO standard. For example we do not find all operations related to the delivery of services (e.g. on-line communication between the crew and with the ATC) explicitly defined as part of ICAO SMS; "availability" of key personnel for safety related tasks (rostering, etc.) is also not explicit and the issue of commitment of the operational level (pilots, ATC, maintenance, etc.) seems at most implicit.

2.3.2 Interviews with airlines

Two interviews were conducted to find out how safety management systems work in practice. The objective of the interviews was to see if the Dutch model covers all the things that safety management systems are doing in the airlines. In order to verify this point, we interviewed two senior flight safety managers responsible for flight safety from two airlines, a world leading airline (A) and a local low fare airline (B[8]). In order to map their system onto ours, we

[8] Airline A is a worldwide airline company which transported more than 74 million passengers and more than 600,000 tons of cargo in 2008. Airline B is a low fare airline company which served over 46 million passengers in 2009.

first asked the interviewees to give a full description of their safety management system without showing them our model, so as to minimize the effect on the mindset of the respondent. In the second part of the interview the interviewees were asked to give comments on an early version of Dutch model (i.e. the safety management model in CATS). After the interview, the experts also provided their safety management manual as supplementary documents. The interview report was sent to the experts for review, so that aspects that they might have thought of after the interviews could be added to the report.

Airlines safety management systems

In the first part of the interview, we asked the interviewees to talk about their safety management systems. They generally described two systems dedicated to flight safety to ensure it is in compliance with the regulations. The first system was the “Accident prevention and flight safety program” which contains occurrence reporting systems, data gathering and analysis tools, and a company wide safety database to identify adverse trends or address deficiencies in the interests of flight safety. When an event occurs, the accident investigation team will carry out an investigation in order to determine the causes of an incident/accident. The contents and the recommendations are done during the final stages of the investigation report. Using the IATA risk matrix, every event in the safety report will be assessed as low, medium, high, or substantial risk, based on a combination of the likelihood of the event and the severity of damage. Every event is recorded in the database with a risk level. Small risks, e.g. birds hit, are put into the database with no immediate safety action required, except to keep tracking the trend and check that the trend does not go in the wrong direction. For medium and high risk, the occurrence report will be discussed with line management by the safety manager to define the corrective actions. In the monthly safety meetings, each division will bring their report and discuss safety concerns and set out corrective action both within and across the borders of divisions. During the meeting, they also discuss organizational factors, for instance problems in training, procedures, and in the organization. If the issue discussed has an impact on another division, they will prepare advice to the quarterly meeting, which is a level higher, including director and post-holders (person responsible for management and supervision from each division) of the company. In this meeting, they only look at a few high risk aspects. During the quarterly meeting, corrective or preventive actions will be discussed and endorsed by the director and post-holders. The corrective action will then be monitored by quality monitoring processeses in the second system.

The second system was the “quality assurance programme” which contains the “monitoring and feedback system” to ensure corrective actions are carried out properly, and the “regulatory compliance process” to ensure all operations are being conducted in accordance with all requirements, standards and procedures.

Comparing these two systems with the Dutch model, their SMSs were equated to our “risk assessment” and “feedback and learning” delivery systems which drive the whole SMS, but hardly mentioned the operational delivery systems for human and hardware, except in so far as they are manifest in the requirements, standards and procedures checked in the compliance process. We were surprised by the fact that what they called SMSs was, in our eyes, a very partial view ignoring the operational processes – even the rather limited set in the ICAO requirements and seeing the SMS in a very formal way (compliance with complex regulatory requirements by prescribing measures to prevent recurrence). They thought that their description and the safety manual almost covered the whole safety management system. We could conclude that the SMSs in practice concentrated almost exclusively on the first two headings of the ICAO requirements, and did not even have as much coverage of the operational aspects as the other two headings in ICAO.

In the second part of the interview, we specifically asked the interviewees to talk about what and how they deliver resources and controls specified in the Dutch model to the online people. In both cases, the respondents had great difficulty in thinking in conceptual terms about safety management systems or seeing things in term of our management model. We had to indicate more specifically what each of the delivery systems contains. They then pointed out that it is the *line operations* who are actually carrying out all of the functions in the delivery systems integrated in their normal activities. One of the interviewees gave an excellent definition of what he thought line and safety department should do.

> *"Flight safety department is the section within the organization that sees things happening and analyzes things. But we can't make safety progress ourselves. We are the mirror with thorough investigations, but the line managers have to come up with the correct actions to justify the recommendations, tradeoff between safety recommendations and other issues (such as money), and make decisions to make flight safer."*

However, the safety managers did not define all of the activities where safety is influenced in the line operations as part of their safety management systems. As a safety management group or safety service they do not see it as their role to be responsible for making flying safer; it is the line managers who are (because they have the budget and the people to improve safety). The only bit clearly defined as being part of the SMS was the "feedback and learning" and "risk assessment", whilst they saw the rest of the delivery systems as being so integrated in what people in the line is doing that they were unable to be separated out in the way that we do for the Dutch model.

Whilst we applaud the degree of integration of safety in line operations, it was surprising that the informants did not emphasize more the actual content of what the line operations do to achieve safety. It had been expected that the flight safety managers would have a clear overview of how everything fits together (how operations do it and how operations may fail so that the safety department can fulfill their monitoring and learning role). But they did not succeed in articulating that despite their roles as monitor of the whole proactive and reactive system which should mean they have a clear overview of this. Hence, in a sense the mapping (for human and hardware delivery systems) did not really work. On the other hand, neither of the two interviews identified aspects of what was being done for safety management in practice that did not fit the Dutch model.

Another attempt at mapping:

After the interviews it became apparent that the whole range of human and hardware delivery systems had been left out because they lie in the line. To get into that aspect and validate this part of the system, it would have been preferably to use observation research as the most unobtrusive way of observing the phenomena in relation to our system. But, it would have been a huge tasks for me to observe, since safety business processes are so embedded in everything that is being done in the organization. Due to the time constraints, we made a further attempt through interviewing a safety professional from the flight safety department in Airline A, who already has some academic and research training in SMS. This interviewee has also been a captain or first officer for almost 25 years. As well as a flight instructor, from 1998 he was also senior-type-rating examiner which includes pilot and instructor training and examination. In 2005, the interviewee started as flight safety investigator in the flight safety department of Airline A.

For each human delivery system we asked the interviewee to freely talk about his company's management process in terms of those headings of "procedure", "competence", "availability", "commitment", and "communication". The interviewee was asked, for instance, to talk about the process of delivering the aircraft operating procedures. Without showing the interviewee our framework, I mapped what the interviewee said about their processes onto our "barrier life cycle management", i.e. the block diagram developed in ARAMIS project (see Figure 2-6 as an example and the others in Appendix A). Next, we presented the interviewee with the block diagrams and asked the interviewee what he thought of it and if he could indicate steps (blocks) or flows (between blocks) that had not been identified in our model, but were present in their systems.

- The interviewee indicated that the model which they have in mind is usually less explicit in the safety context than the Dutch model, which is deliberately designed to be explicitly checked, assessed, and gaps found. He considered that it was fair enough that managers or people flying usually do not spend much time to think and talk about the theoretical concepts for their practical work and probably do not see a value in making it more explicit.
- Mapping what he told on to the Dutch model, it turned out that our model accommodates their systems well. "Commitment & motivation" was the subject which matches least well, but this was probably because their "commitment & motivation" activities are the most inexplicit subject of all in their system. So, there is potential for company A to develop this aspect more explicitly. However, this discrepancy is not a criticism of the Dutch model, which is relatively more comprehensive; it is more a criticism of the SMS of airline A.
- From this second mapping attempt, I also concluded that the "promulgate and train" step is an important one for all of the risk control delivery systems involving man. After the contributions of commitment, communication, procedure, availability, knowledge and skill have been designed as risk control measures, management needs to communicate with the workforce about what these measures are and train them to be able to perform the designed actions. The ICAO SMS also identifies it as an important element to support its core operational activities (see "training and education" under "safety promotion" in Section 2.3.1); indeed "promulgation and training" are the means to deliver resources from management to the online workforce. This management task, which is an explicit step in the "procedure" and "knowledge & skill" delivery systems, is not so explicitly modelled in the others in the Dutch model (see Appendix A). Therefore, in Chapter 7, we will make this factor more explicit in the new formulation.

It is important to note that the safety business processes in the airlines are so integrated that they are not explicit in the organization; therefore to ask the interviewee to freely talk about the full extent of his company's safety management without giving any direction proved impossible. The only way we were able to get the interviewee to describe them was to present to him with the headings of the delivery systems and then ask him to think how it was managed in his company. However, what we really wanted was going the other way round by demonstrating that our model accommodates everything from their system. In actual fact, the interviewees did not come up with any category which did not fit our model. But, this was not a very robust test because they were starting from our model to look at theirs. However, as far as it goes, this mapping of two actual SMSs onto the Dutch model does not reveal any important gaps in it.

2.4 Findings and suggestions

Tracing the development of the modelling through earlier projects has shown that the Dutch model has a long development history. Some parts of it are consistent features that have stood the test of time. However, behind this relative consistency in the formulation of the model, the review of the development of the Dutch model in Section 2.1 showed two critical problems that need to be resolved in this research.

- First, none of the previous projects had been finished and quantified in such a way that the original objectives of management modelling were realized, fully taken into account the block diagrams for their influences. This was partly due to the limited time scales available for the projects, but also because the concept model was not easy to apply in the past and therefore needs some simplification.
- Second, the Dutch model is still too conceptual and generic in respect to resolving (preventing and coping with) human and technical errors. Classifying errors in a meaningful way is therefore essential to record such data in a way amenable to the detection of trends in incident occurrence, or in identifying different ways in which the system could fail. Thus, individual process analysis must be performed to devise more specifically the resources and controls needed. Put simply, error analysis is an essential component of safety management. This part of the analysis will be the focus of the next two chapters.

Next, the comparisons with HFACS and SoTeRiA showed that the well-defined elements forming part of those models can essentially be satisfactorily mapped to Dutch delivery systems. There are some ill-defined elements, particularly safety culture and safety climate, but also relating to higher level corporate aspects, of which their definition are still so vague that they cannot be easily compared, or they relate to aspects of management and organisation at a higher system level than the Dutch model currently attempts to deal with. To accommodate the first aspects in our model, the definitions of safety culture/climate and whether we need to devise a specific delivery system for it will be discussed in Chapter 7. Dealing with any other more generic and higher level concepts is left to future work beyond this thesis.

Finally, the intention of the two interviews with airlines was to map their safety management system specifically on to the Dutch model. Initially, what we were expecting to find was a fully development and documented SMS, which we could take and look at it in detail and see whether everything we saw there fitted comfortably into our system. However, we did not find that. Instead, what we found was a very partial view. The main lesson is that they do not identify a whole section of the Dutch model covering all proactive efforts to plan and control hazards as being part of their safety management system. Their SMS label covers only the risk assessment, feedback, and learning loop as the two core safety management activities that ICAO recommends. This does not mean that airlines are not doing those tasks, only that they see them as so much a part of operations that they are not identified as safety management.

Currently, basic elements of safety management system may be in place in the airlines, but they are not explicitly and systematically identified as such in the operations and technical systems. What is needed is to move the SMS a step forward to make explicit the safety concerns and actions in each of the processes of the business, via the application of management controls to all aspects of the business processes critical to safety.

Although the interviews ended up with not being the test they had been exactly planned to be, in the first interview and the mapping with ICAO SMS we did verify that the “risk assessment” and “feedback and learning” elements of the Dutch model do fit. For the rest of the Dutch model, the elements were only partially identified as part of the ICAO standard. Only during the interview, were those elements indicated as being carried out in the line management, monitored by the “quality assurance system” with only the reactive monitoring being seen by “accident prevention and flight safety program” as theirs. In the second interview, due to the way it was finally set up, the remaining part of the Dutch model was partially validated based on what the interviewee told us about the management process in his airline, in the sense that no important operational and management process elements were discovered, with the possible exception of safety culture/climate, which did not fit somewhere in the Dutch model.

The unexpected lesson that we learned was that what are currently identified by airlines as their SMSs are predominately reactive in nature. What is needed is to move the traditional perspective toward a proactive view and to link processes in the line (operations and maintenance) explicitly with the control of barriers, so that safety can be explicitly traced through the organization. However, there is still a big difficulty in getting people to use an explicit management model in aviation. It meant that the process of validation of our model could not be easily done. However, we end up with an interesting conclusion in its own right, as well as one for future study.

3 Human performance

It is now acknowledged widely that during operations, human performance factors have a dominant influence on the safety of aviation operations. Estimates in the literature indicate that somewhere between 70 and 80 percent of all aviation accident can be attributed, at least in part, to human error (Shappell and Wiegmann, 1996). However human error is only the symptom of deficiencies in the architecture of the system and is often a result of a chain of events such as is described in Reason's Swiss Cheese model (Reason, 1990). Due to this, human error analysis and risk control measures modifying human behaviour are an essential component, or rather target of safety management.

In Chapter 2, we have formulated the management influences in terms of a safety management model. In risk modelling, the link between the management model and the accident scenarios/event tree modelling normally runs through the underlying models of human and technical failure. Therefore human models need to be able to interface not only with the events, but also with the management model. Moreover, individual process analysis needs to be performed to understand what resources and controls of human behaviour and performance are needed to create effective management functions.

Research concerning the human models in aviation will be reviewed in the first section of the chapter (Section 3.1). The essential human factors which the management have influence and control are the subject of the next section of the chapter (Section 3.2), identified from the existing accident databases and reporting systems. Those factors will be mapped on to the existing Dutch management model in order to see if the Dutch model can support the control functions to the human factors identified in them. Where they do not match, additional functions will be added to the Dutch model in Chapter 7. Afterwards the factors formulated in the current (probabilistic) quantification models are discussed in Section 3.3. This describes the current modelling situations in general. The last section of the chapter provides some conclusions on the human performance studies in aviation in general, and on the suggestions done for the link with the Dutch model in particular.

3.1 Cognitive frameworks of task performance and human error

3.1.1 Underlying mechanism

Information processing models have been the dominant models of human performance in psychology and human factors for some time. They have been developed, for example by Fitts (1954), Miller (1956), Berliner et al. (1964), and Wickens (1984, 1992) over a period of many years. The Wicken's version is probably the most accepted version of information processing, also in aviation. The principal feature of this approach emphasises input, processing, output and feedback loops, which are depicted in Figure 3-1. Sensory information is received by the body's various receptor cells and converted into neural impulses and stored in a system of sensory registers. The information received is then filtered and processed, with the processing (e.g. working & long term memory, judgment and decision making) occurring inside the human, leading to external actions (e.g. physical actions and speech) as a result. These outputs are continually monitored by feedback to the senses and memory to allow constant adaptation.

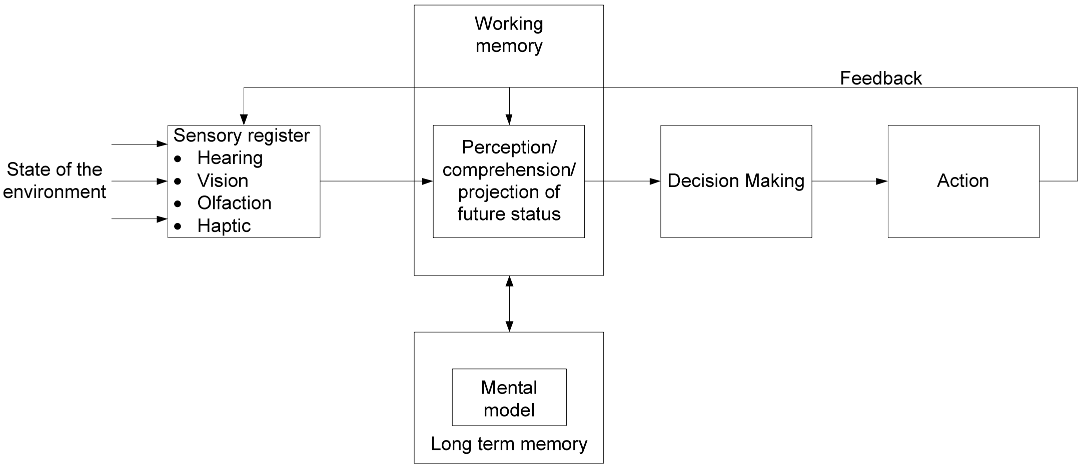


Figure 3-1 Generic model of human information processing (after Wickens 1992)

This model comprises a number of cognitive functions which may influence human information processing and errors can happen at any or many points during this process. Wickens (1992) notes that “information flow need not start with the stimulus […], sometimes our decisions or responses are internally triggered by “thoughts” in working memory” [p.20]. Nevertheless, all human actions (especially that of the flight crew) are performed in a specific context. Hence, human actions can only be described meaningfully in reference to the details of the context that accompanied and produced them (Dekker, 2006). This context is the combined effect of aircraft conditions, weather, traffic conditions, and performance shaping factors (PSFs) that together create a situation where an unsafe act is likely to manifest itself (or not). Swain (1983) introduced the term PSFs and it has become a common notion in the evaluation of human reliability. Swain defined PSFs as "any factor that influences human performance", and further differentiated PSFs into internal and external factors. By internal, he means operator characteristics that would affect the performance of a task, such as stress, fatigue, knowledge, personality, experiences, and attitudes. External PSFs include factors external to the individual, such as man-machine interface design, training programs, organizational structure and rewards, and written procedures. But in Swain’s definition and most projects, PSFs are usually mixed with contextual factors and organizational factors. So, PSFs have served in the past as a catch-all for explaining less-than-adequate human performance in complex systems without any clear definition and organizational hierarchical classification.

3.1.2 The factors essential for linking to safety management

The PSFs defined by Swain could interact with the cognitive factors at any or many points during the process illustrated in Figure 3-1. They affect the way in which human error becomes manifest. But they do not always sit easily within human performance models, they are not always clearly identified in the cognitive frameworks of task performance and human error models.

Among many ways to categorize human error (e.g. Reason, 1990; Rasmussen, 1982), Isaac et al. (2002) and Shorrock and Kirwan (2002) have used the cognitive domains of information processing to demonstrate the PSFs and their relationship with the different levels of human cognitive error types (see Figure 3-2).

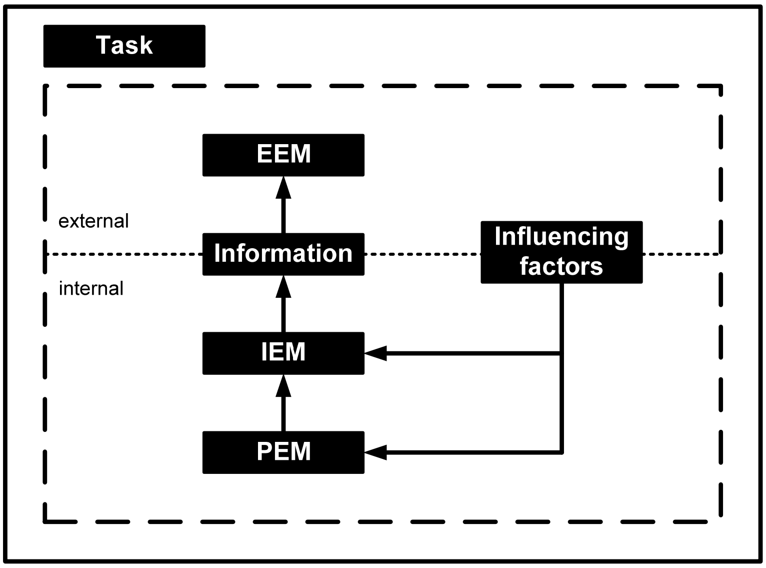


Figure 3-2 Relationship between the error types and the influencing factors

External error modes (EEMs) describe what happens in the real world and what error occurs, in terms of the external and observable manifestation of the error. EEMs can be formulated based on logical outcomes of erroneous actions, in terms of timing, sequencing, selection, quality, and so on (Shorrock and Kirwan, 2002). But EEMs (e.g. action too late, action too long) do not tell us anything about the cognitive origins of the error.

Internal error modes (IEM) describe what cognitive function failed or could fail, and in what way it failed. IEMs relate specifically to the "functions" of the cognitive domain. Following Wicken's model of human information processing, Shorrock and Kirwan (2002) classified four cognitive domains for air traffic controllers, namely "Perception", "Memory", "Judgment, planning and decision making", and "Action execution" (see Table 3-1). Each of the cognitive domains contains several functions, for instance, "action execution" was divided into "timing", "positioning", "selection", and "communication". Then each of the functions were combined with a keyword, such as early, late, incorrect, etc. to describe the internal manifestation of the error (such as late selection) within each cognitive domain.

The information box between EEM and IEM describes the subject matter or topic of the error, and the terms within the taxonomy relate specifically to the IEMs. For instance, what information did the flight crew misperceive, forget, misrecall, or miscommunicate? This is an important taxonomy because it highlights specific areas for error reduction. For instance, it does not help in knowing that a large number of memory failures occur if one cannot pinpoint what information is being forgotten, or alternatively what is being miscommunicated.

An error may be described as "incorrect decision making" for IEM, but going a level deeper, one might find that this was due to inappropriate expectations, or "stimulus overload". Such findings led to the creation and differentiation of "internal error mode" (IEM) from "psychological error mechanisms" (PEM). PEMs describe how the error occurred in terms of the psychological nature of the IEM within each cognitive domain. Thus, PEMs provide a very fine level of detailed information for error reduction and mitigation.

Table 3-1 Shorrock and Kirwan's (2002) taxonomy

Label	Selected examples	
External error mode (EEM)	**Selection and quality** • Omission • Action too much • Action too little • Action in wrong direction • Wrong action on right object • Right action on wrong object • Wrong action on wrong object • Extraneous act **Timing and sequence** • Action too long • Action too short • Action too early • Action too late • Action repeated • Mis-ordering	**Communication** • Unclear information transmitted • Unclear information recorded • Information not sought/obtained • Information not transmitted • Information not recorded • Incomplete information transmitted • Incomplete information recorded • Incorrect information transmitted • Incorrect information recorded
Internal error mode (IEM)	**Perception** • No detection (visual) • Late detection (visual) • Misread • Visual misperception • Misidentification • No identification • Late identification (visual) • No detection (auditory) • Hearback error • Mishear • Late auditory recognition **Memory** • Forget to monitor • Prospective memory failure • Forget previous actions • Forget temporary information • Misrecall temporary information	**Judgment, planning and decision making** • Misprojection • Poor decision • Late decision • No decision • Poor plan • No plan • Under-plan **Action execution** • Selection error • Positioning error • Timing error • Unclear information transmitted • Unclear information recorded • Incorrect information transmitted • Incorrect information recorded • Information not transmitted • Information not recorded

	• Forget stored information • Misrecall stored information
Psychological error mode (PEM)	**Perception** • Expectation bias • Spatial confusion • Perceptual confusion • Perceptual discrimination failure • Perceptual tunnelling • Stimulus overload • Vigilance failure • Distraction/preoccupation **Memory** • Similarity interference • Memory capacity overload • Negative transfer • Mislearning • Insufficient learning • Infrequency bias • Memory block • Distraction/Preoccupation **Judgment, planning and decision making** • Incorrect knowledge • Lack of knowledge • Failure to consider side- or long-term effects • Integration failure • Misunderstanding • Cognitive fixation • False assumption • Prioritisation failure • Risk negation or tolerance • Risk recognition failure • Decision freeze **Action execution** • Manual variability • Habit intrusion • Spatial confusion • Perceptual confusion • Functional confusion • Dysfluency • Misarticulation • Inappropriate intonation • Thoughts leading to actions • Environmental intrusion • Other slip • Distraction/preoccupation

The list of different levels of human cognitive error types and the causal connection between PSFs in Isaac et al. (2002) and Shorrock and Kirwan's (2002) models provided an excellent foundation for classifying cognitive errors from PSFs in a meaningful way. By extension from Figure 3-2, a human performance model for flight crew can be depicted as Figure 3-3.

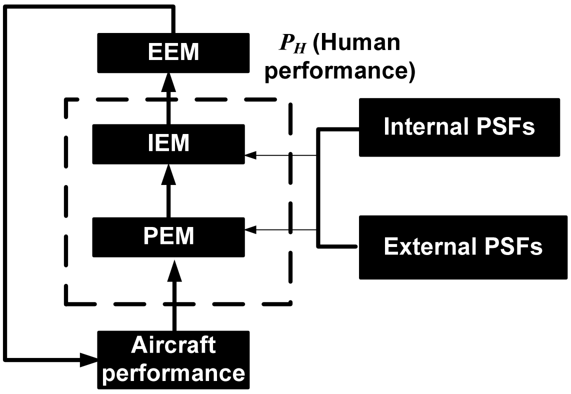


Figure 3-3 Human performance model

Internal PSFs are defined as physiological and psychological factors of the flight crew (e.g. fatigue, knowledge, personality, experiences, and attitudes) which influence the flight crew performance. External PSFs (e.g. man-machine interface design, written procedures) are defined as factors directly involved in the execution of a flight but external to the flight crew which influence the flight crew performance (in many cases equivalent to the outputs of the Dutch model's delivery systems). The external factors should be directly linked to the execution of a flight, but should not mix with a deeper set of organizational factors in a hierarchical structure, such as selection of appropriate staff or manning. Both the internal and external PSFs influence psychological mechanism within each cognitive domain (PEM); and the manifestation of the error will result in internal error modes (IEM) whose effect can eventually be observed as what error occurred (EEM).

Thus, in summary, the human performance can be formulated as a function of internal and external PSFs,

$$P_H = f(\text{internal } PSFs, \; ex\text{ternal } PSFs) \qquad (2.1)$$

where f can be seen as cognitive processing from PEM to IEM (the process within the dash box in Figure 3-3); P_H is the human performance including any errors, which can be observed as EEM. To yield effective error counter-measures, the analyst should classify the PEMs or IEMs in the human error mechanism for EEM[9] and identify what PSFs have aggravated the occurrence of the errors.

From a management point of view, it is considered more effective to modify the situations and threats which people are facing, rather than trying to influence directly their behavioural processes (i.e. the processes of handling the flow and transforming the information into action). To alter how a person's cognitive process functions is usually the last strategy

[9] A clear distinction between PEM and IEM may require significant understanding of the psychological aspects of an error and the available information about PEM from accident/incident investigations. To yield effective error counter-measures, the analyst should be able to classify the IEM or at least the cognitive domain which it is in.

management would look to for mitigation measures, simply because it is difficult to modify by management effectively. However, one should be aware that Shorrock and Kirwan's model does not use the dimension of Reason's classification dealing with "violations" and "errors", so violations are not explicitly included in Table 3-1[10]. Since their model does not explicitly use the distinction between violations and errors, it has limitation for us because what we are interested in is the link to management actions. There is a very different action from management if the unsafe act is due to errors, as compared to a violation. So, it lacks an important aspect to link it to those management influences.

Next, to be capable of shaping design and safety-related interventions from a management point of view, in the following section we will look at the human factors classifications coming from different approaches in the accident and incident investigation schemes for both influencing factors in our model in Figure 3-3.

3.2 Accident/incident investigation schemes

The influencing factors identified in this research must be clearly defined within an organisational hierarchical classification in such a way that management can devise resources and controls to ensure the continued correct functioning of human information processing and action. Taking this into account:

- The influencing factors should be clearly defined and as comprehensive as possible. They should come from a comprehensive search within the classification of the accident/incident reporting systems, data collection tools, and human factors taxonomies in aviation.
- The influencing factors should be directly linked to the execution of a flight. They can be either internal or external to the flight crew, but should not mix with a deeper set of organizational factors in a hierarchical structure. In other words they should be formulated as outputs of our Dutch model delivery systems, not as the delivery system process itself.
- The influencing factors should be classified in such a way as to reduce the multitude of error possibilities into a manageable set to model and to influence.

In order to create our taxonomy of influencing factors, a number of widely used accident/ incident reporting system, aviation human factors taxonomies, and data collection tools mentioned in the research of Beaubien and Baker (2002) and Stolzer et al. (2008) were preliminary reviewed against the criteria mentioned above. These tools are

Accident/incident reporting systems

- Aviation safety reporting systems (ASRS)
- Confidential human factors incident reporting programme (CHIRP)
- ICAO accident/incident reporting data reporting system (ADREP)

Taxonomies of human performance and human factor

- Human factors analysis and classification system (HFACS)
- Line Operations Safety Audit (LOSA)

Data collection tools

- The British Airways safety information system (BASIS)
- Aviation casual contributors for event reporting systems (ACCERS)

[10] One cannot add violation because it fits in all of the factors in Table 3.1., e.g. one can have an 'action too much' which is a violation; one can also have 'poor decision' which is a violation. So Shorrock and Kirwan's model already incorporates the notion implicitly.

It had been expected that all tools for this purpose would have some sort of explicit model as their basis. But after reviewing the taxonomies summarized in the research, it was found that many of the tools use text narratives, or do not build their human factors taxonomy according to any particular human cognitive model. For instance, Beaubien and Baker (2002) concluded that the data fields in ASRS were culled from those commonly used to describe previous accidents and incidents, but had not been developed according to any particular theory of human error; whilst CHIRP was modelled directly on the ASRS system. Moreover, for neither BASIS nor ACCERS is there documentation showing whether they were developed based on any particular model, and no clear detailed model is apparent. Hence these four taxonomies were discarded from our list. Consequently, only those systems containing fields of causal/contributing factors or explicitly distinguishing error types are relevant for review here. These are HFACS, ADREP and LOSA, which will be dealt with in the following sections.

3.2.1 Human Factors Analysis and Classification System (HFACS)

HFACS (Wiegmann & Shappell, 2003) was specifically developed to define the latent and active failures implicated in Reason's Swiss cheese model (Reason, 1990) so that it could be used as an accident investigation and analysis tool. As described diagrammatically in Figure 2-9, the HFACS framework consists of 4 levels of failures: 1) Unsafe Acts, 2) Preconditions for Unsafe Acts, 3) Unsafe Supervision, and 4) Organizational Influences. A brief description of the major components and causal categories can be found in Section 2.2.1 and Appendix B. Chapter 2 dealt with levels 3 and 4; here we deal with levels 1 and 2.

HFACS is one of the most commonly referenced tools of this kind in aviation. There are a number of desirable qualities of the HFACS's classification according to our criteria. First, the underlying factors are classified in a manageable set. Second, the structure is organized hierarchically, which makes it suitable for our purpose to develop management support relative to them. In one study (Wiegmann & Shappell, 2001), 319 National Transportation Safety Board (NTSB) human causal factors were identified using HFACS taxonomy and no additional category codes had to be added. This suggests that it is relatively comprehensive and requires few, if any, additional fields.

However, it does have some ambiguities. Despite the authors' claim that HFACS has passed a content validation and an inter-rater validation (which tests whether the users are able to identify similar causal factors and reach the same conclusions during the course of an accident/incident investigation by using the HFACS classification), it does not necessarily guarantee a hierarchical classification. The distinction between categories distributed between the active (Level 2: Preconditions for Unsafe Acts) and latent threats to safety (Level 3: Unsafe Supervision & Level 4: Organizational Influences) may not be clear enough. For instance in Figure 3-4, "procedures" have been classified in the highest level as organizational influences in HFACS; whilst it is true that airlines are responsible to ensure the good quality of procedures, the procedures themselves are involved significantly in diagnosing a situation and acting on it in the pilots' information processing. In this respect "good procedures" should be also considered as preconditions for unsafe acts.

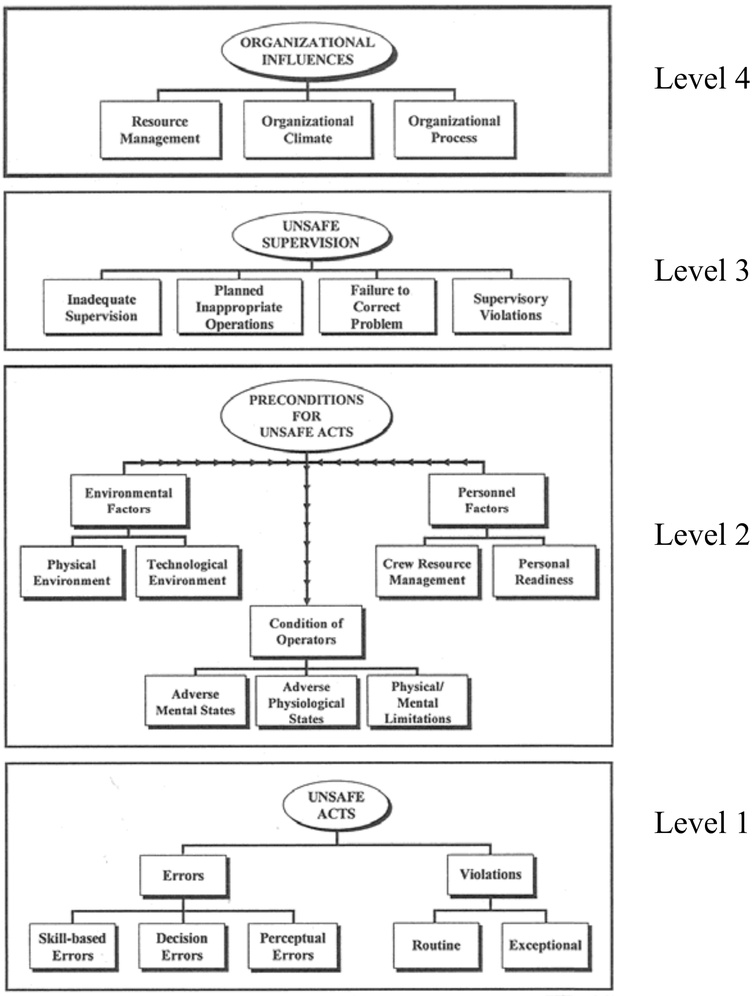


Figure 3-4 Human Factor Analysis and Classification System (HFACS)

Besides, some of the factors in the selected examples of level 2 are not explicitly formulated as factors influencing the flight crew performance and explain why a human error took place. For instance, "failure to conduct adequate briefing" and "failure to communicate and coordinate" are given as examples of underlying causes in HFACS. However, they are still phrased as observable productions of human actions rather than factors influencing the process of those actions. In this respect, they do not precisely help managers to devise relevant control measures to them. These failures could be due to the poor quality of the communication equipment, lack of interpersonal skills, or due to an unfavourable trans-cockpit authority gradient. So, it requires caution when using such factors as preconditions of human performance without clear identification. Although there are a few categories of causal factors in HFACS that needed to be resolved, HFACS is the one among the widely used human factors taxonomies in aviation which has a relatively clear hierarchical classification scheme which is classified into a manageable set to model and to influence. Thus, we will take and build on its taxonomies for our modelling.

3.2.2 ICAO Accident/Incident Data Reporting System (ADREP /ECCAIRS)

The worldwide accident/incident data reporting system (ADREP) was established after ICAO Accident Investigation and Prevention (AIG) meeting in 1974. ICAO has recommended its member States to use ECCAIRS (European Co-ordination Centre for Aviation Incident

Reporting Systems), developed by the European Commission, which allows States to share safety information about accidents and/or incidents, based on the ICAO-developed ADREP 2000 taxonomy (ICAO, 2000). States are therefore applying ECCAIRS as a tool to report accidents and serious incidents to ICAO. According to Directive 2003/42/EC (Official Journal of the European Union, 2003) of the European Parliament, pilots, air traffic controllers, airport managers, aviation maintenance technicians and aircraft ground handlers are mandated to report occurrences[11] to the competent authorities and EU member states are required to put in place a mechanism to collect, evaluate and store these aviation occurrences in a database. Hence the ECCAIRS's database serves as a European repository of occurrence data to which all ECCAIRS users have access.

In the ADREP taxonomy, human errors are classified into two levels of failures: 1) errors in operating the aircraft (what is the human error?) and 2) the underlying causes (why a human error took place?).

1) Errors in operating the aircraft are coded into 122 descriptive factors in the ADREP taxonomy, which are grouped into 5 categories:

- Flight crew's perception/judgment (perception)
- Flight crew's decision error (judgment, planning, and decision making)
- Flight crew's operation of equipment error (action execution)
- Flight crew's aircraft handling error (action execution)
- Crew action in respect to flight crew procedures (violation)

In each category, human errors are described in more detail. For instance, that the flight crew erroneously decide to initiate a flight, or the flight crew misinterpret the percevied warning. Factors mentioned at this level can be mapped onto the IEM/PEM categories defined in Section 3.1.2. For a complete listing of ADREP's human errors, refer to ICAO (2000).

2) The underlying causes that can be mapped onto the PSFs in our model are clustered into 5 categories:

- Human
- Human-environment interface
- Human-hardware/software interface
- Human-system support interface
- Human-human interface[12]

These categories are detailed into four levels of sub-categories, in more than 250 explanatory factors at the greatest level of detail in the ADREP taxonomy. Table 3-2 gives some selected examples of the detailed causes fitting into each of the 5 main categories.

The advantage of this taxonomy for the underlying causes is that it is relatively comprehensive, including a large and extensive list of factors that can be used to describe a wide range of PSFs from minor to more serious occurrences. But, the disadvantage of ADREP classification is that its classification scheme is more like an archive rather than an error framework that can be used for accident investigation and data analysis. For instance, there is no hierarchical distinction between levels within the underlying causes. Company

[11] 'Occurrence' means operational interruption, defect, fault or other irregular circumstance that has or may have influenced flight safety and that has not resulted in an accident or serious incident.

[12] These five are similar to the SHEL model (Edwards, 1972).

regulatory issues and weather information are put in the same category ("interface between human and the work environment") and in the same level in the hierarchy. While these are less of a problem for the original use of the taxonomy (coding of accidents and incidents), for the causal model it is an undesirable characteristic. The lack of multi-level modelling of these concepts causes ambiguities to the user when trying to distinguish the lower level failures and the more global organizational processes that govern the work activity. This is corroborated by Cacciabue (2000) who states that the overwhelming size of ADREP classification makes it of little use for data analysis.

Table 3-2 Selected examples of underlying causes- the explanatory factors

Human being	Flight crew's operation of auxiliary power unit
Personal size	Flight crew's operation of electrical system
Loss of consciousness/fainting	Workplace seat design inadequate
Impairment-chronic alcohol abuse	Inadequate information/data sources
Fatigue-rest/duty time	User friendliness/usability
Psychological-confirmation bias	Reliability of automation
Experience of route	**Interface between human and system support**
Interface between human and the work environment	Standard Operating Procedures
Landing/take-off site infrastructure	Emergency and abnormal procedures
Visibility from workspace/workplace	Company procedures
Cultural issues	Simulator training
Operational control personnel policies	**Interface between humans**
High workload due to staff/skills shortage	Interface between humans in relation to surveillance
Interface between the human and the hardware/software	Interface between humans in relation to cross-checking
Flight crew's operation of air conditioning	Interface between humans in relation to the use of teletype communications

In addition, while the majority of factors included in ADREP's PSFs are quite comprehensive, the PSFs classified under "psychological limitations" (one sub-category under "human being") are rather confusing (see Table 3-3). In this taxonomy, ADREP seems to mix the cognitive error modes (perception, decision making, action) with the underlying PSFs that create a situation where an unsafe action is likely to manifest itself. According to the model of Isaac et al. (2002) and Shorrock and Kirwan (2002), the factors in Table 3-3 certainly belong to the cognitive error modes not the PSFs. Besides, such a classification also causes confusion to the coders, whether they should classify the cognitive errors here (PSFs) or under the first level of failures in the ADREP taxonomy (descriptive factors). This indicates that the ADREP classification needs some reconciling.

As an industry-wide standard for accident and incident data collection, while not perfect, ADREP does provide a solid basis for data analysis. In order to assist States in developing safety data collection, analysis and exchange capabilities, ICAO has recently developed a safety data management training course based on the ECCAIRS aimed at promoting

ECCAIRS as a tool to code, enter, analyse and extract safety data. However, from our experience it is currently difficult to do the data analysis due to the fact that the underlying model of taxonomy is not clearly stated and hierarchically organized. We strongly recommend to ICAO that a "modified" or "simplified" multi-level structure accompanied by clear definitions should be developed and distributed to the analysts. This would certainly improve data quality, encourage reporting, enhance usage, and allow identification of systematic shortcomings.

Table 3-3 Psychological limitations in ADREP's PSFs (factor code: 103000000)

Factor code	Factor subject	Factor code	Factor subject
103010000	Action or lack of action	103040000	Perception & monitoring
103010100	Action-slip	103040100	Psychological-perception
103010300	Action-mistake	103040200	Psychological-attention
103010400	Procedure violation	103040300	Psychological-monitoring
103010500	Timing	103040302	Monitoring displays
103010700	Psychological error-other	103040303	Monitoring outside world
103020000	Psychological planning	103040305	Monitoring a person
103020101	Pre-flight planning	103040500	Psychological-vigilance
103020102	In-flight planning	103040600	Psychological-distraction
103020200	Action-preparedness	103040700	Channelized attention
103020300	ATC planning	103040800	Attention habituation
103030000	Information processing	103040900	Attention-other
103030100	Action-decision making		
103030400	Mis-recognition		
103030500	Misunderstanding		
103030600	Assumption incorrect		
103030700	False hypothesis		
103030800	Confirmation bias		
103030900	Mind set/expectancy		
103031000	Psychological-habituation		

3.2.3 Line Operations Safety Audit (LOSA)

The Line Operations Safety Audit (LOSA) is a direct observation of task performance from online behaviour audits. The underlying conceptual framework is known as Threat and Error Management (TEM) (Klinect, 2005; ICAO, 2002). In LOSA, trained observers fly along in the cockpit and record the types of "threats" and "errors" being made, and how flight crews manage these situations to maintain safety during normal operations.

The conceptual foundations of LOSA consist of the error, the threat, and the error and threat management. "Error" is defined as flight crew actions or inactions that lead to a deviation from correct performance which reduce safety margins and increase the probability of adverse operational events on the ground or during flight (see Appendix D, error codes). Our study of LOSA's "error" codes demonstrates that their "error" codes are formulated in aviation language and on the basis of standard operational procedure (SOP). "Threats" are external situations increasing the operational complexity of the flight, which fall outside the influence of the flight crew (see Appendix D, threats codes). LOSA data shows how often particular deviations occur and are corrected. The results of the audit are currently used to inform and help carriers (mainly in America, Asia, and Oceania) to define and improve their crew resource management (CRM) training.

A study of LOSA's "error" codes shows that their classifications of flight crew errors focus mainly on things observable by the auditors, so LOSA's modelling philosophy is mainly based on the EEM (external error mode) rather than internal psychological error modes, which are not directly observable. "Threat" codes give us rich sources for elicitation of PSFs, particularly the external ones, because LOSA's threat codebook considers almost exclusively the directly observable influencing factors visible in the online audits, such as environmental factors, threats induced by other service providers, and unexpected aircraft malfunction. However, there is no information on internal threats existing or happening inside a person, such as inappropriate experience or knowledge, that could be used for building our internal PSFs.

3.2.4 Interim conclusion

When searching through the taxonomies of these data collection tools, we found that data collection tools concerning the human factors in aviation are extensive, but relatively unsystematic. Among these, perhaps HFACS and LOSA are the most organized tools of this kind according to our studies. All of these models treat human errors as a general category, without classifying their external error modes or delving into psychological error modes. We have uncovered some lack of clarity and ambiguity in relation to the category of "procedure" in the PSFs of HFACS, which need to be resolved and also some factors which are in fact observable productions of human actions rather than influencing factors. In LOSA audits, the internal (physiological and psychological) factors related to the flight crew are not included in its taxonomy since it only counts directly observable influencing factors visible during the online audits.

Therefore, to build a taxonomy which is comprehensive to a sufficient extent, which has a hierarchical classification and is limited to a manageable set, based on the theoretical considerations from Isaac et al. (2002) and Shorrock and Kirwan (2002) we reduced the taxonomies of influencing factors from HFACS and LOSA to a coherent and manageable set. Next, the taxonomies were supplemented with the factors from the extensive lists from ADREP and checked to ensure that, as far as possible, they were comprehensive and mutually exclusive. Table 3-4 gives the descriptions and the selected examples of each category. The last column links them to the concepts at a management level of the delivery systems. A complete set of contents from ADREP PSFs were mapped onto the existing delivery systems in Appendix E to see if the Dutch model can support the control functions linked to them.

The mapping results from Table 3-4 and the complete table in Appendix E show that there is a good match, apart from the last category (including traffic configuration, possible flight delays, weather) for which we need to add a management function which influences them to the current Dutch model. With this modification the essential human factors can be supported by the management control functions. However, although there is generally a good match, there are four remarks for the current Dutch model. First, concerning the delivery system of "competence and suitability" in Table 3-4, it seems desirable to split this delivery system into its two component parts of "competence" and "suitability", since these are managed quite separately in airlines.

Second, communication and coordination between flight crew, ATC, and operation centre fit better under the communication and coordination delivery system. But, data and information

Table 3-4 Influencing factors for flight crew

Definition	Description	Selected examples	Delivery systems
Internal PSF: physiological and psychological factors of the flight crew which might influence human information processing	-Technical and interpersonal skills that match the requirements of performance	• Experience for complex situation • Technical skill • Communication& coordination between flight crew, ATC, and operation center	• Competence • Communication&coordination
	-Physical fitness to perceive, process, respond to information and feedback information	• Human physical and sensory limitation • Medical illness • Hypoxia • Physical fatigue • intoxication	• Suitability • Availability
	-Psychological fitness to perceive, process, respond to information and feedback information	• Mental fatigue due to sleep loss or other stressors • Emotional state • Personality characteristics (complacency, overconfidence)	• Suitability
	-Decision to choose one from several possible courses of action and decide to commit to safety procedure above other personal and organisational goals	• Pressure to achieve • Personal objectives • Incentives, needs • Perceptions of organization's beliefs and attitudes (manifested in actions, policies, and procedures, affect its safety performance)	• Motivation to commit to safety
external PSF: Factors external to the flight crew which might influence human information processing	-Clear and relevant guides to adequate performance	• Operational manual • Checklist • Charts • Standard operating procedures • Emergency and abnormal procedures	• Procedure
	-Tools and materials relate to physical activity designed scientifically to match human factors (include working postures, materials handling, repetitive movements, work related musculoskeletal disorders, workplace layout, safety and health)	• Checklist layout • Display/interface characteristics • Automation/alerts/warning	• Techonology-Man-machine interface (Instruments& workplace design)
	-Data and information	• Information from ATIS • Information from operation center	• Communication&coordination
	-Environment in which the action needs to be performed	• Traffic configuration o Traffic density	• No relevant management functions in the current Dutch

		o Traffic complexity (e.g. Runway length, Runway crossing, Runway condition, Runway slipperiness, Terrain at/near airport) • Possible flight delays • Ambient environment o Wind shear o Cross wind o Visibility in flight o Visibility at airport o Turbulence o Icing o Light condition	model. Need to add under "Workload" delivery system (will explain in Section 7.2.7).

should be explicitly identified under this management system because they are often the main purpose of the communication. Communication can take place either verbally or non-verbally (e.g. through written message, gestures, etc).

Third, in previous projects we have been inconsistent in deciding exactly where fatigue is covered in the model. It can be considered either under suitability as a temporary health issue, or under the availability of competent and suitable personnel. Since the fatigue management provided by airlines is managed under cockpit manning and rostering, we consider it better for this project to see it as an aspect of availability.

Finally, the most important observation resulted from mapping is that it echoes the criticism of the delivery systems in Section 2.1.5, namely that they are still too vague and generic in respect to the output of specific influencing factors. In Chapter 7, we will tailor the delivery systems for each of the new/modified categories of the factors found in the accident and incident analysis and make the delivery systems less vague and generic. The full discussion is to be found in Section 7.2.

3.3 Human factors in reliability-oriented techniques or PSA driven methods

We now turn to categories of human factors used in the techniques applied particularly in quantitative risk analysis. Human performance modelling for Probabilistic Safety Assessment (PSA) started in earnest in the 1970s. As human operators play an essential role at the operational level of any risk-bearing activity and their behaviour is different from and more complex than hardware performance and failure, a new approach was developed for Human Reliability Analysis (HRA) (Kirwan & Ainsworth, 1992; Kirwan, 1994). This used 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 PSFs (performance shaping factors) (THERP; Swain & Guttmann, 1983). This technique calculates the human error probabilities by identifying the sorts of PSFs external or internal to the individual. Dependencies and interactions between PSFs are treated with influence diagrams or Bayesian Belief Networks (BBNs).

The objective of this part of this thesis is to review the PSFs formulated at the human factor level in HRA. Because aviation is a different activity from other process industries, which has a great deal of behavioural barriers interacting with sophisticated hardware, we choose only to review techniques for this section that are currently used in aviation as methods. This limits this section to CATS (Ale, 2009), SoTeRiA (Mohaghegh, 2007; Mohaghegh & Mosleh, 2009) and IRP (EUROCONTROL, 2006). The quantification technique is not the focus of this section; it is the classification framework and how they have been formulated to be used in quantification.

3.3.1 Causal Model for Air Transport Safety (CATS)

The human performance models in CATS focused on those human actions that can influence the accident scenarios, but were also designed to link to managerial and organizational influences on human performance. As mentioned in the introduction of this thesis, CATS used the structure of the management model developed in earlier studies (i.e. I-Risk, ARAMIS, WORM – see Chapter 2 for a detailed explanation) to categorize the influencing factors and to cluster them. The default expectation was that all of the management elements would be relevant (seen as delivering or managing the success mode of each of the influencing factors) to each human performance area modelled.

Hence it was expected that the human performance model would contain all management functions, categorized under the Dutch model headings of "competence", "communication & coordination", "technology-interface", "availability", "procedure", and "commitment & conflict resolution" (see Figure 3-5). Within the CATS project, the formulation of the human performance models has treated human errors as a general category, without classifying their external error modes or delving into psychological error modes.

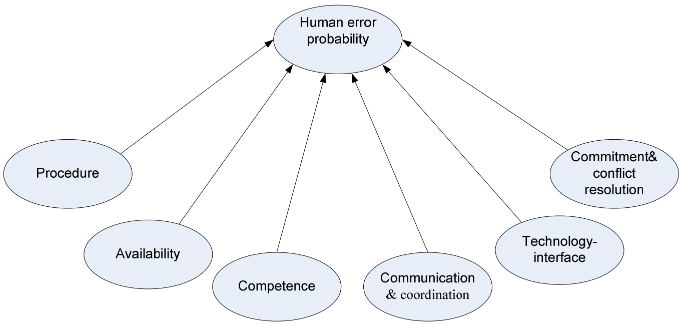


Figure 3-5 Scheme of human performance model in CATS

The formulation of the influencing factors which were initially selected to be modelled in CATS are presented in the first column in Table 3-5. "Pilot attitude" and "procedure" which had been regarded as important in the beginning of the project, but were considered too complicated to represent quantitatively at the stage when the CATS was under development were eventually left out completely. It was concluded in the CATS final report that these factors which were left as issues for later concern in a further phase of CATS should be developed in a subsequent stage. The final lists of selected PSFs used in the flight crew error part of the CATS model are operationalised in column 2 of Table 3-5.

Column 3 and column 4 show the mapping in respect to the list of influencing factors summarized in Table 3-4 and their relevant management influences. Compared to what was summarized in Table 3-4, factors modelled so far in the development phase of CATS have been a very limited subset of actually relevant factors and were restricted in operationalisation so that they could be easily quantified. This meant that the categories of "motivation to commitment to safety" and "procedures", which are found prominently in Table 3-4, were totally left out in the human factors modelling used in CATS.

Table 3-5 Performance shaping factors for flight crew in CATS

Initially selected as important aspects in CATS	Operationalized PSF in CATS	PSFs identified in this research (Table 3-4)	Delivery systems (Table 3-4)
Experience	total number of hours flown	Technical and interpersonal skills that match the requirements of performance	Competence
Training	the number of days since the last type recurrent training		
Intra-cockpit communication	Number of flights in which the pilot and first officer will have a different mother tongue		Communication& coordination
Fatigue	Stanford Sleeping Scale	Physical fitness to perceive, process, respond to information and feedback information	Availability
Pilot attitude	(left out)	Decision to choose one from several possible courses of action and decide to commit to safety procedure above other personal and organisational goals	Motivation to commit to safety
Procedure	(left out)	Clear and relevant guides to adequate performance	Procedure
Technology interface	Four aircraft generations	Tools and materials relate to physical activity designed scientifically to match human factors	Technology-Man-machine interface
Weather	Rainfall rate in mm/hr	Environment in which the action needs to be performed	No relevant management functions in the current Dutch model. Need to add under “Workload” delivery system (to be explained in Section 7.2.7).
Workload	Number of times the crew members have to refer to the abnormal/ emergency procedures		

3.3.2 Socio-technical risk analysis (SoTeRiA)

SoTeRiA (Mohaghegh, 2007; Mohaghegh &Mosleh, 2009) is a hybrid modelling technique for integration of organizational and social aspects with the technical system. Similar to CATS, the events, conditions, and causes of the accident scenarios are modelled through an integration of Event Sequence Diagram (ESD), Fault Tree (FT), and Bayesian Belief Network (BBN).

SoTeRiA used a quantitative example in airline maintenance systems to demonstrate the feasibility and value of its hybrid techniques. A human performance model for maintenance technicians was developed in detail (see Figure 3-6).

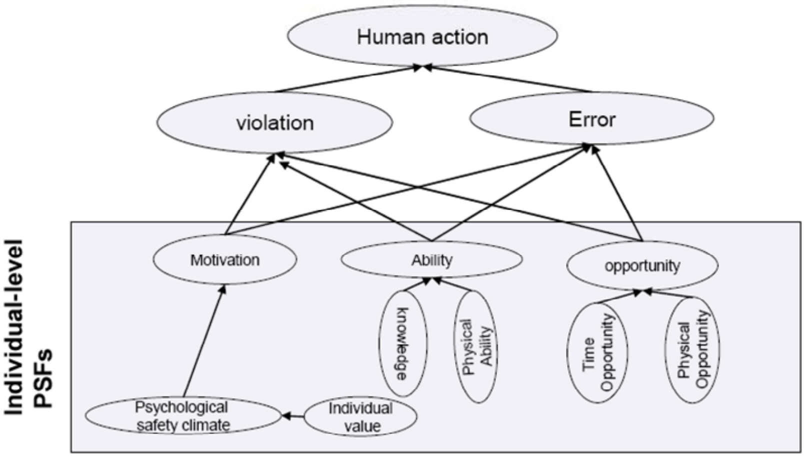


Figure 3-6 Schematic representation of SoTeRiA[13]

In SoTeRiA, the author considered “motivation”, “ability”, and “opportunity” as individual-level PSFs. “Motivation” is defined as most affected by psychological climate and individual values. “Ability” is influenced by knowledge and physical ability. “Opportunity” is seen as some temporal opportunity (or lack of it) such as “time opportunity” (e.g. time pressure due to work schedule) or “physical opportunity” (due to physical working environment such as lighting)” (Mohaghegh, 2007). Table 3-6 shows the important factors of these categories identified in SoTeRiA and how they were finally chosen to be implemented in that model.

Column 1 shows the PSFs taken from Figure 3-6. Column 2 shows the final lists of factors selected in SoTeRiA. For example, “knowledge” and “time opportunity” were modelled as “level of experience” and “time pressure” respectively, both influenced by management modules of “training” and “hiring” in SoTeRiA. “Physical ability” and “physical opportunity” can be placed against to the taxonomies of “suitability” and “technology man-machine interface” in Table 3-4, but they were not represented in the final quantitative version of SoTeRiA. The last column shows the mapping with the existing delivery systems. It confirms the need to split the delivery systems for competence and suitability into its two component parts as noted in the earlier interim conclusion section (3.2.4).

[13] Courtesy of Zahra Mohaghegh copied from Figure 5.13 in p194 of Zahra Mohaghegh’s thesis (Mohaghegh, 2007)

Table 3-6 Selected performance shaping factors for a maintenance technician in SoTeRiA

Selected important aspects in SoTeRiA	Operationalized PSF in SoTeRiA	PSFs identified in this research (Table 3-4)	Delivery systems (Table 3-4)
Motivation	morale	Tentatively mapped	Tentatively mapped
Psychological climate		Decision to choose one from several possible courses of action and decide to commit to safety procedure above other personal and organisational goals	Commitment & motivation
Individual value			
Group safety climate			
Ability			
Knowledge	level of experience	Technical and interpersonal skills that match the requirement performance	Competence
Physical ability	not mentioned	Physical fitness to perceive, process, respond to information and feedback information	Suitability
Opportunity			
Time opportunity (time pressure due to work schedule)	time pressure	Physical fitness (fatigue) to perceive, process, respond to information and feedback information	Availability
Physical opportunity (due to physical working environment such as lighting)	not mentioned	Tools and materials relate to physical activity designed scientifically to match human factors	Technology-Man-machine interface

In contrast to other models, which do not deal with "safety climate" as a set of PSFs, SoTeRiA did try to include variations of safety climate as a set of important factors (e.g. motivation, psychological climate, individual value). In Mohaghegh's publications there are some confusions about what safety climate means, arising largely from a lack of clear definitions of categories and sub-categories (e.g. what does "Motivation" or "Psychological safety climate" mean?). These were left at a higher level of abstraction, and eventually were missing in the final operationalisation in SoTeRiA and further named as "morale" in the maintenance technician commitment module. Although they did not given clear explanations we tentatively map the notions to our commitment dimension. We will come back to this issue in Chapter 7 to explore how the concept of safety climate influences behaviour and relates to safety management.

3.3.3 Eurocontrol Integrated Risk Picture (IRP)

Eurocontrol's Integrated Risk Picture (IRP) (EUROCONTROL, 2006) was initiated in the Safety Research and Development section at the Eurocontrol Experimental Centre (EEC). It was developed for air traffic management (ATM) in Europe, showing the relative safety priorities in the gate-to-gate ATM cycle, and the safety impacts of future ATM developments. The IRP aimed at generating feedback for improvement of system safety performance, i.e. where to invest in safety. Moreover, it identified critical R&D to deliver the safety targets, rather than blue sky R&D.

The research includes a baseline IRP for ATM as it is in 2005, and a benchmark prediction of how it will be in 2012. The IRP has been developed using techniques of fault trees, event trees and influence diagrams. Each accident category is represented as a separate fault tree. A barrier failure concept was chosen as the basis for fault tree development. The fault tree elements represent the distinct causal factors such as technical failures and human errors, which are the immediate causes of failure of the barriers against accidents.

Like the other two models, air traffic controller performance is presented as one of the common cause influences associated with a task in the ATM model. They indicate that the most important human factors underlying the controller performance in tactical separation are "reliability", "resources", "teamwork", "competence", and "human-machine interface". Cross referencing the categories they give, these map quite well onto Table 3-4 and later onto the Dutch model delivery systems. But there is no further development of the human performance model or description of any of these factors in the report. It is therefore impossible to cross reference it more fully.

3.3.4 Conclusion on PSA models

The three models we describe in this section show that the categories in Table 3-4, which we use as comparator, manage to cover pretty well all the factors from the models.

The factors identified from these three models show their potential to be linked to the management factors. However, the studies also demonstrate that the human factors formulated in the current (probabilistic) quantification models typified by CATS (either for flight crew, maintenance technicians or ATM) are still very partial, due to the strong focus on quantifiability. That means that the factors selected at the human factor level in the generic quantification models like the ones mentioned above are only a very limited subset of the potential and actual influencing factors related to qualitative analysis and shown in Table 3-4. From this limited subset of the influences even more are then left out of the model entirely because it is argued that they cannot be quantified successfully. This means that the factors considered in the risk models such as CATS are not a comprehensive overview of factors that have the potential to influence flight crew performance.

3.4 Overall Conclusions

As mentioned at the beginning of this chapter, the objective of this chapter was to identify the essential human factors which the management have to support. Initially, what we were expecting to find was more accident databases and reporting systems with clear classification schemes, which we could look at in detail in order to see whether we could take and build on those taxonomies which were sufficiently comprehensive and had a hierarchical classification for aviation. But we did not find any really suitable. What we found was that data collection tools concerning human factors are relatively unsystematic. Among these, only HFACS, LOSA and ADREP data met our basic criteria. Based on these three models, we built a human factor taxonomy for this research, as shown in Table 3-4. Factors were mapped on to the existing Dutch management model to see if the Dutch model can support the control functions linked to the human factors identified there. Comparing the human factors with the Dutch model, there is a good match, apart from the topics of "workload" and "competence and suitability". Additional functions for "workload" need to be added to the Dutch model and "competence and suitability" need to be split in Chapter 7.

The factors formulated in the current (probabilistic) quantification models were also discussed in this chapter. The current modelling in respect of quantifying the important set of the human factors covering the internal and external factors of the flight crew in these approaches seem in general quite limited. In sum, the factors identified from these models covered only the limited categories of factors that Table 3-4 recommends. Therefore, the most important improvement we can propose for human performance modelling in HRA is to get a better understanding of the relationship between the qualitatively generally well understood notions and then translate these into real, observable and thus quantifiable influences on risk and risk reduction. Chapters 5 and 6 are designed to develop a quantification method that could do so.

In Chapter 7, we will also reconsider the topics of safety culture and safety climate which we found extensively in SoTeRiA, to see if they fit sufficiently into the existing delivery systems.

4 Technical performance

In Chapter 1, we briefly described the modelling techniques for technical failure in the CATS project. The modelling is based on a combination of three techniques: Event Sequence Diagrams (ESD), Fault Trees (FT) and Bayesian belief nets (BBNs) (Figure 4-1). ESDs delineate the possible accident scenarios. FTs describe the events, conditions and causes of the scenarios. Each cause of a barrier failure in an FT is a base event. The base events of the fault trees include events representing technical failures and events representing human (un)reliability.

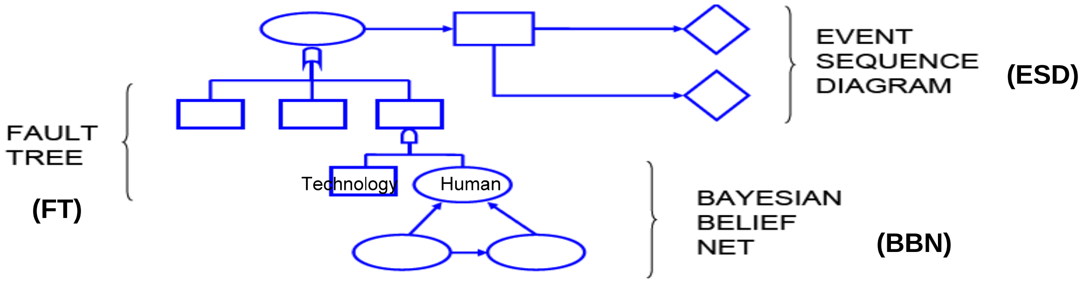


Figure 4-1 The basic constituents of CATS

In Chapter 3, the human factors and their influences on human reliability have been discussed. In this chapter we consider the technical performance of the aviation system and how the way it is represented can provide links to the management system modelled in Chapter 2. This chapter works out these links in principle, but will not go into great detail, as these aspects of management have not yet been worked out to an operational level in the current state of the CATS project. We concentrate in this chapter on the aircraft as system component, with its design, use and maintenance. However, exactly the same reasoning and approach can be applied for the ATC hardware/technology or that of the airport.

The term “technical failure” signifies malfunctioning leading to unsatisfactory performance of an engineered system which, if uncorrected, will lead the aircraft to crash or injuries to occur to the occupants. In general, malfunctions occur when a component or structure is no longer able to withstand the conditions that are imposed on it during operation or it is asked to perform (well) beyond its design base.

Failure of an aircraft engineering component is a phenomenon which has been known to occur in many major accidents/incidents in aviation history. A striking example is the Space Shuttle Challenger disaster (NASA, 1986) on January 28, 1986, when the Space Shuttle broke apart 73 seconds into its flight, leading to the deaths of its seven crew members. After an O-ring seal in its right solid rocket booster failed at lift-off, an explosion over the Atlantic Ocean off the coast of central Florida caused the entire vehicle to disintegrate. The direct cause was the O-ring failure, but its design had contained this potentially catastrophic flaw since 1977, which the designers and users had failed to address properly.

It is comparatively rare for an accident/incident to occur exclusively due to the malfunctioning of a technical component; often there is a human involvement in the semi-automated system. This partial automation can be clumsy for human to interact with, making it difficult to program, monitor, or verify, especially during periods of high workload. This is related to Man-Machine Interface (MMI) problem. An example is the following accident in which neither pilot was aware that the autothrottle system had disengaged with the thrust levers at idle during an instrument landing system (ILS) approach to Bournemouth Airport, England (AAIB, 2009). Why the pilots did not see the flashing red light on the instrument panel that warns of autothrottle disengagement was unanswered, although it stated: “It is likely that flight crews are subconsciously filtering out what is perceived as a nuisance message”. The report said that the 737 did not have, and was not required to have, an aural indication of autothrottle disengagement. As a result, AAIB recommended that Boeing and the U.S. Federal Aviation Administration review the effectiveness of the autothrottle system disengagement warning in a number of series 737s and improve them if necessary. Although there was no direct malfunction of the component of the auto-throttle, its technology interface with the crew was an important factor which should have received more consideration in the design.

For the purposes of this thesis we consider these as two broad categories of technical performance:

- one deals with the correct technical functioning and design of the technology, leading to satisfactory performance of an engineered system. This includes the hardware and software onboard;
- the other deals with the interaction with the human operator, the ergonomics and human-centred design of the MMI.

In general, three main aspects (design, operations and maintenance) have to be dealt with to ensure satisfactory performance of an aircraft system over its entire design life in relation to both the technical functioning and the MMI (Figure 4-2).

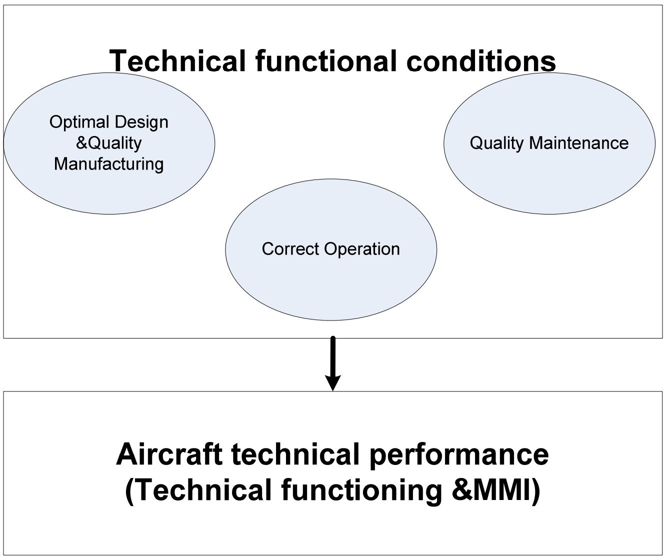

Technical functional conditions
Optimal Design &Quality Manufacturing
Quality Maintenance
Correct Operation
Aircraft technical performance (Technical functioning &MMI)

Figure 4-2 Aircraft technical performance model

A satisfactory performance (in a safety perspective) of an aircraft system can be formulated as:

$$P_T=f\,(a,b,c,d) \qquad (4.1)$$

P_T= aircraft technical safety performance (including functioning and MMI)
a = safe design and manufacturing

b= safe operation
c= safe maintenance
d= others (e.g. weather conditions)

Deficiency in any of these critical stages will hamper safety performance of the aircraft before the end of its design life. Safe design, safe manufacturing, and safe maintenance lie in the supporting processes that need error prevention and quality control. Safe operation concerns online operation of aircraft system within the design limits. This also requires an operating environment that matches with the design specifications in respect of subjects such as weather.

Figure 4-3 shows the human-aircraft system interaction model, which integrates the aircraft technical model with the human performance described in Figure 3-3. From Sections 4.1 to 4.3, we discuss each of the aspects in Equation 4.1. We do not have a separate section on the MMI life cycle in this discussion, but deal with both technical functioning and MMI throughout the four aspects of the life cycle.

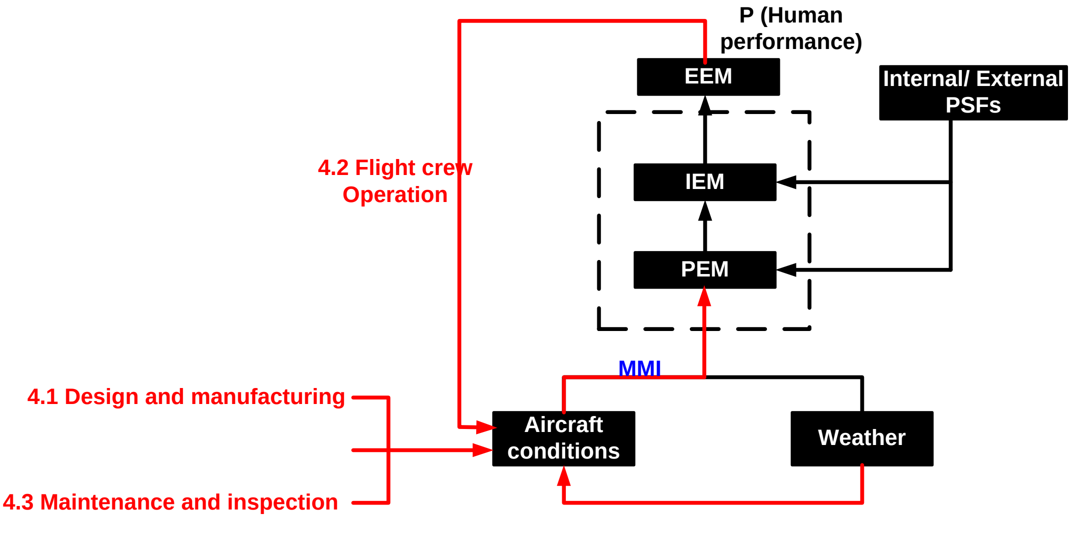

P (Human performance)
EEM
Internal/ External PSFs
4.2 Flight crew Operation
IEM
PEM
MMI
4.1 Design and manufacturing
Aircraft conditions
Weather
4.3 Maintenance and inspection

Figure 4-3 Human-aircraft system

4.1 Design and manufacturing

A number of accidents can be traced to errors in the conceptual design and manufacturing phases of aircraft. For example, on September 8, 1994 (NTSB, 1994a), a 737-300 aircraft of USAir lost control at about 6,000 feet (1830 meters) during an approach for landing at Pittsburgh International Airport, Pennsylvania. All five crew members and 127 passengers were killed and the airplane was destroyed by impact forces and fire. The National Transportation Safety Board (NTSB) determined that the probable cause of the USAir flight accident was the Boeing 737's rudder malfunction, including rudder reversals and the inadequacy of the 737 rudder system design. One year later, the Board recommended numerous rudder design changes to older 737s and Boeing agreed to retrofit older 737s with a new rudder system design.

Design and manufacturing process

The aircraft design cycle starts from specifying performance definitions, requirements, and system specifications. A conceptual design is developed to meet these requirements by taking modelling and technological ideas into consideration. After that, feasibility studies are carried out to review several alternatives against the conceptual design by considering their performance and cost analysis. In the last stage, the final design will be prototyped for simulation testing and performance evaluation.

For each aircraft model which passes its simulation certification testing and type approval, an entire fleet of airplanes with the same design is assumed to be built with the same geometry, loads, and material properties based on the assumed model from the design system. Even if we can assume perfect design and correct material selection, the system performance is still not guaranteed. This is because the system components may not meet the design requirements during the manufacturing phase. There are large numbers of process stages in the manufacturing and production of components ranging from melting of the alloy, casting, mechanical working, heat treatment, and metal joining, through to finish machining (Reddy, 2004). Each process stage has to be carefully monitored to ensure its correctness. To ensure the quality of manufacturing, the designer must be aware of manufacturability and incorporate this into the design procedure. The final component should be tested against all the system specifications under all possible operational conditions to make certain that the components are indeed of good quality. Hence, for aircraft functional safety it is important to control critical manufacturing elements by thorough quality assurance. Figure 4-4 shows a general aircraft design and manufacturing process as described above.

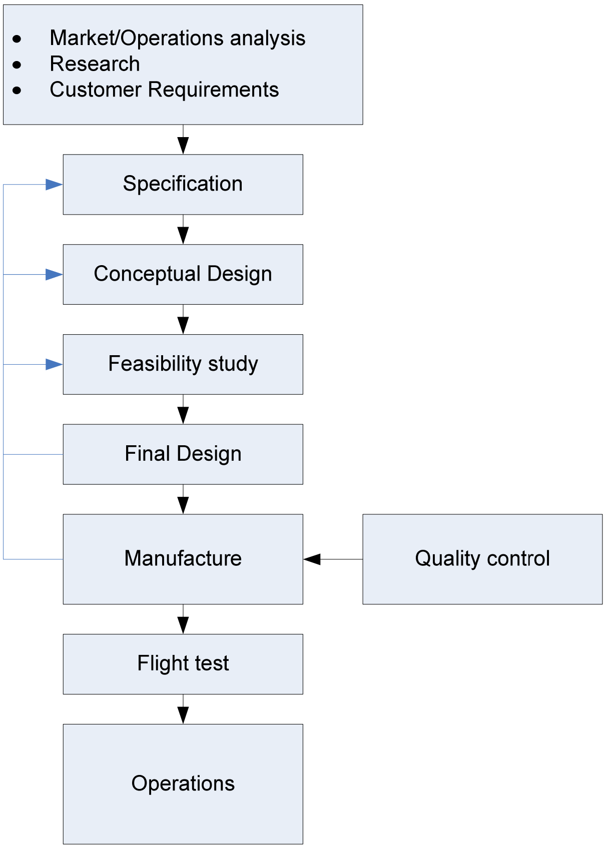


Figure 4-4 Aircraft design process

Failures in any block of this process[14] may cause structural failure and result in a catastrophe. A striking story of a conceptual design failure is the Turkish Airlines Flight 981 accident (BEA, 1976), which caused the worst air disaster up to that time (before the Tenerife Disaster event of 1977 and the crash of Japan Airlines Flight 123 in 1985) resulting in the deaths of all 346 on board. This accident was due to the failure of the cargo door closing mechanism. The crash resulted from the poor design of the rear cargo door latch system which was a new design at that time to swing the cargo door outward (instead of inward), allowing more storage space in the cargo area. However, this design allowed the cargo door to blow off by the pressure inside the cargo area when the locking pins were not fully engaged as a result of a poor design of the closing mechanism[15].

However, what also led to the Turkish airline disaster was the failure in modification of the design or redesign after a previous accident. The failure mechanism of Turkish Flight 981 was discovered in an accident several years previously. After this accident, the director of product engineering of Convair (a McDonnell Douglas subcontractor during the early 1970s), Dan Applegate, wrote a memorandum to his management noting that "the airplane demonstrated an inherent susceptibility to catastrophic failure when exposed to decompression of the cargo compartment in 1970 ground tests". He pointed out a potentially fatal crash would seem imminent and changes were needed. However, changing or upgrading an airplane design is more complex than we might imagine. To make a simple change, the process may take at least two years. Therefore, there was debate about who (McDonnell Douglas or the sub-contractor) should take the responsibility for such conceptual system design decisions and end up paying for the changes (Fielder & Birsch, 1992). Meanwhile, McDonnell Douglas announced a number of minor changes to the system, but nothing fundamental was done to change the design. This eventually led to the Turkish Airlines disaster due to the same technical fault Applegate had foreseen two years prior. McDonnell Douglas's reputation and the reputation of the DC-10 were seriously harmed after this event. Thus, failures in modification can also cause structural failure and result in a catastrophe.

Design in CATS

When the current phase of the CATS project finished, "design" and "manufacturing" of the aircraft equipment had not been modelled in CATS. Since the processes in Figure 4-4 provide a vital contribution to safe operation they need to be linked to the management modelling in the future for "technology". Further development of the modelling of design could use the framework of this model and draw on other models of the design process, such as those in the special issue of Safety Science v45 (1&2) by Kjellén (2007), Drogoul et al. (2007), Kirwan (2007), and by Hale et al. (2007), in order to make more progress in this earliest intervention in containing technical failures.

4.2 Flight crew operation (Malfunction due to crew action or inaction)

On the one hand the development of safer and more reliable aircraft can only be achieved if it pays close attention to the skills and limitations inherent in humans operating the aircraft as

[14] Design failure such as inaccurate modelling of physical phenomena, errors in structural analysis, and errors in load calculations or manufacturing defects such as freckle defect and embedded particles below the coated surface may cause internal cracks in component, which will lead to component failure during operation.

[15] The blowing out of the cargo door led to the instantaneous loss of pressurization of the cargo area, which led to collapse of the cabin floor. This led in turn to loss of control of the aircraft, because the control cables for the rear control surfaces of the DC-10 were routed through the cabin floor.

they interact with the technology. This is reflected in the importance of aircraft cockpit design and human interfaces for controls and displays. On the other hand, while manoeuvring an aircraft, whether extensively automated or not, pilots must know the aircraft's structural and aerodynamic operating limits specified by the designers for the operations phase (see Figure 4-5).

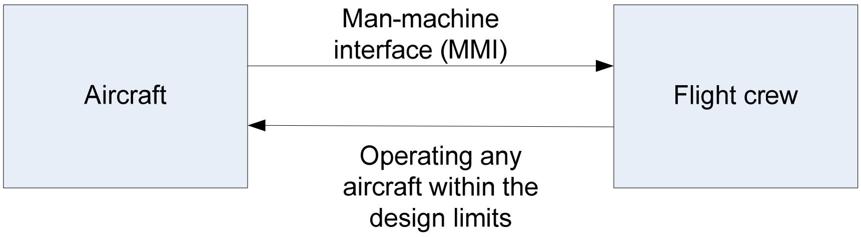

Man-machine interface (MMI)
Aircraft
Flight crew
Operating any aircraft within the design limits

Figure 4-5 Interaction between aircraft and flight crew

Operating aircraft system beyond the design limits (e.g. loads, airspeed, and altitude) and not following the procedures in this respect may cause irrecoverable damage to the components and shorten the lifespan of the airframe. To ensure that the aircraft structure is capable of withstanding all the loads imposed on it and performing its take-off with sufficient acceleration, the maximum takeoff weight (MTOW) is set at which the pilot of the aircraft is allowed to attempt to take off. Exceeding the structural limits (overstress of the aircraft) has a lasting effect on safety when it causes damage. If the overstressing could result in loss of ultimate integrity, it becomes a latent threat to safety if not detected. For example in an accident in 1992, an aircraft of Mongolian Airline, which was designed to hold only 17 passengers, was allowed to take 26 persons on board, which led to an accident killing all people onboard. Another striking accident of this kind was the American Airlines Flight 587 (NTSB, 2001), the second-worst aviation accident in United States history. The NTSB determined that the probable cause of this accident was the in-flight separation of the vertical stabilizer as a result of the loads beyond ultimate design that were created by the first officer's unnecessary and excessive rudder pedal inputs, which caused excedance of design stresses. Flawed manoeuvring training by the airline and poor rudder system design were major contributing factors to this overstressing, the board said.

In addition, flight crew might be inclined not promptly to report damage to the structural strength of the aircraft to the airline in order to avoid blame. This can be another problem which can cause damage to aircraft integrity. Hence, it is important for an airline to create a no blame culture rather than making its employees feel that they have to cover up human errors during mandatory mishap-checks, for fear of losing their jobs.

Following procedures in respect to operating the aircraft system within the design limits is consistently expected. However, pilots must have the ability and authority to deviate when essential, especially in an emergency situation. Normally pilots (except test pilots or military aviators) will not even approach and certainly not exceed the boundary of the flight envelope[16]. But, in an emergency situation, instrumentation to show and "defend" the envelope boundary, but equipped with an override, would probably reduce air accidents by indicating to the pilots where the boundary is, whilst helping them make a quick evasive

[16] The flight envelope is a concept designed for the purpose of warning the pilot away from making control commands that would force the aircraft to exceed its defined limiting operating conditions. It constitutes several variables, such as airspeed, altitude, G force, pitch, bank, and loading variables.

manoeuvre in response to extreme situations. An example of the latter was the FedEx flight 705 (Hirschman, 1997) hijacked by an employee passenger travelling in the jump-seat, who tried to take over the plane for the purpose of a suicide attack on April 7, 1994. During the fighting, the flight crew performed extreme aerial manoeuvres beyond the designed capabilities of the aircraft to keep the hijacker off balance. Although severely injured, the flight crew eventually landed the plane safely.

All the human factors issues mentioned above (pilot commitment to the load limit, aircraft manoeuvring training, pilot commitment to report damage to aircraft structural strength, or pilot ability to deviate from the design limits) can be found in Table 3-4 in Chapter 3 and can be mapped to the Dutch safety management model from there.

Operations in CATS

CATS dealt extensively and adequately with the pure technology failures of performance and functioning, but had some shortcomings in relation to the links to the human factors (MMI) and the human and organisational failures underlying the technical failures. In CATS, some of the "unsafe operations" which have an immediate unsafe interaction with aircraft performance (e.g. autopilot incorrectly used by flight crew, or brakes not applied correctly) were modelled in the base events of FTs. These actions often are the direct causes of technical system malfunctions, or have a deterministic impact in an accident. However, some operations which may or may not cause immediate damage to the aircraft, such as overstressing the aircraft by exceeding the structural limits, were not explicitly linked to the aircraft equipment failure nor clearly represented in the current CATS model.

In the further development of the model, there is a need for greater clarity in what should be modelled as an immediate cause to be placed in the base event of the FT and what should be modelled as underlying the malfunction of the technology and modelled in the human factors and management models. Currently the guideline is not clear. However, whether these human factors are plugged into the base events of the FT or into another level deeper underneath the malfunction of the technology, the human factors model devised for flight crew in Chapter 3 can be linked in.

4.3 Maintenance and inspection

After approval of the design and manufacturing, a sound aircraft maintenance system supports the continuing airworthiness of the aircraft. Maintenance can be defined as the process of ensuring that a system continually performs its intended function at its designed-in level of reliability and safety. There are two types of maintenance: scheduled maintenance and unscheduled maintenance. Scheduled maintenance is a preventive form of maintenance conducted at preset intervals to inspect and ensure that the aircraft is air-worthy. Unscheduled maintenance is needed after any failure event or the discovery of any unexpected technical wear or anomaly. Such breakdown maintenance actions are designed to restore the functionality of the system and put systems back into service.

4.3.1. Maintenance program (Kinnison, 2004;Kroes et al., 1993)

Aircraft maintenance is a complicated and costly business and is characterized by large amounts of regulations, procedures and documentation. 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 must have an approved aircraft maintenance program. This program is initially produced by the Maintenance Review Board (MRB) consisting of manufacturers and aviation authorities. Taken from there, it is the responsibility of the Engineering Department at the airline to package these tasks into workable units and ensure that all task limits are met (time, cycle, etc). The maintenance program should be customized for the individual airline, depending on the airline's fleet size, route structure, aircraft utilization and from years of operational experience.

The actual planning and scheduling of all aircraft maintenance activity within the airline is a detailed and complex process. Most airlines have departments with titles such as, "Production Planning and Control Department (PP&C)", (whose tasks are shown in Figure 4-6) dedicated to the detailed forecasting, planning, and control of aircraft maintenance tasks and inspections for individual aircraft. Aircraft maintenance is usually scheduled in "checks" of varying proportions, ranging from daily checks ("A" checks) to heavy maintenance checks ("D" checks)[17]. The PP&C Department estimates the maintenance workload for the long term and the short term based on the existing fleet and business plans and on any known changes in these for the forecast period. Then, it schedules all aspects of these checks including manpower, parts, supplies, and facilities. When the aircraft comes in, coordination with flight operations, ground handling and support activities has to take place in order that everything will be ready. Once the actual contents of a check are known, the PP&C plans which particular aircraft is due for which check and prepares and produces all of the "task cards" and documentation necessary to the technicians.

To support the technicians in carrying out the tasks correctly, all tasks have a reference to the relevant section of the maintenance manual. It is the responsibility of Engineering Department to develop maintenance manuals with inputs from the practices of the maintenance organisation.

Before the aircraft is scheduled to arrive in the maintenance workshop, the completed work package for that aircraft should be delivered to the technicians. The work package specifies all the tasks of maintenance and inspections which must be conducted for a scheduled check on a particular aircraft. It contains a list of contents and task cards. They describe information pertaining to the type of aircraft, specify the repair job, procedure for repair, and note additional materials required. However, this planning and scheduling process is not as straightforward as it might seem. Many works will be predictable beforehand, but the rest are raised only as the results of the inspections during the actual performance of maintenance. After the inspection and maintenance tasks have been done, a second maintenance person (the inspector who is responsible for monitoring the work progress and output of the work) may re-inspect a repair before the item is closed out. In order to keep the schedule, in certain pre-defined situations it is possible for the technicians to notify planned jobs which have not been done and therefore defer the maintenance to a later check (deferred maintenance) or notify additional personnel to complete the non-critical task. All scheduled items and additional items must be either certified and signed off as complete or logged as deferred.

[17] For instance, The "A" check for a 747-400 is done every 600 flight hours; the "B" check every 1200 flight hours, the "C" check every 5000 flight hours or 18 months and the "D" checks between 25,000 and 27,500 flight hours for the first time, and subsequently every 25,000 flight hours or 6 years.

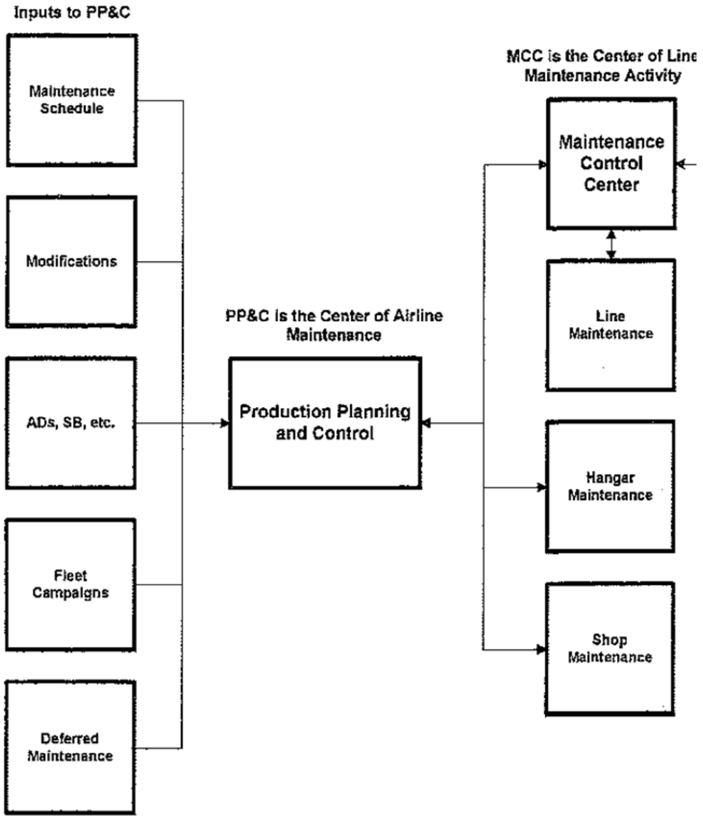


Figure 4-6 Functions controlling maintenance (taken from Kinnison, 2004)

In addition, from time to time, the manufactures develop modifications and improvements for their systems. These are issued as service bulletins (SBs), and service letters (SLs). If a safety or airworthiness issue is involved, the mandatory modification is issued as airworthiness directives (AD) by the authority. Once the decision to incorporate the changes has been made by airlines, their engineering departments have to provide the necessary information needed by maintenance to accomplish the modification. So together with the feedback from the line maintenance and the modification from manufactures and authorities, that is listed on the left side of Figure 4-6. In that case the PP&C has to alter the plan and integrate these into the maintenance activities for future checks.

The abovementioned scheduled and unscheduled maintenance is divided between the hangar, the shops, and the line as necessary (shown in Figure 4-6). “Line maintenance” is those tasks done without removing the aircraft from the flight schedule, i.e. daily, 48 hours and transit checks, and in most airlines the “A” and “B” checks. In line maintenance, if a discrepancy occurs, normally it will be written up in the aircraft maintenance logbook. While the aircraft is standing at the gate in-between two flights, the maintenance personnel have to troubleshoot, analyze the problem, and perform the corrective actions as soon as they can to minimize delay on the ground. “Hangar maintenance” refers to the maintenance which is done on out-of-service aircraft. This kind of work is scheduled as “C” and “D” checks, heavy maintenance visits, modifications of aircraft or aircraft system by service bulletins (SBs), service letters (SLs), and airworthiness directives (AD) or special inspections required by the airline, or the authority. “Shop maintenance” is usually done on an out-of-service aircraft which requires removal of components and equipment from the aircraft for maintenance. The removal and replacement task is usually done by line or hangar maintenance personnel. The removed unit is then sent to the appropriate shop for repair.

To ensure that work has been performed in accordance with the applicable standards of airworthiness, all work performed will be periodically reviewed by the quality assurance department, and with the aviation authority, who will also, from time to time, carry out audits.

4.3.2. Human error in aviation maintenance

While many technical defects are corrected during various maintenance stages, others can go unnoticed in the checks before the plane is handed back to the operating company. In 1979, an engine fell off an American Airlines Flight 191 (NTSB, 1994b) while it was taking off from Chicago Airport. All 271 people on board and 2 people on the ground were killed. The probable cause was that the engine pylon had been stressed due to an incorrect engine removal procedure. The United States National Transportation Safety Board (NTSB) discovered that instead of following the maintenance procedure as recommended by McDonnell Douglas to remove the engine prior to the removal of the engine pylon, American Airlines (among others) used the short cut procedure by removing the pylon assembly as a whole using a forklift truck. Though this was a cheaper and quicker method, it was extremely difficult to hold the engine assembly straight while it was being removed and this eventually caused cracking in the process.

This shows that, behind the technical aspects of maintenance (inspection intervals, maintenance methods, etc.) lie again the human factors relating to the maintenance planners, inspectors and fitters carrying out the work. Therefore, the maintenance process (see Figure 4-7) for the example of the maintenance delivery system in the ARAMIS model that did that) requires good procedures, competent and committed people available to carry out the processes described in the block diagram in collaborative teams, using good functioning and user-friendly hardware. This means that modelling needs to split maintenance down into these steps and then use the human delivery systems related to each significant step for deeper modelling.

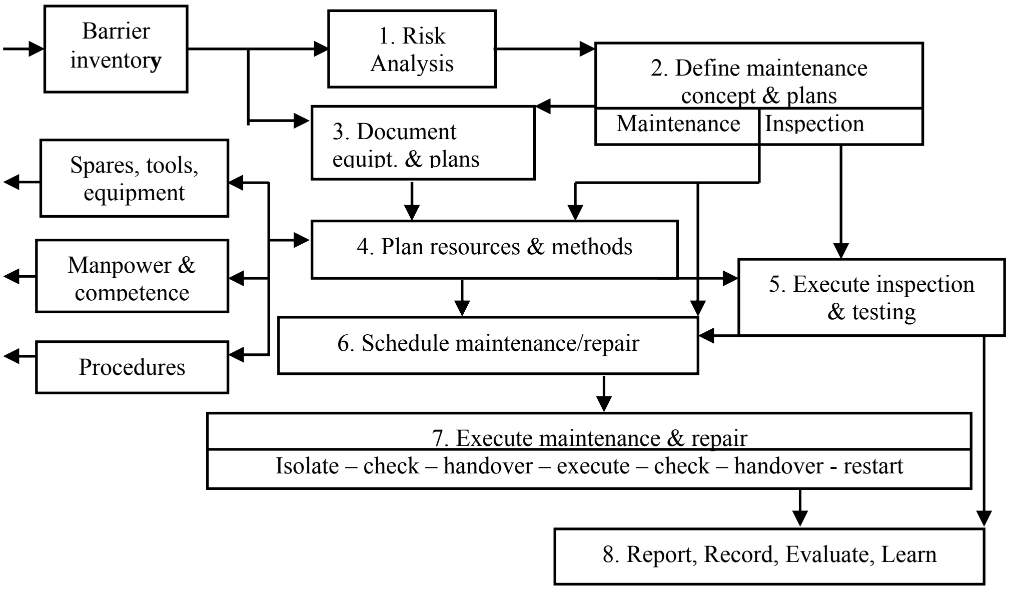
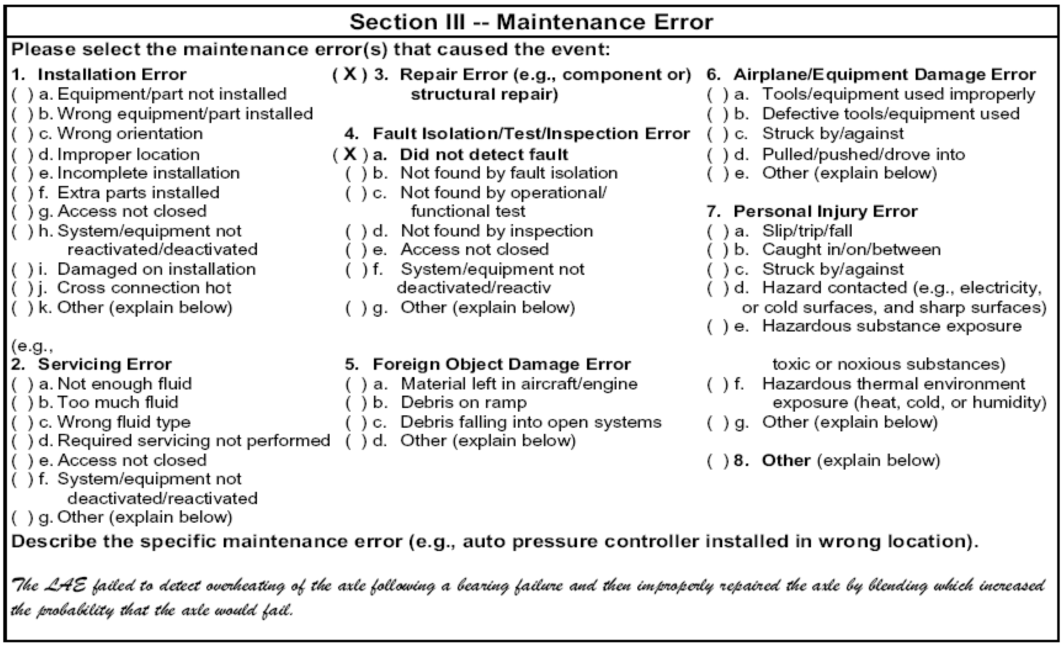

Barrier inventory
1. Risk Analysis
2. Define maintenance concept & plans
Maintenance
Inspection
3. Document equipt. & plans
Spares, tools, equipment
Manpower & competence
Procedures
4. Plan resources & methods
5. Execute inspection & testing
6. Schedule maintenance/repair
7. Execute maintenance & repair
Isolate – check – handover – execute – check – handover - restart
8. Report, Record, Evaluate, Learn

Figure 4-7 Maintenance process (ARAMIS, 2004)

There are certainly many more accidents which could illustrate problems at all of the different points in the inspection and maintenance cycle. According to Hobbs and Williamson (2003), a total of 619 safety occurrences involving aircraft maintenance were reported in their study of which 96% were related to the actions of maintenance personnel. This shows the importance of modelling it in a full aviation system model. Since the tasks related to the maintenance of the aircraft are also carried out by people, it is possible, if desired, to analyse these tasks more deeply using the framework identified in Chapter 3. Boeing, in cooperation with several airlines and the FAA, has developed an error reporting system called the Maintenance Error Decision Aid (MEDA) (Allen & Rankin, 1995). It aims to tackle the underreporting of human error in aviation maintenance and to provide a standard method for analysing errors. Figure 4-8 shows its error classification.



Section III -- Maintenance Error

Please select the maintenance error(s) that caused the event:

1. Installation Error
( ) a. Equipment/part not installed
( ) b. Wrong equipment/part installed
( ) c. Wrong orientation
( ) d. Improper location
( ) e. Incomplete installation
( ) f. Extra parts installed
( ) g. Access not closed
( ) h. System/equipment not reactivated/deactivated
( ) i. Damaged on installation
( ) j. Cross connection hot
( ) k. Other (explain below)

(e.g.,

2. Servicing Error
( ) a. Not enough fluid
( ) b. Too much fluid
( ) c. Wrong fluid type
( ) d. Required servicing not performed
( ) e. Access not closed
( ) f. System/equipment not deactivated/reactivated
( ) g. Other (explain below)

(X) **3. Repair Error (e.g., component or) structural repair)**

4. Fault Isolation/Test/Inspection Error
(X) **a. Did not detect fault**
( ) b. Not found by fault isolation
( ) c. Not found by operational/ functional test
( ) d. Not found by inspection
( ) e. Access not closed
( ) f. System/equipment not deactivated/reactiv
( ) g. Other (explain below)

5. Foreign Object Damage Error
( ) a. Material left in aircraft/engine
( ) b. Debris on ramp
( ) c. Debris falling into open systems
( ) d. Other (explain below)

6. Airplane/Equipment Damage Error
( ) a. Tools/equipment used improperly
( ) b. Defective tools/equipment used
( ) c. Struck by/against
( ) d. Pulled/pushed/drove into
( ) e. Other (explain below)

7. Personal Injury Error
( ) a. Slip/trip/fall
( ) b. Caught in/on/between
( ) c. Struck by/against
( ) d. Hazard contacted (e.g., electricity, or cold surfaces, and sharp surfaces)
( ) e. Hazardous substance exposure toxic or noxious substances)
( ) f. Hazardous thermal environment exposure (heat, cold, or humidity)
( ) g. Other (explain below)

( ) **8. Other** (explain below)

Describe the specific maintenance error (e.g., auto pressure controller installed in wrong location).

The LAE failed to detect overheating of the axle following a bearing failure and then improperly repaired the axle by blending which increased the probability that the axle would fail.

Figure 4-8 Maintenance error

Maintenance error in Figure 4-8 is broken down into several classifications: improper installation, improper servicing, improper/incomplete repair, improper fault isolation/ inspection/testing, actions causing foreign object damage, actions causing surrounding equipment damage, and actions causing person injury. Although different from the classification made by Shorrock and Kirwan (2002) (see Table 3-1), these map onto the external error modes (EEMs) of human error describing what error occurs in terms of the external and observable manifestation of the error. The first four topics also map to the execution steps in step 7 of the ARAMIS maintenance model.

Boeing's MEDA further identifies 10 categories of influencing factors that are considered useful in indicating where changes are needed to reduce human error and break down the causality of an incident. They are (Rankin et al., 2000):

- Information-written or computerized information used by maintenance technicians in their job, e.g. maintenance manuals, service bulletins, and maintenance tips,
- Equipment, tools, and parts,
- Airplane design and configuration,
- Job and task,
- Technical knowledge and skills,
- Factors effecting individual performance, e.g. physical health, fatigue, time constraints, and personal events,
- Environment and facilities,

- Organizational environment issues, e.g. quality of support from other Maintenance and Engineering organizations, company policies and processes, and work force stability,
- Leadership and supervision, e.g. planning, organizing, prioritizing, and delegating work,
- Communication, e.g. written and verbal communication between people and between organizations.

When it comes mapping those categories of influencing factors to the human factors figure (Figure 3-3) and table (Table 3-4) in Chapter 3, we do not see major differences with the human factors relating to flight crew and ATC. What is common between the human factor categories listed for MEDA and that in Chapter 3 is the technical and interpersonal skills that match the requirement performance, the physical and psychological fitness to do the work, clear and relevant guides to adequate performance, and tools and materials that are available and designed scientifically to match human factors. The table in Chapter 3 covers quite well the factors identified in maintenance errors. What is specific for maintenance, however, is the importance of “communication and coordination” between maintenance technicians, which particularly relies on written information (e.g. work cards) supported by verbal handovers, because they have to work together and coordinate activities between shifts. Maintenance operations are frequently characterized by asynchronous communications such as technical manuals, memos, Advisory Circulars, Airworthiness Directives, workcards, and other non-immediate formats. Another difference lies in leadership. Unlike the flight crew, which is generally small (2-3 people), maintenance operations are characterized by large teams working on disjointed tasks. A maintenance task may require multiple teams (line, hanger, shop) each with their own responsibilities. Therefore, supervisors or team leaders routinely serve as intermediaries among many points of the organization. The only human factor that is not incorporated in Boeing’s list but in ours is “commitment”.

Maintenance in CATS

In CATS, a maintenance technician model was linked with aircraft equipment failures, despite the fact that the maintenance technician has a much more indirect influence on the system than crew and ATCs (whose decisions and actions are directly in the causal chain). Maintenance technicians are involved in the common node activities of the root organizational factors. A maintenance technician performance model (see Figure 4-9) was built by determining relevant performance shaping factors (PSFs) using a similar approach to that for the direct human reliability models stated in Section 3.3.

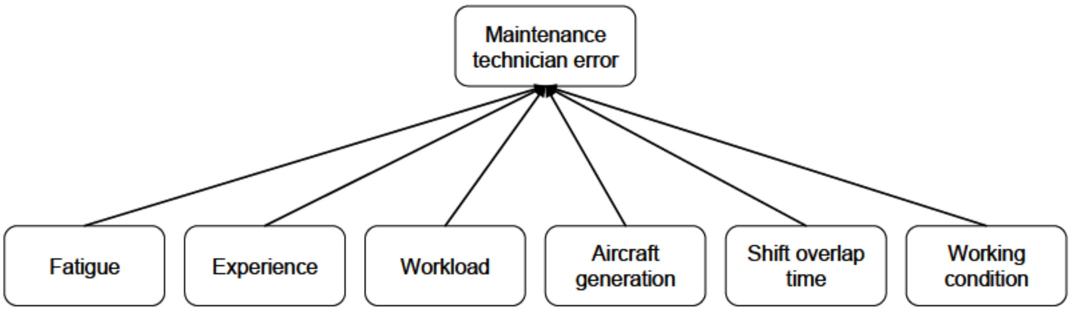


Figure 4-9 Maintenance technician performance model

However, of the three human performance models formulated in CATS (i.e. flight crew, maintenance technician, and ATC), the maintenance technician model was relatively simple, and some of the influences require further refinement. As a similar approach was taken as for the flight crew, the first column in Table 4-1 shows the influencing factors which were initially selected to be modelled in the maintenance technician model in CATS. The factors listed in the second and third columns show the final lists of selected PSFs used in the maintenance error part of the CATS model. “Complexity of the required diagnosis and response” which measures the overall complexity of the task at hand and “composition of the team” of maintenance technicians which is expected to influence team member coordination and cooperation were considered too complicated to represent quantitatively at the stage when the CATS was under development and therefore were eventually left out completely.

Moreover, it was considered impossible to assess maintenance procedures on their ambiguity, compatibility, etc. Therefore, when considering the effect of “suitability of procedures” on maintenance technician performance, it was decided to consider four different aircraft generations by assuming that aircraft of newer design are easier to maintain than aircraft of older ones. Furthermore, all of the delivery systems also need to be operationalised and link into the PSFs in Figure 4-9, in order to model more specifically where in the steps of the delivery systems the most failures occur. Hence, the maintenance model needs further extension if it is to be used for more detailed estimating and evaluating of this aspect of risk and safety.

Table 4-1 Performance shaping factors for maintenance technician in CATS

Initially selected as important aspects for maintenance technician in CATS	Node	Description
Special fitness needs	Fatigue	7 levels of fatigue, 1 means wide awake, 7 is close to sleep onset
Applicability and suitability of training / experience	Experience	Number of years working as a maintenance technician in current position
Workload, time pressure and stress	Workload	Delay in release to service of the aircraft
Suitability of procedures	Aircraft generation	Four generations of aircraft, with 4 being the most recent
Accessibility and operability of the equipment to be manipulated	Aircraft generation	Four generations of aircraft, with 4 being the most recent
Communications and whether one can be easily heard	Shift overlap time	Overlap time of two subsequent maintenance shifts
Environment in which the actions need to be performed	Working condition	Whether the work is performed at the ramp (outside) or in the hangar (inside)
Complexity of the required diagnosis and response	(left out)	
Composition of the team	(left out)	

4.4 Proposals

In this chapter, three aspects (design & manufacturing, flight crew operation and maintenance) were considered and their processes were modelled to ensure satisfactory performance of an aircraft system over its entire design life in relation to both the technical functioning and the man-machine interface. The use phase has been dealt with in CATS for the failures of the functioning of the technology, and its interaction with the human factor in the technology interface has been dealt with under the human factor in Chapter 3. It is the design and maintenance phases which need further development.

It should be noted that the design and maintenance processes (delivery systems) mentioned in this chapter are deliberately made generic and simple, in order to cover a broad perspective. It is possible to link these management processes to technical failure once that failure has been linked to the generic steps of that technology life cycle. We could then analyse the design task as set out in Figure 4-4, and the maintenance task as set out in Figure 4-7 and link these to the delivery systems providing the availability, competence, commitment, coordination and procedures relevant to each step. The latter approach is a modification of and an improvement on the previous approach, but may be inherently more time consuming and need more information for decomposition. Moreover, it makes the model more complicated as it iterates deeper into another level of human errors and organisation. The depth of the modelling is a decision needed to be determined by the importance of the influences and the user requirements in the model design stage. The availability of data is also another issue. Currently, we suggest that, only if the tasks related to the provision and maintaining of the aircraft are proven to be safety critical, should such an analysis be done.

5 Quantification methods of the SMS

One of the objectives of CATS is to incorporate into quantitative risk modelling the management factors as set out in the Dutch safety management system as influences for determining the failure rates related to human performance or technical failure at the lower level in the accident analysis. Hence, the objective of this chapter is to find a suitable technique to quantify the size of different management influences on risk. Incorporating the management factors into probabilistic risk assessment is a highly complex problem requiring careful work to resolve. Before going to the mathematical heart of quantification for management in risk modelling, we first briefly review the requirements and challenges to integrate managerial influences in risk modelling in the current existing frameworks.

5.1 Requirements and challenges to integrate managerial influences in quantified risk modelling

In other industries in the past a number of frameworks have been tried to incorporate quantification of managerial and organizational factors into probabilistic safety assessment. This problem has been tackled in MACINE (Embrey, 1992) in the field of rail transport, WPAM (Davoudian et al., 1994a, b), Omega Factor Model (Mosleh et al., 1997), Yu et al. (2004) in nuclear power plant, SAM (Paté-Cornell & Murphy, 1996) in hazardous material transport, STAMP (Leveson, 2004) in aerospace and aviation, ORIM (Øien, 2001) in offshore installation, SoTeRiA (Mohaghegh, 2007; Mohaghegh & Mosleh, 2009) in aviation maintenance, and Trucco et al. (2008) in maritime transport.

In general, integrating the quantification of managerial influences in risk modelling needs:

(1) a theoretical model of organizational performances that affect risk;
(2) a theoretical model of the link between an organization and the technical risk model
(3) a suitable technique for the quantification (see Øien (2001) and Mohaghegh (2007)) including the availability of data.

(1) We addressed the issue of management models in Chapter 2. We indicated in Chapter 1 that the management model for use in the CATS project had been made before the start of this thesis. We have examined the concurrent validity of the Dutch safety management system model in Chapter 2 of this PhD research, with respect to other existing frameworks used in aviation. The Dutch model has been found to be reasonably well supported.

(2) We addressed the link between the technical and human performance risk models and the management model in chapters 3 and 4. With respect to behavioural parameters (PSFs), the selected human factors, the way they are formulated in the risk models and the existing data available about them largely determines to what extent the management factors can be factored in. As demonstrated in Chapter 3, the PSFs considered in the current quantified risk models are mostly very partial representations of all the influences which are theoretically relevant. Given that they are the elements that management factors have to link to, we can expect that the management factors which do not fit with the strict requirement to be quantifiable from data in such cases have to be left out. This, therefore, may exclude management factors that potentially have a great influence on human performance, and creates a great challenge for management modellers to capture in quantification the overall implications of management changes to prevent accidents.

Moreover, another potential challenge to fully represent the management influences in quantification is that a management model is not designed for probabilistic risk assessment (PRA) purpose. Management consists essentially of continuing processes of measuring performance and quality on a regular basis and of adjusting an initial plan to reach the organisation's intended goal. This is a dynamic process, leading to optimisation if done well or to decay and decline if not. Traditional quantified risk analysis, on the other hand, looks for specific failure terms and causal chains and cannot cope with dynamic loops in its quantification. So, there are inherent differences in conceptualisation and modelling philosophy. Because of this inherent difference and the fact that an optimal quantification technique for safety management modelled as a control function currently does not exit to our best knowledge, it is no big surprise to see that organizational factors are often treated as simple influences on the failures (PSFs) in the causal chain and modelled in simple failure terms to fit the PRA. This results in a certain amount of reduction for a management model when linking to a risk model.

(3) Currently there are two major quantification methods to incorporate management factors into risk models. Some of the existing studies use Bayesian Belief Networks (BBNs) and some studies use System Dynamics alone or combined with BBNs. The BBN method provides a framework of the logical relationships between variables and nicely considers the uncertainty in the dependencies between variables. But BBNs do not allow feedback loops in the modelling formulation. Besides, when the data are not available, the expert judgment linked to the BBNs becomes exponentially more complex as the variables in the models increase. Some studies (Leveson, 2004; Yu et al., 2004; Mohaghegh et al., 2009) have tried to model management and organizational factors deterministically by using System Dynamics or have combined this with the probabilistic approach of BBNs to overcome the difficulties of BBNs. Each of these recent studies has tried, in some way, to decrease the deficiencies of the quantification techniques for risk frameworks. From Section 5.3 on, we will look at the two major approaches (BBNs and System Dynamics) to see what each of them contributes to the modelling and quantification of safety management model of CATS. The advantages and disadvantages of each will be discussed.

A major limitation in any quantification process, which we need to look at first, is, however, the availability of data. If data can be found which directly indicate the probability of management failures leading to technical or behavioural failures, this would greatly simplify the task. We therefore first investigate the direct data available about organizational factors. These have not yet been systematically investigated for quantification purposes. Through the discussion of what data are available, we will investigate the possibility of using them for quantification and show why direct data about organizational factors is so difficult.

5.2 How much data is available?

To quantify changes in management practice on the calculated probabilities of risk in the CATS model, data were needed about failures in the management delivery systems in relation to online failures. Four types of hard data (ADREP, LOSA, EU-OPS, IOSA) were collected to investigate whether it would be feasible to use any of these data sources to quantify the relationship between the safety management system and the online failures. In this section the experience of working with these databases will be discussed, revealing their current shortcomings. A proposal to improve them so that they would be better sources of data can be found in Section 7.4, with the other findings of this thesis.

5.2.1 ADREP

The first source was the study of the original accident and incident reports from the ICAO Accident/Incident Reporting System (ADREP) (ICAO, 2000). The Ministry of Transport, Public Works and Water Management in the Netherlands supplied the ADREP database together with the ECCAIRS[18] software which is used by EASA to store occurrence data. We analysed accident/incident data world-wide from 1996 to 2007. This period provided a dataset that is large enough for quantification and is considered representative for "current" air transport.

As mentioned in Section 3.2.2, the ADREP taxonomy has a standard structure, which has been adopted by ICAO member states throughout the world. This structure classifies human errors into two levels of failures: errors in operating the aircraft and the underlying causes which describe why a human error took place.

1) Errors in operating the aircraft: these errors are coded into 122 descriptive factors in the ADREP taxonomy, which can be mapped onto the IEM/PEM categories from the earlier sections (3.1.2):
   - Flight crew's perception/judgment (perception)
   - Flight crew's decision error (judgment, planning, and decision making)
   - Flight crew's operation of equipment error (action execution)
   - Flight crew's aircraft handling error (action execution)
   - Crew action in respect to flight crew procedures (violation)

2) Underlying causes: the underlying causes can be mapped onto the influencing factors in our model. The causes are clustered into 5 categories.
   - Human
   - Human-environment interface
   - Human-hardware/software interface
   - Human-system support interface
   - Human-human interface[19]

   These categories are detailed into four levels of sub-categories, with more than 250 explanatory factors at the greatest level of detail in the ADREP taxonomy.

Note, in ADREP, each accident contains several entries, representing the time sequence of the events in it (Table 5-1). For each event (in each row) ADREP can classify and record only one related human error type (descriptive factor in columns 2&3) and one related underlying cause (explanatory factor in columns 4&5). Therefore, one accident could have multiple human errors and underlying causes. However, ADREP does not have any representation of the levels of management (our delivery systems) that directly affect those underlying factors, except the information of who should be responsible for the underlying factors. In order to assist our management modelling, information about this organization/person (in columns 6&7) was extracted and the connections between PSFs and the delivery systems were built to identify the managerial risk control failures within airlines (see the mapping table in

[18] ECCAIRS (European Co-Ordination Centre for Aviation Incident Reporting Systems) is the software package developed by the EU that was used to access the reported accidents/incidents data. The European Aviation Safety Agency uses the ECCAIRS software to store accidents/incidents data. The ICAO ADREP 2000 taxonomy has been implemented in ECCAIRS.

[19] These five are similar to the SHEL model (Edwards, 1972).

Appendix E). When presenting these queries to the database, it was found that making detailed queries and export of data in ECCAIRS was extremely user-unfriendly (Ale et al., 2009). The basic problem is that although one can select a set of data fairly easily using a built-in query language, to export to other type of format the user has to select each attribute parameter value for export individually. These problems have been further elaborated in a report (Bellamy, 2007) which has been issued to a contact at joint research centre of European commission. Despite these problems, considerable effort was expended on assessing the potential usefulness of the data.

Table 5-1 Examples of ADREP data categories

Accident file #	Descriptive factor subject	Descriptive factor subject	Explanatory factor subject	Explanatory factor subject	Organization /person	Organization /person
1	12210300	Perception-other aircraft	201050300	Visibility from workplace	10101	Pilot
1	12210300	Perception-other aircraft	101020700	Sensory threshold	10101	Pilot
1	12210300	Perception-other aircraft	103080600	Situational awareness	10101	Pilot
2	22060000	ATM monitoring	201050300	Visibility from workplace	20402	ATCOs and FIS staff
2	26010000	Communication failures	501010500	Human interface-language	10104	Pilot of other a/c
2	22060000	ATM monitoring	201010400	Obstructions to vision	20500	Aerodrome personnel
2	22060000	ATM monitoring	201010100	Taxiway/runway	20501	Management
2	22060000	ATM monitoring	103031000	Psychological-habituation	20402	ATCOs and FIS staff

The results

18427 data points (events) from 5876 accidents were exported from the initial search query. In only 2418 events (13%) from 536 accidents (9.1%) were data recorded about both descriptive (what was the human error?) and explanatory factors (why did the human error take place?). These percentages are relatively small compared to the large panel of 5876 accidents. Out of 2418 events, 916 occurrences were related to flight crew error. We would expect an increasing attention to why a human error took place and more efforts for the investigation of management in recent years, as the subject of management has become more prominent in aviation safety. In order to find out whether through the years there had been such an increase, we counted the number of accidents which record any such factors and divided them by the total number of accidents in each year[20], to get a rate. The finding shows that underlying causes did not show an increase in recording through the years as our expectation.

The ADREP taxonomy contains many ambiguous elements due to its extensive listings without a clear error framework. To try to understand these, a great deal of effort had to be invested to decode the accident narratives which explain the causes in sufficient detail to support behavioural analysis at the level of detail and structure used in this research. The results reported here mainly focus on human errors: showing the most dominate factors at

[20] We assumed that, on average, in each of the accidents there should be at least one managerial failure causing the accident to happen.

level 1 and level 2 of the model in this thesis, and the connection we were able to make from organizational influences (level 3) all the way to accident occurrence.

- Table 5-2 shows the frequencies of flight crew errors based on our sample of accident and incident data from the ADREP database. Most of them (40%) are related to action execution.

Table 5-2 Flight crew errors-descriptive factors

Descriptive factors subject	**Frequency**	**Percentage**
Flight crew's operation of equipment error (action execution)	171	19%
Flight crew's aircraft handling error (action execution)	194	21%
Crew action in respect to flight crew procedures (violation)	261	28%
Flight crew's perception/judgment (perception)	159	17%
Flight crew's decision error (judgment, planning, and decision making)	131	14%
Total	916	100%

- Further analysis showed that there are 7 groups of PSFs that dominate, providing 80% of the underlying causes in the ADREP data:
  - psychological limitations
  - human physiology,
  - company, management, manning or regulatory issues,
  - interactions, team skills crew/team resource management training,
  - communication,
  - experience, knowledge and recency, and
  - physical environment (see Table 5.3 for the full list).

As mentioned above, the ADREP's underlying causes are detailed into four levels of sub-categories. With more than 250 explanatory factors in the ADREP taxonomy, Table 5.3 only shows the higher level of the explanatory factors in ADREP. More detailed sub-categories under each factor's subject can be found in Appendix E.

Table 5-3 Underlying causes- explanatory factors

Explanatory factor subject	Explanatory factor subject	Frequency
103000000	psychological limitations	379
102000000	human physiology	68
203000000	company, management, manning or regulatory issues	67
502000000	interactions, team skills crew/team resource management training	65
501000000	communication	61
105000000	experience, knowledge and recency	43
201000000	physical environment	43
402000000	human interface-training	39
301050000	human and hardware interface	23
503000000	supervision	20
306000000	operational material	18
401000000	human interface-standard operating procedures	18
101000000	human sensory limitation	17
204000000	operational task demands	14
304000000	automation/automatic systems	13
504000000	regulatory activities	12
104000000	personal workload management	8
305000000	automatic defenses/warnings	4
202000000	psychological factors	2
505000000	Other human-human interface	2
	Total	916

- Finding the underlying management factors that contribute to human errors is the final goal. As explained above, ADREP does not offer direct management information. Hence, each PSF in Table 5.3 was analysed to link with management failures on the basis of the mapping resulting from the table in Appendix E. Almost all of these can be mapped onto the delivery systems. But as discussed in Section 3.2.2, the PSFs classified under “psychological limitations” mix with the cognitive errors in ADREP taxonomy, so only limited factors in this category can be well mapped onto our categories.

  Figure 5-1 shows the relative contribution of delivery systems[21] to five types of flight crew error. According to this information, competence (and to a lesser extent suitability) are dominant delivery systems for most of these types of flight crew error. Besides these two, the research also shows significant differences for the most important delivery systems across error types. For violation (action execution in respect to standard procedure), communication and coordination is the most influential. Some of the accidents were caused by miscommunication between pilot and ATC so that the pilot violates the ATC instruction. For perception/judgment it is the technology interface

[21] The delivery system categorisation used here is the one which was developing through the thesis and which is described in detail in Chapter 7, which sub-divided some of the delivery systems in the original Dutch model (e.g. competence and suitability) and added systems to cover areas not dealt with earlier (e.g. workload).

design, which is supposed to provide the right information to the flight crew, but does not, and for decision errors it is communication and coordination, whilst for operation of equipment (e.g. load sheet calculations, auto land) there is an important contribution related to conflict resolution concerning pressures to achieve and supervision problems.

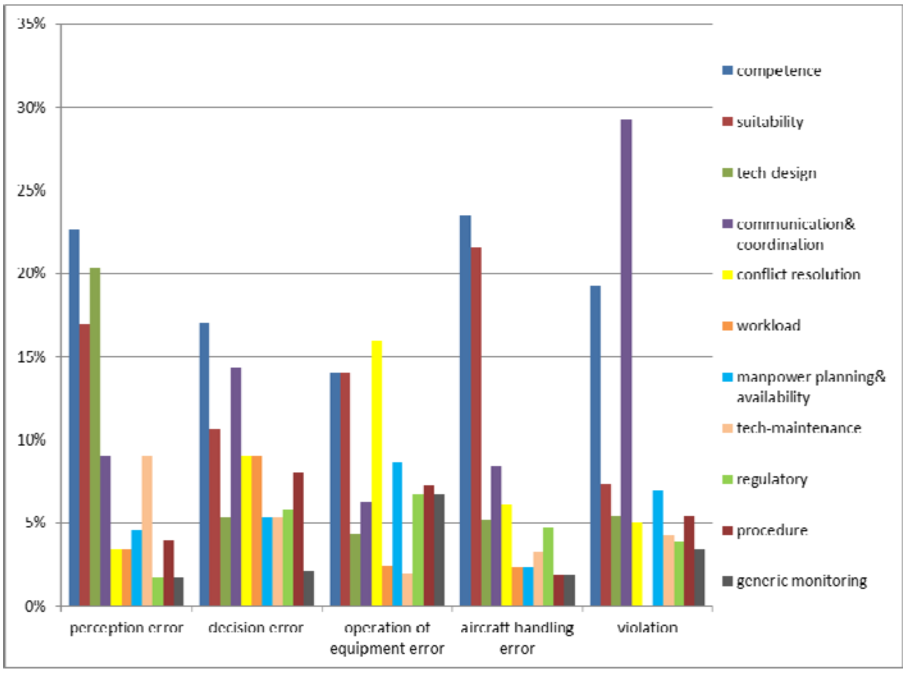


Figure 5-1 Relative contribution of delivery systems to flight crew error types

Notice that the five error types are described on a generic level rather than being very specific. The approach in this 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 and when more detail and accurate data is available. It is also important to note that all of the analysis in this sub-section was done using only the accident/incident database, without exposure data. This means that all results are only about flights that ended in an accident/incident and so got into the database. Whether a certain management deficiency is likely to be material in producing a given error type and how many of those errors the safety management system has prevented before the accident happened can only be answered if exposure data are known. Exposure data can potentially be gathered from the Line Operation Safety Audit (LOSA) database which is populated with data from the performance of pilots in normal[22] operations.

5.2.2 LOSA

LOSA is a direct observation of flight crew performance from online audits. In the accident analysis, it serves as a source of exposure data. The underlying conceptual framework has been discussed in Section 3.2.3. In this sub-section, we briefly review the framework and expand the focus of the error structure to compare with the models used here and to the other sources of potential data. This is important to integrate different sources of data for management studies. The data categories found in LOSA consist of error, threat, error management, and threat management.

[22] Normal is a relative term here, since pilots know they are being observed in LOSA flights and so may not behave typically.

- Error is defined as flight crew actions or inactions that lead to any deviation from correct performance which reduces safety margins and increases the probability of adverse operational events on the ground or during flight. To be able to compare exposure data with accident/incident data, additional work has to be done to compare the LOSA error structure with the ADREP one, to see whether the taxonomies are compatible. Studies of LOSA's error codes (see Appendix D) demonstrated that their classifications of flight crew errors focus mainly on things observable by the auditors, so their error codes are associated with the external error modes (EEM), which is different from ADREP's human error codes focusing on PEM and IEM.
- Threat is defined as an external situation increasing the operational complexity of the flight external to the influence of the flight crew. Events such as malfunctions or ATC controller errors are considered threats because they are not the result of actions by the flight crew. LOSA's threat codebook shows a relatively high count for direct observable influencing factors visible in the online audits, such as environmental factors, threats induced by other service providers, and unexpected aircraft malfunction. These factors can be well mapped to the external PSFs and environmental factors in our model but there is no information contained in ADREP database on internal threats existing or happening inside a person.

Information containing 15,941 errors from 4,306 observations collected in 27 LOSA projects from 2002 to 2007 was supplied by the LOSA Collaborative[23]. The dataset was analysed and the statistics are presented in Figure 5-2. In this figure, we can clearly see what the underlying structure is in their framework. Errors are classified into 13 types. The consequences of errors which conventionally gave the information for the right-hand side of a bow-tie model were also supplied by the LOSA Collaborative. Errors can be "detected and well managed", "detected but ignored", or "not detected" at all. In addition, any of those three states might lead to three mutually exclusive error outcomes: undesired aircraft states (UAS), additional error, and inconsequential result. The occurrence of an UAS means that the errors were not well managed and result in the position, speed, attitude, or configuration of an aircraft being in an undesired state, which clearly reduces safety margins. This in turn can lead to the error not being detected but not causing any consequence (inconsequential), or can lead to addition flight crew error (additional error). Once there is an additional error in any state, the coder has to link back to the error box and identify a new error type.

To compare the occurrence of the same deviations in accidents and incidents, it is important to be able to compare these data within a common taxonomic structure. In addition, these data should, under suitable confidentiality conditions, be available to the analysts. Neither condition is true for LOSA data at present. We are particularly interested in LOSA data about how different types of threat and error are managed by which kind of delivery systems during flight. This information is contained in the "Error Management Description" and "Threat Management Description" (see Figure 5-3 and Figure 5-4 for examples) respectively in the error and threat observations in the raw data. However, this critical information for making use of LOSA data for our management studies are recorded in narratives. Due to confidentiality issues relating to the airlines concerned, this part of the qualitative information was in the end not made available to us.

[23] www.losacollaborative.org

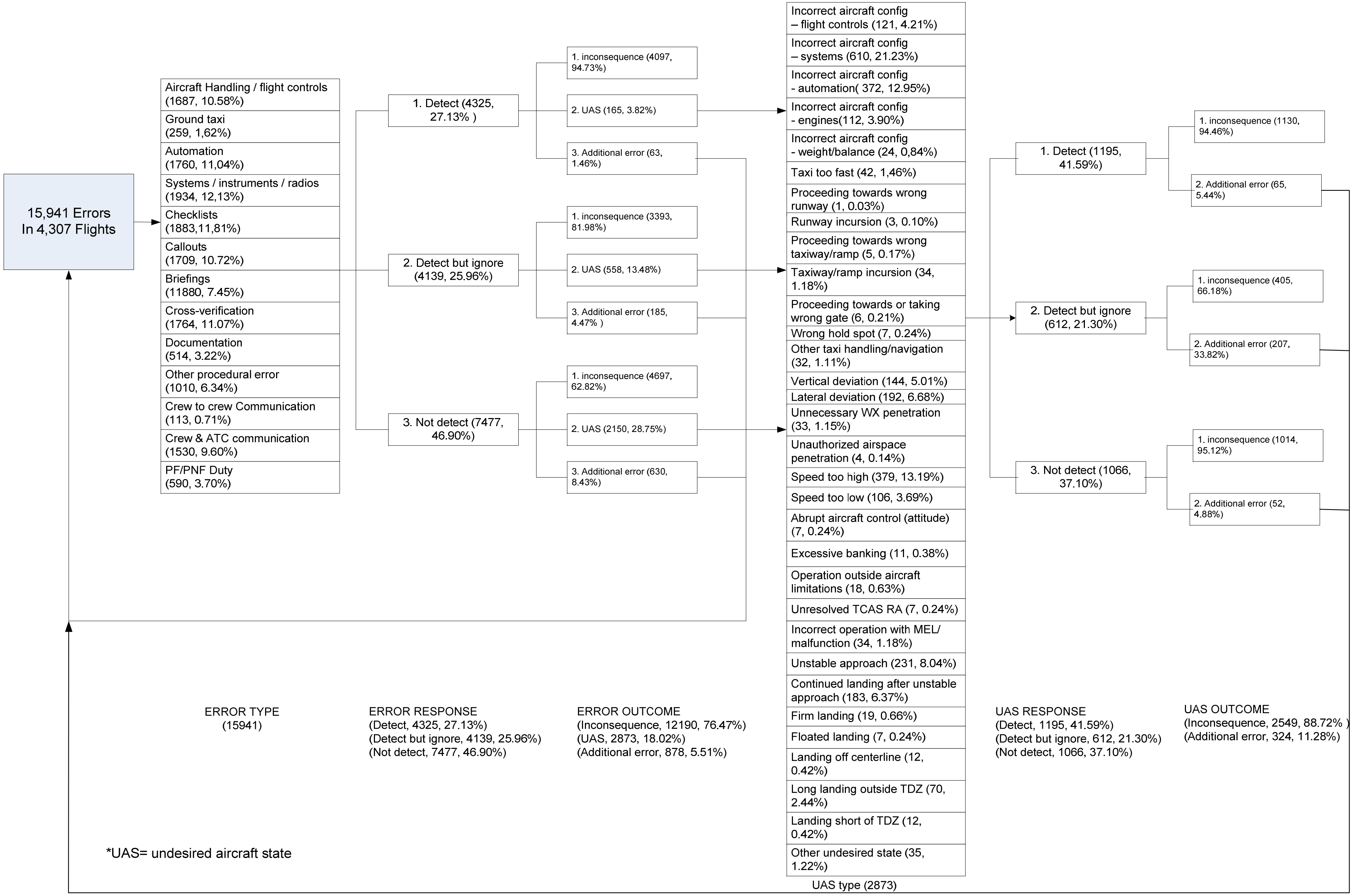
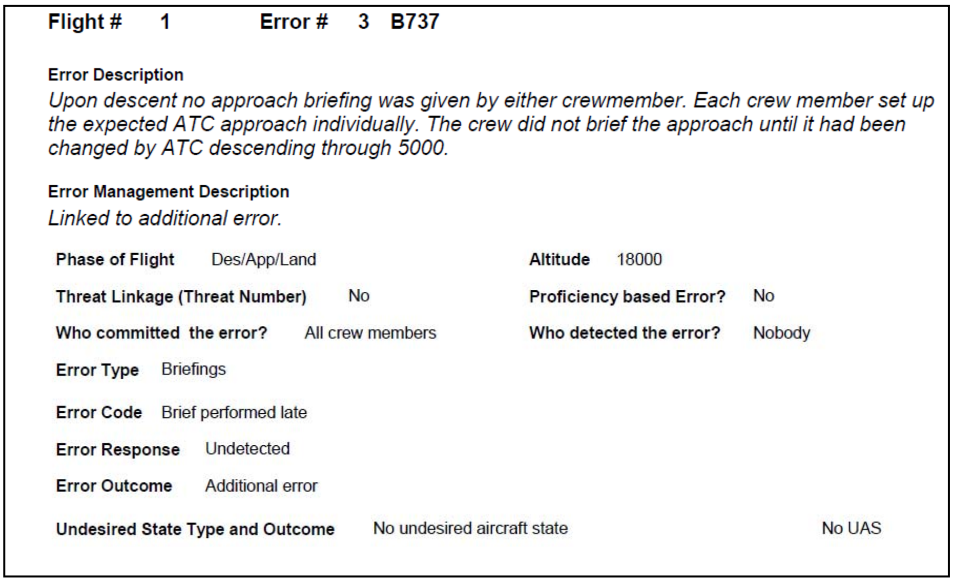

15,941 Errors
In 4,307 Flights
Aircraft Handling / flight controls (1687, 10.58%)
Ground taxi (259, 1,62%)
Automation (1760, 11,04%)
Systems / instruments / radios (1934, 12,13%)
Checklists (1883,11,81%)
Callouts (1709, 10.72%)
Briefings (11880, 7.45%)
Cross-verification (1764, 11.07%)
Documentation (514, 3.22%)
Other procedural error (1010, 6.34%)
Crew to crew Communication (113, 0.71%)
Crew & ATC communication (1530, 9.60%)
PF/PNF Duty (590, 3.70%)
1. Detect (4325, 27.13%)
2. Detect but ignore (4139, 25.96%)
3. Not detect (7477, 46.90%)
1. inconsequence (4097, 94.73%)
2. UAS (165, 3.82%)
3. Additional error (63, 1.46%)
1. inconsequence (3393, 81.98%)
2. UAS (558, 13.48%)
3. Additional error (185, 4.47%)
1. inconsequence (4697, 62.82%)
2. UAS (2150, 28.75%)
3. Additional error (630, 8.43%)
Incorrect aircraft config – flight controls (121, 4.21%)
Incorrect aircraft config – systems (610, 21.23%)
Incorrect aircraft config - automation(372, 12.95%)
Incorrect aircraft config - engines(112, 3.90%)
Incorrect aircraft config - weight/balance (24, 0,84%)
Taxi too fast (42, 1,46%)
Proceeding towards wrong runway (1, 0.03%)
Runway incursion (3, 0.10%)
Proceeding towards wrong taxiway/ramp (5, 0.17%)
Taxiway/ramp incursion (34, 1.18%)
Proceeding towards or taking wrong gate (6, 0.21%)
Wrong hold spot (7, 0.24%)
Other taxi handling/navigation (32, 1.11%)
Vertical deviation (144, 5.01%)
Lateral deviation (192, 6.68%)
Unnecessary WX penetration (33, 1.15%)
Unauthorized airspace penetration (4, 0.14%)
Speed too high (379, 13.19%)
Speed too low (106, 3.69%)
Abrupt aircraft control (attitude) (7, 0.24%)
Excessive banking (11, 0.38%)
Operation outside aircraft limitations (18, 0.63%)
Unresolved TCAS RA (7, 0.24%)
Incorrect operation with MEL/ malfunction (34, 1.18%)
Unstable approach (231, 8.04%)
Continued landing after unstable approach (183, 6.37%)
Firm landing (19, 0.66%)
Floated landing (7, 0.24%)
Landing off centerline (12, 0.42%)
Long landing outside TDZ (70, 2.44%)
Landing short of TDZ (12, 0.42%)
Other undesired state (35, 1.22%)
UAS type (2873)
1. Detect (1195, 41.59%)
2. Detect but ignore (612, 21.30%)
3. Not detect (1066, 37.10%)
1. inconsequence (1130, 94.46%)
2. Additional error (65, 5.44%)
1. inconsequence (405, 66.18%)
2. Additional error (207, 33.82%)
1. inconsequence (1014, 95.12%)
2. Additional error (52, 4.88%)
ERROR TYPE
(15941)
ERROR RESPONSE
(Detect, 4325, 27.13%)
(Detect but ignore, 4139, 25.96%)
(Not detect, 7477, 46.90%)
ERROR OUTCOME
(Inconsequence, 12190, 76.47%)
(UAS, 2873, 18.02%)
(Additional error, 878, 5.51%)
UAS RESPONSE
(Detect, 1195, 41.59%)
(Detect but ignore, 612, 21.30%)
(Not detect, 1066, 37.10%)
UAS OUTCOME
(Inconsequence, 2549, 88.72%)
(Additional error, 324, 11.28%)
*UAS= undesired aircraft state

Figure 5-2 LOSA data structure and statistical analysis



Flight # 1 **Error #** 3 **B737**

Error Description

Upon descent no approach briefing was given by either crewmember. Each crew member set up the expected ATC approach individually. The crew did not brief the approach until it had been changed by ATC descending through 5000.

Error Management Description

Linked to additional error.

Phase of Flight Des/App/Land **Altitude** 18000

Threat Linkage (Threat Number) No **Proficiency based Error?** No

Who committed the error? All crew members **Who detected the error?** Nobody

Error Type Briefings

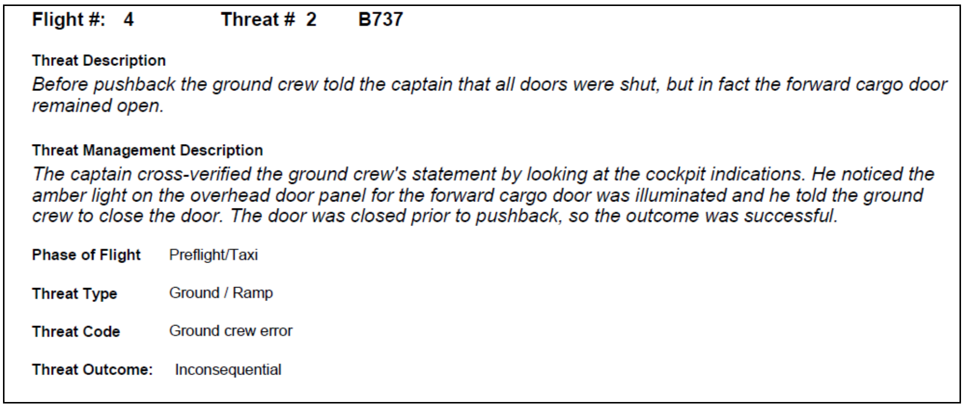
Error Code Brief performed late

Error Response Undetected

Error Outcome Additional error

Undesired State Type and Outcome No undesired aircraft state No UAS

Figure 5-3 Example of LOSA raw data for error (LOSA Collaborative, 2007)



Flight #: 4 **Threat #** 2 **B737**

Threat Description

Before pushback the ground crew told the captain that all doors were shut, but in fact the forward cargo door remained open.

Threat Management Description

The captain cross-verified the ground crew's statement by looking at the cockpit indications. He noticed the amber light on the overhead door panel for the forward cargo door was illuminated and he told the ground crew to close the door. The door was closed prior to pushback, so the outcome was successful.

Phase of Flight Preflight/Taxi

Threat Type Ground / Ramp

Threat Code Ground crew error

Threat Outcome: Inconsequential

Figure 5-4 Example of LOSA raw data for threat (LOSA Collaborative, 2007)

In summary, ADREP and LOSA construct their error frameworks in different ways. The difference is that, whereas ADREP addresses the perception, decision, action, and violation of flight crew errors, LOSA focuses on external observation of the error mode. Moreover, even though the threats in the LOSA structure parallel those external PSFs in ADREP taxonomies, we were unable to get access to this part of the LOSA data, due to the threats being mostly recorded in narratives. In our model, we need both information about threats and threat management to serve as the link between human factors and management influences.

5.2.3 EU-OPS and IOSA

So far, in the previous two sub-sections we have considered access to real management data in a way which has not been possible in the past. However, it has proven that up to now only a limited amount of relevant information was available. Besides, the data sources presented in the previous sections do not have direct information about the performance of a company's safety management system. A potential source of data directly about failures of different elements of the safety management system is the audit results from the airlines conducted with different authorities' audit tools, e.g. IATA Operational Safety Audit (IOSA) and under European Community (EU) regulations for the operation of commercial air transport (EU-OPS). In the course of the CATS project, we investigated whether it would be feasible to use the results of these audits as data for the quantification of the management influences.

5.2.3.1 EU-OPS

Any commercial passenger and cargo airline within the European Union flying jet or propeller aircraft has to comply with EU-OPS standard. The main purpose of mapping the EU-OPS audit with our SMS is (a) to find management related data, and (b) to consider how these data could be couple to the CATS model to assist the inspector in risk prioritising. In such a way, the inspector can go beyond compliance inspections and inspect critical aspects of the company safety management process in a more efficient and effective way.

The regulations in EU-OPS were analysed according to the CATS delivery system definitions and classified according to the delivery systems scheme defined in CATS. Table 5-4 gives the classification for each subpart of the EU-OPS, which was derived in an exercise together with the senior inspector from Inspectie Verkeer en Waterstaat, the regulator in the Netherlands, going through all of the classifications. As shown in Table 5-4, the regulations in subparts N and Q in the table are mostly related to the competence and availability delivery systems. The inspector was pleased that the regulations were classifiable in this way, so that he has more grip on how a long list of regulations is systematically related to the management functions in a company. Since EU-OPS was specified by the European Parliament and the European Council in 2006 and implemented in 2008, just before the end of the current phase of work on the CATS model, we were not able to get enough data from it for analysis. But it was considered quite promising if we could finish the classification for the other subparts and collect audit data on the basis of results per article of the regulation, which in turn could be classified according to the delivery system scheme of the management influences described earlier. Then if this could be compared with the accident/incident data (provided by the same airlines), this could be useful in investigating which regulation or delivery system could have priority because of its impact on the risk output. In addition, it assists with the mapping of the regulation onto CATS in order to support a more risk based approach to inspections.

5.2.3.2 IOSA

With the similar purpose of mapping, another attempt to find management related data was to map our SMS with the broader IOSA programme. The IOSA Programme is an internationally recognised and accepted evaluation system designed to assess the operational management and control systems of an airline. About 330 IATA airline members use this audit. IOSA contains the results of a number of audits per airline. These results are confidential. The basic audit cycle is 24 months, but there are provisions that permit a reduced interval under certain circumstances.

In IOSA’s framework, there are 7 operational domains:

- flight operation,
- operational control and flight dispatch,
- aircraft engineering and maintenance,
- cabin operations,
- aircraft ground handling,
- cargo operations, and
- operational security.

Table 5-4 EU-OPS classification

		Delivery systems						
		Procedures	Competence	Availability	Commitment	Communication	Technology function	Technology Man-machine interface
SUBPART N	FLIGHT CREW							
OPS 1.940(a)(1)	Composition of Flight Crew			X				
OPS 1.940(a)(2)	Composition of Flight Crew			X				
OPS 1.940(a)(3)	Composition of Flight Crew		X					
OPS 1.940(a)(4)	Composition of Flight Crew			X				
OPS 1.940(a)(5)	Composition of Flight Crew					X		
OPS 1.940(a)(6)	Composition of Flight Crew			X				
OPS 1.940(a)(7)	Composition of Flight Crew		X					
OPS 1.940b)	Composition of Flight Crew			X	X			
OPS 1.943	Initial Operator's Crew Resource Management (CRM) training		X					
OPS 1.945	Conversion Training and checking		X					
OPS 1.950	Differences Training and Familiarisation Training		X					
OPS 1.955	Nomination as commander		X					
OPS 1.960	Commanders holding a Commercial Pilot Licence		X					
OPS 1.965	Recurrent Training and Checking		X					
OPS 1.968	Pilot qualification to operate in either pilot's seat		X					
OPS 1.970	Recent experience		X					
OPS 1.975	Route and Aerodrome Competence Qualification		X					
OPS 1.978	Alternative Training and Qualification Programme		X					
OPS 1.980	Operation on more than one type or variant		X					
OPS 1.981	Operation of helicopter and aeroplane							X
OPS 1.985	Training Records		X					
SUBPART Q	FLIGHT AND DUTY TIME LIMITATIONS AND REST REQUIREMENTS							
OPS 1.1090 1.	Objective and scope			X				
OPS 1.1090 2.	Objective and scope			X				
OPS 1.1090 3.	Objective and scope			X				
OPS 1.1090 4.	Objective and scope				X			
OPS 1.1090 5.	Objective and scope			X				
OPS 1.1095	Definitions	NR	NR	NR	NR	NR	NR	NR
OPS 1.1100	Flight and duty limitations			X				
OPS 1.1105	Maximum daily flight duty period (FDP)			X				
OPS 1.1110	Rest			X				
OPS 1.1115	Extension of flight duty period due to in-flight rest			X				
OPS 1.1120	Unforeseen circumstances in actual flight operations — commander's discretion			X				
OPS 1.1125	Standby			X				
OPS 1.1130	Nutrition			X				
OPS 1.1135	Flight duty, duty and rest period records			X				

Each operation is audited with respect to "management and control", "training and quantification", "line operation", and "operation engineering requirements & specifications". All of the items in the checklists in flight operation in IOSA (IATA, 2007) were analysed according to the CATS delivery systems definitions and classified according to the delivery systems scheme defined in CATS. Table 5-5 shows the classification for the 352 items in the IOSA's flight operation list. Judging by the percentage distributions, a lot of audit topics are related to "procedure" and "competence & suitability". It is interesting to note that a different pattern has been found from accident/incident data (see the delivery systems that influence the underlying causes in the ADREP data, Figure 5-1) where "procedure" is not the most important factor in the accident database or across different error types, but rather the delivery systems of competence (and suitability), communication and coordination and conflict resolution. Our model suggests that the focus of the audits may not match with the potential risk. It would appear that the many failures of management of procedures do not apparently lead to a commensurate number of accidents or incidents. This, in turn, raises important issues of appropriateness and effectiveness of regulation as a mechanism to manage and control the risk. This preliminary finding provides a direction for future research—that is, we need to make the link and comparison between exposure data (e.g. IOSA, EU-OPS, LOSA) and accident data (ADREP) if they can be made available, so that the aspects of safety management system in the companies can be checked proactively according to risk prioritising.

Table 5-5 IOSA flight operations classification

	Competence &suitability	Availability	Communication & coordination	Procedure	Commitment	Tech- function	Tech-interface	Total
Management&control	13	2	3	16	1	4	--	39
Training&qualification	62	--	--	--	--	--	--	62
Line operation	39	6	30	119	3	10	7	214
Engineering requirements&specifications	--	--	--	9	--	28	--	37
Total	114	8	33	144	4	42	7	352
percentage	32%	2%	9%	41%	1%	12%	2%	100%

Finally, although the IOSA project could produce a lot of valuable information on management relating to a specific company, it was finally not possible to obtain the audit data from IOSA for confidentiality reasons, even after a long period of negotiation. It is quite understandable that data such as IOSA are not meant for public use. But as we have advocated through this thesis, if under specific conditions (e.g. maintaining confidentiality or use of aggregated data) the data owners could release the information for scientific research and modelling, this will help with the empirical data collection and the data analysis in the field.

The conclusions from these very time-consuming experiences of trying to work with the four types of hard data is that lack of data is a serious problem in management risk modelling, largely due to confidentiality problems, missing data, and lack of clear, consistent and recognisable causal frameworks. In Chapter 7, we will advocate changes needed for these data sources in Section 7.4.

Consequently, within this thesis we were finally unable to use objective data and had no alternative but to employ structured expert judgment in CATS. From the next section on, we step back to considering the available quantification techniques. In Section 5.3, we start by reviewing the modelling techniques of System Dynamics as it has significant advantages by incorporating increased levels of complexity through the presence of feedback loops that the BBNs approach cannot offer. Section 5.4 then turns to the BBN approach. After looking at the strengths and weaknesses of each of the two techniques, we propose a simpler addition to overcome some of their weaknesses and add to the toolbox of available methods for quantifying management influences.

5.3 System Dynamics

As mentioned above, one of the limitations in the BBNs is the inability to represent system feedback. This means BBNs describe relationships as one-way causal influences at a particular instant in time or as net influences on eventual steady-state conditions. To better representing the learning and feedback loops, Leveson (2004), Yu et al. (2004) and Mohaghegh (2007) have tried to model management and organizational factors deterministically by using System Dynamics.

System Dynamics is a methodology that "integrates knowledge (mostly descriptive) about the real world, with the concepts of how feedback structures cause all change through time". It uses the art of computer simulation for dealing with systems that are too complex for mathematical analysis. What makes using System Dynamics different from other approaches to studying complex systems is the use of feedback loops. This provides a common foundation that can be applied wherever one wants to understand and influence how things change over time. These elements help describe the nonlinear character of the feedback systems in which we live (Forrester, 1961). System Dynamics has been applied to several complex dynamic systems, including psychological, social, technological and business aspects (Sterman, 2000).

The creation of a decision support system based on a System Dynamics model consists of several steps.

1. Define the problem boundary and identify key variables.
2. Identify the most important stocks and flows. The stocks in a system tell decision makers where things are and provide them with the information about the flows needed for action.
3. Identify the feedback loops and draw causal loop diagrams that link the stocks and flows: a causal loop diagram consists of variables connected by arrows which denote the causal influences among the variables. For instance, the birth rate is determined by both the population and the fractional birth rate. Each causal link is assigned a polarity, either positive (+) or negative (-) to indicate how the dependent variable changes when the independent variable changes (see Figure 5-5).
4. Write the equations that determine the flows. Each variable has to be identified as a quantifiable property of the system that changes over time. For instance, in Figure 5-5, "birth rate" has 2 flows going into it, that is the "fractional birth rate" and the "size of the population". The equation that changes the "birth rate" can be defined as birth rate = fractional birth rate × population[24], where we have an equation which combines all the things going into it.
5. Estimate the parameters and initial conditions. These can be estimated using statistical methods, expert opinion, market research data or other relevant sources of information.
6. Simulate the model and analyze the results, often done via computer simulation

[24] The term "birth rate" refers to the number of people born per time period. "Fractional birth rate" describes birth rate per person. "Population" refers to the number of people per time period.

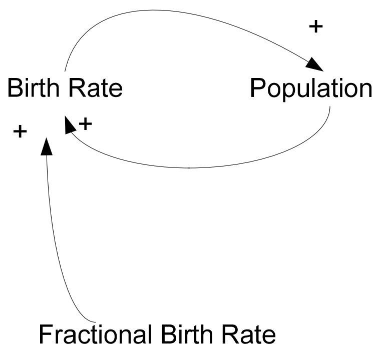


Figure 5-5 A simple example of a System Dynamics

The application of a System Dynamics methodology in safety management is quite new. Few studies have tried to model management and organizational factors deterministically using System Dynamics to better represent the learning and feedback loops.

Leveson (2004) developed a new accident model for engineering safer systems by using System Dynamics to replace the traditional chain-of-events model for both organizational and technical systems. In this study, an accident is viewed as a control problem, which results from inadequate control of external disturbances, component failures, or dysfunctional interactions among system components. The whole system is modelled as interrelated components that are kept in a state of dynamic equilibrium by feedback loops of information and control. This means a system in this conceptualization is not a static design, but a dynamic process that is continually adapting to achieve its ends and to react to changes in itself and its environment.

Yu et al. (2004) and Mohaghegh (2007, 2009) tried to use System Dynamics to model the effects of organizational factors as the deeper and more fundamental causes of accidents and incidents, and have linked these models with probabilistic risk models. Yu et al. (2004) link the System Dynamics model of the organization directly to PRA models of accidents for the assessment of the effects of organizational factors on nuclear power plant safety. They indicated that a System Dynamics approach can effectively show cause and effect relationships among factors. However, the interconnection they propose between PRA and System Dynamics is not clarified in their paper. Mohaghegh (2007, 2009)[25] has explicitly combined the probabilistic approach of BBNs and a deterministic approach using System Dynamics to represent the effects of organizational factors on airline maintenance system safety. In this study, the events, conditions, and causes of the accident scenarios are modelled through an integration of Bayesian Belief Networks (BBN), Event Sequence Diagrams (ESD), and Fault Trees (FT). System Dynamics is added to the bottom layer of the technical risk model to depict deterministic and dynamic causation mechanisms from the management and organizational models to the risk model. This approach has been applied in the aviation safety domain, focusing on airline maintenance systems. The example demonstrates that this approach can be used to analyse the dynamic effects of organizational factors on system risk.

System Dynamics, qualitatively speaking, has very good “representational features” for complex systems. The approach visually describes the relationships between the variables which makes them intuitively understandable. Also, it allows the researcher to handle increased levels of complexity in risk modelling through the presence of feedback loops. With

[25] For qualitative modelling, see Section 2.2.2 & Section 3.3.2 for more discussion.

its advantage of having such representations, a number of interlinked steps in the block diagrams of the Dutch model (which can be pictured as cycles) could be effectively coded using System Dynamics. How a company takes systematic and effective actions for one step may determine the performance of the next step it feeds into. Therefore, one could formulate those "steps" as stocks and "influences between steps" as flows. The interlinked steps must communicate effectively, and incorporate monitoring, feedback, and correction. The "learning and improvement" identified in the diagrams can also be rigorously represented as feedback loops. Using such a technique the Dutch management systems model could be effectively and systematically modelled with minimum reduction in respect to the risk model. We recommend this as a step for future study.

However, as soon as we start to derive equations for the flows which would be needed to quantify the System Dynamics models for the effects of organizational factors on risk, the technique hits problems. At the moment, the studies trying to model management and organizational factors in risk models using System Dynamics either do not mention the quantification methods associated with it or the equations have no basis in data or proof of fitness. For example, to incorporate "level of training" and "change in technician commitment" in the System Dynamics of safety, Mohaghegh (2007) defined the equations respectively as

$$\text{Level of training} = \frac{\text{experience gap} \times \text{ relative management commitment to safety}}{\text{time to provide training}}$$

$$\text{Change in technician commitment} = \frac{\text{Safety goal - Technician commitment to safety}}{\text{Time to change personal commitment to safety}}$$

Whilst the representations of the variables in the equations and the structures of the equations (e.g. multiplication or subtraction) do suggest the dynamic behaviour of the system, these two equations were rather arbitrarily assigned. They were obtained through non-structured human judgment- e.g. interviews with employees, surveys. Thus, it could be argued whether the formulations adequately present the real life meaning, let alone what dimensions are relevant for quantifying the terms of the equation or finding a common metric to make the mathematical manipulations work and where the data can be found to populate the graphs. Consequently, applying these arbitrary equations may affect the accuracy of the management model, and eventually impact the explanatory and predictive power of the resulting risk model. Hence, in this respect, System Dynamics models have received widespread criticism as "measurement without data" (Nordhaus, 1973).

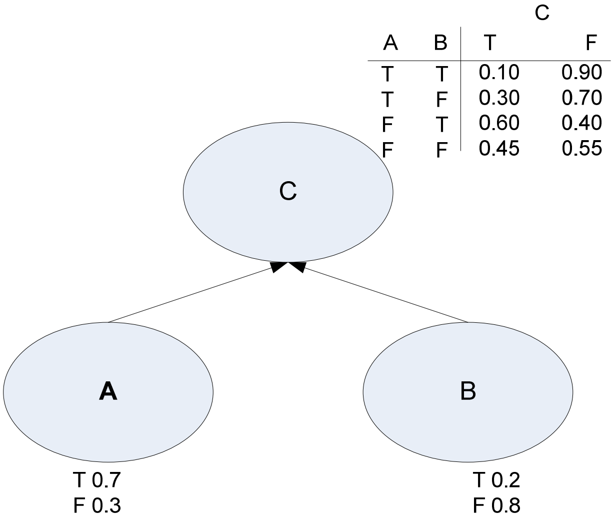
In conclusion, System Dynamics is thought to be a promising approach to capture a wide range of soft management influences and incorporate feedback loops that play an important role in maintaining or improving the functioning of the risk control measures. However, it does not currently solve the quantification problem of management. For this reason, it was decided to look back at the BBN approach to assess its further potential.

5.4 Bayesian Belief Nets (BBNs)

5.4.1 General description

Somewhat similar to the modelling language of System Dynamics, Bayesian Belief Net (BBN) is a directed graph which provides a framework of the logical relationships between variables, but it also captures the *uncertainty* in the dependencies between these variables using conditional probabilities. A BBN is a directed *acyclic* graph, meaning it contains no cycles or feedback loops. Hence the limitation of the BBN approach is the inability to represent system feedback[26]. In a BBN, the nodes of the graph represent random variables and the arcs represent conditional dependencies among the corresponding random variables. The arcs are directed (as influences) from the parent node to the child node. Nodes which are not connected represent variables which are conditionally independent of each other. BBNs are useful tools in making inferences about uncertain states when limited information is available. BBNs are frequently used for making diagnoses, with applications to medical science as well as various engineering disciplines (Jensen, 1996).

To show the main components and the steps used when building up a BBN model, an example of a belief net with three nodes is presented in Figure 5-6. Suppose that there are two events A and B which could influence event C. Suppose all three variables have two possible values, T (for true) and F (for false). The situation can be modelled with a Bayesian network. Assume that the probability tables for A and B and the conditional probability table for C are given as in the tables next to the nodes in Figure 5-6. It can be seen that, for node C, the conditional probability table lists the probabilities that this node takes each of its values, for all combinations of the parent's values (A and B). For instance, given node A is true and B is true the probability that C is true is 0.1.



		C	
A	B	T	F
T	T	0.10	0.90
T	F	0.30	0.70
F	T	0.60	0.40
F	F	0.45	0.55

Figure 5-6 Simple example of BBN

The construction of a BBN consists basically of the following steps

1. List the relevant variables by starting with the objective of the analysis and describe the factors that might influence these objectives.
2. Describe the different variables in precise terms. Each factor can be in one of a number of different states. For instance, "quality of training" might be in one of the states "good quality" or "bad quality". The states should be exclusive and exhaustive.
3. Construct the qualitative influence model using a directed acyclic graph. This considers the relations existing between the variables. But as a directed acyclic graph, no directed path through the graph can return to its own starting point.

[26] Theoretically, these types of influences can be modelled with Dynamic Bayesian Belief Nets (Ghahramani, 1997), but up to now the application of this approach is still not well-understood.

4. Quantify the network. This includes assigning conditional probabilities for each variable given each possible combination of states of the variables in the parent nodes. The conditional probabilities can either be derived from historical data or elicited from experts in the field.
5. After a Bayesian Belief Net is completed for the variables and their relationships, it can be used to answer probabilistic queries about them. This is known as "*inference*". For example, the network can be used to find out updated knowledge of the state of a subset of variables (the evidence variables) when other variables have not yet been observed.

There is often a severe limit on the availability of data to quantify the failure probabilities in step 4. Expert opinion in the form of subjective probabilities has been a dominant source for failure probabilities. Experts may have valuable knowledge about the system and parameters within the system in their specific field of interest. This knowledge is not certain. Experts may have an "idea" of the system or "true" value of certain parameters, but it is always with a certain level of subjective confidence or degree of belief. Hence, one of the questions that arise when performing expert judgement is how to elicit and utilize expert opinions in a more reliable way. Cooke (1991) and Cooke and Goossens (2000) presented a structured expert judgment procedure to combine expert's opinion in a scientifically defensible way. The overall goal of the method developed by those authors is to achieve rational consensus in the resulting assessments, so doing to enable the information of diverse experts or stakeholders to be incorporated into the process by which the results are reached and that the process itself optimises performance as measured by valid performance criteria. Structured expert judgment has been successfully applied in a series of studies in the field of the nuclear industry, chemical and gas industries, chemical toxicity, ground water pollution, volcano eruptions, dikes and dams, aerospace and aviation, and in the health, banking, and occupational sectors (Cooke and Goossens, 2000).

As stated in Chapter 3, human error probabilities in risk analysis can be quantified by BBNs. Within a BBN framework, it is possible to model the organizational and management factors that affect human error probabilities. Currently, most of the existing organizational factors frameworks (e.g. Embrey, 1992; Paté-Cornell & Murphy, 1996; Mosleh et al., 1997; Øien, 2001; Trucco et al., 2008) suggest using BBNs (or more precisely discrete BBNs) as a modelling technique to model human factors and to incorporate management factors into risk models. CATS is so far the only study to use distribution free continuous BBNs to quantify the human performance model. To find a suitable technique for the Dutch safety management model, discrete BBNs and distribution free continuous BBNs are discussed in more detail in the next section.

5.4.2 Discrete BBNs

In discrete BBNs, the nodes included in the network represent discrete variables, e.g. "yes" and "no" or "true" and "false", or "bad", "medium", and "good". As mentioned in the steps for constructing a BBN, quantification of the BBNs requires quantification of the conditional probabilities for each variable given each possible combination of states of the variables in the parent nodes. The main drawback of the discrete BBN is the excessive assessment and maintenance burden in applying it in data-sparse environments. Specifying a conditional probability table is a simple process as long as the child nodes do not have too many parents. But the numbers of probabilities that have to be assessed and maintained for a child node increases exponentially with the number of the parent nodes and with the number of states that each parent nodes can take.

For instance, Embrey (1992) proposed a model of accident causation which describes the interrelationships between management factors, immediate causes (PSFs) and operator errors in the rail transportation sector. Figure 5-7 shows a qualitative influence model taken from Embrey (1992). Management and organizational factors are expressed in terms of factors such as "task complexity" and "assignment of job role". Expert judgement is used to assess the conditional probabilities for each variable given each possible combination of states of the variables in the parent nodes, and of the unconditional probabilities of the states of each node without parents.

In Embrey's case, "time pressure" is influenced by "staffing levels", "task complexity", and "assignment of job roles", and each of the variables is discretized to have two possible values. The conditional probability table for the "time pressure" contains $2 \times 2^3 = 16$ entries which must be acquired and maintained. The conditional and unconditional probabilities used to quantify "time pressure" are shown in Appendix F. If the "time pressure" were to be rated with a five-point scale "1-5", and three more variables were found to be relevant to "time pressure" and added to the parent nodes, "time pressure" would have five possible states and six parents. If each node is then discretized to have two possible values, the conditional probability table for the child node contains $5 \times 2^6 = 320$ entries which must be acquired and maintained. This assessment burden can only be reduced by reducing the number of parent nodes and/or simplifying the discretization of the nodes. This stresses the main disadvantage of applying discrete BBN methodologies in a highly complex system.

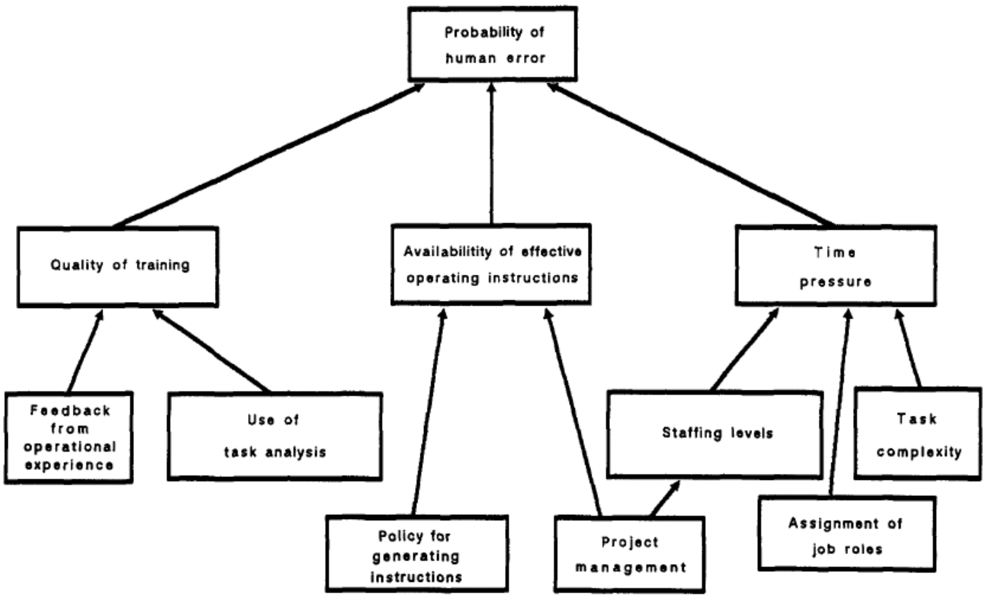

Probability of human error
Quality of training
Availabilitity of effective operating instructions
Time pressure
Feedback from operational experience
Use of task analysis
Staffing levels
Task complexity
Policy for generating instructions
Project management
Assignment of job roles

Figure 5-7 A BBN for human error probability

In addition, discrete BBNs are not very flexible when learning from new data which become available. Discrete BBNs take the unconditional probabilities from the experts only for the parent nodes, but the unconditional probabilities for the child nodes are computed from the conditional probability tables which are usually obtained from experts. For instance in Embrey's case (Figure 5-7), the unconditional probability that "time pressure" is "high" or "low" is calculated as the sum of the products of each high and low probability with the corresponding conditional (joint) probabilities. (See calculation A4 for "time pressure" in Appendix F). From the computation, the unconditional probability that "time pressure" is high or low is equal to 0.4 and 0.6, respectively. But, if the unconditional probabilities for this child node later become available from data, their values may be different from the computed ones. In quantification with expert judgment, it would be impractical to configure the elicitation of the conditional probability tables from experts such that it can comply with the unconditional probabilities of the child node obtained from data.

Moreover, discrete BBNs are not very flexible with respect to changes in modelling. In the case of management factors, which could best be seen as managing a set of actions taken by managers to deliver resources and controls to the barriers in order to reduce risk to an acceptable level, these management actions can change from time to time by learning from different sorts of information (such as accident/incident data, audit data, or adaptations to organisational changes, etc.), or deteriorate through loss of interest by senior management or disillusionment among the workforce. If we add one management factor to the parent nodes, we have to re-do all the assessment for the child of this node.

5.4.3 Distribution free continuous BBNs

This stresses some of the deficiencies of discrete BBNs. To overcome these limitations, distribution free continuous BBNs have been developed. The continuous version of BBNs solves the problem of assessing and maintaining huge numbers of probabilities. Kurowicka and Cooke (2004) introduced distribution-free continuous BBNs using vines with copula[27] that represent (conditional) independence as zero (conditional) correlation. Nodes are associated with arbitrary continuous, invertible distribution functions and arcs are expressed in terms of (conditional) rank correlations. The quantification of the continuous BBNs involves assigning a one-dimensional marginal distribution[28] to each node and a (conditional) rank correlation[29] to each arc in BBN. Figure 5-8 show a quantified version of distribution free continuous BBN applied in the professional software. The node shows the marginal distribution (as well as the name of the node and the mean and the standard deviation of the distribution); the arc represents the probabilistic dependency between the variables, which are indicated by rank correlations.

In this version of BBNs, the assessment burden is reduced to a one-dimensional distribution for each node and for each arc a conditional rank correlation. Complexity is reduced to a linear function of the number of parent nodes, rather than exponential. Using this approach, adding variables does not require re-assessing the "influence"[30] of the child given all of its parents, but only requires assessing the new "influence" of the parent node on the child node, given the already existing nodes. When deleting variables, the remaining rank correlations can be re-computed using formulas (Morales, 2010). This reduces the flexibility problems with respect to changes.

[27] Copula is a multivariate distribution that various general types of dependence can be represented.

[28] Given two variables X and Y and their joint distribution, the marginal distribution of X is the probability distribution of X averaging over the values of Y. It is typically calculated by summing (if Y is discrete) or integrating (if Y is continuous) the joint distribution over Y.

[29] Rank correlation (or Spearman's rank correlation) between two variables measures the degree of correspondence between the ranking values of the two variables. When the higher values of one variable go together with the higher values of the other variables, they are positively rank correlated. If the higher ranked values of one of the variables correspond to lower ranked values of the other variable, they are negatively rank correlated.

[30] This "influence" in distribution free continuous Bayesian Belief Nets is called the partial regression coefficient

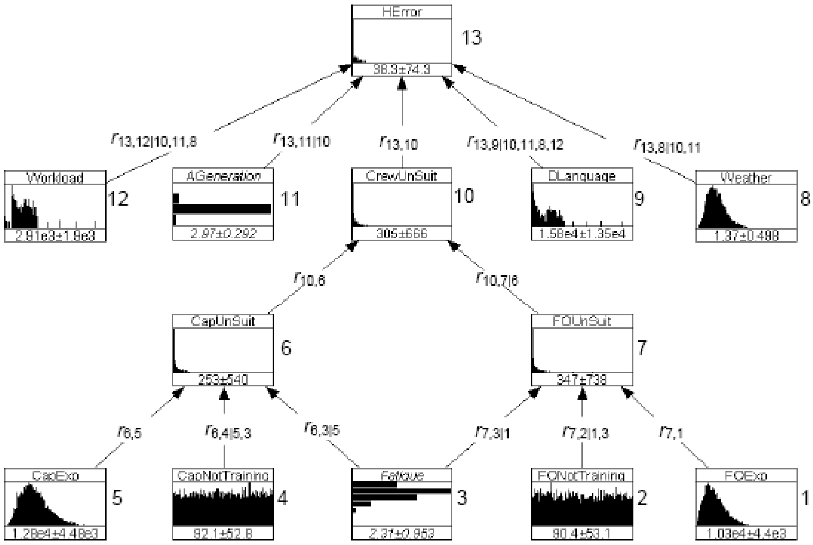


Figure 5-8 Flight Crew Performance model, quantified

As the BBN representation is very suitable for describing dependencies among components, it was decided to use distribution free continuous BBNs to mathematically represent the whole CATS model (from technical failures to human factors). The ESD's and the FT's in the CATS model were converted into a single large BBN. Using a BBN, the interdependencies inside the ESD's, FT's and BBN's and also between them can be rigorously modelled. Introducing variables with continuous distributions largely avoids the computational explosion of quantifying a highly complex problem using the discrete BBNs methodologies. This also allows a consistent handling and proper account of interdependencies and uncertainties throughout the model. This combination of system wide representation makes the CATS model a unique and potentially powerful instrument. But it came at a price in the CATS model since converting ESD's and FT's into a single BBN makes the model less transparent and the mathematical computation is highly complex. Therefore, we should be very careful only to apply the approach to those problems in which such a functionality is indeed necessary (Roelen, 2008).

No literature had previously tried to incorporate management and organizational factors into a probabilistic risk model using continuous BBN. The following section therefore presents the experience gained from one of the elicitation exercises I was involved with (i.e. flight crew performance model) within the CATS framework. The lessons learned which are relevant to this thesis are about the method of elicitation rather than the actual outcome of the expert judgement itself in terms of the derived risk figures.

5.4.3.1 Lesson learned from CATS

As has been discussed above, the process of building a BBN model includes constructing the qualitative influence model and quantifying the network.

In Section 3.3.1, we have shown the rationale behind selection of the relevant variables for the flight crew performance model. A list of nine PSFs was initially selected to be modelled in CATS. In order to make these performance shaping factors amenable for quantification, they were translated into proxy quantities. The initial nine variables had to be reduced to seven because two variables ("pilot attitude" and "procedure") were considered too complicated to represent quantitatively at the stage when the CATS project was under development. The model variables and their definitions taken from the flight crew performance model in CATS can be found in Table 3-5 in Chapter 3.

Having defined the model variables, each variable described previously is represented as a node in the model. Their (inter)relations are defined in the model structure as arcs, presented in Figure 5-9. To facilitate quantification of the model with expert judgment, it is required to keep the number of parent nodes limited to 5 or 6 maximum. Hence, “experience”, “training”, and “fatigue” for captain and for first officer were linked separately under three intermediate nodes namely “flight crew suitability”, “captain suitability” and “first officer suitability”, instead of going straight into “flight crew error”, to reduce the number of its parent nodes.

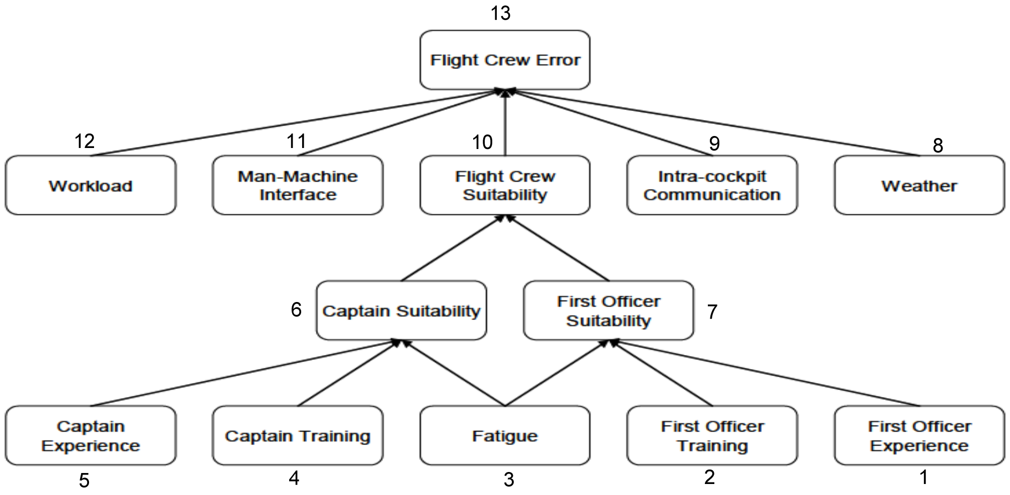


Figure 5-9 Flight crew performance model structure

The next step in the process is to quantify the network for the BBN application. The quantification of the network requires assessing a marginal distribution for each node and a conditional rank correlation for each arc. For nodes’ quantification, Table 5.6 lists the nodes, their definition, and how the marginal distribution of each node is derived. They were quantified either based on data or on expert judgment. Detailed information of the marginal distribution for each node can be found in the CATS final report (Ale et al., 2009).

For arc quantification, the rank correlations between the nodes can be computed from data or can be obtained from experts. Since no data were available in the current study, all rank correlations between nodes were retrieved by expert judgment using a conditional quintile approach, following the elicitation procedures described in Morales et al. (2008). A typical question asked to an expert to elicit the unconditional rank correlation, r_{xy}, between two variables X and Y is as follows:

> *Suppose that the variable X was observed above its q^{th} quantile. What is the probability that also variable Y will be observed above its q^{th} quantile?*

Table 5-6 Performance shaping factors for flight crew in CATS

Node #	Relevant variables	Objectively quantifiable units	Basis for quantification of the marginal distribution
1	First Officer Experience	Total number of hours flown by the First Officer	Data
2	First Officer Training	The number of days since the last type recurrent training for the First Officer	Data
3	Fatigue	Stanford Sleeping Scale	Data
4	Captain Training	The number of days since the last type recurrent training for the Captain	Data
5	Captain Experience	Total number of hours flown by the Captain	Data
6	Captain Suitability	Likelihood that the Captain fails a proficiency check	Expert judgment
7	First Officer Suitability	Likelihood that the First Officer fails a proficiency check	Expert judgment
8	Weather	Rainfall rate in mm/hr	Data
9	Intra-cockpit communication	Number of flights in which the pilot and first officer have a different mother tongue	Expert judgment
10	Crew Suitability	Likelihood the Captain or the First Officer fails a proficiency check	Based on Captain and First Officer suitability
11	Man-machine interface	Four aircraft generations; 1, 2, 3, or 4	Data
12	Workload	Number of times the crew members have to refer to the abnormal/ emergency procedures	Data
13	Flight Crew Error	Likelihood that the flight crew makes an unrecovered error that is potentially hazardous for the safety of the flight	From the associated Fault Tree

To assess the conditional rank correlation, $r_{yz|x}$, between variable Y and Z given the variable X, the expert is asked to answer question:

> *Suppose that not only variable X but also variable Z were observed above their q^{th} quantile. What is the probability that also variable Y will be observed above its q^{th} quantile?*

In the flight crew performance model in CATS, *q* is set equal to 0.5 (the median) and the conditional probability is translated to rank correlations by assuming the minimum information copula realizing the joint distribution. All additional required conditional rank correlations are obtained by asking increasing nested questions. For instance, to assess the required conditional rank correlations between variables 8, 9, 10, 11, 12, and 13 shown in Figure 5-9, the final question became (Ale et al., 2009):

> *Suppose that you select 3,200,000 flights at random. Suppose that out of the 3,200,000 you select 1,600,000 for which crew suitability is at least equal to its median value and out of those 1,600,000 you select 800,000 for which aircraft generation is also at least equal to its median value. Additionally suppose that out of these 800,000 you select 400,000 for which weather is also at least equal to its median value and out of these 400,000 you select 200,000 for which abnormal situation is also at least equal to its median value. Finally out of the last 200,000 you select 100,000 for which also mother tongue difference between pilot and first office 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 the experts to have a certain level of statistical knowledge and also to be able to hold this whole complex question, with all of its sub-clauses, in their minds. This may not be directly related to what it requires to be a field expert. During the elicitation, training sessions for the experts in the area of assigning probabilities did take place separately for each expert. One expert out of five admitted to difficulties to understand the method and to answer the complex questions.

After the elicitation, to combine the expert results for model quantification the Classical Model developed by Cooke (1991) and Cooke and Goossens (2000) for expert judgment was used to construct a weighted combination of expert probability assessments. During the elicitation, each of the experts were asked about a set of variables, called seed variables, whose true values are observable quantities and are known in advance by the analyst. This set of seed variables is used to measure and validate the expert performance in uncertainty quantifications. Derived from experts' calibration and information performance, the weight to be given to each expert in combining the expert elicitation results are computed based on the expert's answer to these seed questions. The method to measure performance of experts and combine their judgments is addressed more fully in Cooke (1991).

In the flight crew performance elicitation, 8 seed questions were asked to the experts. Below are some examples of the seed questions used in flight crew performance model.

- What is the probability per flight of a fire on board the aircraft?
- What was the probability of an unstable approach, for KLM flights only, between 1998 and 2001?
- What is the probability per flight of flight crew impairment or incapacitation?
- What was the size of the world fleet of large (100 passenger seats or more) commercial jets in 2003?

Since the weights are determined on the basis of seed variables, the choice of meaningful seed variables is critical. These seed variables must be drawn from the expert's area of expertise. To be able to observe a significant difference in calibration, preferably 8 to 10 seed variables are required for each panel. The fewer the number of seed variables, the less robust the calibration scores are likely to be.

In practice, in this case, no combination was required for the opinion since the optimized weight combination gave all the weight to one expert. This means that the rank and conditional rank correlations were taken as the opinion of one single expert. This is not an unusual situation, as it has been commonly seen from the expert elicitation done by Cooke and Goossens (2008). Giving some experts zero weight simply means that their knowledge was already contributed by other experts and adding their expertise would only add noise. Mathematically speaking it is correct to have the results dominated by one assessment, because this method indicates the best combination according to the performance criteria. But practically one may wonder whether the level of statistical knowledge of the experts had an influence on their performance. Even if the experts can answer the seed variable correctly, it is not guaranteed that they will answer the conditional rank correlation questions reliably, due to the high complexity of the questions shown above. This justifies some real concern regarding the application of this method.

Although the expert judgment procedure has been successfully applied to many studies, experience from this particular study shows a number of potential challenges relating to the application of this method to the management quantification.

First is the availability and robustness of the seed variables for safety management. Our management model depicts the management influence as the provision of resources and controls for installing and maintaining risk controls/barriers for a particular operational process. The availability of information which can be used as seed variables in this respect is severely limited. In the light of the accident/incident analysis with the Adrep data, we have shown the severe limitation of extracting sufficient data about management practices and influences (Section 5.2). Although we might generate a few seed variables from Adrep data, or based on the simulation data conducted by research institutions, for instance, from fatigue countermeasures research or workload countermeasures studies, it is very questionable whether these sub-field variables sufficiently represent safety management domain knowledge across a wide range of management issues. Besides, given the dearth of available data, to what extent can the performance on seed variables be said to robustly predict the performance on any variable of interest? The calibration scores can only be computed using the set of selected seed variables and the question which therefore remains is whether the expert scores would remain constant if different sets of seed variables were to be used. The right choice for the calibration questions is also raised by Hanea (2009) in a study modelling fire safety in buildings using BBNs. The robustness analysis on her seed variable selection concluded that the choice of calibration questions does influence the scores of the experts and leads to significant changes in the re-computed optimized weight for experts. This means selection of the performance criteria can result in a different estimation of probabilities. Therefore, using the little available data from management, we are not confident that the answers will be reliable estimates of the probabilities we want to obtain from a group of experts.

Perhaps the major disadvantage of this expert elicitation is the limitation on the number of parent nodes. As we have shown in constructing the BBNs model for flight crew, it was practically required to keep the number of parent nodes limited to 5 or 6 maximum in order for experts to mentally process the complex questions, and even this number produces highly complex combinations of sub-clauses in the elicitation question. This feature is of course due to the sophisticated consideration of the interdependencies between variables. But as we have explained in the beginning of this chapter, the goal of CATS was to try to represent as much as possible of the complex processes of safety management, as shown in the Dutch model, so that managers would recognise in them the set of decisions they have to make between actions to enhance safety, Ideally we would like to build a probabilistic model for management, and BBNs offers this possibility. However due to the large number of delivery systems and management steps within the deliveries, the complexity of the current problem exceeds the capability of the BBN method to model all influences probabilistically. Hence it is unrealistic to use only the expert elicitation that the BBN method involves with the current study.

5.5 An additional method for quantification

As always, different techniques have different strengths and weakness. The BBN approach has a systematic quantification, which sophisticatedly captures the uncertainty in the dependencies between variables; but it gets much too complex as the number of parent nodes increases and it cannot cope with feedback. On the other hand, System Dynamics incorporates a wide range of soft variables and allows feedback loops in a more complex system than BBNs are able to do; but it currently lacks a formal estimation method in its application to safety. Therefore, at present what seems to be needed is a technique which can help bridge the gap. For this reason we decided to develop a supplementary method by combining paired comparison with distribution free continuous BBNs--in order to get round the complexity of the expert elicitation that the BBN method involves when the parent nodes increase while still being able to incorporate as complete as possible of soft management influences into the risk model.

5.5.1 Paired comparisons

Paired comparisons are psychological scaling models which can be traced back to the studies leading to the Weber–Fechner law which attempts to describe the relationship between the physical magnitudes of stimuli and the perceived intensity of the stimuli. Paired comparisons were originally introduced for studying psychological responses by Thurstone (1927). Bradley (1953) developed a variant that became popular when applied to market research (e.g. consumer tests, preference and choice behaviour), social choice or public choice (e.g. voting systems). Later it was applied to assess human error probabilities (Seaver and Stillwell, 1983; Comer et al., 1984), to assess failure probabilities (Goossens et al., 1989), to assess landfill technology failures (Rodić, 2000), and to assess safety management options (Hale et al., 1999, 2000).

A simple example can explain the idea behind this method. Suppose a number of experts are available to assess the capital of 4 airlines, KLM (KL), British Airways (BA), United Airlines (UA), and Cathay Pacific (CX). We assume that each expert has some internal value for the companies' capital, but the experts are unable to verbalize how much money they have reliably. However, they are able to give their opinion on whether one company has more capital than the other.

To handle this example with the simplest model of Thurstone, we assume that the internal values over the population of the experts for each airline's capital are normally distributed, with mean μ and standard deviation σ^2. We assume that the mean value μ assessed by a group of experts for each airline is equal to the true capital the airline possesses. The distributions of the internal values for each airline are independent of each other, and all of them have standard deviation σ^2. In the paired comparison method, experts are asked to choose between alternatives pairwise according to whether they think the one airline in the pair has more capital than the other. If the expert chooses KL over BA, then the value drawn for KL is larger than the value for BA. Each expert has to judge each pair of airlines once. In this case, with 4 companies involved, there are 6 (C_2^4) comparisons to be made by each expert.

Let K, B, U, and C be independent normal variables distributed as the internal values of KL's, BA's, UA's, and CX's capital in the expert population, with means of k, b, u, and c. The distributions of the internal values of each airline are shown in Figure 5-10.

The probability that the expert favours K over B is determined by the internal value he chooses for K and B, which is determined by the relative distance of k from b. In the distribution here, the relative distance between k and b is small, thus it is quite possible that the expert may choose K over B as having the higher capital. On the other hand, since the relative distance between b and c is great, it is very unlikely that the expert will choose C over

B as greater. By using the experts' judgments to estimate these probabilities, we can estimate the relative distances between k, b, u, and c.

The experts' pairwise comparison can be examined for consistency and concordance. Measures of goodness of fit have been defined for both the normal and the exponential model. Via simulation it is also possible to generate confidence bounds. However, it should be noted that the ease with which the objects can be compared is determined by the time needed to make all of the paired comparisons, which rise factorially with the number of objects. Hence we still have a limitation of numbers of nodes in our causal trees, but the hypothesis is that this number is not so limited as with the BBNs.

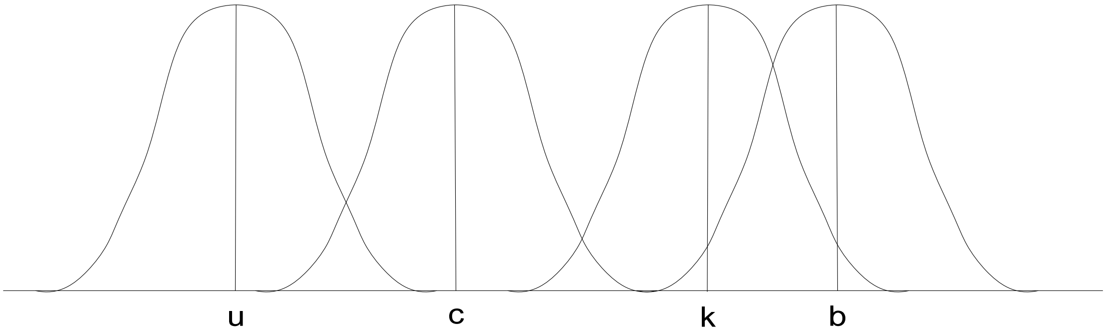


Figure 5-10 Distributions of the internal values of four airlines

The drawback of the method of paired comparisons in risk assessment is that it does not lead to true quantitative value estimations but results only in a relative scale of objects to be compared with one or more degrees of freedom. The degrees of freedom depend on the paired comparison model chosen for estimating failure probabilities. The assumption of normal distribution in the Thurstone (1927) model leads to a scale with two degree of freedom. The model of Bradley (1953) based on the exponential distribution leads to a scale with one degree of freedom. This means that, to be able to yield an assessment of actual failure probabilities, one or two empirical values are required to transfer the relative scales to their absolute values.

Hale et al. (1999, 2000) used paired comparison to assess the relative importance of management factors on risk control in the field of chemical industries, in particular related to maintenance management in major hazard chemical plants. Seven technical maintenance parameters in the risk model were selected for assessment; the 8 management factors were derived from the management model of the I-Risk project and specific influences under each of the 8 headings, which were relevant to each parameter, were derived from discussion with experts. The relative importance of these management factors related to each technical parameter was obtained from these expert judgements, making it possible to differentiate within the set of management influence weightings specific to the different parameters. But data from other sources about the absolute importance of management influences was not available in order to transfer the management weighting to error probabilities. Therefore, it was not possible to conclude the real size of the management influences on risk. Relative weightings are, however, useful for choosing between action options, even without absolute values of risk reduction.

In the next section, we propose a technique which combines the methods of paired comparisons and BBNs to overcome this difficulty.

5.5.2 Paired comparisons combined with distribution free continuous BBNs approach

As discussed in the beginning of this chapter, we have listed three general requirements to integrate managerial influences in risk modelling. Based on these three requirements: (a) a link between risk model and management model, (b) a theoretical model of organizational performances that affect risk, and (c) a suitable technique for the quantification, our quantitative methodology is decomposed into the following steps:

Step 1: Identify the target parameters in the risk model which the management model link into
Step 2: Generate management actions and use card sorting to reduce the preliminary list to a manageable set
Step 3: Design a quantification technique, including

- Assess the impact of the management actions on the parameter using paired comparisons
- Assess the states of the management actions in the situation to be modelled
- Calculate the total management influences on the parameter in the risk model

Step 4: (Re)calculate the total effect on the risk

Each step is described in more detail below.

Step 1: Identify the target parameters in the risk model which the management model link into
CATS, like most existing frameworks, used PSFs in a human model as links between the technical system and the management model. Based on the Dutch management model the aim of safety management should be to minimize pre-conditions for failure and to avoid creating contexts in which there are greater opportunities for human errors to become manifest in the workplace. This is done through the provision of resources and controls to the functioning of barriers/risk controls. Therefore the link between the management model and the human performance model could be best formulated through influencing factors.

As discussed in the beginning of this chapter, the range of management influences is determined by the PSFs given. Hence, the operationalized definitions of these influences in the BBNs should be clearly identified for safety management modellers, so that they can devise the management factors which will later to be used in the expert elicitation of the relative importance of management factors on managing the defined parameter.

Step 2: Generate management actions and use card sorting to reduce the preliminary
Given the detailed definitions of the parameters and what they include and exclude, the next step is to describe the scenarios which could lead to deviations from an optimal value for the parameter and which therefore have to be managed (e.g. poor quality of procedure due to its layout or due to its readability, clarity and depth of information). As complete a set as possible of management actions directly relevant to the parameter and to the management of those scenarios need to be generated by discussion with experts, interviews and observation, supplemented by literature. The implication of formulating management factors as a series of actions is that the management factors included in the model therefore can be seen by managers and regulators as things they can influence by taking specified actions. So they can know how their actions could influence the probabilities of the failure at the operational level and eventually influence safety. These management actions are generated under the heading of the delivery systems implemented in CATS and broken down further in the ARAMIS

block diagrams described in Chapter 2. For instance, to generate management actions for a good quality of procedure one could start from the steps of procedure delivery based on the generic process described in ARAMIS: 1) collect information over state of the art, 2) receive certificated procedures from aircraft manufacturers; 3) write company specific procedures; 4) approve procedures; 5) train the users in the procedures; 6) use the procedures; 7) monitor their use; 8) modify/maintain/enforce. Best practice for each of the steps can be collected from literature, laboratory experiments, safety audits, regulations and recommendations from international experts.

If the initial list of the potential management actions relevant to each parameter, which is collected in this way, is too long for use in the later paired comparison exercise, expert sorting can be a useful way to reduce the large set of management actions to a manageable number of important influences. The preliminary list of management actions can be presented to a group of experts in the specific field. They are first asked to add any management actions they miss in the list and group any influences whose management is inseparable. Then each of the experts is asked to do a card-sort of the actions per parameter into three groups - very important, marginally important and in between. The actions placed by all experts in the "marginally important" category can then be eliminated at least from initial modelling, so as to concentrate on the influences that all experts agree are very important. Depending on how much the list had to be reduced, the middle category can be excluded also, or included. The actions whose level of importance is not agreed between experts (e.g. some put it in marginally important and some put it in between) can remain in the list to be compared by using the paired comparison method. In this way we can fairly quickly get to a manageable number of important influences. However, but we should be very careful only to apply it to those situations in which such a reduction is really necessary, because it might delete management actions considered important by the subsequent paired comparison by the experts. Practice shows that 12 to a maximum of 15 influences can readily be compared.

Step 3.1: Design a quantification technique--Assess the impact of the management actions on the parameter using paired comparisons (the weighting process and the anchoring process)

In the CATS project, since there is not sufficient operational data to assess the weights, expert elicitations were done to assess the impact of the management actions on each of the influencing factors or PSFs in the flight crew performance model. In the following we describe the application of paired comparison in combining several expert beliefs on the relative importance of a certain parameter when a set of management actions have to be judged.

The weighting process

The Bradley-Terry (BT) model (Bradley, 1953) was used to ask experts to compare N management actions (*MAs*) pairwise and indicate their judgment about which action of the pair is more likely to improve the state of a PSF. In defining the questions, the experts were asked to judge which factors would give a relatively greater improvement in the parameter if they were to be improved, or which could give a greater degradation if neglected. Relative importance in these two directions may not be symmetrical. In the exercise for this thesis, we defined the default baseline as "no management action is implemented" and we asked the experts to judge which factors would give a relatively greater improvement in the parameter if they were to be implemented. Formulating the question in this way fits better with the logic of focusing on "risk-reducing measures" and the objective of CATS to provide guidance on which measures to take to improve safety.

Let MA_1, ..., MA_n denote the management actions which are designed to improve the state of a PSF X_i. Experts are asked a series of paired comparison questions as to which management action is more likely to improve the state of X_i. If expert *e* prefers management action MA_i to management action MA_j, (i.e. $MA_i > MA_j$), then MA_i is judged by *e* more probably than MA_j to improve the state of X_i, which we denote as $V_{ie} > V_{je}$, where V_{ie} is the internal value of expert *e* for MA_i.

When experts compare a large number of factors, especially if the actual influence of some of them is quite similar, it is not surprising that a few circular triads or intransitivities may result, for example A1>A2, A2>A3, but A3>A1. As the number of circular triads increases, we would begin to doubt whether the expert has sharply defined underlying preferences. David (1963) specifies what should be taken to be a maximum number of acceptable circular triads in an expert's preferences. Kendell (1962) has developed a statistic to be used to test the null hypothesis that an expert answered in a random fashion versus s/he has defined underlying preferences. Based on his equations, if the random preference hypothesis for any expert cannot be rejected at the 5% level of significance (p-value exceeds 0.05), the expert has to be dropped from the analysis, which means s/he does not contribute to the answers for an actual preference structure.

In addition to the above analysis, the agreement of the experts as a group can be statistically tested. One test involves the coefficient of agreement (*u*) (Kendell, 1962), another the coefficient of concordance (*W*) (Siegel, 1956). These two analyses test if all agreements of experts are due to chance (H_o), or there is a sufficient consensus among experts to be used. For both statistics, the hypothesis that all agreements are due to chance should be rejected at the 5% level of significance in order for us to have confidence in the expert estimates.

After eliminating the experts who fail in the circular triad test, the data is assessed with the significance tests using *u* and/or *W* for the set of experts. When these tests are passed, the Bradley-Terry model assumes each MA_i is associated with a true scale value V_i and that the probability r_{ij} that MA_i is preferred over MA_j can be written as

$$r_{ij} = \frac{V_i}{V_i + V_j} \tag{5.1}$$

where $\sum_{i=1}^{n} V_i = 1$. The estimates of $V_i, \ldots, V_j$ can be obtained by using the maximum likelihood (David, 1963), but only up to a scaling constant. In addition, the goodness of fit test provided in Bradley (1953) can be used to test the appropriateness of the BT model.

As mentioned, the BT model only gives us the relative importance of $V_i, \ldots, V_j$, which produces a list of "weighting factors" of MAs for improving the state of X_i. To transfer this relative importance to the answer we are seeking, "the absolute values of X_i which can be improved by MA_i", a reference value or anchoring value has to be supplied. This means if the true value r_i (that is the true value of MA_i influence on a PSF X_i) is known or can be estimated then the true value of r_j for the others MAs can be found by setting

$$r_j = \frac{r_i \times V_j}{V_i} \qquad (5.2)$$

Knowledge about the reference values or anchoring values should preferably be obtained from operational data. However, measurable operational data for any of the MA_i influences on the PSF are currently absent, so we have to resort to estimating these data also from expert judgment.

The anchoring process

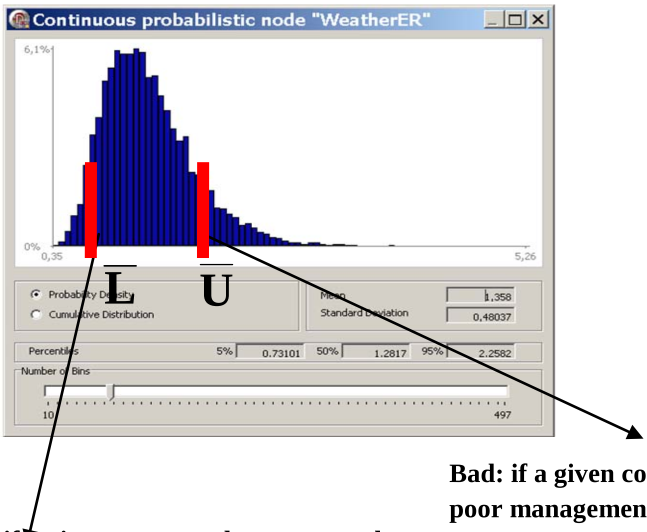
We assume that good management can support the control functions of a human or technical risk control/barrier. So, for example, good management guarantees a lower fatigue level of the flight crew by better scheduling of flights to prevent flight crew members becoming too fatigued, or by providing good sleeping facilities on stopover, etc. and bad management does the reverse. As our end goal is to estimate to what extent that good and bad safety management can improve or deteriorate the fatigue level. Where we already have the probability density function of that PSF as collected from empirical data for use in the continuous BBNs, we can use the range between two estimated values (the lower bound of the mean given that all the management actions for PSF X_i are rated as effectively managed and the upper bound of the mean given that all the management actions for PSF X_i are rated as poorly managed) on the probability density function of that PSFs as an anchor point to transfer the relative weighting scale obtained from paired comparison into a true value.

To do so, we define these two anchoring points, $\overline{U}$ and $\overline{L}$ on the probability density function of a PSF X_i, where

- $\overline{L}$ denotes the estimated value of X_i given that $\forall MA_i = 1$ from all the experts. This represents the situation that all the management actions designated as important for X_i are *implemented effectively* throughout their life cycle.
- $\overline{U}$ denotes the estimated value of X_i where $\forall MA_i = 0$ from all the experts, that is all the management control measures for X_i are *poorly managed* in the organization.

Figure 5-11 illustrates the concept of these two values in relation to the probability density functions of a PSF. After the relative importance of management actions have been compared for each parameter, each expert is asked to give his/her judgment on these two values given that all the management actions for PSF X_i are rated as either effectively managed or poorly managed[31]. The interval between these two anchoring points $\left|\overline{L} - \overline{U}\right|$ indicates the total range of the PSF which can be explained by all the defined management actions. The bigger the interval the more the PSF can be influenced by the management. The terms "implemented effectively" and "poorly managed" are subjective and will be likely to be determined by the range of experience of the experts being used, based on what they have seen (experienced) and heard about well and poorly managed aviation organisations.

[31] Whether it is lower bound or upper bound depends on the definition of the PSFs. If the PSF is formulated as a failure term in BBN where a lower value is expected after a favorable management intervention, it is the lower bound which is asked for.



● Probability Density
○ Cumulative Distribution

Mean 1,358
Standard Deviation 0,48037

Percentiles 5% 0.73101 50% 1.2817 95% 2.2582

Number of Bins 10 497

Bad: if a given company has very poor management

Good: if a given company has very good management

Figure 5-11 Expert judgments on truncation points

Using this value ($\left|\overline{L}-\overline{U}\right|$) as reference value, the true value r_i (the influence of MA_i on the PSF X_i) can be calculated by assigning the weighting values linearly between the two anchoring points as

$$r_i = \left|\overline{L}-\overline{U}\right| \times V_i \qquad (5.3)$$

If we do not have objective data on the upper and/or lower bounds, we again have to fall back on expert judgement. This is the case for many of the PSFs in CATS. We return to this in Chapter 6.

Step 3.2: Assess the states of the management actions in the situation to be modelled (the rating process)

In the approach we designed for CATS, there are two state values of a management action in the situation to be modelled. It is determined by how well it has been performed on average by an organization. Assume R_i represents the rating value over a management action, MA_i.

- R_i =1, if an organization decides to implement MA_i effectively, by providing resources and controls to it, detecting potential deviations, and improving/maintaining its functioning through its life cycle.
- R_i=0, for the rest of the states where an organization does not decide to implement it or any of the tasks has failed.

We could have rated it into several states (e.g. 5-point, where 1 represents the lowest (worst) score, and 5 represents the highest (best) score). The use of a large number of states would make the quantification process more nuanced. Although there is nothing in the methodology that prevents such an analysis, due to our focus on the weighting process and for simplicity sake, we only focused on these two states at least in the first instance.

Step 3.3: Calculate the total management influences on the parameter in the risk model

Given the true value of MA_i's effect on PSF X_i obtained from step 3.1 and the rating values obtained from step 3.2, we can estimate the new value of PSF X_i given different combinations of MAs having been applied as

$$x_i' = \begin{cases} \overline{U} - \sum_{i=1}^{n} r_i R_i, & \overline{U} \geq \overline{L} \\ \overline{L} + \sum_{i=1}^{n} r_i R_i, & \overline{U} \leq \overline{L} \end{cases}, \sum_{i=1}^{n} V_i = 1,\ R_i \in \{0,1\}, \qquad \begin{matrix} (5.4a) \\ (5.4b) \end{matrix}$$

x_i' is the new value of X_i after the management influences are applied. The application of the equations depends on the definition of the PSF $X_{i.}$. If the PSF is formulated as a failure term in BBN where a lower value is expected after a favorable management intervention (such as "fatigue" which is defined in Stanford Sleepiness Scale[32]), (5.4*a*) applies. If a higher value is expected after a favorable management intervention, such as duration of training or quality of training, (5.4*b*) will apply.

Step 4: (Re)calculate the total effect on the risk

Given the current state of a PSF X_i has been observed, by conditionalizing on its new value x_i' we could infer the probabilities of human error and recalculate the total risk due to the influence of different management actions.

5.6 Summary and conclusion

In this chapter, the analysis of the available objective data on management failures from ADREP (accidents) and LOSA, IOSA and EU-OPS (exposure) showed that these sources currently do not provide useful data for modelling purposes, either because their data models do not encompass management factors consistently, or because potentially valuable data were not made available for confidentiality reasons. A number have potential, if modified and/or if access to them could be achieved without breaching confidentiality. Confrontation between the data from the accidents and from the audit results was shown already to raise interesting questions about the focus of the audits and the relative importance of different delivery systems.

The modelling techniques of System Dynamics and BBNs as contributions to the quantification of SMS in CATS were reviewed, and the advantages and disadvantages of these were discussed in Section 5.3 and Section 5.4 respectively. As we have demonstrated in this chapter, the complexity of the current problem exceeds the capability of these two modelling techniques in quantifying the SMS in CATS. Hence, a new quantification method linked to BBNs and paired comparisons is proposed in Section 5.5 in order to take into account more of the richness of the Dutch theoretical model and overcome the complexity of the expert elicitation that the BBN method involves while linking it to the risk model. This

[32] Stanford Sleepiness Scale is a measurement of sleepiness with score from 1 to 7, where 1 signifies "feeling active, vital, alert, or wide awake" and 7 stands for "almost in reverie, sleep onset soon, losing struggle to remain awake".

paired comparison method provides relative weightings of influences, which can be converted to absolute influences if anchoring data is available or elicited from experts. The main disadvantage of the paired comparison method is that it does not take any account of dependencies between the influences. It assumes them to be independent. In the next chapter, we turn to a critical discussion of this proposed method based on two experiments. The pros and cons of applying this new method to quantify the management influences on the outcome of the human errors will be discussed further there.

6 Using paired comparisons to quantify the effect of management influences in CATS

A supplementary method was proposed in the previous chapter to help quantify the Dutch management model in CATS. In this chapter, through two experiments the feasibility of applying this proposed quantification method to quantify the management influences on the outcome of the human errors will be discussed.

The first experiment was operationalised on three of the pre-defined variables in the flight crew human performance model (HPM) within the CATS project–fatigue, weather and workload. These variables were precisely defined with objective quantifiable units. The second experiment was a trial to explore the feasibility of using the same technique on a qualitative variable, the quality/effectiveness of emergency procedures, so as to quantify the relative management influences on it in relation to risk. To test this, this "soft" variable (which was found theoretically relevant in the beginning of the project but was considered too difficult to be modelled in the flight crew HPM and so was totally left out of the main CATS project) was re-introduced into the model after the CATS project.

The quantification method developed in Section 5.5 in Chapter 5 is recapitulated here for application:

- Identify the targeted model variable(s) in the flight crew HPM which the management model links into;
- Generate relevant management actions to influence the variable(s) and use card sorting to reduce the preliminary list to a manageable size;
- Assess the impact of the management actions on the variable(s) using paired comparisons;
- Assess the states of the management actions in the situation to be modelled;
- Calculate the total management influences on the HPM's variables;
- (Re)calculate the total effect on the risk.

In Sections 6.1 and 6.2, the two experiments will be described following the order of these steps. Then, the combined results of the experiments are discussed briefly. We expect that the findings from these two exercises will give more insight into the application of this additional method which can subsequently lead to modification of the total CATS method. Recommendations for future work will be proposed in Section 6.3 for continuing exploration of this interesting field. The recommendations will be taken on board in a new model in Chapter 7.

6.1 First experiment

In the first experimental study, the proposed quantification method was applied to the flight crew HPM within the CATS project to quantify the size of the various management influences on flight crew error probability.

6.1.1 Experiment design

Identify the targeted model variables in the flight crew HPM

As explained in the beginning of Chapter 5, the operationalized definitions of the PSFs in the HPMs largely determine the range of management influences to be incorporated in the risk model. Thus, the PSFs' definitions should be clearly identified in order for the management modellers to be able to specify the management control measures to influence the variables as defined in the BBNs. In Chapter 5 Section 5.4, we have demonstrated how each PSF is quantified in the model structure of the BBN network. As the definitions of the variables are essential in this experiment and their graphical representation will be used later to demonstrate the total management influences on human errors, the definitions of the variables and the model structure are recapitulated in Table 6-1 and Figure 6-1 respectively for easy reference.

Table 6-1 Performance shaping factors for flight crew in CATS

Node #	Relevant variables	Objectively quantifiable units
1	First Officer Experience	Total number of hours flown by the First Officer
2	First Officer Training	The number of days since the last type recurrent training for the First Officer
3	Fatigue	Stanford Sleeping Scale; where 1 signifies "feeling active and vital; wide awake" and 7 stands for "almost in reverie; sleep onset soon; losing struggle to remain awake"
4	Captain Training	The number of days since the last type recurrent training for the Captain
5	Captain Experience	Total number of hours flown by the Captain
6	Captain Suitability	Likelihood that the Captain fails a proficiency check
7	First Officer Suitability	Likelihood that the First Officer fails a proficiency check
8	Weather	Rainfall rate in mm/hr
9	Intra-cockpit communication	Number of flights in which the pilot and first officer have a different mother tongue
10	Crew Suitability	Likelihood the Captain or the First Officer fails a proficiency check
11	Man-machine interface	Four aircraft generations; 1, 2, 3, or 4
12	Workload	Number of times the crew members have to refer to the abnormal/ emergency procedures
13	Flight Crew Error	Likelihood that the flight crew makes an unrecovered error that is potentially hazardous for the safety of the flight

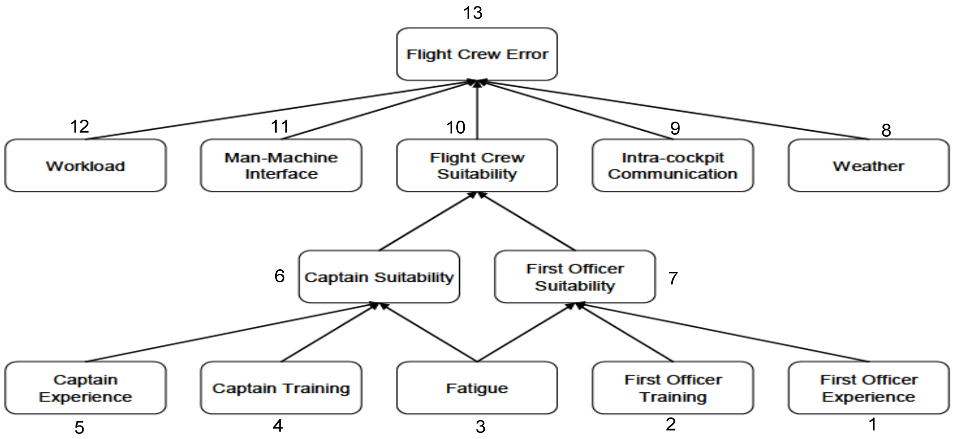


Figure 6-1 Flight crew performance model structure

Management factors have to be operationalized as actions for each of the PSFs in Table 6.1 except for the intermediate nodes (6, 7, and 10) that were mainly designed to facilitate quantification of the model. From the remaining ten PSFs, three were chosen which could be influenced by a number of management factors. These were fatigue, weather and workload.

The remaining PSFs in Table 6-1 (i.e. experience, training, intra-cockpit communication and technology interface) were defined in a way that is relatively difficult for management to influence. For instance, because "pilot training" was defined as "the number of days since last type recurrent training", management can only influence this through the planned frequency of retraining, ensuring that the pilot is freed to attend the retraining and ensuring that the flight crew is not made up of two pilots who have not had refresher training for a long time. To estimate the influences of training policy on human error in the quantified model of the BBNs, one can simply conditionalize on the days since last training of a captain (nodes 4) and/or a officer (nodes 2), and the BBNs will calculate the associated risk. In such case, we do not need expert judgement to assess the effect of different policies since we already have data on the actual distribution of days since last training and the estimates of the effect of this on error rate. However, this definition leaves out many aspects of the training which management can influences, such as the quality of the design of the training, its scope, the quality of the instructors and the simulators, etc.

Similarly, for "man-machine interface" the only management influence in CATS directly on the operationalisation of this node (aircraft generation) is the management decision about purchasing new generation aircraft and phasing out old generation ones. This influence does not require any expert judgement, since we already have estimates of how many aircraft of each generation there are, and what effect the difference in generations has on error rate. To assess the effect of changing policy on fleet composition simply means changing the relative numbers of each generation. However aircraft generation is a very limited proxy for the man-machine interface, which can be influenced by far more management processes in the design and testing process.

Also, for "pilot experience" defined as "the total number of hours flown", there is no scope for other management influences to improve pilot knowledge and skill, apart from selecting candidates with a high minimum level of experience, or avoiding both members of the cockpit crew having low experience of flying the given aircraft type. Last, for "communication and coordination between flight crew", the node only represents language compatibility between flight crew. With this "mother tongue difference", management can either influence it by preventing scheduling pilots of different mother tongue from operating together, or only hiring the pilots who speak the same language. But the other aspect of communication and coordination–e.g. interpersonal skill to work as a group, which is the subject of extensive Crew Resource Management training, is not included in this formulation.

Generate relevant management actions to influence the variable(s)

Although many of the factors in Table 6-1 were not formulated in a way that could link to the management model, we were still able to use paired comparison expert judgments to assess the relative importance of management influences for "fatigue", "weather", and "workload".

Using the block diagrams (Appendix A) of the delivery systems for the Dutch management model, management actions were generated for each of the PSFs by discussion with experts, interviews and observation, supplemented by literature review. After an internal circulation within the research team, the list was presented to one of the experts who was asked to add any management actions he missed or eliminate any if he felt were not appropriate. He did not add or delete any items to the list, but did group some action items whose management was inseparable. The management actions finally chosen for “fatigue”, “weather”, and “workload” as well as their link to the delivery system diagrams are presented in Appendix G, and are to be found in the results section of this chapter (6.1.3.1) in Tables 6-3, 6-4 and 6-5. In total, we generated

- 14 management factors to prevent fatigue (node 3 in Table 6-1);
- 13 management factors to prevent the plane encountering bad weather condition en-route and so reduce the weather risk (node 8 in Table 6-1);
- 4 factors for managing the aircraft system malfunction which might cause the crew members to have to refer to the abnormal/ emergency procedures (A/E procedure) section of the aircraft operation manual during flight and so increase their workload (node 12 in Table 6-1).

The number of influences defined is within the suggested maximum of 15, so no card sorting was necessary to reduce the lists. The management factors in the tables were put in pairs and presented to the experts in three separate questionnaires.

6.1.2 Elicitation procedure

Assess the impact of the management influences
Seven expert pilots who all had a minimum of 2,500 hours experience in flying were asked to carry out the paired comparison exercise. Each expert was given the three questionnaires to fill in and their opinions were elicited independently.

The elicitations were run to a standard format. Initially the purpose of the elicitation was introduced and the protocol was discussed step by step to allow the experts to have a chance to pose questions about definitions, steps and the list of management actions. Some experts commented on the limited definition of “weather”[33] and a majority of the experts commented on the limited definition of “workload”. They indicated that the definitions as such did not reflect reality, and pointed out that such strict formulation of the definitions made it difficult to think about the management influences in relation to the variables. For example, all of the time the experts are working on the workload questionnaire, they have to remind themselves that they are being asked to rate the influences just on the basis of the number of times they have to refer to the abnormal and emergency procedure, and not on the far broader interpretation related to operator’s capacity and task demand which we usually give to workload. However, they were told that this aspect of the task could not be changed and were asked to continue and do their best.

After that, each expert was asked to work through the 91 [(c_2^{14})] + 78 [(c_2^{13})] + 6 [(c_2^{4})] paired comparison questions for fatigue, weather, and workload, respectively. For each pair, the expert had to indicate which action would give a relatively greater improvement (less risk) in the parameter in question. After the elicitation of the relative importance, each expert was asked to estimate the anchoring points, $\overline{U}$ and $\overline{L}$, for all three variables in order for analysts to transfer the relative weighting scale obtained from paired comparisons into a true value. For the fatigue scale, each expert was given the probability density function of fatigue as collected for use in the CATS BBN from a field study of 12,965 samples (Roelen et al., 2007) (see Figure 6.2). Based on the information provided, the experts were asked to give their judgment on the value of the fatigue level on the Stanford Sleeping Scale if all of the 14

[33] In CATS, only rainfall rate was considered. But weather influences aircraft safety in a complex way, e.g. lightning, microbursts, fog, heavy rain, etc.

management actions were to be as effectively managed as possible, and the same question was also asked for all the management actions being as poorly managed as imaginable. The distributions available for weather and workload were also the probability density functions as collected for use in the CATS BBN (Roelen et al., 2007). The graphs for them are shown in Appendix H.

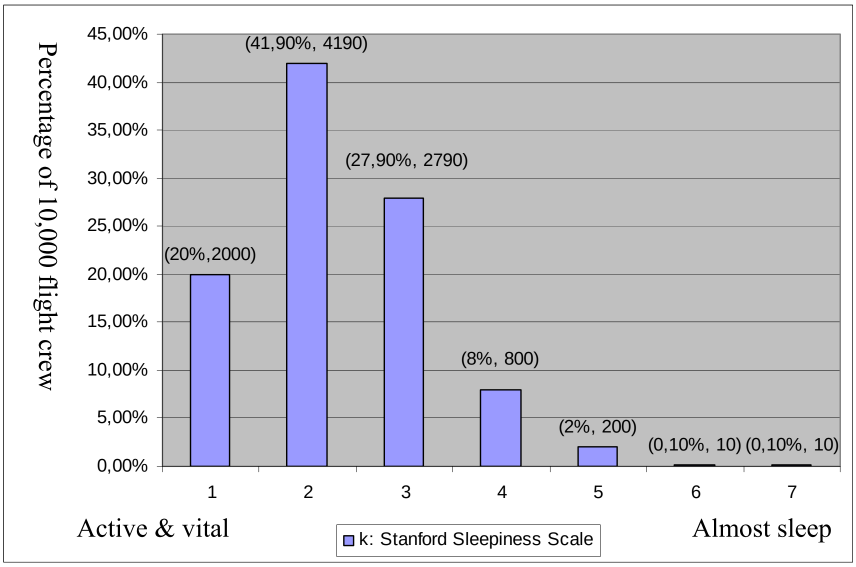


Figure 6-2 Distribution of flight crew fatigue

6.1.3 Data collection and analyses

In this subsection, the results will be reported in three parts. The first part is the rank orders and weights from the results of the paired comparisons. The second part is the results on the anchoring points. The third part is to measure the current state of management actions.

6.1.3.1 Weightings

The paired comparison data collected from the 7 experts were tested for inconsistencies by analyzing the number of circular triads. Experts with more than a threshold number of circular triads on a given variable (dependent on the number of paired comparisons to be made on the variable) were removed from the analysis. Three experts (e3, e4, e6) were removed from the "weather" section, and one expert (e4) was removed from each of the "workload" and "fatigue" sessions (see Table 6-2, the shaded boxes).

Table 6-2 Number of circular triads for fatigue, weather, and workload

Expert	Fatigue (n=14)	Weather(n=13)	Workload(n=4)
e1	32	32	0
e2	16	11	0
e3	4	34	0
e4	48	64	1
e5	6	17	0
e6	37	67	0
e7	20	3	0
Circular triad threshold	40	34	1

For each elicitation the data were fed into the Unibalance program[34]. The program assigns rank orders and weights to the management actions depending on how often they are rated as the most important in a pair. The coefficients of agreement (*u*) and the coefficient of concordance (*W*) of the remaining experts passed the statistical test at the 5% level indicating that there is a reasonable consensus on the rank orders and on the weights of each pair of actions across all the retained experts. In addition, the goodness of fit test used to test the appropriateness of the BT model could not be rejected at the 5% level of significance. Table 6-3, 6-4 and Table 6-5 give the rank orders and weights assigned to the management actions for the variables tested. They also show which of the delivery systems from the Dutch SMS model each influence falls under.

Focusing on the untransformed weighting values in Table 6-3 to Table 6-5, the results of a "group opinion" on the rank orders can already provide the airline management with a basic knowledge of the importance in managing threats in flight crew errors. With all of the lists, the airline management can identify the strengths and weaknesses in the management actions which they currently take for fatigue, weather information and workload factors, and allocate resources and controls accordingly to improve safety. This would involve auditing their current practice, especially on crew rest periods and facilities and whether the set criteria are good enough and are not being violated. We also found that some management actions which our experts said in discussion are currently not given much attention by management (such as providing a feedback system and occurrence reporting system to adapt current schedules) are considered (by the pilot experts) to be more effective and more important in influencing fatigue than the current concerns of the airline of putting in place things such as an active noise production system as alert function which allow the crews to create an alarm to detect incipient drowsiness.

In addition, the list of actions can provide an inspector or auditor with a basic knowledge of the operation of an organization (i.e., how the organization is supposed to operate in respect to managing a threat). During a field inspection, the list of management actions developed in this study (if further agreed by the inspection authorities) can provide a basic checklist of indicators which the inspector can use to evaluate the quality and efficiency of SMS performance in the airline and provide a basis for prioritization. From the ranking, it has proved that some of the important pressures can come from conflicts in the commitment to safety in the higher hierarchy; the case study in fatigue has shown that management policies should not be overridden in practice by over-scheduling tired pilots, since this is considered very important.

[34] http://risk2.ewi.tudelft.nl/oursoftware/11-unibalance

Table 6-3 Management influences on fatigue (n=14)

Item #	Management action	Weighting %	Type of influence
14	Ensure that management policy is not overridden in practice by over-scheduling tired pilots	18.1%	Commitment
4	Provide comfortable accommodation for getting good sleep at stopovers	15.8%	Man/machine interface (workplace design)
2	Set a minimum rest period after each flight and a minimum period free of all duty after a given number of consecutive days of duty	11.2%	Availability
5	Create a suitable crew rest environment and an appropriate place for a nap in multicrew aircraft	10.9%	Man/machine interface (workplace design)
1	Set maximum hours per flight duty period and cumulative duty period	9.7%	Availability
13	Require good communication between flight crew members to openly discuss fatigue and their current ability to carry on work and, if necessary, to rotate flight tasks with other crew members	6.7%	Communication
6	Provide several days off for the flight crew to adjust to a new sleep/wake schedule	5.7%	Availability
10	Provide and use good fatigue assessment tools to objectively discover pilots with relatively high fatigue and performance decrement	5.2%	Technology function& Man-machine interface
3	Set an average sleep requirement of 8 hours in a 24-hour period	4.8%	Availability
7	Provide a feedback system and occurrence reporting system, whose data are used to adapt schedules	4.4%	Availability
12	Provide equipment designs to improve work condition to reduce operator's on line fatigue and discomfort	3.2%	Man-machine interface
11	Provide a technical alert system that informs pilots if they are falling asleep during operations (e.g. active noise production)	1.9%	Technology function & Man-machine interface
8	Require crew to attend an education and training module that helps pilots to understand the cause and effect of fatigue, and teaches pilots how to minimize fatigue and its effects (e.g. NASA nap, use of bright light exposure to minimizing circadian rhythm)	1.7%	Competence
9	Check alcohol and drug consumption for a suitable period before flying	0.8%	Suitability

u=0.05, W =0.34, p=0.0383

Table 6-4 Management influences on weather (n=13)

Item #	Management action	Weighting %	Type of influence
8	Equip aircraft with an airborne weather radar system capable of detecting thunderstorms and other potentially hazardous weather conditions	23,8%	Technology-function
12	Management is committed to continuous improvement in instrumentation, information provision and (joint) training to develop collaborative solutions to weather constraint issues	13,5%	Commitment
9	Ensure flight crew, before entering the proximity of adverse weather, explicitly discuss weather conditions, instructions, alternate airports, hazards and experience	13,0%	Communication
10	Ensure Captain or FO monitors and, where necessary, challenges whether the other takes unnecessary risks in going through bad weather and take immediate action to correct deviations	11,9%	Commitment
4	Define minimum weather criteria to meet operational requirements and policies for preflight weather avoidance (e.g. alternate airport, choosing flight paths and landing routes)	11,4%	Procedure
13	Train flight crew members to enhance their decision making in adverse weather and environmental conditions	7,3%	Competence
6	Ensure flight crew, prior to each flight, complete a review of weather information (including en-route and departure, destination and alternate airports)	6,5%	Procedure
7	Ensure flight crew monitor weather information en route (ATIS, ASOS/AWOS, ATC, etc.), and, where necessary, reanalyze their flight plan	3,8%	Procedure
3	Enhance communication between pilot and dispatcher about weather conditions to maintain safe operational control	3,1%	Communication
5	Create a daily strategic plan of operations based on known or forecasted weather two to six hours in the future	2,4%	Procedure
1	Collaborate with the ATC System Command Center for constant information exchange about weather on route (pilot and ATC)	2,0%	Communication
2	Provide weather information from approved sources to the dispatcher and pilot	1,3%	Communication
11	Management rewards strict adherence to weather-related procedures and takes disciplinary action against violations	0,1%	Commitment

u=0.15, W =0.45, p=0.0005

Table 6-5 Management influences on workload (n=4)

Item #	Management action	Weighting %	Type of influence
2	Malfunction due to poor, incomplete or missed maintenance or errors in maintenance	64,1%	Tech-function
4	Malfunction due to external factors	28,6%	None
1	Malfunction due to inherent design	4,6%	Tech-function
3	Malfunction due to crew action or inaction	2,7%	Competence/ Commitment

u=0.40, W =0.59, p=0.0025

6.1.3.2 Anchoring points

After the elicitation of the relative importance, each expert was subsequently asked to estimate the anchoring points to transfer the relative weighting scale obtained from the paired comparisons into a true value. This was only really appropriate for two of the three factors. As explained, the weather definition from the BBNs was so strictly formulated that one expert refused to give his judgment of the management influences on the weather distribution of the rainfall rate that the plane encounters en route. However, the other experts were happy to take a looser concept of the weather variable and see it as the risk of encountering or coping with such a rainfall rate being affected by the influences.

There is no evidence that the expert who did not pass the consistency test in paired comparison exercise had a strong disagreement on the estimated value of the anchoring points with the rest of the experts. Treating all the experts equally, anchoring points for fatigue, weather, and workload were obtained by averaging the values estimated from all seven experts and shown in Table 6-6.

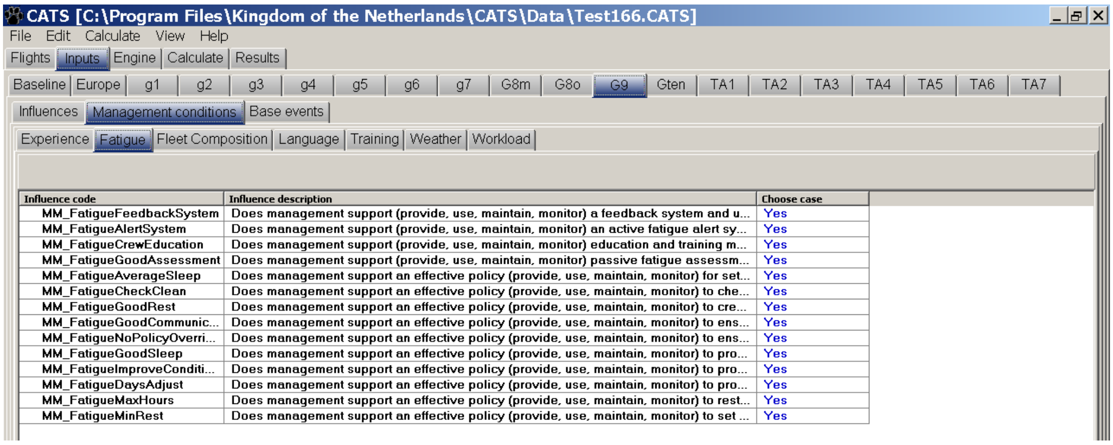
Table 6-6 Estimated anchoring points from seven experts

	$\overline{L}$ (estimated value of the variable given that all the relevant management actions are implemented effectively)	$\overline{U}$ (estimated value of the variable given that all the management actions are poorly managed)
Fatigue	1.73 (SSS)	3.2 (SSS)
Weather	1.02 (mm/hrs)	1.877 (mm/hrs)
Workload	2025 (flights /100,00 flight)	4475 (flights /100,00 flight)

6.1.3.3 The states of the management actions

To assess the states of the management actions means to measure their quality at a given point in time. This measurement will mainly be determined by expert judgments through the use of qualitative tools such as safety audit tools. In CATS, this assessment for each management action is currently scored by the end users of the CATS model (e.g. airline management, regulators) based on what they have seen (experienced) and heard described in the management systems.

In CATS, this rating process was integrated into the CATS software to support risk calculations for the end users. This means that the potential users can easily enter their current state of the management actions in the simulation, and test whether their current risk control is unsatisfactory or could be significantly improved. The user can also simulate the influences of his decisions on risk via manipulating his choices. This can provide a signal of areas where the risk could be lowered most effectively. Figure 6-3 illustrates the user interface for setting up cases. In the figure all of the management conditions are set to their optimal (rated as effective). It should be noted that the current rating scale is just in two states: yes (implemented effectively) or no (not implemented). But as noted earlier, it can potentially be rated on a scale, e.g. from 1 to 5, where 1 represents the lowest (worst) score, and 5 represents the highest (best) score. This could be done in an extension of CATS model. A large number of states will make the quantification process more comprehensive.


CATS [C:\Program Files\Kingdom of the Netherlands\CATS\Data\Test166.CATS]
File Edit Calculate View Help
Flights Inputs Engine Calculate Results
Baseline Europe g1 g2 g3 g4 g5 g6 g7 G8m G8o G9 Gten TA1 TA2 TA3 TA4 TA5 TA6 TA7
Influences Management conditions Base events
Experience Fatigue Fleet Composition Language Training Weather Workload

Influence code	Influence description	Choose case
MM_FatigueFeedbackSystem	Does management support (provide, use, maintain, monitor) a feedback system and u...	Yes
MM_FatigueAlertSystem	Does management support (provide, use, maintain, monitor) an active fatigue alert sy...	Yes
MM_FatigueCrewEducation	Does management support (provide, use, maintain, monitor) education and training m...	Yes
MM_FatigueGoodAssessment	Does management support (provide, use, maintain, monitor) passive fatigue assessm...	Yes
MM_FatigueAverageSleep	Does management support an effective policy (provide, use, maintain, monitor) for set...	Yes
MM_FatigueCheckClean	Does management support an effective policy (provide, use, maintain, monitor) to che...	Yes
MM_FatigueGoodRest	Does management support an effective policy (provide, use, maintain, monitor) to cre...	Yes
MM_FatigueGoodCommunic...	Does management support an effective policy (provide, use, maintain, monitor) to ens...	Yes
MM_FatigueNoPolicyOverri...	Does management support an effective policy (provide, use, maintain, monitor) to ens...	Yes
MM_FatigueGoodSleep	Does management support an effective policy (provide, use, maintain, monitor) to pro...	Yes
MM_FatigueImproveConditi...	Does management support an effective policy (provide, use, maintain, monitor) to pro...	Yes
MM_FatigueDaysAdjust	Does management support an effective policy (provide, use, maintain, monitor) to pro...	Yes
MM_FatigueMaxHours	Does management support an effective policy (provide, use, maintain, monitor) to rest...	Yes
MM_FatigueMinRest	Does management support an effective policy (provide, use, maintain, monitor) to set ...	Yes

Figure 6-3 User interface of CATS software

6.1.4 Management effect on the risk

Calculate the new values of PSFs given different management actions have been applied

Assuming that one has audited the present states of the management actions, one can estimate the values of the corresponding PSFs with the aid of equation (5.3) and (5.4 a):

$$Fatigue'(SSS) = 3.20 - [(3.20 - 1.73) \times [0.181 \times R_1 + 0.158 \times R_2 + ... + 0.008 \times R_{14}] \qquad (6.1)$$

$$Weather'(\text{mm/hrs}) = 1.877 - [(1.877 - 1.02) \times [0.238 \times R_1 + 0.135 \times R_2 + ... + 0.001 \times R_{13}] \qquad (6.2)$$

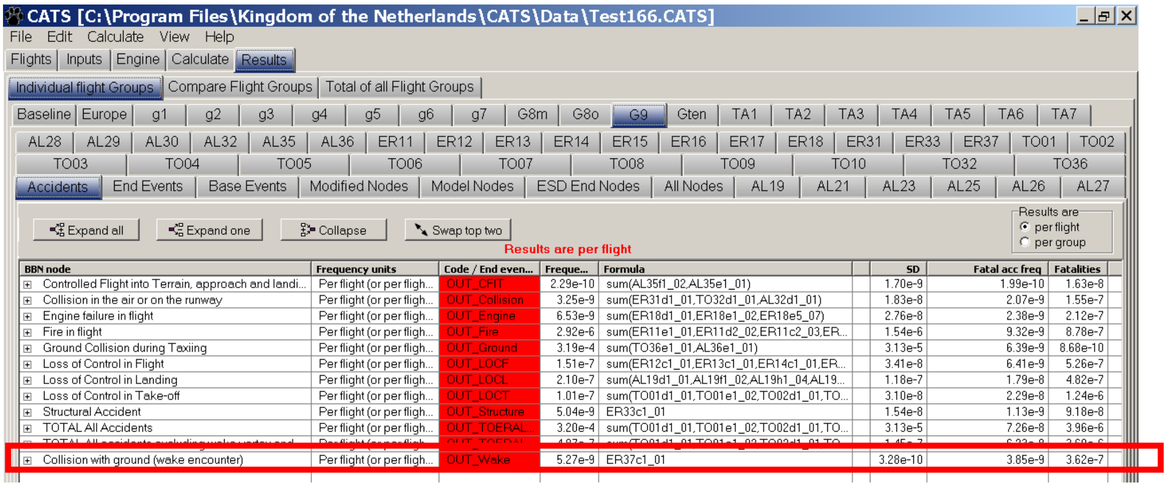
$$Workload'(\text{flights /100,00 flight}) = 4475 - [(4475 - 2025) \times [0.641 \times R_1 + 0.286 \times R_2 + ... + 0.027 \times R_4]$$

(6.3)

R_i represents the rating value over a management action MA_i. If all the management actions are implemented effectively throughout their life cycle, the ratings for the airlines are the best that they can be, i.e. $\forall R_i = 1$. Then, the value of the PSF goes to its estimated minimum (i.e. $\overline{L}$ in Figure 5-11). On the other hand, if the ratings are the worst that they can be, i.e. $\forall R_i = 0$ the quality of the corresponding PSF deteriorates to its estimated maximum (i.e. $\overline{U}$ in Figure 5-11). Other scenarios will fall between these two values.

(Re)calculate the total effect on the risk

Given the current value of the PSFs, the management effect on human error probability and risk is calculated using the BBN algorithm. Figure 6-4 and Figure 6-5 show the numerical calculation of total accident frequency with different management influences. With all the management items implemented effectively throughout their life cycle, the total accident frequency is equal to 3.20 e-4. Whilst, if the management quality is in its worst state, the accident rate increases to 1.51e-2, which is two orders of magnitude difference between good management and bad management.

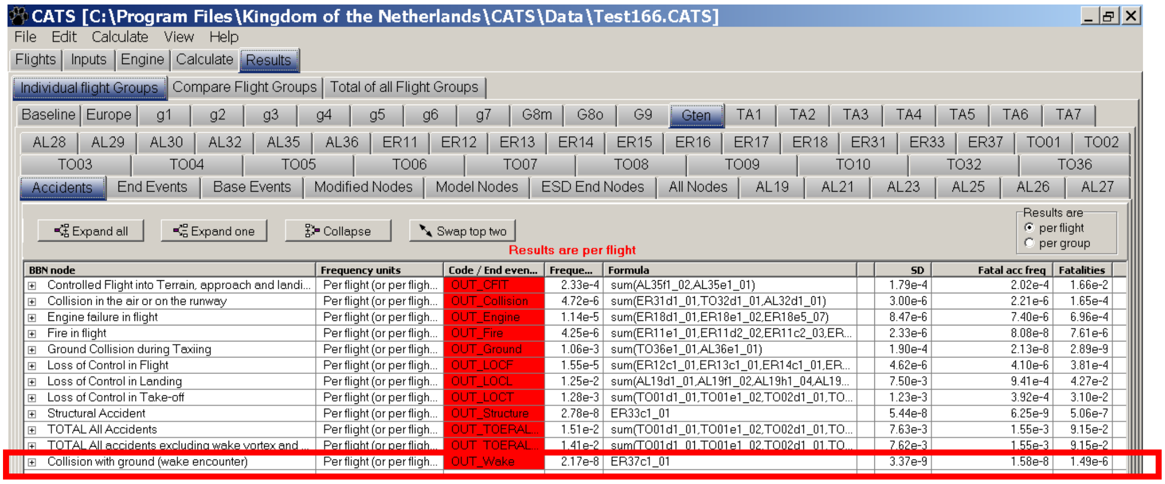

CATS [C:\Program Files\Kingdom of the Netherlands\CATS\Data\Test166.CATS]
File Edit Calculate View Help
Flights Inputs Engine Calculate Results
Individual flight Groups Compare Flight Groups Total of all Flight Groups
Baseline Europe g1 g2 g3 g4 g5 g6 g7 G8m G8o G9 Gten TA1 TA2 TA3 TA4 TA5 TA6 TA7
AL28 AL29 AL30 AL32 AL35 AL36 ER11 ER12 ER13 ER14 ER15 ER16 ER17 ER18 ER31 ER33 ER37 TO01 TO02
TO03 TO04 TO05 TO06 TO07 TO08 TO09 TO10 TO32 TO36
Accidents End Events Base Events Modified Nodes Model Nodes ESD End Nodes All Nodes AL19 AL21 AL23 AL25 AL26 AL27
Expand all
Expand one
Collapse
Swap top two

Results are per flight

Results are
● per flight
○ per group

BBN node	Frequency units	Code / End even...	Freque...	Formula		SD	Fatal acc freq	Fatalities
⊞ Controlled Flight into Terrain, approach and landi...	Per flight (or per fligh...	OUT_CFIT	2.29e-10	sum(AL35f1_02,AL35e1_01)		1.70e-9	1.99e-10	1.63e-8
⊞ Collision in the air or on the runway	Per flight (or per fligh...	OUT_Collision	3.25e-9	sum(ER31d1_01,TO32d1_01,AL32d1_01)		1.83e-8	2.07e-9	1.55e-7
⊞ Engine failure in flight	Per flight (or per fligh...	OUT_Engine	6.53e-9	sum(ER18d1_01,ER18e1_02,ER18e5_07)		2.76e-8	2.38e-9	2.12e-7
⊞ Fire in flight	Per flight (or per fligh...	OUT_Fire	2.92e-6	sum(ER11e1_01,ER11d2_02,ER11c2_03,ER...		1.54e-6	9.32e-9	8.78e-7
⊞ Ground Collision during Taxiing	Per flight (or per fligh...	OUT_Ground	3.19e-4	sum(TO36e1_01,AL36e1_01)		3.13e-5	6.39e-9	8.68e-10
⊞ Loss of Control in Flight	Per flight (or per fligh...	OUT_LOCF	1.51e-7	sum(ER12c1_01,ER13c1_01,ER14c1_01,ER...		3.41e-8	6.41e-9	5.26e-7
⊞ Loss of Control in Landing	Per flight (or per fligh...	OUT_LOCL	2.10e-7	sum(AL19d1_01,AL19f1_02,AL19h1_04,AL19...		1.18e-7	1.79e-8	4.82e-7
⊞ Loss of Control in Take-off	Per flight (or per fligh...	OUT_LOCT	1.01e-7	sum(TO01d1_01,TO01e1_02,TO02d1_01,TO...		3.10e-8	2.29e-8	1.24e-6
⊞ Structural Accident	Per flight (or per fligh...	OUT_Structure	5.04e-9	ER33c1_01		1.54e-8	1.13e-9	9.18e-8
⊞ TOTAL All Accidents	Per flight (or per fligh...	OUT_TOERAL...	3.20e-4	sum(TO01d1_01,TO01e1_02,TO02d1_01,TO...		3.13e-5	7.26e-8	3.96e-6
⊞ TOTAL All accidents excluding wake vortex and ...	Per flight (or per fligh...	OUT_TOERAL...	4.87e-7	sum(TO01d1_01,TO01e1_02,TO02d1_01,TO...		1.45e-7	6.23e-9	3.69e-6
⊞ Collision with ground (wake encounter)	Per flight (or per fligh...	OUT_Wake	5.27e-9	ER37c1_01		3.28e-10	3.85e-9	3.62e-7

Figure 6-4 Good management influences


CATS [C:\Program Files\Kingdom of the Netherlands\CATS\Data\Test166.CATS]
File Edit Calculate View Help
Flights Inputs Engine Calculate Results
Individual flight Groups Compare Flight Groups Total of all Flight Groups
Baseline Europe g1 g2 g3 g4 g5 g6 g7 G8m G8o G9 Gten TA1 TA2 TA3 TA4 TA5 TA6 TA7
AL28 AL29 AL30 AL32 AL35 AL36 ER11 ER12 ER13 ER14 ER15 ER16 ER17 ER18 ER31 ER33 ER37 TO01 TO02
TO03 TO04 TO05 TO06 TO07 TO08 TO09 TO10 TO32 TO36
Accidents End Events Base Events Modified Nodes Model Nodes ESD End Nodes All Nodes AL19 AL21 AL23 AL25 AL26 AL27
Expand all
Expand one
Collapse
Swap top two

Results are per flight

Results are
● per flight
○ per group

BBN node	Frequency units	Code / End even...	Freque...	Formula		SD	Fatal acc freq	Fatalities
⊞ Controlled Flight into Terrain, approach and landi...	Per flight (or per fligh...	OUT_CFIT	2.33e-4	sum(AL35f1_02,AL35e1_01)		1.79e-4	2.02e-4	1.66e-2
⊞ Collision in the air or on the runway	Per flight (or per fligh...	OUT_Collision	4.72e-6	sum(ER31d1_01,TO32d1_01,AL32d1_01)		3.00e-6	2.21e-6	1.65e-4
⊞ Engine failure in flight	Per flight (or per fligh...	OUT_Engine	1.14e-5	sum(ER18d1_01,ER18e1_02,ER18e5_07)		8.47e-6	7.40e-6	6.96e-4
⊞ Fire in flight	Per flight (or per fligh...	OUT_Fire	4.25e-6	sum(ER11e1_01,ER11d2_02,ER11c2_03,ER...		2.33e-6	8.08e-8	7.61e-6
⊞ Ground Collision during Taxiing	Per flight (or per fligh...	OUT_Ground	1.06e-3	sum(TO36e1_01,AL36e1_01)		1.90e-4	2.13e-8	2.89e-9
⊞ Loss of Control in Flight	Per flight (or per fligh...	OUT_LOCF	1.55e-5	sum(ER12c1_01,ER13c1_01,ER14c1_01,ER...		4.62e-6	4.10e-6	3.81e-4
⊞ Loss of Control in Landing	Per flight (or per fligh...	OUT_LOCL	1.25e-2	sum(AL19d1_01,AL19f1_02,AL19h1_04,AL19...		7.50e-3	9.41e-4	4.27e-2
⊞ Loss of Control in Take-off	Per flight (or per fligh...	OUT_LOCT	1.28e-3	sum(TO01d1_01,TO01e1_02,TO02d1_01,TO...		1.23e-3	3.92e-4	3.10e-2
⊞ Structural Accident	Per flight (or per fligh...	OUT_Structure	2.78e-8	ER33c1_01		5.44e-8	6.25e-9	5.06e-7
⊞ TOTAL All Accidents	Per flight (or per fligh...	OUT_TOERAL...	1.51e-2	sum(TO01d1_01,TO01e1_02,TO02d1_01,TO...		7.63e-3	1.55e-3	9.15e-2
⊞ TOTAL All accidents excluding wake vortex and ...	Per flight (or per fligh...	OUT_TOERAL...	1.41e-2	sum(TO01d1_01,TO01e1_02,TO02d1_01,TO...		7.62e-3	1.55e-3	9.15e-2
⊞ Collision with ground (wake encounter)	Per flight (or per fligh...	OUT_Wake	2.17e-8	ER37c1_01		3.37e-9	1.58e-8	1.49e-6

Figure 6-5 Bad management influences

6.1.5 Discussion

In this exercise, experts were generally comfortable and enjoyed having their expertise extracted in this manner. Five of the seven experts, who also participated in the BBN elicitation for flight crew error, considered that the approach described in this chapter was relatively an easier and more intuitive elicitation method for them. Paired comparisons is also a feasible method to do with the interviews in real applications. On average it took approximately one hour for an expert to finish the 91 paired comparisons questions.

Although most experts were doing fine with the consistency test, some experts indicated that prior to the elicitation the strict formulations made it difficult for them to judge the relative management influences on a given parameter. Whether the strict formulation of the variable has an effect on the number of the circular triads was raised after the elicitation. The results presented in Table 6-2 show that experts generally performed less well in the weather session than in the fatigue session. Although the threshold in the weather session is lower than in the fatigue session, more experts were removed from the weather analysis which indicated that some experts had more unstable preferences there. This may partially explain the lower consistencies from the results of the participants. However, it should be noted that the results are not directly comparable because the studies were carried out on a different list of management actions and also on a different parameter. Moreover, there may be other reasons for experts to have (more) circular triads (David, 1963). The expert may be less competent to answer the weather question than the fatigue question; or, the questions in the weather session may be generally more difficult than those in the fatigue section. Another explanation may be that in this case there is little or no significant difference in the objects with respect to preference, or, there is indeed no valid ordering of the objects (equally important) even when they differ markedly. Therefore, a case study on a proper weather definition with the same pairs of questions to the same expert group would be a first step towards verifying this statement, if we are to check this hypothesis properly.

6.2 Second experiment

As one of the important findings in Section 3.3, the analysis of CATS, SoTeRiA, and IRP showed that the human factors formulated in the current (probabilistic) quantification models are still very partial. Some of the human factors are left out of the model because it is argued that they cannot be quantified successfully. “Procedures” is one of these factors, which were initially identified as important influencing factors in CATS that contributes to flight crew error probabilities, but in the final version of the CATS project were not incorporated into the model. The difficulty was that several dimensions which are relevant to the effective use of procedures (e.g. quality, accessibility, usability, and effectiveness) were considered too difficult to model and quantify in numerical units. Therefore, this variable was totally left out of the model.

It was decided to take the opportunity of this experiment to examine the issue of procedure use and to consider this variable from a broad viewpoint, and quantify the impact of management influences on it using the same method proposed in this research. If this method is feasible, it can provide a first step in bridging the gap between qualitatively generally understood notions and a quantitative modelling technique. Therefore, the question arises:

- can the paired comparison method be effectively applied to a qualitative variable?

However there is an additional question relating to combining the paired comparison results with the quantification studies of the BBN network. In the quantitative cases as shown in the previous experiment, it is fairly easy to find distributions for PSFs (which were precisely defined with objective quantifiable units) and then anchor the weighting results from the paired comparisons into such distributions. But, in qualitative cases, how to build the distributions for “soft” variables is less obvious. Without the node being precisely defined and its probability density function being demonstrated, we are not able to progress to the step of asking experts to give their judgment on the anchoring points. Therefore, one will need to provide some solutions to this question, if one wants to incorporate the quantitative management influences into the human models. In the second experiment, we will only focus on finding the rank orders and the relative weightings for “procedure”. An idea to quantify its distribution will be discussed in Section 7.3 in Chapter 7.

6.2.1 Experiment design

Identify the targeted model variables in the flight crew HPM

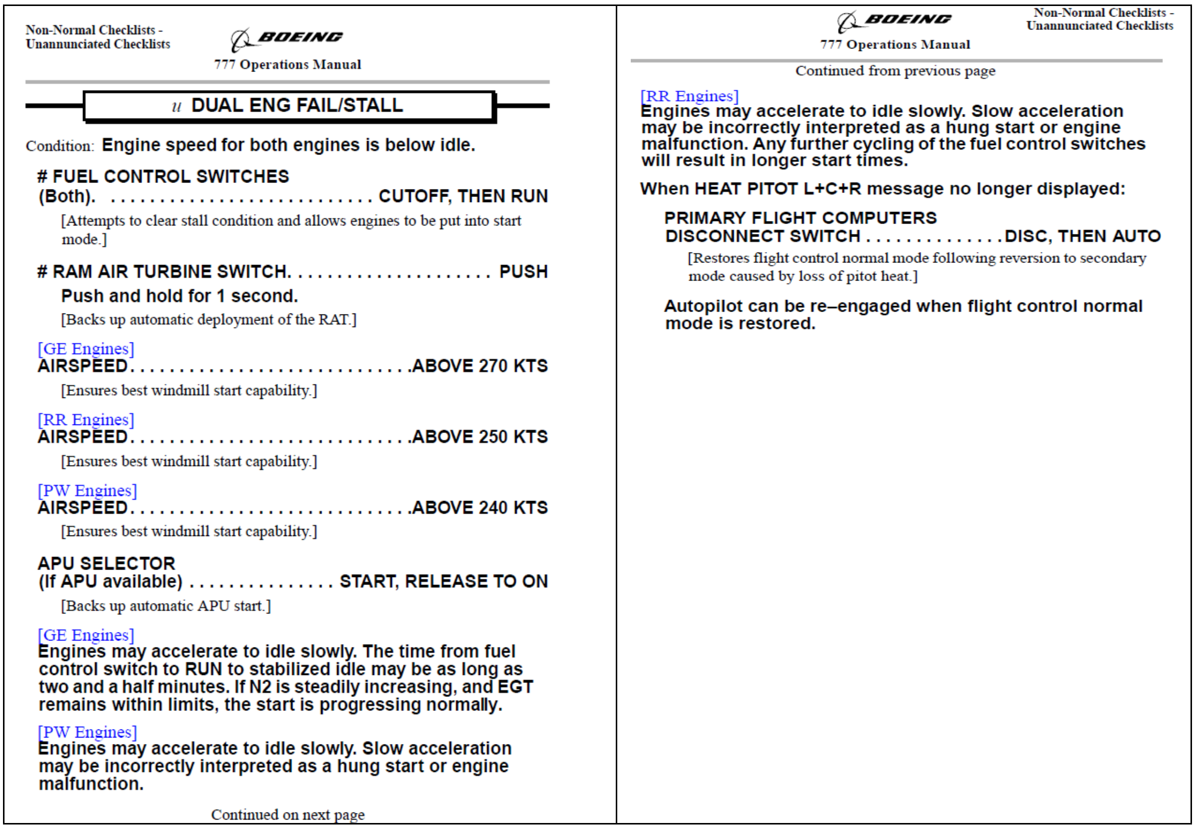
In complex system such as aviation, operations, training and standardization largely depend on an elaborate set of procedures which are specified and mandated by the regulator and/or the operational management of organization. These procedures are intended to provide guidance to the pilots, to ensure a logical, efficient, safe, and predictable (standardized) means of carrying out the mission, both in normal and abnormal situations. Accurate, readable, clear and up-to-date procedures influence how the primary process is executed in operating an aircraft. Besides, procedures are involved significantly in diagnosing a situation (such as using a checklist for complex and emergency situations) and acting on it. However, in some operations these procedures can become a motley collection of things with little coherency in terms of internally consistency and operational logic, or simply inadequate for the task at hand. When operating rules and procedures are deficient, not only will this make it difficult for pilots to obtain what they want from procedures, but it may also lead to deviations and may cause tragic human and technical consequences.

The airframe manufacturer is the first to design procedures for a new aircraft. The airline bases its own procedures on those of the manufacturer, but can deviate from manufacturer's by amending procedures to meet company needs. The company needs are usually determined by the company's philosophy of operation, which is largely influenced by the individual philosophies of the top managers in combination with e.g. economic factors, new generation of aircraft, airport policies or by the company's culture. However, exceptions to the above are abnormal and emergency procedures (A/E procedures). In the abnormal and emergency cases, most airlines adopt the manufacturer's procedures and modify them to a lesser extent if at all (Degani, 1994). This is certainly because they are more critical than the standard operating procedures. Because of the lesser extent to which such uncertain organizational factors are attached to A/E procedures during the design process, we decided for reasons of simplicity to concentrate in this exercise only on abnormal and emergency procedures.

In practice, immediate actions in response to certain emergency or abnormal situations (e.g. engine fire) are carried out from memory, after which the action taken is confirmed and subsequent actions read off by referring to the checklists in a quick reference handbook[35] extracted from the A/E procedures. Two copies of the quick reference handbook must be provided on the flight deck so that both pilots have access to a copy. The checklist forms part of the operations manual and airlines have to ensure that the checklist extracted from the A/E procedures provides clear guidance to pilots during the abnormal/emergency situations. An example of non-normal checklists for dual engine failure/stall can be seen in Figure 6-6.

In the expert elicitation, we asked experts to identify the relative importance of different management actions on improving the *effectiveness* of A/E procedures to reduce flight crew error rates. Before working with such a qualitative variable, we needed to ensure that judges focused their attention on the same features of the variable. Therefore, we defined the characteristics of an effective A/E procedures as *"accurate, safe, clear, up-to date and easy to read and understand."* This definition is used in the paired comparison expert elicitation.

[35] A quick reference handbook is a handbook containing extracts from the Operations Manual which may need to be referred to quickly and/or frequently, usually including Emergency and Abnormal procedures.



Non-Normal Checklists - Unannunciated Checklists

BOEING

777 Operations Manual

u DUAL ENG FAIL/STALL

Condition: Engine speed for both engines is below idle.

FUEL CONTROL SWITCHES
(Both). . . . . CUTOFF, THEN RUN

[Attempts to clear stall condition and allows engines to be put into start mode.]

RAM AIR TURBINE SWITCH. . . . . PUSH

Push and hold for 1 second.

[Backs up automatic deployment of the RAT.]

[GE Engines]
AIRSPEED. . . . . .ABOVE 270 KTS

[Ensures best windmill start capability.]

[RR Engines]
AIRSPEED. . . . . .ABOVE 250 KTS

[Ensures best windmill start capability.]

[PW Engines]
AIRSPEED. . . . . .ABOVE 240 KTS

[Ensures best windmill start capability.]

APU SELECTOR
(If APU available) . . . . START, RELEASE TO ON

[Backs up automatic APU start.]

[GE Engines]
Engines may accelerate to idle slowly. The time from fuel control switch to RUN to stabilized idle may be as long as two and a half minutes. If N2 is steadily increasing, and EGT remains within limits, the start is progressing normally.

[PW Engines]
Engines may accelerate to idle slowly. Slow acceleration may be incorrectly interpreted as a hung start or engine malfunction.

Continued on next page

BOEING

777 Operations Manual

Non-Normal Checklists - Unannunciated Checklists

Continued from previous page

[RR Engines]
Engines may accelerate to idle slowly. Slow acceleration may be incorrectly interpreted as a hung start or engine malfunction. Any further cycling of the fuel control switches will result in longer start times.

When HEAT PITOT L+C+R message no longer displayed:

PRIMARY FLIGHT COMPUTERS
DISCONNECT SWITCH . . . . .DISC, THEN AUTO

[Restores flight control normal mode following reversion to secondary mode caused by loss of pitot heat.]

Autopilot can be re–engaged when flight control normal mode is restored.

Figure 6-6 Non-normal checklist for dual engine failure/stall

Generate relevant management actions to influence the variable(s)

Procedures are not developed on their own. Nor are they inherent in the equipment. Management can influence the quality, accessibility and usability of procedures through a range of decisions and influences in the business process. Procedures have to be specified, made explicit, certificated and in place for all safety-critical activities in the organization. A flight training programme should be established to ensure that all flight crew are adequately trained to perform their assigned procedural tasks. An explicit evaluation system with suitable performance indicators for monitoring the success of procedures and goals should be also in place. All of these steps are covered under the Dutch model's "procedure" delivery system and should be managed by organizations. Hence, management actions for improving the effectiveness of A/E procedure to be used in real-time emergencies were generated in Table 6-7 using this logic. The same method of devising the influences was used as in the first experiment, with use of the delivery system to structure the search for influences, discussion with domain experts, the study of literature and the testing of the initial list of influences on a friendly expert. This resulted in 16 influences being defined,

The initial list of the management actions derived for A/E procedure was considered too long for use in the paired comparison exercise. Hence, expert sorting was used in order to reduce the 16 preliminary management actions to a manageable number of important influences. Two expert pilots were asked to add any management actions they missed, group any influences whose management was inseparable and do a card sort of the influences into three groups. The items were discussed at the start of the sorting to clarify any issues about what was being asked. No management actions were added or grouped; only the phrasing was modified. Each of the experts then did a card sort into three groups - very important, marginally important and in between. Influences placed by all two in the "marginally important" category were eliminated.

Table 6-7 Management actions for improving the effectiveness of A/E procedures

1	Ensure aircraft manufacturers provide good quality of A/E procedures to airlines
2	Ensure aircraft manufacturers provide to airlines a suitable set of procedures for all relevant A/E situations
3	Design A/E procedures and improve their quality by validating them in a flight simulator using line pilots
4	Ensure airline set and meet clear criteria for readability, clarity, depth of A/E procedures
5	Publish the A/E procedures in the designated common language(s)
6	Effectively train all relevant A/E procedures in the simulator
7	Ensure feedback from the training department to the flight technical department for any necessary changes to the A/E procedures
8	Ensure prompt communication of, and uniformly training in any changes made to A/E procedures
9	Ensure that documented procedures are carried onboard for each flight and located in a way easily accessible by the flight crew
10	Provide a feedback system for flight crews to identify and report changes needed to the A/E procedures or the need for more training in simulators
11	Use flight data and audit data to monitor the use of A/E procedures and indicate changes needed to them or the need for more training in simulators
12	Properly distribute any changes of A/E procedures to all flight crews, by identifying the dates and the version of operational documents
13	Provide operational feedback to aircraft manufacturers and regulators for better design of the A/E procedures
14	Standardize the format for the outline, general rules, checklist names, and standard text for A/E procedures across fleets
15	Ensure airlines amend the A/E procedures from aircraft manufacturers to adapt them to the specific company needs and style
16	Provide a clear prioritisation of the critical tasks in each AE procedure

Three management actions were eliminated, one of which related to "standardize the procedures across fleets" (item 14); and two related to "customizing A/E procedures from the manufacturer" (item 15 &item 16). The latter which had been eliminated by both of the experts corresponds to our expectations from the literature[36]. The final list of 13 influences to be used in the paired comparison is that from items 1 to 13 in Table 6-7.

6.2.2 Elicitation procedure

Assess the impact of the management influences
In order to offer the possibility to compare the experiment with the first experiment above, with variables that were more quantitatively defined, the elicitation session was held with the same expert group and the same protocol used in the first experiment. The definition of the variable and what is considered to be an "effective" A/E procedure was discussed explicitly

[36] Although according to the literature, in most cases airlines do not modify the procedures from the manufacturers, to be certain we put these two management actions in the preliminary list for card sorting. If they were not suitable or not important for this case, we expected that they would be deleted.

with each of the experts at the start of the elicitation to ensure this qualitative characteristic was interpreted consistently by the experts in the experiment. The experts were asked to work individually through a set of 78 (c_2^{13}) paired comparisons, in which each management action was compared to each other. The experts indicated for each pair which action would have more impact on the improvement of effectiveness of A/E procedures to be used in real-time emergencies.

After the previous experiment, it was felt that if the experts were asked to give some information about the elicitation it would help understand better the cause of any circular triads and help the analysts improve the design of experiments. Hence, after the elicitation experts were explicitly asked

- whether it was difficult to determine relative importance for a qualitative variable defined as such; and
- during the exercise which comparisons they had the most difficulties with, and what the main reason was for that.

6.2.3 Data collection and analyses

The experts were tested for inconsistencies by analyzing the number of circular triads (see Table 6-8). In this exercise, two experts (e2 and e7) were removed from the analysis for not passing the threshold for the circular triads. The coefficients of agreement (*u*) and concordance (*W*) of the remaining experts passed the statistical test at the 5% level which means there is a reasonable consensus across the remaining expert group. In addition, the goodness of fit test used to test the appropriateness of the BT model could not be reject at the 5% level of significance. Table 6-9 gives the rank orders and weights assigned to the management actions for improving the effectiveness of A/E procedure. Table 6-10 shows the rating summed for the generic management steps in “procedure delivery”.

Table 6-8 Number of circular triads for A/E procedure

Expert	A/E procedures (n=13)
e1	7
e2	37
e3	25
e4	20
e5	2
e6	29
e7	46
Circular triad threshold	34

Table 6-9 Management influences on procedure (n=13)

	Item name	Weighting %	Type of influence and its steps
9	Ensure that documented procedures are carried onboard for each flight and located in a way easily accessible by the flight crew	53.20%	Procedure-use
4	Ensure airline set and meet clear criteria for readability, clarity, depth of A/E procedures	8.29%	Procedure-provide
6	Effectively train all relevant A/E procedures in the simulator	8.29%	Procedure-train
13	Provide operational feedback to aircraft manufacturers and regulators for better design of the A/E procedures	6.26%	Procedure-change/maintain
2	Ensure aircraft manufacturers provide to airlines a suitable set of procedures for all relevant A/E situations	4.30%	Procedure-specify
7	Ensure feedback from the training department to the flight technical department for any necessary changes to the A/E procedures	4.14%	Procedure-change/maintain
1	Ensure aircraft manufacturers provide good quality of A/E procedures to airlines	3.45%	Procedure-provide
11	Use flight data and audit data to monitor the use of A/E procedures and indicate changes needed to them or the need for more training in simulators	3.20%	Procedure-monitor
3	Design A/E procedures and improve their quality by validating them in a flight simulator using line pilots	2.98%	Procedure-provide
5	Publish the A/E procedures in the designated common language(s)	2.21%	Procedure-provide
10	Provide a feedback system for flight crews to identify and report changes needed to the A/E procedures or the need for more training in simulators	1.90%	Procedure-monitor
8	Ensure prompt communication of, and uniformly training in any changes made to A/E procedures	1.01%	Procedure-train
12	Properly distribute any changes of A/E procedures to all flight crews, by identifying the dates and the version of operational documents	0.79%	Procedure-change/maintain

u=0.0974, W =0.38, p = 0.0054

Table 6-10 Relative weighting (%) of generic influences within procedure delivery system

Management steps	**Weighting %**
Procedure-use	53.20%
Procedure-provide	16.93%
Procedure- change/maintain	11.19%
Procedure-train	9.30%
Procedure-monitor	5.10%
Procedure-specify	4.30%
Grand Total	100.00%

Focusing on the untransformed values in Table 6-9, “ensure that documented procedures are carried onboard for each flight and located in a way easily accessible by the flight crew” is deemed to be the most important: that the pilot can reach the checklist, take it out and actually use it. But, with hindsight this may be a too obvious influence to include, its importance swamping as it does all others. The importance is followed by two equal influences “clear criteria for readability, clarity, depth” and “effective training”. As the real world occurrences of abnormal and emergency situations are rare, the more practice one has the more one gets prepared and one’s skill level is maintained adequately. However, in real cases the number of abnormal and emergency situations is rather small so there is a limited amount of information which airlines can distribute to their flight crews to learn from. Besides, airlines rarely change the content of A/E procedure by themselves, but it is more important for them to feed changes back to aircraft manufacturers, after which it is up to them to revise the procedures (item 13). Under these circumstances, the influence of item 12 was therefore considered marginal by the experts. But, it should be born in mind that this judgment is crucially determined by the original list of possible influences to the designed variable. So, this conclusion only applies to the abnormal and emergency procedures, and was not tested for other procedures (the standard operating procedures) in aviation.

6.2.4 Findings and discussions

This paired comparison exercise went smoothly. All the experts were more content with the A/E procedure defined in this way than with the definitions in the previous experiments. Experts indicated that this variable gave a better representation of reality and facilitated the judgment of the relative importance of the management influences on the variable. So, they generally felt quite comfortable to judge the relative importance of management influences to this qualitative variable. The coefficients were acceptable and the number of experts with results of insufficient consistency was similar to the previous quantitative variable studies.

After the elicitation when the experts were specifically asked to identify the difficulties they had, some experts (e2, e3 and e7) indicated that this paired comparison exercise was more difficult than the previous one (the fatigue case). They all claimed that this was due to the more generic level of the management items in terms of functionalities and these experts were later proved to have low consistencies (high circular triads) within the expert group.

In addition discussion with all the participants showed that this method of explicitly referring to the management life cycle may, however, be difficult for the experts, because of the circular nature of the influences being judged. As has been pointed out, it is characteristic of the Dutch model that, in order to maintain or improve the system performance, one must incorporate monitoring and feedback into the management process and correct those prerequisite actions on a timely basis. If the expert considers that the actual change gives more effect, but will not occur unless the monitoring reveals that need, he must decide which to mark as most important. For example, item 1 and item 13 are contiguous steps in the processes which feed into each other. Item 13 feeds back to item 1 and improves the quality of the latter, but item 1 is the prerequisite of item 13. In this case, the judge must mentally construct some function of the relevant characteristic and use this as a basis of comparison. Expert 1, although aware of the relation between these two actions, was able to rationalise to himself and us that procedure is a developing system, so despite the fact that the initial procedure may not be ideal the feedback and learning system is indeed the key element which improves initial results to a better quality. Nevertheless, not every expert may make this same rationale when confronted with such a question. Some experts may be unstable in this respect when declaring their preference. The occurrence of the circular triads which includes these two items with any other third item was found to be high in the data analysis.

Overall, concerning the main question addressed in this experiment, the results indicate that it is not a problem to apply paired comparison method to a qualitative variable, but the problem is in fact the ability to define and select management items that are relevant to the variable and not so interacting that it is impossible to distinguish and rate them.

6.3 Findings and suggestions

The results of both experiments in this chapter showed that in general paired comparisons is relatively a more easy and intuitive elicitation method than the complex BBN questions. The experiments demonstrated that the method designed in this research can be applied both for quantitative variables and qualitative cases. It is particularly useful that the "soft" variables could be modelled more closely to the reality of what can be influenced by management in clearly demonstrable ways, an approach not allowed for by the restricted numerical definitions imposed by CATS. This ability to model in more extensive and nuanced ways leads to greater understanding of the importance of management influences on human factors.

Table 6-11 compares the pros and cons of the two methods reviewed in Chapter 5 and this additional paired comparison method.

Table 6-11 comparison between quantification methods

	Paired comparisons + distribution free continuous BBNs	**Distribution free continuous BBNs**	**System Dynamics**
Advantages	Quick and easy to implement and experts are comfortable with the elicitation	Captures the uncertainty in the dependencies between variables	Incorporates a wide range of soft variables
	Copes easily with qualitative variables and their management influences		Allows feedback loops in a complex system
	Simple elicitation questions for experts		
Dis-advantages	Compresses feedback loops into independent influences in ways confusing	Cannot cope with feedback	Lacks a formal estimation method in its application to safety
	Cannot deal (easily)with dependencies	Gets much too complex as the number of parent nodes increase (maximum 5 or 6 parents nodes)	

From the table, it can be seen that each of the methods has its advantages and disadvantages. The paired comparison approach has been proven to be a useful tool to differentiate a set of management options to reduce human error. One virtue of this model lies in its role as consensus builder, another is that it can be used especially in cases when the decision problem is less tangible and the objects to be compared are (currently) impossible or impracticable to measure in relevant and objective ways. It seems to recommend itself particularly in ordering the influences into relative strengths, which could be used to reduce a complex, multi-factor set of influences to a number manageable for the more rigorous BBN analysis.

In the first experiment, the combination with distribution free continuous BBNs allows us to quantify risk both in terms of the current quality of the management actions in the supporting process that influence the outcome of the human factors, and in the causal relations between them to be expressed. It can also be used to compute the changes resulting from the management actions of interest, given information about other actions. In qualitative cases, to build a distribution of how much absolute effect the variable has is less obvious. More research into this aspect will be discussed in the next Chapter (Section 7.3).

The findings from the experiments in this chapter yield a way to modify the procedure and give a direction to continue exploration of this interesting field. The recommendations to create an experimental design in which a satisfactory ranking could be achieved are as followed:

1. The variable should be clearly formulated and realistic to the problem, no matter whether it is qualitatively or quantitatively formulated.
2. The variable should be so defined that the management influences can change them.
3. Management influences should also be clearly defined and (preferably) as specific as possible to avoid strong self-interpretation by the judges and should truly reflect all the things management can do to change the variable.
4. The deliberate inclusion of feedback cycles (as shown in the second experiment) may have come at the cost of more circular triads in an expert's preference structure, as some participants tended to hesitate when confronted with this situation. Thus, explicit cycles in the management items should be avoided.
5. Finally, discuss with each expert prior to the rating exercise about the phrasing of the management items, but beware introducing individual interpretation differences.

7 Integrated model and proposal for the future

As demonstrated at the beginning of this thesis, there is a need for accident analysts and safety regulators to have models that represent possible causal event sequence scenarios that include technical, human, and organisational factors. The CATS model was intended to provide a comprehensive model of all of the relevant levels in Rasmussen's hierarchy (Figure 7.1) in order to support the identification and implementation of risk control measures at all levels. However, as we noted earlier, things did not turn out exactly as envisaged and planned. As a result a number of holes were left in the model for later filling and some issues were resolved in ways which strongly limited the breadth and nuance catered for in the model. This left the room for this thesis to make proposals for a more articulate model, where it is possible to represent causal sequences and the full range of influences, but which can be connected to current or feasible future practice in safety data collection and analysis.

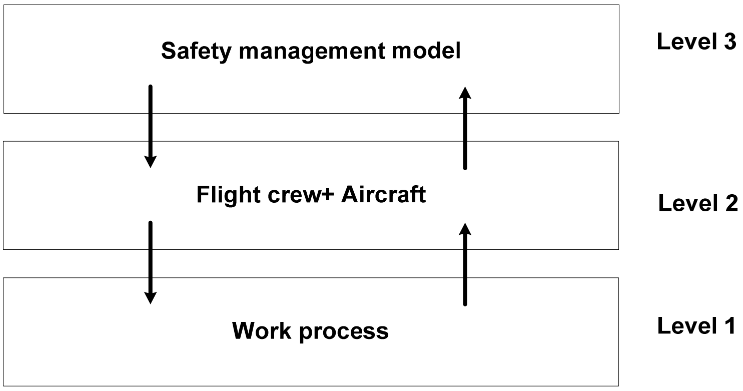


Figure 7-1 Hierarchical model

In the first section of this thesis (Ch2-Ch4), the modelling of human factors and technology failures at the lower level (1) in Rasmussen's hierarchical model (Figure 7-1) and a safety management model which can connect with them were discussed. A critical discussion of the CATS quantification method linked to the BBNs and some experiments in developing a supplementary method to get round the complexity of the expert elicitation that BBN method involves were discussed in the second section of this research (Ch5-Ch6).

The findings of the previous chapters which require improvements are listed here. We will concentrate on giving the improvements for each of the findings in the following sections:

1. Chapter 2 reviewed the development of the Dutch model. In the past the delivery systems were developed as resources and controls to the barriers, in order to have the barrier functions put and kept in place for their whole life cycle. The functions of the barriers were often modelled as observable human and technical actions (e.g. to detect, diagnose and act on threats). In the previous projects, the analysis of the individual factors underlying these actions did not play an important role in the development of the management system. Only with CATS, dealing with aviation (where human factors play a much more important role) was it felt necessary to understand in more detail the individual factors underlying the actions and interactions of human and hardware in order to link management into them. However, the current Dutch model did not give modellers sufficient clarity about where the management controls should go in relation to these individual factors. So in Chapter 2 (development of the Dutch model) and Chapter 3 (observations resulted from the mapping table), we have

criticized the Dutch model for being still too conceptual and generic in respect to resolving (preventing and coping with) human and technical errors. In the light of this, Section 7.1 provides a reflection on the basic hierarchical model in Figure 7.1 and particularly addresses the way in which the SMS needs to be linked with the technical and human factors. We conclude that to resolve these errors deeper analysis of the individual factor is an essential step to create effective management functions. Hence the set of factors which we came up with in Chapter 3 and Chapter 4 is where the management model should link in.

2. In Chapter 3, we summarised the whole list of human factors from accident analysis and from the rest of the project in Table 3.4. Factors in that table were mapped onto the existing Dutch management model to see if the Dutch model can support the control functions linked to them. Comparing the human factors with the Dutch model, we found that there are some changes which need to be made, namely the desirability of splitting the delivery systems for competence and suitability into its two component parts and the need for a workload delivery system, which is not quite the same as the one currently dealing with availability. There was also the need to reconsider the topics of safety culture and safety climate based on the conclusion from the comparisons with HFACS and SoTeRiA, to see if they fit sufficiently into the existing delivery systems, particularly that related to commitment and motivation. So in Section 7.2, we will take a closer look at these delivery systems in particular. For completeness we will also review the other delivery systems (procedures, communication/coordination and hardware/interface), to draw together the issues relating to them and their application to aviation.
3. An additional criticism in Chapter 2 was that, in the previous projects, the block diagrams showing the steps of the delivery systems which constitute their influences were not fully taken into account. This was partly due to the limited time scales available for the projects, but also because the concept model was not easy to apply in the past. So in Section 7.2, we will have a sub-section to build a generic structure of the safety management delivery in order to clarify the current steps in the delivery systems. Since each of the delivery systems should have this structure, we will work that out before the issues in the previous bullet are discussed. So, this generic structure will be presented in the first part of Section 7.2; and then the proposed changes for each delivery system and the operationalization of that general structure in the different delivery systems will be in the second part of Section 7.2.
4. In the review of the current quantification model in Section 3.3, we indicated that there is a strong need to translate the relevant qualitative variables into real, observable and quantifiable influences in risk modelling. Chapters 5 and 6 were designed to develop a quantification method that could assist in doing so. In Chapter 6, a quantification procedure for a qualitative variable (A/E procedure) was designed. But as we noted in that chapter, we need to handle the issues of generating the distributions of the qualitative variables and anchoring the management effects on them for a broader range of influences. Section 7.3 will particularly focus on these two issues.
5. Finally, in Chapter 5 we pointed out that the current limited data sources (ADREP, LOSA, EU-OPS and IOSA) do not sufficiently support the quantification of the causal relationships modelling. Section 7.4 discusses how these data sources could be improved so that they would become better data of sources.

7.1 A generic hierarchical control model for aviation safety

7.1.1 The treatment of hierarchical relations: A general structured approach

As we concluded in Chapter 2, the Dutch management model in the past defined the role of safety management in terms of the management of the life cycles of risk control barriers. This process is conceptualized as designing/selecting them, putting them in place, ensuring their use by the relevant operators, maintaining or modifying them to retain or restore their functioning and replacing or improving them where necessary. This conceptualization worked fine for physical manifestations of barriers. But when the model has to be applied to CATS where there is much more reliance on human skills and competence to act as the barriers to accident scenarios, the Dutch model did not give modellers sufficient clarity about where the management controls should go. Thus, we argued that the links from the Dutch safety management models to human behaviour should be more explicitly worked out.

To develop the link within Rasmussen's hierarchical model, we distinguish three levels of causation by describing a general structured approach based upon "control-internal process-action analysis" in Figure 7-2. Also, we will describe how each level in this qualitative model can be translated into the quantitative structure of CATS:

- Level 1 describes a series of observable *actions* performed by flight crew and aircraft in the flight process from gate to gate. In the CATS model, this is modelled at the level of the event sequence diagrams (ESDs) which formulate the possible deviations at the level of the primary flight process.
- Level 2 is the (hidden) internal process (of the cognitive mechanisms of the human and the equivalent internal functioning of the hardware) which leads to actions and interactions at level 1. As described in Chapter 4, the unsatisfactory actions of an aircraft in level 1 are influenced by the instrumentation design of the technical function and man-machine interface (MMI) in level 2. The failures of these functions in level 2 are often identified or modelled in fault trees, which describe the failure events, conditions and causes related to the hardware. For the failure of human actions, level 2 is where the human factors interact with the underlying cognitive process of the flight crew (see Chapter 3) and where the whole list of PSFs in Table 3.4 sits. Some, but by no means all, of these factors were formulated with BBNs in the CATS model (Section 3.3.1).
- Level 3 describes the safety management system that ensures that the outputs of system 1 (level 1) meet the objectives and goals set by that management system in system 2. In this version of the Dutch model, we see safety management as ensuring that the internal processes in level 2 are working properly and individual factors that interfere with it are managed to an acceptable level. So instead of defining the role of safety management in terms of the management of the life cycles of physical risk control barriers, we focus this version of the Dutch model on managing individual factors and internal processes of the human and hardware. The benefit of doing so is that the SMS can be devised more specifically and we can end up with a model tailored to the issues related to human factors found in the accident analysis.

There are two important features of this generic model. First, there are two distinct systems identified in this risk control problem. One system deals with the direct aspects of the execution of an action (system 1), and the other is a system that should prepare resources and controls for that action (system 2). Second, the link from system 2 to the level below (system 1) is best modelled as one of managing the internal processes of producing actions and also managing the individual factors that interfere with that. This shows how resources and controls or lack of them at a higher level influence lower level performance.

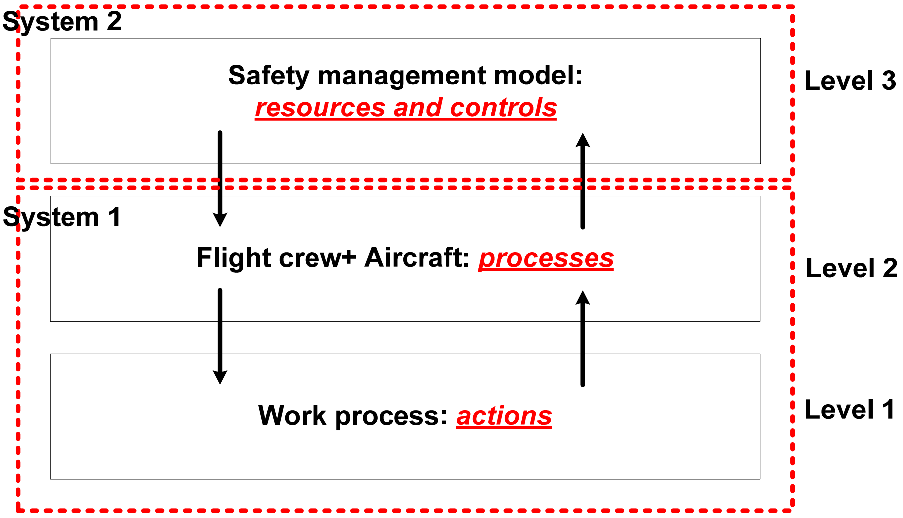


Figure 7-2 Hierarchical model: control-process-action

Since we have resolved the way in which the SMS needs to be linked with the technical and human factors into the structure of Rasmussen's hierarchical model, we can bring the theories and findings of the previous chapters to fit within Rasmussen's structure in order to have a more detailed modelling of the system levels including technical, human, and management systems in well-defined ways.

7.1.2 A detailed modelling of system levels: insights from previous chapters

Figure 7.3 depicts the model which puts the findings from the previous chapters in a unified model.

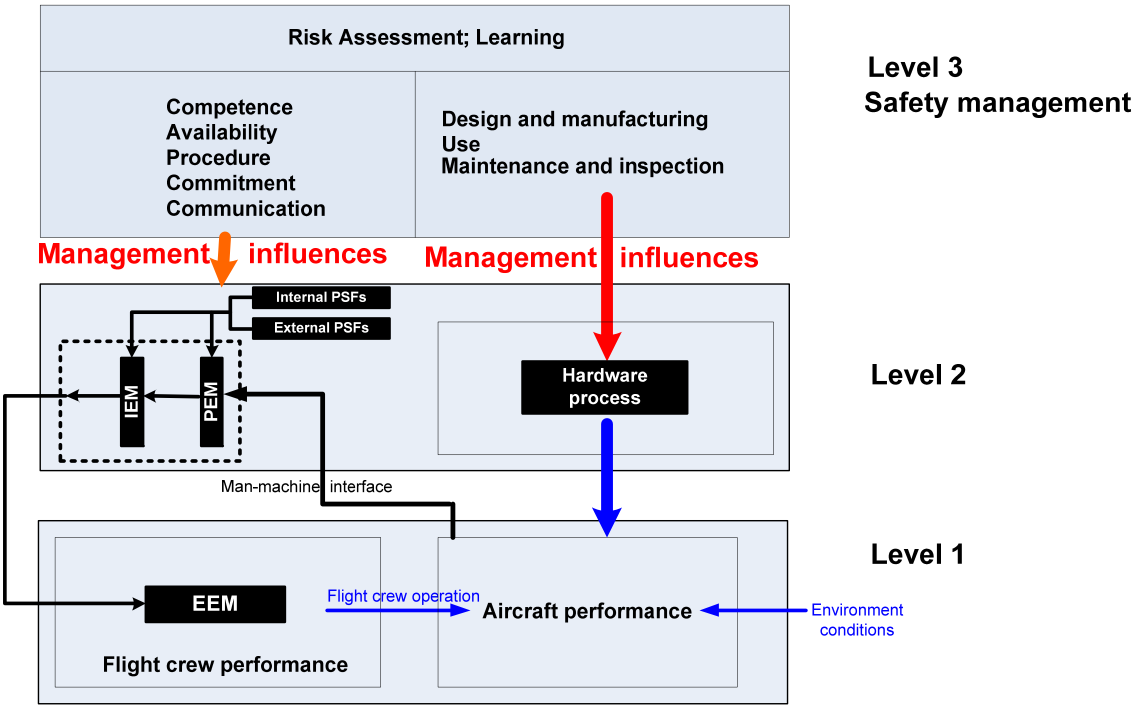


Figure 7-3 General structured model for aviation

From bottom up, the EEM are external observable error modes which are formulated based on logical outcomes of erroneous actions. The dashed box in level 2 indicates the internal "process" of the flight crew operation. PEM and IEM describe what cognitive functions failed or could fail, and in what way they failed. They are classified into four cognitive domains, namely "Perception", "Memory", "Judgment, planning and decision making", and "Action execution" (see Table 3.1 in Section 3.1.2 for a detailed description). Internal and external PSFs in level 2 both influence the pilot's performance through PEM and IEM, aggravating the occurrence of errors, but also influencing error recovery. We argue that these influencing factors are the "threats" that management should first provide resources and controls to manage, because it is considered more effective in the first instance to modify the situations and threats which people are facing, and only in the second instance try to influence directly their behavioural processes or EEM. To alter how a person's cognitive process functions is usually the last strategy management should look to for mitigation measures, simply because it is more difficult to modify effectively by management. The abovementioned influences from PSFs and from aircraft to human performance are indicated with black arrows.

Factors influencing aircraft performances lie in the pilots' online operation of the aircraft within the design limits, environment conditions, and the instrument design. The technical functioning and Man Machine Interface (MMI) all need to be working properly. All of these aspects have to be dealt with to ensure satisfactory performance of an aircraft system over its entire design life. Each of these aspects was discussed in Chapter 4. The influences from these aspects to aircraft performances are indicated with blue arrows in the flowchart.

In level 3, the management model is seen as providing the essential resources and controls to level 2. It adopts the concept of delivery systems and tasks within each delivery system. In ARAMIS, the delivery systems are modelled in detail as a series of steps representing their life cycle, which then can be seen as a set of actions taken by managers to deliver resources and controls to the human and technology in order to reduce risk to an acceptable level.

A key concept that is not explicitly shown in the hardware model in Figure 7.3 is that the hardware delivery systems (design, supply, inspection, maintenance, etc.) deliver their resources and controls to the aircraft through design engineers, maintenance technicians, and their software and tools. In other words, we really need to add the "human" cognitive process and its five delivery systems above the box showing the hardware process in Figure 7.3 to analyse these tasks more deeply for each of the design, manufacturing, inspection and maintenance operators. In principle, this deeper analysis will supplement the current simplified delivery systems for hardware. However, it makes the model more complicated. Currently, we suggest that only if the tasks related to the provision and maintaining of the aircraft are critical to an analysis being done, is it necessary to apply this deeper analysis to that context. In Chapter 4, we discussed some of the human errors and hardware failures in respect of design engineers and maintenance technicians. Errors from maintenance technician are probably the ones which currently most need further modelling development. Maintenance of equipment requires procedures, competent people who are available when and where needed, who communicate well with each other, are motivated to do their jobs safely and competently and use any equipment provided safely and skilfully. These same issues can also be modelled for any sub-task within the hardware delivery system block diagram (e.g. making maintenance plans). But for the software and tools used in maintenance which play a more remote role in the whole accident causal chain we would suggest not to apply the human factors iteration to their provision and maintenance.

Table 7-1 Influencing factors for flight crew (recapitulated from Table 3.4 in Chapter 3)

Definition	Description	Selected examples	Delivery systems
Internal PSF: physiological and psychological factors of the flight crew which might influence human information processing	• Technical and interpersonal skills that match the requirements of performance	• Experience for complex situation • Technical skill • Communication& coordination between flight crew, ATC, and operation center	• Competence • Communication&coordination
	• Physical fitness to perceive, process, respond to information and feedback information	• Human physical and sensory limitation • Medical illness • Hypoxia • Physical fatigue • intoxication	• Suitability • Availability
	• Psychological fitness to perceive, process, respond to information and feedback information	• Mental fatigue due to sleep loss or other stressors • Emotional state • Personality characteristics (complacency, overconfidence)	• Suitability
	• Decision to choose one from several possible courses of action and decide to commit to safety procedure above other personal and organisational goals	• Pressure to achieve • Personal objectives • Incentives, needs • Perceptions of organization's beliefs and attitudes (manifested in actions, policies, and procedures, affect its safety performance)	• Motivation to commit to safety
external PSF: Factors external to the flight crew which might influence human information processing	• Clear and relevant guides to adequate performance	• Operational manual • Checklist • Charts • Standard operating procedures • Emergency and abnormal procedures	• Procedure
	• Tools and materials relate to physical activity designed scientifically to match human factors (include working postures, materials handling, repetitive movements, work related musculoskeletal disorders, workplace layout, safety and health)	• Checklist layout • Display/interface characteristics • Automation/alerts/warning	• Technology-Man-machine interface (Instruments& workplace design)

	- Data and information	- Information from ATIS - Information from operation center	- Communication&coordination
	- Environment in which the action needs to be performed	- Traffic configuration o Traffic density o Traffic complexity (e.g. Runway length, Runway crossing, Runway condition, Runway slipperiness, Terrain at/near airport) - Possible flight delays - Ambient environment o Wind shear o Cross wind o Visibility in flight o Visibility at airport o Turbulence o Icing o Light condition	- No relevant management functions in the current Dutch model. Need to add under “Workload” delivery system (will explain in Section 7.2.7).

Up to now, we have emphasized that we need to link the delivery systems clearly to the whole list of human factors from accident analysis. Chapter 3 (Table 3.4) essentially does that. We recapitulate it in Table 7.1 for easy reference and summarize the main points we found in Chapter 3 and that table which necessarily need to be changed in the delivery systems;

- ✓ The possibility of splitting the delivery system for competence and suitability into its two components;
- ✓ The need for management of workload to cope with the environmental factors and the question about whether that can be put under the availability delivery system or whether it needs to be a separate delivery system; and
- ✓ The need to deal with safety culture and safety climate in relation to commitment and motivation, but also other aspects of the safety management system.
- ✓ The need to give clearer descriptions of what falls under each delivery system and to link this to the detailed steps of the different delivery systems.

In the next section, we will look inside each of the management delivery systems and discuss exactly how they should work.

7.2 New version of the Dutch SMS

Before going deeper into each of the delivery systems and discussing how each of the delivery systems works, we begin with a section about the generic structure of safety management delivery of resources and controls in order to clarify the current steps in the delivery systems. This generic structure has to encompass the theories which have stood the test of time (i.e. deliveries, tasks, and series of steps) inherited from the previous projects.

7.2.1 A generic structure per delivery system: A simplified model

Figure 7.4 draws a generic structure which is later worked out specifically per delivery system to deal with each of the PSFs.

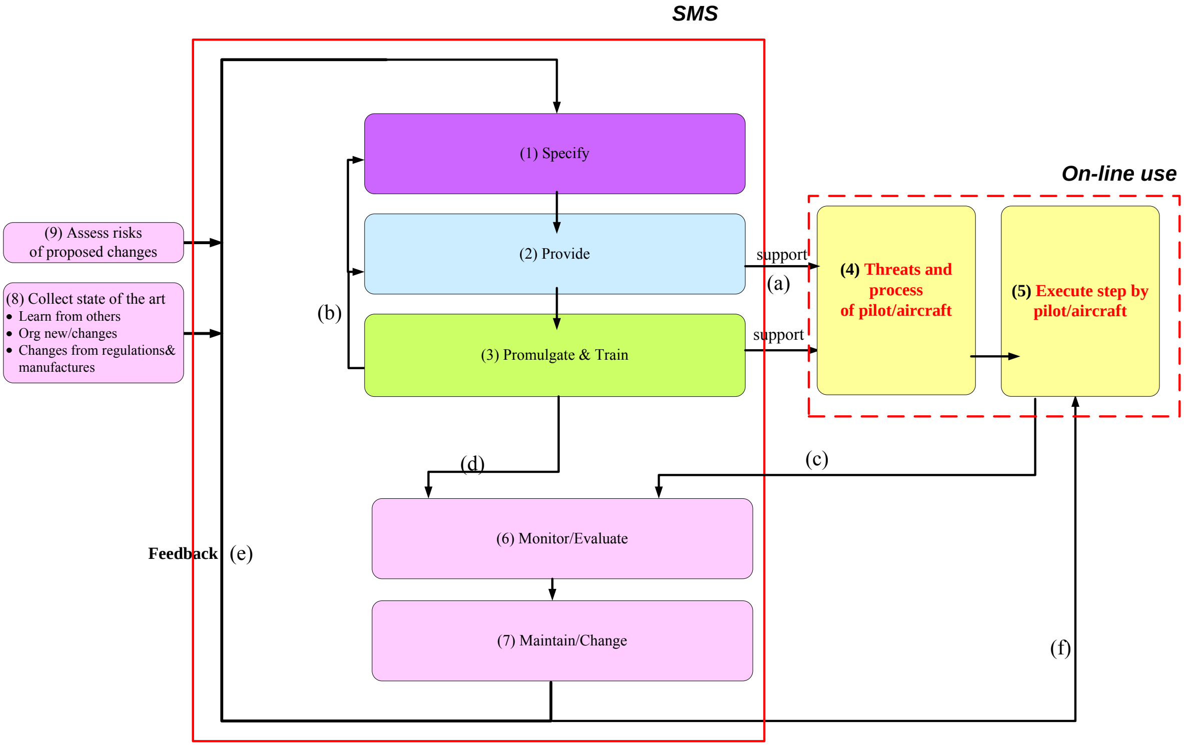

SMS
(1) Specify
(2) Provide
(3) Promulgate & Train
(6) Monitor/Evaluate
(7) Maintain/Change
(9) Assess risks of proposed changes
(8) Collect state of the art
• Learn from others
• Org new/changes
• Changes from regulations& manufactures
On-line use
(4) Threats and process of pilot/aircraft
(5) Execute step by pilot/aircraft
support
(a)
support
(b)
(c)
(d)
Feedback
(e)
(f)

Figure 7-4 General structured safety management model

The safety management system (SMS) is defined as a system which manages the resources and guidance needed to control risks. The management processes which make up that system and manage those risks can be broken down into 9 steps, forming a closed loop of a learning system. "Specify", "provide", "use", "monitor/evaluate", "maintain/change" are concepts taken from the previous Dutch model (see Chapter2). The model above (Figure 7.4) introduces two new or modified concepts to our previous model of the management process: "promulgate and train", and "threats and internal processes". It also makes the important distinction between the internal processes of the management level (steps (1) - (3) & (6) - (9)) and the output of each delivery provided to the execution level (steps (4) & (5)). In this section, each step is discussed and the modifications are introduced as they become relevant in our description of the model.

(1) **Specify:** what is the control measure, what does it have to do and under what circumstances, in collaboration/communication with what other measures? This step includes 2 major tasks

- Define processes, harm scenarios and risk control measures
- Carry out task analysis of behaviour and/or hardware as risk control measure

(2) **Provide:** ensure that the measure is designed, built, procured, installed and adapted/adjusted to its operating circumstances

(3) **Promulgate and train:** this step is especially important for risk control measures involving humans. As noted by the airline (Section 2.3.2), in most cases management have to train their workforce to be able to perform the designated actions after the risk control measure in the hardware and/or behavioural delivery system has been designed. Indeed, promulgation and training are the "means" that *deliver* resources and controls from management to the online workforce. The link (b) indicates the feedback from trainees gathered during the training sessions to the "specify" and "provide" department. This information is particularly valuable to improve the design of the control measures in the delivery systems.

(4) **Threats and internal process:** we consider safety management as the process to provide the resources and controls designed to ensure that the internal processes are working properly to analyse and deal with the threats (individual factors) that interfere with it, so that they are managed to an acceptable level. So, the whole set of the human factors summarized in Table 7.1 and the technical analysis in Chapter 4 can be plugged in directly to this block.

(5) **Actions executed by pilot and aircraft:** this is the direct execution of the primary work process. The technology and the human are fed by different delivery systems to execute online tasks. We need the step to indicate whether the human and technology work or not and do what they are designed to do.

(6) **Monitor/Evaluation:** "monitor" is to detect (potential) deviations from specified functioning and forestall or correct them. This happens along link (c) for monitoring on-line performance, and along link (d) to monitor or audit whether management functions (1) + (2) + (3) are carried out appropriately or as specified. "Evaluation" is the step to assess actual performance against specification. This should also take into account the evaluation at the two levels for on-line performance and management performance.

(7) **Maintain/change:** "maintain" covers actions to restore functioning to the original level and includes repair and revision of hardware, refresher training of competence and reinforcement of motivation and commitment. "Change" covers the application of new insights to improve functioning or replace with better functioning measures. The improvement could follow the feedback (e) from (7) to the management

functions of “specify”, “provide” and “promulgate and train”; it can also be feedback (f) to the workforce, such as discussion about on-line use of procedures, appraisal of safe acts, etc. The feedback loop (e) can be seen as the dynamic learning of a safety management system. This loop usually takes more time and goes through higher level management. However, learning could also take place at a more personal level as shown in loop (f). This is usually through on-line supervision during flight or (de)briefing before or after flight.

(8) **Collect state of the art:** this includes learning from others, changes within the organization which require adaptation of a particular delivery system (or systems), and changes outside the organization, e.g. regulations, aircraft designs, etc.

(9) **Assess risks of proposed changes:** this is the basis for all decisions about risks which are present and the risk control measures to be taken. It is the starting point and foundation of successful risk management.

Based on the logic proposed above, each delivery system needs to contain all of the steps from (1) to (3) and (6) to (9). Each delivery system should be properly managed by these steps within it. These steps suggest a direction to devise management actions to control human factors and technical failures in aviation modelling. The only difference between safety management for technology and for humans is that before the execution step technology does not need the step for “promulgate and training”. That is because technology does not have a choice. It either does what it is programmed to do, or it breaks down during the execution step.

As a final point, it should be noted that showing separate boxes in our model to distinguish the two levels of hierarchy between management and on-line operation emphasises an issue that was not underlined in the previous models. That is, that the output of the delivery systems determines the performance of the online barriers, either through how well they influence and control the human factors or how well they design and maintain the aircraft hardware. In the original model of the block diagrams, these output-input relations were hidden in the cycle. For instance, in the original ARAMIS block diagram for “competence” (see Figure 2.6 in Chapter 2) the “output” of “competent people” was not made explicit between the box dealing with training and the one labelled “monitor task performance”, which starts the steps of the monitoring and improvement part of the cycle. For causal modelling, these relations are important to be explicitly shown, because the SMS is not a closed system but rather an open system designed to provide the resources and controls to the on-line users. One of the primary benefits of doing so is that when one wants to transfer the block diagrams into the model, the blocks can be directly transferred into events and the links can be directly transferred into arcs in the quantification methods e.g. BBNs or system dynamics. The output then goes to the next layer of the human or technology specified in our boxes (4) and (5) in Figure 7.4, where the whole set of the underlying human mechanisms and technical analysis can plug in. Consequently, the integrated models of Figure 7-4 and Figure 7-3 can serve as the basis for the risk modelling which we offer to guide future research.

7.2.2 New and clarified delivery systems

Overall, each delivery system has the generic structure in Figure 7-4. In the next sections, we are going to work this structure out for each of the delivery systems, including the changes coming out of Table 7-1.

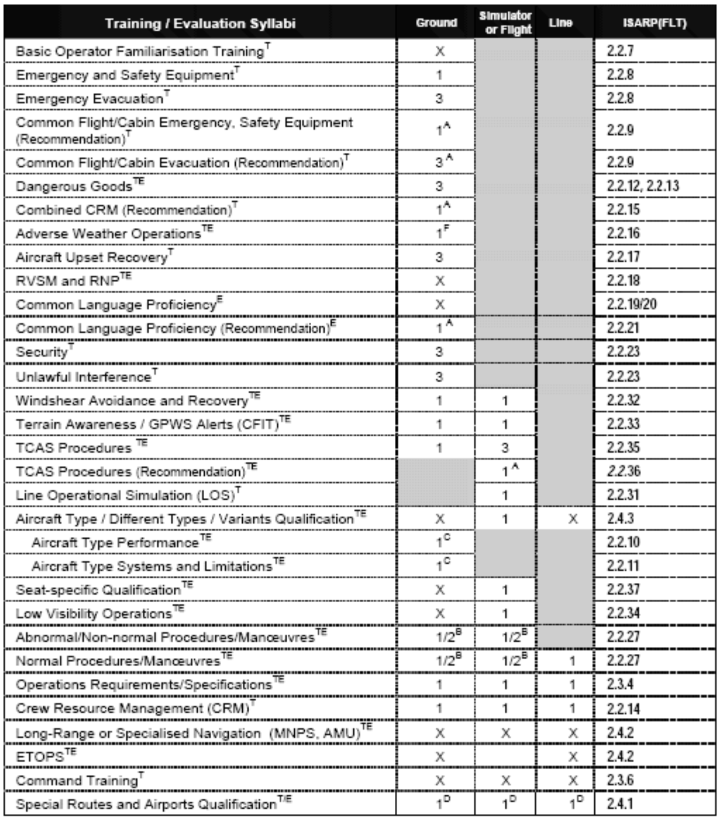
7.2.2.1 Competence and suitability

Competence refers to knowledge and skill to deal with a safety operation, critical procedures, special instrument aids, diagnosis skills, and emergencies/unusual events. Suitability pertains to both physical suitability (colour vision, size limitation, health, etc.) and cognitive qualities of persons (intelligence, field dependence, spatial aptitude etc.). Since the management steps for assuring each of these two outputs are different, focussing respectively on training and regular screening, we suggest splitting them into two delivery systems instead of the one found earlier in the Dutch model.

Competence

Competence refers to knowledge and skills learned through training and experience, although there is an element of selection to assess already acquired training and trainability. In CATS, the experience of a pilot was initially expressed as both total number of hours flown (all types) and number of hours on aircraft type. They were considered to be equally important and are obviously correlated. Because information on total number of hours flown (all types) is relatively easily obtained, total number of hours was selected as a simplification as the appropriate unit to represent pilot experience. Although both factors are certainly relevant, total time (no matter number of hours flown or number of hour on type) does not give the complete picture of a pilot's knowledge and skill. Table 7-2 shows an extended list of the training items (IATA, 2007) that are required to be obtained by a qualified pilot according to the IATA Operational Safety Audit (IOSA). This means the operator should have a ground and flight training programme, should ensure all flight crew members are adequately trained to perform their assigned duties, and evaluated in accordance with the specification in Table 7-2.

Table 7-2 Training and evaluation requirements (cited from IOSA)



Training / Evaluation Syllabi	Ground	Simulator or Flight	Line	ISARP(FLT)
Basic Operator Familiarisation Training^{T}	X			2.2.7
Emergency and Safety Equipment^{T}	1			2.2.8
Emergency Evacuation^{T}	3			2.2.8
Common Flight/Cabin Emergency, Safety Equipment (Recommendation)^{T}	1^{A}			2.2.9
Common Flight/Cabin Evacuation (Recommendation)^{T}	3^{A}			2.2.9
Dangerous Goods^{TE}	3			2.2.12, 2.2.13
Combined CRM (Recommendation)^{T}	1^{A}			2.2.15
Adverse Weather Operations^{TE}	1^{F}			2.2.16
Aircraft Upset Recovery^{T}	3			2.2.17
RVSM and RNP^{TE}	X			2.2.18
Common Language Proficiency^{E}	X			2.2.19/20
Common Language Proficiency (Recommendation)^{E}	1^{A}			2.2.21
Security^{T}	3			2.2.23
Unlawful Interference^{T}	3			2.2.23
Windshear Avoidance and Recovery^{TE}	1	1		2.2.32
Terrain Awareness / GPWS Alerts (CFIT)^{TE}	1	1		2.2.33
TCAS Procedures ^{TE}	1	3		2.2.35
TCAS Procedures (Recommendation)^{TE}		1^{A}		*2.2.*36
Line Operational Simulation (LOS)^{T}		1		2.2.31
Aircraft Type / Different Types / Variants Qualification^{TE}	X	1	X	2.4.3
Aircraft Type Performance^{TE}	1^{C}			2.2.10
Aircraft Type Systems and Limitations^{TE}	1^{C}			2.2.11
Seat-specific Qualification^{TE}	X	1		2.2.37
Low Visibility Operations^{TE}	X	1		2.2.34
Abnormal/Non-normal Procedures/Manœuvres^{TE}	1/2^{B}	1/2^{B}		2.2.27
Normal Procedures/Manœuvres^{TE}	1/2^{B}	1/2^{B}	1	2.2.27
Operations Requirements/Specifications^{TE}	1	1	1	2.3.4
Crew Resource Management (CRM)^{T}	1	1	1	2.2.14
Long-Range or Specialised Navigation (MNPS, AMU)^{TE}	X	X	X	2.4.2
ETOPS^{TE}	X		X	2.4.2
Command Training^{T}	X	X	X	2.3.6
Special Routes and Airports Qualification$^{T/E}$	1^{D}	1^{D}	1^{D}	2.4.1

Table 7-2 is a useful table which shows the extent of training, but it is still too complex and difficult for all training elements in column 1 to be operationalized separately. Therefore, some grouping is necessary. The extended list of the items can be reduced to the smaller number of types of competence (see Table 7-3), classified into four groups: normal flying (including commonly occurring difficulties), emergencies, other difficult situations, and crew management issues, of which the last category fits better under the communication and coordination delivery system and will be dealt with under the heading of that delivery system. Interviews with pilots and trainers would need to indicate whether this initial grouping proposal is a sensible set of distinctions, or whether, for example to merge the second and third categories or distinguish within the "normal flying" category those issues dealing with avoiding approaching the boundaries of the safe flight envelope from those far from those boundaries – see Table 7-3.

Table 7-3 Training and evaluation requirements classified into categories

Normal flying	• Basic operator familiarisation training • Common flight/Cabin evacuation (recommendation) • Common flight/Cabin Emergency, Safety Equipment (recommendation) • Dangerous goods • Security • Adverse weather operations • Windshear avoidance and recovery • Terrain awareness/GPWS alters • TCAS procedure • Line operational simulation (LOS) • Seat specific qualification • Low visibility operations • Normal procedure manoeuvres • Operations requirements/specifications
Emergency	• Emergency and safety equipment • Emergency evacuation • Unlawful interference • Abnormal/Non-normal procedures/Manoeuvres • Aircraft upset recovery
Special qualification (flying over difficult terrain or into special airport)	• ETOPS • Aircraft type/different types/variants qualification • Special routes and airports qualification • Long range or specialised navigation (MNPS, AMU) • RVSM and RNP
Crew management issues	• Combined CRM (Recommendation) • Crew resource management (CRM) • Command training • Common language proficiency

Figure 7.5 shows the modified block diagrams for the competence delivery system for aviation. Based on the block diagrams described in Figure 7.5, we can generate the management actions under the heading of each block diagram and further measure the quality of the different types of training (ground, flight or simulator) suggested in Table 7-2. These

management actions would include whether full motion flight simulators or advanced training devices (block 4) are used to enable a pilot to meet with more emergencies or fly into difficult airports, and how often the management schedule the training (block 5) and evaluation (block 6) for each element. The elements of training in ground, simulator or in flight should, in any case satisfy the requirements and frequencies suggested in the second to fourth columns in Table 7.2. There is also an issue related to lack of basic ability or proficiency to fly. For instance, a particular incident report referred to a pilot who failed the proficiency checks, got re-trained and finally passed a test on the same subject after a number of times (AAIB, 2001). Looking into his training history, we could argue that this pilot indeed performed below standard and seemed to have innate difficulties in this particular type of the skill, an aspect which should be controlled under box 3 in Figure 7.5.

It should be noted that we need to make a generic distinction in all delivery systems between the PSF and the delivery system governing it. For the "competence" delivery system the yellow shaded box ("competent staff") in Figure 7.5 is the PSF at individual level and the other selection and training boxes are the way that it is delivered.

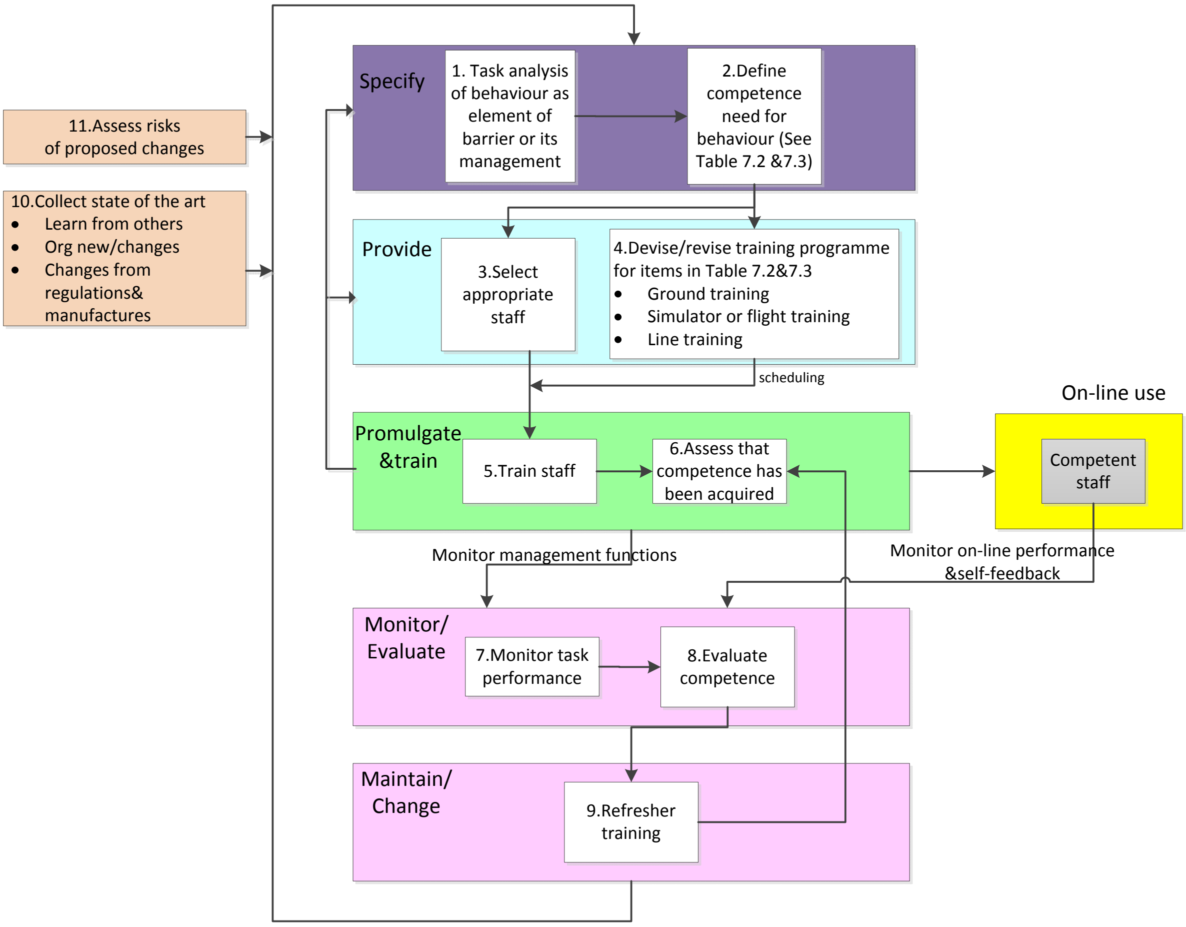

Specify
1. Task analysis of behaviour as element of barrier or its management
2.Define competence need for behaviour (See Table 7.2 &7.3)
11.Assess risks of proposed changes
10.Collect state of the art
Learn from others
Org new/changes
Changes from regulations& manufactures
Provide
3.Select appropriate staff
4.Devise/revise training programme for items in Table 7.2&7.3
Ground training
Simulator or flight training
Line training
scheduling
On-line use
Promulgate &train
5.Train staff
6.Assess that competence has been acquired
Competent staff
Monitor management functions
Monitor on-line performance &self-feedback
Monitor/ Evaluate
7.Monitor task performance
8.Evaluate competence
Maintain/ Change
9.Refresher training

Figure 7-5 Competence delivery system

7.2.2.2 Suitability

Factors that are known to influence physical and mental suitability can be split into short term conditions and long term conditions. The short term covers factors like drugs and alcohol, fatigue, and health, covered by screening programs and on-site monitoring; as opposed to the long(er) term covered by selection and annual medical checks. As said in the beginning of

Chapter 7.2, these issues of "individual fitness for the task" are only marginally influenceable by training. Therefore, an additional delivery system in Figure 7.6 (including initial examination, initial selection, later checking and monitoring) was developed for suitability to cover both inherent aspects such as colour blindness and physical size, as well as more malleable aspects such as health, physical fitness and personality.

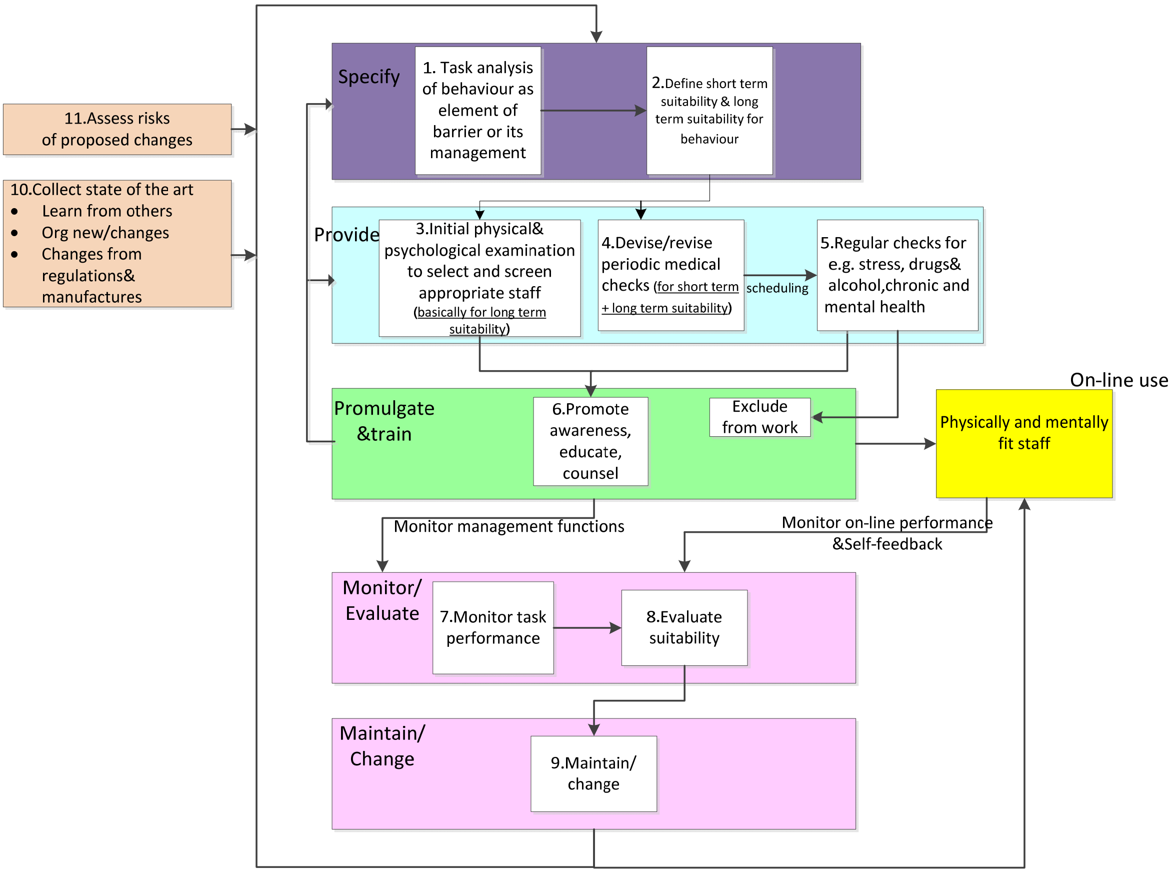

11.Assess risks of proposed changes
10.Collect state of the art
• Learn from others
• Org new/changes
• Changes from regulations& manufactures
Specify
1. Task analysis of behaviour as element of barrier or its management
2.Define short term suitability & long term suitability for behaviour
Provide
3.Initial physical& psychological examination to select and screen appropriate staff (basically for long term suitability)
4.Devise/revise periodic medical checks (for short term + long term suitability)
scheduling
5.Regular checks for e.g. stress, drugs& alcohol,chronic and mental health
Promulgate &train
6.Promote awareness, educate, counsel
Exclude from work
On-line use
Physically and mentally fit staff
Monitor management functions
Monitor on-line performance &Self-feedback
Monitor/ Evaluate
7.Monitor task performance
8.Evaluate suitability
Maintain/ Change
9.Maintain/ change

Figure 7-6 Suitability delivery system

For long term suitability (box 3), all certified pilots have to undergo physical examination by a government designated medical examiner and meet the appropriate medical standards to get their medical certificate. A number of studies (e.g. Shaw & Sichel (1971)) have shown that some people might have *personality issues in accident proneness* which might make them more likely to suffer accidents than others. Some accidents are more to do with perceptual skill which includes behaviour, cognition and affect that cause wrong responses. However, whether accident proneness actually exists as a distinct, persistent and independently verifiable physiological or psychological syndrome is still under question, because the exact nature of accident proneness has always been unclear and there are lots of factors intertwined with accident proneness such as carelessness, absent mindedness, use of alcohol and drug, etc.

For short term suitability (box 4), crew members are subjected to regular periodic medical examinations to ensure their continuing health and excluded from flying if they fail them. There is also a role for alertness among crew of their fellow crew members' short-term suitability to function. Factors which influence short term suitability, which can cause pilot incapacitation and may contribute to accidents/incidents, are stress, drugs and alcohol, chronic and mental health (e.g. heart attacks, emotion). In previous projects (e.g. IRISK, ARAMIS, WORM) we have been inconsistent in deciding exactly where fatigue is covered in the model.

It can be considered either under suitability (as a temporary impairment to be diagnosed) or under the availability of competent personnel. Since the fatigue management provided by airlines is managed under cockpit manning and rostering, we consider it better for this project to see it as an aspect of availability. So we will discuss fatigue in more detail under that heading.

Stress relates to other broader issues, such as personal situation, divorce, death of close relation, organizational loss of job, etc. These are issues for a regular check (box 5) - often best done by colleagues and supervisors rather than just medical staff - monitoring (box 7) and evaluation (box 8) of suitability. Drugs and alcohol are also obvious issues in aviation, and there are policies in place for those. Many countries have developed and implemented regulations for drugs and alcohol testing programs for safety sensitive personnel in the aviation industry. To proactively manage drug and alcohol misuse, a policy for testing is needed, together with processes developed by airlines in conjunction with employees, to promote education, and ensure feedback is in place to manage reports of substance impairment of flight crew.

7.2.2.3 Manpower planning and availability

Since the 1990s, *fatigue* has been a major issue in national and international transportation research programs. The issue of fatigue is quite often mentioned as a factor in accidents and is considered as an important factor in any discussion of pilot cognitive performance. Flight crew fatigue has been the subject of extensive research (Simon & Valk, 1993; Simons et al., 1994; Simons & Valk, 1997; Simons & Valk, 1998; Valk & Simons 1996) and data on fatigue levels is available from that research. In CATS, fatigue was quantified using the Stanford Sleepiness Scale (SSS). SSS measures are highly correlated with flying performance and threshold of information processing speed during periods of intense fatigue.

As mentioned in "competence and suitability", fatigue management provided by airlines can best be managed under cockpit manning and rostering. Currently in aviation prescriptive flight and duty time limitations focus on limiting the duration of work and on rest breaks, but they ignore many other factors such as bio-rhythm factors of the flight crew, the quality of on-board sleep (e.g. on-board bunk, rest on first class seat), and the balance between waking exertion and restorative sleep. ICAO indicated that current regulations do not address fatigue management in a manner that maximizes operational efficiencies and safety. Several recent scientific research studies on the causes of fatigue-related impairment (e.g. Mann, 1999; Gander et al., 1998; Akerstedt, 2000; Gander, 2001; Dawsan & McCulloch, 2005; Flight Safety Foundation, 2005; Rosekind et al., 2006; Wesensten, et al., 2004) have indicated a number of important factors:

1) Task-related factors such as duty hours,
2) Circadian factors that drive daily cycles in mental and physical performance, and
3) Sleep-related factors that relate to amount of sleep obtained both daily and on an on-going basis.

Thus, in order to investigate greater operational flexibility than the strict flight duty time limitation, in 2006 the ICAO Operations Panel formed a fatigue risk management subgroup to develop Fatigue Risk Management System (FRMS) guidelines material as the next step to update current regulations. In 2008, ICAO introduced FRMS to Annex 6 in a Working Paper. The new FRMS is a more flexible system which focuses on an adaptive, data-driven, continuous improvement programme for managing fatigue. If we recall our fatigue experiment in Section 6.1, the results in our experiment also indicated that more flexible countermeasures, such as a suitable crew rest environment and an appropriate place for a nap in multi-crew aircraft are indeed seen to be more effective than the current prescriptive duty time and rest scheme limitations.

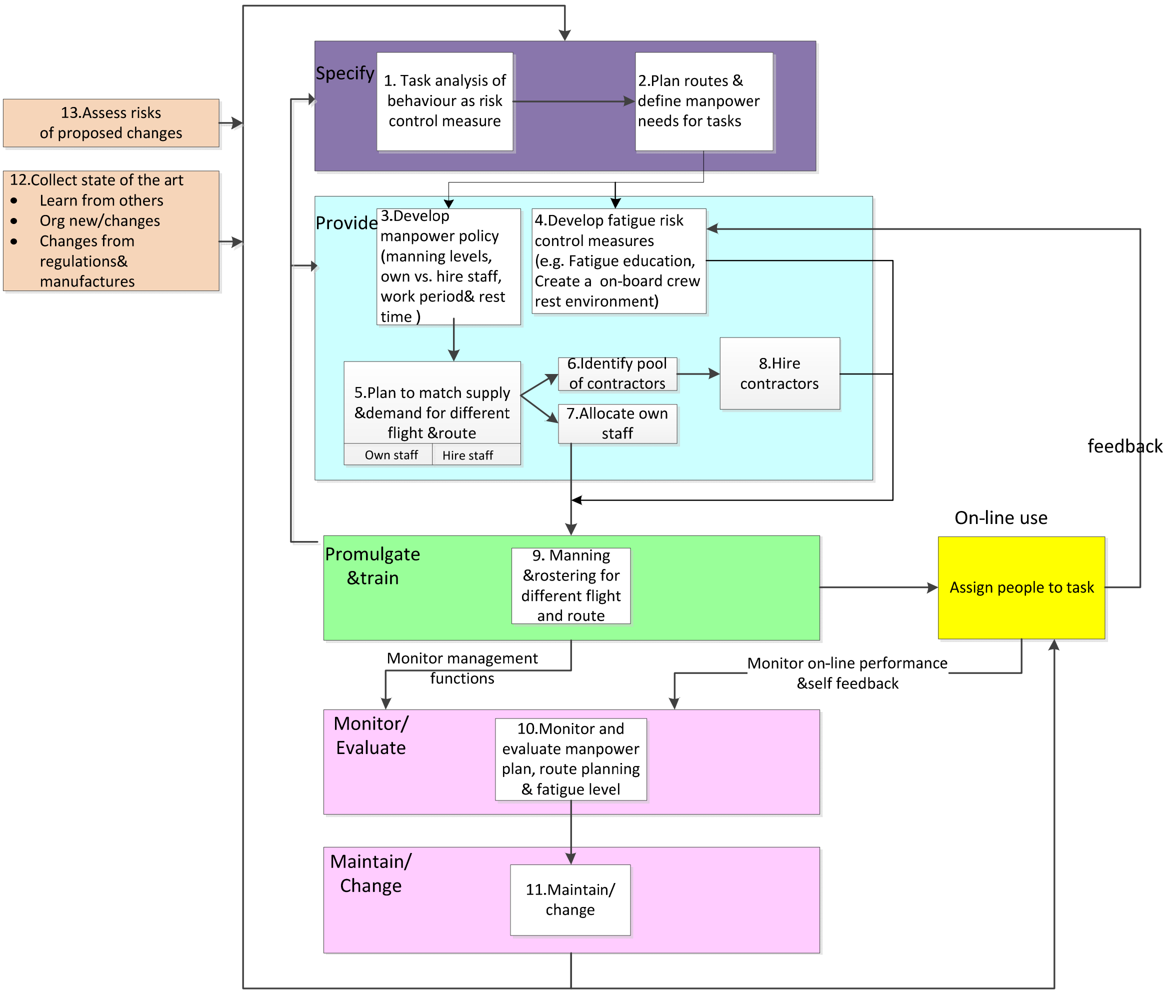

13.Assess risks of proposed changes
12.Collect state of the art
Learn from others
Org new/changes
Changes from regulations& manufactures
Specify
1. Task analysis of behaviour as risk control measure
2.Plan routes & define manpower needs for tasks
Provide
3.Develop manpower policy (manning levels, own vs. hire staff, work period& rest time)
4.Develop fatigue risk control measures (e.g. Fatigue education, Create a on-board crew rest environment)
5.Plan to match supply &demand for different flight &route
Own staff
Hire staff
6.Identify pool of contractors
7.Allocate own staff
8.Hire contractors
feedback
On-line use
Promulgate &train
9. Manning &rostering for different flight and route
Assign people to task
Monitor management functions
Monitor on-line performance &self feedback
Monitor/ Evaluate
10.Monitor and evaluate manpower plan, route planning & fatigue level
Maintain/ Change
11.Maintain/ change

Figure 7-7 manpower planning & availability delivery system

Although the aspects of manpower planning may not at the moment be so relevant to the cockpit[37], the delivery system of manpower planning and availability in ARAMIS does cover the manning issues which deal in the aviation environment with how many replacement personnel we need for all relevant tasks. This is important in long haul flights, whilst manning levels, hiring of contractors for safety-sensitive functions and the size of work station responsibilities (boxes 5-8 in Figure 7.7) are much more important for aircraft controllers and aircraft maintenance for safety operations. In Figure 7.7 we combine fatigue risk management (block 4) mentioned above with the delivery system of manpower planning & availability in ARAMIS. This leaves the issue of workload, which is related, but cannot be solved in the short term by allocating more manpower, to be dealt with in the next section.

7.2.2.4 Workload

In the context of flight crew activities, workload can be defined as all the physical and mental effort required to fly an aircraft. It includes planning, thinking, navigation, communication,

[37] In the past, there has been considerable discussion about how many people are needed, up until the advent of automation has reduced flight crew members from three to two.

and controlling the aircraft (Stein & Rosenberg, 1983). Workload can be defined as the difference between the amount of resources demanded by the task situation and the amount of resources available by the operator to perform in the task situation. Hence, workload is not only sensitive to multiple characteristics of a task, i.e. task demand, but also to characteristics of the operator, i.e. operator capacity (Hart, 1987; Hancock et al., 1995). Operator capacity is highly influenced by the fatigue and competence (training and experience) of the flight crew and by their communication with each other. These two aspects are already represented in other nodes of the flight crew operator model in CATS. To avoid double counting only the aspects of task demand were considered in the modelling of workload in CATS.

A straightforward determinant for air traffic controller's workload could be simplified as the number of aircraft for which the controller is responsible in a specified time and sector and the complexity of the manoeuvres normally (and in emergencies) taking place in that (type of) sector. But pilot workload is a more complex issue. Factors that influence task demand are

- traffic density,
- traffic complexity,
- weather,
- possible flight delays,
- amount of information available, and
- the technical status of the aircraft.

Although each of the aspects mentioned does play a role in reality, in CATS it was difficult to clearly define and quantify all of these aspects. So the probability of an abnormal technical status of the aircraft such that the flight crew has to apply the abnormal or emergency procedures from the Aircraft Operations Manual directly was considered the proxy for task demand.

When we look back in the past projects, workload was treated under "availability" in the ARAMIS projects and I-risk project. Because both projects were dealing with the chemical industry, the workload was very much linked to either the provision of enough control room or maintenance people. In these cases, workload was determined by how much work there was compared to how many people to do it. In the context of flight crew activities there is still an element of that, but there is less possibility of solving it by adding new people, especially during the flight, although this is an option for air traffic control and aircraft maintenance. As explained above, workload in flight crew activities is far more complex than in the chemical industry so not all of what comes out of Table 7.1 can fit under availability. Therefore, those factors need to be dealt with as a separate delivery system.

Figure 7.8 shows a possible block diagram for workload. The main distinction is between normal and what we regard as abnormal factors (e.g. traffic, weather, delays) that influence flight crew task demand. The block diagram for manpower planning and availability as shown in the previous delivery system (Figure 7.7) covers the aspects of route planning, cockpit manning and rostering. This part of the influences is something that can be planned for. However, there are situations where we cannot plan for the increased crew workload - e.g. sudden flight plan change or below-minima weather conditions. So, the only way of dealing with it is to avoid it or cope with it as best we can. This is described in Figure 7.8. This involves providing information (boxes 3 & 4) about extremes and training (box 5) to enable the pilot to cope with more extremes. This system is new and has not been incorporated in any of the delivery systems in the previous project and should be covered here. As the management of availability plays a role that could potentially determine the flight crew task demand, these two aspects (Figures 7.7 and 7.8) should be handled closely together.

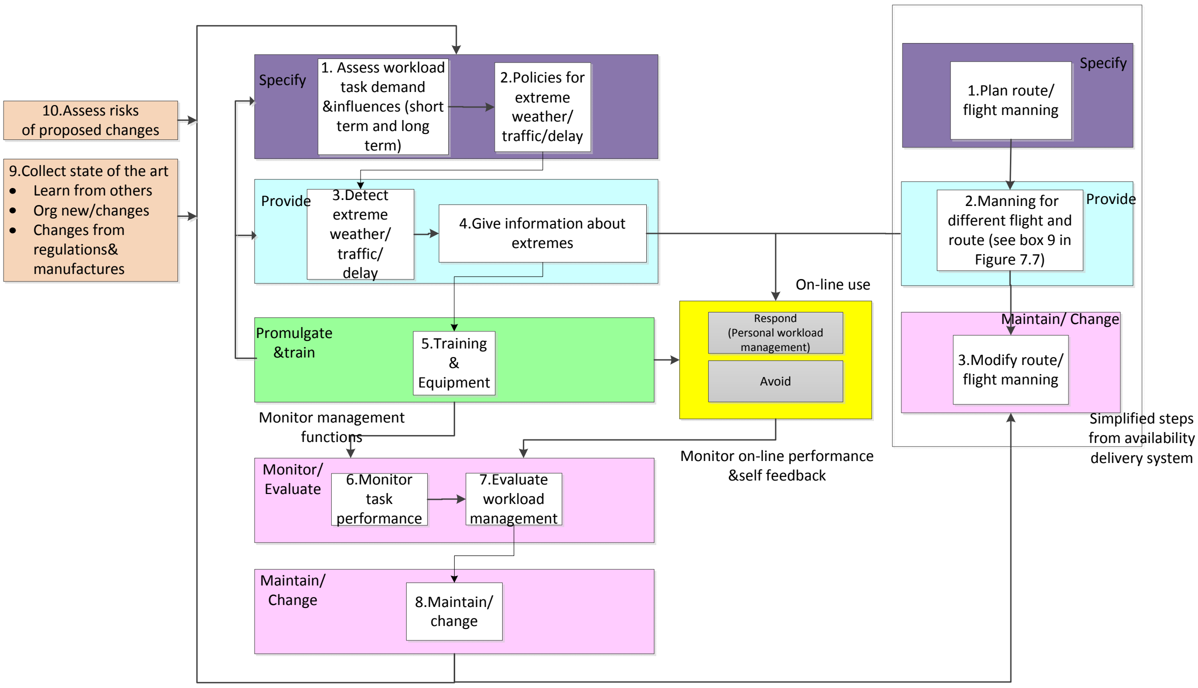

10.Assess risks of proposed changes
9.Collect state of the art
Learn from others
Org new/changes
Changes from regulations& manufactures
Specify
1. Assess workload task demand &influences (short term and long term)
2.Policies for extreme weather/ traffic/delay
Provide
3.Detect extreme weather/ traffic/ delay
4.Give information about extremes
On-line use
Promulgate &train
5.Training & Equipment
Respond (Personal workload management)
Avoid
Monitor management functions
Monitor on-line performance &self feedback
Monitor/ Evaluate
6.Monitor task performance
7.Evaluate workload management
Maintain/ Change
8.Maintain/ change
Specify
1.Plan route/ flight manning
2.Manning for different flight and route (see box 9 in Figure 7.7)
Provide
Maintain/ Change
3.Modify route/ flight manning
Simplified steps from availability delivery system

Figure 7-8 Workload delivery system

7.2.2.5 Procedures, rules, checklists, and goals

Procedures exist to cover many types of sequences and flight crew actions in normal and abnormal situations. Deficiencies in procedures or the complexity or other inadequacies of procedures may make it difficult for air crew to obtain what they want from procedures. This may lead to deviations from procedures by flight crew and increase the likelihood of online human errors. In the aviation system, procedures, rules, checklists and goals are considered to have a dominant influence on the safety of aviation operations, although the limited analysis of ADREP data in Chapter 5 does call this into question to an extent. Although it is apparent that they play an important role in aviation safety, none of the current risk models in aviation incorporates them as a PSF. In CATS, even though the use of emergency procedures was used as PSF, it was incorporated not to represent the quality of those procedures, but as a proxy for workload. As explained in the second experiment in Chapter 6, the difficulty to incorporate procedures into risk modelling was that the dimensions relevant to the procedures were relatively qualitative and soft (e.g. quality, accessibility, usability, and effectiveness) so that they are considered difficult to model and quantify in numerical units. In addition, there are many different procedures to be considered. There are the airport’s arrival and departure procedures that are different for each runway at each airport. There are also different procedures between Boeing and Airbus, or between KLM and BA. Because of these many procedures, it is practically impossible to assess each procedure on different characteristics like readability, clarity, depth of procedures, etc. Therefore some priority needs to be established as to which procedures are most critical.

For each aircraft type the aircraft manufacturer will write procedures for normal and abnormal tasks. We could therefore try to look at groupings of procedures and consider general differences between for instance abnormal and emergency procedure (A/E procedure) and standard operating procedures (SOPs, including normal checklists). In most cases, an airline will base its own procedures on those of the manufacturer, but can deviate, for instance to be in accordance with the standard operating procedures that apply to the airline’s own philosophy across all the aircraft types. For A/E procedures this deviation is usually very small, while for SOPs it may be greater, because there are also other issues related to the structure of the operational environment which could prompt procedure changes by the airlines. These factors include noise abatement or the effect of short and long haul operation (Degani, 1994). In developing the normal procedures, airlines tend to address more operational environmental factors surrounding the aircraft so that the process of conducting the normal tasks by the pilots is more efficient.

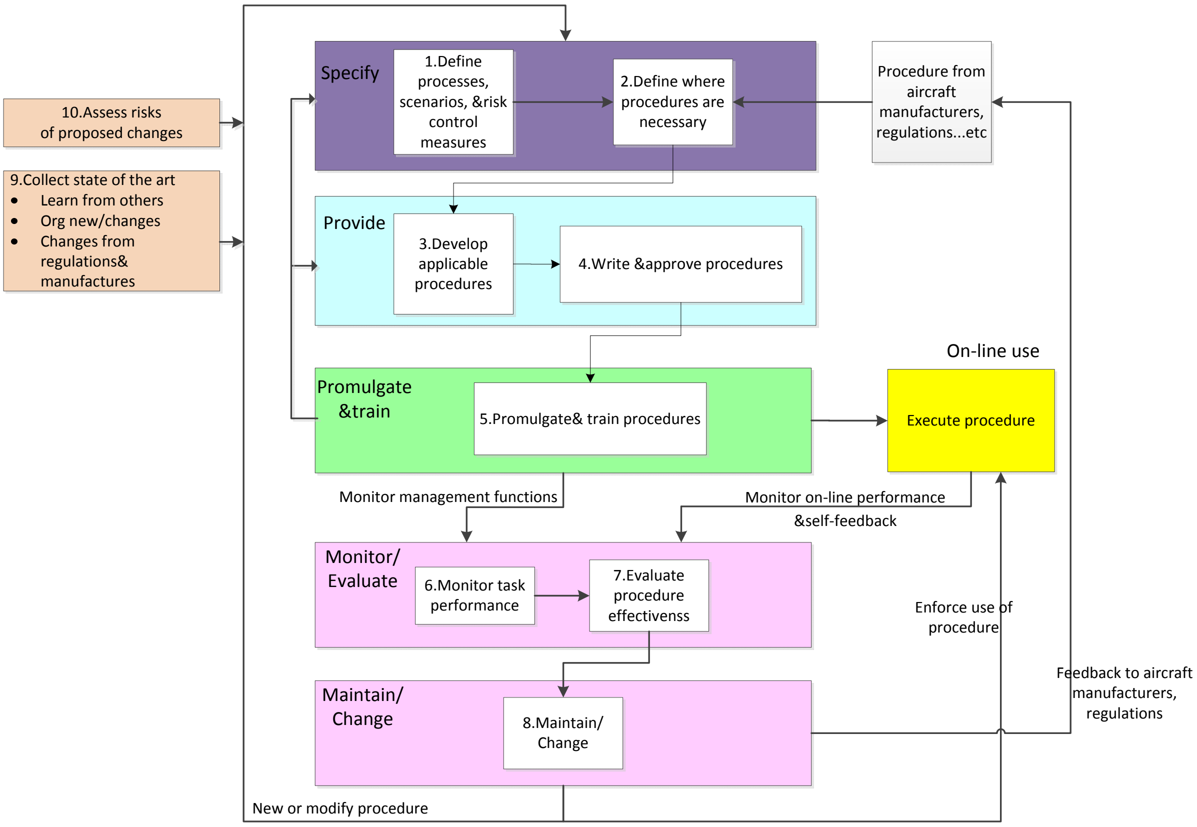

10.Assess risks of proposed changes
9.Collect state of the art
Learn from others
Org new/changes
Changes from regulations& manufactures
Specify
1.Define processes, scenarios, &risk control measures
2.Define where procedures are necessary
Procedure from aircraft manufacturers, regulations...etc
Provide
3.Develop applicable procedures
4.Write &approve procedures
On-line use
Promulgate &train
5.Promulgate& train procedures
Execute procedure
Monitor management functions
Monitor on-line performance &self-feedback
Monitor/ Evaluate
6.Monitor task performance
7.Evaluate procedure effectivenss
Enforce use of procedure
Feedback to aircraft manufacturers, regulations
Maintain/ Change
8.Maintain/ Change
New or modify procedure

Figure 7-9 Procedure delivery system

In the paired comparison exercise reported in Chapter 6, we focused on the effectiveness of abnormal/emergency procedure to avoid the complexity usually reflected in normal procedures. The management influences which we derived for abnormal/emergency procedure in Chapter 6 came from the existing delivery system block diagram from ARAMIS, which seemed to work quite well. Therefore the current block diagram does not need significant changes for the management of procedures, only the clarification provided by translating the steps into specify, provide, promulgate/train, use and monitor/maintain (see Figure 7.9 for the new block diagram). However, SOPs also need to be considered as a PSF in the complete model, using the same approach. So, it is just a question of repeating the quantification process in Chapter 6 with the other types of procedures which are relevant. But, in developing the normal procedures we expect that when generating relevant management actions for normal procedure, the quality of the translation process by airlines (to adapt procedure to the specific company need and style, or give clear prioritisation of the critical tasks in each procedure) will play a more essential part in procedure development than it did for A/E procedures.

7.2.2.6 Communication and coordination

Communication refers to exchange of information and instructions between people within the steps of a primary business activity. Communication occurs either verbally (face-to-face, or through communication channels such as radio or satellite/mobile telephone) or non-verbally (gestures, or by passive written messages emanating from data link, e-mail, memo, briefing). Coordination covers those mechanisms designed to ensure the smooth interaction of actions between individuals and groups working on a joint safety critical task or responsible for the correct functioning of a given risk control measure.

Communication and coordination within flight crews & communication and coordination between flight crews and cabin crews

Flight crew communication and coordination is an essential prerequisite for safe flight. Without good *interpersonal skills* the flight crew cannot work together as a team. This includes clear and concise face-to-face communication during briefing, crosscheck and verbal confirmation. The flight crews not only need to develop such communication skills, but also a range of *behaviour associated with teamwork*. They also need to develop a *shared mental model of the problem solving* in the course of the flight in order to allow problem solving to be shared and allocated among crew members. Since nowadays the cockpit doors have to be closed and locked during flight, it makes communication between flight deck and cabin crew more difficult. For instance, lack of communication between cabin crew and flight crew of a flight of Emerald Airways in 2005 nearly caused the aircraft to crash (AAIB, 2007). Shortly after the aircraft takeoff, a hydraulic connection associated with the forward left door stairs sprang a leak and caused the forward part of the passenger cabin to fill with hydraulic fluid mist. The cabin crew wrongly diagnosed the mist as “smoke”, and moved a number of passengers to seats towards the rear of the cabin without inform the flight crew. This caused the aircraft's center of gravity to be dangerously altered under the weight of people moving to the back. Although the aircraft managed to make a safe landing, it caused minor injuries to thirty-three passengers and four crew members. Hence, it is clear that in aviation communication and coordination are not confined to intra-cockpit, but also related to those with cabin crew to enhance communication and coordination and avoid confusion. This is also relevant in the case of terrorist action, air rage or medical emergency.

A steep *trans-cockpit authority gradient* may exist when the captain has much more experience than the first officer and/or has a personality or culture that emphasises hierarchy. The most notorious accident illustrating the danger of steep authority gradient was the 747 disaster in Tenerife in 1977. The KLM airplane did not have clearance to take off while a Pan Am airplane was still on the runway. The KLM captain advanced the throttles and the co-pilot quickly advised the captain that ATC clearance had not yet been given. Despite the warning from the co-pilot, the captain began the takeoff roll and the subordinate co-pilot this time did not have the confidence to make a second attempt to stop him. The two airplanes collided on the runway killing 583 people. After that, the steep hierarchical gradient between crew members was strongly discouraged. More emphasis was placed on *decision making in the team* by Crew Resource Management (CRM), to prevent or minimize communication and coordination errors. Training in this therefore also belongs here.

In CATS, only the intra-cockpit communication was modelled, and then only one small aspect, namely the native language (i.e. difference in mother tongue) of the captain and the first officer which was intended to represent the language compatibility and communication between captain and first officer.

Communication and coordination from pilot to ATC and from pilot to operation centre
Communication is not only important within the cockpit but also from the cockpit to ATCs and flight operations. Much of the communication (e.g. callouts, callback, etc.) between cockpit and air traffic controllers is done with the use of verbal communication through *communication channels* such as radio or satellite/mobile telephone. In order to have communication, both the ATCs and the pilots must share a *common code* (e.g. standard phrases, terminology) and *speak the same language*, so that the meaning or information contained in the message can be interpreted without error. As the sender of a message, it is vital for one to expect some kind of *response* from the person one is communicating with. In aviation a considered and detailed reply with a read-back of the key parts of the instruction to show mutual understanding is more important than just simple acknowledgment that his message has been received with a colloquial phrase such as "OK" or “received”. Another aspect of the channel of communication is the medium used to convey the message, so the quality of the communication equipment (e.g. communications equipment malfunctions) and busyness on the communication frequency (e.g. message can be blocked by mutual interference on the radio frequency which causes a whistling sound and then is inaudible to the flight crew) are also vital during the communication.

Communication and coordination apply equally to ATC operations. Air traffic controllers need to coordinate their activities with those of other controllers in their own organisation as well as controllers in other organisations, e.g. when an aircraft is handed over from one sector to another. There is also a vital communication process during shift changeovers. Much of the communication between air traffic controllers is done with the use of “flight strips”. Each strip contains information on a particular aircraft and is passed over from one controller to another. Voice communication among air traffic controllers is also relevant. When two controllers are located in the same room (e.g. the control tower) this voice communication will usually be done without additional equipment, but when two controllers are located in different rooms, some type of equipment such as telephone, radio, fax or a combination will have to be used for voice communication. Because of the low ambient noise levels in control rooms, voice communication among air traffic controllers located in the same room is not considered a problem. However, when some types of equipment are used to communicate there is a possibility of disturbance of the signal. Moreover, there are also issues such as task allocation and backup when one controller gets overloaded. These should be handled by the supervisor shifting work from one controller to the other. Communication and coordination within ATCs is a very important area, almost warranting a separate section. However, it is not appropriate to go into too much detail in this section since the focus of this project has been on flight crew. But it should be noted that CRM training should also apply to intra-ATC communication and coordination.

A flight dispatcher is a person responsible for planning and monitoring the progress of an aircraft journey in the operations centre. In contrast to verbal communication with ATCs or other flight crew, communication between dispatchers is primarily based on written information. Ground-air communications has increased as the advent of the data link technology (e.g. development of the satellite communication) and associated procedures has gradually taken place. Currently it is much easier to contact the operation center than it used to be. Pilots do not need to rely solely on long range radio transmission, but can get much more on-line information from the operations center. The operations center, for instance, can provide flight crew with a package for briefing with technical information about the route and weather information before the flight. The operations center can also decide whether the flight should be diverted, delayed, or canceled in respect to safety. Moreover, with consistent contact with the operations center, dispatchers can provide information about on-route weather and advise the flight crew of any circumstances that might affect flight safety. In the case of emergency (e.g. a technical failure), the pilot can also call the operations centre for technical assistance. Thus, both non-verbal and verbal information are different ways by which pilots and dispatchers communicate with each other.

Figure 7.10 shows the communication and coordination delivery system revised from ARAMIS block diagram. The crew management training issues were considered to fit better under this delivery system than under competence. It should be noted that this protocol links closely to the protocol on the management of hardware and software of the communication channels (block 5). It also links to the protocol on procedures and rules (block 4), since many communication processes are, or should be, formalised and subject to procedures.

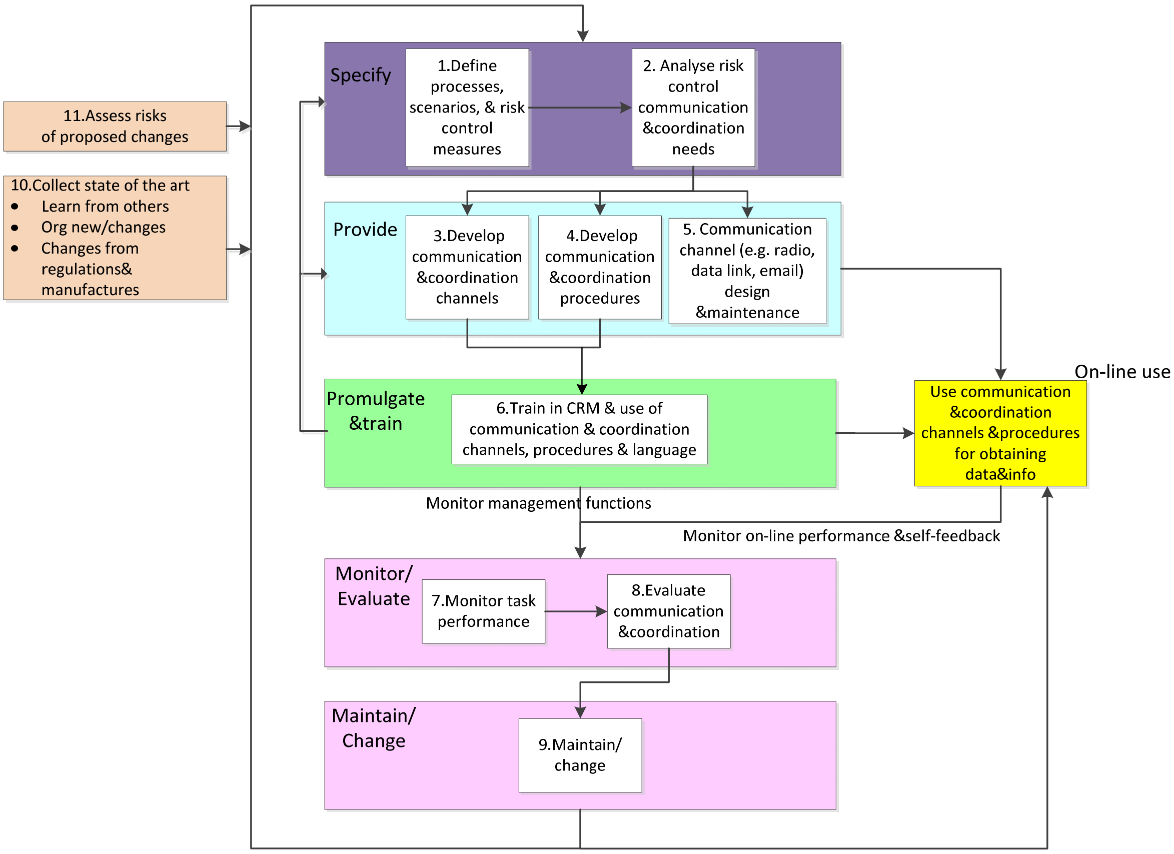

11.Assess risks of proposed changes
10.Collect state of the art
Learn from others
Org new/changes
Changes from regulations& manufactures
Specify
1.Define processes, scenarios, & risk control measures
2. Analyse risk control communication &coordination needs
Provide
3.Develop communication &coordination channels
4.Develop communication &coordination procedures
5. Communication channel (e.g. radio, data link, email) design &maintenance
Promulgate &train
6.Train in CRM & use of communication & coordination channels, procedures & language
On-line use
Use communication &coordination channels &procedures for obtaining data&info
Monitor management functions
Monitor on-line performance &self-feedback
Monitor/ Evaluate
7.Monitor task performance
8.Evaluate communication &coordination
Maintain/ Change
9.Maintain/ change

Figure 7-10 Communication and coordination delivery system

7.2.2.7 Man-machine interface

In the case of flight crew, instruments and workplace design should include the ergonomic quality of the human system interface in the flight deck and the clarity of instrumentation. The quality of the interface between machine (the aircraft or ATC equipment) and its human operator (the flight crew or ATC) has greatly improved over the years. The proxy variable used for the support given by the interface to decision making by pilots in CATS was the aircraft generation. A better man machine interface design has been taking workload away from the flight crew as each generation passes. So in principle “aircraft generation” was a good proxy in CATS. This can be illustrated by comparing the cockpit of a first generation commercial jet transport aircraft like the De Havilland DH-106 Comet, with that of a modern jet airliner like the Boeing 777. When considering the effect of technological advances on the safety of air transport it is common to consider four different generations of aircraft since the introduction of the jet engine. First generation aircraft were typically designed in the 1950s. Most of the aircraft were certified before 1965 according to British Civil Airworthiness Requirements (BCAR's) or other certification bases. Jet engines were still very new, and the aircraft had very limited cockpit automation, simple navigational aids and limited approach equipment. Examples are the DH Comet, Fokker F-27 and Boeing 707. Second generation aircraft, designed in the 1960s and 1970s, have more reliable engines. The aircraft were certified between 1965 and 1980, but not yet based on common JAR-25/FAR-25 rules. Cockpit equipment was more advanced, with better auto pilots, auto throttles, flight directors and better navigational aids. Third generation aircraft (e.g. Fokker 50 and Boeing 737-700), designed in the 1980s and 1990s, typically show consideration for human factor aspects in the cockpit. Electronic Flight Instrument Systems (EFIS) and improved auto pilots are used. Furthermore, the aircraft are equipped with ACMS data systems and high-by-pass engines designed according to higher certification standards. Fourth generation aircraft like the Airbus A 320 and Boeing 777 have full glass cockpits and digital fly-by-wire systems. Those different aircraft generations provide a convenient classification for the quality of the man-machine interface in CATS (Ale et al., 2009). Research has shown that the probability of flight crew error is significantly reducing for subsequent aircraft generations (Roelen and Wever, 2002).

However, competence has gradually moved from direct flying (how to fly) to supervisory control (how to operate the system) as we have moved towards modern aircraft generations. So knowledge and skills to fly can be different between different aircraft generations. Since working with equipment is also a skill to be learned, there is a link between man-machine interface and competence and suitability.

As mentioned above, the proxy variable (aircraft generation) used in CATS was quite suitable to cluster many technology changes. But what it does not offer is the possibilities of looking at the effect of changes in the design process or in the changes in the way in which the interface is maintained. So, aircraft generation does not offer a formulation in terms of these two delivery systems and of other decisions that can be made apart from just buying new aircraft. Hence, in the future if we want to model the effects of the design process or the influences of maintenance improvement, it would be necessary not to use this proxy but to expand it. Although we have not looked in detail at these two delivery systems in this thesis, the delivery systems developed for ARAMIS (see Figure 7.10 and Figure 7.11) would be the starting point. It should be noted that the two delivery system block diagrams with the design and the maintenance functions deal with both technical functioning and MMI, and the steps in them can also be classified according to the "specify" to "maintain/change" headings in our generic structure. The translation into that format would be work for the future.

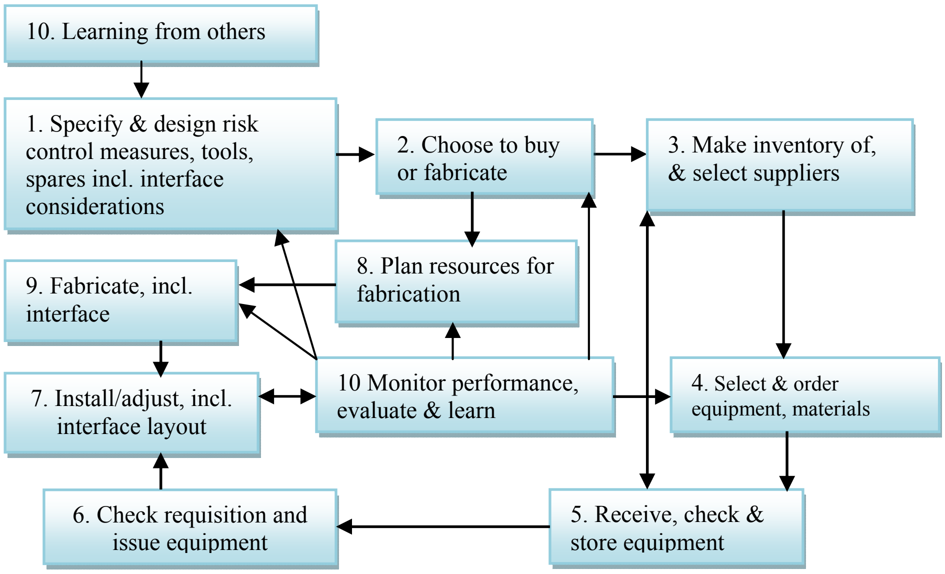

10. Learning from others
1. Specify & design risk control measures, tools, spares incl. interface considerations
2. Choose to buy or fabricate
3. Make inventory of, & select suppliers
8. Plan resources for fabrication
9. Fabricate, incl. interface
10 Monitor performance, evaluate & learn
4. Select & order equipment, materials
7. Install/adjust, incl. interface layout
6. Check requisition and issue equipment
5. Receive, check & store equipment

Figure 7-11 Delivery system of equipment and interface design, purchase, construction, installation and adjustment

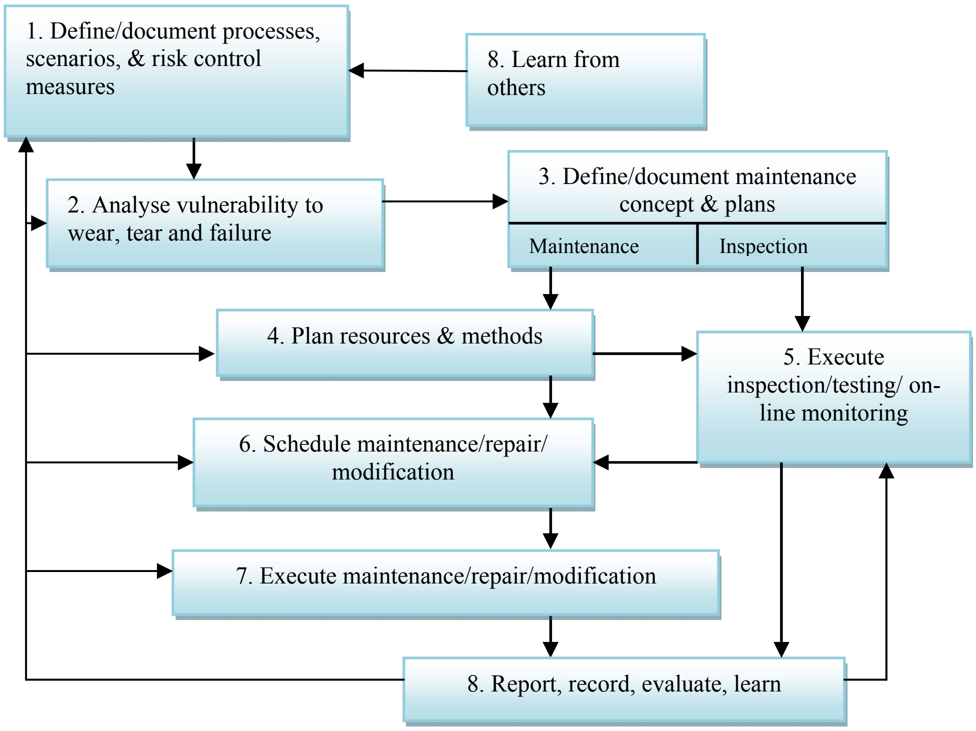

1. Define/document processes, scenarios, & risk control measures
8. Learn from others
2. Analyse vulnerability to wear, tear and failure
3. Define/document maintenance concept & plans
Maintenance
Inspection
4. Plan resources & methods
5. Execute inspection/testing/ on-line monitoring
6. Schedule maintenance/repair/ modification
7. Execute maintenance/repair/modification
8. Report, record, evaluate, learn

Figure 7-12 Delivery system of inspection, testing and maintenance of hardware and interface

7.2.2.8 Commitment to safety

As stated in Section 3.3.1, two influencing factors ("commitment to safety" and "procedure") were initially considered important in the beginning of the CATS project, but were completely left out because they were considered too complicated to represent quantitatively. It was concluded in the CATS final report that these factors should be developed in a further phase of CATS. For procedure, we have taken the opportunity in one of the experiments in Chapter 6 to examine the issue of procedure use and quantify the impact of management influences on it. However, pilot commitment to safety above other personal and organisational goals, has long been a difficult PSF to model properly in risk assessment. Therefore, we will discuss this issue more thoroughly in this section to understand what are the factors influencing commitment and motivation, and particularly on how this factor can be combined in a risk model to influence work behaviour in safety.

Commitment and motivation theory

Commitment or motivation to safety has been a difficult PSF to model properly, in part because they are both difficult concepts to properly define and partly because there are many complex and multidimensional social and psychological factors relating to them. Pinder (1998) and Meyer & Herscovitch (2001) provided definitions that nicely accommodate the different theoretical perspectives in the explanation of work motivation and commitment:

- *Work motivation* is a set of energetic forces that originates both within as well as beyond an individual's being, to initiate work-related behavior, and to determine its form, direction, intensity, and duration (Pinder, 1998).
- *Commitment* is a force that binds an individual to a course of action that is of relevance to a particular target (Meyer & Herscovitch, 2001).

In definition, commitment and motivation are related concepts. More specifically, it could be suggested that commitment is one component of motivation (Meyer et al., 2004).

Theories of work motivation have been widely applied to explain task performance. Many theories have been set forth to explain employee motivation (see Kanfer, 1990; Pinder, 1998). Locke's (1997) general model of the motivation process is perhaps one of the most comprehensive and complete up to date (c.f. Meyer et al., 2004). The causal connections it proposes are well supported by empirical evidence (Locke, 1997; Pinder, 1998). However, motivation researchers seldom address commitment as an energizing force for motivated behaviour. For this reason, Meyer et al. (2004) incorporate Meyer and Herscovitch's (2001) model of workplace commitment as part of Locke's general motivation process. With the integration of the motivation and commitment processes, Meyer et al.'s model offers an excellent foundation for detailing the PSF for commitment and combining it into our model. We present a simplified depiction for the motivation and commitment process as described by Meyer et al. (2004) in Figure 7.13.

Goal setting is a powerful way of motivating people. The value of goal setting is well recognized so that management systems have goal setting basics incorporated within them. Goals can be self-generated or assigned by the employers, but we presume that all consciously motivated behaviour is goal-oriented. Goals derive from basic human *needs*, *personal values*, *personality*, *incentives*, and *self-efficacy*[38]. *Goal regulation* mediates the effects of the antecedents of needs, values/personality, incentives and self-efficacy. It is defined as different psychological states or underlying mindsets reflecting the reasons and purpose for a course of action. There are two important components of goal regulation, *perceived locus of causality* and *perceived purpose*. The term *perceived locus of causality* refers to a person's beliefs about why he or she is pursuing a particular goal. This can vary

[38] Self-efficacy is a person's belief about their capabilities of achieving designated levels of performance shaped through experience and socialization. For example, a person with high self-efficacy may engage in a more active safety-related activity when a safety challenge occurs, whereas a person with low self-efficacy may succumb and have no confidence to make any attempt to query with ATC when he is not certain about a given instruction (e.g. take-off clearance)

from internally driven[39] to externally driven[40]. *Perceived purpose* refers to someone's general purpose in trying to approach a given desired end-state. The model identifies two purposes either a *promotion focus* to accomplish goals or "make gains", or a *prevention focus* based on fulfilling obligations and responsibilities and "avoiding losses". As discussed by regulatory theory (Higgins, 1997, 1998), these two purposes influence a person in the decision-making process, and determine the different ways they achieve their goal.

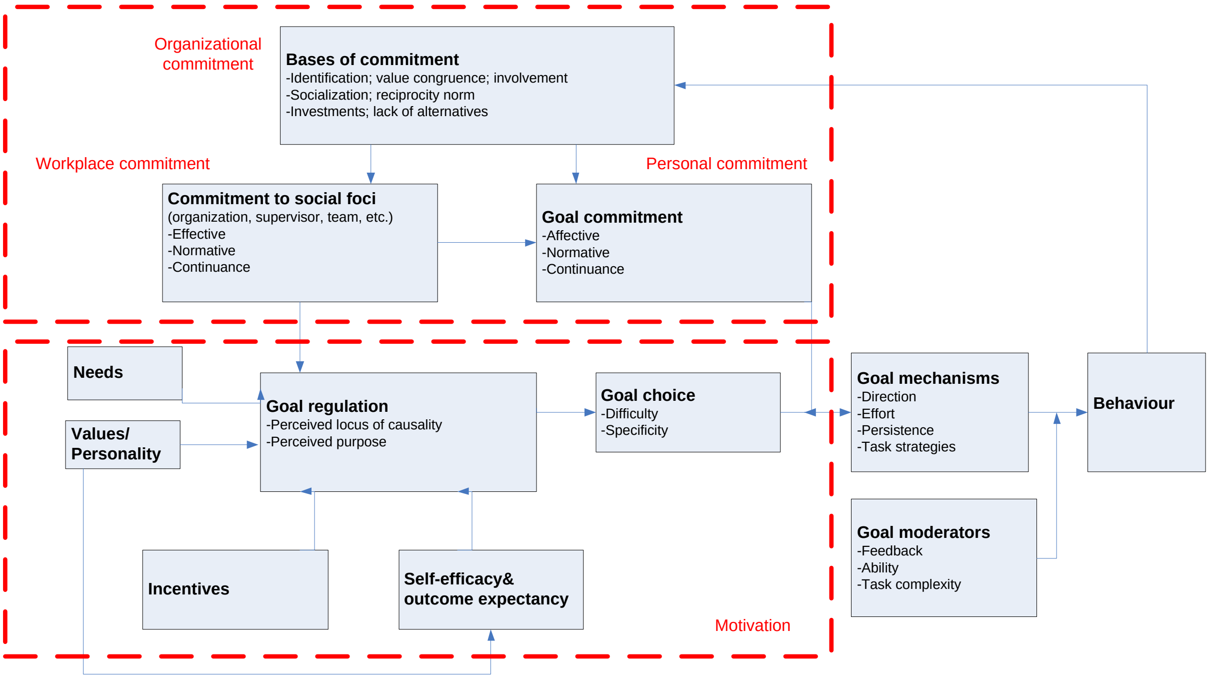


Figure 7-13 A simplified integrated model of employee commitment and motivation (adapted from Meyer et al., 2004)

The goals that individuals choose can differ in *goal difficulty* and *goal specificity*. Goal difficulty is the level of difficulty for employee to achieve the goal. Goal specificity means the goal should be relatively clear and precise in its target. The dashed box at the bottom of Figure 7.13 shows how complex employee motivation is.

As stated, commitment is seen in Meyer's model as a force that binds an individual to a course of action that is of relevance to a safety goal. The binding nature of commitment makes it somewhat distinct from motivation because self-commitment is generally reserved for important actions or decisions that have relatively long term implications (Meyer et al., 2004). In combination with the degree of goal difficulty and goal specificity, the individual determines the direction of behaviour, the intensity of effort put in to it, the degree of persistence to pursue a goal, and the possibility that the individual will develop strategies to achieve a particular goal (see *goal mechanism*). This variable describes the cognitive function of decision making. The strength of the moderating effects from goal commitment on the *goal mechanisms* will be greatest if an employee has a strong affective attachment to the organization or to the workplace or the work team. Therefore, Meyer et al. (2004) highlight the importance of organizational commitment (*bases of commitment*) and group commitment

[39] This refers to intrinsic, integrated, and identified regulation.

[40] This refers to external and introjected regulation.

(*commitment to social foci*, i.e. team, supervisor, etc.) which play important roles in the motivation process of an individual (see the red dash box in on the top of Figure 7-13).

Finally, in order for a goal to be successful, Locke's theory states that there must be a set of *moderators* that are necessary for goal accomplishment: *feedback*, *ability*, and *task complexity*. Goal setting cannot be effective if individuals cannot check the state of their performance in relation to their goal. In aviation, this *feedback* can be achieved in a relatively short time dynamic by e.g. analysing and feeding back information from the flight data recorders or feedback by line check pilots. *Ability* refers to experience and skills that one attains. If one is not competent enough, no chance exists for one to reach a goal. This aspect is dealt with in the "competence delivery systems" (Section 7.2.2.1). Lastly, *task complexity* also moderates the effects of goals because more complex goals require the review of more complex strategies than lower difficulty goals. The amount of work a flight crew has to accomplish in the time available, along with their overall sense of being pressured, may alter their decisions to achieve their safety goal. Whether the employee's *behaviour* matches the safety goal could be observable in the real world.

It should be noted that it is consistently expected that flight crew follow procedures (which are dealt with under the procedure delivery system) in situations which have been anticipated and for which procedures are appropriate. Violations[41] are observable human behaviours which represent a deliberate deviation from the procedures and rules that govern safe flight. Violations can be divided into routine violations and exceptional violations. Routine violations tend to be habitual by nature, such as pilot routinely flying into marginal weather. This is often tolerated by authority (Reason, 1990). On the other hand, exceptional violations are rare occasions that take place in particular circumstances (e.g. equipment failures, emergency) that make violations an instinctive reaction to the situation or a conscious decision to violate, because the existing procedures are not seen as relevant. Not all violations, observed in the "behavioural" block in Figure 7.13, are "bad". If there is a situation which occurs unexpectedly and cannot be handled by existing procedures, a pilot is correct to violate that procedure. This perspective is not included in Locke's model. Besides, it should be noted that pilots must have the *ability* to deviate, especially in an emergency situation, when the procedures do not apply or do not cover the situations that pilots are facing. Aviation operations take place in a dynamic and tightly coupled environment. We cannot assume that in all circumstance, if the procedure is followed, then the result will be valid. So in an emergency, such justified deviation must be accepted. Pilots should be, for instance, permitted to deviate from a low priority procedure (e.g. after take-off check lists) when this conflicts with the high workload induced by an abnormal condition (e.g. engine on fire).

Moreover, it should also be noted that commitment to safety is not the same as commitment to work, which can lead to organisational goals such as productivity, punctuality, or economy being preferred above safety goals. Commitments to these different targets have the potential to conflict with each other. Since "commitment to safety" is quite different to "commitment to work" where employee motivation and commitment theories have most applied, therefore the last two points need to be taken into account in Meyer et al.'s (2004) model in Figure 7.13 for safety applications.

[41] A violation is a planning failure where a deliberate decision to act against a rule or plan has been made. Mistakes, slips and lapses are not considered as deliberate decisions against rule. They should be dealt with by training, covered in the competence delivery system.

PSFs for commitment and motivation

Having described the model variables, each variable can be transferred into a PSF and incorporated into our current qualitative human performance model as depicted in Figure 7.14. Human performance can be modelled as influenced by the internal goal mechanisms along with the other factors discussed in Meyer et al. The qualitative model in Figure 7.14 introduces one modification to Meyer et al. (2004) model: to abandon the box of *goal moderator* because we believe that in our new model we already incorporate *ability* (see discussion in section 7.2.2.1 competence and section 7.2.2.2 suitability) and *task complexity* (see discussion in section 7.2.2.4 workload) that affect human behaviour and performance. There is also a *feedback* loop imbedded in each of the delivery systems. Since all of these factors in Meyer's "goal moderators" are taken care of under the other PSFs in our model, we have eliminated "goal moderators" from the commitment PSFs. However, if we want to quantify the human performance model in Figure 7-14 using BBNs, the feedback from behaviour to the bases of commitment in Meyer's model (which identifies the fact that pursuing a course of action can strengthen commitment to that course of action) is not allowed. This is due to the modelling limitation of constructing feedback loops in the BBNs.

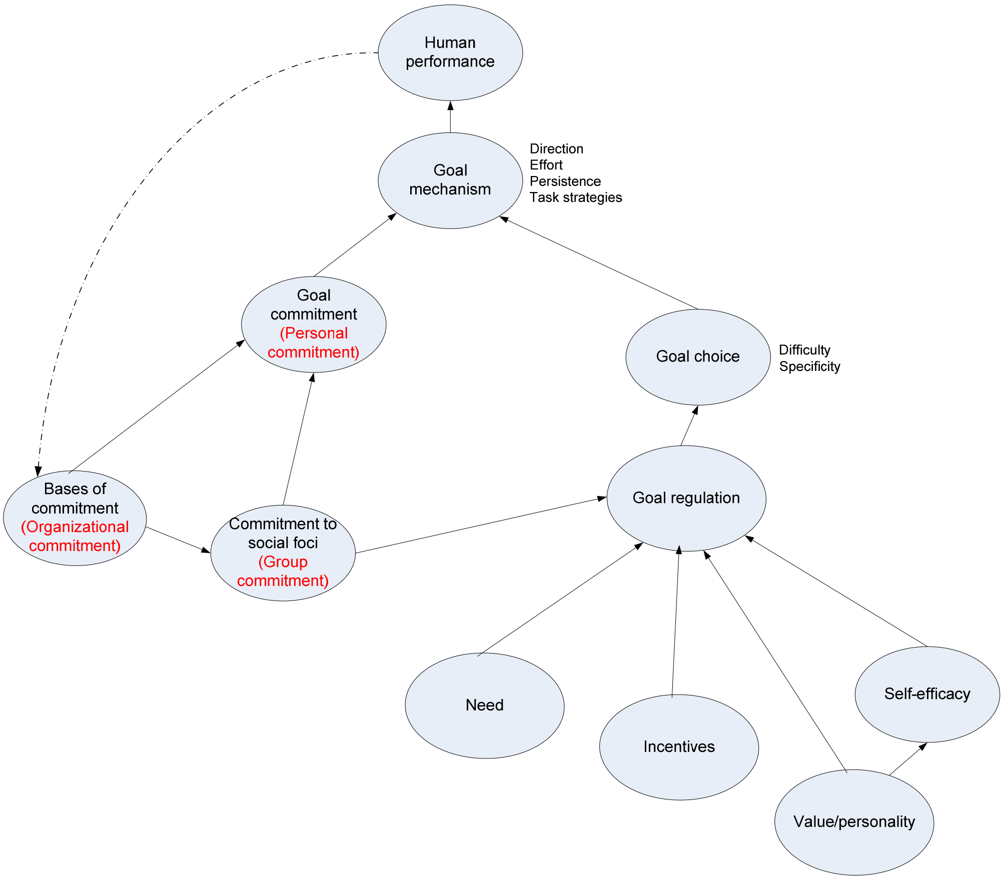


Figure 7-14 PSFs of pilot commitment and motivation

We hypothesise that the model in Figure 7.14 is a good point to start for risk modelling as it covers a lot of PSFs relating to what people consider as relevant to commitment, motivation, and safety culture (see next paragraphs for more discussion on safety culture and safety climate) in the field. But this figure raises two issues:

- First is that it is not clear yet over which PSFs in that model management have influence. When we go through the analysis of safety culture in the next paragraphs, we will examine whether the rest of the items in this figure can find a "management" home.

- The second problem is that a large number of factors involved in Figure 7.14 are still generic. There is certainly a need to operationalize each of the nodes in that diagram to facilitate the quantification. To model what the various circles might contain in practice, a number of examples in Table 7.1 (the long list of PSFs) and more examples listed in the ADREP data in the Appendix which link to commitment and motivation could be fitted into each of the nodes and make the practical implication of that theoretical model.

Safety culture and safety climate

The pilot's decision to choose one from several possible courses of action and decide to commit to safety above other personal and organizational goals is certainly influenced by his personal safety attitudes, but there is also a strong organizational aspect to these influences. Safety culture and safety climate are considered important in this respect also because many researchers believe that both constructs influence personal safety attitudes to some degree and both constructs can predict unsafe behaviour and accidents. Safety climate is a multidimensional construct that covers a wide range of individual attitudes towards, and assessments of the work environment. Safety culture refers to the more fundamental underlying beliefs and values of a group of people in relation to risk and safety.

Flin et al. (2000) indicated that the number of factors making up the safety climate varied from 2 to 19 in the studies they reviewed. Therefore, it is not surprising to see that when involved in the discussions with the experts from the CATS program, there was little progress when we asked questions about: what do you mean by safety culture? How do you recognize it? What is its significance? We repeatedly received only unclear and ill-defined terms which could not be made concrete at all in such a way that they could be handled in the model and measured in practice. In consequence, this area did not produce any fruitful results for risk modelling.

Nevertheless, we argue that if we were to zoom in much more on the cultural content and specify the "cultural" influences on each of the delivery systems, we would be able to represent "culture" influences in risk modelling at least to some extent. To illustrate, recall Flin et al.'s (2000) review of the literature on climate surveys. Although they found the number of factors varied between 2 and 19 in the studies they review, Flin et al. (2000) reported several common themes that are often mentioned in literature. These are attitudes towards and beliefs about:

1) Management/supervision,
2) Safety system,
3) Risk,
4) Work pressure,
5) Competence, and
6) Procedure/rules.

"Management/supervision" in Flin's research refers to management and supervisors' attitudes and behaviour in relation to safety, e.g. management commitment, management support, management attitudes, management activity. This part would appear to be covered by the "organizational commitment" which is indicated in Figure 7.14.

"Safety system" encompasses many different aspects of the organisation's safety management system, including safety statement, safety officials, safety committees, safety policies, and safety equipment. These include specific safety barriers and safety equipment which should be considered as risk control measures not management influences. Safety statements and safety policies are procedures; safety officials, safety committees are actors for functions and not management functions.

"Risk" refers to a number of conceptual issues, namely, self-reported risk taking, perceptions of risk/hazards on the worksite and attitudes towards risk and safety. This is covered to a great extent by the combination of commitment and competence[42].

Finally, "work pressure"[43], competence"[44], and "procedures/rules" are already well defined in our delivery systems scheme.

On the basis of the mapping of the most common "cultural" themes in Flin's studies into our safety management model, we propose that safety cultural aspects referred to in current research can be reflected in the strength of the safety management system, namely the attitudes of those in the organization to those different aspects of the safety management system. This is expressed through the quality and operation of our delivery systems. In other words, safety culture (in the current studies) can be seen as a measure of how seriously a company takes its own management process. Kennedy and Kirwan (1998) noted that safety management is considered to be a manifestation of the overall culture. Mearns et al. (2003) also state that safety management practice is an indicator of the safety culture of upper management. Therefore, in the light of the obvious connection between safety culture and safety management, we argue that more favorable safety management practices are expected to result in an improved safety climate of the general workforce, and vice versa. This means that safety culture is largely dealt with already, at least in principle, in our safety management model.

However, we still can put a question mark on whether we do need culture as a separate delivery system and which PSFs in our detailed human performance model of "commitment and motivation" (Figure 7-14) have management influences, if we take a closer look at the parent nodes (i.e. "need", "incentives", "value/personality", "self-efficacy", "organizational commitment" and "group commitment") in that figure. "Need" and "incentives" are factors that can be coupled with a company's goals influenced by management influences. According to Meyer's theory these influences belong to "work motivation", which we have not covered up to now anywhere in our delivery systems. So, we need to cover these factors under our delivery systems for commitment and motivation. "Personality" is something which can be covered in suitability, whereas "self-efficacy" is influenced by personal experience and is part of competence. We also need to add "organizational commitment" and "group commitment" to the attitudes and beliefs we cover in our delivery system for "commitment and motivation". In some cases, conflicting pressures at the shop floor are created in the management hierarchy, which places incompatible demands on front-line personnel in their roles as risk control measures. These pressures often come from conflicts in the commitment of more senior managers in the hierarchy and how managers treat and motivate their online staff to stick to or violate the rule of safety. These too needs to be covered by that delivery system.

[42] The perception of risk is competence. The self-reported risk taking is a manifestation of commitment, as are the attitudes toward risk and safety.

[43] Work pressure (Flin et al., 2000) largely relates to workload and work pace

[44] Competence (Flin et al., 2000) relates to management selection, training, competence standards and their assessment

We end up with a complex of items, represented in Figure 7.14 which we need to collapse into a manageable delivery system in order to express the management influences which can be deployed to change commitment and motivation. These are represented in the delivery system in Figure 7.15.

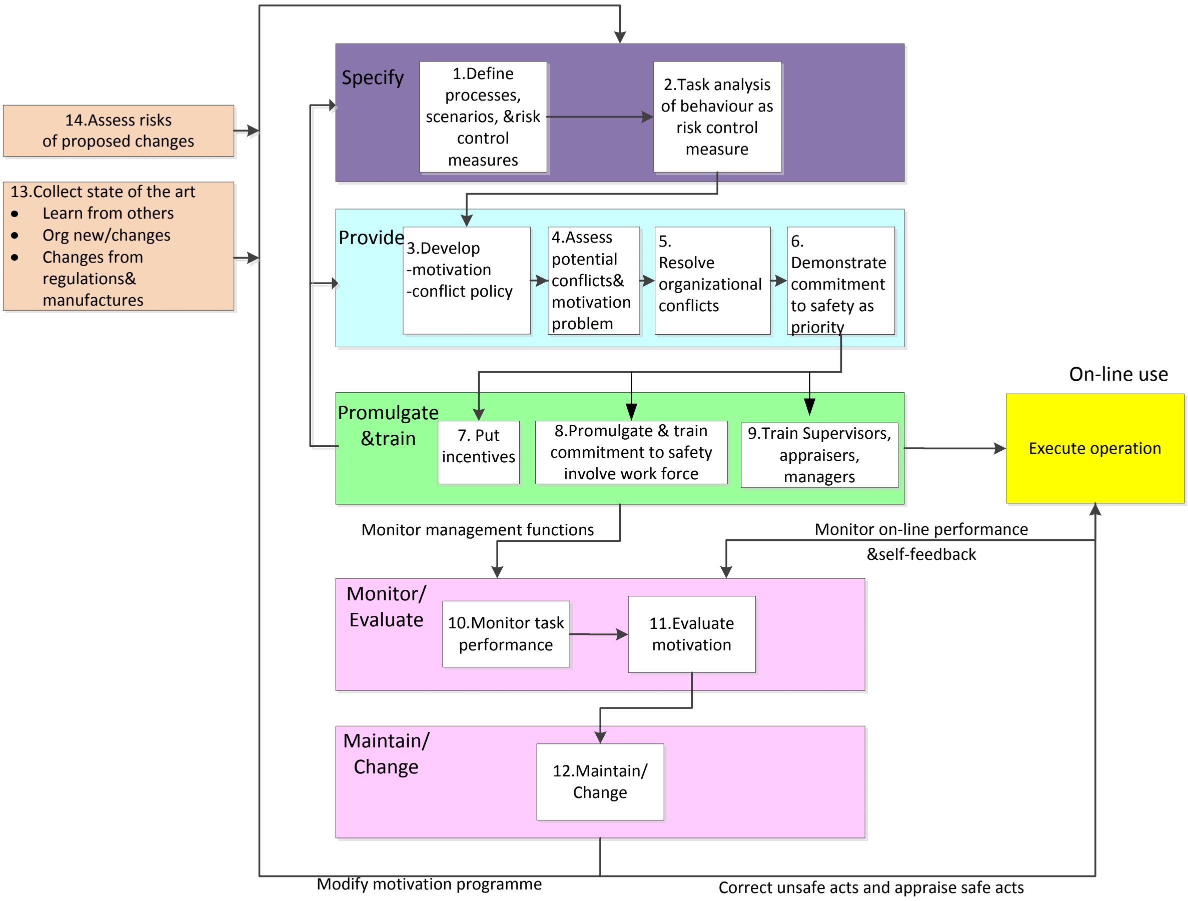

14.Assess risks of proposed changes
13.Collect state of the art
Learn from others
Org new/changes
Changes from regulations& manufactures
Specify
1.Define processes, scenarios, &risk control measures
2.Task analysis of behaviour as risk control measure
Provide
3.Develop -motivation -conflict policy
4.Assess potential conflicts& motivation problem
5. Resolve organizational conflicts
6. Demonstrate commitment to safety as priority
Promulgate &train
7. Put incentives
8.Promulgate & train commitment to safety involve work force
9.Train Supervisors, appraisers, managers
On-line use
Execute operation
Monitor management functions
Monitor on-line performance &self-feedback
Monitor/ Evaluate
10.Monitor task performance
11.Evaluate motivation
Maintain/ Change
12.Maintain/ Change
Modify motivation programme
Correct unsafe acts and appraise safe acts

Figure 7-15 Delivery system for "Commitment and Motivation"

The first 2 boxes (1 & 2) specify the policy of how to achieve commitment and assess the safety culture of the company to see if there is a suitable level of trust, participation and maturity of risk perception and understanding of how safety is achieved. This forms the basis for an analysis, taking into account all of the factors in Figure 7.14, of what all of the behaviours are which can be influenced by motivation and commitment, both in managers and operational personnel and what are the desirable behaviours under foreseeable conditions. Box 3 to 6 then provide the organisational conditions and box 7 to 9 develop the individual incentives necessary for those behaviours to be possible and desirable for those at the sharp end. Box 3 to 6 ("provide") covers many of the aspects of safety culture, including the influences from higher management which may create or resolve conflicts between safety and other goals, while box 7 to 9 ("promulgate and train") considers the incentives impinging directly on individual pilots, ATC, maintenance fitters and their immediate supervisors. In such case, "provide" steps (particularly box 6) cover the development and demonstration of the organizational commitment identified in Figure 7.14, and develops (box 7) the incentives to cater for the needs identified there. "Promulgate and train" influences the individual needs and incentives, whilst the group commitment in Figure 7.14 can be best be influenced by the process of involving the various work groups (and their reference groups) in the analysis of the motivational (safety cultural) drivers (boxes 3 – 5) and the promulgation and training steps (box 8), backed up by a participative style inculcated in and demonstrated by the supervisors (box 9) to encourage involvement. The group commitment also emerges in the social control of the operational execution, manifest in the monitoring and evaluation processes (boxes 10 and 11).

We hope that this gives a place for all of the difficult issues we have discussed under the headings of commitment, motivation and safety culture. However, there certainly needs to be a lot more work done to come to something really definite to operationalize all of the commitment and motivation aspects.

7.3 Enhancing PSF quantification

In Chapters 5& 6, we designed a method to quantify management influences on human factors, both for quantitative units and qualitative cases. As summarized in Chapter 6, we showed that it can be fairly easy to anchor the weighting results from the paired comparisons into risk/human error in quantitative cases. To do this, we need a quantified network of a human error BBN consisting of distribution functions for the nodes and rank correlations for the arcs. As an example, which we can then use as template for discussing how to proceed with qualitative cases, we take the fatigue node. To build a distribution function for that, we have to:

1) Quantify fatigue using a unidimensional scale, in our case the Stanford Sleeping Scale, where 1 signifies “feeling active and vital; wide awake” and 7 stands for “almost in reverie; sleep onset soon; losing struggle to remain awake”. This collapses all of the aspects of fatigue into an externally measurable single dimension.
2) Determine probability densities for real-life flight crew fatigue over this Stanford Sleeping Scale to indicate what current management practices produce as a result. This was done in our case by using a field study of 12,965 samples from the Aviation Medicine Group of TNO (Simon & Valk 1993, Simons et al. 1994, Simons & Valk 1997, Simons & Valk 1998, Valk & Simons 1996).
3) The rank correlation for the arc (the influence of fatigue on risk and the interrelations between the fatigue node and the other human factors) can be computed from data or can be obtained from experts. Since no data were available in the current study in CATS, all rank correlations between nodes in the flight crew model were retrieved by expert judgment (see Section 5.3.1.1 for a more detailed description). Basically, the computation of rank correlation can be taken care of by the modelling techniques of BBNs.
4) Anchor the maximum and minimum effects of the management influences on fatigue using the probability density distribution from 2) as a guide. We asked the experts to give their judgment on the value of the fatigue level on the Stanford Sleeping Scale if all of the management actions were to be as effectively managed as possible, and the same question was also asked for all the management actions being as poorly managed as imaginable. In this way, we could anchor the management influences (weighting results) from the paired comparisons into the risk/human error in the BBN.

However, as we noted in Chapter 6, the main problem in the qualitative “procedure” case and in relation to many other soft variables in the human performance model we do not have agreed unidimensional definitions which are easy to model in this way and to scale or calibrate to their effect on risk/error. Moreover, there is no data on that scale from research. Hence, each of the sub-sections below deals with the question of how to proceed.

7.3.1 Quantifying qualitative variables

When it comes to applying this approach to qualitative cases, we can use the example of the quality of procedures as example to discuss how to fill in the four steps above. All other qualitative PSFs could then, in principle, be handled in the same way.

In Chapter 6, the definition of the effectiveness of procedures as needing to be "accurate, safe, clear, up-to date and easy to read and understand" was used in the paired comparison expert judgment to focus the search for the relative importance of different management actions on improving the effectiveness of A/E procedures. This definition is acceptable to be used in paired comparisons. However, if we want to analyze if and how well this organizational factor is being managed across a number of airlines, we need a scale to determine the effectiveness of the procedure in the same way that the Stanford Sleepiness Scale measures fatigue, where the 7 steps signifying different levels of vitality are a good benchmark, which links a point on the scale with the likelihood of error.

The same approach could be used for procedures. We could produce an ordinal measurement scale that assigns a scale of 1-5 indicating degree of effectiveness with respect to the procedure variable. 1 could represent the lowest (worst) score, namely "useless for guiding the behaviour", and 5 could represent the highest (best) score, namely "perfect for guiding behaviour". We would then have to define what the points in the middle are in terms of a particular level of effectiveness. The next step would be to see whether there is agreement among experts on this synthetic output or functional scale. We would have to give the subject matter experts different procedures and ask them to rate those procedures on the scale of effectiveness. If there were to be an agreement on the scale across the experts (the deviation should not be big), we could claim that the users are able to identify similar rating and reach the same conclusions during the course of rating by using this synthetic scale.

On the other hand, we could define the effectiveness of procedure in terms of characteristics of the procedures, using a number of indicators. For instance, as we did in the experiment in Chapter 5, we could propose the indicators for the effectiveness of procedures as "accurate, safe, clear, up-to date and easy to read and understand". We could split each of the six characteristics, at least in the first instance, into two states (1=yes, 0=no). The measured values of the indicators could be summed and converted into a unidimensional scale of 0-5, where 0 signifies the ultimate lower end of procedure as "not accurate, not safe, not clear, not up-to date and not easy to read and understand" and 5 stands for the top end of "very accurate, safe, clear, up-to date and easy to read and understand". This method gives, on the face of it, a more objective way of evaluating the effectiveness of procedure and gives more information of what experts think is important for effectiveness of procedure.

With these two approaches there is nothing, in principle, to stop us doing exactly the same with all the other factors mentioned in Table 7.1 as was done in the fatigue case. However, we realize that this would entail a large amount of work.

7.3.2 Quantify distributions for qualitative variables

Once the scale is defined we need data on how the variable, in our example case the quality of procedures, is actually distributed according to this scale across a number of procedures or across a number of airlines. Unfortunately, lack of data is a well-known phenomenon in risk modelling (see 5.2 for more discussion). The depth of the modelling is often determined by opportunity. When there are data the models go deeper than when there are not. Human performance modelling is still difficult territory when it is aimed at quantifying the probability of human error in particular settings and this is still an important weakness, not only in CATS. So, this means that the human models need further extension of data collection when they are used for more detailed analyses.

To do so, field studies to evaluate a number of airlines should be carried out. This is what we have seen in the Stanford Sleepiness Scale where large scale experimental studies have been performed by the Aviation Medicine Group of TNO in the Netherlands and data were collected by asking pilots to use this synthetic scale. This generated a distribution of the variable over its scale derived from practice, as shown in Table 6.2 (in Chapter 6). In the light of the obvious similarity in purpose and implication, we argue that it is plausible and warranted to apply the same approach by using well defined synthetic scales defined in 7.3.1 to collect data for procedures and the other PSFs identified in Table 7.1 when necessary. Given the importance of the role of flight crew in the air transport system and technical staff in the maintenance model, this is justified, even though it is a long term task. If the scales are properly developed and trialed with experts, through field studies over a number of years we could have many anchored scales and extensive data which could be used in aviation safety management quantification.

Figure 7.16 shows an example of a (fictional) probability distribution of a possible factor derived based on this approach which could be used in the quantification approach we offer in Section 5.4.

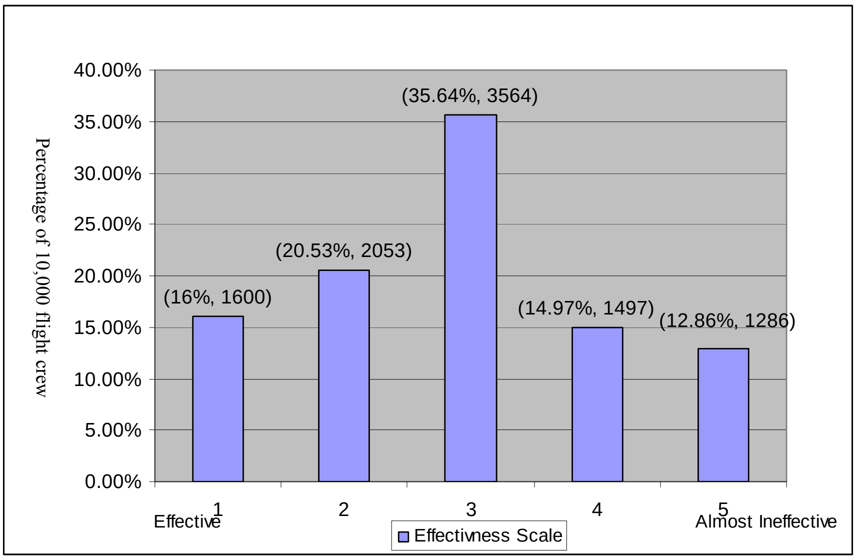


Figure 7-16 Fictional distribution for the actual effectiveness of A/E procedures

7.4 How to improve the availability of the data in aviation?

In Section 5.2, four types of hard data (ADREP, LOSA, EU-OPS, IOSA) were investigated to see whether it would be feasible to use any of these data sources to quantify the relationship between the safety management system and the human errors. The main results shown in that earlier section concluded that only a limited amount of management information was available up to now, largely due to confidentiality problems, missing data, and the lack of clear, consistent and recognisable causal frameworks for data collection. All of these point to the need for changes in the future.

Firstly, in CATS the management information contained in LOSA data and in IOSA audit data were both found to be extremely confidential to the airlines concerned. Even after a long period of negotiation they were finally not made available to us. It is quite understandable that data such as IOSA are not meant for public use. But, if under specific conditions such as maintaining confidentiality or use of aggregated data, the data owners would be prepared to release the information for scientific research, this would greatly help with the empirical data collection and the data analysis in the field.

Secondly, the experience of comparing data across different sources has shown that each data source uses its own classification system or taxonomy as a way of organizing knowledge about a subject matter. That is, different databases have different error classifications, which have unfortunately led to different models of the causal pathways. This means that there is a translation step necessary to be able to populate all information from different database into a common and recognisable causal risk framework. However, currently this framework does not exist in the field (see Chapter 2 to Chapter 4 discussions). This makes the comparison between different sources of data almost impossible.

Certainly, we can consider other sources of data not studied in this thesis like the black box, cockpit voice recorder and simulators. They seem to have rich potential if we can overcome the problems of confidentiality, by collecting only aggregated data and not individually attributable data. But the same issue still remains. So, to make the data analysis and comparisons possible, what we currently need is a model which provides the consistent framework which clarifies problems due to differences in definitions between model elements and data classifications. The general structured model we proposed earlier in this chapter (particularly the model in Figure 7.3 in Section 7.1) certainly does that.

Therefore, we offer our model to serve as the primary basis for mapping. Another more radical solution is to propose changing the way the current systems collect or classify data so that it fits better to our proposed model. As it is the industry-wide standard for accident and incident data collection in aviation, we would suggest the ambiguity in the classification system in ADREP should be resolved and the system needs to modify its current taxonomy by imposing a multi-level structure consisting of EEM, IEM, PEM, PSFs, and organizational factors. This would certainly improve data quality, encourage reporting, enhance usage, and allow identification of systematic shortcomings. We realize that such a proposal will meet with much opposition based on the enormous investment in the existing system.

7.5 Summary

In this chapter, we have concentrated on making proposals for the findings of the previous chapters which require improvements. This includes a generic hierarchical control model for aviation safety, a new version of the Dutch safety management model, some suggestions of quantifying the human factors, and recommendations to improve the availability of the data in aviation to be able to quantify the relationship between the safety management system and the human factors. In next chapter, we will highlight the major points found in this thesis and draw the final conclusions.

8 Summary, conclusions and recommendations

The Dutch Ministry of Transport, Public Works and Water Management recognized the importance of, and demand for a causal model in aviation. A project called CATS (Causal model for Air Transport Safety) was embarked on in 2005 to develop an integrated risk model. Part of this project envisaged linking a safety management model with the technical/human factors model and quantifying the risk implications of different management, as well as technical and procedural changes to prevent accidents. The general structure of the CATS model and its management part (the Dutch management model) had been decided upon before I was taken on to work on this thesis. It was therefore decided that this thesis should take a step back from the CATS project and re-examine the place and role of that management model and its quantification in a more fundamental way, in order to see what recommendations might be proposed in the longer term.

The main research question in this thesis was therefore:

Is it possible to develop a safety management model which can link with the human and technical factors as modelled in CATS, or compatible with it, in such a way that it lends itself to quantification of the contribution of those management factors to the risk?

To answer the research question, the following main steps were taken:

1. The development of the Dutch management model, its assumptions and structure were analysed, and its completeness and appropriateness to be applied to the aviation field were also examined;
2. To link with the human and technical factors, critical analyses were carried out to look at the issue of the human factors and technology failure models at the lower level, and how these might be connected with the management model.
3. The availability of data about management failures was also explored. Since lack of data has proven in this research to be a serious problem in quantifying the probability of management failures leading to technical or behavioural failures, the current quantification methods (linked to the BBNs and system dynamics) were critically discussed and a simpler form of quantifying management influences based on paired comparisons was assessed for its potential to get around the complexity of the expert elicitation that the BBN method involves;
4. Options for improvement on modelling and quantification were explored.

The main results are given below in more detail.

8.1 State of the art in the Dutch management model and need for additional improvement

The Dutch management model has a long development history through projects called I-Risk, ARAMIS and WORM, in which it has been tested a number of times, particularly in the chemical process industry. In the first part of Chapter 2 a review of the history of the development of the Dutch model was made to identify the crucial assumptions made in the development of the Dutch model and to indicate the issues still remaining to be solved. Since the assumption made in the CATS project was that the Dutch model would be suitable for the aviation industry, this was a hypothesis which was not tested in detail in CATS. Therefore, in the second part of Chapter 2 it was decided to compare the Dutch model used in CATS with other relevant models in order to assess the comparative validity of the model and indicate how it needs to be improved.

8.1.1 A review of the history of the development of the Dutch model and issues remaining to be solved

Tracing the development of the modelling through earlier projects (I-Risk, ARAMIS, WORM) has shown that some parts of the Dutch model have consistent features that have stood the test of time, whilst some others have gone through significant changes. A constant feature is the formulation of the primary process of the activity or company as the process of execution either by human or hardware or a particular mix of those that directly controls the activities and the hazards inherent in them. Early versions of the model linked management to the occurrence and control of errors and failures in the safety systems. In the ARAMIS project and later this link was formalised and *barriers* designed to prevent the full range of accident scenarios for the activity were formulated, consisting of the correct functioning of hardware and/or human behaviour. The barriers should detect, diagnose, and act to prevent the primary process leading too directly towards or over the boundary of a safe envelope of operations. The company management exercises control over major hazards by allocating suitable *resources and controls* to ensure the continued correct functioning of that hardware and human behaviour. Therefore, safety management was seen as “delivering” those systematic resources and controls to those barriers, in order to have the barriers put and kept in place for their whole life cycle.

The supply of these resources and controls is achieved by secondary (management) processes. Since safety management was seen as “delivering” resources and controls to the barriers, they were called *delivery systems* in I-Risk. The delivery systems were further detailed in ARAMIS as workings out of the barrier life cycle both for hardware and behaviour. This led to slightly different formulations of a limited number (usually 7) of delivery systems related to delivering hardware and behaviour in different projects. For example, in CATS we devised 2 delivery systems related to hardware (technology interface, technology function) and 5 delivery systems related to behaviour (procedures, availability, competence, commitment, communication). Apart from these delivery systems, the safety management system also has to manage, at a higher system level, the processes of “risk (scenario) identification, barrier selection and specification”, and the process of “monitoring, feedback, learning and change management”. Moreover, in order to clearly show what actions should be taken by managers to deliver resources and controls from the 7 delivery systems to the barriers, each delivery system was modelled as a series of steps which can be pictured as cycles of actions. Whether the company takes systematic and effective actions for each of the steps and links those steps seamlessly together determines the effectiveness of the barrier functions.

However, behind this relative consistency in the formulation of the model, the review of the development of the Dutch model showed two critical problems that needed to be resolved in this research. That is, that none of the previous projects had been finished in such a way that the original objectives of management modelling were realized. This lack of completion was partly due to the limited time scales available for the projects, but also because the concept model was not easy to apply in the past and therefore needed some simplification. Secondly, the current Dutch model did not give modellers sufficient clarity about where the management controls should go in relation to the individual factors. Only with CATS, dealing with aviation (where human factors play a much more important role in achieving safety) was it felt necessary to understand in more detail the individual factors underlying the actions and interactions of human and hardware in order to link management into them. In Section 7.1, we have addressed the way in which the SMS needs to be linked with the technical and human factors, and also clarified the steps in the delivery systems into a generic structure which can be worked out much more easily but still encompasses the essentials of the theory from the previous projects. Briefly the simplified generic structure consists of 7 generic steps forming a closed loop deliberately made per delivery system.

8.1.2 Comparative validation of the Dutch model

The only form of validation of the Dutch model which was feasible was a comparative one. In order to consider this, different models developed for or applied to aviation were mapped onto the Dutch management model to see if it could accommodate all of the insights that they contained. The comparisons with HFACS and SoTeRiA showed that the majority of the well-defined elements forming part of those models can be satisfactorily accommodated within our delivery systems. There are some ill-defined elements, particularly safety culture and safety climate, but also relating to higher level corporate aspects, of which their definition are still so vague that they cannot be easily compared, or they relate to aspects of management and organisation at a higher system level than the Dutch model currently attempts to deal with. To accommodate the first aspects in our model, the definitions of safety culture and safety climate were discussed in Chapter 7. We argued that safety cultural aspects defined in current research can be reflected in the strength of the safety management system through our delivery systems. Therefore, safety culture (in the current studies) can be seen as a measure of how seriously company takes its own management process. So, this means that safety culture is largely dealt with already in our safety management model. Dealing with any other more generic and higher level concepts (second aspect) is left to future work beyond this thesis.

Together with the foregoing tests based on the literature, a further step was taken towards a concurrent validation of the model, through interviews with airlines to see whether what they do in practice fits the Dutch model. These interviews revealed that current safety management systems (SMSs) as defined in the airlines are often not explicit about the complete risk control system, but are seen as clustered around the reactive accident investigation and learning program. The following findings and lessons are especially worth mentioning:

- What is called the SMS in airline is still a very partial and formal view. Airlines often identify only the “risk assessment”, “accident/incident investigation” and “learning” elements as their SMS.
- Comparing this to the Dutch management model, line operations are the ones that actually carry out all of the functions in the delivery systems for human and hardware. However, airlines seem not to define these activities in the line (where safety is influenced proactively) explicitly as part of their SMSs. In further studies, we found that the safety activities in the line operations were so integrated in the processes of the business and considered as simply part of normal business, that it was very difficult to tease them out. A further attempt was performed in this research by deeper questioning based on the Dutch model’s delivery systems. However, this turned out as an exercise to use the Dutch model to validate the existence of elements of normal practice in the company which are relevant to safety, rather than using current practice to validate the Dutch model. In this way, however, we did find that the activities summarised in the delivery systems defined in CATS (2 delivery systems for hardware and 5 delivery systems for behaviour) could accommodate the airlines processes well. However, to be more confident of the match and the validation of the Dutch model, some observational research is more preferable and is suggested to be used in future research.

8.2 The underlying causes that contribute to human performance and aircraft deficiencies in aviation and need for additional improvement in modelling

The link between the management model and the accident scenarios/event tree modelling runs through the underlying models of human and technical failure. These models therefore need to be able to interface not only with the events, but also with the management model.

8.2.1 Human performance model

To provide an overview of the underlying human error mechanism and identify the possible factors that contribute to this underlying human error mechanism, a model (Figure 8-1) based on Isaac et al.'s (2002) and Shorrock and Kirwan's (2002) models was introduced in Chapter3. It shows the external error mode (EEM) as the manifestation of the cognitive process of the flight crew, influenced by PSFs and aircraft performance. To yield effective error counter-measures, different approaches in the accident and incident investigation schemes were also reviewed in that chapter to identify the PSFs. It became clear that many accident/incident investigation tools use text narratives or do not build their human factors taxonomy according to any particular human cognitive model. Also, some models with taxonomies are not comprehensive to a sufficient extent within a hierarchical classification to be able to map to a deeper set of organizational causal factors.

Therefore, Table 8-1 was introduced to provide an overview of the factors for PSFs, which were summarized from different accident and incident investigation schemes, linking them to concepts at the same system level as the delivery systems.

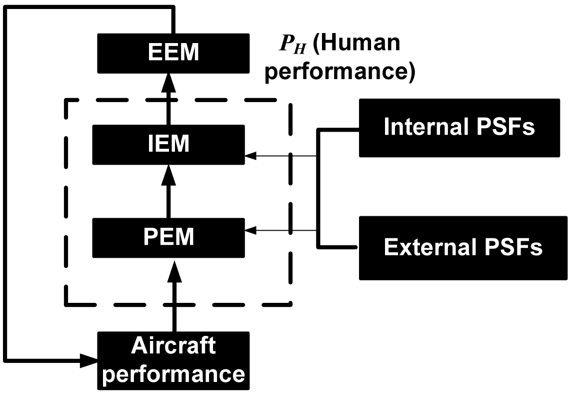


Figure 8-1 Human performance model

If we compare the factors in Table 8.1 with the delivery systems in the Dutch model, there is a good match, apart from the topics of "workload" and "competence and suitability". Additional functions for "workload" needed to be added to the Dutch model and "competence and suitability" needed to be split in Chapter 7. Also this mapping addresses one of the criticisms of the development of the Dutch model in Chapter 2, namely that they are still too conceptual and generic in respect to resolving (preventing and coping with) human and technical errors. One of the contributions of this thesis is therefore to tailor the delivery systems for each of the new/modified categories of the factors found in the accident and incident analysis and make the delivery systems less vague and generic. The full discussion and development is to be found in Section 7.2.2.

Quantification of the PSFs and how that has been formulated in current Human Reliability Analysis (HRA) in aviation was also reviewed in Chapter 3. The following shortcomings were found with the current HRA:

- The factors normally selected in HRA are often a limited subset of the complete set of influencing factors found in Table 8.1.
- The stronger the quantification objective, the more influences are left out of the limited selected subset, because they cannot be quantified rigorously.

This implies that the factors considered in the risk models are often very partial and not a comprehensive overview of factors that have the potential to influence flight crew performance as identified in Table 8.1. Therefore, one of goals of this thesis was to develop a quantification method within which it is possible to represent and quantify more influences of human factors in risk modelling in aviation.

8.2.2 Technical performance model

The qualitative research reported in Chapter 4 indicated four activities influencing aircraft technical performance:

- Design, material section
- Manufacturing
- Maintenance and inspection
- Flight crew operation

The first three aspects lie in the technical support provided by the management system and the last lies in the pilot online operation of the aircraft within the design limits. Figure 8-2 shows the aircraft performance model, which integrates the aircraft technical model with the human interaction described in Figure 8-1.

In Chapter 4, these four aspects (design, maintenance, manufacturing, flight crew operation) were briefly considered and modelled to ensure satisfactory performance of an aircraft system over its entire design life in relation to both the technical functioning and the man-machine interface. The processes relating to each aspect (e.g. design process, inspection and maintenance process, manufacturing process) and their links to the management modelling were discussed in that chapter. Detailed research on this part of the model fell outside the scope of the modelling for this thesis, and has also not been fully accomplished within the CATS project. Hence, that chapter confined itself to laying the groundwork for further development of the modelling design in the future.

Each process contains steps which require competent and committed people available to carry out the processes following the good procedures in collaborative teams, using good functioning and user-friendly hardware. For the modelling of safety in the operating phase of aviation, it is the maintenance technicians who are the human operators that play the most critical roles in making sure that the hardware of the aircraft stays safe. In future modelling it is recommended to use the human performance model proposed in Chapter 3 and summarised in 8.2.1 above to link to these hardware life cycle processes. This deeper analysis would supplement the current simplified process analysis of delivery systems for hardware. However, it makes the model more complicated as it iterates deeper into the organisation. Currently, we suggest that, only if the tasks related to the provision and maintaining of the aircraft are critical, should such an analysis be done.

Table 8-1 Influencing factors for flight crew and mapping of PSFs onto higher level systems

Definition	Description	Selected examples	Delivery systems
Internal PSF: physiological and psychological factors of the flight crew which might influence human information processing	• Technical and interpersonal skills that match the requirements of performance	• Experience for complex situation • Technical skill • Communication& coordination between flight crew, ATC, and operation center	• Competence • Communication&coordination
	• Physical fitness to perceive, process, respond to information and feedback information	• Human physical and sensory limitation • Medical illness • Hypoxia • Physical fatigue • intoxication	• Suitability • Availability
	• Psychological fitness to perceive, process, respond to information and feedback information	• Mental fatigue due to sleep loss or other stressors • Emotional state • Personality characteristics (complacency, overconfidence)	• Suitability
	• Decision to choose one from several possible courses of action and decide to commit to safety procedure above other personal and organisational goals	• Pressure to achieve • Personal objectives • Incentives, needs • Perceptions of organization's beliefs and attitudes (manifested in actions, policies, and procedures, affect its safety performance)	• Motivation to commit to safety
external PSF: Factors external to the flight crew which might influence human information processing	• Clear and relevant guides to adequate performance	• Operational manual • Checklist • Charts • Standard operating procedures • Emergency and abnormal procedures	• Procedure
	• Tools and materials relate to physical activity designed scientifically to match human factors (include working postures, materials handling, repetitive movements, work related musculoskeletal disorders, workplace layout, safety and health)	• Checklist layout • Display/interface characteristics • Automation/alerts/warning	• Techonology-Man-machine interface (Instruments& workplace design)

	• Data and information	• Information from ATIS • Information from operation center	• Communication&coordination
	• Environment in which the action needs to be performed	• Traffic configuration o Traffic density o Traffic complexity (e.g. Runway length, Runway crossing, Runway condition, Runway slipperiness, Terrain at/near airport) • Possible flight delays • Ambient environment o Wind shear o Cross wind o Visibility in flight o Visibility at airport o Turbulence o Icing o Light condition	• No relevant management functions in the current Dutch model. Need to add under “Workload” delivery system (will explain in Section 7.2.7).

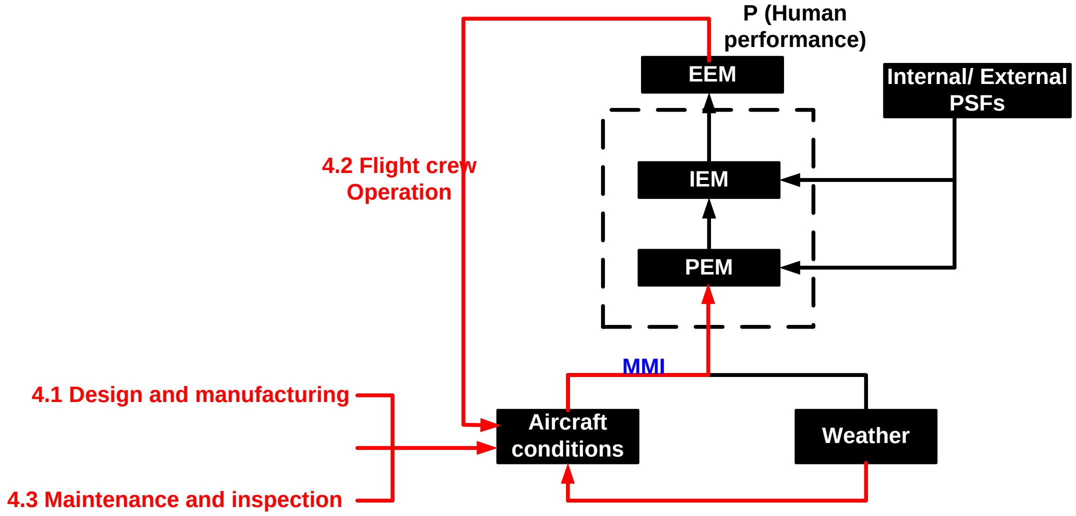

P (Human performance)
EEM
Internal/ External PSFs
4.2 Flight crew Operation
IEM
PEM
MMI
4.1 Design and manufacturing
Aircraft conditions
Weather
4.3 Maintenance and inspection

Figure 8-2 Aircraft technical performance model

8.3 Available data and their problems in risk modelling in aviation

In Section 5.2, four types of hard data (ADREP, LOSA, EU-OPS, IOSA) were investigated to see whether it would be feasible to use any of these data sources to quantify the relationship between the safety management system and the human errors. The conclusions from these very time-consuming experiences of trying to work with the four types of hard data is that

- only a limited amount of management information was available up to now
- this is largely due to confidentiality problems, missing data, and the lack of clear, consistent and recognisable causal frameworks underlying the data collection models.

These two points cry out for changes in the data sources as discussed in Chapter 7.

8.4 Quantification methods of SMSs and their problems in risk modelling in aviation

Currently there are two major methods to incorporate management factors into risk models. Some of the existing studies use Bayesian Belief Networks (BBNs) which provide a framework for modelling the logical relationships between variables and captures the uncertainty in the dependencies between these variables. Most of the existing studies use discrete BBNs. CATS is the only study to use distribution-free continuous BBNs in the technical model to overcome some of discrete BBNs' limitations. Other studies have tried to model management and organizational factors deterministically by using System Dynamics or have combined this with BBNs. Each of these methods was reviewed in Chapter 5, and each method's advantages and disadvantages in its application to management quantification can be grouped as follows:

8.4.1 System Dynamics

- System Dynamics has very good representational features for complex systems. It can present complex systems using feedback loops. So the "learning and improvement" identified in the block diagrams of the Dutch model could be effectively and systematically coded with minimum simplification using System Dynamics for risk modelling.

- At the moment, the equations which would be needed to quantify the System Dynamics models for risk frameworks have no basis in data or evidence. The equations are generally arbitrarily assigned through non-structured human judgment (e.g. interviews with employees, surveys). Applying these arbitrary equations can affect the accuracy of the management model, and eventually impact the explanatory and predictive power of the resulting risk model. Hence, in this respect, System Dynamics models have received widespread criticism as "measurement without data".

8.4.2 Discrete BBNs

- BBNs manage uncertainty by explicitly representing the conditional dependencies between different variables. Dependencies and interactions between different variables are easily modelled in this way.
- A BBN does not allow feedback loops in the modelling formulation, so it cannot represent the learning and feedback loops in the model.
- Discrete BBNs are not flexible with respect to changes in modelling. If we add one management factor to the parent nodes, we have to re-do all the assessments for the child of this node.
- The main drawback of the discrete BBN is the excessive assessment and maintenance burden in specifying a conditional probability table using expert judgment. When directly measured data are not available, the expert judgment linked to the BBNs becomes too complex as the variables in the models increase. The numbers of probabilities that have to be assessed and maintained for a child node increases exponentially with the number of the parent nodes and with the number of states that each parent node can take. So to minimising the excessive assessment burden, both numbers have to be kept to their minimum.

8.4.3 Distribution free continuous BBNs

- The same comments mentioned above apply as for discrete BBNs, except that in this version of BBNs complexity is reduced to a linear function of the number of parent nodes rather than exponential. This version of BBNs also reduces the flexibility problems with respect to changes.
- Similar to the last bullet covered under discrete BBNs, to facilitate quantification of the model with expert judgment, in this version of BBNs it is required to keep the number of management nodes limited to 5 or 6 maximum in order for experts to mentally process the complex (conditional rank correlation) questions, and even then they experience difficulties.
- In distribution free continuous BBNs, each expert is asked about a set of seed variables. Based on the expert's answer to these seed questions, the weight to be given to each expert in combining the expert elicitation results is computed. However, with the dearth of available data from management, it is doubtful if seed variables are available and can be used to robustly predict the expert performance.

Overall, these methods have different strengths and weakness. BBN approaches have a systematic quantification, which sophisticatedly captures the uncertainty in the dependencies between variables; but they get much too complex as the number of parent nodes increases and cannot cope with feedback loops. On the other hand, System Dynamics can incorporate a wide range of soft variables and allows feedback loops in a more complex system than BBNs are able to do; but it currently lacks a formal estimation method in its application to safety.

This can be combined with the important finding from the qualitative research in human factors: we need to represent more influences of human factors in risk models in order to incorporate as complete a set as possible of soft management influences into risk quantification. This requires the development of a supplementary method to be combined with the BBNs to quantify the factors represented in System Dynamics, and reduce in a systematic way the number of variables to be quantified by those BBNs to a manageable number.

8.4.4 Supplementary method (combining paired comparisons with distribution free continuous BBNs)

Paired comparisons are psychological scaling models which give us the relative importance or weighting of a set of variables. Hale et al. (1999, 2000) used paired comparison to assess the relative importance of management factors on risk control in the chemical industries. But they did not successfully connect the management weighting to error probabilities. Therefore, it was not possible to conclude the real size of the management influences on risk. The method designed in this research was intended to solve this problem and get round the complexity of the BBNs.

The quantitative method, recapitulated from Section 5.4.2, can be decomposed into the following steps for each type of delivery system:

-Step 1. Identify the human factors in the technical/human behaviour risk model which the management model will link into

-Step 2. Generate management actions to influence these factors and use card sorting to reduce the preliminary organisational influences to a manageable number (up to about 14)

-Step 3. Use a quantification technique, including

- Assess the impact of the management actions on the human factors using paired comparisons
- Assess the states of the management factors in the situation to be modelled
- Calculate the total management influences on the human factor in the risk model

-Step 4. (Re)calculate the total effect on the risk

The experiments discussed in Chapter 6 studied the feasibility of applying this method to quantify the management influences on two sets of human factors. The first were “fatigue”, “weather” and “workload” that had been precisely defined with objective quantifiable units in the CATS project. The second was the “quality of emergency procedures” that had been considered too difficult to be modelled in the flight crew model in CATS and had been totally left out of the modelling. The findings were:

- Using paired comparison, experts felt more comfortable to judge the relative importance of management influences on the variables than they did with the complex BBN questioning.
- This method can be applied both for cases with quantitative units, and with just qualitative units. It is particularly useful to be applied in the qualitative cases, because it gave a better representation of the reality of the variable and incorporates more influences of human factors in the risk models.
- The technique is therefore capable of producing a relative quantification of as complete a set as possible of the management influences into the risk model.

- However, anchoring of the relative weightings with expert judgment to turn them into absolute weightings is only possible if there is underlying quantification of the factor being influenced. Where this is not present, only relative statements of importance of the factors can be made.
- As with the BBNs the method does not handle cycles of influences well. It can be difficult for the experts to assess the relative importance of contiguous steps in the processes which feed into each other.
- An important potential use is to reduce the number of potential influences on a variable by using a transparent method to identify its main priorities before the use of the more rigorous BBN techniques for final quantification

In conclusion, this method is able to enrich the assessment of influences and provides a filter to reduce qualitative complexity to its main priorities.

8.5 Improving safety management modelling and its quantification in aviation

Even with all the improvements and development achieved in this research, the delivery systems in the original Dutch safety management model were found to be still too broad, generic, and not easy to apply to model all of the issues related to human factors and technical failures found in accident analysis. Thus, the Dutch model needed some changes. In order to improve it, at least the following groups of improvements, discussed in Chapter 7 should be taken into account.

8.5.1 Clarify the hierarchical relations between the SMS and operations

Firstly, to demonstrate explicitly where the links are from management to the lower level individual and technical performance, a general structured model (Figure 8-3) was introduced in this thesis to clarify issues related to human factors and technical failures in the accident analysis.

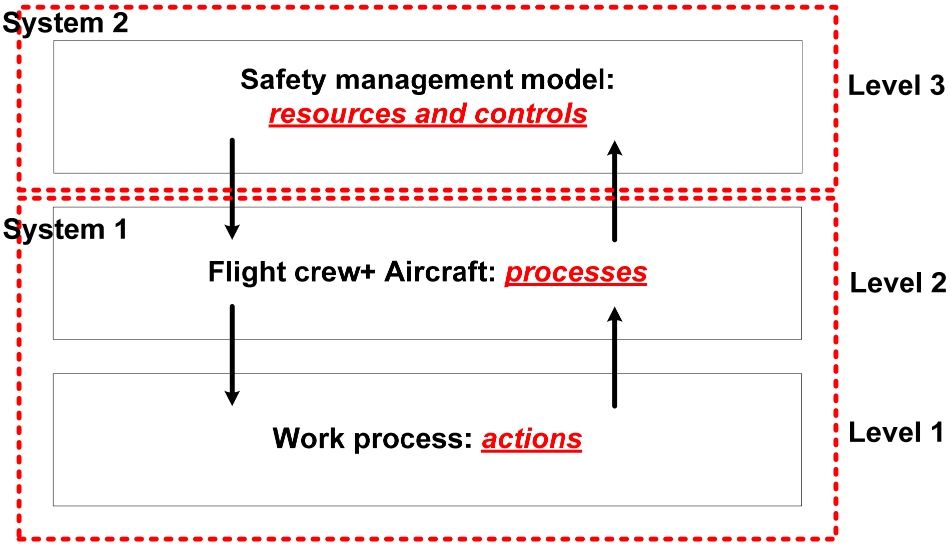


Figure 8-3 Hierarchical model

The hierarchical relations between the SMS and operations are treated as a control process. Two systems in the risk control process are distinguished. System 1 deals with the direct aspects of the execution of an action and is further divided into two levels, the work process itself (level 1), consisting of observable actions to directly control risk performed by flight crew and aircraft in the flight process, and the internal process (of the cognitive mechanisms of the human and the equivalent internal functioning of the hardware) which leads to actions and interactions at level 1.

System 2 is the SMS that prepares resources and controls that ensure that the outputs of system 1 (level 1) meet the objectives and goals set by that management system in system 2. In this version of the Dutch model, we see safety management as ensuring that the internal processes in level 2 are working properly and individual factors that interfere with it are managed to an acceptable level. So instead of defining the role of safety management in terms of the management of the life cycles of risk control barriers, we focus this version of the Dutch model on managing individual factors and internal processes of the human and hardware. The benefit of doing so is that the SMS can be devised more specifically and we can end up with a model tailored to the issues related to human factors found in the accident analysis.

8.5.2 Improve the detailed modelling of the system levels

The theories and findings in this thesis were put together in the general structured model in Figure 8-4 to include technical, human, and SMS factors in an integrated and articulated model. To make the SMS more specific in its task of managing issues related to underlying causes in level 2, this thesis specified an extensive list of the human factors at that level (level 2) (Table 8.1). The list of the human factors and the control functions that need to link to them by the delivery systems was discussed in Section 7.2. This specifies and makes clear what management actions should be done and should be revised in level 3 to resolve human factor issues in level 2.

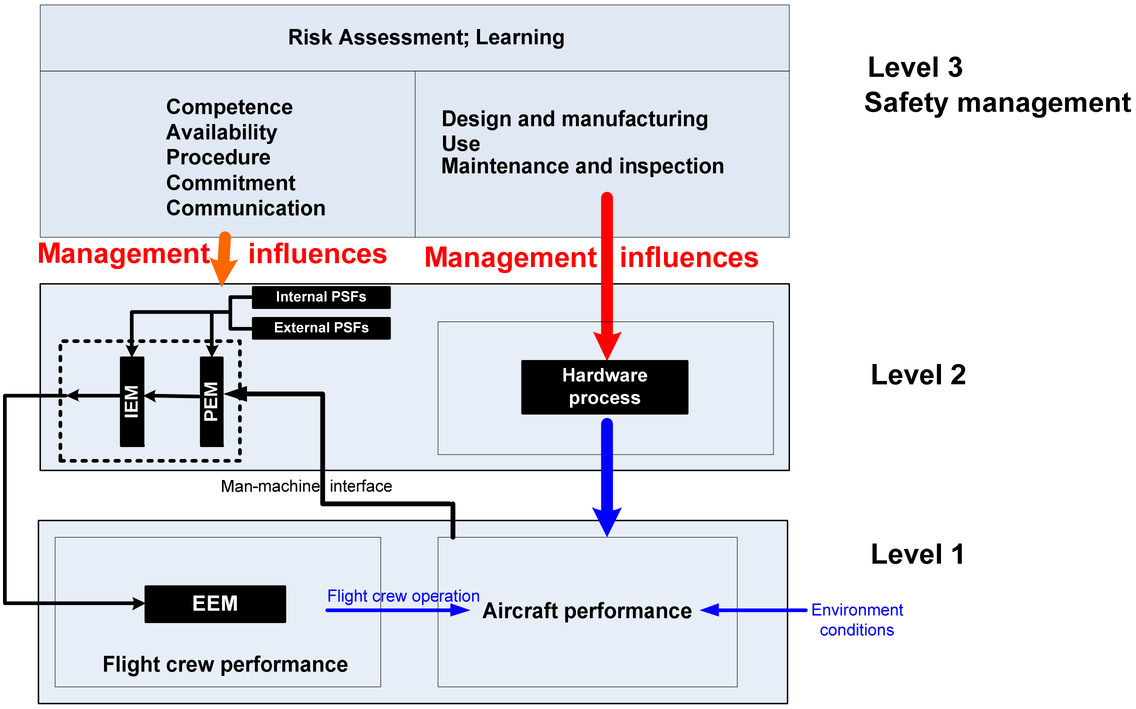


Figure 8-4 General structured model for aviation

8.5.3 Clarify delivery systems into a generic structure

To make the Dutch theoretical model easier to apply to the risk modelling, this thesis simplified the delivery systems (identified in level 3 in Figure 8-4) into a generic structure. Figure 8-5 draws this generic structure which was worked out specifically per delivery system in Chapter 7 to provide resources and controls to the factors identified in Table 8.1.

The management process entails 9 generic steps per delivery system:

(1) specify,
(2) provide
(3) promulgate/train
(4) threats and process of pilot/ aircraft (not management step but operational step)
(5) execute step by pilot/aircraft (not management step but operational step)
(6) monitor/evaluate
(7) maintain/change
(8) collect state of the art
(9) assess risks of proposed changes

All of these steps form a closed loop of a learning system, which has been worked out in flow diagrams. These steps are deliberately made generic to cover both delivery systems of "technology (hardware and software)" and "humans and their behaviour" as ways of operationalising the risk control measures and their functions. As demonstrated in Section 7.2, this generic structure has been proved by applying it to each of the existing delivery system.

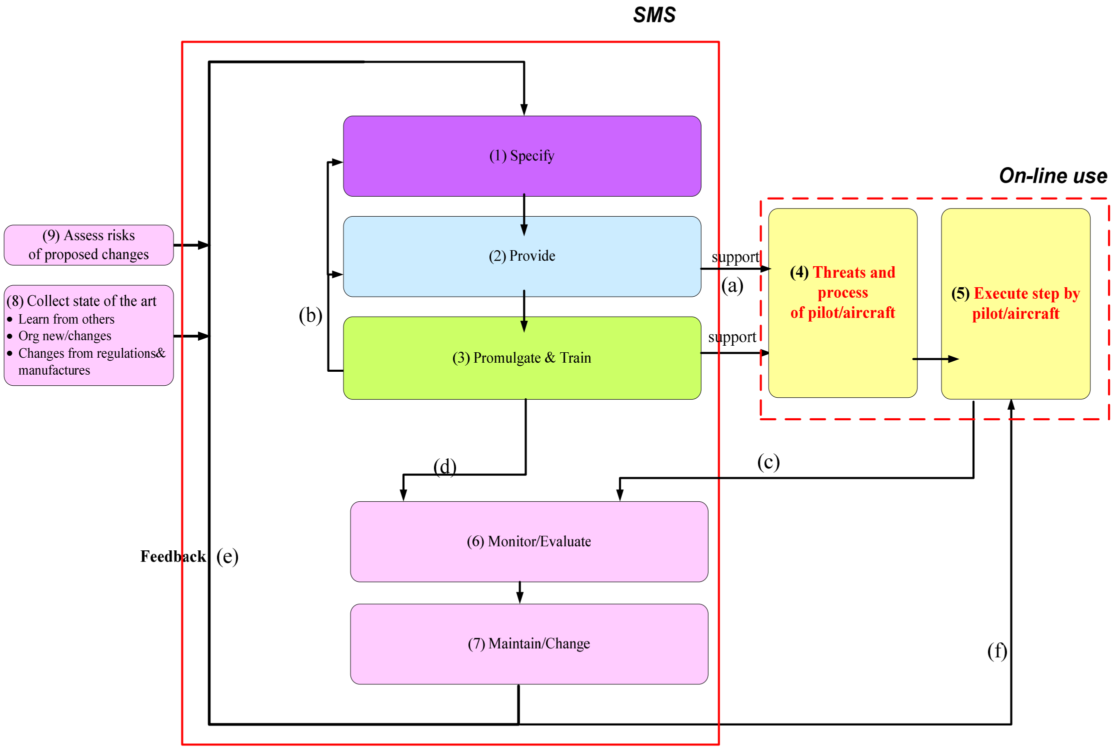

SMS
On-line use
(1) Specify
(2) Provide
(3) Promulgate & Train
(4) Threats and process of pilot/aircraft
(5) Execute step by pilot/aircraft
(6) Monitor/Evaluate
(7) Maintain/Change
(9) Assess risks of proposed changes
(8) Collect state of the art
• Learn from others
• Org new/changes
• Changes from regulations& manufactures
support
(a)
support
(b)
(c)
(d)
Feedback
(e)
(f)

Figure 8-5 General structure of the delivery systems in the safety management model

8.6 Final conclusion

The scope of this thesis was not to develop a new methodology for risk modelling to link all system levels from the technical models of risk and their control through human to management control measures. Instead it started from an existing risk modelling approach and investigated whether the experience of using it can be used for learning and the risk modelling approach can be improved. Its contributions are to clarify the list of human and technical factors to be treated, and to develop and test an additional way of quantifying them (particularly for the soft variables). The thesis also showed from literature how difficult it is to meet the needs of quantification and at the same time incorporate as complete a set of influences as possible in the risk model. This thesis has gone through this process. The difficulty of risk modelling lies in finding the balance between quantification and reality, and in making compromises to meet the project goals. The added value of this thesis is that it made the modellers aware of these conflicts and helped them to know what must be made explicit to achieve the vital objective to be reached.

8.7 Limitations

There are several things that the research presented in this thesis did not take into consideration:

- The model presented in this thesis is a reduced scale model, limited to the research scope of the safety management influences related to the performance of flight crew in the aircraft in the primary process of flight operations. However, this first attempt showed the potential for a much more extended model, in which the same techniques of modelling are used for the many other stakeholders (identified in Figure 1.3 in Chapter 1). Currently there are separate human models for air traffic controllers and maintenance technicians in the CATS model. They also could be developed by adding the management influences for these actors that are considered important but have not yet been developed in CATS. This can be done by using the 3-level analysis identified in the general structured model in Figure 8-4 and detailing it for each of the actors in turn, for example the part of the model concerning the work process of the maintenance technicians. In this way, the model can be made more useful for different groups of stakeholders, showing their own influences, whilst also taking into account the factors that are beyond their main field of interest.

- This thesis did not exhaustively research all theoretical models of safety management, but confined itself to two which are already applied in aviation. The perspective used was an existing safety management model (the Dutch model), which had been developed in earlier studies in the Netherlands and which had been chosen for use in the main CATS project. There are other management models available in the literature (e.g., Pate-Cornell (1996), Davoudian (1994a, b), which vary in their applicability to specific situations, and in their focus and intended use. Their potential contribution to a definitive safety management model for aviation has not been assessed and could give additional insights. However, this thesis has made an intensive study of the justification of the Dutch safety management model applied in risk modelling in aviation, which had not been carried out before. This revealed a number of dead ends or limits to the validation of the model. However, the experience of doing so, which is presented in this thesis may give other researchers with similar objectives an idea of how a theoretical model makes its way into scientific application, and what assumptions and compromises are made or not made.

- The focus for this thesis was only on the influences on flight crew performance from the management of the airlines, as we assumed that they have the most control over the resources and controls delivered to the flight crew. We did not study the way in which the regulator's mandatory requirements develop and link with an airline company's safety management. What has been modelled so far in this thesis and in CATS does not make explicit the link to the tasks of the regulator, which he may wish to see explicitly modelled if the model is to be useful in guiding his decisions. A start can be made on this extension to the modelling based upon the preliminary experiences of mapping from EU-OPS and IOSA in Section 5.2.3. As demonstrated in that section, the task of the regulator can be seen as making mandatory the requirements for organizations to have the primary risk control measures in place. A comparison of the scope of these audits would be interesting. It could be useful to perform risk-based inspections using such audit data on the basis of regulation results, if they are available.

8.8 Recommendations for future work

8.8.1 For qualitative modelling

As has been indicated, one of the main objectives of developing risk models has often been the requirement to quantify influences. This also applies in CATS. However, before quantifying any risk model, qualitative analysis is needed, to gather an in-depth understanding of human and technical behaviour and the reasons and (organisational and system) influences that govern such behaviour. To provide a well-articulated and rigorously constructed qualitative model, the following aspects have to be taken into account:

- First, the qualitative model should contain all of the influences which actors in the system would recognise as options to influence risk. A complete search of literature and analysis of a large sample of accidents and incidents can give a potential set of influences that is as complete as can be achieved[45]. In the modelling we need to establish which are the major influences which we need to retain and which can be left out. If the modellers do not know (as at present) how to quantify the important influencing factors, they should leave the factors in the model; but indicate that these factors will not as yet influence the quantification of the model, by "turning them off". This is a vital principle to resolve the conflict between limiting the model only to easily quantifiable influences and thereby losing many important influences, and swamping the model with unusable factors.
- Moreover, when the quantification that defines the factors in objectively quantifiable units only captures some (small) part of a complete influence (e.g. days since last training is taken as a measure of the whole influence of training, which takes no account of the quality of the training), the modellers should indicate that the quantification is only partial and the factor is seen as a proxy for the total influence.
- Given that a shortcoming in the management factors which are modelled is seen as a contribution to an unsatisfactory risk control, we need to analyse the individual factors for the shortcoming of these specific factors. It is most useful for the users if the model focuses on defining the influences as risk-reducing measures. It is important for management factors to be explicitly modelled as actions which can be taken by managers to influence the problems identified in the human and technical failures, rather than treated as simple failures in the causal chain. If this is the case, then we at least capture the effect of the remote managerial factors explicitly and provide help in risk control measures. As work progresses with the expert judgement exercises to model more of the

[45] 100% completeness is theoretically never attainable, but can be approached if there is a system to update the set of influences based on learning from own and others' experience of accidents and incidents.

qualitative influences in the Dutch management model, this will generate the need for more research into how to formulate the influence of management actions on risk control.

8.8.2 For the methodology of quantification

- First of all, it should be emphasized that further work needs to be done on model validation. The assessment of the anchoring values by expert judgments has been shown to be one of the sensitive parts in the quantification methodology. These anchoring values were the ones obtained by assessing the assumed number of contributions to PSFs of organizational factors given that they were changed to their worst or best possible state. Methods need to be developed to cope with situations in which no objective data exist to anchor the judgements.
- Data collection on objective studies of the influence of given management factors would facilitate ranking and weighting studies, in which the paired comparison results are compared with the results obtained from data. In such cases, the influence of management actions on the PSF assigned on the basis of expert judgment can be more realistic and more objective. The data available can be used not only for validation, but also for quantification.
- Validation of the risk models is only possible to the extent that data are available. Currently independent quantitative validation is almost impossible for management. Therefore other proxies need to be used to assess and maximise the validity of the model where possible, such as comparison with modelling efforts elsewhere in the world and expert and peer review of the results. Even if they are much less comprehensive than a full predictive validation, they can be useful to gain acceptance of the model.
- Obviously, there is a huge need for further data to improve risk modelling. A risk model such as CATS has very large data requirements. For the development of a complete management model, several thousand numbers need to be extracted or estimated. Current major problems are in finding exposure data (where tools such as LOSA, IOSA have great potential) and in overcoming the low amount of the management data recorded on accidents and incidents (ADREP). As we have advocated through this thesis, if company audit data could be released for scientific research it would be of great help to get a better estimate of the probabilities of management events early in the causal chain.
- Another problem for the data collection is the lack of common error classification schemes as a way of organizing human and management factors. Currently the data are held in separate databases built on different classifications. Without a common classification scheme, it proved in CATS that it is almost impossible to compare between different sources of data. Therefore, what seems urgently needed is to develop industry-wide schemes that are comprehensive and compatible with each other. This would allow data held in separate databases to be systematically integrated into the model. In such a way, it would become possible to enhance overall data usage and help identify weaknesses and holes in the aviation system.
- Different modelling and analysis techniques have different strengths and weaknesses. That means they are suitable for use in answering different questions. The use of the paired comparison method to filter and provide relative quantification of management influences in an aviation risk model is quite new. We present this as a first step towards understanding the priorities of management influences and their impact on safety. Hopefully we can combine the advantage of this with the advantages of the other different techniques in the future for quantification. For instance, we can use the method of paired comparison to eliminate very minor management influences to reduce the complexity

problem. Then, the rest of the influences can be formulated in the BBNs in which dependencies and interactions between different variables can be easily modelled. This is the hope that this thesis offers, but its proof will be given only if this potential is realised in practice in the next phase of modelling practice, for example in a renewed CATS.

References

- AAIB (2001). Final Report No. 1793 by the Aircraft Accident Investigation Bureau concerning the accident to the aircraft AVRO 146-RJ100, HB-IXM, operated by Crossair under flight number CRX 3597 on 24 November 2001 near Bassersdorf/ZH. Aircraft Accident Investigation Bureau of Switzerland, Berne, Switzerland.
- AAIB (2007). Aircraft accident report No 1/2007. Report on the serious incident to British aerospace ATP, G-JEMC 10 NM Southeast of Isle of man (Ronaldsway) airport on 23 May 2005. UK Air Accidents Investigation Branch, Hampshire, U.K.
- AAIB (2009). Aircraft accident report No 3/2009. Report on the serious incident to Boeing 737-3Q8, registration G-THOF on approach to Runway 26 Bournemouth Airport, Hampshire on 23 September 2007. UK Air Accidents Investigation Branch, Hampshire, U.K.
- Agency, E. A. S. (2008). Annual safety review 2008.
- Akerstedt, T. (2000). "Consensus statement: fatigue and accidents in transport operations." Sleep Research **9**(395).
- Ale, B. J. M., Bellamy, L. J., Cooke, R. M., Goossens, L. H. J., Hale, A. R., Roelen, A. L. C. and Smith, E. (2006). "Towards a causal model for air transport safety - an ongoing research project,." SAFETY SCIENCE **44**(8): 657-673.
- Ale, B. J. M., Baksteen, H., Bellamy, L. J., Bloemhof, A., Goossens, L. H. J., Hale, A. R., Mud, M., Oh, J. I. H., Papazoglou, I. A., Post, J. and Whiston, J. Y. (2007). "Quantifying occupational risk: The development of an occupational risk model." Safety science **46**: 176-185.
- Ale, B. J. M., Bellamy, L. J., Boom, R. P. v. d., Cooper, J., Cooke, R. M., Kurowicka, D., Lin, P. H., Morales, O., Roelen, A. L. C. and Spouge, J. (2008a). Using a Causal model for Air Transport Safety (CATS) for the evaluation of alternatives. ESREL2008 Valencia.
- Ale, B. J. M., Bellamy, L. J., van der Boom, R., Cooke, R. M., Goossens, L. H. J., Kurowicka, D., Lin, P. H., Roelen, A. L. C., Cooper, H. and Spouge, J. (2008b). Further development of a Causal model for Air Transport Safety (CATS); The complete model. PSAM9. Hong Kong.
- Ale, B. J. M., Bellamy, L. J., Cooke, R. M., Kurowicka, D., Lin, P. H., Morales, O., Roelen, A. L. C. and Spouge, J. (2009). Causal Model for Air Transport Safety (final report). Den Haag, Netherlands, Ministerie van Verkeer en Waterstaat.
- Allen, J., J.A. and Rankin, W. L. (1995). "A Summary of the Use and Impact of the Maintenance Error Decision Aid (MEDA) on the Commercial Aviation Industry. Proceedings of the Flight Safety Foundation 48th Annual International Air Safety Seminar, International Federation of Airworthiness 25th International Conference, and the International Air Transport Association "Managing Safety", Seattle, WA, USA.".
- Amalberti, R. (2001). "The paradoxes of almost totally safe transportation systems." Safety Science **37**(2-3): 109-126.
- BEA (1976). Rapport Final TC-JAV Ermenonville (in french). Bureau d'Enquêtes et d'Analyses, Le Bourget cedex, France.
- Beaubien, J. M. and Baker, D. P. (2002). "A review of selected aviation Human Factors taxonomies, accident/incident reporting systems, and data reporting tools." Applied Aviation Studies **2**(2): 11-36.
- Bellamy, L. J., Wright, M. S. and Hurst, N. W. (1993). History and development of a safety management system audit for incorporation into quantitative risk assessment. International Process Safety Management Workshop, AIChE/CCPS.

- Bellamy, L. J., Papazoglou, I. A., Hale, A. R., Aneziris, O. N., Ale, B. J. M., Morris, M. I. and Oh, J. I. H. (1999). Developmenet of an integrated technical and managmeent risk control and moniitoring methdology for managing and quntifying on-site and off-site risks, I-Risk main report. EU Contruct number ENVA CT96-0243, Ministry of Social Affairs and Employment, Den Haag. Den Haag Ministry of Social Affairs and Employment.
- Bellamy, L. J. (2007). Experiences with using ECCAIRS ADREP Database in the CATS Project. White Queen report 070423-07.
- Bellamy, L. J., Ale, B. J. M., Whiston, J. Y., Mud, M. L., Baksteen, H., Hale, A. R., Papazoglou, I. A., Bloemhoff, A., Damen, M. and Oh, J. I. H. (2008). "The software tool storybuilder and the analysis of the horrible stories of occupational accidents." Safety Science **46**: 186-197.
- Berliner, D. C., Angelo, D., Shearer, J. (1964). Behaviours, measures and instruments for performance and evaluation in simulated environments Symposium on the Quantification of Human Performance. Albuquerque, New Mexico.
- Bradley, R. (1953). "Some statistical methods in taste testing and quality evaluation." Biometrica 9 **9**: 22-38.
- Cacciabue, P. (2000). Human factors insight and reliability data from accident reports: the case of ADREP-2000 for aviation safety assessment. International conference on probabilistic safety assessment and management
- Comer, K., Seaver, D., Stillwell, W., Gaddy, C. (1984). Generating human reliability estimates using expert judgement. NUREG/CR-3688.
- Cooke, R. M. (1991). Experts in Uncertainty: opinion and subjective probability in science, New York: Oxford University Press.
- Cooke, R. M. and Goossens, L. H. J. (2000). Procedures guide for structured expert judgment. Technical Report EUR 18820. Brussels, Belgium, European Commission.
- Cooke, R. M. and Goossens, L. H. J. (2008). "TU Delft expert judgment data base." Reliability Engineering & System Safety **93**(5): 657-674.
- David, H. A. (1963). The method of paired comparison, London: Charles Griffin & Co. Ltd.
- Davoudian, K., Wu, J.-S. and Apostolakis, G. (1994a). "Incorporating organizational factors into risk assessment through the analysis of work processes." Reliability Engineering and System Safety **45**(1-2): 85-105.
- Davoudian, K., Wu, J. S. and Apostolakis, G. (1994b). "The work process analysis model (WPAM)." Reliability Engineering and System Safety **45**(1-2): 107-125.
- Dawson, D. and McCulloch, K. (2005). "Managing fatigue: it's about sleep." Sleep Medicine Reviews **9**(5): 365-380.
- Degani, A. and Wiener, E. L. (1994). Philosophy, policies, procedures, and practices: The Four “P”s of flight deck operations. In N. Johnston, N. McDonald, & R. Fuller (Eds.), Aviation Psychology in Practice (pp. 44-67). Hants, England: Avebury Technical.
- Dekker, S. (2006). The field guide to understanding human error, Ashgate
- Deming, R. (1968). Characteristics of an effective management control system in an organisation Boston.
- Drogoul, F., Kinnersly, S., Roelen, A. and Kirwan, B. (2007). "Safety in design – Can one industry learn from another?" Safety Science **45**(1-2): 129-153.
- EASA (2008). Annual Safety Review 2008. European Aviation Safety Agency, Cologne, Germany.

- Edwards, E. (1972). Man and machine: Systems for safety. In Proc. of British Airline Pilots Associations Technical Symposium. British Airline Pilots Associations, London.
- Embrey, D. E. (1992). "Incorporating management and organisational factors into probabilistic safety assessment." Reliability Engineering and System Safety **38**(1-2): 199-208.
- EUROCONTROL (2006). Integrated Risk Picture for air traffic management in Europe, EEC Note No. 05/06, EUROCONTROL, France.
- FAA (2006). Advisory Circular No. 120-92. Introduction to safety management systems for air operators. Washington DC.
- Fielder, J. H. and Birsch, D. (1992). The DC-10 case: a study in applied ethics, technology, and society, State University of New York Press
- Fitts, P. M. (1954). "The information capacity of the human motor system in controlling the amplitude of movement " Experimental Psychology **47**(6): 381-391.
- Flin, R., Mearns, K., O'conner, P. and Bryden, R. (2000). "Measuring safety Climate: Identifying the common features " Safety Science **34**: 177 - 192.
- Forrester, J. W. (1961). Industrial Dynamics, Pegasus Communications.
- Flight Safety Foundation (2005). "Lessons from the dawn of ultra-long range flight" Flight Safety Digest **24**: 1-45.
- Gander, P. H. (2001). "Fatigue management in air traffic control: the New Zealand approach." Transportation Research Part F **4**(1): 49-62.
- Gander, P. H., Rosekind, M. R. and Gregory, K. B. (1998). "Flight crew fatigue VI: an integrated overview." Aviation, Space, and Environmental Medicine **69**: B49-B60.
- Ghahramani, Z. (1997). "Learning Dynamic Bayesian Networks." Lecture Notes in Computer Science **1387**: 168-197.
- Goossens, L. H. J., Cooke, R. M. and Steen, J. v. (1989). "Expert opinions in safety studies, Philosophy and Technical Social Sciences, vols. 1-5. Delft University of Technology, Delft."
- Haddon, W. (1973). "Energy damage and the ten countermeasure strategies." Human Factors **15**(4): 355-366.
- Hale, A. R., Heming, B. H. J., Carthey, J. and Kirwan, B. (1994). Extension of the model of behaviour in the control of danger. Industrial Ergonomics Group, School of Manufacturing & Mechanical Engineering, University of Birmingham, UK.
- Hale, A. R., Heming, B. H. J., Carthey, J. and Kirwan, B. (1997). "Modeling of safety management system " Safety Science **26**: 121-140.
- Hale, A. R., Costa, M. A. F., Goossens, L. H. J. and Smit, K. (1999). "Relative importance of maintenance management influences on equipment failure and availability in relation to major hazards. In: Schuëller, G.I., Kafka, P. (Eds.), Safety and Reliability, ESREL, A.A. Balkema, Rotterdam ": 1327-1332.
- Hale, A., Goossens, L., Costa, M., Matos, L., Wielaard, P. and Smit, K. (2000). Expert judgement for the assessment of management influences on risk control. Foresight & precaution. . Cottam MP, H. D., Pape RP, Tait J. Rotterdam: Balkema 1077-1082.
- Hale, A. R. and Guldenmund, F. (2004). ARAMIS audit manual, version 1.3. Safety Science Group, Delft University of Technology, Delft, the Netherlands, Safety Science Group, Delft University of Technology, Delft, the Netherlands.
- Hale, A., Kirwan, B. and Kjellén, U. (2007). "Safe by design: where are we now? ." Safety Science **45**(1-2): 305-327.

- Hale, A. R., Bolt, H. and Walker, D. (2010). Classifying underlying causes of fatalities: the case of construction. In Proceedings of the 10th Probabilistic Safety Assessment and Management Conference. Seattle. June.
- Hancock, P. A., Williams, G. and Manning, C. M. (1995). "Influence of Task Demand Characteristics on Workload and Performance." Aviation Psychology **5**(1): 63-86.
- Hanea, D. M. (2009). Human Risk of Fire:Building a decision support tool using Bayesian networks. Ph.D. thesis, Delft University of Technology.
- Hart, S. G. (1987). The prediction and measurement of mental workload during space operations, NASA Space Life Sciences Symposium, Washington D.C.
- Higgins, E. T. (1997). "Beyond pleasure and pain." American Psychologist **52**: 1280-1300.
- Higgins, E. T. (1998). Promotion and prevention: Regulatory focus as a motivational principle. In M. P. Zanna (Ed.), Advances in experimental social psychology (Vol. 30, pp. 1-46). New York: Academic Press.
- Hirschman, D. (1997). Hijacked: The True Story of the Heroes of Flight 705, Delta Publishing.
- Hobbs, A. and Williamson, A. (2003). "Associations between errors and contributing factors in aircraft maintenance." Human Factors **45**(2): 186-201
- Hourtolou, D. and Salvi, O. (2003). ARAMIS project: development of an integrated accidental risk assessment methodology for industries in the framework of SEVESO II directive. Safety and Reliability - ESREL T., B. and P.H.A.J.M., V. G.**:** 575-581.
- Hurst, N. W., Young, S., Donald, I., Dibson, H., & Muyselaar, A. (1996). "Measures of safety management performance and attitudes to safety at major hazard sites." Loss Prevention in the Process Industries **9**(2): 161-172.
- IATA (2007). IOSA Standards Manual Ed 2, International Air Transport Association.
- ICAO (2000). Accident/incident reporting manual (ADREP). International Civil Aviation Organization, Montreal, Canada
- ICAO (2002). "International Civil Aviation Organization, Line Operations Safety Audit (LOSA), DOC 9803-AN/761, Montreal, Canada."
- ICAO (2009). Safety management manual (SMM). International Civil Aviation Organization, Montreal, Canada
- Isaac, A., Shorrock, S. T. and Kirwan, B. (2002). "Human error in European air traffic management: the HERA project." Reliability Engineering and System Safety **75**(2): 257-272.
- Jensen, F. V. (1996). An introduction to Bayesian networks, Springer.
- Kanfer, R. (1990). Motivation theory and organizational psychology. In M. D. Dunnette & L. Hough (Eds.), Handbook of industrial and organizational psychology (2nd ed., Vol. 1, pp. 75-170). Palo Alto, CA: Consulting Psychologists Press.
- Kendall, M. (1962). Rank correlation methods, London: Charles Griffin & Co. Ltd.
- Kennedy, R. and Kirwan, B. (1998). "Development of a Hazard and Operability-based method for identifying safety management vulnerabilities in high risk systems." SAFETY SCIENCE **30**(3): 249-274.
- Kinnison, H. A. (2004). Aviation Maintenance Management, McGraw-Hill, New York.
- Kirwan, B. (1994). A guide to practical HRA London, Taylor and Francis.
- Kirwan, B. (2007). "Safety informing design " Safety Science **45**(1-2): 155-197.
- Kirwan, B. and Ainsworth, L. A. (1992). A guide to task analysis. London, Taylor and Francis

- Kjellén, U. (2007). "Safety in the design of offshore platforms: Integrated safety versus safety as an add-on characteristic." Safety Science **45**(1-2): 107-127.
- Klinect, J. (2005). Line operations safety audit: a cockpit observation methodology for monitoring commercial airline safety performance. PhD. thesis, University of Texas.
- Kroes, M. J., Watkins, W. A. and Delp, F. (1993). Aircraft maintenance and repair McGraw-Hill, New York.
- Kurowicka, D. and Cooke, R. M. (2004). Non-Parametric continuous Bayesian belief nets with expert judgment. Proceedings of the 4th International Conference on Probabilistic Safety Assessment and Management, New York: Springer.
- Leveson, N. (2004). "A new accident model for engineering safer systems." SAFETY SCIENCE **42**(4): 237-270.
- Locke, E. A. (1997). The motivation to work: What we know. In M. L. Maehr & P. R. Pintrich (Eds.), Advances in motivation and achievement (Vol. 10, pp. 375-412), Greenwich, CT: JAI Press.
- LOSA Collaborative (2007). Threat and Error Code Book. Austin, Texas.
- Mann, M. B. (1999). NASA Statement on Pilot Fatigue. Hearing on Pilot Fatigue before the Aviation Subcommittee of the Committee on Transportation and Infrastructure. US House of Representatives.
- Mearns, K., Whitaker, S. M. and Flin, R. (2003). "Safety Climate, safety management practice and safety performance in offshore environments." Safety Science **41**: 641-680.
- Meyer, J. P., Becker, T. E. and Vandenberghe, C. (2004). "Employee commitment and motivation: A conceptual analysis and integrative model." Applied Psychology **89**(6): 991-1007.
- Meyer, J. P. and Herscovitch, L. (2001). "Commitment in the workplace: Toward a general model." Human Resource Management Review **11**: 299-326.
- Miller, G. A. (1956). "The magical number seven, plus or minus two: some limits on our capacity for processing information." Psychological Review **63**(2): 81-97.
- Mohaghegh, Z. (2007). On the theoretical foundations and principles of organizational safety risk analysis. PhD. thesis, University of Maryland.
- Mohaghegh, Z. and Mosleh, A. (2009). "Incorporating organizational factors into probabilistic risk assessment of complex socio-technical systems: Principles and theoretical foundations." Safety Science **47**(8): 1139-1158.
- Morales Nápoles, O. (2010). Bayesian belief nets and vines in aviation safety and other applications. PhD. thesis, Delft University of Technology.
- Morales, O., Kurowicka, D. and Roelen, A. (2008). "Eliciting conditional and unconditional rank correlations from conditional probabilities." Reliability Engineering & System Safety **93**(5): 699-710.
- Mosleh, A., Goldfeiz, E. and Shen, S. (1997). The ω-factor approach for modeling the influence of organizational factors in robabilistic safety assessment. IEEE six annual human factors meeting. Orlando, FL , USA
- NASA (1986). Report of the Presidential Commission on the Space Shuttle Challenger Accident. National aeronautics and space administration, Washington, D.C.
- Nordhaus, W. D. (1973). "World dynamics: measurement without data." Economic Journal **83(**332).
- NTSB (1994a). Aircraft accident report. NTSB/AAR-99/01, Nation Transportaion Safety Board, Washington, D.C.

- NTSB (1994b). Aircraft accident report: American Airlines, Inc.,DC-10-10, N110AA, Chicago-O'Hare International airport, Chicago, Illinois. National Transportaion Safety Board, Washington, D.C.
- NTSB (2001). Aircraft accident report. NTSB/AAR-04/04, Nation Transportaion Safety Board, Washington, D.C.
- Official Journal of the European Union (2003). Directive 2003/42/EC of the European Parliament and of the Council of 13 June 2003 on occurrence reporting in civil aviation, OJ L 167, 04.07.2003.
- OHSAS 18001 (1999). Occupational Health and Safety Series Specification. London: British Standards Institution.
- Øien, K. (2001). "A framework for the establishment of organizational risk indicators." Reliability Engineering & System Safety **74**(2): 147-167.
- Ostroff, C. (1995). "Best practices." Human Resource Management: Ideas and Trends in Personnel(356).
- Paté-Cornell, M. E. and Murphy, D. M. (1996). "Human and management factors in probabilistic risk analysis: The SAM approach and observations from recent applications." Reliability Engineering and System Safety **53**(2): 115-126.
- Pinder, C. C. (1998). Motivation in work organizations, Upper Saddle River, NJ: Prentice Hall.
- Rankin, W., Hibit, R., Allen, J. and Sargent, R. (2000). "Development and evaluation of the Maintenance Error Decision Aid (MEDA) process." International Journal of Industrial Ergonomics **26**(2): 261-276.
- Rasmussen, J. (1997). "Risk management in a dynamic society: A modelling problem." Safety Science **27**: 183-213.
- Rasmussen, J. (1982). "Human errors. A taxonomy for describing human malfunction in industrial installations." Occupational Accidents **4**(2-4): 311-333.
- Rasmussen, J. and Svedung, I. (2000). Proactive risk management in a dynamic society. Swedish Rescue Services Agency, Karlstad, Sweden. Karlstad, Sweden, Swedish Rescue Services Agency.
- Reason, J. (1990). Human Error Cambridge University Press.
- Reddy, A. V. (2004). Investigation of aeronautical and engineering component failures, CRC Press.
- RIVM (2008). The quantification of occupational risk: The development of a risk assessment model and software. Report 620801001/2008. National Institute for Public Health and the Environment.
- Rodić, L. j. (2000). Reliability of landfill technology. Ph.D. thesis. Delft University of Technology.
- Roelen, A. L. C. (2008). Causal risk models of air transport-comparsion of user needs and model capabilities. PhD. thesis, Delft University of Technology.
- Roelen, A. L. C., Bellamy, L. J., Hale, A. R., Molemaker, R. J. and van Paassen, M. M. (2001). Feasibility of the development of a causal model for the assessment of third party risk around airports. National Aerospace Laboratory (NLR), Amsterdam, the Netherlands. Amsterdam, National Aerospace Laboratory (NLR).
- Roelen, A. L. C., van Baren, G. B., Smeltink, J. W., Lin, P. H. and Morales, O. (2007). A generic flight crew performance model for application in a causal model of air transport. NLR-CR-2007-562, National Aerospace Laboratory, the Netherlands.
- Roelen, A. L. C. and Wever, R. (2002). An analysis of flight crew response to system failures, PT-1 Flight crew intervention credit in system safety assessment phase 1 report, NLRCR-2002-547-PT-1, NLR Amsterdam.

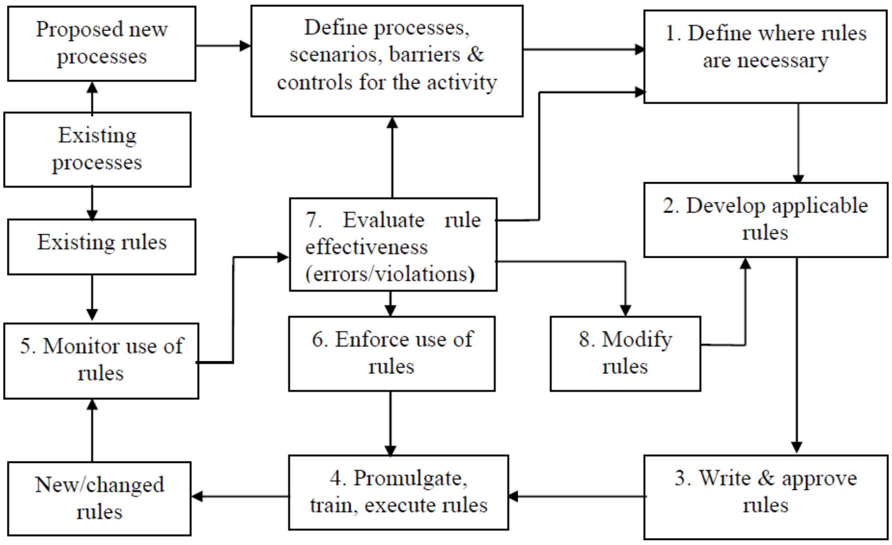
- Rosekind, M. R., Gregory, K. B. and Mallis, M. M. (2006). "Alertness management in aviation operations: enhancing performance and sleep." Aviation, Space, and Environmental Medicine **77**: 1256-1265.
- Seaver, D. and Stillwell, W. (1983). Procedures for Using Expert Judgment to Estimate Human Error Probabilities in Nuclear Power Plant Operations. In: NUREG/CR-2743, U.S. Nuclear Regulatory Commission, Washington, D.C.
- Shappell, S. and Wiegmann, D. (1996). "U. S. Naval Aviation mishaps 1977-92: Differences between single- and dual-piloted aircraft." Aviation, Space, and Environmental Medicine **67**: 65-69.
- Shaw, L. S. and Sichel, H. S. (1971). Accident proneness: Research in the occurrence, causation, and prevention of road accidents, Pergamon.
- Shorrock, S. T. and Kirwan, B. (2002). "Development and application of a human error identification tool for air traffic control." Applied Ergonomics **33**: 319-336.
- Siegel, S. (1956). Nonparametric statistics. , New York, NY: McGraw-Hill
- Simons, M. and 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. and 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. and 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.
- Simons, M., Valk, P. J. L., de Ree, J. J. D., Veldhuijzen van Zanten, O. B. A. and 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.
- Stein, E. S. and Rosenberg, B. L. (1983). The measurement of pilot workload, Report No. DOT/FAA/EM-81/14, FAA Technical Center, Atlantic City Airport, New Jersey, USA.
- Sterman, J. D. (2000). Business Dynamics: Systems thinking and modeling for a complex world, Graw Hill.
- Stolzer, A. J., Halford, C. D. and Goglia, J. J. (2008). Safety Management Systems in Aviation Ashgate.
- Swain, A. D. and Guttmann, H. E. (1983). A handbook of human reliability analysis with emphasis on nuclear power plant applications. NUREG/CR-1278, Sandia National Laboratories, Albuquerque, NM.
- Technica (1988). The Manager Technique. Management Safety Systems Assessment Guidelines in the Evaluation of Risk. London.
- Thurstone, L. L. (1927). "A law of comparative judgment." Psychological Review **34**(4): 273-286.
- Trucco, P., Cagno, E., Ruggeri, F. and Grande, O. (2008). "A Bayesian Belief Network modelling of organisational factors in risk analysis: A case study in maritime transportation." Reliability Engineering & System Safety **93**(6): 845-856.
- Valk, P. J. L. and 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.

- Van den Top, J. (2010). Modelling Risk Control Measures in Railways: Analysing how designers and operators organise safe rail traffic. PhD. thesis, Delft University of Technology.
- Wesensten, N. J., Belenky, G., Thorne, D. R., Kautz, M. A. and Balkin, T. J. (2004). "Modafinil versus caffeine: Effects on fatigue during sleep deprivation." Aviat Space Environ Med **75**: 520-525.
- Wickens, C. (1984). Engineering psychology and human performance. Columbus, OH, USA, Charles E. Merrill.
- Wickens, C. (1992). Engineering psychology and human performance. New York, Harper-Collins.
- Wiegmann, D. A. and Shappell, S. A. (2001). A human error analysis of commercial aviation accidents using the Human Factors Analysis and Classification System (HFACS). DOT/FAA/AM-01/3, U.S. Department of Transportation, Federal Aviation Administration, U.S. Department of Transportation, Federal Aviation Administration,
- Wiegmann, D. A. and Shappell, S. A. (2003). A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System, Ashgate.
- Wilpert, B. (2000). Organizational factors in nuclear safety, Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 5), pp. 1251-1265. Kondo S., Furuta K., editors. Osaka, Japan: Universal Academy Press.
- Yu, J., Ahn, N. and Jae, M. (2004). "A quantitative assessment of organizational factors affecting safety using system dynamics model." Journal of the Korean Nuclear Society **36** (1): 64-72.

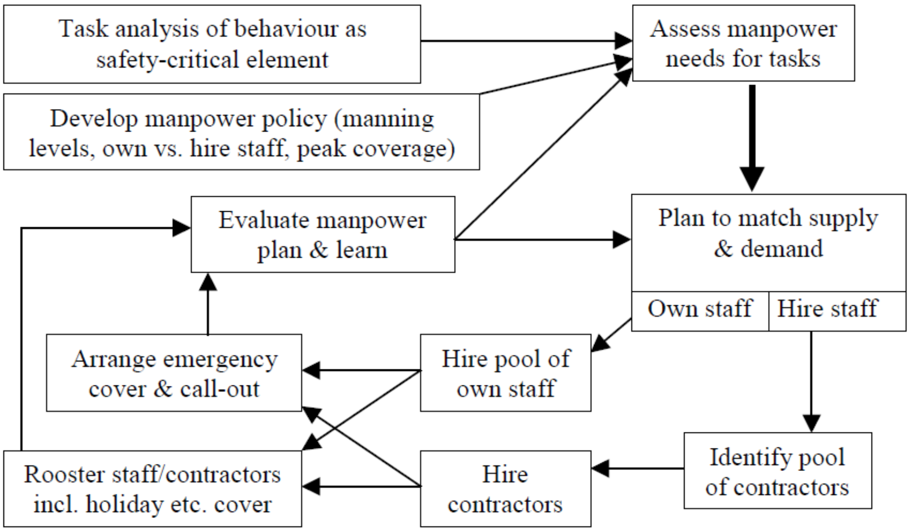
Appendices

Appendix A: Definition of delivery systems of ARAMIS (Hale & Guldenmund, 2004)

- **Procedures, plans, rules and goals:** procedures and rules are specific performance criteria, which specify in detail, usually in written form, a formalised "normative" behaviour or method for carrying out an activity (checklist, task list, action steps, plan, instruction manual, fault-finding heuristic, form to be completed, etc.).

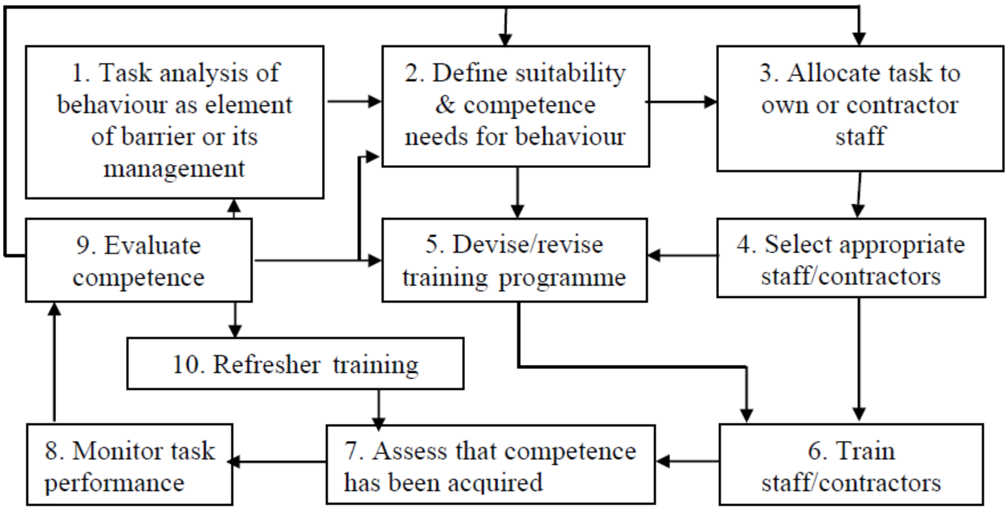

Proposed new processes
Define processes, scenarios, barriers & controls for the activity
1. Define where rules are necessary
Existing processes
Existing rules
7. Evaluate rule effectiveness (errors/violations)
2. Develop applicable rules
5. Monitor use of rules
6. Enforce use of rules
8. Modify rules
New/changed rules
4. Promulgate, train, execute rules
3. Write & approve rules

- **Availability, manpower planning:** covers allocating the necessary time (or numbers) of competent people to the tasks that have to be carried out, at the moment (or within the time frame) when they should be carried out. It also covers the process of planning and allocation of tasks over time, including coverage for:
  - Holidays
  - Sick leave
  - Peak loads
  - Ensuring breaks and rest pauses
  - Limiting overtime and fatigue, incl. in relation to time zone changes


Task analysis of behaviour as safety-critical element
Assess manpower needs for tasks
Develop manpower policy (manning levels, own vs. hire staff, peak coverage)
Evaluate manpower plan & learn
Plan to match supply & demand
Own staff
Hire staff
Arrange emergency cover & call-out
Hire pool of own staff
Rooster staff/contractors incl. holiday etc. cover
Hire contractors
Identify pool of contractors

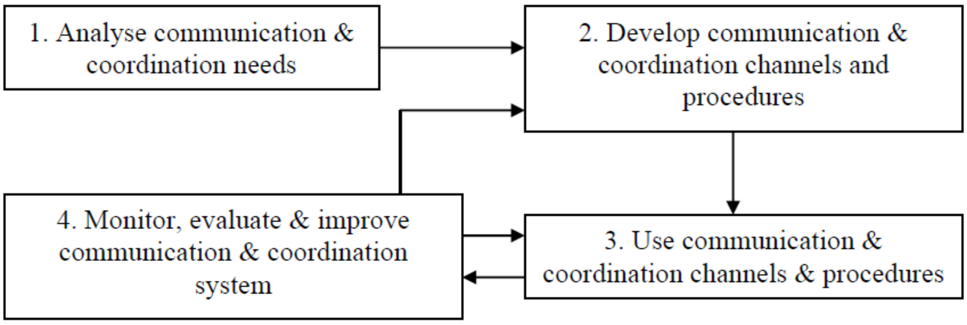
- **Competence:** The knowledge, skills, and abilities of on-line and/or back-up personnel for the safe execution of safety-critical tasks related to risk control. Competence covers the cognitive aspects of behaviour, which can be learned through training, experience and practice.

  Suitability: Physical attributes that are usually more permanent characteristics of an individual, though some can be modified or compensated for over the longer term. They include:
  - Size, strength, dexterity
  - Physical condition, health
  - Visual acuity, colour blindness, hearing


1. Task analysis of behaviour as element of barrier or its management
2. Define suitability & competence needs for behaviour
3. Allocate task to own or contractor staff
4. Select appropriate staff/contractors
5. Devise/revise training programme
6. Train staff/contractors
7. Assess that competence has been acquired
8. Monitor task performance
9. Evaluate competence
10. Refresher training

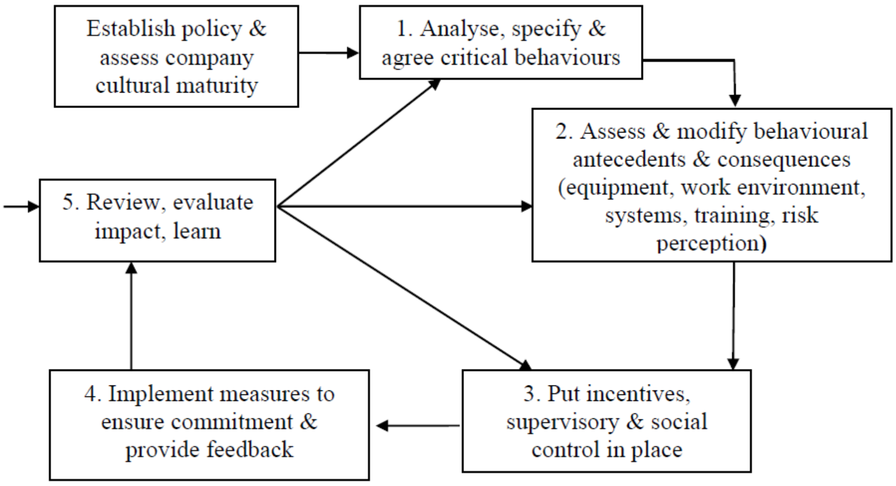
- **Communications:** are those communications, which occur between people within any primary business activity. They are only relevant to this protocol if the activity related to a risk control measure and its functioning is carried out by more than one person (or group), who have to coordinate or plan joint activities. Communications between tasks that are represented in the other parts of the management system are not included here, since they are represented by the continuity of activity within those delivery systems and protocols. Communication occurs either face-to-face, or through communication channels such as (mobile) telephone, data link, radio, e-mail, memo. It can be spoken or written.

  Coordination: covers those mechanisms designed to ensure the smooth interaction of actions between individuals and groups working on a joint task or responsible for the correct functioning of a given risk control measure. These include plans, meetings, authorisation and communication procedures and supervision.


1. Analyse communication & coordination needs
2. Develop communication & coordination channels and procedures
3. Use communication & coordination channels & procedures
4. Monitor, evaluate & improve communication & coordination system

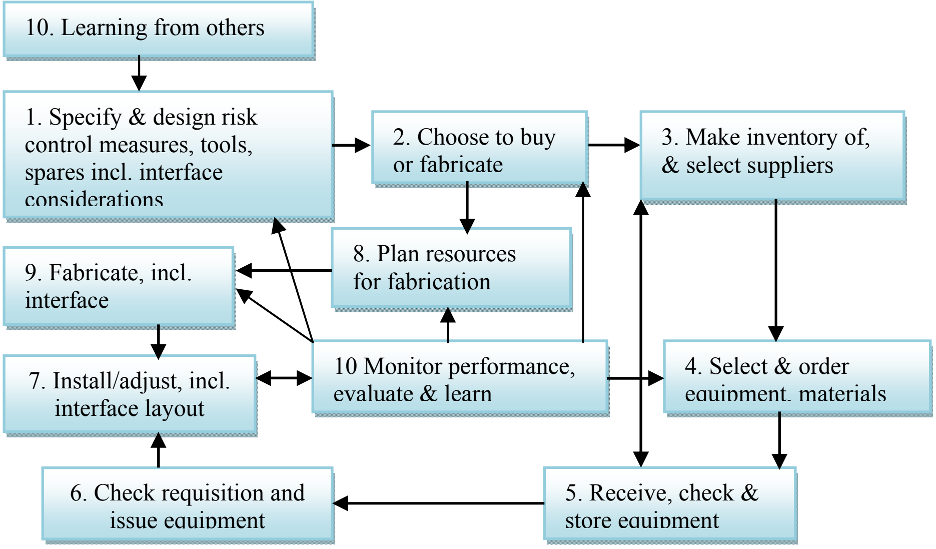
- **Commitment:** the systems for provision of the incentives and motivation that personnel have and/or need to carry out their tasks and activities according to the appropriate safety criteria and procedures. These criteria may be specified by the organisation, derived from external sources (legal, societal criteria) or generated by individuals and/or groups within the discretion allowed to them in the system.

  Conflict resolution: the mechanisms (such as supervision, monitoring, decision-making procedures) by which potential and actual conflicts between safety and other criteria in the allocation and use of personnel, hardware and other resources are recognised, avoided or resolved if they occur.



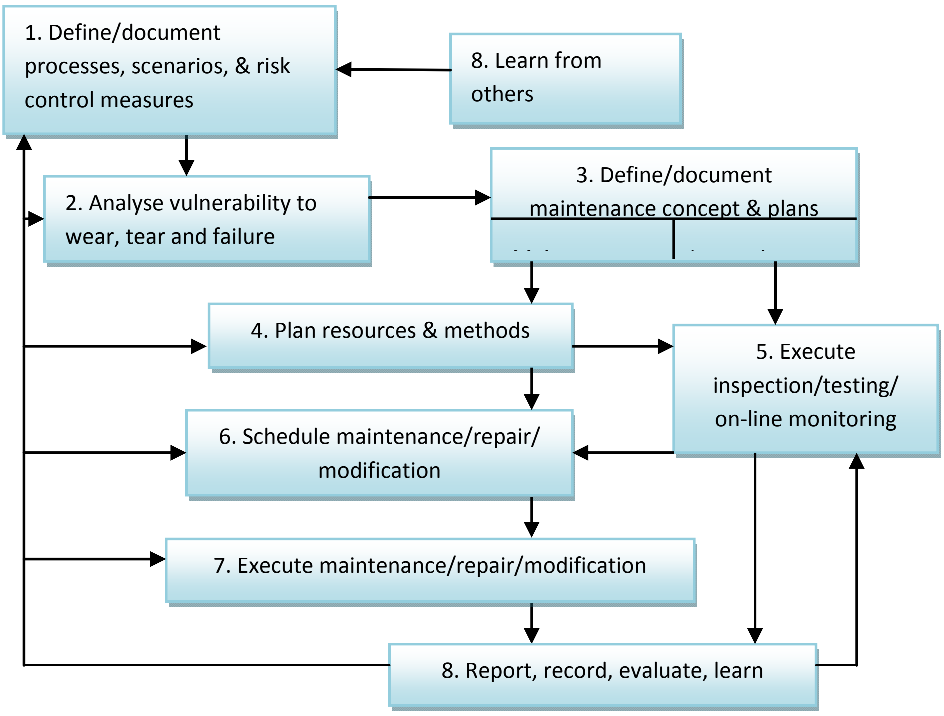
Establish policy & assess company cultural maturity
1. Analyse, specify & agree critical behaviours
2. Assess & modify behavioural antecedents & consequences (equipment, work environment, systems, training, risk perception)
5. Review, evaluate impact, learn
4. Implement measures to ensure commitment & provide feedback
3. Put incentives, supervisory & social control in place

- **Management of equipment and interface specification, design, purchase, installation and adjustment:** covers all hardware and software which has a function within any risk control measure designed to fulfil a safety function in the system. It forms the part of the life cycle from the specification and design of the risk control measures, up to the point where the risk control measure or element has been installed and adjusted and is ready for use or functioning. It also covers the purchase, storage and issue of spare parts and replacements, which will be used in the maintenance and modification phase of the risk control measures and elements. It also covers the step of recording the performance of all of the hardware and software covered, so that its functioning can be assessed and evaluated for the learning process.



10. Learning from others
1. Specify & design risk control measures, tools, spares incl. interface considerations
2. Choose to buy or fabricate
3. Make inventory of, & select suppliers
8. Plan resources for fabrication
9. Fabricate, incl. interface
10 Monitor performance, evaluate & learn
4. Select & order equipment, materials
7. Install/adjust, incl. interface layout
6. Check requisition and issue equipment
5. Receive, check & store equipment

- **Management of inspection, testing and maintenance of hardware and interface:** covers all hardware and software which has a function within any risk control measure designed to fulfil a safety function in the system. It forms the part of the life cycle of these risk control measure elements from the point where they have been installed and adjusted and are ready for use. It covers all the activities which monitor the working of the risk control measures, detect the (chance of) deviation from the designed working and identify the need for work to be done to restore the functioning or replace the risk control measure (elements) with new ones. This process also manages modifications. These can be divided into small modifications, which are carried out at the same time as, and under the same management as the maintenance activities, and modifications of a more major type, which should be dealt with by a change management process.


1. Define/document processes, scenarios, & risk control measures
8. Learn from others
2. Analyse vulnerability to wear, tear and failure
3. Define/document maintenance concept & plans
4. Plan resources & methods
5. Execute inspection/testing/ on-line monitoring
6. Schedule maintenance/repair/ modification
7. Execute maintenance/repair/modification
8. Report, record, evaluate, learn

Appendix B: Selected example of unsafe acts and preconditions in HFACS (Wiegmann & Shappell , 2003)

HFACS unsafe acts of pilot operators	
Errors	**Violations**
Skill-based errors	**Routine**
Breakdown in visual scan	Flew an unauthorized approach
Inadvertent use of flight controls	Violated training rules
Poor technique	Failed to use ATC radar advisories
Over-controlled the aircraft	Inadequate briefing for flight
Omitted checklist item	Failed to comply with departmental manuals
Failed to prioritize attention	Violation of orders, regulations, SOPs
Omitted step in procedure	Failed to inspect aircraft in-flight caution light
Decision Errors	**Exceptional**
Improper procedure	Performed unauthorized acrobatic maneuver
Misdiagnosed emergency	Improper takeoff technique
Wrong response to emergency	Failed to obtain valid weather brief
Inadequate knowledge of systems, procedures	Exceeded limits of aircraft
Inappropriate maneuver	Accepted unnecessary hazard
Perceptual Errors	Not current/qualified for flight
Due to misjudged distance/altitude/airspeed	
Due to spatial disorientation	
Due to visual illusion	

HFACS unsafe aircrew condition	
Adverse Mental Sates	**Crew resource management**
Loss of situational awareness	Failed to conduct adequate brief
Complacency	Lack of teamwork
Stress	Lack of assertiveness
Overconfidence	Poor communication/coordination within &between aircraft, ATC, etc.
Poor flight vigilance	Misinterpretation of traffic calls
Task saturation	Failure of leadership
Alertness	**Personal Readiness**
Mental fatigue	Failure to adhere to crew rest requirements
Distraction	Inadequate training
Channelized attention	Self-medicating
Adverse Physiological States	Overexertion while off duty
Medical illness	Poor dietary practices
Hypoxia	Pattern of poor risk judgment
Physical fatigue	**Physical Environment**
Intoxication	Weather
Motion sickness	Altitude
Physical/mental limitation	Terrain
Visual limitations	Lighting
Insufficient reaction time	Vibration
Information overload	Toxins in the cockpit
Inadequate experience for complexity of situation	**Technological Environment**
Incompatible physical capabilities	Equipment/controls design
Lack of aptitude to fly	Checklist layout
Lack of sensory input	Automation

Appendix C: Human resource functions (Ostroff, 1995)

Selectivity in Recruiting/Hiring (SE)

- Examining various recruiting sources (e.g. want ads, employee referrals, colleges) to determine which provide the most appropriate employees
- Provide information to job applicants that realistically describes the job and company (positive as well as negative aspects)
- Regularly conduct validation studies in the tests, predictors or hiring practices used
- Use hiring procedures or tests that are based on job duties and requirements
- Use hiring procedures or tests to determine who will best fit in with the company's culture and values

Internal Staffing (ST)

- Fill non-entry level position from among present employees who desire promotion or transfers

Contingent Workforce (CW)

- Use nonpermanent workers (e.g. temps, contractors, retirees) in managerial related jobs
- Use nonpermanent workers in professional, non-managerial jobs
- Use nonpermanent workers in low-level jobs, such as secretarial, custodial,
- etc.

Training and Employee Development (TR)

- Conduct formal analyses to determine the training needs throughout the company
- Develop clear specific objectives for what is to be learned in training programs
- Determine the most appropriate method (e.g., lecture, role-playing, hands-on) for teaching particular skills in training program
- Provide training (inside or outside the company) to keep employees' skills up-to-date
- Develop mechanisms to that employees are supported or rewarded for using their newly learned skills on the job
- Provide remedial or basic skills training fir those employees who need it
- Provide programs (e.g. training, mentoring, job rotation) to develop new skills and prepare employee for variety of jobs in the company
- Develop career plans and paths for employee movements in the company
- Counsel or meet with employees to discuss their own career goals and realistic career options
- Have formal orientation programs that provide new employees with information about the job

Appraisal (AP)

- Regularly (at least once a year) conduct appraisals of employees' performance
- Have supervisors/mangers meet with individual employees to give developmental performance feedback
- Develop performance appraisal forms that focus on the relevant duties and specific skills requires for successful job performance
- Train mangers in conducting accurate performance appraisals and giving employees feedback

Compensation and Reward Systems (RE)

<u>Job Based Pay</u>

- Determine pay levels for each job category using a formal job evolution system to compare and order jobs based on skills levels and/or experience
- Determine pay levels for each job or jobs category based on information about the "going rate" in the market

- Group jobs into pay classes or pay grades and determine a pay range for each class
- Formally analyze and determine the most appropriate mix of direct pay and benefits

Individual Merit Rewards

- Link individual employees' rewards, raised or bonuses to how well they perform the job

Contingent Rewards

- Link individual employees' rewards, raises or bonuses to how well the unit or team performs
- Regularly evaluate whether productivity goals and quality standards are being met
- Provide incentives to employees to increase productivity or quality

Organizational-Based Reward

- Use reward and compensation programs that link employees' rewards to how well the company performs (e.g. profit sharing, employee stock ownership plans)

Skill-Based Pay

- Base individual employees' raises/bonuses on a skill-based pay system

Pay Leader

- Adhere to pay policy of being a pay leader (high paying) in the industry or area

Non-financial Rewards

- Encouraging managers/supervisors to use non-financial rewards such as recognition, praise, etc.

Benefit

- Provide health retirement insurance and other benefits to employees
- Have procedures to assist employees in understanding their benefits

Job Analysis (JA)

- Conduct job analyses that describe the tasks performed, behaviors, abilities, knowledge and skills needed, and equipment required to perform the job
- Update job analysis information on a regular basis
- Use standardized, systematic procedures to collect job analysis information

Job Enrichment (JE)

- Design jobs to provide employees with sufficient variety, autonomy and feedback

Team Systems (TS)

- Establish committees/teams of employees who examine productivity and quality problems and provide recommendations for changes
- Utilize autonomous workgroups or self-managed teams who have responsibilities for decisions assigning work, and determining work methods
- Use a total quality management approach to improve productivity and service

Employee Assistance (EA)

- Offer employee assistant programs to help employees deal with personal job related issues such as stress, family problems, substance abuse and financial counseling
- Sponsor or provide fitness programs for employees, such as athletic programs or fitness clubs
- Use alternative work schedule, such as flexible hours, job sharing, part time work or work at home
- Use a flexible benefits package that gives employees in allocating their benefits "dollars" across health, retirement, insurance, child care, etc.
- Provide programs or benefits to help employees balance work and family concerns such as childcare, elder care, referral networks, childcare sick leave, etc.
- Provide outplacement services, such as counseling and job search skills, for employee who are discharged or laid off

Due Process (DP)

- Have a formal grievance procedure or formal complaint resolution system for employees
- Adhere to progressive discipline system in which employees are disciplined in successive steps ranging an oral warning to eventual dismissal
- Have mechanism in place for employees to communicate suggestions or register complaints

Employee voice/Empowerment (EM)

- Have formal procedures for sharing important information with employees
- Involve employees in design and administration of compensation systems, performance evaluation systems, methods for enhancing productivity, etc.
- Regularly survey the opinions of workers regarding their job conditions and satisfaction
- Involve employees in major decisions that will directly affect their work processes

Diversity (DI)

- Provide transition or other programs for employees to understand and accept members from other culture, ethnic background of gender groups
- Develop programs to increase promotion rate of members of protected classes
- Establish goals, time tables and/or other procedures to increase minority representation and diversity in the company
- Conduct adverse impact analyses or analyses to determine if discrimination against members of protected classes exists in hiring or promotion practices

Legal compliance (LC)

- Regularly check for compliance with laws pertaining to discriminations and disabilities
- Regularly check for compliance with laws pertaining to employee safety
- Regularly check for compliance with laws pertaining to employee rights
- Regularly check for compliance with laws pertaining to pay, compensation and benefit

Safety (SA)

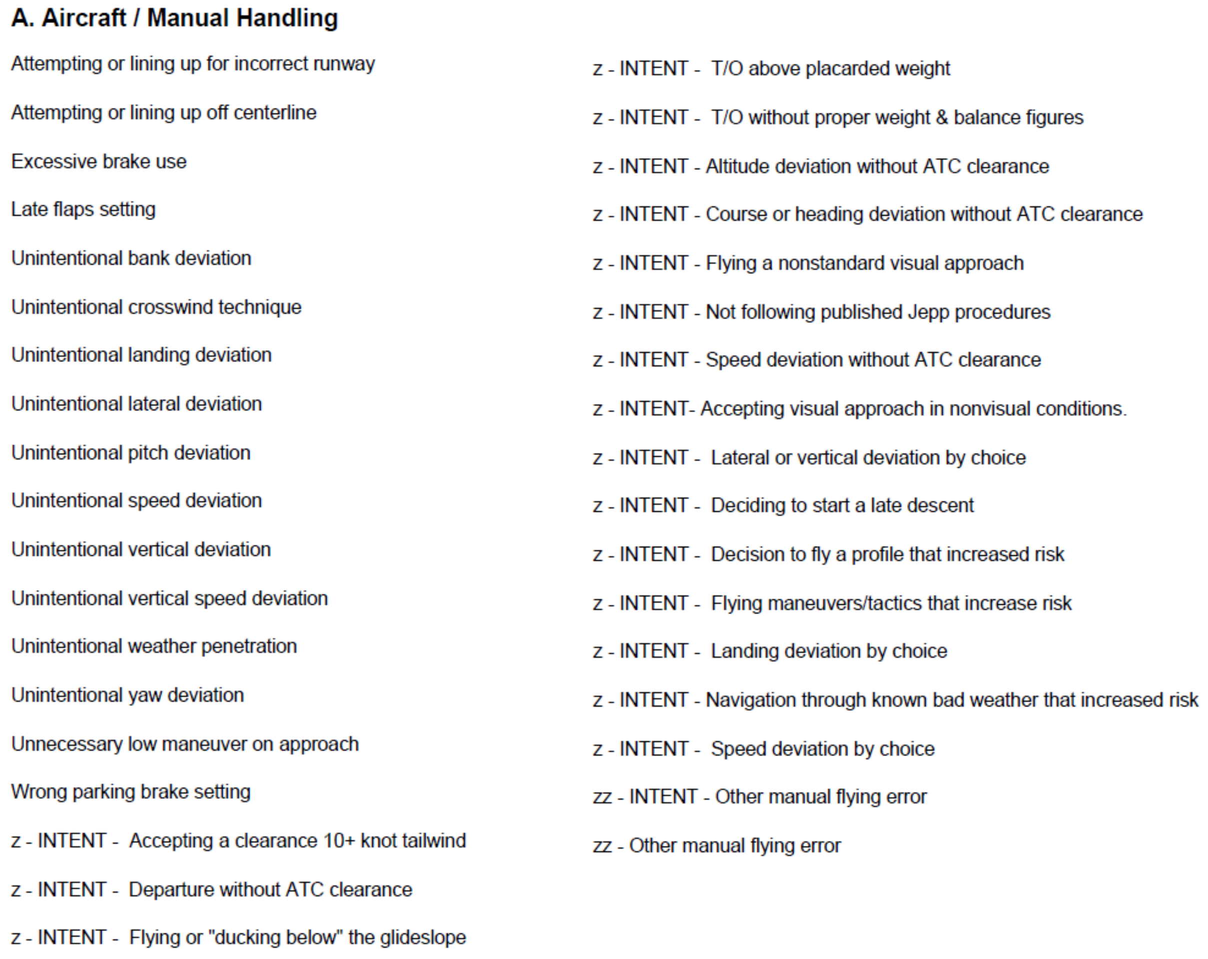
- Maintain an accident record system or use committees of workers or causes of accident and safety hazards
- Train employees to emphasis safe practices in the work place
- Conduct internal safety inspection

Union Relations (UR)

- Employees unionized labor
- Monitor the number of NLRB grievances filed
- Efficiently settle collective bargaining contracts for unionized employees
- Share information with union representatives regarding the companies financial status, work conditions and potential procedural changes

Appendix D: LOSA “error” codes and “threat” codes (LOSA Collaborative, 2007)

Errors



A. Aircraft / Manual Handling

Attempting or lining up for incorrect runway

Attempting or lining up off centerline

Excessive brake use

Late flaps setting

Unintentional bank deviation

Unintentional crosswind technique

Unintentional landing deviation

Unintentional lateral deviation

Unintentional pitch deviation

Unintentional speed deviation

Unintentional vertical deviation

Unintentional vertical speed deviation

Unintentional weather penetration

Unintentional yaw deviation

Unnecessary low maneuver on approach

Wrong parking brake setting

z - INTENT - Accepting a clearance 10+ knot tailwind

z - INTENT - Departure without ATC clearance

z - INTENT - Flying or "ducking below" the glideslope

z - INTENT - T/O above placarded weight

z - INTENT - T/O without proper weight & balance figures

z - INTENT - Altitude deviation without ATC clearance

z - INTENT - Course or heading deviation without ATC clearance

z - INTENT - Flying a nonstandard visual approach

z - INTENT - Not following published Jepp procedures

z - INTENT - Speed deviation without ATC clearance

z - INTENT- Accepting visual approach in nonvisual conditions.

z - INTENT - Lateral or vertical deviation by choice

z - INTENT - Deciding to start a late descent

z - INTENT - Decision to fly a profile that increased risk

z - INTENT - Flying maneuvers/tactics that increase risk

z - INTENT - Landing deviation by choice

z - INTENT - Navigation through known bad weather that increased risk

z - INTENT - Speed deviation by choice

zz - INTENT - Other manual flying error

zz - Other manual flying error

B. Automation

Failure to execute a FMC/FMGC mode when needed

Failure to execute a MCP/FCU mode when needed

Failure to use flight directors (FD)

Improper use of automation

Improper use of flight directors

Manual aircraft control with MCP/FCU mode engaged

Omitted / wrong waypoint or route settings entered in FMC/FMGC

Other MCP/FCU error

Other wrong FMC/FMGC entries

Wrong altitude entered into the FMC/FMGC

Wrong approach selected in FMC/FMGC

Wrong autothrottle setting

Wrong FMC/FMGC format for input

Wrong FMC/FMGC page displayed

Wrong MCP/FCU altitude setting dialed

Wrong MCP/FCU course setting dialed

Wrong MCP/FCU heading set or dialed

Wrong MCP/FCU mode executed

Wrong MCP/FCU mode left engaged

Wrong MCP/FCU navigation select setting

Wrong MCP/FCU speed setting dialed

Wrong MCP/FCU vertical speed / flight path angle

Wrong mode executed in the FMC/FMGC

Wrong mode left engaged in the FMC/FMGC

Wrong present position entered into the FMC/FMGC

Wrong setting on the MCP/FCU autopilot or flight director switch

Wrong speed setting entered into the FMC/FMGC

Wrong weight and balance calculations entered into FMC/FMGC

z - INTENT - Discretionary omission of FMC data

z - INTENT - Nonstandard or wrong MCP/FCU settings

z - INTENT - Nonstandard automation usage

zz - INTENT - Other automation error

zz - Other automation error

C. Flight Controls

Attempting to use INOP controls

Decision to use wrong thrust / power

Failure to engage thrust reversers on landing

Failure to raise or lower landing gear on schedule

Other incorrect switch or lever settings

Wrong autobrake setting

Wrong flaps setting

Wrong speed brakes setting

Wrong spoilers setting

Wrong stab trim settings

Wrong thrust / power settings

Wrong thrust reverser setting

z - INTENT - Failure to arm spoilers

z - INTENT - Use of excessive power on pushback

zz - INTENT - Other flight control error

zz - Other flight control error

D. Systems / Inst / Radio

Failure to respond to GPWS warnings

Failure to respond to TCAS warnings

Failure to set altitude alerter

Incorrect nav display setting

Lack of weather radar use

Wrong ACARS entries

Wrong altimeter settings

Wrong anti-ice setting

Wrong ATC frequency dialed / selected

Wrong ATIS frequency dialed

Wrong bug settings

Wrong display switch setting

Wrong exterior lights/beacon setting

Wrong fuel switch setting

Wrong nav radio frequency dialed

Wrong packs setting

Wrong panel setup for engine start

Wrong radar setting for situation

Wrong radar settings

Wrong TCAS setting

Wrong transponder setting

z - INTENT - Failure to respond to GPWS warnings

z - INTENT - Failure to respond to TCAS warnings

z - INTENT - Using equipment placarded as INOP

z - INTENT - Failure to respond to overspeed warning

z - INTENT - Failure to set altitude alerter

z - INTENT - Omitted login to datalink

z - INTENT - Setting altimeters before transition level

z - INTENT - Unauthorized response to aircraft warning

z - INTENT - Wrong bug settings

zz - INTENT - Other systems / inst / radio error

zz - Other systems / inst / radio error

E. Ground Navigation

Attempting or turning down wrong gate/taxiway/ramp/hold spot

Attempting or turning down wrong runway

Attempting to taxi off centerline

Failure to hold short

Missed gate

Missed runway

Missed taxiway

Pushback without clearing right or left

Taxi in position and hold with unready cabin

Taxi on taxiway / runway with oncoming traffic

Taxi too close to other aircraft

Taxi too fast

Taxi without clearing right or left

Unintentional taxi too fast

Wrong thrust setting during taxi

z - INTENT - Taxi too fast

z - INTENT - Taxi deviation by choice

z - INTENT - Taxi deviation without clearance

zz - INTENT - Other ground navigation error

zz - Other ground navigation error

F. Pilot to Pilot Comm

Crew miscommunication of information

Failure to communicate approach information

Misinterpretation of ATIS

Missed command within crew

Wrong airport communicated

Wrong engine out procedures communicated

Wrong gate assignment communicated

Wrong nav aid communicated

Wrong runway communicated

Wrong taxiway/ramp/gate/hold spot communicated

z - INTENT - Sterile cockpit violation

zz - INTENT- Other pilot/pilot communication error

zz - Other pilot/pilot communication error

G. Crew to External Comm

Crew did not repeat ATC clearance

Crew omitted ATC call

Crew omitted cabin / flight attendant call

Failure to communicate pertinent ATC information

Failure to give readbacks or callbacks to ATC

Failure to verify ATC instructions

Incomplete clearance readback

Misinterpretation of ATC instructions

Misinterpretation of ground instructions

Misinterpretation of tower instructions

Misinterpretation of pilot report (PIREP)

Missed ATC calls

Missed instruction to hold short

Missed taxi instruction

Omitted call signs to ATC

Omitted non-radar environment report to ATC

Omitted position report to ATC

Use of nonstandard ATC phraseology

Wrong position report

Wrong readbacks or callbacks to ATC

Z – INTENT - Accepting an unauthorized LAHSO clearance

z - INTENT - Omitted ATC calls

z - INTENT - Omitted call signs to ATC

z - INTENT - Incomplete call signs

z - INTENT - Omitted non-radar environment report to ATC

z - INTENT - Omitted position report to ATC

z - INTENT - Use of nonstandard ATC phraseology

zz - INTENT Other crew/external communication error

zz - Other crew/external communication error

H. Checklists

Checklist not performed to completion

Checklist performed late or at the wrong time

Completed checklist not called 'complete'

Did not call for checklist

Failure to visually verify settings when called for on a checklist

Missed checklist item

Omitted abnormal checklist

Omitted checklist

Wrong checklist performed

Wrong response to a challenge on a checklist

z - INTENT - Checklist performed from memory

z - INTENT - Completed checklist not called 'complete'

z - INTENT - Omitted abnormal checklist

z - INTENT - Use of nonstandard checklist protocol

z - INTENT - Checklist not performed to completion

z - INTENT - Checklist performed as "to-do" checklist

z - INTENT - Checklist performed late or at wrong time

z - INTENT - No challenge and response on checklist

z - INTENT - Omitted checklist

z - INTENT - Self-initiated checklist by PNF

zz - INTENT - Other checklist error

zz - Other checklist error

I. Callouts

Incorrect climb callouts

Incorrect descent/approach callouts

Incorrect V-speed callouts

Omitted altitude callouts

Omitted climb callouts

Omitted descent/approach callouts

Omitted landing callouts

Omitted speed deviation callout

Omitted vertical deviation callout

Omitted V-speed callouts

z - INTENT - Nonstandard descent/approach callouts

z - INTENT - Omitted climb callouts

z - INTENT - Omitted descent/approach callouts

z - INTENT - Nonstandard climb callouts

z - INTENT - Nonstandard V-speed callouts

z - INTENT - Omitted altitude callouts

z - INTENT - Omitted V-speed callouts

z - INTENT - Other nonstandard calls

zz - INTENT - Nonstandard altitude callouts

zz - INTENT Other callout error

zz - Other callout error

J. Briefings

Brief performed late

Incorrect / incomplete approach briefing

Incorrect / incomplete departure/takeoff briefing

Incorrect / incomplete flight attendant briefing

Omitted approach briefing

Omitted departure review / takeoff briefing

Omitted required engine-out briefing

Omitted required F/A briefing

z - INTENT - Incorrect / incomplete approach briefing

z - INTENT - Late briefing

z - INTENT - Omitted departure/takeoff briefing

z - INTENT - Omitted handover briefing

z - INTENT - Omitted required engine-out briefing

z - INTENT - Incorrect / incomplete departure/takeoff briefing

z - INTENT - Incorrect / incomplete flight attendant briefing

z - INTENT - Omitted approach briefing

z - INTENT - Omitted required flight attendant briefing

zz - INTENT Other briefing error

zz - Other briefing error

K. Cross Verification

Failure to clarify MEL or logbook entry

Failure to cross-verify altimeter settings

Failure to cross-verify automation with raw data

Failure to cross-verify clearance

Failure to cross-verify documentation or paperwork

Failure to cross-verify FMC/FMGC inputs

Failure to cross-verify MCP/FCU/altitude settings

Failure to cross-verify speed before flap selection

Failure to cross-verify takeoff figures/calculations

Failure to detect/correct pulled circuit breakers

Failure to monitor engine start

Failure to verify crew actions

Omitted brake check

Omitted flight control check

Omitted flight mode annunciation

Omitted fuel check

Omitted systems check

z - INTENT - Failure to cross-verify paperwork

z - INTENT - Failure to cross-verify altimeter settings

z - INTENT - Failure to cross-verify FMC/FMGC changes

z - INTENT - Failure to cross-verify manual with paperwork

z - INTENT - Failure to cross-verify MCP/FCU altitude settings

z - INTENT - Nonstandard cross-verification

z - INTENT - Omitted flight mode annunciation (FMA)

zz - INTENT Other cross verification error

zz - Other cross verification error

L. Documentation

Incorrect or failing to make an entry into the logbook

Miscalculation of hold times

Misinterpreted items on flight documentation

Missed items on paperwork (flight plan, NOTAM, dispatch release)

Wrong clearance recorded

Wrong or no ATIS information recorded

Wrong or no fuel information recorded

Wrong or no Jepp page/charts in view

Wrong performance chart used

Wrong runway information recorded

Wrong times calculated in flight plan

Wrong V-speeds recorded

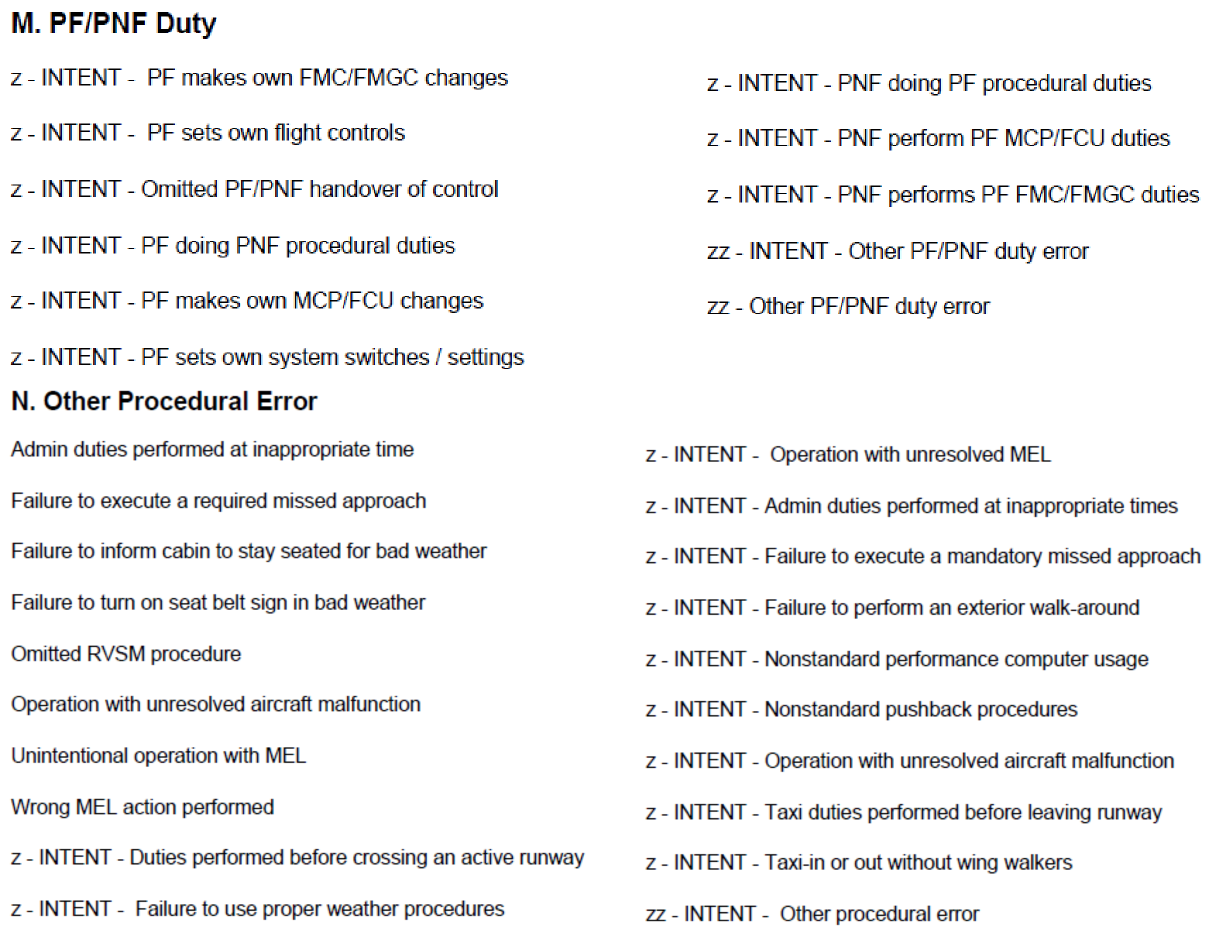
Wrong weight and balance information recorded

z - INTENT - Failure to make logbook entry

z - INTENT - No Jepp pages/charts in view

zz - INTENT - Other documentation error

zz - Other documentation error



M. PF/PNF Duty

z - INTENT - PF makes own FMC/FMGC changes

z - INTENT - PF sets own flight controls

z - INTENT - Omitted PF/PNF handover of control

z - INTENT - PF doing PNF procedural duties

z - INTENT - PF makes own MCP/FCU changes

z - INTENT - PF sets own system switches / settings

z - INTENT - PNF doing PF procedural duties

z - INTENT - PNF perform PF MCP/FCU duties

z - INTENT - PNF performs PF FMC/FMGC duties

zz - INTENT - Other PF/PNF duty error

zz - Other PF/PNF duty error

N. Other Procedural Error

Admin duties performed at inappropriate time

Failure to execute a required missed approach

Failure to inform cabin to stay seated for bad weather

Failure to turn on seat belt sign in bad weather

Omitted RVSM procedure

Operation with unresolved aircraft malfunction

Unintentional operation with MEL

Wrong MEL action performed

z - INTENT - Duties performed before crossing an active runway

z - INTENT - Failure to use proper weather procedures

z - INTENT - Operation with unresolved MEL

z - INTENT - Admin duties performed at inappropriate times

z - INTENT - Failure to execute a mandatory missed approach

z - INTENT - Failure to perform an exterior walk-around

z - INTENT - Nonstandard performance computer usage

z - INTENT - Nonstandard pushback procedures

z - INTENT - Operation with unresolved aircraft malfunction

z - INTENT - Taxi duties performed before leaving runway

z - INTENT - Taxi-in or out without wing walkers

zz - INTENT - Other procedural error

Threats

A. Weather

Crosswind, tailwind, gusty winds or high winds aloft

Icing/snow conditions only

IMC/fog/low visibility conditions

Thunderstorms / turbulence / icing combo

Moderate/severe turbulence only

Windshear

z - Other Weather threat

B. Airport

Airport construction

Contaminated taxiway / runway

Lack of or faded signage / markings

Difficult takeoff/climb or approach airport procedures

Unusable or improperly functioning NAVAIDS

Other runway threats

Other taxiway / ramp threats

z - Other Airport threat

C. ATC

ATC command - challenging clearances, late changes

ATC error

ATC language difficulty

ATC non-standard phraseology

ATC radio congestion

ATC runway change

Similar call signs

z - Other ATC threat

J. Ground Maintenance

Maintenance error

Maintenance event

z - Other Ground Maintenance threat

K. Ground / Ramp

Ground crew error

Ground handling event

z - Other Ground / Ramp threat

D. Environment Ops Pressure

ATIS or ACARS communication

GPWS warning

TCAS TA/RA

Terrain

Traffic (air or ground congestion)

z -Other Environmental Ops Pressure threat

E. Airline Ops Pressure

Crew scheduling event

Operational pressure (delays, OTP, late arriving pilot or aircraft)

z - Other Airline Ops Pressure threat

F. Aircraft

Aircraft malfunction unexpected by crew

Automation event or anomaly

MEL/CDL with operational implications

z - Other Aircraft threat

G. Cabin

Cabin / flight attendant distraction / interruption

Flight attendant error

z - Other Cabin threat

I. Dispatch / Paperwork

Dispatch / paperwork error

Dispatch / paperwork event

z - Other Dispatch / Paperwork threat

L. Manuals / Charts

Chart error

Manual error

z - Other Manuals / Chart threat

M. Other

Other

Appendix E: ADREP and delivery systems mapping

Explanatory factor subject in ADREP		Explanatory factor subject in ADREP	Delivery system failure subject
100000000		**Liverware (human)**	
101000000		Personal physical or sensory limitations	Suitability (physical)
	101010000	Personal physical characteristics	Suitability (physical)
	101020000	Human sensory limitations	Suitability (physical)
	101020100	Vision	Suitability (physical)
	101020700	Sensory threshold	Suitability (physical)
	101030000	Other physical limits	Suitability (physical)
102000000		Human physiology	Suitability (physical)
	102010000	Illness/incapacitation	Suitability (physical)
	102010201	Illness-food poisoning	Suitability (physical)
	102010202	Vertigo/dizziness	Suitability (physical)
	102010400	Heart attack	Suitability (physical)
	102010500	Hypoxia/anoxia	Suitability (physical)
	102010900	Loss of consciousness	Suitability (physical)
	102020000	Human impairment-health/fitness/lifestyle	Suitability (physical)
	102020400	Impairment-food intake	Suitability (physical)
	102021200	Impairment-barred drugs	Suitability (physical)
	102022000	Impairment-psychological	Suitability (psychological)
	102030000	Human fatigue/alertness	Manpower planning and availability
	102030300	Fatigue-chronic	Manpower planning and availability
	102030400	Circadian dysrhythmia	Manpower planning and availability
	102030500	Fatigue-rest/duty time	Manpower planning and availability
	102030900	Fatigue-other	Manpower planning and availability
	102040000	Human vestibular or visual illusions	Competence
	102040200	Human visual illusions	Competence
103000000		Psychological limitations	
	103050000	Psychological-skill/technique/ability	Competence; suitability
	103050200	Psychological-airmanship	Competence
	103050201	Handling of aircraft	Competence
	103050300	Psychological-competence	Competence
	103050500	Psychological-skill	Competence
	103050503	Lack of practice	Competence
	103050700	Reaction time-ability, related to level of skill or ability	Competence; suitability
	103070000	Knowledge acquisition	Competence
	103080000	Situational awareness	Suitability (mental); Competence; Communication
	103080100	Spatial disorientation (e.g. not knowing when the aircraft is straight and level)	Suitability (mental); Competence; Communication
	103080500	Losing the picture (e.g. inexperienced controllers in high workload conditions)	Suitability (mental); Competence; Communication

	103080600	Situational awareness and automation factors (e.g. knowing the mode to which the autopilot is selected)	Suitability (mental); Competence; Communication
	103090000	Personality and attitude factors	Suitability (mental)
	103090302	Confidence in equipment	Suitability (mental)
	103090303	Self confidence problems	Suitability (mental)
	103090400	Complacency factors	Suitability (mental)
	103100000	Mental/emotional state factors	Suitability (mental)
	103100201	Post-incident stress	Suitability (mental)
	103100300	Apprehension problems	Suitability (mental)
	103100400	Personal anxiety problems	Suitability (mental)
	103100500	Personal panic factors	Suitability (mental)
104000000		Personal workload management	Workload
	104010000	Task scheduling	Workload
	104020000	Personal timing of actions	Workload
	104020100	Unforeseen task additions	Workload
	104030000	High workload task shedding	Workload
	104040000	Task allocation	Workload
105000000		Experience & knowledge	Competence
	105010000	**Experience & qualification**	Competence
	105010100	Qualifications	Competence
	105010201	Total hours/years	Competence
	105010202	Experience in position	Competence
	105010203	Experience-aircraft type	Competence
	105010204	Experience of aerodrome	Competence
	105010205	Experience of route	Competence
	105010207	Other experience factors	Competence
	105010300	Use of tools and equipment	Competence
	105020000	**Recency factors**	Competence
	105020200	Recency on aircraft type	Competence
	105020400	Recent experience-route	Competence
	105020600	Experience-operational	Competence
	105030000	**Adequacy of knowledge**	Competence
	105030100	General knowledge	Competence
	105030200	Current knowledge	Competence
	105030300	Regulatory requirements	Competence
	105030500	Aircraft system knowledge	Competence
	105030600	Knowledge of procedures	Competence
	105030601	Company policies	Competence
	105030602	Flight procedures	Competence
	105030603	ATM procedures	Competence
	105030604	Aerodrome procedures	Competence
	105030605	Maintenance procedures	Competence
200000000		**Liveware-environment**	
201000000		Physical environment	
	201010000	Aerodrome/landing/take-off site	Aerodrome- design & maintenance
	201010100	Taxiway/runway characteristics, conditions	Aerodrome- design & maintenance
	201010300	Landing/take-off site infrastructure	Aerodrome- design & maintenance
	201010400	Obstructions to vision on landing site	Aerodrome- design & maintenance
	201010500	Landing take-off site facilities	Aerodrome- design &

			maintenance
	201020000	ATC service	
	201020200	lack of Air Traffic Control (none provided normally or service temporarily suspended for some reason)	Communication and coordination
	201030000	provision of ATS information	
	201030100	ATS weather information (poor, out of date, unavailable)	Communication and coordination
	201040000	Weather/visibility conditions (e.g. disorientation in fog or workload increased by poor weather conditions)	Workload
	201050000	Workspace environment	
	201050300	Visibility from workplace (e.g. Tower window struts obscuring vision or small cockpit window)	Workplace- design
	201051500	Excessive vibration in the work environment	Workplace- design; maintenance
202000000		Psychosocial factors	
	202010000	Job satisfaction	Commitment and motivation
	202020000	Morale/motivation	Commitment and motivation
	202030000	Culture issues	Commitment and motivation
	202040000	Domestic issues (e.g. death of a close relative or divorce)	Suitability (mental)
	202040400	Interpersonal conflicts (e.g. not getting on with a work colleague)	Conflict resolution
203000000		company, management, manning or regulatory issues	
	203010000	Pressure to achieve	Conflict resolution
	203010100	Commercial pressures	Conflict resolution
	203010200	Specific company problem	Conflict resolution
	203010300	Supervision problems (e.g. conflicts between management requirements and operational supervisory responsibilities)	Conflict resolution
	203010400	Managerial operating pressures	Conflict resolution
	203020000	Labour relations factors (factors related to labour relations, e.g. working to rule or strikes)	Conflict resolution
	203020200	Industrial action (e.g. problems arising as a result of ATC strikes)	Conflict resolution
	203030000	Mgmt personnel policy	Conflict resolution
	203030100	Operational personnel policies (e.g. pilots discouraged from making diversions due to cost implications)	Conflict resolution
	203030200	Operational control personnel policies	Conflict resolution
	203030400	Recruitment personnel policies (e.g. recruitment of inappropriate staff)	Suitability; competence
	203030500	Staffing personnel policies (factors related to staffing, general personnel policies and numbers)	Manpower planning and availability
	203030600	Manning/resource allocation personnel policies (the availability of staff and their deployment)	Manpower planning and availability

	203031300	Personnel policies in instructions/directives/orders (e.g. pilots discourage from making go-arounds due to const implications)	Conflict resolution
	203040000	Manning issues	Manpower planning and availability
	203050000	Regulatory authority policies and practice	Regulatory (regulatory are currently not dealt with in our model. They are at higher level)
204000000		operational task demands	
	204010000	Workload task demands	Workload
	204010100	Work overload/task saturation	Workload
	204010800	Additional workload due to unusual/unfamiliar	Workload
	204011000	Additional workload due to poor air traffic flow	Workload
	204020000	Time pressure factors	Commitment and motivation
	204020100	Time pressure while flying	Commitment and motivation
	204030000	Mental pressure during normal operations	Commitment and motivation
	204040000	Training, examination or check situation	Competence
	204040100	Examination, check or training in progress	Competence
	204040101	Flight crew examination, check or training	Competence
	204040102	ATCO examination, check or training	Competence
	204050000	Miscellaneous operational task demands	Workload
	204050100	Task demand caused by other aircraft	Workload
	204050200	Task demand caused by passengers	Workload
	204050300	Task demand due to technical problem/failure	Workload
	204050400	Task demand caused by ground operations	Workload
300000000		Liveware (human)-Hard/software interface	
301000000		human and hardware interface	
	301010000	Workplace equipment/design	Tech-design and manufacturing
	301010100	Design or ergonomics	Tech-design and manufacturing
	301010104	Workplace instruments design unsuitable	Tech-design and manufacturing
	301010105	Workplace electronic display design	Tech-design and manufacturing
	301010109	Workplace controls and displays mislocated	Tech-design and manufacturing
	301010110	Workplace controls and displays	Tech-design and manufacturing
	301020000	Non-flight deck/cockpit aircraft equipment	Tech-design and manufacturing
	301030000	Aircraft maintenance equipment (tool design and reliability)	Aircraft maintenance tech-design and manufacturing; tech-maintenance
	301040000	ATC equipment (tool design and reliability)	ATC tech-design and manufacturing; tech-maintenance

	301050000	Suitability of design/ergonomics for training purposes	Tech-design and manufacturing
	301060000	Suitability of design for maintenance purposes	Aircraft maintenance tech-design and manufacturing
302000000		Inadequate info/data sources (factors related to lack of availability of information , inaccurate information or intermittent information)	

	302010000	Data sources	Communication;
			Tech-design and manufacturing;
			Tech-maintenance

	302010100	Radar information (not availability, inaccurate or intermittent)	Communication;
			Tech-design and manufacturing;
			Tech-maintenance

	302020000	Communication media	Tech-design and manufacturing; Tech-maintenance
303000000		human software/firmware interface	
	303010000	Human firmware interface	Tech-design and manufacturing
	303020000	Software (e.g. software which is not user friendly)	Tech-design and manufacturing
304000000		Automation systems	
	304010000	Automation design philosophies	Tech-design and manufacturing
	304010500	Reliability of automation	Tech-design and manufacturing; Tech-maintenance
	304020000	Use of automation	Competence
	304020100	Training-automation	Competence
	304020300	Use of automation (e.g. overuse of automation)	Competence
305000000		Automatic defenses/warnings (e.g. warning not working, /not available, misleading or too many false alarms)	Tech-design and manufacturing; Tech-maintenance
	305010000	Workplace warnings	Tech-design and manufacturing; Tech-maintenance
	305010300	TCAS	Tech-design and manufacturing; Tech-maintenance
	305020000	ATC alarms/alerts	Tech-design and manufacturing; Tech-maintenance
	305020100	Conflict alert	Tech-design and manufacturing; Tech-maintenance
	305030000	Other defenses/warnings	Tech-design and manufacturing; Tech-maintenance
306000000		Operational material	Procedure
	306010000	Workplace manuals, checklists and charts	Procedure
	306010100	Workplace manuals	Procedure
	306020000	Flight progress strips	Communication and coordination(info)
	306030000	Mtce engineering material (task cards and process sheets)	Procedure
	306040000	Operational documents, charts or checklists	Procedure
	306040200	Other publications	Procedure
	306040400	Written regulations	Procedure
	306040500	Other handbooks/manuals	Procedure

	306040600	Other checklists	Procedure
400000000		Human v system support	
401000000		Human interface-procedures	Procedure
	401010000	Human interface- SOP	Procedure
	401020000	Human interface- abnormal procedure	Procedure
	401030000	Human interface- ATC procedures	Procedure
	401030100	Human interface- ATC operational	Procedure
	401040000	Human interface- aerodrome procedure	Procedure
	401050000	Human interface- mtce procedures	Procedure
	401060000	Human interface- company procedure	Procedure
	401070000	Human interface- other procedures	Procedure
	401080000	Human interface- custom and practice	Procedure
402000000		Human interface- training	
	402010000	Human interface- basic/initial training	Competence
	402020000	Human interface- specific training	Competence
	402030000	Human interface- simulator training	Competence
	402040000	Human interface-on-the-job training	Competence
	402050000	Human interface- emergency training	Competence
	402060000	Human interface- crew/team resource management training	Communication and coordination
	402070000	Human interface- recurrent training	Competence
	402080000	Human interface- route training	Competence
	402090000	Human interface-miscellaneous training	Competence
	402090300	Training in the use of manuals	Procedure (part of promulgate of procedure)
	402090400	Other training	Competence
500000000		the liveware (human)-liveware (human) interface	
501000000		Human v communications	Communication and coordination
	501010000	Human v spoken communications	Communication and coordination
	501010100	Human v communications between crew	Communication and coordination
	501010300	Human v air-ground communications	Communication and coordination
	501010301	Human v ATC-pilot communications	Communication and coordination
	501010400	Human v ground-ground communications	Communication and coordination
	501010500	Human interface-language	Communication and coordination
	501010700	Human v phraseology	Communication and coordination
	501010800	Human v readback/hearback	Communication and coordination
	501010900	Human v call sign confusion	Communication and

			coordination
	501011000	Human v noise interference	Communication and coordination
	501011100	Human v interpretation	Communication and coordination
	501020000	Human v written/read communications	Communication and coordination
	501020100	Human v documentation	Communication and coordination
	501020101	Human v incomplete docs	Communication and coordination
	501020200	Human v flt progress strip	Communication and coordination
	501020700	Human v misreading	Communication and coordination
	501030000	Human v visual signals	Communication and coordination
	501030100	Human v grd-hand signals	Communication and coordination
502000000		Human v team skill/CRM	Communication and coordination (CRM training)
	502010000	Human v team skills	Communication and coordination (CRM training)
	502010100	Human v coordination	Communication and coordination (CRM training)
	502010300	Human v confidence/trust	Communication and coordination (CRM training)
	502010400	Human v cross-checking	Communication and coordination (CRM training)
	502010600	Human v peer pressure	Communication and coordination (CRM training)
	502010800	Human v briefing team	Communication and coordination (CRM training)
	502010900	Human v self feedback	Communication and coordination (CRM training)
	502011000	Human v decision process	Communication and coordination (CRM training)
	502011200	Human v team planning	Communication and coordination (CRM training)
	502011300	Human v managing workload	Communication and coordination (CRM training)
	502020000	Human v formal coordination required by procedure	Communication and coordination
	502030000	Human v team changeover	Communication and coordination
	502040000	Human v other interaction	Communication and coordination
503000000		Human v supervision	
	503010000	Human v ops supervision	Generic "monitoring" steps in the delivery system
	503020000	Human v training supervision (e.g. failing to notice or correct a mistake made by a trainee)	Generic "monitoring" steps in the competence delivery system
	503030000	Human v standards (e.g. a supervisor allowing standards to lapse)	Commitment

504000000		Human v regulatory activities	
	504010000	Regulatory procedures (e.g. requirement to report incidents where safety has jeopardized)	Monitoring, feedback and learning; regulatory
	504020000	Regulatory standards (e.g. regulatory standards are considered to be inadequate)	Procedure; regulatory
	504020100	Design standards (e.g. design standards inadequate)	Procedure; regulatory
	504020200	Certification standards (e.g. certification standards inadequate)	Procedure; regulatory
	504030000	Human interface-regulations (e.g. poor regulation)	Procedure; regulatory
	504040000	Human interface-inspections (e.g. inspections too infrequent)	Monitoring, feedback and learning; regulatory
	504050000	Human interface-monitoring organizations (monitoring the activities of organizations or individual)	Monitoring, feedback and learning; regulatory
	504060000	Human interface-surveillance associated with regulations	Monitoring, feedback and learning; regulatory
	504070000	Human interface-audits (e.g. audits fail to detect problems with an organization)	Monitoring, feedback and learning; regulatory
	504080000	Human interface-checks (e.g. checks not thorough enough)	Monitoring, feedback and learning; regulatory

Appendix F: Influence diagram calculations (Embrey, 1992)

A1	The weight of evidence for assignment of job role	Good	Poor
		0.5	0.5
A2	The weight of evidence for task complexity	High	Low
		0.6	0.4

A3 Staffing levels			
If Project management is	Then weight of evidence for staffing levels being		Weights (project management)
	Adequate	Inadequate	
Effective	0.6	0.4	0.1
Ineffective	0.2	0.8	0.9

Unconditional probability (weighted sum) that staffing levels are

Adequate	Inadequate
0.24	0.76

A4 For Time pressure					
If staffing levels are	Assignment of job is	Task complexity is	Weight of evidence for Time pressure being		Weights
			Low	High	(staffing levels×job roles×task complexity)
Adequate	Good	Low	0.95	0.05	0.072=0.24×0.5×0.6
Adequate	Good	High	0.30	0.70	0.048=0.24×0.5×0.4
Adequate	Poor	Low	0.90	0.10	0.072=0.24×0.5×0.6
Adequate	Poor	High	0.25	0.75	0.048=0.24×0.5×0.4
Inadequate	Good	Low	0.50	0.50	0.23=0.76×0.5×0.6
Inadequate	Good	High	0.20	0.80	0.15=0.76×0.5×0.4
Inadequate	Poor	Low	0.40	0.60	0.23=0. 76×0.5×0.6
Inadequate	Poor	High	0.01	0.99	0.15=0. 76×0.5×0.4

Unconditional probability (weighted sum) that time pressure is

High	Low
0.3981	0.6019

Appendix G: Management actions for "fatigue", "weather", and "workload"

Fatigue

Item	Management actions	Delivery systems
1	Set maximum hours per flight duty period and cumulative duty period	Availability
2	Set a minimum rest period after each flight and a minimum period free of all duty after a given number of consecutive days of duty	Availability
3	Set an average sleep requirement of 8 hours in a 24-hour period	Availability
4	Provide comfortable accommodation for getting good sleep at stopovers	Man/machine interface (workplace design)
5	Create a suitable crew rest environment and an appropriate place for a nap in multicrew aircraft	Man/machine interface (workplace design)
6	Provide several days off for the flight crew to adjust to a new sleep/wake schedule	Availability
7	Provide a feedback system and occurrence reporting system, whose data are used to adapt schedules	Availability
8	Require crew to attend an education and training module that helps pilots to understand the cause and effect of fatigue, and teaches pilots how to minimize fatigue and its effects (e.g. NASA nap, use of bright light exposure to minimizing circadian rhythm)	Competence
9	Check alcohol and drug consumption for a suitable period before flying	Suitability
10	Provide and use good fatigue assessment tools to objectively discover pilots with relatively high fatigue and performance decrement	Technology function& Man-machine interface
11	Provide a technical alert system that informs pilots if they are falling asleep during operations (e.g. active noise production)	Technology function& Man-machine interface
12	Provide equipment designs to improve work condition to reduce operator's on line fatigue and discomfort	Man-machine interface
13	Require good communication between flight crew members to openly discuss fatigue and their current ability to carry on work and, if necessary, to rotate flight tasks with other crew members	Communication
14	Ensure that management policy is not overridden in practice by over-scheduling tired pilots	Commitment

Weather

Item	Management actions	Delivery systems
1	Collaborate with the ATC System Command Center for constant information exchange about weather on route (pilot and ATC)	Communication
2	Provide weather information from approved sources to the dispatcher and pilot	Communication
3	Enhance communication between pilot and dispatcher about weather conditions to maintain safe operational control	Communication
4	Define minimum weather criteria to meet operational requirements and policies for preflight weather avoidance (e.g. alternate airport, choosing flight paths and landing routes)	Procedure
5	Create a daily strategic plan of operations based on known or forecasted weather two to six hours in the future	Procedure
6	Ensure flight crew, prior to each flight, complete a review of weather information (including en-route and departure, destination and alternate airports)	Procedure
7	Ensure flight crew monitor weather information en route (ATIS, ASOS/AWOS, ATC, etc.), and, where necessary, reanalyze their flight plan	Procedure
8	Equip aircraft with an airborne weather radar system capable of detecting thunderstorms and other potentially hazardous weather conditions	Technology-function
9	Ensure flight crew, before entering the proximity of adverse weather, explicitly discuss weather conditions, instructions, alternate airports, hazards and experience	Communication
10	Ensure Captain or FO monitors and, where necessary, challenges whether the other takes unnecessary risks in going through bad weather and take immediate action to correct deviations	Commitment
11	Management rewards strict adherence to weather-related procedures and takes disciplinary action against violations	Commitment
12	Management is committed to continuous improvement in instrumentation, information provision and (joint) training to develop collaborative solutions to weather constraint issues	Commitment
13	Train flight crew members to enhance their decision making in adverse weather and environmental conditions	Competence

Workload

Item	Management actions	Delivery systems
1	Malfunction due to crew action or inaction	Tech-function
2	Malfunction due to poor, incomplete or missed maintenance or errors in maintenance	Tech-function
3	Malfunction due to crew action or inaction	Competence/ Commitment
4	Malfunction due to external factors	None

Appendix H: Distribution of rainfall rate and Distribution of number of time the crew members have to refer to the A/E procedure (Ale, 2009)

Distribution of rainfall rate

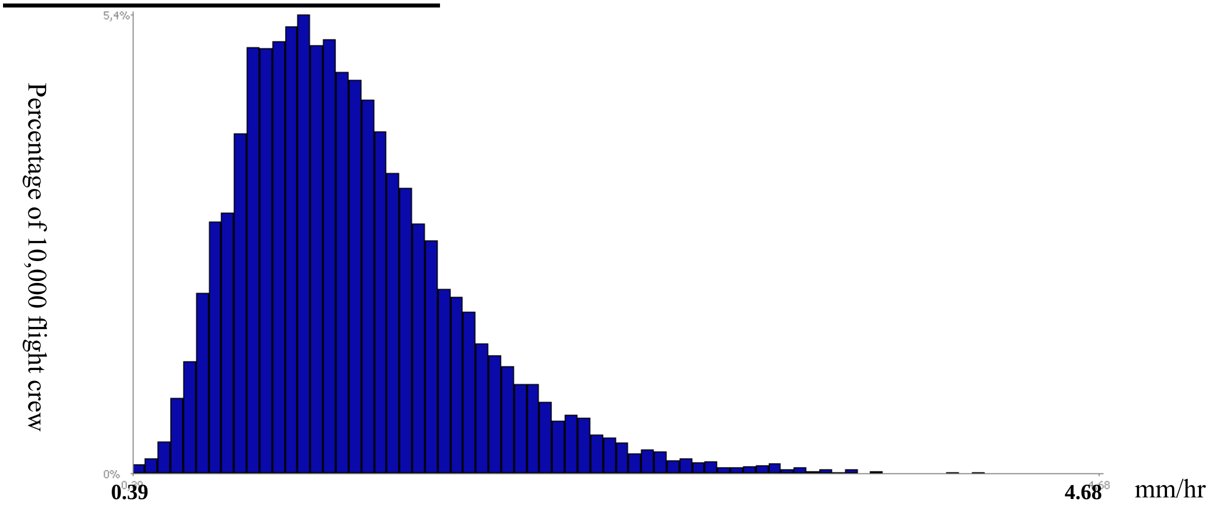


Figure H-1 Rainfall rate (mm/hr) translated into airborne weather radar in the cockpit

Distribution of number of time the crew members have to refer to the A/E procedure section of the aircraft operation manual during flight

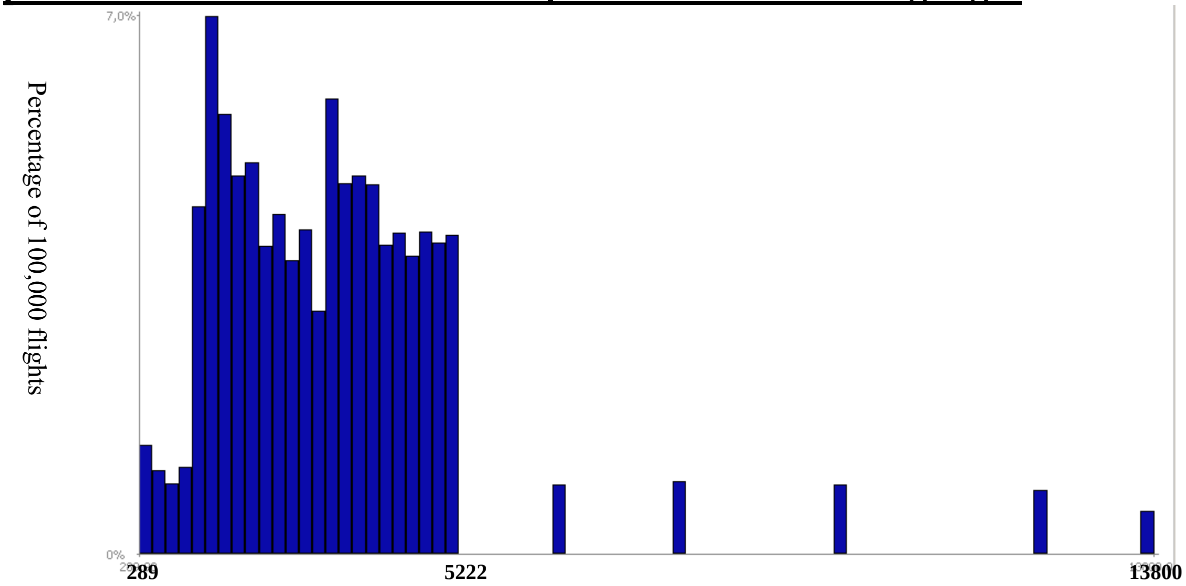


Figure H-2 Times the crew members have to refer to the abnormal/emergency procedures section per 100,000 flight

Summary

Aviation safety is so well developed that the current rate of accidents for scheduled commercial aircraft involving passenger fatalities in the region of the 27 EU Member States plus Iceland, Liechtenstein, Norway and Switzerland is approximately 3.6 accidents in 10 million flights. Because of the low accident rate, individual organizations such as airlines or airports cannot rely on information from accidents to drive their safety improvements. They need a causal model to identify latent failures further back in time and up into the higher management levels. CATS (Causal Model for Air Transport Safety) was set up in 2005 by the Dutch Ministry of Transport, Public Works and Water Management as a project to develop an integrated risk model of air transport for the whole flight cycle from (departure) gate to (arrival) gate. The experience gained in CATS is used to illustrate the development of a safety management model in probabilistic risk assessment, and to indicate what should be improved in the longer term.

The general structure of the CATS model and its management part had been decided before I was taken on to work on this thesis. Modification to it within the time scale of the project could only be partial. Moreover, the strong emphasis on quantification of CATS, as is the case with other risk models, led to a number of restrictions in what was eventually modelled in the management section of that model. Therefore, this PhD took a step back and examined the assumptions made in arriving at the original intentions for management modelling in CATS and the decisions made in putting them into practice. This involved:

1. Revisiting the development of the Dutch management model used in CATS and critically examining its structure, the assumptions built into that structure, and its completeness and appropriateness to be applied to the aviation field.
2. Critically analysing the issue of the human factors and technology failures at the lower system level, and how these might be connected with a management model.
3. Exploring the availability of data within the aviation industry about management failures in order to quantify the risk implications of different management changes to prevent accidents. The findings show that only a limited amount of management information is available up to now.
4. Critically assessing quantification methods linked to the BBNs and system dynamics and proposing a simpler form of elicitation method to get round the complexity of the expert elicitation that the BBN method involves.
5. Putting together the areas of improvement on management modelling and quantification (Chapter 7) to form a new integrated model that is proposed for further testing and eventual use in an extension of CATS.

The main conclusions and implications of the thesis are presented as follows:

1) ***A strength of the Dutch management system is its formulation of safety management influences as a set of actions (a management process) "delivering" the systematic resources and controls to the barriers of hardware and/or human behavior, which can be taken by managers to influence risk. The concurrent validity of the Dutch safety management model with respect to other existing frameworks used in aviation was***

examined and shown to be reasonably well supported. However, it needs some simplification and more focus on how it links with human and technical errors (ch2).

CATS built on earlier projects (I-Risk, ARAMIS, WORM), to model 2 delivery systems related to hardware (technology interface, technology function) and 5 delivery systems related to behaviour (procedures, availability, competence, commitment, communication). Apart from these delivery systems, the safety management system (SMS) also has to manage, at a higher system level, the processes of "risk identification, barrier selection and specification", and the process of "monitoring, feedback, learning and change management".

However, behind this relative consistency in the formulation of the model, the review of the development of the Dutch model showed two critical problems that needed to be resolved: 1) none of the previous projects had been completed the original objectives of management modelling, partly due to the limited time scales available for the projects, but also because the concept model needed some simplification; 2) the current Dutch model did not give modellers sufficient clarity about how the management controls should link to the individual factors. The way in which the SMS needs to be linked with the technical and human factors is dealt with in ch7.1, which also clarifies the steps in the delivery systems into a generic structure which can be worked out much more easily for aviation safety (Section 7.2.1).

Different models developed for or applied to aviation were mapped onto the Dutch management model to provide a concurrent and face validation to the Dutch model used in CATS. The comparisons with HFACS and SoTeRiA showed that the majority of the well-defined elements forming part of those models can be satisfactorily accommodated within our delivery systems. There are some ill-defined elements of those models whose definitions are still so vague that they cannot be easily compared. These are particularly safety culture and safety climate, but also aspects of management and organisation at a higher system level than the Dutch model currently attempts to deal with. We argued that safety cultural aspects can be reflected in the strength of the SMS through our delivery systems (Section 7.2.2.8). Therefore, safety culture (in the current studies) can be seen as a measure of how seriously company takes its own management process. Dealing with any other more generic and higher level concepts is left to future work beyond this thesis.

Studies in two airlines were conducted to find out if the Dutch model also covers all the things that SMSs are doing in the airlines (Section 2.3.2). These showed that line operations are the ones that actually carry out all of the functions in the delivery systems. Such safety business processes are so integrated in what people in the line are doing that airlines seem not to define these activities explicitly as part of their SMSs (where safety is influenced proactively). The only aspects clearly defined as being part of their SMSs are the "feedback and learning" and "risk assessment" elements defined in our delivery systems.

2) ***An overview of the human factors from different accident and incident investigation schemes was constructed in order to see if the Dutch model can support the control functions to all of these human factors. The human factors identified were mapped to concepts at the same system level as the delivery systems. This showed that two management functions were missing and needed to be added to the new version of the Dutch model. We also found that the human factors formulated in the current probabilistic quantification models seem in general quite limited, due to the strong focus on quantifiability. Therefore, the factors considered in risk models such as CATS are not a comprehensive overview of factors that have the potential to influence flight crew performance (ch3).***

   No really suitable human factor classification scheme was found which is sufficiently comprehensive within a hierarchical classification to be able to map to a deeper set of organizational causal factors. Data collection tools concerning human factors are relatively unsystematic in respect of our criteria. Hence, a new human factor taxonomy (Table 3.4) was built for this research. Mapping the human factors onto the Dutch model shows that there is a good match but additional functions for “workload” needed to be added to the Dutch model and “competence and suitability” needed to be split. Also, the mapping shows that the current Dutch model is too conceptual and generic in respect to resolving (preventing and coping with) human errors. The whole analysis in Section 7.2.2 tailors the delivery systems for each of the new/modified categories of the factors found in the accident and incident analysis.

   The human factors formulated in the current probabilistic quantification models were also reviewed. The current modelling of the human factors of the flight crew seems in general quite limited. Therefore, the most important improvement we can propose for human performance modelling in HRA is to get a better understanding of the relationship between the qualitatively generally well understood notions and their translation into real, observable and thus quantifiable influences on risk and risk reduction.

3) ***As the management aspects of aircraft deficiencies have not been worked out to an operational level in the CATS project, the processes of “design and manufacturing”, “safe operation by flight crew”, and “maintenance” were modelled, but only conceptually, to ensure satisfactory performance of an aircraft system over its entire design life in relation to both the technical functioning and the man-machine interface (ch4). This work could be used for further development of the modelling in CATS.***

4) ***Four types of hard data on aviation performance (ADREP, LOSA, EU-OPS, IOSA) were critically analysed to show that only a limited amount of management information has been available up to now (Section 5.2).*** This is largely due to confidentiality problems, missing data, and the lack of clear, consistent and recognisable causal frameworks underlying the data collection models. Therefore, it is not possible to use any of these data sources to quantify the relationship between the SMS and the human errors. It would be of great help to get a better estimate of the probabilities of management events early in the causal chain if company audit data could be released for

scientific research. There is an urgent need to develop industry-wide data collection and analysis schemes that are comprehensive and compatible with each other. This would allow data held in separate databases to be systematically integrated into the model. In such a way, it would become possible to enhance overall data usage and help identify weaknesses in the aviation system.

5) ***Currently Bayesian Belief Networks (BBNs) and System Dynamics are two major quantification methods to incorporate management factors into risk models. These methods have different strengths and weakness in quantification (Section 5.3 & Section 5.4). A supplementary method (combining paired comparison with distribution free continuous BBNs) is proposed in this research to help quantify the Dutch management model in CATS (Section 5.5).*** The experiments show that in general paired comparisons are relatively an easier and more intuitive elicitation method than the complex BBN questions. The method designed in this research can be applied both for quantitative variables and qualitative cases. It is particularly useful that the "soft" variables could be modelled more closely to the reality of what can be influenced by management in clearly demonstrable ways. However, the paired comparison method does not take account of dependencies between the management influences. It assumes them to be independent. It is therefore of most use in screening and prioritising management influences.

6) ***As demonstrated in the preceding sections the Dutch model needs three major changes to provide a comprehensive model of all of the relevant levels in causal chain: a) clarification of the hierarchical relations between the SMS and operations; b)improvement in the detailed modelling of each system level; and c) clarification of the generic structure of the delivery systems, which is much simpler and easier to apply.***

a) A general structured model (Figure 7.2 in Section 7.1.1) is introduced in this thesis to clarify the relationship between the SMS and operations (human factors and technical failures) in the accident analysis. The hierarchical relations between them are treated as a control process. In this version of the Dutch model, we treat safety management as ensuring that the internal processes at the operational level (human and technology) are working properly and individual factors that interfere with them are managed to an acceptable level.

b) The theories and findings of the SMS (ch2), human (ch3), and technical factors (ch4) were put into an integrated and articulated model (Figure 7.3 in Section 7.1.2). Level 3 identifies the management model seen as providing the essential resources and controls to level 2. This management model adopts the concept of delivery systems and tasks within each delivery system. Level 2 is the (hidden) internal cognitive mechanisms of the human and the equivalent internal functioning of the hardware, which leads to actions and interactions at level 1. To make the SMS more specific in its task of managing issues related to underlying causes in level 2, an extensive list of the human factors at that level is specified (Table 7.1 in Section 7.1.2) and the control functions that need to be linked to them by the delivery systems is extensively discussed (Section 7.2). The behaviour of the aircraft (level 1) is influenced by the instrumentation design of the technical function and the man-machine interface (MMI) in level 2.

c) The delivery systems are simplified into a generic structure (Figure 7.4 in Section 7.2.1), worked out specifically per delivery system to provide resources and controls to the human factors identified in the previous point.

To conclude, this thesis re-examines the place and role of the human and management models and their quantification in a more fundamental way. Based on the experience of CATS, this thesis shows the challenge of quantifying management influences in risk modelling in aviation, but also makes proposals for improvement: a generic hierarchical control model for aviation safety, a list of human and technical factors to be treated in risk modelling in aviation, an additional way of quantifying safety management in risk model, and recommendations to improve the availability of the data in aviation to be able to quantify the relationship between the SMS and the human factors. These recommendations could eventually be used in an extension of CATS, or in the other research with similar objectives.

Samenvatting

Luchtvaartveiligheid is zo goed ontwikkeld dat de huidige ongevalsratio voor geplande commerciële vluchten met dodelijke slachtoffers onder passagiers in het domein van de 27 EU-lidstaten plus IJsland, Liechtenstein, Noorwegen en Zwitserland ongeveer 3,6 ongevallen per 10 miljoen vluchten is. Vanwege dit lage aantal ongevallen is het lastig voor individuele organisaties zoals luchtvaartmaatschappijen of luchthavens om gebruik maken van informatie afkomstig van ongevallen om hun veiligheidsverbeteringen te sturen. Zij hebben behoefte aan een causaal model om latente storingen verder terug in de tijd en tot op hogere management niveaus te identificeren. CATS (Causal Model for Air Transport Safety ofwel Causaal Model voor Luchttransportveiligheid) is in 2005 opgezet door het Nederlandse Ministerie van Verkeer en Waterstaat als een project om een geïntegreerd risicomodel voor luchttransport voor de gehele vluchtcyclus te ontwikkelen, van (vertrek)poort tot (aankomst)poort. De ervaring opgedaan in CATS wordt gebruikt om de ontwikkeling van een veiligheidsmanagementmodel voor probabilistische risicobepaling te illustreren en om aan te geven wat op de langere termijn verbeterd zou moeten worden aan de huidige gang van zaken.

De algemene structuur van het CATS-model en het managementgedeelte was reeds vastgelegd voordat ik aangenomen werd om aan dit proefschrift te werken. Het was slechts mogelijk om, binnen de tijdsspanne van het project, delen ervan te wijzigen. Bovendien leidde de sterke nadruk op kwantificering binnen CATS, zoals ook het geval is met andere risicomodellen, tot een aantal beperkingen in het managementgedeelte van het model. In dit proefschrift is daarom een stap terug en zijn de aannamen onderzocht die zijn gemaakt bij de oorspronkelijke bedoelingen voor het modelleren van het management in CATS en de beslissingen die zijn gemaakt om deze in praktijk te brengen. Deze betroffen:

1. Heroverwegen van de ontwikkeling van het Nederlandse managementmodel dat wordt toegepast in CATS en het kritisch beschouwen van haar structuur, de aannamen die geleid hebben tot deze structuur, zowel de volledigheid als de juistheid ervan voor toepassing in de luchtvaartsector.
2. Kritische analyse van storingen veroorzaakt door menselijke factoren en technologie op het lagere systeemniveau en hoe deze kunnen worden verbonden met een managementmodel.
3. Verkennen van de beschikbaarheid van data binnen de luchtvaartindustrie over het falen van management, om de risico implicaties te kwantificeren van verschillende veranderingen in management, om ongevallen te voorkomen. Bevindingen tonen aan dat tot nu toe slechts een beperkte hoeveelheid management informatie beschikbaar is.
4. Kritisch evalueren van kwantificeringsmethoden gekoppeld aan de BBNs en systeemdynamiek. Een eenvoudigere vorm van de elicitatiemethode wordt voorgesteld om de complexiteit van expertbevraging te omzeilen, die de BBN-methode momenteel nodig heeft.
5. De gebieden voor verbetering van managementmodellering en kwantificering (Hoofdstuk 7) bij elkaar brengen om een nieuw geïntegreerd model te bouwen dat voor verder onderzoek wordt voorgesteld en uiteindelijk kan worden gebruikt in een uitbreiding van CATS.

De belangrijkste conclusies en implicaties van het proefschrift kunnen als volgt worden gepresenteerd:

1) **Een sterkte van het Nederlandse managementsysteem is de formulering van veiligheidsmanagementinvloeden als een set van maatregelen (een management proces) dat systematisch middelen en controles "levert" aan de hardware barrières en/of menselijk gedrag, die door managers kunnen worden gebruikt om risico te beïnvloeden. De concurrente validiteit van het Nederlandse veiligheidsmanagementmodel met betrekking tot andere, bestaande raamwerken die momenteel in de luchtvaart worden gebruikt, is onderzocht. Het is aangetoond dat het model redelijk goed wordt ondersteund. Het Nederlandse managementsysteem zal echter meer vereenvoudigd moeten worden met een grotere aandacht voor de relatie tussen menselijke en technische fouten (Hoofdstuk 2).**

CATS bouwde voort op eerdere projecten (I-Risk, Aramis, WORM) om leveringssystemen (*delivery systems*) te modelleren. Twee systemen met betrekking tot hardware (technologie-interface, technologie-functie) en vijf gerelateerd aan gedrag (procedures, beschikbaarheid, competentie, inzet, communicatie). Afgezien van deze leveringssystemen dient het veiligheidsmanagementsysteem (VMS) op een hoger systeemniveau ook processen te beheren van de "risico (scenario) identificatie, barrière selectie en specificatie", en processen als "monitoren, terugkoppeling, leren en verandermanagement".

Echter bleken achter deze betrekkelijke consistentie in de formulering van het model, twee kritieke problemen bij de herziening van de ontwikkeling van het Nederlandse model naar voren te komen, die opgelost moesten worden: 1) geen van de vorige projecten heeft de oorspronkelijke doelstellingen van het modelleren van het management afgemaakt, deels vanwege de beperkte tijdspanne die beschikbaar is voor dergelijke projecten, maar ook omdat het conceptmodel meer vereenvoudiging nodig had; 2) het huidige Nederlandse model gaf de modelbouwers niet voldoende duidelijkheid over hoe de managementcontroles met de individuele factoren verbonden dienen te worden. De wijze waarop het VMS moet worden verbonden met de technische en menselijke factoren wordt behandeld in Hoofdstuk 7.1, waarin ook de stappen in de leveringssystemen worden duidelijk gemaakt zodat deze een stuk eenvoudiger voor luchtvaartveiligheid uitgewerkt kunnen worden (Hoofdstuk 7.2.1).

Verschillende modellen die ontwikkeld zijn of toegepast in de luchtvaart, zijn vergeleken met het Nederlandse managementmodel, gebruikt bij CATS om een concurrente en indruks validiteit te bepalen. De vergelijkingen met HFACS en SoTerRia toonden aan dat de meerderheid van de goed gedefinieerde elementen die deel uitmaken van deze modellen in voldoende mate kunnen worden ondergebracht binnen onze leveringssystemen. Er zijn een aantal slecht gedefinieerde elementen binnen deze modellen waarvan de definities nog steeds zo vaag zijn dat ze niet gemakkelijk kunnen worden vergeleken. Dit geldt met name voor veiligheidscultuur en veiligheidsklimaat, maar ook voor aspecten van management en organisatie op een hoger niveau dan het Nederlandse model systeem momenteel probeert te hanteren. We voerden aan dat de veiligheidsculturele aspecten kunnen worden gereflecteerd in de kracht van het VMS via onze leveringssystemen (Hoofdstuk 7.2.2.8). Daarom kan veiligheidscultuur (in de huidige studies) worden gezien als een maat hoe serieus een bedrijf haar eigen managementproces neemt. Omgaan met andere, meer generieke concepten op een hoger niveau is werk voor de periode na dit proefschrift.

Er zijn studies bij twee luchtvaartmaatschappijen uitgevoerd om te achterhalen of het Nederlandse model alle zaken omvat die VMS'en uitvoeren in luchtvaartmaatschappijen (Hoofdstuk 2.3). Deze toonden aan dat zij die werkzaam zijn in de operatie ook degenen zijn die daadwerkelijk alle functies in de leveringssystemen uitvoeren. Dergelijke veiligheidswerkprocessen zijn zo geïntegreerd in de werkzaamheden van mensen in de operatie, zodat luchtvaartmaatschappijen deze activiteiten niet expliciet lijken te definiëren als onderdeel van hun VMS'en (waar veiligheid proactief beïnvloed wordt). De enige aspecten die duidelijk omschreven worden als een onderdeel van hun VMS'en zijn de "terugkoppeling en leren" en "risicobeoordeling" elementen gedefinieerd in onze leveringssystemen.

2) **Een overzicht van de menselijke factoren (*human factors*) van verschillende ongevallen- en incidentenschema's werd samengesteld om te bezien of het Nederlandse model de controlefuncties kan ondersteunen van al deze menselijke factoren. De geïdentificeerde menselijke factoren werden toegewezen aan concepten op hetzelfde systeemniveau als de leveringssystemen. Dit toonde aan dat twee managementfuncties ontbraken en toegevoegd moesten worden aan de nieuwe versie van het Nederlandse model. We vonden ook dat de menselijke factoren zoals geformuleerd in de huidige probabilistische kwantificeringsmodellen in het algemeen vrij beperkt lijken, als gevolg van de sterke aandacht voor kwantificeerbaarheid. Daarom geven de factoren die in risicomodellen zoals CATS worden beschouwd geen uitputtend overzicht van factoren die de prestaties van cockpitpersoneel zou kunnen beïnvloeden (Hoofdstuk 3).**

Er is geen goed geschikt classificatiesysteem van menselijke factoren gevonden dat voldoende dekkend is binnen een hiërarchische indeling om deze te kunnen toewijzen aan een diepere reeks van organisatorische oorzakelijke factoren. Instrumenten voor gegevensverzameling met betrekking tot menselijke factoren zijn relatief onsystematisch ten aanzien van onze criteria. Vandaar dat een nieuwe taxonomie voor menselijke factoren (Tabel 3.4) voor dit onderzoek is ontwikkeld. Een vergelijking van de menselijke factoren met het Nederlandse model toont een goede overeenkomst aan, maar er moesten extra functies voor 'werkdruk' aan het Nederlandse model toegevoegd worden en 'bekwaamheid en geschiktheid' moest worden opgesplitst. Uit de vergelijking blijkt eveneens dat het huidige Nederlandse model te conceptueel en te generiek is met betrekking tot het oplossen van (het voorkomen van en het omgaan met) menselijke fouten. De gehele analyse in Hoofdstuk 7.2.2 maakt de leveringssystemen op maat voor elk van de nieuwe/gewijzigde categorieën van de factoren in de ongevals- en incidentanalyse.

De menselijke factoren opgenomen in de huidige probabilistische kwantificeringsmodellen zijn ook beoordeeld. De huidige modellering van de menselijke factoren van de bemanning lijkt in het algemeen vrij beperkt. Daarom is de meest belangrijke verbetering die we kunnen voorstellen de modellering van menselijke prestaties in HRA om een beter begrip te krijgen van de relatie tussen de kwalitatief, en over het algemeen goed begrepen, noties en de vertaling ervan in echte, waarneembare en dus meetbare invloeden op risico's en risicobeperking.

3) **Daar de managementaspecten van gebreken van vliegtuigen niet op een operationeel niveau zijn uitgewerkt in het CATS-project, zijn de processen "ontwerp en fabricage", "veilige bediening door het cockpitpersoneel" en "onderhoud" slechts conceptueel gemodelleerd, om over de gehele ontwerpcyclus een bevredigende uitvoering van een vliegtuigsysteem te verzekeren in relatie tot zowel het technisch**

functioneren als de mens-machine-interface (Hoofdstuk 4). Dit werk zou kunnen worden gebruikt voor de verdere ontwikkeling van de modellering in CATS.

4) **Vier typen van harde gegevens over luchtvaartprestaties (ADREP, LOSA, EU-OPS, IOSA) werden kritisch geanalyseerd om te laten zien dat slechts een beperkte hoeveelheid managementinformatie tot nu beschikbaar is (Hoofdstuk 5.2).** Dit is grotendeels te wijten aan problemen m.b.t. vertrouwelijkheid, ontbrekende gegevens, en het ontbreken van duidelijke, consistente en herkenbare causale raamwerken die ten grondslag liggen aan de dataverzamelingsmodellen. Het is daarom niet mogelijk om enige van deze bronnen te gebruiken om de relatie te kwantificeren tussen het VMS en menselijke fouten. Om tot een betere inschatting te komen van de waarschijnlijkheid van management gebeurtenissen in het begin van de causale keten zou het zeer behulpzaam zijn als auditgegevens van bedrijven kunnen worden vrijgegeven voor wetenschappelijk onderzoek. Er is een dringende behoefte aan het ontwikkelen van industrie-brede dataverzamelings- en analyseschema's die uitvoerig en onderling uitwisselbaar zijn. Hierdoor zouden data die momenteel zijn opgeslagen in aparte databases systematisch kunnen worden geïntegreerd in het model. Op deze manier zou het mogelijk worden om het algehele gebruik van data te bevorderen en te helpen bij het identificeren van zwakke plekken in het luchtvaartsysteem.

5) **Momenteel zijn Bayesiaanse Belief Networks (BBNs) en Systeem Dynamiek twee belangrijke kwantificeringsmethoden om managementfactoren op te nemen in risicomodellen. Deze methoden hebben verschillende sterke en zwakke punten in de kwantificering (Hoofdstuk 5.3 & 5.4). Een aanvullende methode (het combineren van gepaarde vergelijking met de distributievrije, continue BBNs) is in dit onderzoek voorgesteld om het kwantificeren van het Nederlandse management model in CATS te vergemakkelijken (Hoofdstuk 5.5)**. De experimenten tonen aan dat in het algemeen gepaarde vergelijkingen relatief makkelijker zijn en een meer intuïtievere uitlokkingsmethode is dan de ingewikkelde BBN-vragen. De methode die in dit onderzoek is ontwikkeld kan worden toegepast voor zowel de kwantitatieve variabelen als de kwalitatieve gevalsbeschrijvingen. Het is bijzonder nuttig dat de 'zachte' variabelen in overeenkomst met de realiteit konden worden gemodelleerd van hetgeen door het management op duidelijk aantoonbare manieren kan worden beïnvloed. Echter, de gepaarde vergelijkingsmethode houdt geen rekening met de afhankelijkheden tussen managementinvloeden. Het veronderstelt dat deze onafhankelijk zijn (Hoofdstuk 6). Het is daarom van het meeste nut bij het screenen en prioriteren van managementinvloeden.

6) Zoals in de voorgaande paragrafen is aangetoond heeft het Nederlandse model drie belangrijke wijzigingen nodig om te voorzien in een alomvattend model van alle relevante niveaus in de causale keten: a) verduidelijking van de hiërarchische relaties tussen het VMS en bedrijfsactiviteiten; b) verbetering in de gedetailleerde modellering van ieder systeemniveau; en c) verduidelijking van de generieke structuur van de leveringssystemen, die veel eenvoudiger is en gemakkelijker is toe te passen.

a) Een algemeen gestructureerd model (Figuur 7.2) is in dit proefschrift geïntroduceerd om de relatie te verduidelijken tussen het VMS en (menselijke factoren en technische storingen) bij de ongevalsanalyse. De hiërarchische verhoudingen tussen deze worden behandeld als een controleproces. In deze versie van het Nederlandse model, behandelen we veiligheidsmanagement als het toezien dat de interne processen op het operationele niveau (de mens en technologie) correct werken en de individuele factoren die hiermee interfereren tot een aanvaardbaar niveau worden beheerst.

b) De theorieën en bevindingen van het VMS (Hoofdstuk 2), menselijke (Hoofdstuk 3), en technische factoren (Hoofdstuk 4) zijn in een geïntegreerd en geleed model verwerkt (Figuur 7.3). Niveau 3 beschouwt het managementmodel als het verstrekken van de essentiële middelen en controles naar niveau 2. Dit managementmodel neemt het concept van leveringssystemen aan en de taken binnen elk leveringssysteem. Niveau 2 betreft de (verborgen) interne cognitieve mechanismen van de mens en de vergelijkbare interne werking van de hardware, die tot acties en interacties leiden op niveau 1. Om het VMS meer specifiek in zijn taak te maken van het beheren van kwesties in verband met de onderliggende oorzaken op niveau 2, is een uitgebreide lijst van de menselijke factoren op dat niveau opgesteld (Tabel 8.1) en de functies die daaraan moeten worden gekoppeld door de leveringssystemen, wordt uitvoerig besproken (Hoofdstuk 7.2). Het gedrag van het vliegtuig (niveau 1) wordt beïnvloed door het ontwerp van de instrumentatie van de technische functie en de mens-machine interface (MMI) op niveau 2.

c) De leveringssystemen zijn vereenvoudigd tot een generieke structuur (Figuur 7.4), specifiek uitgewerkt per leveringssysteem om controles en middelen te verstrekken aan de menselijke factoren uit het vorige punt.

Tot slot, dit proefschrift onderzoekt opnieuw op een meer fundamentele manier de plaats en rol van de mens en managementmodellen en hun kwantificering. Gebaseerd op de ervaring van CATS, toont dit proefschrift de uitdaging van het kwantificeren van managementinvloeden aan in risicomodellen in de luchtvaart, maar doet ook voorstellen voor verbetering: een generiek hiërarchisch controlemodel voor luchtvaartveiligheid, een lijst van menselijke en technische factoren die behandeld dienen te worden in risicomodellering in de luchtvaart, een extra manier voor het kwantificeren van veiligheidsmanagement in het risicomodel, en aanbevelingen ter verbetering van de beschikbaarheid van data in de luchtvaart om de relatie te kunnen kwantificeren tussen het VMS en menselijke factoren. Deze aanbevelingen zouden uiteindelijk kunnen worden gebruikt in een uitbreiding van CATS, of in ander onderzoek met vergelijkbare doelstellingen.

Acknowledgements

It has been my dream to combine my profound interest in academic matters with an exploration of the world. I did not know where I would end up and so I waited to see where my dreams would take me. Four years after my post-graduate studies, I found myself in a beautiful flat low-lying country, ready to start on my new academic journey and experience life in a different country.

The journey so far has been full of joy. In that connection there are many people who I would like to thank and acknowledge.

First, I am very grateful to my promoter, Andrew Hale, who made the beginning of this journey possible. Your support and encouragement was essential to the completion of this dissertation. I am still amazed about the way you commit yourself to supervising your PhD candidates, how you are always available to answer questions and how, in my case, you always found time to provide feedback, no matter where you were or how busy you were. I have to say that I have been very fortunate to have you as my promoter.

I would also like to thank all my colleagues from the Safety Science Group who have helped to make my work such a pleasant endeavour; in particular Marieke Kluin deserves special recognition for being such a cheerful roommate! Thanks also go to Hinke Andriessen, Ellen Jagtman, Frank Guldenmund, Carla van Dongen, and Erika van Verseveld, who have helped me on many occasions, especially in the last few months leading up to my defence.

My heartfelt thanks also go to the Dutch Ministry of Transport, Public Works and Water Management, which not only sponsored the CATS (Causal Model for Air Transport Safety) project but also provided me with a rich environment in which to conduct the research presented in this thesis. The CATS people that I particularly want to thank are Coen van Gulijk, Roger Cooke, Dorota Kurowicka, Oswaldo Morales-Napoles, Linda Bellamy, Dan Ababei, John Spouge, John Cooper and Rob van der Boom. Special thanks go to Ben Ale, the project leader who gave me the freedom and the confidence to pursue my own ideas and the necessary guidance for those ideas to reach fruition; Alfred Roelen, you also shared with me your knowledge of aviation and conducted several informal discussions during which you answered my various queries. Seven internal and external CATS experts, Thomas Bos, Arthur Dijkstra, Arun Karwal, Wim Huson, Udo Dees, Hessel Benedictus and Robert Tump, were kind enough to devote their time to going through all the paired comparison questions with me. Even though sometimes the paired comparison questions were very long and some of the pairs looked quite similar, you remained patient and did a wonderful job. I want to thank all of you experts for your time and patience. Without your information this thesis would not have been possible at all. One of you particularly needs an extra mention because you were kind enough and enthusiastic enough to provide me with valuable information during several long interviews in a small café at Schiphol Airport. However, since I promised to keep this confidential, I will not mention your name. But many others listed here know who you are and I want to thank you again for your willingness to share with me your airline experience.

I thank my paranymphs, Hinke Andriessen and Marieke Kluin, who support me and have agreed to be the “best women” at my defence (apart from me). I really appreciate everything you have done for me. I hope you don't need to act as a physical shield if the debate becomes too heated!!

Also in my private life, I want to thank the people who provided all the necessary relaxation outside the university. My friends, Haibo, Chintan, Jason, Nana, Chien-Ching, Kai-Fan, Yi-Chen, Min Zhang, Huai-Wen, Yuan-Tse, and Chung-Kai. My life in the Netherlands would not have been so easy without you. I would also like to thank father Lai's family. Your company made me feel very at home.

Furthermore, I would also like to thank Professor Chaug-Ing Hsu and Professor Kai Kao who encouraged me to go abroad and who believed in me. You were not only professors at my university in Taiwan, but also important mentors along the way.

My deepest thanks go to my family. My father who set the bar high for my academic career and dedication to science; and my dear mother, thank you for your support and advice, not only concerning my thesis but also all the big and small things in life. Tingchieh, thank you for all the long-distance calls, photos, and videos designed to help reduce my homesickness. I treasure lovely memories of when you visited me in the Netherlands in 2009. Mingwei, thank you for your permanent support and for being my private doctor without charging me a penny. I would also like to thank my parents-in-law for their kind support, especially during my pregnancy in the last year of my PhD research.

Moreover, I also want to thank my little daughter, Rozen. You have been a wonderful baby and a good time manager. You were patient enough to wait until the last moment (literally!), until your mother had submitted her final thesis and had obtained a signature from her promoter before coming into this world. Without your patience, this thesis would not have gone so smoothly. Finally, the last sentences are reserved for Bill. Your love and support (especially technical ones) have helped me to achieve this milestone. I could never finish this thesis without you. (Although you did try to slow me down through the pregnancy, I succeeded ☺). I am very grateful that you have joined me on my journey and that you are there to enrich my life!

Pei-Hui Lin
Delft, June 2011

About the author

Pei-Hui Lin was born in Tainan, Taiwan on May 30th, 1976. After graduating from Chia-Chi Girls' High School in 1995, she went on to study Transportation Engineering and Management at the National Chiao Tung University in Hsinchu, Taiwan. She graduated in 2001 and her M.Sc. thesis was on airline network modelling in conjunction with airport noise charges. A paper based on this work subsequently won the best paper award in the 5th Network Conference in Taiwan. Later it was turned into a journal paper and published in the international journal of Transportation Research-part D.

After graduating from university, she first worked as a consultant in the transportation industry in Taiwan. Soon after that she worked for the Department of Transportation in Taipei, where she was in charge of the city's traffic organization, planning and improvement. Later, she worked as a researcher at the national Industrial Technology Research Institution where she was responsible for providing pragmatic and creative solutions for the transportation industry in Taiwan.

In December 2005, she started on her PhD research at Delft University of Technology where she became engaged in the CATS (Causal Model for Air Transport Safety) project. At the end of 2010, her first child, Rozen, was born. Since April 2011, Pei-Hui has been a postdoc within the Safety Science Group at Delft University of Technology.

NGInfra PhD thesis series on infrastructures

1. Strategic behavior and regulatory styles in the Netherlands energy industry
   Martijn Kuit, 2002, Delft University of Technology, the Netherlands.
2. Securing the public interest in electricity generation markets, The myths of the invisible hand and the copper plate
   Laurens de Vries, 2004, Delft University of Technology, the Netherlands.
3. Quality of service routing in the internet: theory, complexity and algorithms
   Fernando Kuipers, 2004, Delft University of Technology, the Netherlands.
4. The role of power exchanges for the creation of a single European electricity market: market design and market regulation
   François Boisseleau, 2004, Delft University of Technology, the Netherlands, and University of Paris IX Dauphine, France.
5. The ecology of metals
   Ewoud Verhoef, 2004, Delft University of Technology, the Netherlands.
6. MEDUSA, Survivable information security in critical infrastructures
   Semir Daskapan, 2005,Delft University of Technology, the Netherlands.
7. Transport infrastructure slot allocation
   Kaspar Koolstra, 2005, Delft University of Technology, the Netherlands.
8. Understanding open source communities: an organizational perspective
   Ruben van Wendel de Joode, 2005, Delft University of Technology, the Netherlands.
9. Regulating beyond price, integrated price-quality regulation for electricity distribution networks
   Viren Ajodhia, 2006, Delft University of Technology, the Netherlands.
10. Networked Reliability, Institutional fragmentation and the reliability of service provision in critical infrastructures
    Mark de Bruijne, 2006, Delft University of Technology, the Netherlands.
11. Regional regulation as a new form of telecom sector governance: the interactions with technological socio-economic systems and market performance
    Andrew Barendse, 2006, Delft University of Technology, the Netherlands.
12. The Internet bubble - the impact on the development path of the telecommunications sector
    Wolter Lemstra, 2006, Delft University of Technology, the Netherlands.
13. Multi-agent model predictive control with applications to power networks
    Rudy Negenborn, 2007, Delft University of Technology, the Netherlands.
14. Dynamic bi-level optimal toll design approach for dynamic traffic networks
    Dusica Joksimovic, 2007, Delft University of Technology, the Netherlands.
15. Intertwining uncertainty analysis and decision-making about drinking water infrastructure
    Machtelt Meijer, 2007, Delft University of Technology, the Netherlands.
16. The new EU approach to sector regulation in the network infrastructure industries
    Richard Cawley, 2007, Delft University of Technology, the Netherlands.
17. A functional legal design for reliable electricity supply, How technology affects law
    Hamilcar Knops, 2008, Delft University of Technology, the Netherlands and Leiden University, the Netherlands.
18. Improving real-rime train dispatching: models, algorithms and applications
    Andrea D'Ariano, 2008, Delft University of Technology, the Netherlands.
19. Exploratory modeling and analysis: A promising method to deal with deep uncertainty
    Datu Buyung Agusdinata, 2008, Delft University of Technology, the Netherlands.
20. Characterization of complex networks: application to robustness analysis
    Almerima Jamaković, 2008, Delft University of Technology, Delft, the Netherlands.
21. Shedding light on the black hole, The roll-out of broadband access networks by private operators
    Marieke Fijnvandraat, 2008, Delft University of Technology, Delft, the Netherlands.
22. On stackelberg and inverse stackelberg games & their applications in the optimal toll design problem, the energy markets liberalization problem, and in the theory of incentives
    Kateřina Staňková, 2009, Delft University of Technology, Delft, the Netherlands.
23. On the conceptual design of large-scale process & energy infrastructure systems: integrating flexibility,reliability, availability,maintainability and economics (FRAME) performance metrics
    Austine Ajah, 2009, Delft University of Technology, Delft, the Netherlands.
24. Comprehensive models for security analysis of critical infrastructure as complex systems
    Fei Xue, 2009, Politecnico di Torino, Torino, Italy.
25. Towards a single European electricity market, A structured approach for regulatory mode decision-making
    Hanneke de Jong, 2009, Delft University of Technology, the Netherlands.
26. Co-evolutionary process for modeling large scale socio-technical systems evolution
    Igor Nikolić, 2009, Delft University of Technology, the Netherlands.
27. Regulation in splendid isolation: A framework to promote effective and efficient performance of the electricity industry in small isolated monopoly systems
    Steven Martina, 2009, Delft University of Technology, the Netherlands.

28. Reliability-based dynamic network design with stochastic networks
    Hao Li, 2009, Delft University of Technology, the Netherlands.
29. Competing public values
    Bauke Steenhuisen, 2009, Delft University of Technology, the Netherlands.
30. Innovative contracting practices in the road sector: cross-national lessons in dealing with opportunistic behaviour
    Mónica Altamirano, 2009, Delft University of Technology, the Netherlands.
31. Reliability in urban public transport network assessment and design
    Shahram Tahmasseby, 2009, Delft University of Technology, the Netherlands.
32. Capturing socio-technical systems with agent-based modelling
    Koen van Dam, 2009, Delft University of Technology, the Netherlands.
33. Road incidents and network dynamics, Effects on driving behaviour and traffic congestion
    Victor Knoop, 2009, Delft University of Technology, the Netherlands.
34. Governing mobile service innovation in co-evolving value networks
    Mark de Reuver, 2009, Delft University of Technology, the Netherlands.
35. Modelling risk control measures in railways
    Jaap van den Top, 2009, Delft University of Technology, the Netherlands.
36. Smart heat and power: Utilizing the flexibility of micro cogeneration
    Michiel Houwing, 2010, Delft University of Technology, the Netherlands.
37. Architecture-driven integration of modeling languages for the design of software-intensive systems
    Michel dos Santos Soares, 2010, Delft University of Technology, the Netherlands.
38. Modernization of electricity networks: Exploring the interrelations between institutions and technology
    Martijn Jonker, 2010, Delft University of Technology, the Netherlands.
39. Experiencing complexity: A gaming approach for understanding infrastructure
    Geertje Bekebrede, 2010, Delft University of Technology, the Netherlands.
40. Epidemics in Networks: Modeling, Optimization and Security Games. Technology
    Jasmina Omić, 2010, Delft University of Technology, the Netherlands.
41. Designing Robust Road Networks: A general method applied to the Netherlands
    Maaike Snelder, 2010, Delft University of Technology, the Netherlands.
42. Simulations of Energy Transitions
    Emile Chappin, 2011, Delft University of Technology, the Netherlands.
43. De ingeslagen weg. Een dynamisch onderzoek naar de dynamiek van de uitbesteding van onderhoud in de civiele infrastructuur .
    Rob Schoenmaker, 2011, Delft University of Technology, the Netherlands
44. Safety Management and Risk Modelling in Aviation: the challenge of quantifying management influences.
    Pei-Hui Lin, 2011 Delft University of Technology, the Netherlands.

Order information: info@nextgenerationinfrastructures.eu