TR diss
2202

LEAST SQUARES FILTERING AND TESTING FOR GEODETIC NAVIGATION APPLICATIONS

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus, prof.drs. P.A. Schenck,
in het openbaar te verdedigen ten overstaan van een commissie
aangewezen door het College van Dekanen
op dinsdag 6 april 1993 te 14.00 uur

door

Martin Alfred Salzmann

geboren te Thalwil (Zwitserland),
geodetisch ingenieur.

BIBLIOTHEEK TECHNISCHE UNIVERSITEIT
Promotiecommissie
DELFT

Dit proefschrift is goedgekeurd door de promotoren:
prof. dr. ir. P.J.G. Teunissen en prof. dr. R.F. Rummel.

Abstract

Least Squares Filtering and Testing for Geodetic Navigation Applications

This thesis deals with the data processing, testing, and design procedures for use in dynamic systems, particularly integrated navigation systems, and provides a unified theoretical framework for these procedures. The data processing procedure — the Kalman filter — is analysed from a least squares point of view as this method provides a better understanding of some aspects of the Kalman filter, especially the cases where correlation between the observables is present and for non-linear filtering.

The testing procedure is derived from the theory of hypothesis testing in linear models and is based on generalized likelihood ratio tests, which are shown to be optimal within a certain class of tests. The testing procedure is especially suited for additive model errors in the functional model, and consists of three parts, namely detection, identification, and adaptation (DIA). In the detection step the overall validity of the model is checked and possible model errors are identified in the identification phase. The adaptation step is required to maintain the optimality of the real-time filter in the presence of model errors. The detection and identification steps correspond to the testing procedure used in geodetic network analysis. The DIA procedure allows local and global (covering several epochs) testing and can be implemented recursively, and consequently very efficiently. The DIA procedure constitutes the quality-control step in an overall data processing strategy for dynamic systems.

A design study in which the quality of the system is quantified should precede the implementation of the DIA procedure. In the design procedure the important aspects to consider are the quality of the estimation result under nominal conditions (described by the precision) and the sensitivity of the estimation result to undetected model errors (called the reliability). Reliability of the dynamic system is directly related to the implemented testing procedure. Measures for precision and reliability are discussed. The optimization of the design has to be performed with respect to the precision *and* reliability, and some suggestions for a design procedure for use in dynamic systems are made.

The DIA and design procedures are validated in an extensive simulation study based on a simple linear model and a (hydrographic) navigation system. This study shows that the quality of the system mainly depends on the precision of the observables and the level of integration in the dynamic system. The quality of a system can be improved by using more precise observables and by increasing the level of integration. Based on the design study recommendations on the window lengths of the tests can be given. It is shown that tests for slips require longer window lengths than tests for outliers. The detection and identification steps of the DIA procedure work very well, even in the presence of multiple errors. The adaptation procedure is validated for local tests. Adaptation works well for outliers, but an adaptation procedure for slips should be implemented with care. The testing and design studies show that the correct specification of likely model errors by means of alternative hypotheses is a

crucial element in model validation techniques for dynamic systems.

The least squares (Kalman) filter and the DIA procedure with its associated design procedure are important building blocks of a real-time data processing and quality assurance procedure for dynamic systems.

Acknowledgements

First of all I wish to thank prof. Peter Teunissen for the supervision of this thesis project, and the many discussions that have lead to an improvement of this thesis.

I am much obliged to Dave Wells of the University of New Brunswick for giving me the opportunity to work in Fredericton for a year.

I wish to thank Frank Kenselaar, Anton Kösters, Jos Mulder, Christian Tiberius, and Ronald Verhoeven for the technical discussions and/or their willingness to read parts of the manuscript. Above all I like to thank my fellow graduate students, the research associates, and all other colleagues at the Delft Geodetic Computing Centre for providing an atmosphere in which it is very enjoyable to work.

About the Author

Martin Salzmann was born in 1962 in Thalwil (Switzerland). From 1980 to 1986 he was a student at the department of Geodetic Engineering of Delft University of Technology (DUT). In 1986 he graduated with a thesis on quality aspects of cadastral surveys. From 1986 onwards he has been a research associate at the department of Geodetic Engineering of DUT working on topics related to Kalman filtering. In the period August 1987 to August 1988 he was detached at the department of Surveying Engineering of the University of New Brunswick (Canada). Currently he is participating in the project for a new national technical manual for the cadastral service of the Netherlands.

Samenvatting

Kleinste kwadraten filters en toetsing voor toepassingen in de geodetische navigatie

In dit rapport worden aspecten van de gegevensverwerkings-, toetsings- en ontwerpprocedures voor dynamische systemen (en in het bijzonder voor geïntegreerde navigatiesystemen) behandeld. Doelstelling van het onderzoek is te komen tot een eenduidig theoretisch kader voor de bovenstaande procedures.

De gegevensverwerking, geïmplementeerd met behulp van het Kalman filter, wordt geanalyseerd, uitgaande van de kleinste-kwadraten vereffening. De methode van de kleinste-kwadraten verschaft een beter inzicht in verschillende aspecten van het filter, met name voor de gevallen met correlatie tussen de waarnemingsgrootheden en niet-lineaire waarnemingsvergelijkingen.

De toetsprocedure wordt afgeleid uit de theorie van de hypothesentoetsen in lineaire modellen en is gebaseerd op 'generalized likelihood ratio' toetsen, welke optimaal zijn in een bepaalde klasse van toetsen. De procedure is bijzonder geschikt voor modelfouten met een additief karakter en bestaat uit drie delen, detectie, identificatie en adaptatie en wordt derhalve de DIA-procedure genoemd. In de detectiefase wordt de algemene geldigheid van de modelbeschrijving getoetst, waarna in de identificatiefase de modelfout wordt opgespoord. De adaptatiestap is nodig om in 'real-time' de optimale eigenschappen van de schatters te kunnen waarborgen. De detectie- en identificatiefase komen grotendeels overeen met de toetsprocedure voor landmeetkundige netwerken. De DIA-procedure is geformuleerd voor toetsen die betrekking hebben op een of meerdere tijdstippen en kan recursief en dus efficiënt worden geïmplementeerd. De DIA-procedure garandeert de 'real-time' kwaliteitsbeheersing in de totale gegevensverwerking voor dynamische systemen.

Een ontwerpstudie, waarin de kwaliteit van het systeem gekwantificeerd wordt, hoort vooraf te gaan aan de implementatie van de toetsprocedure. In de ontwerpprocedure wordt de kwaliteit van het systeem onder nominale omstandigheden (beschreven met precisie) en de invloed van onontdekte fouten op de schatters (de zogenaamde betrouwbaarheid) onderzocht. De betrouwbaarheidsbeschrijving is direkt gerelateerd aan de gekozen toetsprocedure. Aan de hand van kwaliteitsmaten voor de precisie en betrouwbaarheid wordt een voorstel gedaan voor een ontwerp-procedure voor dynamische systemen.

De DIA- en ontwerp-procedures worden onderzocht in een uitgebreide simulatiestudie die betrekking heeft op een simpel lineair model en een (hydrografisch) navigatiesysteem. Het blijkt dat de kwaliteit van het systeem vooral wordt bepaald door de precisie van de waarnemingsgrootheden en de mate van integratie van het systeem. Op basis van de ontwerpstudie worden aanbevelingen gedaan over de lengte van de toetsvensters. Toetsen voor blunders hebben een korter toetsvenster nodig dan toetsen voor zogenaamde slips. De detectie- en identificatiestappen van de DIA-procedure werken uitstekend, zelfs indien meerdere fouten optreden. De adaptatie-stap is onder-

zocht voor lokale toetsen en werkt goed voor blunders. Adaptatie voor slips vereist een nauwkeurige modellering. De specificatie van waarschijnlijke modelfouten met behulp van alternatieve hypothesen is een essentieel onderdeel van de model-validatie technieken voor dynamische systemen.

Het kleinste kwadraten filter en de DIA-procedure met de daaraan gerelateerde ontwerp-procedure vormen het fundament voor een gegevensverwerkings- en kwaliteitsbeheersingsprocedure voor dynamische systemen.

Contents

Abstract i

Acknowledgements ii

About the Author ii

Samenvatting iii

1 Introduction 1

2 A Least Squares Approach to Kalman Filtering 7

2.1 Introduction 7
2.2 The System Model and the Linear Kalman Filter 8
2.3 Least Squares Adjustment 11
2.4 The Linear Kalman Filter — A derivation based on least squares for a model with observation equations 15
2.5 The Linear Kalman Filter — A derivation based on least squares for a model with condition equations 19
2.6 Alternative Noise Models 25
2.7 Model Nonlinearities 38
2.8 Concluding Remarks 44

3 A Testing Procedure for Use in Dynamic Systems 47

3.1 Introduction 47
3.2 Hypothesis Testing 48
3.3 A Testing Procedure 54
3.4 Detection 61
3.5 Identification 63
3.6 Adaptation 68
3.7 The B-method of Testing 79
3.8 Recapitulation of the DIA Procedure 81
3.9 Other Techniques for Model Validation 82
3.10 Concluding Remarks 85

4 Design of Dynamic Systems **87**
4.1 Introduction 87
4.2 Design and Quality Assurance 88
4.3 Precision 91
4.4 Reliability 93
4.5 A First Step towards a Design Procedure for Integrated Navigation Systems 107
4.6 Reliability and RAIM 108
4.7 Concluding Remarks 110

5 Design Computations **111**
5.1 Introduction 111
5.2 Quality Measures 112
5.3 Design Setup — Linear Models 114
5.4 Design Results — Linear Models 114
5.5 Design Setup — Navigation Models 120
5.6 Design Results — Navigation Models 126
5.7 Reliability Studies for Integrated Navigation Systems 154
5.8 Concluding Remarks 155

6 DIA Test Computations **157**
6.1 Introduction 157
6.2 Detection and Identification 158
6.3 Adaptation 173
6.4 Concluding Remarks and Recommendations 189

7 Conclusions **191**

References **195**

A The Predicted Residual **201**
A.1 The Predicted Residual 201

B Observation Equations **203**
B.1 Observation Equations 203

C Software Implementation of the Adaptation Procedure **207**
C.1 Adaptation Procedure — Software Implementation 207

Chapter 1

Introduction

This thesis deals with various aspects of high-precision real-time dynamic positioning. Particular attention is paid to the data processing algorithms for real-time position determination and aspects of quality assurance of dynamic positioning. In the following we will introduce the problem definition and will successively discuss the historical context of the current research, the elements of geodetic and navigational methodologies which are at the basis of the research, and the main points of our investigations.

Problem Definition

In society there exists an increasing demand for the real-time accurate determination of position and velocity. Although this trend is partly technology driven, demanding positioning requirements exist in the fields of traffic management (think of the growing traffic densities at sea, on land, and in the air) and in the (more geodesy oriented) fields of photogrammetry, satellite positioning, and resource exploration. The general trend is that the position and velocity not only have to be determined precisely, but also that the quality of the estimation process to obtain position estimates is assured. One of the research objectives at the Delft Geodetic Computing Centre is the development of a real-time data processing procedure for geodetic positioning systems. The data processing covers the estimation of the unknown parameters and the quality assurance of the estimation process. The task we face is the development of a real-time, optimal estimation and testing procedure for use in geodetic, dynamic positioning systems. The optimization of such a procedure is based on a design procedure which will also be considered herein. The procedure is based on a unified framework of the theory of least squares and hypothesis testing in linear models.

Context of the Current Research

This research should be seen in the context of three professional activities, namely land surveying, hydrography, and navigation, all of which are concerned with the determination of position. Position determination can be considered as a process of taking

measurements and computing one's position. Thereupon the land surveyor and hydrographer may use these positions for mapping purposes, whereas the navigator is generally interested in where he is going. Until recently the three professional communities had largely different working methods. The land surveyor and hydrographer, however, shared their methodologies of data processing (by means of least squares adjustment) and quality control. The hydrographer and navigator both worked in a dynamic environment and often used the same positioning systems, and the land surveyor and the navigator, finally, were hardly aware of each other's existence. In the past twenty years this situation has gradually changed because land surveyors, hydrographers, and navigators have had to face similar problems. Land surveying has become much more 'dynamic' and faces a growing demand for (nearly) real-time results. Navigation, at the other end of the spectrum, has to face rapidly increasing accuracy requirements, which have long been a primary concern in land surveying and hydrography. Last but not least, all communities are increasingly relying on the same positioning system, namely the Global Positioning System (GPS). In short land surveyors, hydrographers, and navigators are all becoming, in part, precise positioning specialists. One has now arrived at a point where land surveying, hydrography, and navigation share the following problems:

- The dynamics of the measurement process (even static satellite positioning relies on moving satellites).
- The availablity of a large, continuous stream of data.
- The requirement of real-time quality control.
- The trend towards fully automated data processing.

Considering the problem definition our research is primarily related to the aspects of data processing and quality control.

The Synergism of Geodetic Adjustment Theory, Quality Control and Navigation Techniques

We have seen that precise positioning problems in dynamic environments can be approached from two sides. Firstly one can start from the techniques used in land surveying and adapt them to a dynamic environment. Secondly one can extend the navigation methodologies with the adjustment and testing procedures of land surveying. In this thesis we will use elements of both disciplines. Our starting point is the adjustment and testing procedure of (mathematical) geodesy.

In land surveying it is common practice to work with redundant measurement setups. Surveyors have long been aware that precise measurements do not automatically provide accurate estimation results if errors in the data or other model misspecifications are not detected. Redundant data allow the testing for possible model misspecifications. Besides it is often not possible to reoccupy a measurement station. To obtain

consistent results from a redundant measurement setup surveyors use the least-squares adjustment procedure. BAARDA [1968] was the first to introduce a *systematic* testing procedure in geodesy, namely the B-method of testing. In the B-method tests of various dimensions, encompassing the same alternative hypothesis, have an equal detection power for that specific alternative hypothesis. Closely related to testing is the concept of reliability. Reliability describes the sensitivity of the estimation result to errors that have not been identified by the testing procedure, and can thus be considered as a measure of the quality under the alternative hypothesis. Geodetic measurements are expensive, and consequently design procedures have been developed to optimize the estimation procedure with respect to precision and reliability. The simultaneous optimization (or design) with respect to quality (quality comprises precision and reliability) is part of a larger quality assurance cycle, which also includes quality control (generally implemented by means of a testing procedure) and validation of the adjustment results. In the following we will frequently use elements of geodetic adjustment and testing theory. The reader interested in the ideas underlying geodetic adjustment and testing theory is referred to [BAARDA, 1967, 1968, 1977] and [KOCH, 1988].

From the navigation community we 'borrow' the experience of data processing in real-time and more specifically the Kalman filter [KALMAN, 1960]. The Kalman filter is an estimation (or rather adjustment) procedure to recursively estimate the state (parameters) of a dynamic system. Although the filter has not been developed especially for navigation purposes, it has found its widest use in the navigation environment. The Kalman filter is a well documented estimation algorithm and is covered by numerous textbooks (we refer, for example, to [JAZWINSKI, 1970; GELB, 1974; MAYBECK, 1979,1982]). Developments in the field of navigation have been recorded by KAYTON [1990].

With the Kalman filter, the B-method of testing and its associated design procedure we have available the tools to develop our data processing and design methodology for dynamic systems.

On the Adjustment and Testing for Dynamic Systems

We begin our investigations by considering the optimal estimation procedure for dynamic systems. In our effort to provide a unified framework for the data processing of dynamic systems we consider the recursive estimation of parameters from the viewpoint of least squares. For linear models the least squares adjustment provides us with the best estimators within the class of linear unbiased estimators (see, e.g., [KOCH, 1988]). Best means that one obtains estimators with minimum variance. We will show that the Kalman filter (and many other results that are known from filter theory) can be derived by the method of least squares directly. Under the working (or null) hypothesis the Kalman filter thus constitutes an optimal estimation procedure. We do not derive new filter algorithms, but show that the filter results can be obtained by a simple methodology familiar to surveyors.

After the data processing algorithm (i.e. the Kalman filter as based on the least

squares adjustment) has been established, we consider the operation of the data processing scheme in the presence of model misspecifications, that is under an alternative hypothesis (model misspecifications are specified as alternative hypotheses). Tests to detect and identify model errors will be derived using the theory of hypothesis testing in linear models. Generally one will strive for the most powerful testing method, that is for tests which give the best reliability. For the cases we consider most powerful tests do not exist; a useful subset is however provided by the class of uniformly most powerful invariant (UMPI) tests. It can be shown that the generalized likelihood ratio tests which are used in geodetic testing procedures are UMPI-tests. Our testing procedure also includes the adaptation for model errors. The estimation of the model errors can be derived from the least squares procedure applied to the model under the alternative hypothesis. After adaptation one can often revert to the processing under the null hypothesis.

The implementation of the optimal estimation and the most powerful (under certain conditions) testing procedure in a dynamic system requires a careful design with respect to the precision and reliability criteria. A design study is necessary to give a qualitative description of the system.

Contribution of this Report

This report provides a unified framework for the adjustment and testing procedures for (geodetic) navigation systems based on the theory of least squares and hypothesis testing in linear models. The detection, identification, and adaptation (DIA) procedure is derived and especially the aspect of adaptation is closely investigated. The design procedure for dynamic systems is extended with the aspect of reliability and we will provide a first step towards the generalization of the design procedure for geodetic networks to geodetic navigation systems. A systematic design study for an integrated navigation system based on the measures of precision and reliability is presented. This analysis helps to understand the properties of the various design measures we use and demonstrates how they can be used in (navigation) system design. We extensively investigate the performance of the (local) adaptation procedure and establish its usefulness. Part of the work presented herein has been reported previously in [TEUNISSEN AND SALZMANN, 1988, 1989; TEUNISSEN 1990a, 1990b; SALZMANN 1990, 1991].

Outline of this Report

In CHAPTER 2 we investigate the Kalman filter in the context of least squares adjustment. The chapter deals with the data processing under the null hypothesis. Using the least squares approach many well-known results in filter theory can be derived in a unified manner. The Kalman filter is the data processing algorithm that is used throughout our research. CHAPTER 3 is devoted to the derivation of the DIA procedure. Based on the theory of hypothesis testing in linear models a testing strategy for model misspecifications in the functional model is investigated. We pay much attention

to the adaptation step of the procedure. In CHAPTER 3 we actually consider the data processing under an alternative hypothesis. Quality assurance of dynamic systems is discussed in CHAPTER 4, where the design of dynamic systems with respect to precision and reliability is considered. The concept of reliability is closely related to the testing procedure implemented parallel to the filter. The design methodology proposed in CHAPTER 4 is subsequently applied to a simple linear model and an integrated navigation system in CHAPTER 5. The simple linear model primarily serves to explain certain phenomena that are found in the analysis of the more sophisticated navigation model. In a simulation study in CHAPTER 6 we then apply the DIA procedure to datasets related to the aforementioned linear and navigation models. We discuss the detection and identification phases separately from the adaptation procedure. The conclusions and recommendations resulting from our research are given in CHAPTER 7.

Chapter 2

A Least Squares Approach to Kalman Filtering

2.1 Introduction

In this chapter we consider algorithms for the real-time estimation of parameters in dynamic systems and especially the Kalman filter. The introduction of the Kalman filter [KALMAN, 1960] was an important event in the development of estimation theory. Basically the Kalman filter facilitates the *recursive* estimation of the parameters (or states) of linear, time-varying systems. Like many estimation methods the Kalman filter can be derived from various points of view. SORENSEN [1970a] has given a lucid account of the historical development of the Kalman filter in the context of least squares estimation. In this chapter we will pursue this least squares approach to Kalman filtering and the reasons for this are twofold. Firstly surveyors are very familiar with least squares estimation in general. Secondly, and more importantly, we believe that the principle of least squares constitutes a framework for a unified, comprehensive, and self-contained treatment of filtering problems. Moreover the least squares estimators are the best estimators within the class of linear unbiased estimators. In the following we will show that the least squares approach leads to major results in filtering theory, known from the literature, directly.

Filter algorithms can be derived using probabilistic and deterministic methods (for a discussion the reader is referred to SORENSON [1970b] and MELSA AND COHN [1978]). It is well known that in case one considers *linear* time-varying systems, the least squares, maximum likelihood, minimum mean square error, and maximum a posteriori estimation principles all lead to the same estimator, if the observables are Gaussian. SWERLING [1971] demonstrated that to obtain the Kalman filter the Gaussian assumption does not necessarily have to be made in every step in the derivations. In our least squares approach the stochastic model of the observables is characterized by their first and second (central) moments. Although the careful reader might argue that a derivation of the filter algorithms based on such a limited stochastic model is of little use,

we think that in practice the specification of the first two moments is already difficult enough. If actually the observables can be assumed to be Gaussian (as is often done in the literature), the least squares estimators are equivalent to the 'classical' Kalman filter estimators.

In case the time-varying system under consideration is nonlinear, different estimation principles result in different estimation procedures. Since the theory of nonlinear least squares adjustment is quite well developed, we opt to follow the least squares approach for nonlinear problems as well. The application of nonlinear least squares theory leads to nonlinear filter solutions in a very straightforward manner. Besides one circumvents the cumbersome, nonlinear propagation of probability density functions (see, e.g., [SORENSON, 1970b]).

We will limit ourselves to models formulated in *discrete-time*. In practice the system model, which consists of a dynamic and a measurement model, may be given in continuous-time. Our investigations are based on sampled data systems and thus the continuous to discrete-time conversion of the measurement model does not need to be considered. For the conversion of continuous-time dynamic models into equivalent discrete-time models we refer to [MAYBECK, 1979] or [DECARLO, 1989].

2.1.1 Overview of this Chapter

In Section 2.2 we give an outline of the (linear) dynamic and measurement models and the 'classical' Kalman filter, which is the algorithm commonly used in practice. At this point we introduce the least squares representation of the model and we will pay particular attention to the description of the system noise. In Section 2.3 we give an introductory overview of least squares adjustment. A first derivation of the Kalman filter based on the model with observation equations is given in Section 2.4. Next an alternative derivation, based on the model with condition equations is presented in Section 2.5. In Sections 2.4 and 2.5 we also consider cases related to multiple epochs to provide a link with smoothing. The impact of various stochastic models of the observables is investigated in Section 2.6. In Section 2.7 nonlinearities in the model underlying the Kalman filter and iterative solution strategies are considered. Finally some concluding remarks are given in Section 2.8.

2.2 The System Model and the Linear Kalman Filter

Before the least squares approach to Kalman filtering is discussed, we briefly outline the linear Kalman filter and the model it is based on. We then introduce the system model on which we base our least squares approach and we indicate how in our least-squares approach the system noise (or disturbance) can be looked upon as a discrete-time observable.

2.2.1 The Kalman Filter

We assume that the discrete time dynamic model can be described by the following difference equation

$$\underline{x}_k = \Phi_{k,k-1}\underline{x}_{k-1} + \underline{w}_k \quad , \tag{2.1}$$

where an underscore indicates that a quantity is a random variable and with

$k-1, k$	time indices with $k = 1, 2, 3, \ldots$
$\underline{x}_k$	$n \times 1$ vector of state variables
$\Phi_{k,k-1}$	$n \times n$ state transition matrix
$\underline{w}_k$	$n \times 1$ vector of system noise. (In the literature the contribution of the system noise is sometimes given as $\underline{w}_{k-1}$.)

The measurement model is given by the following equation:

$$\underline{y}_k = A_k\underline{x}_k + \underline{e}_k \quad , \tag{2.2}$$

where

$\underline{y}_k$	$m_k \times 1$ vector of observables
A_k	$m_k \times n$ design matrix
$\underline{e}_k$	$m_k \times 1$ vector of measurement noise.

Observations are not necessarily available at equidistant time intervals and furthermore the number of observations (m_k) may vary with time.

Also the stochastic part of the model has to be specified and at this point we make the customary assumptions that the initial state $\underline{x}_0$ is distributed as $N(x_0, P_0)$ and is uncorrelated with $\underline{w}_k$ and $\underline{e}_k$ for all k. $\underline{w}_k$ is distributed as $N(0, Q_k)$ and $\underline{w}_k$ is uncorrelated with $\underline{w}_l$ for $k \neq l$; $\underline{e}_k$ is distributed as $N(0, R_k)$ and $\underline{e}_k$ is uncorrelated with $\underline{e}_l$ for $k \neq l$; and $\underline{w}_k$ is uncorrelated with $\underline{e}_l$ for all k, l. The matrices P_0 and R_k are positive definite; Q_k is positive semi-definite.

Depending on the application one has in mind, one might wish to obtain an estimate of the state at a certain time k, which depends on all observations taken up to and including time $k + l$. If $l < 0$ the estimation process is called *prediction*. The state estimate then depends on the observations taken prior to the desired time of estimation. If $l = 0$ the process is called *filtering* and in this case the state estimate depends on all the observations prior to and at time k. Finally, if $l > 0$ the process is called *smoothing* and the state estimate depends on the observations taken prior to, at, and after time k.

Since we have primarily real time applications in mind, we shall restrict ourselves in this section to recursive prediction and filtering. The problem we are faced with is to estimate the state at time k using a linear estimator based on all observations up to and including time k. Furthermore the estimate must be 'best' in a certain sense. KALMAN [1960] was the first to solve this problem for the model given by (2.1) and (2.2) using the minimum mean square error criterion.

The (recursive) Kalman filter basically consists of two parts: the *time update* and the *measurement update*. The time update of the state estimate and its associated error covariance matrix are given as:

$$\hat{\underline{x}}_{k|k-1} = \Phi_{k,k-1}\hat{\underline{x}}_{k-1|k-1} \tag{2.3}$$

$$P_{k|k-1} = \Phi_{k,k-1}P_{k-1|k-1}\Phi_{k,k-1}^{\mathrm{T}} + Q_k \ . \tag{2.4}$$

Equation (2.3) gives the estimate of the state at time k using all observations prior to time k. The time update equation is also known as the one-step prediction equation. In [KALMAN, 1960] the predicted state is interpreted as the conditional mean of $\underline{x}_k = \Phi_{k,k-1}\underline{x}_{k-1} + \underline{w}_k$ based on the observables $\underline{y}_i$, for $i = 1, \ldots, k-1$, and $\underline{x}_0$. Since the system noise is assumed to be independent of $\underline{y}_l$ (for all l) and $\underline{x}_0$, its conditional mean equals its unconditional mean, which was assumed to be zero, and thus the time update of the state *estimate* reads $\hat{\underline{x}}_{k|k-1} = \Phi_{k,k-1}\hat{\underline{x}}_{k-1|k-1}$. The measurement update of the state estimate and its associated error covariance matrix are given as:

$$\hat{\underline{x}}_{k|k} = \hat{\underline{x}}_{k|k-1} + K_k(\underline{y}_k - A_k\hat{\underline{x}}_{k|k-1}) \tag{2.5}$$

$$P_{k|k} = (I - K_kA_k)P_{k|k-1} \ , \tag{2.6}$$

where

$$K_k = P_{k|k-1}A_k^{\mathrm{T}}(R_k + A_kP_{k|k-1}A_k^{\mathrm{T}})^{-1} \tag{2.7}$$

is the so-called Kalman gain matrix. The measurement update equation is also known as the filter equation. Equations (2.3) to (2.7) constitute the well-known Kalman filter.

An important role in the filter process is played by the so-called *predicted residual*. The predicted residual is defined as the difference between the actual system output and the predicted output based on the predicted state:

$$\underline{v}_k = \underline{y}_k - A_k\hat{\underline{x}}_{k|k-1} \ . \tag{2.8}$$

The predicted residual represents the new information brought in by the latest observable $\underline{y}_k$. (Therefore the predicted residuals are also called innovations in the literature.) Under the working hypothesis that the mathematical model is specified correctly, the predicted residual has well defined statistical properties, viz:

$$\underline{v}_k \sim N(0, Q_{v_k}) \ , \tag{2.9}$$

where

$$Q_{v_k} = (R_k + A_kP_{k|k-1}A_k^{\mathrm{T}}) \ . \tag{2.10}$$

Note that the predicted residual and its covariance matrix are available during each measurement update.

2.2.2 The System Model in the Least Squares Approach

In the following sections and chapters we will use a model definition which differs slightly from the one given above. We will use the following measurement model:

$$E\{\underline{y}_k\} = A_k \boldsymbol{x}_k \;\; ; \quad D\{\underline{y}_k\} = R_k \;, \tag{2.11}$$

with $k = 1, 2, \ldots$, and where $E\{.\}$ and $D\{.\}$ are the operators of mathematical expectation and dispersion. Note that in (2.11) $\boldsymbol{x}_k$ is a vector of unknown parameters. The observation equation of the initial state is given as:

$$E\{\underline{\boldsymbol{x}}_0\} = \boldsymbol{x}_0 \;\; ; \quad D\{\underline{\boldsymbol{x}}_0\} = P_0 \;. \tag{2.12}$$

We thus only explicitely model the first and second (central) order moments of the observables $\underline{\boldsymbol{x}}_0$ and $\underline{y}_k$. The system noise vector $\underline{w}_k$ was used in (2.1) to model the uncertainty in the state transition and thus constitutes a *disturbance* to the deterministic part of the dynamic model. If we assume that the system noise, which from now on we will denote as the disturbance $\underline{d}_k$, is an observable quantity, it follows from (2.1) that we can formulate the observation equation of the disturbances as

$$E\{\underline{d}_k\} = \boldsymbol{x}_k - \Phi_{k,k-1}\boldsymbol{x}_{k-1} \;\; ; \quad D\{\underline{d}_k\} = Q_k \;. \tag{2.13}$$

In practice the vector $\underline{d}_k$ is *not* observable. The disturbance can be modelled as a random function, that on the average is zero-mean (otherwise the propagation of the state vector in time would not lead to unbiased predictions). We therefore interpret the zero-mean value of the disturbance $\underline{d}_k$ as the *sample* value d_k. In the models (2.11) to (2.13) the system state is considered as a deterministic parameter as a result of which (2.13) can also be interpreted as an observation equation. With equations (2.11) to (2.13) we have now specified a discrete-time model which can be tackled by the least squares method. Because we define the state as a vector of unknown parameters, the matrices $P_{k|k-1}$ and $P_{k|k}$ have to be interpreted as covariance matrices of the predicted and filtered state and not as error covariance matrices.

Note that the model given by (2.11) to (2.13) is equivalent to the description given by (2.1) and (2.2) if $\underline{y}_k$ and $\underline{d}_k$ are Gaussian and are distributed as $N(A_k\boldsymbol{x}_k, R_k)$ and $N(\boldsymbol{x}_k - \Phi_{k,k-1}\boldsymbol{x}_{k-1}, Q_k)$ respectively; $\underline{d}_k$ and $\underline{y}_k$ are mutually uncorrelated; $\underline{d}_k$ is uncorrelated with $\underline{d}_l$ for $k \neq l$; $\underline{y}_k$ is uncorrelated with $\underline{y}_l$ for $k \neq l$; and $\underline{\boldsymbol{x}}_0$ is distributed as $N(\boldsymbol{x}_0, P_0)$.

2.3 Least Squares Adjustment

In this section a synopsis of least squares adjustment will be given. This section serves to introduce the least squares algorithms of interest and to familiarize the reader with the notation we will use in the sequel. We consider least squares adjustment for models with observation equations and models with condition equations. We furthermore discuss the concept of $\underline{y}^R$-variates. The notational conventions introduced in the previous section are maintained.

2.3.1 Model with Observation Equations

The model with observation equations is given as

$$E\{\underline{y}\} = Ax \;\; ; \;\; D\{\underline{y}\} = Q_y \;, \tag{2.14}$$

where $\underline{y}$ is a $m \times 1$ vector of observables; x is a $n \times 1$ vector of unknowns; A is a $m \times n$ design matrix of rank n; and Q_y is a $m \times m$ covariance matrix of the observables of rank m. In the sequel we will often use the shorthand notation $E\{\underline{y}\} = Ax \;;\; Q_y$ instead of (2.14). The observables are written as functions of the unknowns by means of the observation equations. In practice a sample y of the observable $\underline{y}$ is given, and one estimates the unknown vector x. The least squares estimation procedure for the model with observation equations is summarized in Table 2.1.

normal equations
$(A^{\mathrm{T}}Q_y^{-1}A)\hat{\underline{x}} = A^{\mathrm{T}}Q_y^{-1}\underline{y}$
estimators
$\hat{\underline{x}} = (A^{\mathrm{T}}Q_y^{-1}A)^{-1}A^{\mathrm{T}}Q_y^{-1}\underline{y}$ $\hat{\underline{y}} = P_A\underline{y}$ $\hat{\underline{e}} = P_A^{\perp}\underline{y}$
covariance matrices
$Q_{\hat{x}} = (A^{\mathrm{T}}Q_y^{-1}A)^{-1}$ $Q_{\hat{y}} = P_AQ_yP_A^{\mathrm{T}}$ $= P_AQ_y = Q_yP_A^{\mathrm{T}}$ $Q_{\hat{e}} = P_A^{\perp}Q_yP_A^{\perp\mathrm{T}}$ $= P_A^{\perp}Q_y = Q_yP_A^{\perp\mathrm{T}}$
orthogonal projectors
$P_A = A(A^{\mathrm{T}}Q_y^{-1}A)^{-1}A^{\mathrm{T}}Q_y^{-1}$ $P_A^{\perp} = I - A(A^{\mathrm{T}}Q_y^{-1}A)^{-1}A^{\mathrm{T}}Q_y^{-1}$

Table 2.1: Estimation procedure for the model with observation equations.

In many cases the observation equations are nonlinear, i.e.:

$$E\{\underline{y}\} = A(x) \;\; ; \;\; D\{\underline{y}\} = Q_y \;, \tag{2.15}$$

where $A(.)$ is a map of $\mathcal{R}^n$ into $\mathcal{R}^m$. Generally this nonlinear model is approximated using a first order Taylor expansion for the observation equations evaluated at an

approximate value $\boldsymbol{x}_0$ of $\boldsymbol{x}$:

$$A(\boldsymbol{x}) = A(\boldsymbol{x}_0) + \partial_x A(\boldsymbol{x}_0)\Delta\boldsymbol{x} + o(\| \Delta\boldsymbol{x} \|) ,$$

where $\Delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}_0$ and $\partial_x A(\boldsymbol{x}_0)$ is the Jacobian of $A(.)$ evaluated at $\boldsymbol{x}_0$. Using this approximation (we do not go into the discussion on the justification of this approximation at this point) one obtains an estimation procedure for a model with linearized observation equations, if one substitutes $\partial_x A(\boldsymbol{x}_0)$ for A, $\Delta\boldsymbol{x}$ for $\boldsymbol{x}$, and $\Delta\boldsymbol{y} = \boldsymbol{y} - A(\boldsymbol{x}_0)$ for $\boldsymbol{y}$ in the equations given in Table 2.1. Estimators of the parameters $\boldsymbol{x}$ and the observations $\boldsymbol{y}$ are then obtained as $\hat{\underline{\boldsymbol{x}}} = \boldsymbol{x}_0 + \widehat{\underline{\Delta\boldsymbol{x}}}$ and $\hat{\underline{\boldsymbol{y}}} = A(\boldsymbol{x}_0) + \widehat{\underline{\Delta\boldsymbol{y}}}$. To obtain improved estimates the estimation procedure may be iterated, using the latest parameter estimates as new approximate values.

2.3.2 Model with Condition Equations

An *equivalent* (or dual) representation of the model with observation equations can be given by the model with condition equations, where one has to specify the *conditions* the observables have to fulfill. The model with condition equations is given as

$$B^{\mathrm{T}} E\{\underline{\boldsymbol{y}}\} = 0 \;\; ; \;\; D\{\underline{\boldsymbol{y}}\} = Q_y , \tag{2.16}$$

where B^{T} is a $b \times m$ matrix of condition coefficients of rank b. The number of condition equations b is equal to the number of redundant observations $m - n$. Instead of (2.16) we often use the shorthand notation $B^{\mathrm{T}} E\{\underline{\boldsymbol{y}}\} = 0 \;;\; Q_y$. Due to the stochastic nature of the observables (as described by the covariance matrix Q_y) the condition equations are usually not fulfilled. This is described by the $b \times 1$ vector of *misclosures* $\underline{t}$ which is defined as

$$\underline{t} = B^{\mathrm{T}} \underline{\boldsymbol{y}} \;\; ; \;\; Q_t = B^{\mathrm{T}} Q_y B . \tag{2.17}$$

The covariance matrix of the misclosures follows directly from applying the error propagation law. The least squares estimation procedure for the model with condition equations is summarized in Table 2.2.

In practice one often has to deal with nonlinear condition equations

$$B^{\mathrm{T}}(E\{\underline{\boldsymbol{y}}\}) = 0 \;\; ; \;\; D\{\underline{\boldsymbol{y}}\} = Q_y , \tag{2.18}$$

where $B^{\mathrm{T}}(.)$ is a map of $\mathcal{R}^m$ into $\mathcal{R}^b$. If we approximate the nonlinear condition equations by a first order Taylor expansion at y_0 (the best value for y_0 being the observation y itself), we obtain the linearized model with observation equations

$$\partial_y B^{\mathrm{T}}(y_0) E\{\underline{\Delta y}\} = 0 \;\; ; \;\; D\{\underline{\Delta y}\} = Q_y ,$$

where $\Delta y = y - y_0$ and $\partial_y B^{\mathrm{T}}(y_0)$ is the Jacobian of $B^{\mathrm{T}}(.)$ evaluated at y_0. The estimation procedure given in Table 2.2 can be used for the model with linearized condition equations by substituting $\partial_y B^{\mathrm{T}}(y_0)$ for B^{T} and Δy for y. The estimator for the observations y then reads $\hat{\underline{y}} = y_0 + \widehat{\underline{\Delta y}}$.

The connection of the models with observation equations and condition equations is given in Table 2.3. In the nonlinear case it holds that $B^{\mathrm{T}}(A(\boldsymbol{x})) = 0$.

estimators
$\underline{\hat{y}} = P^{\perp}_{Q_yB}\underline{y}$ $\underline{\hat{e}} = P_{Q_yB}\underline{y}$
covariance matrices
$Q_{\hat{y}} = P^{\perp}_{Q_yB}Q_yP^{\perp\mathrm{T}}_{Q_yB}$ $= P^{\perp}_{Q_yB}Q_y = Q_yP^{\perp\mathrm{T}}_{Q_yB}$ $Q_{\hat{e}} = P_{Q_yB}Q_yP^{\mathrm{T}}_{Q_yB}$ $= P_{Q_yB}Q_y = Q_yP^{\mathrm{T}}_{Q_yB}$
orthogonal projectors
$P_{Q_yB} = Q_yB(B^{\mathrm{T}}Q_yB)^{-1}B^{\mathrm{T}}$ $P^{\perp}_{Q_yB} = I - Q_yB(B^{\mathrm{T}}Q_yB)^{-1}B^{\mathrm{T}}$

Table 2.2: Estimation procedure for the model with condition equations.

$\mathrm{rank}(A)$	$=$	n
$\mathrm{rank}(B^{\mathrm{T}})$	$=$	b
b	$=$	$m-n$
B^{T} A	$=$	0
$b\times m$ $m\times n$		$b\times n$
P_A	$=$	$P^{\perp}_{Q_yB}$
$P^{\perp}_A$	$=$	P_{Q_yB}

Table 2.3: Relation between models with observation and condition equations.

2.3.3 $\underline{y}^R$-Variates

Now the least squares estimators based on the models with observation and condition equations have been derived, we introduce an important concept in adjustment theory, namely the $\underline{y}^R$-variates [BAARDA, 1967]. $\underline{y}^R$-Variates are observables that are either stochastically or functionally related to another set of observables $\underline{y}$. In practice one makes a distinction between the following types of $\underline{y}^R$-variates: *free* variates which

are correlated with $\underline{y}$-variates; *derived* variates which are functions of $\underline{y}$-variates; and *constituent* variates, variates of which the $\underline{y}$-variates are functions. In the sequel we are mainly concerned with constituent variates. In the notation associated with $\underline{y}^R$-variates the functional relationship between the $\underline{y}$-variates and the constituent $\underline{y}^R$-variates is given as:

$$\underline{y} = \Lambda \underline{y}^R \quad . \tag{2.19}$$

The least squares estimators of the corrections of the $\underline{y}$- and $\underline{y}^R$-variates are related by the well-known formula

$$\underline{\hat{e}}^R = Q_{Ry} Q_y^{-1} \underline{\hat{e}} \quad . \tag{2.20}$$

Equation (2.20) is valid for all three types of $\underline{y}^R$-variates, but is derived here for the case of constituent variates. Application of the propagation laws of variance and covariance to (2.19) gives

$$Q_y = \Lambda Q_R \Lambda^{\mathrm{T}} \quad \text{and} \quad Q_{Ry} = Q_R \Lambda^{\mathrm{T}} \; .$$

The model with condition equations (2.16) can, using (2.19), also be written as

$$B^{\mathrm{T}} \Lambda E\{\underline{y}^R\} = 0; \quad D\{\underline{y}^R\} = Q_R,$$

which, by applying the least squares estimation procedure for models with condition equations, leads to the following least squares estimator of the corrections:

$$\underline{\hat{e}}^R = Q_R \Lambda^{\mathrm{T}} B (B^{\mathrm{T}} \Lambda Q_R \Lambda^{\mathrm{T}} B)^{-1} B^{\mathrm{T}} \Lambda \underline{y}^R \; .$$

Noting that $Q_R \Lambda^{\mathrm{T}} = Q_{Ry}$, $B(B^{\mathrm{T}} \Lambda Q_R \Lambda^{\mathrm{T}} B)^{-1} B^{\mathrm{T}} = B(B^{\mathrm{T}} Q_y B)^{-1} B^{\mathrm{T}} = Q_y^{-1} P_{Q_y B}$, and $\Lambda \underline{y}^R = \underline{y}$, (2.20) follows immediately.

2.4 The Linear Kalman Filter — A derivation based on least squares for a model with observation equations

The objective of this and the succeeding section is to show that the prediction, filtering, and smoothing formulas found in the literature can easily be derived using the least squares approach. First we present the model we use for our derivations. This model takes into account the system state at times $k-1$ and k and is a model with observation equations. The Kalman filter model is expressed in state-space and the states coincide with the parameters of a model with observation equations.

The linear model of observation equations from which the linear Kalman filter can be derived is given as[1]

$$E\{\begin{pmatrix} \underline{\hat{x}}_{k-1|k-1} \\ \underline{d}_k \\ \underline{y}_k \end{pmatrix}\} = \begin{pmatrix} I & 0 \\ -\Phi_{k,k-1} & I \\ 0 & A_k \end{pmatrix} \begin{pmatrix} x_{k-1} \\ x_k \end{pmatrix} \; ; \tag{2.21}$$

[1] whenever matrices appear with missing elements, then those elements are zero

$$\begin{pmatrix} P_{k-1|k-1} & & \\ & Q_k & \\ & & R_k \end{pmatrix} ,$$

which is of the form $E\{\underline{y}\} = Ax \; ; \; Q_y$. The following derivation is closely patterned after [TEUNISSEN AND SALZMANN, 1988].

Prediction

For prediction one considers the estimation of the state without the use of the observables $\underline{y}_k$. Then (2.21) reduces to

$$E\{\begin{pmatrix} \hat{\underline{x}}_{k-1|k-1} \\ \underline{d}_k \end{pmatrix}\} = \begin{pmatrix} I & 0 \\ -\Phi_{k,k-1} & I \end{pmatrix} \begin{pmatrix} x_{k-1} \\ x_k \end{pmatrix} \; ; \begin{pmatrix} P_{k-1|k-1} & \\ & Q_k \end{pmatrix} . \tag{2.22}$$

Equation (2.22) can be solved immediately as there is no redundancy in this model. Hence the available estimate $\hat{x}_{k-1|k-1}$ of x_{k-1} cannot be improved upon. The least squares estimator of x_k, which is denoted by $\hat{\underline{x}}_{k|k-1}$, follows directly from inverting the design matrix of (2.22). The inverse of the design matrix is:

$$\begin{pmatrix} I & 0 \\ -\Phi_{k,k-1} & I \end{pmatrix}^{-1} = \begin{pmatrix} I & 0 \\ \Phi_{k,k-1} & I \end{pmatrix} .$$

Hence we obtain the estimator of the *predicted* state

$$\hat{\underline{x}}_{k|k-1} = \Phi_{k,k-1}\hat{\underline{x}}_{k-1|k-1} + \underline{d}_k \; . \tag{2.23}$$

As we take the sample value d_k of the disturbance $\underline{d}_k$ to be equal to zero the least squares estimate of the predicted state is

$$\hat{x}_{k|k-1} = \Phi_{k,k-1}\hat{x}_{k-1|k-1} \; . \tag{2.24}$$

The covariance matrix of the estimator of the predicted state is obtained by application of the error propagation law

$$P_{k|k-1} = \Phi_{k,k-1}P_{k-1|k-1}\Phi_{k,k-1}^{\mathrm{T}} + Q_k \; . \tag{2.25}$$

Equations (2.24) and (2.25) constitute the *time update* (or prediction) equations of the Kalman filter (cf. eqs. 2.3 and 2.4).

Filtering

If we include the observables at time k in the estimation process we obtain the estimator of the filtered state. This leads to the following model with observation equations:

$$E\{\begin{pmatrix} \hat{\underline{x}}_{k|k-1} \\ \underline{y}_k \end{pmatrix}\} = \begin{pmatrix} I \\ A_k \end{pmatrix} x_k \; ; \begin{pmatrix} P_{k|k-1} & \\ & R_k \end{pmatrix} . \tag{2.26}$$

Straightforward application of the least squares algorithm (cf. Section 2.3) gives for the estimator of the state at time k and its covariance matrix

$$\underline{\hat{x}}_{k|k} = (P_{k|k-1}^{-1} + A_k^{\mathrm{T}} R_k^{-1} A_k)^{-1} (P_{k|k-1}^{-1} \underline{\hat{x}}_{k|k-1} + A_k^{\mathrm{T}} R_k^{-1} \underline{y}_k) \tag{2.27}$$

$$P_{k|k} = (P_{k|k-1}^{-1} + A_k^{\mathrm{T}} R_k^{-1} A_k)^{-1} \ . \tag{2.28}$$

Application of the matrix inversion lemma

$$(P_{k|k-1}^{-1} + A_k^{\mathrm{T}} R_k^{-1} A_k)^{-1} = P_{k|k-1} - P_{k|k-1} A_k^{\mathrm{T}} (R_k + A_k P_{k|k-1} A_k^{\mathrm{T}})^{-1} A_k P_{k|k-1}$$

to (2.27) and (2.28) gives after some rearrangements the *measurement update* (or filter) equations (2.5) to (2.7). We thus derived the Kalman filter algorithm using standard least squares methods.

In the literature also an alternative form of the Kalman gain matrix is given. Although this alternative form is identical to (2.7), we will nevertheless briefly indicate how it can be derived directly using the least squares approach. If one inserts (2.28) into (2.27) one obtains

$$\underline{\hat{x}}_{k|k} = P_{k|k} P_{k|k-1}^{-1} \underline{\hat{x}}_{k|k-1} + P_{k|k} A_k^{\mathrm{T}} R_k^{-1} \underline{y}_k \ ,$$

which can be rearranged as

$$\begin{aligned} \underline{\hat{x}}_{k|k} &= P_{k|k}(P_{k|k-1}^{-1} + A_k^{\mathrm{T}} R_k^{-1} A_k) \underline{\hat{x}}_{k|k-1} + P_{k|k} A_k^{\mathrm{T}} R_k^{-1} (\underline{y}_k - A_k \underline{\hat{x}}_{k|k-1}) \\ &= \underline{\hat{x}}_{k|k-1} + P_{k|k} A_k^{\mathrm{T}} R_k^{-1} (\underline{y}_k - A_k \underline{\hat{x}}_{k|k-1}) \ , \end{aligned}$$

so that the alternative form of the gain matrix is (cf. eq. 2.7):

$$K_k = P_{k|k} A_k^{\mathrm{T}} R_k^{-1} \ . \tag{2.29}$$

Smoothing

In the derivations above we separately derived the time and measurement update equations of the Kalman filter. It is, however, also possible to consider model (2.21) as a single adjustment problem. The estimator of the filtered state ($\underline{\hat{x}}_{k|k}$) will be identical to the form given by (2.27), but an estimator of the *smoothed* state at time $k-1$ is obtained, because we also take information *after* time $k-1$ into account.

At this point two angles of attack are possible to obtain an estimator of the smoothed state at time $k-1$. The direct approach is to start from (2.21) and to derive the normal equations from the observation equations

$$\begin{pmatrix} P_{k-1|k-1}^{-1} + \Phi_{k,k-1}^{\mathrm{T}} Q_k^{-1} \Phi_{k,k-1} & -\Phi_{k,k-1}^{\mathrm{T}} Q_k^{-1} \\ -Q_k^{-1} \Phi_{k,k-1} & Q_k^{-1} + A_k^{\mathrm{T}} R_k^{-1} A_k \end{pmatrix} \begin{pmatrix} \underline{\hat{x}}_{k-1} \\ \underline{\hat{x}}_k \end{pmatrix} =$$

$$\begin{pmatrix} P_{k-1|k-1}^{-1} \underline{\hat{x}}_{k-1|k-1} + \Phi_{k,k-1}^{\mathrm{T}} Q_k^{-1} \underline{d}_k \\ Q_k^{-1} \underline{d}_k + A_k^{\mathrm{T}} R_k^{-1} \underline{y}_k \end{pmatrix} \ . \tag{2.30}$$

By solving the normal equations (which requires some lengthly algebraic manipulations, including repeated use of the matrix inversion lemma) one obtains the estimator of the one-step delayed smoothed state and its covariance matrix as

$$\hat{\underline{x}}_{k-1|k} = \hat{\underline{x}}_{k-1|k-1} + J_{k-1}(\hat{\underline{x}}_{k|k} - \hat{\underline{x}}_{k|k-1}) \tag{2.31}$$

$$D\{\hat{\underline{x}}_{k-1|k}\} = P_{k-1|k-1} + J_{k-1}(P_{k|k} - P_{k|k-1})J_{k-1}^{\mathrm{T}} \tag{2.32}$$

with $J_{k-1} = P_{k-1|k-1}\Phi_{k,k-1}^{\mathrm{T}}P_{k|k-1}^{-1}$. The estimator of the smoothed state and its covariance matrix are equivalent to the forms found in the literature for a one-step delayed smoothed solution. The direct least squares approach via the normal equations makes additionaly available the covariance between the estimator of the filtered state and the estimator of the smoothed state. A drawback of the derivation via the normal equations is that explicit use is made of the inverse of the covariance matrix of the disturbances (Q_k). This inverse does not necessarily exist as Q_k is only required to be positive *semi*-definite. Therefore we look for an alternative derivation which avoids the explicit computation of the inverse of Q_k.

This alternative, more elegant, derivation is obtained using the $\underline{y}^R$-variates. In the case of smoothing we have to deal with so-called *constituent* variates. From equation (2.23) it follows that $\hat{\underline{x}}_{k-1|k-1}$ and $\underline{d}_k$ are the constituent variates of $\hat{\underline{x}}_{k|k-1}$, or

$$\hat{\underline{x}}_{k|k-1} = \begin{pmatrix} \Phi_{k,k-1} & I \end{pmatrix} \begin{pmatrix} \hat{\underline{x}}_{k-1|k-1} \\ \underline{d}_k \end{pmatrix} \quad . \tag{2.33}$$

In the notation associated with $\underline{y}^R$-variates (2.33) can be written as $\underline{y} = A\underline{y}^R$. After the least squares estimator of the corrections to the $\underline{y}^R$-variates is obtained, the estimator of the smoothed state follows automatically. The estimator of the corrections is given by (2.20):

$$\hat{\underline{e}}^R = Q_{Ry}Q_y^{-1}\hat{\underline{e}} \quad ,$$

where

$$\begin{aligned} Q_y &= P_{k|k-1} \text{ (cf. eq. 2.25)} \\ Q_{Ry} &= \begin{pmatrix} P_{k-1|k-1}\Phi_{k,k-1}^{\mathrm{T}} \\ Q_k \end{pmatrix} \\ \hat{\underline{e}} &= \hat{\underline{x}}_{k|k-1} - \hat{\underline{x}}_{k|k} \quad , \end{aligned}$$

and thus

$$\hat{\underline{e}}^R = \begin{pmatrix} P_{k-1|k-1}\Phi_{k,k-1}^{\mathrm{T}}P_{k|k-1}^{-1} \\ Q_kP_{k|k-1}^{-1} \end{pmatrix} (\hat{\underline{x}}_{k|k-1} - \hat{\underline{x}}_{k|k}) \quad .$$

If we restrict ourselves to the estimator of the correction to the $\underline{y}^R$-variate $\hat{\underline{x}}_{k-1|k-1}$ one obtains as the estimator for the smoothed state at time $k-1$:

$$\hat{\underline{x}}_{k-1|k} = \hat{\underline{x}}_{k-1|k-1} + P_{k-1|k-1}\Phi_{k,k-1}^{\mathrm{T}}P_{k|k-1}^{-1}(\hat{\underline{x}}_{k|k} - \hat{\underline{x}}_{k|k-1}) \quad , \tag{2.34}$$

which is identical to (2.31).

In this section it was demonstrated that the recursive prediction, filter, and (one-step) smoothing equations can be derived using a least squares approach based on model (2.21). If, however, model (2.21) is extended to multiple epochs, derivations based on the least squares approach do not result automatically in convenient recursive formulations of the prediction, filtering, and smoothing problems anymore. A drawback of the derivation based on models with observation equations is that analytic solutions are hard to obtain if many epochs are considered simultaneously, because every additional epoch increases the dimension of the normal matrix (which is usually a completely full matrix and has to be inverted) by the dimension of the state vector. Therefore we also investigate an alternative approach based on the model with condition equations.

2.5 The Linear Kalman Filter — A derivation based on least squares for a model with condition equations

In Section 2.3 it was stated that for a model with observation equations an equivalent model with condition equations can be found. The derivation of the Kalman filter based on a model with condition equations provides us with an alternative to the approach with observation equations, which becomes rather intractable for more than two epochs.

Prediction

To obtain the prediction formula we start from (2.22). For the prediction case one cannot obtain a model with only condition equations because not all unknowns can be eliminated. After eliminating $\boldsymbol{x}_{k-1}$ one obtains

$$\begin{pmatrix} \Phi_{k,k-1} & I \end{pmatrix} E\{\begin{pmatrix} \underline{\hat{\boldsymbol{x}}}_{k-1|k-1} \\ \underline{d}_k \end{pmatrix}\} = \boldsymbol{x}_k \ , \tag{2.35}$$

which leads directly to the prediction formula (2.23).

Filtering

If the observations at time k are included in the estimation process one obtains from (2.26), after eliminating the unknown $\boldsymbol{x}_k$:

$$\begin{pmatrix} -A_k & I \end{pmatrix} E\{\begin{pmatrix} \underline{\hat{\boldsymbol{x}}}_{k|k-1} \\ \underline{y}_k \end{pmatrix}\} = 0 \tag{2.36}$$

which is of the form $B^{\mathrm{T}} E\{\underline{y}\} = 0$, and where B^{T} is a coefficient matrix of dimension $m_k \times (n+m_k)$. It can easily be verified that the property $B^{\mathrm{T}} A = 0$ holds. Application of the least squares algorithm for a model with condition equations gives for the estimators

of the corrections

$$\begin{pmatrix} \hat{\underline{e}}_x \\ \hat{\underline{e}}_y \end{pmatrix} = \begin{pmatrix} -P_{k|k-1}A_k^{\mathrm{T}}(R_k + A_k P_{k|k-1} A_k^{\mathrm{T}})^{-1} \\ R_k(R_k + A_k P_{k|k-1} A_k^{\mathrm{T}})^{-1} \end{pmatrix} (\underline{y}_k - A_k \underline{\hat{x}}_{k|k-1}) \tag{2.37}$$

and thus for the filtered state estimator (cf. eq. 2.5):

$$\begin{aligned} \underline{\hat{x}}_{k|k} &= \underline{\hat{x}}_{k|k-1} + P_{k|k-1}A_k^{\mathrm{T}}(R_k + A_k P_{k|k-1} A_k^{\mathrm{T}})^{-1}(\underline{y}_k - A_k \underline{\hat{x}}_{k|k-1}) \\ &= \underline{\hat{x}}_{k|k-1} + K_k(\underline{y}_k - A_k \underline{\hat{x}}_{k|k-1}) \ . \end{aligned} \tag{2.38}$$

Smoothing

If one starts from model (2.21) a model with condition equations can be obtained by subsequent elimination of x_k (via $y' = y_k - A_k d_k$) and x_{k-1} (via $y'' = y' - A_k \Phi_{k,k-1} \hat{x}_{k-1|k-1}$), as follows:

$$\begin{pmatrix} -A_k\Phi_{k,k-1} & -A_k & I \end{pmatrix} E\{\begin{pmatrix} \underline{\hat{x}}_{k-1|k-1} \\ \underline{d}_k \\ \underline{y}_k \end{pmatrix}\} = 0; \begin{pmatrix} P_{k-1|k-1} & & \\ & Q_k & \\ & & R_k \end{pmatrix}, \tag{2.39}$$

where B^{T} is of dimension $m_k \times (2n + m_k)$. One can easily verify that the property $B^{\mathrm{T}}A = 0$ holds. If one follows a direct approach based on the least squares algorithm given in Section 2.3, one obtains the estimator (2.31) for the state at time $k-1$ after some algebraic manipulations.

Here we use the alternative solution based on $\underline{y}^R$-variates (we repeat the derivation, because in the model based on condition equations the estimators of the least squares corrections are given in a somewhat different form). The constituent variates can be described by (2.33). For the model with condition equations it follows from (2.37) with

$$\hat{\underline{e}}^R = Q_{Ry} Q_y^{-1} \hat{\underline{e}} \ ,$$

where Q_y and Q_{Ry} are as defined before, but

$$\begin{aligned} \hat{\underline{e}} &= -P_{k|k-1}A_k^{\mathrm{T}}(R_k + A_k P_{k|k-1} A_k^{\mathrm{T}})^{-1}(\underline{y}_k - A_k \underline{\hat{x}}_{k|k-1}) \\ &= -K_k(\underline{y}_k - A_k \underline{\hat{x}}_{k|k-1}) \ , \end{aligned}$$

that the estimator of the corrections to the $\underline{y}^R$-variate $\underline{\hat{x}}_{k-1|k-1}$ is

$$\begin{aligned} \hat{\underline{e}}_x^R &= -P_{k-1|k-1}\Phi_{k,k-1}^{\mathrm{T}} P_{k|k-1}^{-1} K_k(\underline{y}_k - A_k \underline{\hat{x}}_{k|k-1}) \\ &= -P_{k-1|k-1}\Phi_{k,k-1}^{\mathrm{T}} P_{k|k-1}^{-1}(\underline{\hat{x}}_{k|k} - \underline{\hat{x}}_{k|k-1}) \ . \end{aligned}$$

This results in an estimator identical to (2.31) and (2.34), namely

$$\underline{\hat{x}}_{k-1|k} = \underline{\hat{x}}_{k-1|k-1} + P_{k-1|k-1}\Phi_{k,k-1}^{\mathrm{T}} P_{k|k-1}^{-1}(\underline{\hat{x}}_{k|k} - \underline{\hat{x}}_{k|k-1}) \ . \tag{2.40}$$

2.5.1 Multiple epoch solutions

We now try to expand our derivation of the Kalman filter to more than two epochs using the model with condition equations. If one is able to find an analytic form valid for any number of epochs this will greatly facilitate all subsequent derivations. The model with condition equations is basically non-parametric and hence a link with the parametric state-space concept of the Kalman filter might cause some problems.

If we expand model (2.21) to, for example, three epochs, one obtains:

$$E\{\begin{pmatrix} \hat{\underline{x}}_{k-1|k-1} \\ \underline{d}_k \\ \underline{y}_k \\ \underline{d}_{k+1} \\ \underline{y}_{k+1} \end{pmatrix}\} = \begin{pmatrix} I_n & 0 & 0 \\ -\Phi_{k,k-1} & I_n & 0 \\ 0 & A_k & 0 \\ 0 & -\Phi_{k+1,k} & I_n \\ 0 & 0 & A_{k+1} \end{pmatrix} \begin{pmatrix} x_{k-1} \\ x_k \\ x_{k+1} \end{pmatrix} ;$$

$$\begin{pmatrix} P_{k-1|k-1} & & & & \\ & Q_k & & & \\ & & R_k & & \\ & & & Q_{k+1} & \\ & & & & R_{k+1} \end{pmatrix} . \qquad (2.41)$$

In order to write model (2.41) as a model with condition equations we have to eliminate the parameters. We begin by eliminating x_{k-1} and x_{k+1}, so that (2.41) reduces to:

$$E\{\begin{pmatrix} \underline{d}_k + \Phi_{k,k-1}\hat{\underline{x}}_{k-1|k-1} \\ \underline{y}_k \\ \underline{y}_{k+1} - A_{k+1}\underline{d}_{k+1} \end{pmatrix}\} = \begin{pmatrix} I_n \\ A_k \\ A_{k+1}\Phi_{k+1,k} \end{pmatrix} x_k .$$

By eliminating x_k one then arrives at the familiar model $B^{\mathrm{T}}E\{\underline{y}\} = 0$, where

$$B^{\mathrm{T}} = \begin{pmatrix} -A_k\Phi_{k,k-1} & -A_k & I_{m_k} & 0 & 0 \\ -A_{k+1}\Phi_{k+1,k-1} & -A_{k+1}\Phi_{k+1,k} & 0 & -A_{k+1} & I_{m_{k+1}} \end{pmatrix} ,$$

where m_k and m_{k+1} are the number of observations at time k and $k+1$ respectively, and

$$\underline{y} = (\hat{\underline{x}}^{\mathrm{T}}_{k-1|k-1}, \underline{d}^{\mathrm{T}}_k, \underline{y}^{\mathrm{T}}_k, \underline{d}^{\mathrm{T}}_{k+1}, \underline{y}^{\mathrm{T}}_{k+1})^{\mathrm{T}} .$$

For an arbitrary number of epochs (starting at time $k-1$) the model with condition equations can be written as:

$$B^{\mathrm{T}}E\{\underline{y}\} = 0; \quad Q_y , \qquad (2.42)$$

where

$$B^{\mathrm{T}} = \begin{pmatrix} -A_k\Phi_{k,k-1} & -A_k & I_{m_k} & 0 & \cdots \\ -A_{k+1}\Phi_{k+1,k-1} & -A_{k+1}\Phi_{k+1,k} & 0 & -A_{k+1} & I_{m_{k+1}} & \cdots \\ -A_{k+2}\Phi_{k+2,k-1} & -A_{k+2}\Phi_{k+2,k} & 0 & -A_{k+2}\Phi_{k+2,k+1} & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} ,$$

$$\underline{y} = (\underline{\hat{x}}_{k-1|k-1}^{\mathrm{T}}, \underline{d}_k^{\mathrm{T}}, \underline{y}_k^{\mathrm{T}}, \underline{d}_{k+1}^{\mathrm{T}}, \underline{y}_{k+1}^{\mathrm{T}}, \ldots)^{\mathrm{T}}$$

and

$$Q_y = \begin{pmatrix} P_{k-1|k-1} & & & & & \\ & Q_k & & & & \\ & & R_k & & & \\ & & & Q_{k+1} & & \\ & & & & \ddots & \end{pmatrix} .$$

The vector of misclosures $\underline{t}$ is defined as:

$$\underline{t} = B^{\mathrm{T}} \underline{y} \ , \tag{2.43}$$

where the elements of the closure vector $\underline{t}$ are:

$$\begin{aligned}
\underline{t}_k &= \underline{y}_k - A_k(\underline{d}_k + \Phi_{k,k-1}\underline{\hat{x}}_{k-1|k-1}) \\
\underline{t}_{k+1} &= \underline{y}_{k+1} - A_{k+1}\underline{d}_{k+1} - A_{k+1}\Phi_{k+1,k}\underline{d}_k - A_{k+1}\Phi_{k+1,k-1}\underline{\hat{x}}_{k-1|k-1} \\
&= \underline{y}_{k+1} - A_{k+1}(\underline{d}_{k+1} + \Phi_{k+1,k}(\underline{d}_k + \Phi_{k,k-1}\underline{\hat{x}}_{k-1|k-1})) \\
\underline{t}_{k+2} &= \underline{y}_{k+2} - A_{k+2}(\underline{d}_{k+2} + \Phi_{k+2,k+1}(\underline{d}_{k+1} + \Phi_{k+1,k}(\underline{d}_k + \Phi_{k,k-1}\underline{\hat{x}}_{k-1|k-1}))) \\
&(etc.)
\end{aligned}$$

Model (2.42), together with (2.43), gives a complete description of the multiple epoch filtering and smoothing problem. However, the structure of the misclosure vector $\underline{t}$ does not immediately lead to major simplifications in the (analytic) solution of the adjustment problem, because the matrix $B^{\mathrm{T}}Q_yB$ which has to be inverted is completely full. Therefore it is required that this model is further developed. It can be seen from (2.43) that the misclosure vector at time k is equivalent to the predicted residual $\underline{v}_k$ $(= \underline{y}_k - A_k\underline{\hat{x}}_{k|k-1})$. In the Kalman filter the predicted residuals and their covariance matrix are readily available. So if we are able to express all elements of the misclosure vector as functions of the predicted residuals this might lead to major simplifications. In the sequel we will show that the vector of misclosures can indeed be written as a function of the predicted residuals.

In order to write the misclosure vector as a function of the predicted residuals, we start with rewriting the misclosure vector as:

$$\begin{aligned}
\underline{t}_k &= \underline{v}_k \\
\underline{t}_{k+1} &= \underline{v}_{k+1} + A_{k+1}(\underline{\hat{x}}_{k+1|k} - \underline{\hat{x}}_{k+1|k-1}) \\
\underline{t}_{k+2} &= \underline{v}_{k+2} + A_{k+2}(\underline{\hat{x}}_{k+2|k+1} - \underline{\hat{x}}_{k+2|k-1}) \\
&(etc.)
\end{aligned}$$

or

$$\begin{aligned}
\underline{t}_k &= \underline{v}_k \\
\underline{t}_{k+1} &= \underline{v}_{k+1} + A_{k+1}\Phi_{k+1,k}(\underline{\hat{x}}_{k|k} - \underline{\hat{x}}_{k|k-1})
\end{aligned}$$

$$\begin{aligned}
\underline{t}_{k+2} &= \underline{v}_{k+2} + A_{k+2}\Phi_{k+2,k+1}(\hat{\underline{x}}_{k+1|k+1} - \hat{\underline{x}}_{k+1|k-1}) \\
&= \underline{v}_{k+2} + A_{k+2}\Phi_{k+2,k+1}(\hat{\underline{x}}_{k+1|k+1} - \hat{\underline{x}}_{k+1|k} + \hat{\underline{x}}_{k+1|k} - \hat{\underline{x}}_{k+1|k-1}) \\
&= \underline{v}_{k+2} + A_{k+2}\Phi_{k+2,k+1}((\hat{\underline{x}}_{k+1|k+1} - \hat{\underline{x}}_{k+1|k}) + \Phi_{k+1,k}(\hat{\underline{x}}_{k|k} - \hat{\underline{x}}_{k|k-1})) \\
&\quad (etc.)
\end{aligned}$$

which can be verified by simple substitution. Recalling that (cf. eq. 2.5)

$$\hat{\underline{x}}_{i|i} = \hat{\underline{x}}_{i|i-1} + K_i\underline{v}_i \quad ,$$

the elements of the misclosure vector $\underline{t}$ can then be written as

$$\begin{aligned}
\underline{t}_k &= \underline{v}_k \\
\underline{t}_{k+1} &= \underline{v}_{k+1} + A_{k+1}\Phi_{k+1,k}K_k\underline{v}_k \\
\underline{t}_{k+2} &= \underline{v}_{k+2} + A_{k+2}\Phi_{k+2,k+1}K_{k+1}\underline{v}_{k+1} + A_{k+2}\Phi_{k+2,k}K_k\underline{v}_k \\
&\quad (etc.)
\end{aligned}$$

We can thus express the misclosure vector as a linear function of the predicted residuals:

$$\underline{t} = L\underline{v} \quad , \tag{2.44}$$

where

$$L = \begin{pmatrix}
I_{m_k} & 0 & 0 & \cdots \\
A_{k+1}\Phi_{k+1,k}K_k & I_{m_{k+1}} & 0 & \cdots \\
A_{k+2}\Phi_{k+2,k}K_k & A_{k+2}\Phi_{k+2,k+1}K_{k+1} & I_{m_{k+2}} & \cdots \\
\vdots & \vdots & \vdots & \ddots
\end{pmatrix} ,$$

and

$$\underline{v} = (\underline{v}_k^{\mathrm{T}}, \underline{v}_{k+1}^{\mathrm{T}}, \underline{v}_{k+2}^{\mathrm{T}}, \ldots)^{\mathrm{T}} \quad .$$

The matrix L is a lower triangular matrix of full rank, and consequently its inverse can be computed in a straightforward manner. Hence the predicted residuals can also be expressed as a linear function of the misclosures

$$\underline{v} = L^{-1}\underline{t} \quad , \tag{2.45}$$

where

$$L^{-1} = \begin{pmatrix}
I_{m_k} & 0 & 0 & \cdots \\
-A_{k+1}\overline{K}_{k+1} & I_{m_{k+1}} & 0 & \cdots \\
-A_{k+2}F_{k+2}\overline{K}_{k+1} & -A_{k+2}\overline{K}_{k+2} & I_{m_{k+2}} & \cdots \\
-A_{k+3}F_{k+3}F_{k+2}\overline{K}_{k+1} & -A_{k+3}F_{k+3}\overline{K}_{k+2} & -A_{k+3}\overline{K}_{k+3} & \cdots \\
\vdots & \vdots & \vdots & \ddots
\end{pmatrix} ,$$

with

$$\begin{aligned}
F_i &= \Phi_{i,i-1}(I - K_{i-1}A_{i-1}) \\
\overline{K}_i &= \Phi_{i,i-1}K_{i-1} \quad .
\end{aligned}$$

Applying the estimation procedure from Table 2.2 to the model with condition equations (2.42) and using the linear relationship between the misclosure vector and the predicted residuals gives for the least squares estimator of the corrections:

$$\begin{aligned}\hat{\underline{e}} &= Q_y B(B^{\mathrm{T}}Q_y B)^{-1}B^{\mathrm{T}}\underline{y} \\ &= Q_y BL^{-\mathrm{T}}(L^{-1}B^{\mathrm{T}}Q_y BL^{-\mathrm{T}})^{-1}L^{-1}L\underline{v} \\ &= Q_y BL^{-\mathrm{T}}(L^{-1}B^{\mathrm{T}}Q_y BL^{-\mathrm{T}})^{-1}\underline{v}\ . \end{aligned} \tag{2.46}$$

The problem of inverting the matrix $B^{\mathrm{T}}Q_y B$ has been replaced by the problem of inverting the matrix $L^{-1}B^{\mathrm{T}}Q_y BL^{-\mathrm{T}}$. The form of this matrix does not immediately suggest that a major simplification is now possible. However, using the fact that $E\{\underline{t}\underline{t}^{\mathrm{T}}\} = B^{\mathrm{T}}Q_y B$ and the linear relation (2.44) between the predicted residuals and misclosure vector, it follows that

$$\begin{aligned}E\{L\underline{v}\underline{v}^{\mathrm{T}}L^{\mathrm{T}}\} &= B^{\mathrm{T}}Q_y B \\ E\{\underline{v}\underline{v}^{\mathrm{T}}\} &= L^{-1}B^{\mathrm{T}}Q_y BL^{-\mathrm{T}}\ , \end{aligned} \tag{2.47}$$

where the matrix $L^{-1}B^{\mathrm{T}}Q_y BL^{-\mathrm{T}}$ is the covariance matrix of the predicted residuals. It is well-known (for a proof see Appendix A) that

$$E\{\underline{v}_k\underline{v}_l^{\mathrm{T}}\} = 0\ ,\quad k \neq l\ , \tag{2.48}$$

and consequently we arrive at the important result that

$$L^{-1}B^{\mathrm{T}}Q_y BL^{-\mathrm{T}} = \mathrm{diag}(Q_{v_k}, Q_{v_{k+1}}, \ldots)\ . \tag{2.49}$$

This matrix can easily be inverted and is a function of covariance matrices of the predicted residuals only. We now have available a least squares description for the prediction, filtering, and smoothing problem, which renders analytical solutions feasible.

To demonstrate the usefulness of the batch solution in terms of a model with condition equations, we consider the following example. Say, for example, we want to derive the estimator of the state at time $k-1$ at time l ($\hat{\underline{x}}_{k-1|l}$), with $l = k, k+1, \ldots$, using all observables in the time interval $[k, l]$. The least squares estimator of the smoothed state at time l is obtained by:

$$\hat{\underline{x}}_{k-1|l} = \hat{\underline{x}}_{k-1|k-1} - \hat{\underline{e}}_{x_{k-1|k-1}}\ .$$

For the estimator of the least squares corrections (cf. eq. 2.46) we need the first row of the matrix $Q_y BL^{-\mathrm{T}}$. From (2.42), (2.45), and (2.46) it follows by simple substitution that

$$\hat{\underline{x}}_{k-1|l} = \hat{\underline{x}}_{k-1|k-1} + P_{k-1|k-1}\Phi_{k,k-1}^{\mathrm{T}}\sum_{i=k}^{l}\left[\prod_{j=k+1}^{i}F_j^{\mathrm{T}}\right]A_i^{\mathrm{T}}Q_{v_i}^{-1}\underline{v}_i\ . \tag{2.50}$$

Using the definitions of the measurement update (2.5) and the gain matrix (2.7), it immediately follows that

$$A_i^{\mathrm{T}}Q_{v_i}^{-1}\underline{v}_i = P_{i|i-1}^{-1}(\hat{\underline{x}}_{i|i} - \hat{\underline{x}}_{i|i-1})\ ,$$

and inserting this result in (2.50) gives for $l = k$

$$\hat{\underline{x}}_{k-1|k} = \hat{\underline{x}}_{k-1|k-1} + P_{k-1|k-1}\Phi_{k,k-1}^{\mathrm{T}}P_{k|k-1}^{-1}(\hat{\underline{x}}_{k|k} - \hat{\underline{x}}_{k|k-1}) ,$$

which is identical to the one-step smoothed solutions found before. If we now assume that (2.50) holds for $l = m - 1$, and use the identity $(I - K_iA_i) = P_{i|i}P_{i|i-1}^{-1}$, it follows from

$$\hat{\underline{x}}_{k-1|m} = \hat{\underline{x}}_{k-1|m-1} + P_{k-1|k-1}\Phi_{k,k-1}^{\mathrm{T}}\left[\prod_{j=k+1}^{m} F_j^{\mathrm{T}}\right] A_m^{\mathrm{T}}Q_{v_m}^{-1}\underline{v}_m$$

that

$$\begin{aligned}\hat{\underline{x}}_{k-1|m} \;=\; & \hat{\underline{x}}_{k-1|m-1} + \\ & P_{k-1|k-1}\Phi_{k,k-1}^{\mathrm{T}}\left[\prod_{j=k+1}^{m}(\Phi_{j,j-1}P_{j-1|j-1}P_{j-1,j-2}^{-1})^{\mathrm{T}}\right] P_{m|m-1}^{-1}(\hat{\underline{x}}_{m|m} - \hat{\underline{x}}_{m|m-1})\end{aligned}$$

and thus (2.50) can be written as

$$\hat{\underline{x}}_{k-1|m} = \hat{\underline{x}}_{k-1|m-1} + \left[\prod_{j=k}^{m} P_{j-1|j-1}\Phi_{j,j-1}^{\mathrm{T}}P_{j,j-1}^{-1}\right](\hat{\underline{x}}_{m|m} - \hat{\underline{x}}_{m|m-1}) . \qquad (2.51)$$

Equation (2.51) corresponds with the so-called fixed-point smoother [MEDITCH, 1969].

2.6 Alternative Noise Models

In the previous sections we assumed that the measurements and disturbances were mutually uncorrelated and uncorrelated in time. In practice it is likely that correlation is present. In this section we will consider the following types of correlation:

A. Correlation between the disturbances (system noise) over a sample period and the measurements at the end of that interval

$$E\{(\underline{d}_k - E\{\underline{d}_k\})(\underline{y}_j - E\{\underline{y}_j\})^{\mathrm{T}}\} = \begin{cases} S_k & j = k \\ 0 & j \neq k \end{cases} .$$

B. Correlation between the measurements and the disturbances (system noise) over the ensuing sample period

$$E\{(\underline{y}_j - E\{\underline{y}_j\})(\underline{d}_{k+1} - E\{\underline{d}_{k+1}\})^{\mathrm{T}}\} = \begin{cases} S_k & j = k \\ 0 & j \neq k \end{cases} .$$

C. Correlation between the measurement noise at successive epochs.

D. Correlation between the disturbances at successive epochs.

This list is not exhaustive, but in general it will be quite difficult to specify more detailed models for various types of correlation. Correlation between noise sources at successive epochs is also called coloured noise or sequentially time correlated noise.

In practice cases C and D will prevail. Due to the approximation of the 'real world' by a dynamic model some correlation between disturbances at successive epochs will always be present. Also time correlated measurements (measurement noise) occur quite frequently in practice. This type of correlation is often due to the internal data processing in the measurement systems (e.g. receivers of radiopositioning systems) or mechanical damping of the measurement devices (e.g. conventional gyros). Cases C and D are often specified as exponentially time correlated noise because other, more sophisticated, noise models are difficult to derive. Cases A and B cover the correlation between disturbances and measurements and are formulated in a somewhat restricted manner. It is likely that disturbances will have some impact on the measurements (consider, for example, a ship subject to pitch, roll, and heave that is equipped with a satellite antenna installed in top of a (sweeping) mast) and besides this type of correlation will probably not be limited to a single epoch. Tractable, general solutions for models which incorporate correlations between disturbances and measurements for longer time spans are, however, not available. Closed form solutions can be found for cases A and B and therefore these cases are considered.

The objective of this section is to demonstrate that the least squares approach leads to solutions for some of these cases directly. Solutions can be derived in a straightforward manner for cases A and B. Time correlated measurements and disturbances (cases C and D) can be tackled by orthogonalization methods (which constitute differencing schemes between correlated observables to obtain derived observables which are uncorrelated) or state augmentation procedures, in which the correlated noise is modelled in state-space. Both approaches are demonstrated and their relation is shown for time correlated measurements. The results obtained in this section are compared with the solutions found in the literature.

2.6.1 Correlation between Measurements and Disturbances — Case A

If one considers correlation between the disturbances over the sample period and the measurements at the end of that interval, the model with observation equations reads:

$$E\{\begin{pmatrix} \underline{\hat{x}}_{k-1|k-1} \\ \underline{d}_k \\ \underline{y}_k \end{pmatrix}\} = \begin{pmatrix} I & 0 \\ -\Phi_{k,k-1} & I \\ 0 & A_k \end{pmatrix} \begin{pmatrix} x_{k-1} \\ x_k \end{pmatrix} ; \tag{2.52}$$

$$\begin{pmatrix} P_{k-1|k-1} & & \\ & Q_k & S_k \\ & S_k^T & R_k \end{pmatrix} ,$$

where S_k is a $n \times m_k$ matrix. This model is equivalent to the following model with condition equations (cf. Section 2.5):

$$\left(\begin{array}{cc} -A_k & I \end{array} \right) E\{ \left(\begin{array}{c} \hat{\underline{x}}_{k|k-1} \\ \underline{y}_k \end{array} \right) \} = 0 \; ; \; \left(\begin{array}{cc} P_{k|k-1} & S_k \\ S_k^{\mathrm{T}} & R_k \end{array} \right) , \tag{2.53}$$

from which the least squares estimators follow directly as

$$\left(\begin{array}{c} \hat{\underline{x}}_{k|k} \\ \hat{\underline{y}}_k \end{array} \right) = \left(\begin{array}{c} \hat{\underline{x}}_{k|k-1} \\ \hat{\underline{y}}_k \end{array} \right) + \left(\begin{array}{c} P_{k|k-1} A_k^{\mathrm{T}} - S_k \\ S_k^{\mathrm{T}} A_k^{\mathrm{T}} - R_k \end{array} \right) Q_{\overline{v}}^{-1} (\underline{y}_k - A_k \hat{\underline{x}}_{k|k-1}) , \tag{2.54}$$

where

$$Q_{\overline{v}} = (A_k P_{k|k-1} A_k^{\mathrm{T}} + R_k - A_k S_k - S_k^{\mathrm{T}} A_k^{\mathrm{T}}) \; .$$

If no correlation is present this form reduces to the standard Kalman filter solution. The variance of the least squares estimators follows from applying the error propagation law to (2.54):

$$P_{k|k} = P_{k|k-1} - (P_{k|k-1} A_k^{\mathrm{T}} - S_k) Q_{\overline{v}}^{-1} (A_k P_{k|k-1} - S_k^{\mathrm{T}}) \; . \tag{2.55}$$

The time update equations remain unchanged.

2.6.2 Correlation between Measurements and Disturbances — Case B

In case one considers correlation between the measurements and the disturbances over the ensuing sample period, the model with observation equations can be written as:

$$E\{ \left(\begin{array}{c} \hat{\underline{x}}_{k|k-1} \\ \underline{y}_k \\ \underline{d}_{k+1} \end{array} \right) \} = \left(\begin{array}{cc} I & 0 \\ A_k & 0 \\ -\Phi_{k+1,k} & I \end{array} \right) \left(\begin{array}{c} x_k \\ x_{k+1} \end{array} \right) \; ; \tag{2.56}$$

$$\left(\begin{array}{ccc} P_{k|k-1} & & \\ & R_k & S_k \\ & S_k^{\mathrm{T}} & Q_{k+1} \end{array} \right) ,$$

where S_k is a $m_k \times n$ matrix. By orthogonalization of $\underline{d}_{k+1}$ with respect to $\underline{y}_k$ one obtains a model equivalent to (2.56), but with uncorrelated observations:

$$E\{ \left(\begin{array}{c} \hat{\underline{x}}_{k|k-1} \\ \underline{y}_k \\ \underline{d}_{k+1} - S_k^{\mathrm{T}} R_k^{-1} \underline{y}_k \end{array} \right) \} = \left(\begin{array}{cc} I & 0 \\ A_k & 0 \\ (-\Phi_{k+1,k} - S_k^{\mathrm{T}} R_k^{-1} A_k) & I \end{array} \right) \left(\begin{array}{c} x_k \\ x_{k+1} \end{array} \right) \; ; \tag{2.57}$$

$$\left(\begin{array}{ccc} P_{k|k-1} & & \\ & R_k & 0 \\ & 0 & Q_{k+1} - S_k^{\mathrm{T}} R_k^{-1} S_k \end{array} \right) .$$

From (2.57) it can be seen that the estimator of the filtered state at time k remains unchanged, so that one obtains:

$$E\{\begin{pmatrix} \underline{\hat{x}}_{k|k} \\ \underline{d}_{k+1} - S_k^T R_k^{-1} \underline{y}_k \end{pmatrix}\} = \begin{pmatrix} I & 0 \\ (-\Phi_{k+1,k} - S_k^T R_k^{-1} A_k) & I \end{pmatrix} \begin{pmatrix} x_k \\ x_{k+1} \end{pmatrix} ; \quad (2.58)$$

$$\begin{pmatrix} P_{k|k} & 0 \\ 0 & Q_{k+1} - S_k^T R_k^{-1} S_k \end{pmatrix} .$$

Equation (2.58) can be solved immediately as there is no redundancy in the model. The estimator of the predicted state at time $k+1$ reads

$$\begin{aligned} \underline{\hat{x}}_{k+1|k} &= (\Phi_{k+1,k} + S_k^T R_k^{-1} A_k)\underline{\hat{x}}_{k|k} + \underline{d}_{k+1} - S_k^T R_k^{-1} \underline{y}_k \\ &= \Phi_{k+1,k}\underline{\hat{x}}_{k|k} + \underline{d}_{k+1} - S_k^T R_k^{-1}(\underline{y}_k - A_k \underline{\hat{x}}_{k|k}) \\ &= \Phi_{k+1,k}\underline{\hat{x}}_{k|k} + \underline{d}_{k+1} - S_k^T R_k^{-1}(\underline{y}_k - A_k \underline{\hat{x}}_{k|k-1} - A_k K_k(\underline{y}_k - A_k \underline{\hat{x}}_{k|k-1})) \\ &= \Phi_{k+1,k}\underline{\hat{x}}_{k|k} + \underline{d}_{k+1} - S_k^T R_k^{-1}(I - A_k K_k)(\underline{y}_k - A_k \underline{\hat{x}}_{k|k-1}) \\ &= \Phi_{k+1,k}\underline{\hat{x}}_{k|k} + \underline{d}_{k+1} - S_k^T R_k^{-1}[R_k(A_k P_{k|k-1} A_k^T + R_k)^{-1}(\underline{y}_k - A_k \underline{\hat{x}}_{k|k-1})] \\ &= \Phi_{k+1,k}\underline{\hat{x}}_{k|k} + \underline{d}_{k+1} - S_k^T (A_k P_{k|k-1} A_k^T + R_k)^{-1}(\underline{y}_k - A_k \underline{\hat{x}}_{k|k-1}) . \quad (2.59) \end{aligned}$$

For the estimate of the predicted state the sample of $\underline{d}_{k+1}$ is taken equal to zero. The covariance matrix of the predicted state is obtained by applying the error propagation law to (2.59)

$$\begin{aligned} P_{k+1|k} &= \Phi_{k+1,k} P_{k|k} \Phi_{k+1,k}^T + Q_{k+1} \\ &\quad - S_k^T (A_k P_{k|k-1} A_k^T + R_k)^{-1} S_k + \Phi_{k+1,k} K_k S_k + S_k^T K_k^T \Phi_{k+1,k}^T . \quad (2.60) \end{aligned}$$

If no correlation is present this form reduces to the standard Kalman filter (prediction) solution.

Comparison with Solutions found in the Literature

The reader can compare the results we have found for cases A and B with the solutions given in literature (e.g. JAZWINSKI [1970]). The solutions given here differ slightly in the parts containing the correlation terms S_k and S_k^T (note the sign changes in comparison with [ibid., pp.209-212]). This difference is due to our definition of the disturbance $\underline{d}_k$. In our model description the matrices S model the correlation between the observables $\underline{y}$ and $\underline{d}$, where $E\{\underline{d}_k\}$ is defined as $x_k - \Phi_{k,k-1} x_{k-1}$. In most textbooks on Kalman filtering the matrices S model the correlation between quantities $\underline{e}$ and $\underline{w}$, where $\underline{w}$ is the so-called system noise (cf. Section 2.2). If one inserts (2.1) in our definition of the disturbance vector $\underline{d}_k$ one obtains $\underline{d}_k = \underline{x}_k - \Phi_{k,k-1}\underline{x}_{k-1} - \underline{w}_k$, where the disturbance $\underline{d}_k$ and the system noise vector $\underline{w}_k$ have opposite signs. This explains the sign changes in our formulas (2.54), (2.55), (2.59), and (2.60) as compared with [ibid.].

2.6.3 Time Correlated Measurements

Time correlated measurement noise can be handled using the concepts of orthogonalization or state augmentation. We start our discussion on time-correlated measurement noise by considering a vector of observables $\underline{y} = (\underline{y}_k^{\mathrm{T}}, \underline{y}_{k+1}^{\mathrm{T}}, \ldots)^{\mathrm{T}}$ of which the covariance matrix is specified as:

$$D\{\underline{y}\} =$$

$$\begin{pmatrix} R_k & R_k\Psi_{k+1}^{\mathrm{T}} & R_k\Psi_{k+1}^{\mathrm{T}}\Psi_{k+2}^{\mathrm{T}} & \cdots \\ \Psi_{k+1}R_k & (\Psi_{k+1}R_k\Psi_{k+1}^{\mathrm{T}} + N_{k+1}) & (\Psi_{k+1}R_k\Psi_{k+1}^{\mathrm{T}} + N_{k+1})\Psi_{k+2}^{\mathrm{T}} & \cdots \\ \Psi_{k+2}\Psi_{k+1}R_k & \Psi_{k+2}(\Psi_{k+1}R_k\Psi_{k+1}^{\mathrm{T}} + N_{k+1}) & \Psi_{k+2}(\Psi_{k+1}R_k\Psi_{k+1}^{\mathrm{T}} + \\ & & N_{k+1})\Psi_{k+2}^{\mathrm{T}} + N_{k+2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{2.61}$$

where the matrices Ψ_i and N_i (which remain unspecified for the moment) model the correlation of the measurements and a contribution of white noise at time i respectively. If one premultiplies the vector $\underline{y}$ by the square and full-rank matrix

$$P = \begin{pmatrix} I & 0 & 0 & \cdots \\ -\Psi_{k+1} & I & 0 & \cdots \\ 0 & -\Psi_{k+2} & I & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{2.62}$$

one can easily verify that the derived observables $P\underline{y}$ are uncorrelated, that is the elements of $\underline{y}$ can be *orthogonalized* by means of the matrix P. In principle this orthogonalization procedure can be used to derive a filter algorithm for time correlated measurements. It is self evident, however, that the algorithms derived using the orthogonalization approach will be rather involved compared with the 'classical' Kalman filter algorithm. The structure of the orthogonalization matrix (2.62) inevitably results in algorithms in which the estimator of the state at time k also depends on the measurements at time $k+1$, and thus recursiveness is lost. We demonstrate this by an example. If we assume, for example, that the first observation is processed at time k, the model with observation equations for four epochs can be written as

$$E\{\begin{pmatrix} \hat{\underline{x}}_{k-1|k-1} \\ \underline{d}_k \\ \underline{y}_k \\ \underline{d}_{k+1} \\ \underline{y}_{k+1} \\ \underline{d}_{k+2} \\ \underline{y}_{k+2} \end{pmatrix}\} = \begin{pmatrix} I & 0 & 0 & 0 \\ -\Phi_{k,k-1} & I & 0 & 0 \\ 0 & A_k & 0 & 0 \\ 0 & -\Phi_{k+1,k} & I & 0 \\ 0 & 0 & A_{k+1} & 0 \\ 0 & 0 & -\Phi_{k+2,k+1} & I \\ 0 & 0 & 0 & A_{k+2} \end{pmatrix} \begin{pmatrix} x_{k-1} \\ x_k \\ x_{k+1} \\ x_{k+2} \end{pmatrix}; \tag{2.63}$$

$$\begin{pmatrix} P_{k-1|k-1} & & & & & & \\ & Q_k & & & & & \\ & & R_k & 0 & R_k\Psi_{k+1}^{\mathrm{T}} & 0 & R_k\Psi_{k+1}^{\mathrm{T}}\Psi_{k+2}^{\mathrm{T}} \\ & & 0 & Q_{k+1} & 0 & 0 & 0 \\ & & \Psi_{k+1}R_k & 0 & (\Psi_{k+1}R_k\Psi_{k+1}^{\mathrm{T}}+ N_{k+1}) & 0 & (\Psi_{k+1}R_k\Psi_{k+1}^{\mathrm{T}}+ N_{k+1})\Psi_{k+2}^{\mathrm{T}} \\ & & 0 & 0 & Q_{k+1} & 0 & 0 \\ & & \Psi_{k+2}\Psi_{k+1}R_k & 0 & \Psi_{k+2}\times (\Psi_{k+1}R_k\Psi_{k+1}^{\mathrm{T}}+ N_{k+1}) & 0 & \Psi_{k+2}(\Psi_{k+1}R_k\Psi_{k+1}^{\mathrm{T}}+ N_{k+1})\Psi_{k+2}^{\mathrm{T}}+N_{k+2} \end{pmatrix}.$$

Specifying the matrix P as

$$P = \begin{pmatrix} I & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & -\Psi_{k+1} & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & -\Psi_{k+2} & 0 & I \end{pmatrix}, \tag{2.64}$$

a derived model with uncorrelated observables can be obtained from model (2.63) as follows:

$$PE\{\underline{y}\} = PA\boldsymbol{x} \;\; ; \;\; D\{P\underline{y}\} = PQ_yP^{\mathrm{T}}, \tag{2.65}$$

where

$$PE\{\underline{y}\} = E\{\begin{pmatrix} \underline{\hat{x}}_{k-1|k-1} \\ \underline{d}_k \\ \underline{y}_k \\ \underline{d}_{k+1} \\ \underline{y}_{k+1} - \Psi_{k+1}\underline{y}_k \\ \underline{d}_{k+2} \\ \underline{y}_{k+2} - \Psi_{k+2}\underline{y}_{k+1} \end{pmatrix}\}; \; PA = \begin{pmatrix} I & 0 & 0 & 0 \\ -\Phi_{k,k-1} & I & 0 & 0 \\ 0 & A_k & 0 & 0 \\ 0 & -\Phi_{k+1,k} & I & 0 \\ 0 & -\Psi_{k+1}A_k & A_{k+1} & 0 \\ 0 & 0 & -\Phi_{k+2,k+1} & I \\ 0 & 0 & -\Psi_{k+2}A_{k+1} & A_{k+2} \end{pmatrix}$$

$$PQ_yP^{\mathrm{T}} = \mathrm{diag}(P_{k-1|k-1}, Q_k, R_k, Q_{k+1}, N_{k+1}, Q_{k+2}, N_{k+2}) \; .$$

To establish a link with the literature we will now consider a special case of model (2.65) covering three epochs (viz. $k-1, k$ and $k+1$). Since all observables are uncorrelated, the estimator of the filtered state at time k can be obtained in the usual way, and hence model (2.65) can be reduced to

$$E\{\begin{pmatrix} \underline{\hat{x}}_{k|k} \\ \underline{d}_{k+1} \\ \underline{y}_{k+1} - \Psi_{k+1}\underline{y}_k \end{pmatrix}\} = \begin{pmatrix} I & 0 \\ -\Phi_{k+1,k} & I \\ -\Psi_{k+1}A_k & A_{k+1} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_k \\ \boldsymbol{x}_{k+1} \end{pmatrix}; \begin{pmatrix} P_{k|k} & & \\ & Q_{k+1} & \\ & & N_{k+1} \end{pmatrix}. \tag{2.66}$$

It follows from (2.66) that one obtains an estimator of the *smoothed* state at time k. We first eliminate the parameters $\boldsymbol{x}_k$ and $\boldsymbol{x}_{k+1}$ to arrive at an equivalent model with

condition equations, namely

$$\left(\begin{array}{ccc} (\Psi_{k+1}A_k - A_{k+1}\Phi_{k+1,k}) & -A_{k+1} & I \end{array} \right) E\{ \left(\begin{array}{c} \hat{\underline{x}}_{k|k} \\ \underline{d}_{k+1} \\ \underline{y}_{k+1} - \Psi_{k+1}\underline{y}_k \end{array} \right) \} = 0 \ . \qquad (2.67)$$

The estimator of the state at time k can be found by straightforward application of the least squares estimation procedure for models with condition equations. Denoting $(\underline{y}_{k+1} - \Psi_{k+1}\underline{y}_k)$ as $\overline{\underline{y}}_{k+1}$ and $(\Psi_{k+1}A_k - A_{k+1}\Phi_{k+1,k})$ as H_{k+1} one finds for the estimator of the state at time k:

$$\begin{aligned} \hat{\underline{x}}_{k|k+1} &= \hat{\underline{x}}_{k|k} - P_{k|k}H_{k+1}^{\mathrm{T}}(H_{k+1}P_{k|k}H_{k+1}^{\mathrm{T}} + A_{k+1}Q_{k+1}A_{k+1}^{\mathrm{T}} + N_{k+1})^{-1} \times \\ &\quad (\overline{\underline{y}}_{k+1} + H_{k+1}\hat{\underline{x}}_{k|k} - A_{k+1}\underline{d}_{k+1}) \ . \end{aligned} \qquad (2.68)$$

If one considers $A_{k+1}Q_{k+1}A_{k+1}^{\mathrm{T}} + N_{k+1}$ as a single new covariance matrix and recalls that the samples of $\underline{d}_i$ are chosen equal to zero, this form is very similar (but *not* identical) to the 'classical' Kalman filter measurement update. From (2.67) one can also derive the estimator of the disturbances at time $k+1$ (to obtain an actual estimate d_{k+1} is set to zero)

$$\begin{aligned} \hat{\underline{d}}_{k+1} &= \underline{d}_{k+1} + Q_{k+1}A_{k+1}^{\mathrm{T}}(H_{k+1}P_{k|k}H_{k+1}^{\mathrm{T}} + A_{k+1}Q_{k+1}A_{k+1}^{\mathrm{T}} + N_{k+1})^{-1} \times \\ &\quad (\overline{\underline{y}}_{k+1} + H_{k+1}\hat{\underline{x}}_{k|k} - A_{k+1}\underline{d}_{k+1}) \ . \end{aligned} \qquad (2.69)$$

The least squares estimator of the filtered state at time $k+1$ is finally obtained as

$$\hat{\underline{x}}_{k+1|k+1} = \Phi_{k+1,k}\hat{\underline{x}}_{k|k+1} + \hat{\underline{d}}_{k+1} \ . \qquad (2.70)$$

Equations (2.68) to (2.70) can be compared with the solutions given by BRYSON AND HENRIKSON [1968]. Except for some minor differences in notation (our matrix H_{k+1} corresponds to $-H_k^r$ in [ibid.]), the main difference of our results and those in [ibid.] is that we arrive at our algorithm in a straightforward manner using the principle of least squares. In comparing our results with [ibid.] it can be seen that BRYSON AND HENRIKSON explicitely interpret the matrix Ψ_{k+1} from the outset as a state transition matrix, whereas in our derivation the matrix Ψ_{k+1} remains largely unspecified (Ψ_{k+1} only has to be chosen in such a way that the matrix P in (2.64) is of full rank). Furthermore we can avoid the somewhat artificial discussion on the interpretation of the estimators $\hat{\underline{x}}_{k|k+1}$ and $\hat{\underline{x}}_{k+1|k+1}$, as in our approach they can simply be interpreted as least squares estimators of the smoothed state at time k and the filtered state at time $k+1$ respectively.

A second approach to deal with time correlated measurement noise is given by the method of *state augmentation*. In order to follow this approach one has to assume that the time correlated measurement noise can be modelled using a state-space approach. As a consequence the time correlated measurement noise is modelled using a *fixed* number of additional states. Including these additional states in the functional model

then automatically leads to state augmentation. Instead of using the model given by (2.61) the time correlated measurement noise is now described as:

$$E\{\underline{d}_k^e\} = e_k - \Psi_{k,k-1} e_{k-1} \quad ; \quad D\{\underline{d}_k^e\} = N_k \quad , \tag{2.71}$$

with

$$E\{\underline{y}_1\} = A_1 x_1 \quad (\text{or } \underline{y}_1 = A_1 x_1 + \underline{e}_1) \quad ; \quad D\{\underline{y}_1\} = R_1 \quad ,$$

where $\Psi_{k,k-1}$ is a $m \times m$ state transition matrix of the measurement noise. (For the sake of simplicity we assume that all m observables $\underline{y}$ are time-correlated; in practice the number of time-correlated observables may lie in the range 1 to m, but this does not affect the following derivations.) Analogous to the reasoning in Section 2.2 the vector $\underline{d}_k^e$ is considered as an observable, random disturbance vector of which the sample values d_k^e are chosen equal to zero. Using equation (2.71) one can for $k = 3, 4, \ldots$ define the augmented system model as:

$$E\{\underline{\bar{d}}_k\} = \bar{x}_k - \bar{\Phi}_{k,k-1} \bar{x}_{k-1} \quad ; \quad y_k = \bar{A}_k \bar{x}_k \quad , \tag{2.72}$$

where

$$\begin{aligned}
\underline{\bar{d}}_k &= (\underline{d}_k^{\mathrm{T}} \ \underline{d}_k^{e\,\mathrm{T}})^{\mathrm{T}} && (n+m) \times 1 \\
\bar{x}_k &= (x_k^{\mathrm{T}} \ e_k^{\mathrm{T}})^{\mathrm{T}} && (n+m) \times 1 \\
\bar{\Phi}_{k,k-1} &= \begin{pmatrix} \Phi_{k,k-1} & 0 \\ 0 & \Psi_{k,k-1} \end{pmatrix} && (n+m) \times (n+m) \\
\bar{A}_k &= \begin{pmatrix} A_k & I \end{pmatrix} && m \times (n+m) \\
D\{\underline{\bar{d}}_k\} &= \begin{pmatrix} Q_k & 0 \\ 0 & N_k \end{pmatrix} && (n+m) \times (n+m)
\end{aligned}$$

with starting values

$$E\left\{\begin{pmatrix} \underline{x}_0 \\ \underline{d}_1 \\ \underline{y}_1 \\ \underline{d}_2 \\ \underline{d}_2^e + \Psi_{2,1} y_1 \\ y_2 \end{pmatrix}\right\} = \begin{pmatrix} I & 0 & 0 & 0 \\ -\Phi_{1,0} & I & 0 & 0 \\ 0 & A_1 & 0 & 0 \\ 0 & -\Phi_{2,1} & I & 0 \\ 0 & \Psi_{2,1} A_1 & 0 & I \\ 0 & 0 & A_2 & I \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ e_2 \end{pmatrix} \quad ;$$

$$\begin{pmatrix} P_0 & & & & & \\ & Q_1 & & & & \\ & & R_1 & & & \\ & & & Q_2 & & \\ & & & & N_2 & \\ & & & & & 0 \end{pmatrix} .$$

This is a very straightforward procedure which can be implemented using standard Kalman filter software, apart from the fact that the measurement model contains noise-free observations. Therefore care has to exercised, as this approach might lead to numerical ill-conditioning of the covariance matrix of the (augmented) predicted state, thus leading to difficulties in the computation of the covariance matrix of the (augmented) filtered state. Due to the noise free measurement model the rank of the covariance matrix of the estimator of the filtered augmented state (with dimension $n + m$) is merely n.

Comparison of Orthogonalization and Augmentation

It remains to be shown at this point that the orthogonalization and the state augmentation approach lead to identical results for the estimators of the system state. If we assume that the first observation is processed at time k and we consider four epochs, the augmented state model can also be formulated as:

$$E\{\underline{y}\} = Ax \;;\; B^{\mathrm{T}}x = c \;;\; D\{\underline{y}\} = Q_y \;, \tag{2.73}$$

where

$$\underline{y} = (\underline{\hat{x}}_{k-1|k-1}^{\mathrm{T}}, \underline{d}_k^{\mathrm{T}}, \underline{y}_k^{\mathrm{T}}, \underline{d}_{k+1}^{\mathrm{T}}, (\underline{d}_{k+1}^{e} + \Psi_{k+1,k}y_k)^{\mathrm{T}}, \underline{d}_{k+2}^{\mathrm{T}}, \underline{d}_{k+2}^{e\,\mathrm{T}})^{\mathrm{T}}$$

$$A = \begin{pmatrix} I & 0 & 0 & 0 & 0 & 0 \\ -\Phi_{k,k-1} & I & 0 & 0 & 0 & 0 \\ 0 & A_k & 0 & 0 & 0 & 0 \\ 0 & -\Phi_{k+1,k} & I & 0 & 0 & 0 \\ 0 & \Psi_{k+1,k}A_k & 0 & I & 0 & 0 \\ 0 & 0 & -\Phi_{k+2,k+1} & 0 & I & 0 \\ 0 & 0 & 0 & -\Psi_{k+2,k+1} & 0 & I \end{pmatrix}$$

$$B^{\mathrm{T}} = \begin{pmatrix} 0 & 0 & A_{k+1} & I & 0 & 0 \\ 0 & 0 & 0 & 0 & A_{k+2} & I \end{pmatrix}$$

$$x = (x_{k-1}^{\mathrm{T}}, x_k^{\mathrm{T}}, x_{k+1}^{\mathrm{T}}, e_{k+1}^{\mathrm{T}} x_{k+2}^{\mathrm{T}}, e_{k+2}^{\mathrm{T}})^{\mathrm{T}}$$

$$c = (y_{k+1}^{\mathrm{T}}, y_{k+2}^{\mathrm{T}})^{\mathrm{T}}$$

$$Q_y = \mathrm{diag}(P_{k-1|k-1}, Q_k, R_k, Q_{k+1}, N_{k+1}, Q_{k+2}, N_{k+2}).$$

Equation (2.73) is given in a so-called mixed model representation, which is in the form of observation equations with conditions on the parameter vector x. We use the condition equations to eliminate the state variables e_{k+1} and e_{k+2} from the state vector x. Noting that $e_{k+1} = y_{k+1} - A_{k+1}x_{k+1}$ and $e_{k+2} = y_{k+2} - A_{k+2}x_{k+2}$, model (2.73)

can also be written as

$$E\{\begin{pmatrix} \underline{\hat{x}}_{k-1|k-1} \\ \underline{d}_k \\ \underline{y}_k \\ \underline{d}_{k+1} \\ \underline{d}^e_{k+1} - y_{k+1} + \Psi_{k+1,k} y_k \\ \underline{d}_{k+2} \\ \underline{d}^e_{k+2} - y_{k+2} + \Psi_{k+2,k+1} y_{k+1} \end{pmatrix}\} =$$

$$\begin{pmatrix} I & 0 & 0 & 0 \\ -\Phi_{k,k-1} & I & 0 & 0 \\ 0 & A_k & 0 & 0 \\ 0 & -\Phi_{k+1,k} & I & 0 \\ 0 & \Psi_{k+1,k} A_k & -A_{k+1} & 0 \\ 0 & 0 & -\Phi_{k+2,k+1} & I \\ 0 & 0 & \Psi_{k+2,k+1} A_{k+1} & -A_{k+2} \end{pmatrix} \begin{pmatrix} x_{k-1} \\ x_k \\ x_{k+1} \\ x_{k+2} \end{pmatrix}, \qquad (2.74)$$

with Q_y as in (2.73). From (2.71) it follows that $\underline{d}^e_i - y_i + \Psi_{i,i-1} y_{i-1}$ is identical to $-(\underline{y}_i - \Psi_{i,i-1} \underline{y}_{i-1})$ and thus to $-\underline{\overline{y}}_i$. If furthermore the matrices Ψ_i in (2.61) correspond to the state transition matrices $\Psi_{i,i-1}$ in (2.71), the equivalence of the state augmentation approach (resulting in (2.74)) and the orthogonalization approach (2.65) is established.

The main difference between the solutions based on orthogonalization and state augmentation is that for the solution based on orthogonalization the matrices Ψ_i not necessarily have to be state transition matrices. In the orthogonalization approach the condition imposed on the matrices Ψ_i is that the transformation matrix P in model (2.64) is of full rank. A difference in the actual application of the orthogonalization and state augmentation algorithms is that with the former approach one obtains estimators $\underline{\hat{x}}_{k|k+1}$ and $\underline{\hat{x}}_{k+1|k+1}$, whereas with the latter approach, due to the use of the standard Kalman filter algorithm, one only obtains the estimator $\underline{\hat{x}}_{k+1|k+1}$.

Summarizing, two strategies can be followed in case of time correlated measurements that result in identical estimators of the filtered state. The orthogonalization approach is based on the differencing of correlated observables. The differencing scheme leads to a somewhat involved filter algorithm, which automatically makes available a one-step smoothed estimate of the state. The dimension of the state vector remains unchanged. The state augmentation approach is based on the fact that one assumes that the time correlated measurement noise can be modelled in state-space. Because exponentially correlated noise can very easily be modelled in state-space, the state augmentation approach is tailored for this type of noise. A drawback is that the dimension of the state vector is increased to $n + m$, but an advantage is that the standard Kalman filter algorithm can be used.

2.6.4 Time Correlated Disturbances

If a complete description of the dynamic behaviour can be given for a certain system, the description of the dynamic model by a state transition only disturbed by white noise is probably adequate. In practice, however, it is rather unlikely that such a model is completely correct, because

- A model description that takes all disturbances into account can seldom be given (such a model requires complete knowledge of the system dynamics).
- Even if the system can be adequately modelled, a reduced system model is sometimes preferred for computational reasons.

Imperfect modelling is often compensated by increasing the variances of the disturbances leading to so-called sub-optimal filter designs. In practice this procedure often works quite well, but its application actually requires a careful sensitivity analysis (see, e.g., GELB [1974]). It has to be kept in mind, however, that time correlated disturbances are generally due to imperfect modelling of the system dynamics and hence the most obvious approach is to enhance the system model.

If the disturbances are time correlated, the system model can be given as (if one assumes the first observation is processed at time k):

$$E\{\begin{pmatrix} \underline{\hat{x}}_{k-1|k-1} \\ \underline{d}_k \\ \underline{y}_k \\ \underline{d}_{k+1} \\ \vdots \end{pmatrix}\} = \begin{pmatrix} I & 0 & 0 & \cdots \\ -\Phi_{k,k-1} & I & 0 & \cdots \\ 0 & A_k & 0 & \cdots \\ 0 & -\Phi_{k+1,k} & I & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} x_{k-1} \\ x_k \\ x_{k+1} \\ \vdots \end{pmatrix} ; \tag{2.75}$$

$$\begin{pmatrix} P_{k-1|k-1} & & & & \\ & Q_k & 0 & Q_k\Psi_{k+1}^{\mathrm{T}} & \\ & 0 & R_k & 0 & \\ & \Psi_{k+1}Q_k & 0 & \Psi_{k+1}Q_k\Psi_{k+1}^{\mathrm{T}} + N_{k+1} & \\ & & & & \ddots \end{pmatrix} .$$

In the case of time correlated measurements we considered two solution methods, namely orthogonalization and state augmentation. In this section we proceed along similar lines and start with the orthogonalization approach.

Specifying an orthogonalization matrix in an analogous manner to (2.64) and considering three epochs we arrive at a derived model of the form

$$PE\{\underline{y}\} = PAx \;\; ; \;\; D\{P\underline{y}\} = PQ_yP^{\mathrm{T}}, \tag{2.76}$$

where

$$PE\{\underline{y}\} = E\{\begin{pmatrix} \underline{\hat{x}}_{k-1|k-1} \\ \underline{d}_k \\ \underline{y}_k \\ \underline{d}_{k+1} - \Psi_{k+1}\underline{d}_k \end{pmatrix}\} \; ; \; PA = \begin{pmatrix} I & 0 & 0 \\ -\Phi_{k,k-1} & I & 0 \\ 0 & A_k & 0 \\ \Psi_{k+1}\Phi_{k,k-1} & -(\Phi_{k+1,k} + \Psi_{k+1}) & I \end{pmatrix} ;$$

$$PQ_yP^{\mathrm{T}} = \mathrm{diag}(P_{k-1|k-1}, Q_k, R_k, N_{k+1}) \ .$$

Since all observables are uncorrelated we can first compute the estimators of the states $\boldsymbol{x}_{k-1}$ and $\boldsymbol{x}_k$ taking into account the observables $\underline{\hat{\boldsymbol{x}}}_{k-1|k-1}$, $\underline{d}_k$, and $\underline{y}_k$. This boils down to the computation of the estimator of the filtered state at time k and the estimator of the of the one-step smoothed state at time $k-1$ (cf. Sections 2.4 and 2.5). Inserting these estimators into (2.76) results in the model

$$E\{\begin{pmatrix} \underline{\hat{\boldsymbol{x}}}_{k-1|k} \\ \underline{\hat{\boldsymbol{x}}}_{k|k} \\ \underline{d}_{k+1} - \Psi_{k+1}\underline{d}_k \end{pmatrix}\} = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ \Psi_{k+1}\Phi_{k,k-1} & -(\Phi_{k+1,k} + \Psi_{k+1}) & I \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_{k-1} \\ \boldsymbol{x}_k \\ \boldsymbol{x}_{k+1} \end{pmatrix} ; \tag{2.77}$$

$$\begin{pmatrix} P_{k-1|k} & J_{k-1}P_{k|k} & 0 \\ P_{k|k}J_{k-1}^{\mathrm{T}} & P_{k|k} & 0 \\ 0 & 0 & N_{k+1} \end{pmatrix} ,$$

where $J_{k-1} = P_{k-1|k-1}\Phi_{k,k-1}^{\mathrm{T}}P_{k|k-1}^{-1}$. Model (2.77) is not overdetermined and hence the estimator of the predicted state at time $k+1$ follows directly as:

$$\underline{\hat{\boldsymbol{x}}}_{k+1|k} = -\Psi_{k+1}\Phi_{k,k-1}\underline{\hat{\boldsymbol{x}}}_{k-1|k} + (\Phi_{k+1,k} + \Psi_{k+1})\underline{\hat{\boldsymbol{x}}}_{k|k} + \underline{d}_{k+1} - \Psi_{k+1}\underline{d}_k \ . \tag{2.78}$$

The covariance matrix of the estimator can be obtained by application of the error propagation law. Equation (2.78) reduces to the standard Kalman filter prediction solution if $\Psi_{k+1} = 0$.

State augmentation provides a second method to deal with time correlated disturbances. The description of time correlated disturbances in state-space results in a fixed number of additional states (namely n). Using state augmentation one tries to model time correlated disturbances as state variables of a fictitious linear dynamic system which is itself excited by white noise. We assume that the time correlated disturbances for $k = 3, 4, \ldots$ can be modelled as

$$E\{\begin{pmatrix} \underline{d}_k^d \\ \underline{d}_k^d \end{pmatrix}\} = \begin{pmatrix} d_k \\ \boldsymbol{x}_k \end{pmatrix} - \begin{pmatrix} \Psi_{k,k-1} & 0 \\ \Psi_{k,k-1} & \Phi_{k,k-1} \end{pmatrix} \begin{pmatrix} d_{k-1} \\ \boldsymbol{x}_{k-1} \end{pmatrix} ; \ D\{\begin{pmatrix} \underline{d}_k^d \\ \underline{d}_k^d \end{pmatrix}\} = \begin{pmatrix} N_k & N_k \\ N_k & N_k \end{pmatrix} \tag{2.79}$$

with

$$E\{\underline{d}_1\} = \boldsymbol{x}_1 - \Phi_{1,0}\boldsymbol{x}_0 \ ; \ D\{\underline{d}_1\} = Q_1 \ .$$

Model (2.79) is the discrete-time equivalent of the models provided in the literature for the continuous time case with time-correlated disturbances (cf., e.g., [MAYBECK, 1979, Ch.4]). The covariance matrix of the 'extended' vector of disturbances is singular. The vector $\underline{d}_k^d$ is considered as an observable, random disturbance vector of which the sample values d_k^d are taken equal to zero. Using equation (2.79) one can for $k = 3, 4, \ldots$ define the augmented system model as:

$$E\{\underline{\bar{d}}_k\} = \bar{\boldsymbol{x}}_k - \bar{\Phi}_{k,k-1}\bar{\boldsymbol{x}}_{k-1} \quad ; \quad E\{\underline{y}_k\} = \bar{A}_k\bar{\boldsymbol{x}}_k \ , \tag{2.80}$$

where

$$\begin{aligned}
\overline{\underline{d}}_k &= (\underline{d}_k^{d\,\mathrm{T}}\ \underline{d}_k^{d\,\mathrm{T}})^{\mathrm{T}} && 2n \times 1 \\
\overline{\boldsymbol{x}}_k &= (\boldsymbol{d}_k^{\mathrm{T}}\ \boldsymbol{x}_k^{\mathrm{T}})^{\mathrm{T}} && 2n \times 1 \\
\overline{\Phi}_{k,k-1} &= \begin{pmatrix} \Psi_{k,k-1} & 0 \\ \Psi_{k,k-1} & \Phi_{k,k-1} \end{pmatrix} && 2n \times 2n \\
\overline{A}_k &= \begin{pmatrix} 0 & A_k \end{pmatrix} && m_k \times 2n \\
D\{\underline{d}_k\} &= \begin{pmatrix} N_k & N_k \\ N_k & N_k \end{pmatrix} && 2n \times 2n \\
D\{\underline{y}_k\} &= R_k && m_k \times m_k
\end{aligned}$$

with starting values

$$E\{\begin{pmatrix} \underline{\boldsymbol{x}}_0 \\ \underline{d}_1 \\ \underline{y}_1 \\ \underline{d}_2^{\mathrm{d}} \\ \underline{d}_2^{d} \\ \underline{y}_2 \end{pmatrix}\} = \begin{pmatrix} I & 0 & 0 & 0 \\ -\Phi_{1,0} & I & 0 & 0 \\ 0 & A_1 & 0 & 0 \\ \Psi_{2,1}\Phi_{1,0} & -\Psi_{2,1} & I & 0 \\ \Psi_{2,1}\Phi_{1,0} & -(\Psi_{2,1}+\Phi_{2,1}) & 0 & I \\ 0 & 0 & 0 & A_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_0 \\ \boldsymbol{x}_1 \\ d_2 \\ \boldsymbol{x}_2 \end{pmatrix} \quad ;$$

$$\begin{pmatrix} P_0 & & & & & \\ & Q_1 & & & & \\ & & R_1 & & & \\ & & & N_2 & N_2 & \\ & & & N_2 & N_2 & \\ & & & & & R_2 \end{pmatrix} .$$

Estimators of the predicted and filtered augmented state can be obtained by means of the standard Kalman filter algorithm. The equivalence between the orthogonalization and state augmentation approaches can be proven along similar lines as was done for time-correlated measurement noise in the previous section. The proof is trivial once one sees that model (2.79) can be transformed into the following model:

$$E\{\begin{pmatrix} \underline{d}_k^d \\ d_k \end{pmatrix}\} = \begin{pmatrix} d_k \\ \boldsymbol{x}_k \end{pmatrix} - \begin{pmatrix} \Psi_{k,k-1} & 0 \\ 0 & \Phi_{k,k-1} \end{pmatrix} \begin{pmatrix} d_{k-1} \\ \boldsymbol{x}_{k-1} \end{pmatrix} \; ; \; D\{\begin{pmatrix} \underline{d}_k^d \\ d_k \end{pmatrix}\} = \begin{pmatrix} N_k & 0 \\ 0 & 0 \end{pmatrix} . \tag{2.81}$$

If one deletes the parameter d_k from the augmented state model (which can be done using the condition $d_k = \boldsymbol{x}_k - \Phi_{k,k-1}\boldsymbol{x}_{k-1}$ in (2.81)) one finally obtains a model identical to (2.76).

2.6.5 Summary

In this section we considered four possible types of correlation and their impact on the filter solutions from a least squares point of view. We showed that for the cases with mutually correlated measurements and disturbances results could be obtained in a straightforward manner. The cases of time correlated measurements and disturbances could be tackled by means of state augmentation and orthogonalization. For the case of time correlated measurements we explicitely demonstrated that the orthogonalization and augmentation approach lead to identical results. State augmentation results in a larger dimension of the system state, but facilitates the use of standard Kalman filter software as it leaves the model underlying the Kalman filter basically unchanged. Orthogonalization corresponds to the differencing of successive (correlated) observables. Using such differencing schemes, however, requires alternative filter algorithms, which have been derived in this section.

2.7 Model Nonlinearities

The Kalman filter and the equivalent least squares solutions are all based on *linear* models. In the well-known formulation of KALMAN [1960] this implies that both the dynamic model and the measurement model are linear. In the least squares approach this means that the observation equations or the condition equations are linear. In practice we often have to deal with *nonlinear* functional relations, for example:

$$E\{\underline{d}_k\} = \boldsymbol{x}_k - \boldsymbol{\phi}_{k,k-1}(\boldsymbol{x}_{k-1})$$

and

$$E\{\underline{y}_k\} = \boldsymbol{A}_k(\boldsymbol{x}_k) \quad ,$$

where it is assumed that the noise enters in an additive fashion. For many applications the dynamic model is approximated by a linearized model. If one assumes that the state trajectory is rather smooth, one can, by choosing a suitably small update interval, approximate a nonlinear dynamic model by a linear model. Most geodetic measurement models, on the contrary, are nonlinear. It is therefore important to have methods to assess the amount of nonlinearity in nonlinear models. Besides it is useful if one can proof that a linearized model is a sufficient approximation (in that case we can use the estimation procedures for the linearized observation and condition equations).

In this section we first briefly consider the impact of nonlinearities on least squares estimators and discuss the Gauss-Newton iteration scheme. Then nonlinearities in the Kalman filter are discussed starting with nonlinearities in the measurement model. Also the case of a nonlinear dynamic model is briefly discussed. Finally, the case with a combined nonlinear measurement and dynamic model is discussed. The section is concluded with some remarks on the practical aspects of dealing with nonlinearities in the filter models and a brief summary.

2.7.1 Nonlinearity and Least Squares

For models with nonlinear observation equations of the form $E\{\underline{y}\} = A(\boldsymbol{x})$, one can make a distinction between two types of nonlinearity [TEUNISSEN, 1989a]. Firstly one has the nonlinearity related to the chosen parametrization, that is the nonlinearity which is *not* invariant under a change of variables in $A(\boldsymbol{x})$. If one considers, for example, the tracking of a survey vessel with a tachymeter, and the position of the vessel is expressed in polar coordinates relative to the tachymeter, the measurement model is linear. A position expressed in rectangular coordinates will inevitably result in a nonlinear measurement model. Secondly, nonlinearity can be due to intrinsic properties of the map $A(\boldsymbol{x})$, that is the n-dimensional manifold (described by the map $A(\boldsymbol{x})$ for various $\boldsymbol{x}$) embedded in the dataspace $\mathcal{R}^m$ is nonlinear.

In most practical cases it is impossible to find exact nonlinear equations for nonlinear least squares problems. There exist, however, useful approximations that describe the biases (due to nonlinearity) in the least squares estimators. These biases for models based on observation equations and condition equations are given in [TEUNISSEN, 1989a] and [TEUNISSEN AND KNICKMEYER, 1988] respectively. The importance of these measures, which can be analysed in the design phase, is that one can assess the amount of non-linearity in the least-squares estimators. TEUNISSEN AND KNICKMEYER [1988] also provide an approximation of the covariance matrix of the (non-linear) least squares estimator.

2.7.2 The Gauss-Newton Iteration Scheme

For practical applications we must now define a processing strategy. If the bias due to nonlinearities is significant, one has to be careful in applying the linearized least squares approach. Because the models are usually approximated by a first order Taylor expansion, *iterations* are necessary to obtain less biased estimates. Iterations for least squares problems for models specified in terms of observation equations are generally performed using the Gauss-Newton iteration scheme

$$\begin{aligned} \hat{\boldsymbol{x}}_{i+1} &= \hat{\boldsymbol{x}}_i + \\ &\quad (\partial_x A(\hat{\boldsymbol{x}}_i)^{\mathrm{T}} Q_y^{-1} \partial_x A(\hat{\boldsymbol{x}}_i))^{-1} \partial_x A(\hat{\boldsymbol{x}}_i)^{\mathrm{T}} Q_y^{-1} (y - A(\hat{\boldsymbol{x}}_i)) \\ &\quad \text{for } i = 1, 2, \ldots , \end{aligned} \tag{2.82}$$

where i denotes the iteration step and $\partial_x A(\hat{\boldsymbol{x}}_i)$ is the Jacobian of $A(.)$ evaluated at $\hat{\boldsymbol{x}}_i$. For a discussion on other iteration schemes see [TEUNISSEN, 1989b]. To test whether the iteration should be continued or not one needs a termination criterion. A criterion for the Gauss-Newton method, which is invariant to a change of variables, is

$$\|\hat{\boldsymbol{x}}_{i+1} - \hat{\boldsymbol{x}}_i\|_{Q_x} < \epsilon \,, \tag{2.83}$$

with $\|\hat{\boldsymbol{x}}_{i+1} - \hat{\boldsymbol{x}}_i\|_{Q_x} = \sqrt{(\hat{\boldsymbol{x}}_{i+1} - \hat{\boldsymbol{x}}_i)^{\mathrm{T}} Q_x^{-1} (\hat{\boldsymbol{x}}_{i+1} - \hat{\boldsymbol{x}}_i)}$ and where ϵ is a preset tolerance level.

Some additional remarks concerning nonlinearities in least squares problems can be made. Firstly, TEUNISSEN [1985; 1989b] has given a geometric interpretation of nonlinear least squares problems. Secondly, a large drawback of iterative schemes is that the computational load of the algorithms cannot be predicted. Therefore it might be necessary to specify a maximum number of iterations as an additional termination criterion for the iterations. Thirdly, one can approximate nonlinear models by higher order Taylor expansions, but then the estimation procedures have to be adapted. A disadvantage of higher order approximations is, in general, the large computational burden associated with them.

2.7.3 Nonlinear Models in Kalman Filtering

The aspects of nonlinearity discussed so far pertain to least squares problems in general. In this section nonlinearity is discussed for filtering problems. Although in the strict sense Kalman filters are by definition based on linear models, the term Kalman filtering will also be maintained in the case of nonlinear models. We will first consider the case of a nonlinear measurement model. In geodetic practice a nonlinear measurement model is the rule rather than the exception. Then we consider a possible nonlinear dynamic model and finally the situation where both measurement and dynamic model are nonlinear is considered. It will be shown that the methods we discussed for nonlinear least squares problems in general, lead to solutions for the nonlinear filtering problems directly.

Nonlinear Measurement Model

We consider the following model with non-linear observation equations (cf. eq. 2.26):

$$E\{\begin{pmatrix} \underline{\hat{x}}_{k|k-1} \\ \underline{y}_k \end{pmatrix}\} = \begin{pmatrix} x_k \\ A_k(x_k) \end{pmatrix} \; ; \; \begin{pmatrix} P_{k|k-1} & \\ & R_k \end{pmatrix} , \tag{2.84}$$

which is of the form $E\{y\} = A(x)$. If the estimator of the predicted state is based on measurements prior to time k, the matrix $P_{k|k-1}$ is an approximation (due to the non-linearity) of the actual covariance matrix of the estimator of the predicted state. Estimates of the system state can be obtained by the straightforward application of the Gauss-Newton iteration scheme (2.82):

$$\begin{aligned} \hat{x}_{k|k,i+1} &= \hat{x}_{k|k,i} + \\ &\quad P_{k|k,i}[P^{-1}_{k|k-1}(\hat{x}_{k|k-1} - \hat{x}_{k|k,i}) + \partial_x A_k(\hat{x}_{k|k,i})^{\mathrm{T}} R_k^{-1}(y_k - A_k(\hat{x}_{k|k,i})] \end{aligned}$$

for $i = 1, 2, \ldots$, where $P_{k|k,i} = (P^{-1}_{k|k-1} + \partial_x A_k(\hat{x}_{k|k,i})^{\mathrm{T}} R_k^{-1} \partial_x A_k(\hat{x}_{k|k,i}))^{-1}$ and with $\hat{x}_{k|k,0} = \hat{x}_{k|k-1}$. After some rearrangements one finds that

$$\hat{x}_{k|k,i+1} = \hat{x}_{k|k-1} + K_{k,i+1}(y_k - A_k(\hat{x}_{k|k,i}) - \partial_x A_k(\hat{x}_{k|k,i})(\hat{x}_{k|k-1} - \hat{x}_{k|k,i})) , \tag{2.85}$$

where $K_{k,i+1} = P_{k|k-1}\partial_x A_k(\hat{x}_{k|k,i})^T(R_k + \partial_x A_k(\hat{x}_{k|k,i})P_{k|k-1}\partial_x A_k(\hat{x}_{k|k,i})^T)^{-1}$. The covariance matrix of the estimator of the filtered state at the $(i+1)$th iteration step is approximated by

$$P_{k|k,i+1} = (I - K_{k,i+1}\partial_x A_k(\hat{x}_{k|k,i}))P_{k|k-1} \cdot \tag{2.86}$$

The time update equations of the Kalman filter remain unchanged. Note that, except for the update of the filtered state (2.85), the formulas of the measurement update of the linear filter can be used with the design matrix A_k replaced by the Jacobian $\partial_x A_k(.)$. The recursive scheme (2.85) is also obtained if one starts from the model with condition equations (cf. eq. 2.36)

$$(-I \;\; I)E\{\begin{pmatrix} A_k(\underline{\hat{x}}_{k|k-1}) \\ \underline{y}_k \end{pmatrix}\} = 0 \; . \tag{2.87}$$

If no iterations are performed ($i = 0$ in eq. 2.85) and $\hat{x}_{k|k,0} = \hat{x}_{k|k-1}$, the result is called an *extended* Kalman filter (EKF). If no iterations are performed and $\hat{x}_{k|k,0} = \overline{x}_k$, with $\overline{x}_k$ some externally provided (approximate) state trajectory, the result is called a *linearized* Kalman filter (LKF). If iterations are performed ($i = 0, 1, 2, \ldots$ in eq. 2.85) and $\hat{x}_{k|k,0} = \hat{x}_{k|k-1}$ the result is called an *iterated extended* Kalman filter (IEKF).

In general it is difficult to predict the number of iteration steps that is necessary to comply with the convergence test (2.83). As the IEKF follows from the Gauss-Newton method it also has the (local) convergency characteristics of the Gauss-Newton method. TEUNISSEN [1991] has shown that the IEKF has a local linear rate of convergence and derives an upperbound of the linear convergency factor for the IEKF. This upperbound can be evaluated in the design phase of the filter and allows the prediction of the number of required iteration steps.

Nonlinear Dynamic Model

For many applications the dynamic model is approximated by a linear model. This is possible if the state trajectory is smooth enough and a suitably small update interval is chosen. In practice the system noise is often used as a useful, although artificial, method of accounting for (among other things) neglected nonlinearities. In general the dynamic model can be derived from some continuous time model, which models the dynamic behaviour of the system under consideration. Dynamic models often are nonlinear, as forces often act in a nonlinear manner on systems. In this section we assume that an equivalent (nonlinear) representation in discrete-time is available, that is the dynamic model is characterized by nonlinear difference equations. Although the discretization of the system dynamics might be troublesome due to the nonlinearities, this approach is followed in order to maintain our description of the Kalman filter in discrete-time. For some applications the use of so-called continuous-discrete filters, where the dynamic model is given in continuous time (i.e. by differential equations) and measurements are available at discrete-time instants, is to be preferred, for instance when the force model

governing the system is available, but for these type of filters the reader is referred to the literature. We thus assume that the dynamic model can be described as:

$$E\{\underline{d}_k\} = \boldsymbol{x}_k - \phi_{k,k-1}(\boldsymbol{x}_{k-1}). \tag{2.88}$$

From (2.88) the following estimator of the predicted state can be derived

$$\hat{\underline{\boldsymbol{x}}}_{k|k-1} = \phi_{k,k-1}(\hat{\underline{\boldsymbol{x}}}_{k-1|k-1}) + \underline{d}_k \quad . \tag{2.89}$$

This estimator is not unbiased because $E\{\phi_{k,k-1}(\hat{\underline{\boldsymbol{x}}}_{k-1|k-1})\} \neq \phi_{k,k-1}(E\{\hat{\underline{\boldsymbol{x}}}_{k-1|k-1}\})$. An approximation of the covariance matrix of the estimator of the predicted state is given by:

$$P_{k|k-1} = \partial_x \phi_{k,k-1}(\hat{\boldsymbol{x}}_{k-1|k-1}) P_{k-1|k-1} \partial_x \phi_{k,k-1}(\hat{\boldsymbol{x}}_{k-1|k-1})^{\mathrm{T}} + Q_k \ , \tag{2.90}$$

where $\partial_x \phi_{k,k-1}(\hat{\boldsymbol{x}}_{k-1|k-1})$ is the Jacobian of the n-dimensional function $\phi(\boldsymbol{x})$ evaluated at $\hat{\boldsymbol{x}}_{k-1|k-1}$. The solution given above is quite simple, because we assumed the noise enters in an additive fashion. If the nonlinear dynamic model is expanded to, say,

$$E\{\underline{d}_k\} = \phi_{k,k-1}(\boldsymbol{x}_k, \boldsymbol{x}_{k-1}) \ ,$$

also the nonlinear propagation of the noise has to be considered.

Combined Nonlinear Dynamic and Measurement Model

We considered the cases of nonlinear measurement and dynamic models separately. In practice it might occur that both models are nonlinear, that is

$$E\{\begin{pmatrix} \hat{\underline{\boldsymbol{x}}}_{k-1|k-1} \\ \underline{d}_k \\ \underline{\boldsymbol{y}}_k \end{pmatrix}\} = \begin{pmatrix} \boldsymbol{x}_{k-1} \\ \boldsymbol{x}_k - \phi_{k,k-1}(\boldsymbol{x}_{k-1}) \\ A_k(\boldsymbol{x}_k) \end{pmatrix} ; \begin{pmatrix} P_{k-1|k-1} & & \\ & Q_k & \\ & & R_k \end{pmatrix} , \tag{2.91}$$

which is of the form $E\{\underline{y}\} = F(\boldsymbol{x})$. For this model we can partly proceed in the same way as we did for the linear case, namely by executing the time and measurement updates separately. The estimator of the predicted state and its covariance matrix are obtained by application of (2.89) and (2.90) respectively. Based on the estimator of the predicted state one can then apply the Gauss-Newton iteration scheme to model (2.84). An alternative approach is to apply to Gauss-Newton iteration scheme to model (2.91) directly. One then not only improves the nonlinear measurement update (as with the IEKF), but one also improves the time update. (One thereby obtains an estimator of the smoothed state at time $k-1$.) The application of the Gauss-Newton iteration scheme to (2.91) leads to an algorithm identical to the so-called *Iterated Linear Filter-Smoother* [JAZWINSKI, 1970]. The iteration scheme can be given as (the derivation will not be given here):

$$\begin{aligned}
\hat{\boldsymbol{x}}_{k|k-1,i} &= \phi(\hat{\boldsymbol{x}}_{k-1|k,i-1}) + \\
&\quad \partial_x \phi_{k,k-1}(\hat{\boldsymbol{x}}_{k-1|k,i-1})(\hat{\boldsymbol{x}}_{k-1|k-1} - \hat{\boldsymbol{x}}_{k-1|k,i-1}) \qquad (2.92) \\
\hat{\boldsymbol{x}}_{k|k,i} &= \hat{\boldsymbol{x}}_{k|k-1,i} + K_{k,i}[\boldsymbol{y}_k - A_k(\hat{\boldsymbol{x}}_{k|k,i-1}) - \\
&\quad \partial_x A_k(\hat{\boldsymbol{x}}_{k|k-1,i})(\hat{\boldsymbol{x}}_{k|k-1,i} - \hat{\boldsymbol{x}}_{k|k,i-1})] \qquad (2.93) \\
\hat{\boldsymbol{x}}_{k-1|k,i} &= \hat{\boldsymbol{x}}_{k-1|k-1} + J_{k-1,i}(\hat{\boldsymbol{x}}_{k|k,i} - \hat{\boldsymbol{x}}_{k|k-1,i}) \qquad (2.94) \\
&\quad \text{for } i = 1, 2, \ldots
\end{aligned}$$

with

$$\begin{aligned}
K_{k,i} &= P_{k|k-1,i}\partial_x A_k(\hat{\boldsymbol{x}}_{k|k-1,i})^{\mathrm{T}}(R_k + \partial_x A_k(\hat{\boldsymbol{x}}_{k|k-1,i})P_{k|k-1,i}\partial_x A_k(\hat{\boldsymbol{x}}_{k|k-1,i})^{\mathrm{T}})^{-1} \\
P_{k|k-1,i} &= \partial_x \phi_{k,k-1}(\hat{\boldsymbol{x}}_{k-1|k,i-1})P_{k-1|k-1}\partial_x \phi_{k,k-1}(\hat{\boldsymbol{x}}_{k-1|k,i-1})^{\mathrm{T}} + Q_{k-1} \\
J_{k-1,i} &= P_{k-1|k-1}\partial_x \phi_{k,k-1}(\hat{\boldsymbol{x}}_{k-1|k,i-1})P_{k|k-1,i}^{-1} \\
P_{k|k-1,0} &= P_{k|k-1} \\
\hat{\boldsymbol{x}}_{k-1|k,0} &= \hat{\boldsymbol{x}}_{k-1|k-1} \\
\hat{\boldsymbol{x}}_{k|k,0} &= \hat{\boldsymbol{x}}_{k-1|k-1} + \phi(\hat{\boldsymbol{x}}_{k-1|k-1}) \; .
\end{aligned}$$

The covariance matrix of the estimator of the filtered state is given by (2.86) and only has to be computed once the iteration process is terminated. Furthermore the covariance matrix of the smoothed state at time $k-1$ does not have to be computed explicitely. The slight difference with the algorithm given in [ibid.] is due to the fact that we assumed a discretized, nonlinear dynamic model. Note the resemblance of (2.93) with (2.85), and of (2.94) with the smoothing formula given in Section 2.4.

2.7.4 Practical Considerations

Working with nonlinear models also requires the assessment of the amount of nonlinearity. It is, however, very difficult to give general guidelines to judge to what extent linearized models are adequate for Kalman filter applications. TEUNISSEN [1989a] shows that the amount of bias in the estimators is a function of the precision of the observables (as given by $\sigma^2 Q_y$), and the parametrization and nonlinearity of the maps $A(.)$, $B^{\mathrm{T}}(.)$, or $F(.)$. In [ibid.] measures for diagnosing nonlinearity are given. These measures can be computed in the design phase of a filter. The biases of the estimators can often be decreased by the choice of a suitable parametrization, high measurement precision, and low system noise.

KREBS [1980] gives an extensive overview of nonlinear filtering, also covering, among other things, higher order filters. The decision which order of approximation is to be used or which other solution may be more appropriate can usually only be determined by extensive simulations. The same holds for the decision whether iterated or non-iterated strategies are to be preferred. Examples of simulations are given in [KREBS, 1980] and [JAZWINSKI, 1970]. In the case the nonlinearity in the measurement model is predominant, these simulations indicate that the iterated extended Kalman filter (IEKF) is usually the best solution. Good approximate values are available by means

of the estimates of the predicted state. Combined with the (local) linear rate of convergence of the IEKF, the number of iterations will be small.

2.7.5 Summary

The purpose of this section was to show how nonlinearities in the (discrete-time) Kalman filter could be dealt with. The Gauss-Newton iteration scheme is used to obtain actual least squares estimates. The impact of nonlinearities on the iteration schemes for so-called extended Kalman filters was examined more closely for three cases, namely a nonlinear measurement model, a nonlinear dynamic model, and a combined nonlinear dynamic and measurement model. It was shown that the direct application of the Gauss-Newton method resulted in the iterated solution schemes presented in the literature. Also some remarks concerning the evaluation of the biases in the resulting least squares estimators were made. The actual evaluation of biases is, however, to a large extent problem dependent. In this section we limited ourselves to first order approximations of the model nonlinearities. Higher order approximations were not discussed. The decision if such filters are more adequate than the filters presented in this section can only be based on extensive simulations for the problem at hand. For additional information on this topic the reader is referred to the references.

2.8 Concluding Remarks

With the least squares method one has available a powerful tool to deal with prediction, filtering, and smoothing problems. The least squares approach provides us with a unified and simple methodology for filtering problems. Except for the mean and the dispersion of the observables no assumptions concerning their distributional properties have to be made in the derivations. Consequently we can only derive the mean and dispersion of the least squares estimators of the predicted, filtered, and smoothed state. The somewhat limiting assumption of normally distributed observables does not have to be made.

The derivations presented in this chapter are valid for measurement and dynamic models formulated in discrete time. In practice most observations will be available at discrete instants, thereby leading to measurement models formulated in discrete time. The dynamic model is, however, sometimes formulated in continuous time. To apply the algorithms derived in this chapter the dynamic model then has to be discretized, which can be easily achieved for many applications in surveying. Furthermore we did not discuss various filter mechanizations. (A filter mechanization is nothing but a specific form of a filter algorithm.) Various mechanizations have been derived for numerical reasons and the optimization of the computational efficiency. Mechanizations are generally defined independently of the way the filter algorithms are derived, although many mechanizations can be directly derived using the least squares approach. For computational aspects of Kalman filtering the reader is referred to BIERMAN [1977] and CHIN [1983].

Especially in the cases of correlation between observables (Section 2.6) and non-linearities in the dynamic and/or measurement models (Section 2.7) we feel that the interpretation of the state estimators as least squares estimators gives a better understanding of the derived results. For these cases the link with least squares is rarely made in the literature. A further major advantage of the least squares approach is that the dual, equivalent formulations of models with observation equations and models with condition equations facilitate the derivation of filter and smoothing results, as one can choose between different least squares estimation procedures (cf. Section 2.3). If one, for example, considers multiple epochs, it is shown in Section 2.5 that an analytical solution of the smoothing problem can be found based on a model with condition equations, whereas a solution based on the model with observation equations is intractable. We do not claim that the derivation is extremely simple, but still the results follow from a straightforward application of the least squares algorithm. In later chapters it will be seen that this form, covering multiple epochs, is very useful.

Chapter 3

A Testing Procedure for Use in Dynamic Systems

3.1 Introduction

In the previous chapter we have shown how the real-time estimation of parameters in dynamic systems can be executed in a least squares framework. Based on the principles of least squares we arrived at the well-known Kalman filter algorithm. The estimation results are optimal if all assumptions underlying the mathematical model hold. Misspecifications in the model will invalidate the results of the estimation procedure and thus also any conclusion based on them. In practice model misspecifications may occur frequently, and hence the estimation procedure should be supported by a model validation technique. Model misspecifications may be due to erroneous observation and dynamic models (i.e. the functional model is not designed properly), errors in the observables (e.g. blunders in sensor outputs), or an incorrectly specified stochastic model (e.g. the covariance matrices of the observables are not correct). In this chapter we will develop an efficient model validation procedure for misspecifications of the functional part of the filter model, and especially errors that can be modelled as additive effects (generally denoted as slips). In addition to misspecifications in the stochastic model, slip-type model errors constitute the most frequently occurring type of model error in geodetic practice. We assume that the structure of the observation and dynamic model is correct. Methods that deal with the estimation of parts of the functional model (usually denoted system identification methods in the literature) are e.g. discussed in [GOODWIN AND PAYNE, 1977]. The analysis of model misspecifications in the stochastic model fits in the framework of the theory of variance component estimation and is often classified as 'adaptive filtering' (see, e.g., [CHIN, 1979]). In the literature the terminology is somewhat ambiguous as the terms system identification and adaptive filtering are sometimes used interchangeably.

Real-time estimation requires a real-time testing procedure, which can detect and isolate model misspecifications and can be used in conjunction with the Kalman filter.

We will focus on tests related to slips in the functional model, i.e. so-called slippage tests. The testing procedure will consist of three steps, namely detection, identification, and adaptation and consequently it is called the DIA procedure. The detection and identification steps were introduced into geodetic testing theory by TEUNISSEN AND SALZMANN [1988;1989] and subsequently extended with the adaptation step by TEUNISSEN [1990a;1990b]. Although we strive for a real-time testing procedure, it will be obvious that certain model errors can only be detected with a certain delay (e.g. 'soft' failures). Therefore we consider the concepts of local and global testing. Local tests only take one epoch of the dynamic system into account and are thus genuinly real-time, whereas global tests take several epochs into account. The DIA procedure can be implemented in recursive form and thereby global tests can efficiently be accomodated in the procedure.

All results are presented within the framework of the theory of hypothesis testing in linear models (see, e.g., [KOCH, 1988]). This approach facilitates the derivation of the testing procedure using generalized likelihood ratio tests and provides a link with the testing procedure for geodetic networks (the so-called B-method) developed by BAARDA [1968]. Furthermore the testing procedure will serve as the basis of the reliability description of dynamic systems in the next chapter. The DIA procedure will be based on the sequence of predicted residuals (innovations), which have well defined statistical properties if the Kalman filter operates at an optimum.

3.1.1 Overview of this Chapter

We begin by reviewing the concept of hypothesis testing in Section 3.2, where we also discuss some aspects of the optimality of the generalized likelihood ratio tests we use in the sequel. We then specialize these findings to the filter model in Section 3.3 where we also discuss the concepts of local and global testing. Detection, Identification, and Adaptation are discussed in Sections 3.4, 3.5, and 3.6 respectively. The feasibility of the B-method of testing for dynamic systems is considered in Section 3.7, and an overview of the DIA procedure is given in Section 3.8. A brief review of and comparison with related testing procedures found in the literature is given in Section 3.9. Finally some concluding remarks are given in Section 3.10.

3.2 Hypothesis Testing

The testing procedure for dynamic systems we will discuss in this chapter is based on the theory of hypothesis testing, which is described in, for instance, [MELSA AND COHN, 1978; KOCH, 1988; CASELLA AND BERGER, 1990]. The objective of hypothesis testing is to decide, based on the actual observations (or, in statistical terms, a sample from the population), which of two complementary hypotheses (generally called the null (or working) hypothesis (H_0) and the alternative hypothesis (H_A)) is true. A hypothesis is a statement about a population parameter and we will limit ourselves to hypotheses related to the *mean* of random variables.

In testing hypotheses one can make two types of errors. Firstly, one can reject the null hypothesis when in fact it is true; this is called an error of the first kind and its probability (α) is often denoted as the false alarm probability. The test is then said to have a level of significance or size α. Secondly, one can accept the null hypothesis when in fact it is false; this is a so-called error of the second kind and its probability is sometimes called the probability of missed detection. In statistics one rather works with the probability that the null hypothesis is rejected correctly with a certain test, and the probability of correctly rejecting H_0 is called the power of the test. A test can be optimized with respect to the level of significance or the power of the test, but unfortunately not simultaneously to both. Our testing procedure will be based on the well-known Neyman-Pearson principle (see, e.g., [MELSA AND COHN, 1978]), which states that for a test with a *fixed* level of significance, one should maximize the power of the test. Actually we try to find the most powerful (MP) test within the class of tests with a certain level of significance.

For simple hypotheses (which means that the distribution parameters are completely specified under H_0 and H_A), it follows from the Neyman-Pearson theorem that the so-called likelihood ratio (LR)-test is a MP test. Although we consider cases based on composite hypotheses, it is worth investigating whether likelihood ratio tests for composite hypotheses (usually denoted generalized likelihood ratio (GLR) tests) are MP-tests within a certain class of tests as well. The generalized likelihood ratio test statistic is defined as (see, e.g., [CASELLA AND BERGER, 1990]):

$$\lambda(y) = \frac{\max\limits_{\theta \in \Theta_0} p_{\underline{y}}(y|\theta)}{\max\limits_{\theta \in \Theta} p_{\underline{y}}(y|\theta)}, \tag{3.1}$$

where $p_{\underline{y}}(y|\theta)$ is the probability density function of the observable $\underline{y}$ given the parameter-(-vector) θ, and where Θ and Θ_0 denote the entire parameter space and the parameter space under H_0 respectively. From (3.1) follows that $\lambda(y)$ is in the closed interval $[0, 1]$. Based on the GLR-test statistic the GLR-test is then defined as:

$$\text{reject } H_0 \text{ if } \lambda(y) < a \text{ ; accept } H_0 \text{ if } \lambda(y) \geq a \text{ ,} \tag{3.2}$$

with $0 < a < 1$. If a GLR-test is most powerful for all $\theta \in \Theta$, then the test is called uniformly most powerful (UMP). In the cited references it is stated that in general the GLR-test is a good test, but it is not necessarily UMP. For the cases we will consider the GLR-test is optimal within a certain class of tests, which will be discussed shortly, and can generally be specified and evaluated in a straightforward manner. Therefore we will restrict ourselves to testing procedures based on GLR-tests.

A consequence of the use of likelihood ratio tests is that we have to specify the distribution of the observables, whereas in the previous chapter we could limit ourselves to the specification of the mean and dispersion of the observables. In the following we will assume that the observables are normally distributed, an assumption which is often

adequate in geodetic practice, and in that case the linear least squares estimators we derived for the predicted, filtered, and smoothed state are also maximum likelihood estimators.

In the following we will establish some properties of the GLR-tests and we will show under which conditions they are optimal. Unfortunately the GLR-tests we consider are not UMP. In practice often no UMP tests of a certain size α exist within the class of all tests for composite hypothesis testing problems. Therefore we restrict the class of tests to be considered by means of the concept of invariance and try to find a UMP size α test for this restricted class. Then for the cases that fulfill the invariance conditions, the test is called a uniformly most powerful invariant (UMPI) size α test. ARNOLD [1981; Ch7 and Ch13] has proven that the GLR-tests used in this report (all of which are tests about the mean) are actually UMPI size α tests. His proof is, however, very technical and requires mathematics beyond the scope of this report. Therefore we will try in the following to make plausible, rather than prove, that the GLR-tests we use are actually UMPI.

3.2.1 On Properties of the Generalized Likelihood Ratio Tests

This section serves to illustrate that the GLR-tests we use in this report are uniformly most powerful invariant (UMPI) tests. We will begin with a very simple example and we will then successively address the concept of invariance, hypotheses related to subsets of the parameter vector, and sufficiency. Finally, we will arrive at the tests familiar from geodetic testing theory. In the following we assume that the size of the tests has been fixed and consequently all tests are tests with size α.

We begin by considering an $m \times 1$ vector of observables $\underline{y}$ that is distributed as follows:

$$\underline{y} \sim N(E\{\underline{y}\}, \sigma^2 I) \,, \tag{3.3}$$

with σ^2 known and

$$H_0 : E\{\underline{y}\} = 0 \;;\; H_A : E\{\underline{y}\} \neq 0 \,. \tag{3.4}$$

For (3.4) no UMP test exists (see, e.g., [CASELLA AND BERGER, 1990; Ch8]), and by restricting the class of tests by the principle of invariance, we try to find a UMP test in this restricted class. We consider three types of invariance, namely distribution invariance, testing procedure invariance and invariance of the hypothesis problem. If we consider, for example, the transformation

$$\underline{y}' = R\underline{y} \,, \tag{3.5}$$

where R is an orthogonal matrix (i.e. $R^{\mathrm{T}}R = I$), then the problem is said to be distribution invariant if the form of the distribution is invariant under the transformation. The transformed observable $\underline{y}'$ is distributed as $N(E\{\underline{y}'\}, \sigma^2 I)$, which is indeed of the same form as (3.3). After the transformation (3.5) the null and alternative hypothesis read $H_0 : E\{\underline{y}'\} = 0$ and $H_A : E\{\underline{y}'\} \neq 0$ and are similar to (3.4), so that the hypothesis problem is also invariant under (3.5). Finally the testing procedure is said to be testing

procedure invariant if for any sample of $\underline{y}$ that leads to rejection of H_0 also the sample of $\underline{y}'$ leads to rejection of H_0 and vice versa. The test must therefore be invariant under orthogonal transformations and consequently the critical region of the test must be (hyper-)spherical with its centre at the origin (we already saw that an orthogonal transformation maintains the distribution and hypothesis problem invariance). Such a critical region can be obtained if we consider test statistics of the form $T(\underline{y}) = (\underline{y}^{\mathrm{T}}\underline{y})/\sigma^2$ and we will consider optimality properties of this test statistic.

If $\underline{y}$ is distributed according to (3.3), then $T(\underline{y}) = \underline{y}^{\mathrm{T}}\underline{y}/\sigma^2$ is distributed as

$$T(\underline{y}) \sim \chi^2(m, \lambda) \text{ with } \lambda = \frac{E\{\underline{y}\}^{\mathrm{T}} E\{\underline{y}\}}{\sigma^2}.$$

In [ARNOLD, 1981] and [CASELLA AND BERGER, 1990] it is stated that if the ratio

$$\frac{p_{T(\underline{y})}(T(y)|\lambda_A)}{p_{T(\underline{y})}(T(y)|\lambda_0)} \tag{3.6}$$

is an increasing function of $T(y)$ for $\lambda_A > \lambda_0$, the test that rejects H_0 if $T(y) > k_\alpha$ (where k_α is the critical value of the test and is a function of the size α) is UMP for testing $H_0 : \lambda = \lambda_0$ versus $H_A : \lambda_A > \lambda_0$. Inserting $T(\underline{y})$ in (3.6) and using the definition of the χ^2-probability density function, it can be seen that the ratio

$$\frac{p_{T(\underline{y})}(T(y)|\lambda_A)}{p_{T(\underline{y})}(T(y)|\lambda_0)} = \exp(\frac{\lambda_A}{2}) \sum_{j=0}^{\infty} \frac{\left(\frac{\lambda_A}{2}\right)^j T(y)^j \Gamma(\frac{m}{2})}{j! 2^j \Gamma(\frac{m}{2} + j)} \tag{3.7}$$

(where $\Gamma(.)$ denotes the gamma-function) is an increasing function of $T(y)$ for all $\lambda_A > 0$. In our particular example (3.4) corresponds to $H_0 : \lambda = 0$ and $H_A : \lambda > 0$, and thus the test

$$\text{reject } H_0 \text{ if } T(y) = \frac{y^{\mathrm{T}}y}{\sigma^2} > k_\alpha \text{ ; accept } H_0 \text{ if } T(y) = \frac{y^{\mathrm{T}}y}{\sigma^2} \leq k_\alpha \text{ ,} \tag{3.8}$$

is indeed a UMP test. Having asserted that (3.8) is a UMP test for the case $H_0 : \lambda = 0$ versus $H_A : \lambda > 0$, this means that (3.8) is a UMPI-test for (3.4), and is invariant under the orthogonal transformation (3.5). We will now show that (3.8) is also a GLR-test. Based on (3.4) and the fact that $\underline{y}$ is normally distributed, the GLR test statistic (3.1) is

$$\lambda(y) = \frac{(2\pi)^{-\frac{m}{2}}(\sigma^2)^{-\frac{m}{2}} \exp(-\frac{1}{2\sigma^2} y^{\mathrm{T}}y)}{(2\pi)^{-\frac{m}{2}}(\sigma^2)^{-\frac{m}{2}}} = \exp(-\frac{1}{2\sigma^2} y^{\mathrm{T}}y). \tag{3.9}$$

The GLR-test (3.2) directly follows from (3.9), and can be written as

$$\text{reject } H_0 \text{ if } \frac{y^{\mathrm{T}}y}{\sigma^2} > \ln a^{-2} \text{ ; accept } H_0 \text{ if } \frac{y^{\mathrm{T}}y}{\sigma^2} \leq \ln a^{-2} \text{ ,} \tag{3.10}$$

and is identical to the UMPI-test (3.8).

We thus have shown that the test we have derived using the invariance restrictions coincides with the GLR-test, and with (3.6) the GLR-test is also shown to be UMPI for (3.4). In the subsequent discussions we will assume that σ^2 is known and is equal to one, and we will now consider some more general cases.

In the testing problems we will consider the specification of the hypotheses is usually related to subsets of the parameter vector. Assume, for example, that the vector $\underline{y}$ can be partitioned in two subsets and is distributed as

$$\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \end{pmatrix} \sim N(\begin{pmatrix} E\{\underline{y}_1\} \\ E\{\underline{y}_2\} \end{pmatrix}, \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}), \tag{3.11}$$

with

$$H_0 : E\{\underline{y}_2\} = 0 \; ; \; H_A : E\{\underline{y}_2\} \neq 0 \; . \tag{3.12}$$

If we consider a transformation of the form $\underline{y}' = R\underline{y} + r$ (where R is an orthogonal matrix and r is an $m \times 1$ vector) then $\underline{y}'$ is distributed as:

$$\begin{pmatrix} \underline{y}'_1 \\ \underline{y}'_2 \end{pmatrix} \sim N(\begin{pmatrix} R_{11}E\{\underline{y}_1\} + R_{12}E\{\underline{y}_2\} + r_1 \\ R_{21}E\{\underline{y}_1\} + R_{22}E\{\underline{y}_2\} + r_2 \end{pmatrix}, \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}). \tag{3.13}$$

If $R_{21} = 0$, $r_2 = 0$, and R_{22} is orthogonal, it follows that the hypothesis problem is invariant, because then (cf. eq. 3.12)

$$H_0 : E\{\underline{y}'_2\} = 0 \; ; \; H_A : E\{\underline{y}'_2\} \neq 0. \tag{3.14}$$

The observables $\underline{y}'_1$ and $\underline{y}'_2$ are uncorrelated and hence $T(\underline{y}'_2) = \underline{y}'^{\mathrm{T}}_2 \underline{y}'_2$ is also a GLR test statistic, because with (3.14) the maximization of the likelihood ratio over all parameters (cf. eq. 3.1), is identical to the maximization over the parameters $E\{\underline{y}'_2\}$. Based on (3.10) the GLR test

$$\text{reject } H_0 \text{ if } \frac{\underline{y}'^{\mathrm{T}}_2 \underline{y}'_2}{\sigma^2} > \ln a^{-2} \; ; \; \text{accept } H_0 \text{ if } \frac{\underline{y}'^{\mathrm{T}}_2 \underline{y}'_2}{\sigma^2} \leq \ln a^{-2} \; , \tag{3.15}$$

is a UMPI test for (3.12), and is invariant under the transformation (3.13) with the conditions given above. Besides it is apparent that the testing procedure is invariant under a wide range of transformations related to $E\{\underline{y}_1\}$.

Next we consider the case where additional parameters $E\{\underline{y}_3\}$ are part of the model

$$\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \end{pmatrix} \sim N(\begin{pmatrix} E\{\underline{y}_1\} \\ E\{\underline{y}_2\} \\ E\{\underline{y}_3\} \end{pmatrix}, \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix}), \tag{3.16}$$

with

$$H_0 : E\{\underline{y}_2\} = 0 \text{ and } E\{\underline{y}_3\} = 0 \; ; \; H_A : E\{\underline{y}_2\} \neq 0 \text{ and } E\{\underline{y}_3\} = 0. \tag{3.17}$$

In this case model (3.16) can be reduced to model (3.11) by sufficiency, because the parameters contained in $E\{\underline{y}_3\}$ have no impact on the testing problem at hand. Consequently the UMPI test for (3.17) is identical to the GLR test (3.15).

Finally we will consider the case generally encountered in geodetic practice, where the m-dimensional vector of observables $\underline{y}$ is distributed as

$$\underline{y} \sim N(E\{\underline{y}\}, Q_y) \,, \tag{3.18}$$

with

$$H_0 : E\{\underline{y}\} \in R(A) \;;\; H_A : E\{\underline{y}\} \in R(A:C) \,, \tag{3.19}$$

and where $R(A) \subset R(A:C) \subset \mathcal{R}^m$ and we assume that $\dim R(A) = n$ and $\dim R(C) = b$ with $n+b \leq m$ ($R(X)$ denotes the range space of the matrix X). Based on the findings given above we will indicate how a UMPI test for (3.19) can be found. The basis vectors of the spaces $R(A)$, $R(C)$, and $R(A:C)^\perp$ span $\mathcal{R}^m$. The range spaces $R(A)$ and $R(C)$ which span $R(A:C)$ are not necessarily orthogonal. By means of the transformation $\overline{C} = P_A^\perp C$, where $P_A^\perp$ is the orthogonal projector onto $R(A)^\perp$, the range space $R(A:C)$ can also be thought to consist of the subsets $R(A)$ and $R(\overline{C})$ with $R(A) \perp R(\overline{C})$. Given these definitions one can now consider a matrix in $\mathcal{R}^m$, say, R for which holds that $R^{\mathrm{T}} Q_y^{-1} R = I$ and that can be partitioned as (R_1, R_2, R_3), where each R_i has the property that $R_i^{\mathrm{T}} Q_y^{-1} R_i = I$. The columns of R_1, R_2, R_3 lie in the spaces $R(A)$, $R(\overline{C})$, and $R(A:\overline{C})^\perp$ respectively. One can then consider the transformation

$$\underline{y}' = R^{\mathrm{T}} Q_y^{-1} \underline{y} \,, \tag{3.20}$$

where the vector $\underline{y}'$ is distributed as (cf. eq. 3.18):

$$\begin{pmatrix} \underline{y}'_1 \\ \underline{y}'_2 \\ \underline{y}'_3 \end{pmatrix} \sim N\left(\begin{pmatrix} E\{\underline{y}'_1\} \\ E\{\underline{y}'_2\} \\ E\{\underline{y}'_3\} \end{pmatrix}, \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} \right) . \tag{3.21}$$

The null and alternative hypotheses are now specified as (cf. eq. 3.19):

$$H_0 : E\{\underline{y}'_2\} = 0 \text{ and } E\{\underline{y}'_3\} = 0 \;;\; H_A : E\{\underline{y}'_2\} \neq 0 \text{ and } E\{\underline{y}'_3\} = 0 \,. \tag{3.22}$$

This testing problem is similar to (3.17) and can be reduced by sufficiency to (3.14), for which the GLR test statistic (of the UMPI test) was shown to be $\underline{T} = \underline{y}_2'^{\mathrm{T}} \underline{y}'_2$. Using (3.20) and the fact that $R(R_i) \perp R(R_j)$ for $i, j = 1, 2, 3$ with $i \neq j$, it follows that

$$\begin{aligned} \underline{T} &= \underline{y}_2'^{\mathrm{T}} \underline{y}'_2 \\ &= \underline{y}^{\mathrm{T}} Q_y^{-1} R_2 R_2^{\mathrm{T}} Q_y^{-1} \underline{y} \,. \end{aligned} \tag{3.23}$$

Taking into account the facts that $R(\overline{C}) = R(R_2)$ and that $R_2^{\mathrm{T}} Q_y^{-1} R_2 = I$, it follows that $R_2 R_2^{\mathrm{T}} Q_y^{-1}$ can be written as $R_2 (R_2^{\mathrm{T}} Q_y^{-1} R_2)^{-1} R_2^{\mathrm{T}} Q_y^{-1}$, which corresponds to the

definition of the projector P_{R_2} (cf. Section 2.3) and thus to the projector $P_{\overline{C}}$. Inserting the definition of $P_{\overline{C}} = \overline{C}(\overline{C}^{\mathrm{T}}Q_y^{-1}\overline{C})^{-1}\overline{C}^{\mathrm{T}}Q_y^{-1}$ with $\overline{C} = P_A^{\perp}C$ in (3.23), one obtains

$$\underline{T} = \underline{y}^{\mathrm{T}}Q_y^{-1}P_A^{\perp}C(C^{\mathrm{T}}P_A^{\perp\mathrm{T}}Q_y^{-1}P_A^{\perp}C)^{-1}C^{\mathrm{T}}P_A^{\perp\mathrm{T}}Q_y^{-1}\underline{y}\,,$$

which, using the results given in Table 2.1, can also be written as

$$\underline{T} = \underline{\hat{e}}^{\mathrm{T}}Q_y^{-1}C(C^{\mathrm{T}}Q_y^{-1}Q_{\hat{e}}Q_y^{-1}C)^{-1}C^{\mathrm{T}}Q_y^{-1}\underline{\hat{e}}\,. \tag{3.24}$$

Equation 3.24 is the well-known test statistic for b-dimensional hypotheses based on the model of observation equations (see, e.g., [Kok, 1984]). Using the results of Section 2.3, the test statistic (3.24) formulated in terms of the equivalent model of condition equations is:

$$\underline{T} = \underline{t}^{\mathrm{T}}Q_t^{-1}B^{\mathrm{T}}C(C^{\mathrm{T}}BQ_t^{-1}B^{\mathrm{T}}C)^{-1}C^{\mathrm{T}}BQ_t^{-1}\underline{t}\,. \tag{3.25}$$

With (3.24) and (3.25) we have available the GLR test statistics of the UMPI test for (3.19). In the next section we will specialize this result to the Kalman filter model.

3.3 A Testing Procedure

The purpose of this section is to introduce the concepts of a testing procedure for use in dynamic systems and to apply the results of the previous section to the Kalman filter model. We start by discussing the concepts of local and global testing. We then discuss the specification of the alternative hypotheses in terms of predicted residuals. As stated in Section 3.1 we are primarily interested in model errors that can be classified as slips. We will see that the Kalman filter provides a linear relationship between the predicted residuals and slip-type errors. The straightforward specification of model errors in terms of the predicted residuals combined with the well defined statistical properties of the predicted residuals (cf. eq. 2.9 and Appendix A) render the predicted residuals extremely suitable for model validation purposes. Moreover the predicted residuals (and their covariance matrices) are readily available from the Kalman filter. The GLR-tests are based on the predicted residuals and correspond to the GLR-tests which were described in the previous section.

3.3.1 Local and Global Model Testing

In the class of slippage tests we make a distinction between *local* model testing and *global* model testing. We speak of local model testing when the tests performed at time k only depend on the predicted state at time k and the observations at time k. If the test takes more than one epoch into account we speak of global model testing. The difference between local and global testing is depicted in Fig. 3.1. From the definition it follows that in contrast with the global tests, the local tests can be executed in real-time. One way to perform global tests is to apply a batch type solution for a batch of collected data. Better results than for local tests can be expected as smoothing is

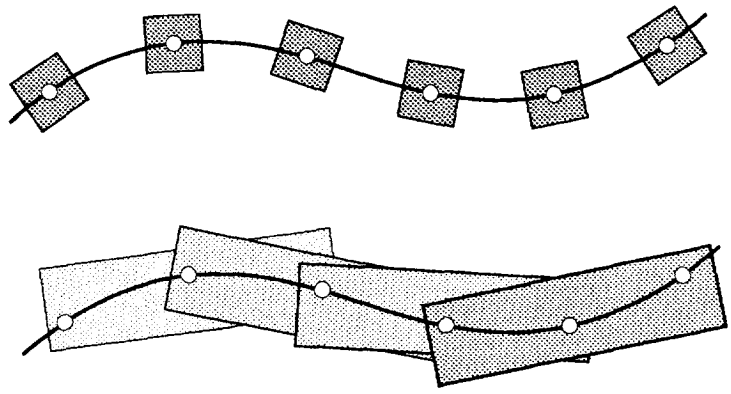


Figure 3.1: Local (top) and Global (bottom) test statistics.

involved. The disadvantage of batch solutions is however, that the test statistics are only available with a delay, and more importantly, the recursiveness, which makes the Kalman filter algorithm so attractive is lost. In the sequel we will show that the global tests can also be executed in recursive form. Besides we think that a small delay in detection and identification is acceptable, because after all it may be more important in practice to detect a possible misspecification with a delay, than not to detect it at all.

3.3.2 Specification of Alternative Hypotheses

If we limit ourselves for the moment to a batch of observables in the time interval from l to k, we will consider model errors of additive nature to the vector of observables $\underline{y}^{\mathrm{T}} = (\underline{\hat{x}}_{l|l-1}^{\mathrm{T}}, \underline{y}_l^{\mathrm{T}}, \underline{d}_{l+1}^{\mathrm{T}}, \underline{y}_{l+1}^{\mathrm{T}}, \ldots, \underline{d}_k^{\mathrm{T}}, \underline{y}_k^{\mathrm{T}})^{\mathrm{T}}$. An outlier in a single observable $\underline{y}_i$ (with $l \leq i \leq k$), for example, can be parametrized as

$$E\{\underline{y}_i\} = A_i x_i + c_y \nabla \ ,$$

where c_y is a $m_i \times 1$ vector and ∇ is an unknown parameter. In general terms a slip type error in the observables in the time interval $[l, \ldots, k]$ can be parametrized as follows:

$$\underset{\sum_{i=l}^{k}(n+m_i)\times 1}{\nabla y} = \underset{\sum_{i=l}^{k}(n+m_i)\times b}{C_y} \ \underset{b\times 1}{\nabla} \ , \tag{3.26}$$

with the full rank matrix C_y known, the b-vector ∇ unknown, and $1 \leq b \leq \sum_{i=l}^{k}(n+m_i)$. Based on (3.26) one can specify the null and alternative hypotheses as (cf. eq. 3.19):

$$\begin{aligned} H_0 : \ \underline{y} &\sim N(Ax, Q_y) \\ H_A : \ \underline{y} &\sim N(Ax + \nabla y, Q_y) \ . \end{aligned} \tag{3.27}$$

Note that we assume that the observables are normally distributed. In the previous section we have seen that the appropriate test statistics for (3.27) are given by (3.24) or (3.25). The latter form is obtained if we assume that the Kalman filter has been derived from a (linear) model of condition equations, viz:

$$B^{\mathrm{T}}E\{\underline{y}\} = 0 \;\; ; \;\; D\{\underline{y}\} = Q_y \; . \tag{3.28}$$

This shows that if the model error ∇y is an element of the null-space of B^{T} (i.e. $\nabla y \in N(B^{\mathrm{T}})$) the testing procedure cannot detect slips in the observables, because then $B^{\mathrm{T}}(E\{\underline{y}\} + \nabla y) = 0$, and therefore we assume in the sequel that ∇y is not an element of $N(B^{\mathrm{T}})$. (This is in accordance with the specification of the alternative hypothesis in (3.19).) Instead of directly using the test statistics (3.24) or (3.25), we will first show how the model error defined by (3.26) can be written in terms of the predicted residuals.

From (3.28) follows that (3.27) can also be written as:

$$\begin{aligned} H_0 : \;\; \underline{t} &= B^{\mathrm{T}}\underline{y} && \sim \;\; N(0, B^{\mathrm{T}}Q_yB) \\ H_A : \;\; \underline{t} &= B^{\mathrm{T}}(\underline{y} + C_y\nabla) && \sim \;\; N(B^{\mathrm{T}}C_y\nabla, B^{\mathrm{T}}Q_yB) \; , \end{aligned} \tag{3.29}$$

where $\underline{t}$ is the $M \times 1$ (with $M = \sum_{i=l}^{k} m_i$) vector of misclosures and $B^{\mathrm{T}}C_y$ is a known $M \times b$ matrix. In Section 2.5.1 we established the one-to-one relationship between the predicted residuals and the vector of misclosures, namely

$$\underline{t} = L\underline{v} \; , \tag{3.30}$$

with L a square and full-rank matrix. Equation (3.30) enables us to specify the hypotheses (3.29) as function of predicted residuals, viz:

$$\begin{aligned} H_0 : \;\; \underline{v} &\sim \;\; N(0, Q_v) \\ H_A : \;\; \underline{v} &\sim \;\; N(\nabla v, Q_v) \; , \end{aligned} \tag{3.31}$$

where $Q_v = L^{-1}B^{\mathrm{T}}Q_yBL^{-\mathrm{T}} = \mathrm{diag}(Q_{v_l}, \ldots, Q_{v_k})$, and the $M \times 1$ vector ∇v can be parametrized as

$$\nabla v = C_v\nabla \; , \tag{3.32}$$

with

$$C_v = L^{-1}B^{\mathrm{T}}C_y \; . \tag{3.33}$$

The $(M \times b)$ matrix C_v can be partitioned as

$$C_v = (C_{v_l}^{\mathrm{T}}, C_{v_l+1}^{\mathrm{T}}, \ldots, C_{v_k}^{\mathrm{T}})^{\mathrm{T}} \; ,$$

and is assumed to be known and of full rank b. In Section 3.5 we will show how the matrices C_{v_j} can be computed efficiently in a recursive manner. The hypotheses (3.31)

cover the time interval $[l, \ldots, k]$ and are called *global* hypotheses. If the hypotheses are limited to a single epoch k, the hypotheses (3.31) reduce to the *local* hypotheses:

$$\begin{aligned} H_{0_k}: \quad \underline{v}_k &\sim N(0, Q_{v_k}) \\ H_{A_k}: \quad \underline{v}_k &\sim N(\nabla v_k, Q_{v_k}) \,, \end{aligned} \tag{3.34}$$

where ∇v_k can be parametrized as $C_{v_k}\nabla$.

3.3.3 Test Statistics for Use in Dynamic Models

With (3.31) and (3.34) we have specified the null and alternative hypotheses in terms of the predicted residuals, and we now introduce the test statistics of the DIA procedure. Using the relation between the misclosure vector and the predicted residuals (3.30), we can write the GLR test statistic (3.25) of the UMPI tests for (3.31) and (3.34) as:

$$\underline{T} = \underline{v}^{\mathrm{T}} Q_v^{-1} L^{-1} B^{\mathrm{T}} C_y (C_y^{\mathrm{T}} B L^{-\mathrm{T}} Q_v^{-1} L^{-1} B^{\mathrm{T}} C_y)^{-1} C_y^{\mathrm{T}} B L^{-\mathrm{T}} Q_v^{-1} \underline{v} \,. \tag{3.35}$$

By thereupon inserting (3.33) in (3.35) the GLR test statistic can be written as:

$$\underline{T} = \underline{v}^{\mathrm{T}} Q_v^{-1} C_v (C_v^{\mathrm{T}} Q_v^{-1} C_v)^{-1} C_v^{\mathrm{T}} Q_v^{-1} \underline{v} \,, \tag{3.36}$$

and consequently the UMPI (GLR) test for testing H_0 against H_A reads:

$$\boxed{\text{Reject } H_0 \text{ in favour of } H_A \text{ if } v^{\mathrm{T}} Q_v^{-1} C_v (C_v^{\mathrm{T}} Q_v^{-1} C_v)^{-1} C_v^{\mathrm{T}} Q_v^{-1} v \;>\; k_\alpha \,,} \tag{3.37}$$

where k_α is the critical value. The critical value of the test can be derived from the distribution of the test statistic (3.36). From (3.31) and (3.36) it follows that $\underline{T}$ is distributed as:

$$\begin{aligned} H_0: \quad \underline{T} &\sim \chi^2(b, 0) \\ H_A: \quad \underline{T} &\sim \chi^2(b, \lambda) \end{aligned} \tag{3.38}$$

with $1 \leq b \leq M$. The non-centrality parameter

$$\lambda = \nabla^{\mathrm{T}} C_v^{\mathrm{T}} Q_v^{-1} C_v \nabla, \tag{3.39}$$

is obtained by substituting $C_v\nabla$ for $\underline{v}$ in (3.36).

The GLR test (3.37) could also have been obtained by inserting the likelihood functions of the predicted residuals under H_0 and H_A (cf. eq. 3.31)

$$\begin{aligned} H_0: \quad p_{\underline{v}}(v) &= (2\pi)^{-M/2} |Q_v|^{-1/2} \exp\{-\tfrac{1}{2} v^{\mathrm{T}} Q_v^{-1} v\} \\ H_A: \quad p_{\underline{v}}(v|\nabla) &= (2\pi)^{-M/2} |Q_v|^{-1/2} \exp\{-\tfrac{1}{2} (v - C_v\nabla)^{\mathrm{T}} Q_v^{-1} (v - C_v\nabla)\} \,, \end{aligned} \tag{3.40}$$

in (3.1) and (3.2), where (3.1) has been replaced by the likelihood ratio

$$\lambda(v) = \frac{p_{\underline{v}}(v)}{\max\limits_{\nabla \in \mathcal{R}^b} p_{\underline{v}}(v|\nabla)} \,. \tag{3.41}$$

With

$$\max_{\nabla \in \mathcal{R}^b} p_{\underline{v}}(v|\nabla) = p_{\underline{v}}(v|\hat{\nabla}), \tag{3.42}$$

where

$$\underline{\hat{\nabla}} = (C_v^{\mathrm{T}} Q_v^{-1} C_v)^{-1} C_v^{\mathrm{T}} Q_v^{-1} \underline{v} \,, \tag{3.43}$$

is the maximum likelihood estimator of ∇ (which because of the assumption of normally distributed observables is also the least squares estimator), the GLR test given by (3.2) can then be written as:

$$\text{Reject } H_0 \text{ if } v^{\mathrm{T}} Q_v^{-1} v - (v - C_v \hat{\nabla})^{\mathrm{T}} Q_v^{-1} (v - C_v \hat{\nabla}) > \ln a^{-2}, \tag{3.44}$$

which after inserting (3.43) in (3.44) is identical to (3.37).

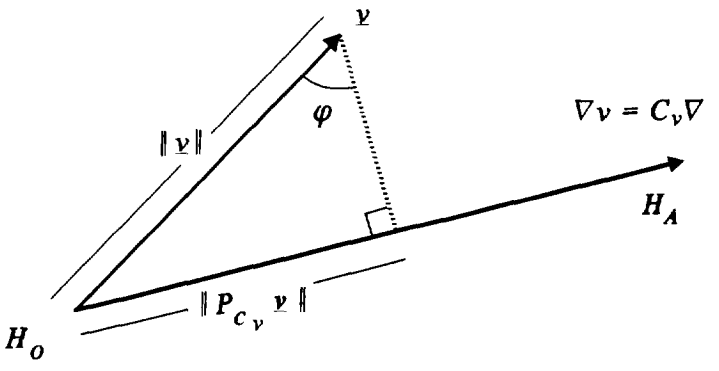


Figure 3.2: Predicted residual space with metric Q_v^{-1} and the two hypotheses H_0 and H_A.

If one assumes that in $\mathcal{R}^M$ the inner product is defined by Q_v^{-1}, the test statistic $\underline{T}$ (given by eq. 3.36) can be interpreted geometrically (cf. Fig. 3.2) as the square of the length of the vector that follows from projecting $\underline{v}$ orthogonally on the range space of C_v, viz:

$$\underline{T} = \|P_{C_v} \underline{v}\|^2 \,, \tag{3.45}$$

where we have used the definition of the orthogonal projector

$$P_{C_v} = C_v (C_v^{\mathrm{T}} Q_v^{-1} C_v)^{-1} C_v^{\mathrm{T}} Q_v^{-1} \,.$$

If C_v is square and of full rank (i.e. $\mathrm{rank}(C_v) = M$), (3.45) reduces to $\underline{T} = \|\underline{v}\|^2$.

The likelihood ratio test (3.37) is based on the fact that the complete covariance matrix Q_v is known. In the case the covariance matrix Q_v is known up to an unknown scale factor (i.e. Q_v can be written as $\sigma^2 \overline{Q}_v$ with $\overline{Q}_v$ known), one can still compute the likelihood ratio (although the likelihood functions under H_0 and H_A (cf. eq. 3.40) are

not valid anymore). Instead the following test statistic [TEUNISSEN, 1990b] has to be used (cf. Fig. 3.2):

$$\sin^2 \underline{\phi} = \frac{\underline{v}^T Q_v^{-1} C_v (C_v^T Q_v^{-1} C_v)^{-1} C_v^T Q_v^{-1} \underline{v}}{\underline{v}^T Q_v^{-1} \underline{v}}. \tag{3.46}$$

Under the null and alternative hypothesis the test statistic $\sin^2 \underline{\phi}$ is distributed as

$$\begin{aligned} H_0: \quad \sin^2 \underline{\phi} &\sim B(b, M, 0) \\ H_A: \quad \sin^2 \underline{\phi} &\sim B(b, M, \lambda) \end{aligned} \tag{3.47}$$

where $B(f_1, f_2, \lambda)$ is the Beta-distribution with f_1, f_2 degrees of freedom and non-centrality parameter λ. The corresponding test reads:

$$\text{Reject } H_0 \text{ in favour of } H_A \text{ if } \sin^2 \phi > B_\alpha(b, M, 0). \tag{3.48}$$

Instead of the test statistic $\sin^2 \underline{\phi}$ one may also take $\cos^2 \underline{\phi}$ or $\tan^2 \underline{\phi}$ as test statistic. Because of their functional dependency, they will give identical outcomes for the testing. It should be noted that since $\sin^2 \underline{\phi} \equiv 1$ for $b = M (= \sum_{i=l}^{k} m_i)$ (cf. eq. 3.45 and Fig. 3.2), the test statistic $\sin^2 \underline{\phi}$ is only applicable for $1 \leq b < M$. In the following we will assume the complete covariance matrix Q_v is known.

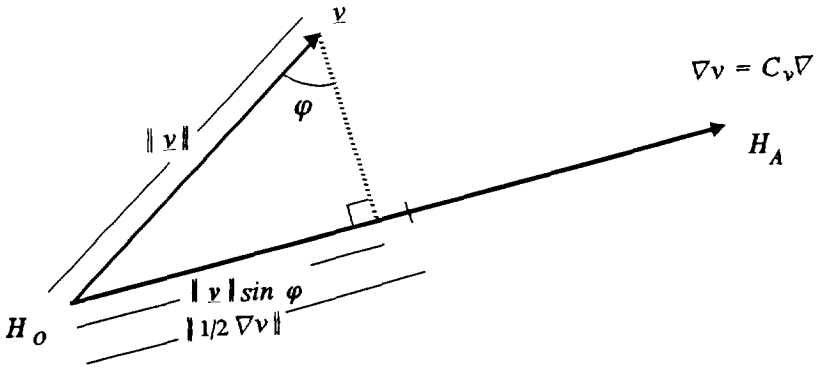


Figure 3.3: Predicted residual space with metric Q_v^{-1} and the two hypotheses H_0 and H_A with ∇ known.

In our derivations so far the vector ∇ was assumed to be unknown under H_A. A special case occurs if besides the matrix C_v also the vector ∇ is known under H_A. In this case the problem reduces to one of *discriminant analysis*. A geometric interpretation is given in Fig. 3.3. If ∇ is known the generalized likelihood ratio test reduces to a simple likelihood ratio test of the following form (cf. eq. 3.44):

$$\text{Reject } H_0 \text{ if } v^T Q_v^{-1} v - (v - C_v \nabla)^T Q_v^{-1} (v - C_v \nabla) > \ln a^{-2}, \tag{3.49}$$

where a is a positive constant. After some rearrangements one arrives at the following test

$$\text{Reject } H_0 \text{ in favour of } H_A \text{ if } \nabla^T C_v^T Q_v^{-1}(v - \frac{1}{2}C_v\nabla) > \ln a^{-1}. \tag{3.50}$$

The test statistic

$$\underline{T} = \nabla^T C_v^T Q_v^{-1}(\underline{v} - \frac{1}{2}C_v\nabla) \tag{3.51}$$

is a linear function of $\underline{v}$ and consequently $\underline{T}$ is distributed as:

$$\begin{aligned} H_0: \quad \underline{T} &\sim N(-\tfrac{1}{2}\nabla^T C_v^T Q_v^{-1} C_v\nabla, \nabla^T C_v^T Q_v^{-1} C_v\nabla) \\ H_A: \quad \underline{T} &\sim N(+\tfrac{1}{2}\nabla^T C_v^T Q_v^{-1} C_v\nabla, \nabla^T C_v^T Q_v^{-1} C_v\nabla)\ . \end{aligned} \tag{3.52}$$

If both hypotheses are equally likely, a is chosen equal to one, which corresponds to the case that H_0 is rejected if $\|v\| \sin\phi < \frac{1}{2}\|\nabla v\|$ (cf. Fig. 3.3). The decision rule for discriminating between H_0 and H_A then becomes:

$$\text{Accept } H_0 \text{ if } \nabla v^T Q_v^{-1}(v - \frac{1}{2}\nabla v) < 0 \text{ ; accept } H_A \text{ otherwise.} \tag{3.53}$$

Although this test is conceptually very simple, the possibility of two fully specified hypotheses H_0 and H_A very rarely occurs in practical applications.

Local and Global GLR Tests

Now the generalized likelihood ratio test has been derived we reconsider the *local* and *global* tests. The test statistic for the local test at time k follows from (3.36) as:

$$\underline{T}^k = \underline{v}_k^T Q_{v_k}^{-1} C_{v_k}(C_{v_k}^T Q_{v_k}^{-1} C_{v_k})^{-1} C_{v_k}^T Q_{v_k}^{-1} \underline{v}_k \ . \tag{3.54}$$

The local test of size α is now as follows:

$$\text{Reject } H_{0_k} \text{ if and only if } T^k > \chi_\alpha^2(b_k, 0)\ , \tag{3.55}$$

where $\chi_\alpha^2(b_k, 0)$ is the upper α probability point of the central χ^2-distribution with b_k degrees of freedom, with $1 \leq b_k \leq m_k$.

The appropriate global test statistic for testing the two (global) hypotheses given by (3.31) is

$$\underline{T}^{l,k} = \underline{v}^T Q_v^{-1} C_v(C_v^T Q_v^{-1} C_v)^{-1} C_v^T Q_v^{-1} \underline{v}\ , \tag{3.56}$$

and the global test of size α is as follows:

$$\text{Reject } H_0 \text{ if and only if } T^{l,k} > \chi_\alpha^2(b, 0)\ , \tag{3.57}$$

with $1 \leq b \leq M(= \sum_{i=l}^{k} m_i)$. The test statistic (3.56) is not yet in a form which is suitable for real-time applications. From the fact that the predicted residuals of

different epochs are uncorrelated (cf. Appendix A) it follows that (3.56) can also be written as:

$$\underline{T}^{l,k} = [\sum_{i=l}^{k} C_{v_i}^{\mathrm{T}} Q_{v_i}^{-1} \underline{v}_i]^{\mathrm{T}} [\sum_{i=l}^{k} C_{v_i}^{\mathrm{T}} Q_{v_i}^{-1} C_{v_i}]^{-1} [\sum_{i=l}^{k} C_{v_i}^{\mathrm{T}} Q_{v_i}^{-1} \underline{v}_i] \ . \tag{3.58}$$

It will be seen that this simplification is essential and facilitates the derivation of global test statistics with batch type properties in recursive form.

In the following section we will develop local and global test statistics for certain classes of alternative hypotheses. For most practical applications two particular forms of $\underline{T}^k$ and $\underline{T}^{l,k}$ are of special importance. Overall model tests correspond to the case that $b_k = m_k$ in (3.55) or $b = \sum_{i=l}^{k} m_i$ in (3.57). Slippage tests for identification purposes are often related to one-dimensional hypotheses, i.e. $b_k = 1$ or $b = 1$ in (3.55) and (3.57) respectively.

3.4 Detection

In the detection step of the DIA procedure one checks the overall validity of the null hypothesis. Therefore the tests associated with this phase are called *overall model tests*, and are used for detecting possible unspecified model errors in H_0.

We first consider local overall model (LOM) tests. If b_k is chosen to be equal to m_k, the vector ∇v_k of (3.34) remains completely unspecified. The matrix C_{v_k} becomes a square, non singular (and thus invertible) matrix and can be eliminated from (3.54). Consequently the LOM test statistic becomes

$$\underline{T}^k = \underline{v}_k^{\mathrm{T}} Q_{v_k}^{-1} \underline{v}_k \ . \tag{3.59}$$

The local overall model (LOM) test for testing the hypothesis H_{0_k} versus H_{A_k} is:

$$\boxed{\text{Reject } H_{0_k} \text{ if and only if } T^k = v_k^{\mathrm{T}} Q_{v_k}^{-1} v_k > \chi_\alpha^2(m_k, 0) \ .} \tag{3.60}$$

In a similar fashion the global overall model (GOM) test can be derived. If b is chosen equal to $\sum_{i=l}^{k} m_i$ the matrix C_v in (3.32) becomes square and invertible and can thus be eliminated from (3.56). Using the property that the predicted residuals are uncorrelated between epochs the test statistic (3.56) can be written as:

$$\underline{T}^{l,k} = \sum_{i=l}^{k} \underline{v}_i^{\mathrm{T}} Q_{v_i}^{-1} \underline{v}_i \ , \tag{3.61}$$

which under H_0 and H_A is distributed as given in (3.38). From (3.61) it follows that the GOM test statistic can also be computed recursively as:

$$\underline{T}^{l,k} = \underline{T}^{l,k-1} + \underline{T}^k \ . \tag{3.62}$$

The GOM test statistic reduces to the LOM test statistic (3.59) for $l = k$. The test statistic covers the complete mathematical model up to k if l is chosen equal to one. The global overall model (GOM) test for testing the hypothesis H_0 versus H_A is:

$$\boxed{\text{Reject } H_0 \text{ if and only if } T^{l,k} = \sum_{i=l}^{k} v_i^{\mathrm{T}} Q_{v_i}^{-1} v_i > \chi^2_\alpha(\sum_{i=l}^{k} m_i, 0)\,.} \tag{3.63}$$

3.4.1 Practical Considerations

Although the global detection test statistics can be computed efficiently in real-time (the predicted residual $\underline{v}_k$ and its covariance matrix Q_{v_k} are readily available during each measurement update), there may still be the practical problem of a delay in time of detection. Besides the detection test statistics cannot be directly compared.

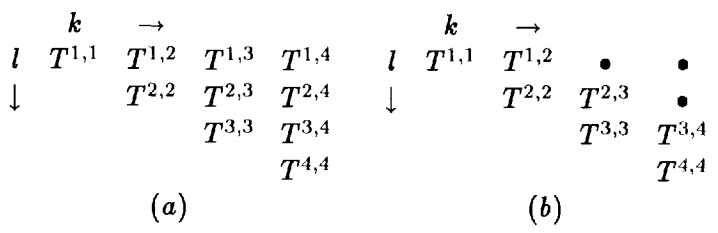


Figure 3.4: The detection test statistic $\underline{T}^{l,k}$ with (a) no window, (b) a moving window with $N = 2$.

From a practical point of view it is impossible to compute all detection test statistics $T^{l,k}$ starting at $l = 1$ for all $k \geq l$. This situation is shown in Fig. 3.4a. In order to reduce the number of computations and the delay time of detection, it is worthwhile to introduce a moving *window* of length N by constraining l to $k - N + 1 \leq l \leq k$. This is shown in Fig. 3.4b. With this window the delay time of detection is at the most equal to $N - 1$. When choosing N one of course has to make sure that the detection power of the test statistic $\underline{T}^{k-N+1,k}$ is still sufficient. This is typically a problem one should take into consideration when designing the filter. The choice of windows for the detection tests has to be in accordance with those of the identification tests, which are discussed in the next section.

Once a window for the detection tests has been chosen one has to be careful in drawing conclusions from the test results. First of all it is impossible to specify the *type* of model error which caused the rejection of the null hypothesis. Furthermore it is difficult to infer at which epoch the model error occurred, i.e. the *time of occurence* l. Under the null hypothesis the test statistics $\underline{T}^{l,k}$ with l in the interval $k - N + 1 \leq l \leq k$ are all distributed as $\chi^2(b, 0)$, but the degrees of freedom b are different for each overall model test and consequently the test statistics cannot be directly compared. A comparison can be made if one normalizes the test statistics with respect to the critical

values of the associated test. For each test the critical value k_α can be computed from the central χ^2-distribution with b degrees of freedom. The 'largest' detection test statistic $T^{l,k}_{\max}$ is the one for which holds that

$$\frac{T^{l,k}_{\max}}{k_\alpha} > \frac{T^{l,k}}{k_\alpha};\ \forall\ l\ \epsilon\ [k-N+1 \leq l \leq k]. \tag{3.64}$$

The choice of the size α of the tests (3.60) and (3.63) has not been discussed yet. For the moment we assume we follow the B-method of testing [BAARDA, 1968], which will be discussed in Section 3.7.

3.5 Identification

In the detection phase of the DIA-procedure we test if the null hypothesis is valid without specifying a particular alternative hypothesis. If the null hypothesis has been rejected one has to search for possible model misspecifications. In the identification phase one must specify alternative hypotheses which could account for the rejection of the null hypothesis by the overall model tests. The specification of possible alternative hypotheses is application dependent and is one of the most difficult tasks in hypothesis testing. One must consider which *types* of model errors are likely to occur and also if local identification test statistics are sufficiently powerful to identify the relevant, likely hypothesis. It is self-evident that the identification of certain types of model errors, e.g. 'soft' sensor failures, requires global tests. In the following we will discuss local and global test statistics for identification purposes. It will be shown how the global test statistics can be computed recursively. We then discuss the identification test procedure. Because it is very difficult to specify generally useful multi-dimensional alternative hypotheses, we limit ourselves to one-dimensional alternative hypotheses. It should be noted that all one-dimensional test statistics (unless indicated otherwise) can be (based on eq. 3.58) generalized to multi-dimensional test statistics.

3.5.1 Identification Test Statistics

If b_k is chosen equal to 1, the matrix C_{v_k} of (3.54) reduces to a vector, which will be denoted by c_{v_k}, and the vector ∇_k reduces to a scalar. In this case the *local* test statistic can be written as:

$$\underline{T}^k = (\underline{t}^k)^2\ , \tag{3.65}$$

with[1]

$$\underline{t}^k = \frac{c^{\mathrm{T}}_{v_k} Q^{-1}_{v_k} \underline{v}_k}{\sqrt{c^{\mathrm{T}}_{v_k} Q^{-1}_{v_k} c_{v_k}}}\ , \tag{3.66}$$

which under H_{0_k} is distributed as $N(0,1)$. This local test statistic can be used to identify particular one-dimensional misspecifications in H_{0_k}, such as a slippage in the

[1] The lower case kernel letter t will be used for one-dimensional slippage test statistics.

mean of the predicted state, a slippage in the mean of the observables, or a slippage in the mean of a combination of the observables and the predicted state. Hence we call (3.66) a *local slippage* (LS) test statistic. If one, for instance, suspects sensor failures or outlying observations one can follow the *datasnooping* approach [BAARDA, 1968] by choosing m_k number of vectors c_{v_k} of the form

$$\begin{array}{cccccc} c_i & = & (0 & \cdots & 1 & \cdots & 0)^{\mathrm{T}} \\ m_k \times 1 & & 1 & & i & & m_k \end{array} , \tag{3.67}$$

for $i = 1, \ldots, m_k$.

If b is chosen equal to one in (3.56) one obtains the *global* one-dimensional identification test statistic. The matrices C_{v_i}, with $i = l, \ldots, k$, reduce to vectors, which are denoted as c_{v_i}. The corresponding one-dimensional *global slippage* (GS) test statistic reads (cf. eq. 3.58):

$$\underline{t}^{l,k} = \frac{\sum_{i=l}^{k} c_{v_i}^{\mathrm{T}} Q_{v_i}^{-1} \underline{v}_i}{(\sum_{i=l}^{k} c_{v_i}^{\mathrm{T}} Q_{v_i}^{-1} c_{v_i})^{1/2}} , \tag{3.68}$$

which under H_0 is distributed as $N(0,1)$. Note that this test statistic reduces to the one-dimensional LS test statistic (3.66) for $l = k$. The one-dimensional global slippage test statistic $\underline{t}^{l,k}$ can be used to identify particular one-dimensional global misspecifications in H_0. In order to be able to use the test statistic $\underline{t}^{l,k}$ in real-time we need a recursive scheme for updating the vectors c_{v_i}, $i = l, \ldots, k$. Various cases, depending on the choice of alternative hypothesis, can be considered. We will consider the following four cases:

a) A jump in the state vector at time l.

b) A permanent slip in the state vector that starts at time l.

c) A single slip (outlier) in the vector of observables at time l.

d) A sensor slip that starts at time l.

We use the one-dimensional vector c_x to specify the type of model error in the dynamic model (cases a and b) and similarly the vector c_y for the specification of model errors in the observations (cases c and d). The choice of the vector c_y is often rather straightforward, whereas the choice of the vector c_x is dependent on the specification of the state vector. By combining the time and measurement update equations of the Kalman filter, *recursive* schemes for the vectors c_{v_i}, $i = l, \ldots, k$ corresponding to cases a) to d) can be derived.

The recursive scheme one obtains for a jump in the state vector (which at time l manifests itself as a disturbance $d_i \neq 0$) reads

$$\boxed{\begin{array}{lcl} c_{v_i} & = & -A_i X_{i,l}, \quad i = l, \ldots, k \\ X_{i+1,l} & = & \Phi_{i+1,i}(I - K_i A_i) X_{i,l}; \; X_{l,l} = c_x \end{array}} \tag{3.69}$$

The recursive scheme for a permanent slip in the state vector (which after time l manifests itself as a systematic disturbance $d_i \neq 0$) reads

$$\boxed{\begin{array}{lcl} c_{v_i} & = & -A_i X_{i,l}, \quad i = l, \ldots, k \\ X_{i+1,l} & = & c_x + \Phi_{i+1,i}(I - K_i A_i) X_{i,l}; \; X_{l,l} = c_x \end{array}} \tag{3.70}$$

The recursive scheme for an outlier in the vector of observables at time l reads

$$\boxed{\begin{array}{lcl} c_{v_i} & = & \begin{cases} c_y & \text{for } i = l \\ -A_i X_{i,l} & \text{for } i = l+1, \ldots, k \end{cases} \\ X_{i+1,l} & = & \Phi_{i+1,i}(X_{i,l} + K_i c_{v_i}); \; X_{l,l} = 0 \end{array}} \tag{3.71}$$

The vector c_y is not specified explicitely here, but can, e.g., be chosen as in (3.67). Finally the recursive scheme for a slip in the jth sensor reads

$$\boxed{\begin{array}{lcl} c_{v_i} & = & c_j - A_i X_{i,l}, \quad i = l, \ldots, k \\ X_{i+1,l} & = & \Phi_{i+1,i}(X_{i,l} + K_i c_{v_i}); \; X_{l,l} = 0 \end{array}} \tag{3.72}$$

where

$$c_j = (0 \; \cdots \; \underset{j}{1} \; \cdots \; 0)^{\mathrm{T}} .$$

The matrices $X_{i,l}$ (here $n \times 1$ vectors), with $i > l$, in (3.69) to (3.72) describe the response of a model error on the predicted state $\hat{x}_{i|i-1}$. With eqs. 3.69 to 3.72 we are now able to compute the one-dimensional global test statistic (3.68) recursively.

It will be clear that our global recursive test statistics are more sensitive to global model errors than the local test statistics. The difference in detection power between the local and global one-dimensional slippage tests for a particular model error follows when one compares the noncentrality parameters ∇t^k and $\nabla t^{l,k}$ of the two test statistics. For the local test statistic (3.66) we have (cf. eq. 3.39)

$$\nabla t^k = (c_{v_k}^{\mathrm{T}} Q_{v_k}^{-1} c_{v_k})^{1/2} \nabla , \tag{3.73}$$

and for our global test statistic (3.68) we have

$$\nabla t^{l,k} = (\sum_{i=l}^{k} c_{v_i}^{\mathrm{T}} Q_{v_i}^{-1} c_{v_i})^{1/2} \nabla . \tag{3.74}$$

Since the matrices Q_{v_i}, $i = l, \ldots, k$ are positive definite, this result shows that $\nabla t^{l,k}$ is an increasing function of k and that $\nabla t^{l,k} > \nabla t^k$ for $k > l$. Hence, with increasing k the detection power of the global test increases and is never less than the detection power of the local test.

3.5.2 Identification Procedure

After detection the most likely model error has to be identified. If more than one alternative hypothesis is specified, one has to determine the *type* and the *time of occurence* of the model error. Because we only consider one-dimensional alternative hypotheses it is possible to compare the test statistics directly. The most likely model error is the one corresponding to the largest test statistic. If the largest test statistic is larger than the critical value of the test, the corresponding alternative hypothesis is declared valid. In the next step one can estimate the model error related to the identified alternative hypothesis. Estimation of the model error is discussed in the next section on adaptation. Summarizing, the identification procedure consists of three steps:

1. A search for the largest test statistic among all specified alternative hypotheses.
2. A check if the largest test statistic is larger than the critical value of the test.
3. If a particular H_A is declared valid the associated model error is estimated.

In a more formal manner the procedure can be given as (assuming only one-dimensional alternative hypotheses are specified):

$$\boxed{\begin{array}{ll} 1. & |t^{l,k}|_{\max} = \max\limits_{k-(N_{\max}-1)\le l\le k-L_{\min}} |t^{l,k}| \\ 2. & \text{If } |t^{l,k}|_{\max} > N_{\alpha/2}(0,1) \text{ declare the corresponding } H_A \text{ valid} \\ 3. & \text{If } H_A \text{ is valid then estimate } \hat{\nabla}^{l,k} \text{ associated with } |t^{l,k}|_{\max} \text{ (cf. Section 3.6),} \end{array}} \tag{3.75}$$

where $(N_{\max}-1)$ and $L_{\min}$ are the maximum delay and minimal lag of the test statistics respectively.

3.5.3 Practical Considerations

From a computational point of view it is impossible to compute all test statistics starting at $l = 1$ (cf. Fig. 3.5a). In order to reduce the number of computations and the delay time of detection, it is worthwhile to consider introducing a moving *window* of length N by constraining l to $k - N + 1 \le l \le k$ (the delay time of detection is then at the most equal to $N - 1$ (cf. Fig. 3.5b)). Finally, the test statistic $\underline{t}^{l,k}$ may be too insensitive for identifying model errors if $l > k - L$. One then can limit oneself to the computation of the test statistics with delay $N - 1$ and lag L (cf. Fig. 3.5c). The design procedure (cf. Chapter 4) will provide information on the actual detection power of the envisaged tests, and is a useful tool in choosing the window size of the tests.

By limiting the computation of the test statistics to a certain window the additional computational load associated with the identification procedure is limited. The additional computations required for the identification step are basically given by the recursive schemes of (3.69) to (3.72), but all the quantities required for these computations

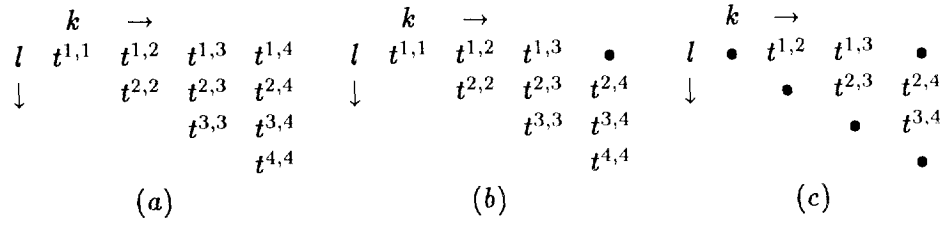


Figure 3.5: The one-dimensional recursive identification test statistic $\underline{t}^{l,k}$ with (a) no window, (b) a moving window with $N = 3, L = 0$, and (c) a moving window with $N = 3, L = 1$.

are provided by the filter. The number of operations related to eqs. 3.69-3.72 grows linearly with the window lengths of the tests and the number of different alternative hypotheses considered. If one limits oneself to local identification the computational costs are almost negligible compared with those of the filter operations.

In practice we have to deal with different types of alternative hypotheses. Therefore it seems unlikely that the delay and lag of all identification tests are chosen identical. In general one will try to design the testing procedure in such a way that real-time corrective action remains as much as possible. Consequently the testing procedure will usually also be based on local tests, even if the detection power of the local tests is small.

The detection tests (cf. Section 3.4) were also based on windows. The DIA procedure is based on the fact that identification takes place after a model error has been detected. Therefore the largest delay of the detection tests (cf. Fig. 3.4) should be chosen at least as large as the largest delay for the identification tests.

The identification procedure as given by (3.75) should be used with care. If, for example, frequently a model error is detected and simultaneously no specific model error can be identified, one should reconsider the choice of alternative hypotheses. If besides the actual model error cannot be specified as a model error of additive nature, it cannot be identified in an optimal manner by our slippage tests. The identification procedure is based on the assumption that model errors are sufficiently separated in time to allow for individual detection and identification. If it is likely that model errors can occur (nearly) simultaneously the procedure given by (3.75) has to be refined. Furthermore we tacitly assumed that the testing parameters for every identification test (the size α and the power γ) are identical. The choice of the testing parameters will be considered in more detail in Section 3.7, where we discuss the application of the B-method of testing. Finally the comparison of the various test statistics is not that straightforward if one also considers multi-dimensional alternative hypotheses.

3.6 Adaptation

The final step in the DIA procedure is adaptation and is needed to eliminate the bias in the filtered state. The moment a model error is detected and identified, the real-time operation of the Kalman filter requires that corrective action is taken immediately. Herein lies the major difference with the testing procedures for geodetic networks, where adaptation is generally performed off-line. The requirement of real-time operation necessitates that an automatic adaptation procedure is devised, which maintains, as much as possible, the optimality of the Kalman filter. After detection of a model error using local tests, real-time adaptation is possible. In many cases, however, a model error will be identified with a certain *delay of detection* (this is a direct consequence of the concept of global testing). In the event of delayed detection at time k the state estimate is biased in the time interval l (the time of occurence of the model error) to k. We have to find a strategy for handling the delayed detection of a slip, but in practice we will often opt for a simple approach that resets the filter at time k, leaving the state estimates biased in the time interval l to k. The rationale of this approach, which will also be followed here, is that for real-time applications we are primarily interested in the present estimate of the state. Besides the bias in the state estimate is probably small as the model error cannot be detected until time k. It will be clear that optimal results are obtained if one is able to design a filter that is capable of following the correct alternative hypothesis at the correct time of occurence. Such an approach requires a whole bank of filters, each one taylored for a particular alternative hypothesis. The efficiency of the DIA-procedure is largely due to the implementation of a detection and identification procedure parallel to a single filter. We will therefore restrict ourselves to adaptation strategies that also operate parallel to a *single* filter.

We will first consider the estimation of model errors after identification and show how the biased state can be adapted. We will then discuss the optimal adaptation procedure, which requires a continuous updating of the filtered state. We will show that the adaptation for outliers can also be performed in an optimum manner by performing the adaptation step only once and reverting to the filter under the null hypothesis. For slips such a simple, optimal adaptation procedure is only possible for the so-called partially constant state space models [TEUNISSEN, 1992]. If one wants to continue the filter under the null hypothesis after a single adaptation step for slips, generally only approximate (suboptimal) solutions are available. We assume that adaptation is performed for one-dimensional alternative hypotheses only, and hence scalar model errors ∇ are considered. The generalization to b-dimensional error vectors is straightforward.

3.6.1 Estimation of Model Errors

At time k the estimator of the model error ∇ with a time of occurence l reads:

$$\underline{\hat{\nabla}}^{l,k} = \frac{\sum_{i=l}^{k} c_{v_i}^{\mathrm{T}} Q_{v_i}^{-1} \underline{v}_i}{\sum_{i=l}^{k} c_{v_i}^{\mathrm{T}} Q_{v_i}^{-1} c_{v_i}} , \tag{3.76}$$

with

$$Q_{\hat{\nabla}^{l,k}} = (\sum_{i=l}^{k} c_{v_i}^{\mathrm{T}} Q_{v_i}^{-1} c_{v_i})^{-1} ,$$

where the vectors c_{v_i} are computed according to (3.69) to (3.72). Note the close relationship with the test statistic (3.68). The estimator (3.76) follows from (3.43) directly if we take the uncorrelatedness of the predicted residuals between epochs into account, and is equivalent to the least squares estimator based on the following model of observation equations:

$$E\{\begin{pmatrix} \underline{v}_l \\ \underline{v}_{l+1} \\ \vdots \\ \underline{v}_k \end{pmatrix}\} = \begin{pmatrix} c_{v_l} \\ c_{v_{l+1}} \\ \vdots \\ c_{v_k} \end{pmatrix} \nabla \; ; \; \begin{pmatrix} Q_{v_l} & & & \\ & Q_{v_{l+1}} & & \\ & & \ddots & \\ & & & Q_{v_k} \end{pmatrix} . \tag{3.77}$$

From (3.77) it can be easily seen that the estimator of the model error can be computed in a recursive manner. At epoch l the recursive bias estimator is initialized as:

$$\underline{\hat{\nabla}}^{l,l} = (c_{v_l}^{\mathrm{T}} Q_{v_l}^{-1} c_{v_l})^{-1} c_{v_l}^{\mathrm{T}} Q_{v_l}^{-1} \underline{v}_l \quad , \tag{3.78}$$

with

$$Q_{\hat{\nabla}^{l,l}} = (c_{v_l}^{\mathrm{T}} Q_{v_l}^{-1} c_{v_l})^{-1} . \tag{3.79}$$

After initialization the estimator $\underline{\hat{\nabla}}^{l,i}$, with $l < i \leq k$, and its covariance matrix are computed recursively as:

$$\underline{\hat{\nabla}}^{l,i} = \underline{\hat{\nabla}}^{l,i-1} + G_i(\underline{v}_i - c_{v_i}\underline{\hat{\nabla}}^{l,i-1}) \tag{3.80}$$

$$Q_{\hat{\nabla}^{l,i}} = (I - G_i c_{v_i}) Q_{\hat{\nabla}^{l,i-1}} \quad , \tag{3.81}$$

where

$$G_i = Q_{\hat{\nabla}^{l,i-1}} c_{v_i}^{\mathrm{T}} (Q_{v_i} + c_{v_i} Q_{\hat{\nabla}^{l,i-1}} c_{v_i}^{\mathrm{T}})^{-1} , \tag{3.82}$$

is the $1 \times m_i$ gain vector.

It can be shown that the estimator of the bias and the estimator of the state vector under H_0 are uncorrelated[2]

$$E\{(\hat{\underline{\nabla}}^{l,k} - E\{\hat{\underline{\nabla}}^{l,k}\})(\hat{\underline{x}}^0_{k|k} - E\{\hat{\underline{x}}^0_{k|k}\})^{\mathrm{T}}\} = 0 \; . \tag{3.83}$$

This important relationship will not be proven here. If one, however, recalls that the Kalman filter can be written as a least-squares adjustment problem and the error estimators are a function of the least-squares residuals, which are uncorrelated with the least-squares estimators of the unknowns, it follows intuitively that (3.83) is true.

3.6.2 Optimal Adaptation

With the recursive schemes (3.69) to (3.72) we have available the 'response' matrices $X_{i,l}$ (in our situation of one-dimensional model errors the matrices $X_{i,l}$ reduce to $n \times 1$ vectors), which describe the response to a model error (starting) at epoch l on the estimator of the predicted state at epoch i, viz. $\hat{\underline{x}}_{i|i-1}$. The 'response' matrix of the model error on the filtered state estimator $\hat{\underline{x}}_{i|i}$ at epoch i is given as

$$\overline{X}_{i,l} = \Phi_{i,i+1} X_{i+1,l} \; , \tag{3.84}$$

where $X_{i+1,l}$ is obtained through (3.69), (3.71) or (3.72). If the model error is associated with a slip in the state vector (cf. eq. 3.70), one obtains

$$\overline{X}_{i,l} = \Phi_{i,i+1}(X_{i+1,l} - c_x) \; . \tag{3.85}$$

At and after the time of identification k, the filtered state can be updated as

$$\boxed{\hat{\underline{x}}^{\mathrm{a}}_{i|i} = \hat{\underline{x}}^0_{i|i} - \overline{X}_{i,l}\hat{\underline{\nabla}}^{l,i}} \; , \tag{3.86}$$

with $i \geq k$, where the estimator $\hat{\underline{\nabla}}^{l,i}$ is computed with (3.76) or (3.80). Error propagation applied to (3.86) yields for $i \geq k$:

$$\boxed{\begin{aligned} P^{\mathrm{a}}_{i|i} &= P^0_{i|i} + \overline{X}_{i,l} Q_{\hat{\nabla}^{l,i}} \overline{X}^{\mathrm{T}}_{i,l} \\ P_{\hat{x}^{\mathrm{a}}_{i|i},\hat{\nabla}} &= -\overline{X}_{i,l} Q_{\hat{\nabla}^{l,i}} \end{aligned}} \tag{3.87}$$

where use has been made of the fact that $\hat{\underline{x}}^0_{i|i}$ and $\hat{\underline{\nabla}}^{l,i}$ are uncorrelated. Equations (3.86) and (3.87) constitute the optimal adaptation equations for the model errors described by (3.69) to (3.72).

Actually the estimators obtained by the Kalman filter operating under H_0 when combined with the recursive bias estimator given by (3.78) to (3.82) and the update

[2]The superscripts 0 and a indicate that an estimate is obtained under H_0 and H_A respectively.

equations (3.86) and (3.87) are identical to the estimators that would have been obtained by the filter operating under H_A. FRIEDLAND [1968] was the first to show that the approach based on the filter operating under the null-hypothesis and a parallel bias filter is equivalent to an augmented filter based on the extended model:

$$E\{\underline{\tilde{d}}_k\} = z_k - F_{k,k-1} z_{k-1} \; ; \; E\{\underline{y}_k\} = \tilde{A}_k z_k \; , \tag{3.88}$$

with

$$z_k = \begin{pmatrix} x_k \\ \nabla \end{pmatrix} \; ; \; F_{k,k-1} = \begin{pmatrix} \Phi_{k,k-1} & B_k \\ 0 & 1 \end{pmatrix} \; ; \; D\{\underline{\tilde{d}}_k\} = \begin{pmatrix} Q_k & 0 \\ 0 & 0 \end{pmatrix} \; ; \; \tilde{A}_k = \begin{pmatrix} A_k & C_k \end{pmatrix} .$$

If one considers slips in the state vector $B_k = c_x$ and $C_k = 0$, whereas for a slip in observation j $B_k = 0$ and $C_k = c_j$. The (minor) difference between the error estimation in the DIA procedure and FRIEDLAND's estimator lies in the fact that we do not assume the bias is present from the outset. The separated-bias approach clearly shows that one can actually filter under the null hypothesis and simultaneously estimate possible model errors, i.e. explicit filtering under the alternative hypothesis using an augmented filter is not necessary.

3.6.3 Adaptation for Outliers

Although it may not be directly apparent from (3.86), the adaptation procedure for outliers is also optimal for the estimator of the state if adaptation is performed only once and after that one reverts to the filter under H_0. This result follows immediately if one formulates the filtering problem as a batch-type least-squares adjustment problem with observation equations. If one considers an outlier at epoch l, the unknown parameter ∇ appears only in a single observation equation[3]

$$E\{\underline{y}_l\} = A_l x_l + c_y \nabla \; ; \; D\{\underline{y}_l\} = R_l \; .$$

After adaptation at time k one obtains the following set of observation equations (cf. eqs. 3.86 and 3.87):

$$E\left\{\begin{pmatrix} \underline{\hat{x}}^{\mathrm{a}}_{k|k} \\ \underline{\hat{\nabla}}^{l,k} \\ \underline{d}_{k+1} \\ \underline{y}_{k+1} \\ \vdots \end{pmatrix}\right\} = \begin{pmatrix} I & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 1 \\ -\Phi_{k+1,k} & I & \cdots & 0 \\ 0 & A_{k+1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} x_k \\ x_{k+1} \\ \vdots \\ \nabla \end{pmatrix} \; ; \tag{3.89}$$

$$\begin{pmatrix} P_{\hat{x}^{\mathrm{a}}_{k|k}} & P_{\hat{x}^{\mathrm{a}}_{k|k},\hat{\nabla}} & 0 & 0 & \cdots \\ P_{\hat{\nabla},\hat{x}^{\mathrm{a}}_{k|k}} & Q_{\hat{\nabla}^{l,k}} & 0 & 0 & \cdots \\ 0 & 0 & Q_{k+1} & 0 & \cdots \\ 0 & 0 & 0 & R_{k+1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} .$$

[3] The same holds if one considers a jump in the state vector.

It can be seen from (3.89) that the parameter ∇ is present in a single observation equation only and can thus be considered as a free y^R-variate (cf. Section 2.3.3). One can thus revert to the filter under H_0, because a better estimator of ∇ for $i \geq k$ will have no impact on the state estimators $\underline{\hat{x}}_{i|i}$. This means that at the time of adaptation $\underline{\hat{x}}^{\mathrm{a}}_{k|k}$ is taken as the new initial state (with covariance matrix $P^{\mathrm{a}}_{k|k}$) and the 'standard' Kalman filter can be used for all following epochs. This is of course very convenient from a computational point of view, because neither the bias filter (cf. eqs. 3.78 to 3.82) nor the adaptation steps (3.86) and (3.87) have to be continued. The estimator $\underline{\hat{\nabla}}^{l,i}$ can still be improved as a y^R-variate (because of the correlation between the observables $\underline{\hat{x}}^{\mathrm{a}}_{k|k}$ and $\underline{\hat{\nabla}}^{l,k}$ in (3.89)) and will be identical to the estimator given by (3.80).

In the following we will illustrate that the results of the optimal adaptation procedure and the filter reverted to H_0 after adaptation are identical by showing that the estimators for the bias at epoch $k+1$ are the same whether one adapts for an outlier at epoch k or at $k+1$.

If one assumes adaptation has taken place at epoch k and the one-step prediction to epoch $k+1$ is performed, one obtains the following model of observation equations for the measurement update at $k+1$:

$$E\{\begin{pmatrix} \underline{\hat{x}}^{\mathrm{a}}_{k+1|k} \\ \underline{\hat{\nabla}}^{l,k} \\ \underline{y}_{k+1} \end{pmatrix}\} = \begin{pmatrix} I & 0 \\ 0 & 1 \\ A_{k+1} & 0 \end{pmatrix} \begin{pmatrix} x_{k+1} \\ \nabla \end{pmatrix} ; \begin{pmatrix} P_{\hat{x}^{\mathrm{a}}_{k+1|k}} & P_{\hat{x}^{\mathrm{a}}_{k+1|k},\hat{\nabla}} & 0 \\ P_{\hat{\nabla},\hat{x}^{\mathrm{a}}_{k+1|k}} & Q_{\hat{\nabla}^{l,k}} & 0 \\ 0 & 0 & R_{k+1} \end{pmatrix}, \tag{3.90}$$

where

$$\begin{aligned}
\underline{\hat{x}}^{\mathrm{a}}_{k+1|k} &= \Phi_{k+1,k}(\underline{\hat{x}}^{0}_{k|k} - \overline{X}_{k,l}\underline{\hat{\nabla}}^{l,k}) + \underline{d}_{k+1} \\
&= \underline{\hat{x}}^{0}_{k+1|k} - X_{k+1,l}\underline{\hat{\nabla}}^{l,k} \\
P^{\mathrm{a}}_{k+1|k} &= \Phi_{k+1,k}(P^{0}_{k|k} + \overline{X}_{k,l}Q_{\hat{\nabla}^{l,k}}\overline{X}^{\mathrm{T}}_{k,l})\Phi^{\mathrm{T}}_{k+1,k} + Q_{k+1} \\
&= P^{0}_{k+1|k} + X_{k+1,l}Q_{\hat{\nabla}^{l,k}}X^{\mathrm{T}}_{k+1,l} \\
P_{\hat{\nabla},\hat{x}^{\mathrm{a}}_{k+1|k}} &= -Q_{\hat{\nabla}^{l,k}}\overline{X}^{\mathrm{T}}_{k,l}\Phi^{\mathrm{T}}_{k+1,k} \\
&= -Q_{\hat{\nabla}^{l,k}}X^{\mathrm{T}}_{k+1,l}\,.
\end{aligned}$$

In equation (3.90) $\underline{\hat{\nabla}}^{l,k}$ is a free y^R-variate and using the relation $\underline{\hat{e}}^R = Q_{Ry}Q_y^{-1}\underline{\hat{e}}$ (cf. eq. 2.20) the estimator of the bias at epoch $k+1$ is obtained as

$$\begin{aligned}
\underline{\hat{\nabla}}^{l,k+1} &= \underline{\hat{\nabla}}^{l,k} - \underline{\hat{e}}_{\hat{\nabla}^{l,k}} \\
&= \underline{\hat{\nabla}}^{l,k} - P_{\hat{\nabla},\hat{x}^{\mathrm{a}}_{k+1|k}}(P^{\mathrm{a}}_{k+1|k})^{-1}\underline{\hat{e}}_{\hat{x}_{k+1|k}}\,,
\end{aligned} \tag{3.91}$$

where

$$\underline{\hat{e}}_{\hat{x}_{k+1|k}} = -P_{\hat{x}^{\mathrm{a}}_{k+1|k}}A^{\mathrm{T}}_{k+1}(Q^{0}_{v_{k+1}} + c_{v_{k+1}}Q_{\hat{\nabla}^{l,k}}c^{\mathrm{T}}_{v_{k+1}})^{-1}(\underline{v}_{k+1} - c_{v_{k+1}}\underline{\hat{\nabla}}^{l,k})$$

follows directly from solving model (3.90) for $\boldsymbol{x}_{k+1}$ using the method of condition equations. Inserting $\underline{\hat{e}}_{\hat{x}_{k+1|k}}$ in (3.91) and using the matrix inversion lemma finally yields:

$$\underline{\hat{\nabla}}^{l,k+1} = \underline{\hat{\nabla}}^{l,k} + G_{k+1}(\underline{v}_{k+1} - c_{v_{k+1}}\underline{\hat{\nabla}}^{l,k}) \ ,$$

which is identical to (3.80). In a similar way one can show (after some lengthly algebraic manipulations) that

$$\begin{aligned} \underline{\hat{\boldsymbol{x}}}^{\mathrm{a}}_{k+1|k+1} &= \underline{\hat{\boldsymbol{x}}}^{\mathrm{a}}_{k+1|k} - \underline{\hat{e}}_{\hat{x}_{k+1|k}} \\ &= \underline{\hat{\boldsymbol{x}}}^{0}_{k+1|k+1} - \overline{X}_{k+1,l}\underline{\hat{\nabla}}^{l,k+1} \ , \end{aligned}$$

which indeed corresponds with the optimal estimator of the adapted state (cf. eq. 3.86).

3.6.4 Adaptation for Slips

For slips we cannot revert to the filter under H_0 and obtain optimal estimators of the filtered states after a single adaptation step. If we consider, for example, slips in the observations, the observation equations (after adaptation at time k) are:

$$E\{\begin{pmatrix} \underline{\hat{\boldsymbol{x}}}^{\mathrm{a}}_{k|k} \\ \underline{\hat{\nabla}}^{l,k} \\ \underline{d}_{k+1} \\ \underline{y}_{k+1} \\ \vdots \end{pmatrix}\} = \begin{pmatrix} I & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 1 \\ -\Phi_{k+1,k} & I & \cdots & 0 \\ 0 & A_{k+1} & \cdots & c_y \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_k \\ \boldsymbol{x}_{k+1} \\ \vdots \\ \nabla \end{pmatrix} \ , \qquad (3.92)$$

and now the parameter ∇ appears in every observation equation of $\underline{y}_i$, with $i > k$. Therefore $\underline{\hat{\nabla}}^{l,k}$ cannot be treated as a free y^R-variate any longer, and we cannot revert to filtering under H_0 in an optimal way. A notable exception exists for slips in the so-called partially constant state-space model, where filtering under H_0 after a single adaptation step is possible.

3.6.5 Adaptation for Slips in the Partially Constant State Space Model

Assume that under the null hypothesis the observables (at an epoch i) are related to a constant bias z as follows:

$$E\{\underline{y}_i\} = A_i\boldsymbol{x}_i + C_i z \ ; \ D\{\underline{y}_i\} = R_i \ , \qquad (3.93)$$

where z is a vector of parameters with $1 \leq \dim(z) \leq m$[4]. The $(m_i \times \dim(z))$-matrix C_i indicates which observables are related to the constant bias state z. TEUNISSEN [1992] was the first to consider the adaptation for slips for models given by (3.93), which he calls partially constant state space (PCSS) models, which are of special importance in

[4] This means that not every observable $\underline{y}_i$ is necessarily related to a bias state z.

GPS data processing. In the following we will follow TEUNISSEN's derivation (based on a batch type least-squares adjustment problem) and show how one can revert to the filter under H_0 after a single adaptation step for a slip in an observable $\underline{y}_i$. We consider the case where a slip occurs at time l in observation j (cf. eq. 3.72):

$$E\{\underline{y}_l\} = A_l x_l + C_l z + c_j \nabla \; ; \; D\{\underline{y}_l\} = R_l \; . \tag{3.94}$$

After adaptation at time k one obtains the following set of observation equations (cf. eqs. 3.89 and 3.92):

$$E\{\begin{pmatrix} \hat{\underline{x}}^{\mathrm{a}}_{k|k} \\ \hat{\underline{z}}^{\mathrm{a}}_{k|k} \\ \hat{\underline{\nabla}}^{l,k} \\ \underline{d}_{k+1} \\ \underline{y}_{k+1} \\ \vdots \end{pmatrix}\} = \left(\begin{array}{ccc|cc} I & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & I & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -\Phi_{k+1,k} & I & \cdots & 0 & 0 \\ 0 & A_{k+1} & \cdots & C_{k+1} & c_j \\ \vdots & \vdots & \ddots & \vdots & \vdots \end{array}\right) \left(\begin{array}{c} x_k \\ x_{k+1} \\ \vdots \\ \hline z \\ \nabla \end{array}\right) \; ; \tag{3.95}$$

$$\begin{pmatrix} P_x & P_{x,z} & P_{x,\nabla} & 0 & 0 & \cdots \\ P_{z,x} & P_z & P_{z,\nabla} & 0 & 0 & \cdots \\ P_{\nabla,x} & P_{\nabla,z} & P_\nabla & 0 & 0 & \cdots \\ 0 & 0 & 0 & Q_{k+1} & 0 & \cdots \\ 0 & 0 & 0 & 0 & R_{k+1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} ,$$

where P_x, $P_{x,z}$, $P_{x,\nabla}$ etc. are shorthand notations for $P_{\hat{x}^{\mathrm{a}}_{k|k}}$, $P_{\hat{x}^{\mathrm{a}}_{k|k},\hat{z}^{\mathrm{a}}_{k|k}}$, $P_{\hat{x}^{\mathrm{a}}_{k|k},\hat{\nabla}}$, and with (cf. eq. 3.86)

$$\begin{pmatrix} \hat{\underline{x}}^{\mathrm{a}}_{k|k} \\ \hat{\underline{z}}^{\mathrm{a}}_{k|k} \end{pmatrix} = \begin{pmatrix} \hat{\underline{x}}^{0}_{k|k} \\ \hat{\underline{z}}^{0}_{k|k} \end{pmatrix} - \begin{pmatrix} \overline{X}^{x}_{k,l} \\ \overline{X}^{z}_{k,l} \end{pmatrix} \hat{\underline{\nabla}}^{l,k} \; . \tag{3.96}$$

After adaptation for the slip at time k we are (considering model (3.93)) not so much interested in the adapted bias estimate $\hat{z}^{\mathrm{a}}_{k|k}$ as well as in the 'new' bias state which is corrected for the slip and which can be obtained by the following full rank transformation:

$$\begin{pmatrix} \overline{z} \\ \nabla \end{pmatrix} = \begin{pmatrix} I & c_j \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z \\ \nabla \end{pmatrix} \; . \tag{3.97}$$

Inserting the inverse transform

$$\begin{pmatrix} z \\ \nabla \end{pmatrix} = \begin{pmatrix} I & -c_j \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \overline{z} \\ \nabla \end{pmatrix} \tag{3.98}$$

(which does not affect the covariance matrix of the observables) into the observation equations (3.95) yields:

$$E\{\begin{pmatrix} \underline{\hat{x}}^{\mathrm{a}}_{k|k} \\ \underline{\hat{z}}^{\mathrm{a}}_{k|k} \\ \underline{\hat{\nabla}}^{l,k} \\ \underline{d}_{k+1} \\ \underline{y}_{k+1} \\ \vdots \end{pmatrix}\} = \left(\begin{array}{ccc|cc} I & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & I & -c_j \\ 0 & 0 & \cdots & 0 & 1 \\ -\Phi_{k+1,k} & I & \cdots & 0 & 0 \\ 0 & A_{k+1} & \cdots & C_{k+1} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{array}\right) \left(\begin{array}{c} \boldsymbol{x}_k \\ \boldsymbol{x}_{k+1} \\ \vdots \\ \hline \\ \overline{z} \\ \nabla \end{array}\right) . \tag{3.99}$$

The parameter ∇ now only appears in the observation equations of the observables $\underline{\hat{z}}^{\mathrm{a}}_{k|k}$ and $\underline{\hat{\nabla}}^{l,k}$. Performing the full-rank transformation

$$E\{\begin{pmatrix} \underline{\hat{z}}^{\mathrm{a}}_{k|k} \\ \underline{\hat{\nabla}}^{l,k} \end{pmatrix}\} = \begin{pmatrix} I & c_j \\ 0 & 1 \end{pmatrix} E\{\begin{pmatrix} \underline{\hat{z}}^{\mathrm{a}}_{k|k} \\ \underline{\hat{\nabla}}^{l,k} \end{pmatrix}\} \tag{3.100}$$

finally reduces the observation equations (3.95) to:

$$E\{\begin{pmatrix} \underline{\hat{x}}^{\mathrm{a}}_{k|k} \\ \underline{\hat{z}}^{\mathrm{a}}_{k|k} \\ \underline{\hat{\nabla}}^{l,k} \\ \underline{d}_{k+1} \\ \underline{y}_{k+1} \\ \vdots \end{pmatrix}\} = \left(\begin{array}{ccc|cc} I & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & I & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -\Phi_{k+1,k} & I & \cdots & 0 & 0 \\ 0 & A_{k+1} & \cdots & C_{k+1} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{array}\right) \left(\begin{array}{c} \boldsymbol{x}_k \\ \boldsymbol{x}_{k+1} \\ \vdots \\ \hline \\ \overline{z} \\ \nabla \end{array}\right) ; \tag{3.101}$$

$$\begin{pmatrix} P_x & P_{x,z} + P_{x,\nabla} c_j^{\mathrm{T}} & P_{x,\nabla} & 0 & 0 & \cdots \\ P_{z,x} + c_j P_{\nabla,x} & P_z + P_{z,\nabla} c_j^{\mathrm{T}} + c_j P_{\nabla,z} + c_j P_\nabla c_j^{\mathrm{T}} & P_{z,\nabla} + c_j P_\nabla & 0 & 0 & \cdots \\ P_{\nabla,x} & P_{\nabla,z} + P_\nabla c_j^{\mathrm{T}} & P_\nabla & 0 & 0 & \cdots \\ 0 & 0 & 0 & Q_{k+1} & 0 & \cdots \\ 0 & 0 & 0 & 0 & R_{k+1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

where the state vector element $\overline{z}$ contains the original bias and the contribution of the slip. In model (3.101) $\underline{\hat{\nabla}}^{l,k}$ has become a free y^R-variate (corresponding to the situation of the adaptation for outliers) and thus has no impact on the state estimators $\underline{\hat{x}}_{i|i}$ and $\underline{\hat{z}}_{i|i}$ for $i > k$. We can thus revert to the filter model under H_0 after a single adaptation step. As we proceed with the filter procedure with the 'new' bias vector $\overline{z}$, care has to be taken that the adaptation step is executed correctly. Starting from (3.96) and applying the transformation (3.100) yields:

$$\begin{pmatrix} \underline{\hat{x}}^{\mathrm{a}}_{k|k} \\ \underline{\hat{z}}^{\mathrm{a}}_{k|k} \end{pmatrix} = \begin{pmatrix} \underline{\hat{x}}^{0}_{k|k} \\ \underline{\hat{z}}^{0}_{k|k} \end{pmatrix} - \begin{pmatrix} \overline{X}^{x}_{k,l} \\ \overline{X}^{z}_{k,l} - c_j \end{pmatrix} \underline{\hat{\nabla}}^{l,k} . \tag{3.102}$$

The covariance matrix of the adapted state follows from (3.102) as

$$\begin{pmatrix} P_x^a & P_{x,\bar{z}}^a \\ P_{\bar{z},x}^a & P_{\bar{z}}^a \end{pmatrix}_{k|k} = \begin{pmatrix} P_x^0 & P_{x,z}^0 \\ P_{z,x}^0 & P_z^0 \end{pmatrix}_{k|k} + \begin{pmatrix} \overline{X}_{k,l}^x \\ \overline{X}_{k,l}^z - c_j \end{pmatrix} Q_{\hat{\nabla}^{l,k}} \begin{pmatrix} \overline{X}_{k,l}^{x^{\mathrm{T}}} \\ (\overline{X}_{k,l}^z - c_j)^{\mathrm{T}} \end{pmatrix}^{\mathrm{T}}. \tag{3.103}$$

Summarizing, we have shown that for the particular case of the partially constant state space model and slips modelled according to (3.94), one can revert to the filter under H_0 after a single adaptation step has been performed. The practical importance of the PCSS model is its applicability to GPS, where the measured phase observables are available up to a constant, unknown bias (called ambiguity). Furthermore GPS phase observables may be contaminated by slips (called cycle slips) due to receiver tracking errors, atmospheric disturbances, and signal interruptions.

3.6.6 Adaptation for Slips; A Suboptimal Solution

We have shown that it is impossible to revert to the filter under H_0 in case of slips in the observables (except in the important case of the partially constant state space model). If one nevertheless wants to continue the operation of the filter under the null hypothesis, one has to resort to suboptimal filter solutions. For practical applications the following ad-hoc strategy can be useful. After the identification of a slip at time k one takes $\underline{\hat{x}}_{k|k}^a$ (which at time k is unbiased) with its covariance matrix $P_{k|k}^a$, as the new initial state and proceeds with filtering under the null hypothesis. This method neglects the correlation between $\underline{\hat{x}}_{k|k}^a$ and $\underline{\hat{\nabla}}^{l,k}$ and is based on the assumption that the bias is estimated well enough at time k. To prevent the accumulation of bias in the state estimator one has to correct the observables from time k onwards. For a slip in the state vector or a slip in an observation channel the observables with their covariance matrices have to be corrected as follows:

	slip in state (cf. eq. 3.70)		slip in observations (cf. eq. 3.72)	
under H_0	$\underline{d}$	Q	$\underline{y}$	R
after adaptation	$\underline{d} - c_x\underline{\hat{\nabla}}^{l,k}$	$Q + c_x Q_{\hat{\nabla}^{l,k}} c_x^{\mathrm{T}}$	$\underline{y} - c_y\underline{\hat{\nabla}}^{l,k}$	$R + c_y Q_{\hat{\nabla}^{l,k}} c_y^{\mathrm{T}}$

(3.104)

Note that the sample values of $\underline{d} - c_x\underline{\hat{\nabla}}^{l,k}$ are not zero.

3.6.7 Practical Considerations

In this section we consider a number of practical aspects related to the adaptation procedure and we will briefly discuss the computational load of the (optimal) adaptation procedure and the biasedness of the test statistics after adaptation.

Computational Load

The additional computational load caused by the adaptation procedure is rather limited. The close relationship between the test statistics and the error estimates (cf. eqs. 3.68 and 3.76 and note that the terms in the numerator and denominator are identical) makes the latter ones available at a very low computational cost. The response matrix of the model error on the filtered state is computed using (3.84), and is closely related to the response matrix for the predicted state, which is computed recursively in the identification procedure.

If one uses the optimal adaptation procedure as given by (3.86) and (3.87), one actually switches to a filter operating under the alternative hypothesis using the recursive bias estimation scheme given by (3.80) to (3.82). A drawback of the optimal method is that the memory of the bias filter grows while the window length of the tests remains fixed. Furthermore a separate bias filter has to be implemented parallel to the Kalman filter, and only one model error at a time can be conveniently handled. Compared to the exact procedure the approximate adaptation procedure for slips (described in Section 3.6.6) is computationally very attractive, although one has to take care that at every time update corrections to the disturbances are applied and that likewise the measurements are corrected at measurement updates. Overall the computational requirements of the adaptation procedure do not seem to be prohibitive.

Biasedness of Test Statistics

After adaptation one has to remove the bias from the test statistics (or the predicted residuals on which they are based), in order to allow the identification strategy to proceed automatically. Those test statistics that are (partly) based on the predicted residuals in the time interval l to k will be biased. After the adaptation step the identification test statistics can be adjusted for the model error as follows:

$$t^{l,k}_{\text{adjusted},j} = \frac{t^{l,k}_j - \rho_{ji} t^{l,k}_i}{(1-\rho^2_{ji})^{1/2}} \tag{3.105}$$

where ρ_{ji} (cf. Chapter 4) is the correlation coefficient between $t^{l,k}_i$ (the test statistic associated with the most likely H_A) and each other test statistic denoted symbolically as $t^{l,k}_j$. The correlation coefficient ρ_{ji} cannot be easily computed in a recursive manner and hence (3.105) is of limited practical use if the time of occurence l and the time of detection k do not coincide for both hypotheses. We therefore suggest to reinitialize the testing procedure after adaptation with local tests (i.e. tests with a window length of one). This strategy backfires if model errors occur (nearly) simultaneously and should thus be implemented with care. The major drawback of the strategy is that it is rather heuristic than theoretically sound. From (3.105) it is apparent that it is impossible to devise a testing procedure which can cope with slips and outliers for a single observation type simultaneously using local tests only; in that case the absolute value of the correlation coefficient is always equal to one, that is the outlier and slip

hypothesis cannot be separated. If only local tests are implemented, one can cope with simultaneous errors using an iterated approach (cf. [TEUNISSEN, 1990b]) which takes the correlation between the test statistics into account.

3.6.8 Alternative Approaches to Adaptation

We have discussed solutions (some exact, some suboptimal) to the adaptation problem. In addition to the discussed procedures some other adaptation schemes might be considered, which only require a single additional filter at most. We will briefly discuss two approaches which will not be pursued any further, but are included for the sake of completeness.

Parallel (lagged) Kalman Filter

If the main objective of the filter procedure is to obtain a dataset free of errors in the observations, one can follow a simple approach. Assume one can ascertain that all likely model errors can be identified and removed with a delay $N-1$ (which corresponds to a window length of N). Then a parallel filter operating with a delay N, using information provided by the real-time filter and operating on the original (temporarily stored) data, leads to optimal results. The real-time filter has to operate in conjunction with detection and identification tests and the data in the time interval $[k-N+1, k]$ have to be stored. The method seems particularly suited if only model errors in the observations are expected and no real-time solutions are required. The lagged filter can operate without a full detection and identification test procedure. The main advantage of the lagged filter is its simplicity. A possible application of the method would be the elimination of outliers from the original dataset. Drawbacks are that no real-time estimates are available and errors in the dynamic model cannot be handled easily (a jump in the state vector, for example, cannot be resolved by deleting an observation from the dataset).

Backward Filtering

Instead of opting to use a lagged filter, one can also implement an algorithm that filters 'backwards' once a model error has been identified, removes the model error, and filters 'forward' to obtain a better real-time solution. Although this method is theoretically feasible, the computational load of the backward filter strategy cannot be predicted and hence the method is deemed unpractical for real-time applications.

3.6.9 Summary

In this section we have developed a real-time adaptation procedure, which maintains the optimality of the estimators. Computational considerations ruled out the implementation of bank of Kalman filters, each one taylored for a particular hypothesis. The adaptation procedure given by (3.86) and (3.87) is optimal for both outliers and slips

(we conciously neglected the bias in the state estimators in the time interval l to $k-1$). For outliers it was shown that one can revert to filtering under the null hypothesis after adaptation, while maintaining the optimality of the state estimators. For slips such an approach is only feasible for the partially constant state space model. In all other cases of the adaptation for slips reverting to filters under the null hypothesis will result in suboptimal solutions. Practical considerations showed that the adaptation strategies, which are based on resetting the window lengths of the tests, will be hampered if model errors occur (nearly) simultaneously. In case the model errors occur separated in time it is to be expected the proposed procedures operate well.

3.7 The B-method of Testing

This section is meant to provide a link with the testing procedure for geodetic networks developed by BAARDA [1968]. This so-called B-method has been used very successfully in network applications and therefore one should consider if it can also be implemented for dynamic systems.

The essential element of the B-method of testing is that an error related to a particular alternative hypothesis should be detected with equal probability by all tests, which encompass that particular alternative hypothesis. This means that the non-centrality parameter λ is equal for all tests. BAARDA suggested to fix λ by specifying the level of significance α and the power of the test γ for one-dimensional tests. One keeps the power of the test fixed at γ_0 for all tests, because one is interested in the model error that can be detected with a probability γ_0. This line of thought implies that the level of significance α of a multi-dimensional test is computed from the level of significance α_0 of the one-dimensional tests. If one fixes the level of significance of the one-dimensional test, the level of significance of the multi-dimensional test can be computed from the inverted power function:

$$\lambda = \lambda(\alpha_0, \gamma_0, 1) = \lambda(\alpha, \gamma_0, b) \ , \tag{3.106}$$

where b is the dimension of the multi-dimensional test.

Up to this point we tacitly assumed that the DIA procedure is based on the B-method. It has been indicated at the end of Section 3.4 that the level of significance of the overall model tests is based on the level of significance of the one-dimensional tests via the B-method of testing, and thus an error with a magnitude related to the noncentrality parameter $\lambda_0 = \lambda(\alpha_0, \gamma_0, 1)$ is detected with equal probability (namely γ_0) by all tests. The size of a model error associated with λ_0 is called minimal detectable bias and is a measure of the detectability of a one-dimensional hypothesis. The minimal detectable bias is further discussed in Chapter 4.

Following the reasoning of the B-method we assume that identification is only performed after a model error has been detected. In the implementation of the DIA procedure one has to take care that the method is implemented correctly. If, for example, the identification of a model error with a delay of four coincides with the detection

of some model error with a delay of two, the detection test does not encompass the identification test. In practice such a mismatch will most likely occur for model errors that are about the size of the minimal detectable bias.

Theoretically the coupling given by (3.106) is strictly valid for model errors of the size of the minimal detectable biases. If one, for instance, considers model errors considerably larger than the minimal detectable biases, the probability of identification of the model error is larger than the probability of detection, but because large model errors are always detected, the B-method is still valid. The point we want to stress is that the coupling of the one- and multi-dimensional tests is not generally valid and should, depending on the application at hand, not be applied blindfoldedly. These remarks do not limit the usefulness of the B-method in the DIA procedure, and experience has shown that in general it works well. In practice difficulties due the direct link between detection and identification can be expected if the model errors are smaller than the minimal detectable biases (but if one wants to detect such errors the testing procedure should be reconsidered anyway). Also in cases where a certain model error is incorporated in, for example, one-, two-, and, three-dimensional hypotheses the comparison of the various identification test statistics might be difficult.

A more serious problem is that the B-method of testing is based on the assumption that the level of significance of all one-dimensional tests is identical. The actual choice of the level of significance is discussed in Section 3.7.1. If the level of significance is not identical, we suggest to base the B-method on the largest level of significance of all one-dimensional hypotheses considered. This may result in a large level of significance of the overall model tests, but we think it is better to use the B-method in a conservative manner and rather allow false alarms than missed detections.

Despite its limitations we think that the use of the B-method for testing procedures in dynamic systems can be justified. The DIA procedure enables the realization of an automatic model validation technique and any implementation of an automatic strategy requires a decision mechanism. We think the B-method is a useful tool to implement such a decision mechanism in the transition of the detection to identification phase of the DIA procedure. To what extent the B-method provides the ultimate solution cannot be determined yet and should be based on further experiments.

3.7.1 Choice of the Level of Significance

The B-method assumes that the level of significance α_0 for all one-dimensional tests is equal. Such an assumption can readily be made for testing procedures in geodetic network adjustments. Conventional control networks are usually designed in such a way that the direction and distance measurements have an approximately equivalent contribution to the final network solution. In GPS networks baseline components are needed in all three (cartesian) coordinate directions. Therefore there is no reason to test the separate components with different levels of significance.

The sensor suite of an integrated navigation system, on the other hand, may consist of a range of different sensor types. Furthermore it is likely that different types of

model errors are specified (for example slips and outliers). The magnitude of the level of significance for each test is of course application dependent and is not necessarily identical for all tests. A vehicle location system, for example, is generally based on a digital map (which can be considered as a sensor type with a low measurement rate and a low suspectibility to model errors) and a magnetic compass (a sensor with a high measurement rate and high error rate). For this particular case it does not seem proper to test a single reading of both 'sensors' with the same level of significance. Generally sensors with high measurement rates allow one to discard measurements more easily (which amounts to testing with a large level of significance (e.g. $\alpha_0 = 0.05$ instead of the $\alpha_0 = 0.001$ used in control network adjustments). The actual choice of the level of significance and power of the tests should take the characteristics of the sensors and the impact of undetected errors into account. The testing parameters should be chosen based on a trade-off between the costs of missed detection and false alarm.

3.8 Recapitulation of the DIA Procedure

To facilitate further reading we summarize the findings of the previous sections. One of the most difficult tasks in designing a testing procedure is the specification of likely model misspecifications. The choice of alternative hypotheses is *application dependent.* One has to ascertain if the likely model misspecifications can really be modelled as additive effects (slips) in the functional model. If, for example, the stochastic model is specified incorrectly, one has to recourse to adaptive filtering techniques. Even if the model misspecifications can be modelled as slips, one has to decide which alternative hypotheses have to be specified. Also the choice of the level of significance and the power of the test depends on the application. The performance of the testing procedure can be analysed using the *design* procedure discussed in the next chapter, and this procedure should be used to determine the window lengths and testing parameters of the tests. We assume that in the design phase the testing parameters, the window lengths of the tests and the alternative hypotheses have been specified. The DIA procedure is summarized in Table 3.1.

We assume that the detection and identification steps are coupled via the B-method of testing, which was discussed in Section 3.7. If the frequent detection of model errors does not coincide with subsequent identifications, one should seriously consider other types of mismodelling, such as an incorrectly specified stochastic model or an under-parametrized dynamic model. The estimation of the model error can be considered as a part of either the identification or the adaptation procedure. After adaptation for outliers one can revert to the filter under H_0. Adaptation for slips can be implemented using the exact procedure given by (3.86) and (3.87), or by an approximate, sub-optimal procedure (cf. eq. 3.104).

	Local Testing		Global Testing
DETECTION	Local Overall Model (LOM) Test(-statistic) (3.59) (3.60)	Test statistic Test	Global Overall Model (GOM) Test(-statistic) (3.61) (3.63)
IDENTIFICATION	Local Slippage (LS) Test(-statistic) (3.66)	Test statistic	Global Slippage (GS) Test(-statistic) (3.68)
	Determine most likely model error via the identification procedure (3.75)		
	Estimation of model error (3.76) or (3.80)		
ADAPTATION	Adaptation procedure (3.86) and (3.87)		

Table 3.1: Overview of the DIA procedure

3.9 Other Techniques for Model Validation

The purpose of this section is to provide a link between the DIA-methodology presented in this chapter and other model validation techniques. The DIA procedure has been developed along the line of hypothesis testing in linear models (cf., e.g., [KOCH, 1988]). It has been shown that the alternative hypotheses can be specified in terms of predicted residuals and that all test statistics can be computed recursively.

Since the early seventies numerous investigations have been performed under the headings 'Failure Detection Identification and Recovery' (FDIR) and 'Robust Kalman Filtering'. The presentation in this section is not meant to be (and cannot be) exhaustive as the literature abounds with model validation techniques. It should, however, provide the reader an entry to the relevant literature. General overviews of model validation techniques are given by, e.g., WILLSKY [1976] and BASSEVILLE [1988]. We will limit ourselves to methods that deal with model errors of additive nature (slips) and techniques that are closely related from an algorithmic point of view. We will briefly discuss the use of the predicted residual in model validation techniques, the GLR approach, a multiple model approach based on parallel filters, the impact of decentralized models on model validation, and 'robust' Kalman filtering.

3.9.1 The Use of the Predicted Residual

The predicted residual naturally presents itself as a tool for model validation of Kalman filters due to its well defined statistical properties under the null hypothesis. The local and global detection tests ((3.60) and (3.63) respectively) are functions of the predicted residuals, and are frequently encountered in the literature on FDIR.

The use of predicted residuals for performance analysis of Kalman filters was introduced by MEHRA AND PESCHON [1971]. Their methodology focusses on the overall performance analysis of the Kalman filter and basically consists of a monitoring procedure for the predicted residuals.

3.9.2 The GLR Approach

The GLR (Generalized Likelihood Ratio) approach is a systematic methodology to detect, identify, and adapt for slip-type model errors and is closely related to the DIA procedure. The GLR approach was developed in the seventies by a group of researchers at MIT [WILLSKY, 1976; WILLSKY AND JONES, 1976; CHANG AND DUNN, 1979]. The algorithm of the DIA procedure itself is almost identical to the (recursive) Failure Detection Identification and Recovery (FDIR) methods based on the Generalized Likelihood Ratio approach derived by the MIT-group. WILLSKY [1976] notes that "GLR performance [is] quite outstanding for failures that can be modelled as additive effects." The aforementioned researchers also provide a link between the error (or bias) estimation procedure and the separate bias estimation concept of FRIEDLAND [1968], which was further elaborated by IGNAGNI [1981;1990]. Also FRIEDLAND [1983] acknowledges the link of the bias estimation procedures with the FDIR methods. Although the DIA algorithm is to a large extent identical to that of the FDIR GLR procedure, there exist a number of important differences. Firstly, the GLR method was never extended with a comprehensive design procedure (although some preliminary results on the detectability and separability of alternative hypotheses are provided by [BUENO ET AL, 1976]). Secondly, the detection and identification steps are not coupled and hence detection and identification can occur independently. Finally, the adaptation (recovery) step is not really well documented.

3.9.3 Multiple Model Approach

The multiple model approach (see, e.g., [WILLSKY, 1986; BROWN AND HWANG, 1987]) is based on the assumption that for every alternative hypothesis a Kalman filter solution is computed. At every filter cycle one then chooses the most likely filter solution. One thus operates a bank of Kalman filters. Such an approach can be very useful if one, for example, operates one filter under the null hypothesis of a constant velocity dynamic model and one filter under the alternative hypothesis of a constant acceleration dynamic model. As soon as an acceleration is detected one can switch to the filter operating under the alternative hypothesis. An advantage of the multiple model approach is that no model degradation occurs if one switches from one filter mode to the other. A possible disadvantage is the heavy computational load, especially if one has to check for many model misspecifications in the observation model.

3.9.4 Decentralized Quality Control

The present testing procedure is based on the premise that the data processing is done by means of a single Kalman filter. If a navigation system consists of many subsystems it may be advantageous (or even necessary) to process the data of each sensor separately and to obtain an overall solution by merging the results of each subset. One then uses the concept of 'federated' or 'decentralized' filters, which was developed by CARLSON [1988; 1990]. In a decentralized setup it is possible to perform

so-called local quality control (i.e. one implements a DIA procedure for each separate (or local) filter). The advantage of such an approach is that model misspecifications are limited to the local filter model and hence smearing effects are limited. A disadvantage is that the redundancy of the local filters is generally quite low. Decentralized quality control is discussed by KERR [1987] and LOOMIS ET AL. [1988]. A theoretical problem arises if one wants to implement a DIA procedure for the overall filter which merges the output of the local filters, because one should take into account the results of the local testing procedures. The situation is similar to the testing procedure in a second phase network adjustment, when the original observations are not available anymore.

3.9.5 Robust Filtering

Some authors (e.g. BORUTTA [1988]) have investigated the use of robust estimation techniques for geodetic applications. An introduction to robust estimation techniques is given by HAMPEL ET AL., [1986]. One could also consider to 'robustify' the Kalman filter for model misspecifications. Investigations dealing with 'robust' Kalman filters are reported by, among others, MASRELIEZ AND MARTIN [1977] and PEÑA AND GUTTMAN [1988, 1989]. Robust filter algorithms have been derived primarily for additive model errors in the observables (which usually are modelled by assuming a mixed distribution for the observables using a so-called variance inflation model). We do not consider robust estimation, but a Kalman filter combined with the DIA procedure can also be considered a 'robust' filter. To illustrate the performance of a robust filter and a DIA supported Kalman filter, we compare results obtained by the DIA approach with results obtained by the robust filter developed by PEÑA AND GUTTMAN [1988]. In [ibid.] a one-dimensional state space model with a scalar observation equation $E\{\underline{y}_k\} = x_k$ for 31 epochs and two simulated outliers is considered. In Fig. 3.6 we have reproduced (on the left) the true state used for the simulation (solid line), the simulated data (dots) and the estimates obtained by the standard Kalman filter (dashed line). On the right we have depicted the true state (solid line), the robust filter estimate (dotted line), and the filter estimate after adaptation for outliers (dashed line). All filters are based on the system model given in [ibid.]. Although the results for this single case presented in Fig. 3.6 by no means constitute the equivalence of the robust and DIA based filter, it can be seen that the performance of both methods is comparable.

3.9.6 Discussion

In the geodetic and navigation practice model errors can often be modelled as slips in the functional model. We have developed a procedure to cope with such errors *automatically*. The recursive DIA method seems to be very suited for (almost) real-time model validation and adaptation purposes. It is based on a single Kalman filter with a parallel operating DIA procedure. Its low computational cost and its sound theoretical base render the DIA procedure a very attractive model validation technique for slip-type model errors.

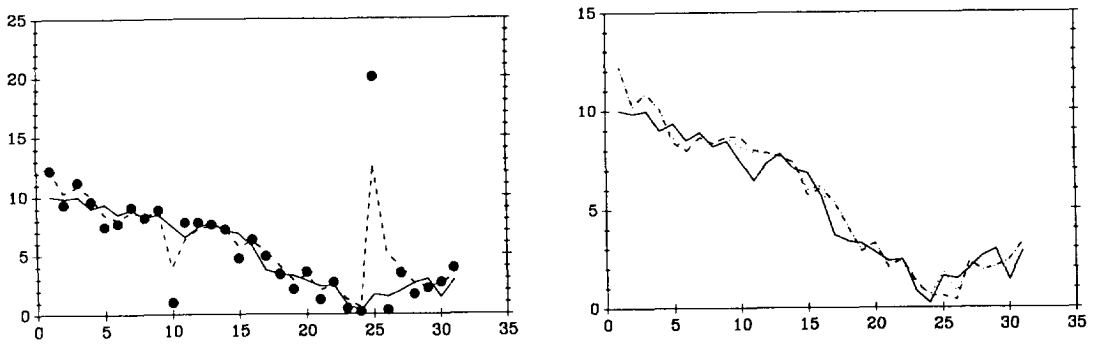


Figure 3.6: Comparison of the estimates of the Kalman filter, the robust filter, and the DIA supported Kalman filter for the example in PEÑA AND GUTTMAN [1988].

Of the alternative methods the GLR approach is the most closely related to the DIA procedure. A major difference is that the DIA procedure is based on a *unified* procedure in which detection, identification, and adaptation are coupled. Particularly the link between detection and identification by means of the B-method has never been applied (to our knowledge) in testing procedures for dynamic systems. Furthermore we think that the implementation of any model validation technique should be preceded by an extensive *design procedure*. As such the design procedure (discussed in Chapter 4) should be considered to be part of the DIA methodology.

3.10 Concluding Remarks

The optimal properties of the filter estimators can only be guaranteed in real-time if a testing procedure is operating synchronously with the filter. We have derived such a testing procedure based on the theory of hypothesis testing, and we demonstrated some optimality properties of the generalized likelihood ratio tests on which the procedure is based. The testing procedure is optimized for model misspecifications that can be modelled as additive effects (or slips), and allows local and global testing. Local tests are genuinly real-time and are based on a single epoch, whereas global tests operate with a (small) delay, but cover a number of epochs. The testing procedure consists of three steps, namely detection, identification, and adaptation (DIA). The DIA procedure can be implemented recursively and is thus very efficient. In the identification step not only the type of model misspecification, but also the time of its occurence is determined. We suggested to couple the detection and identification step of the procedure by means of the B-method, but the implementation of this coupling will require some further investigations. The choice of the testing parameters for each specific alternative hypothesis will have to be established partly by measurement experiments, and will depend

on, among other things, the costs associated with type one and type two errors and the error rate of the sensor suite in question. After a model error has been identified, the state has to be adapted to maintain the optimality of the real-time filter results. Adaptation for outliers is very simple and requires only a single adaptation step, after which one can revert to the filter under the null hypothesis. Exact adaptation for slips, on the other hand, generally cannot be implemented that easily, because it requires continuous adaptation after identification. Approximate adaptation schemes for slips can be devised but their suboptimality has to be established in a design study for each particular case. A drawback of the current implementation of the DIA procedure is that during adaptation the detection and identification test statistics are biased (due to the model error), and therefore an efficient methodology to remove the bias in the predicted residuals has to be found. Despite this drawback the DIA procedure is a very efficient testing procedure for slip-type model errors and will be used throughout this report. The DIA procedure is summarized in Section 3.8.

In the next chapter we will consider (based on the Kalman filter and the DIA procedure) how the design of a dynamic system can be optimized. An important design aspect will be reliability, which describes the sensitivity of the estimation result to model misspecifications. We will see that the reliability of a system depends on the testing procedure that is implemented.

Chapter 4

Design of Dynamic Systems

4.1 Introduction

In the present chapter we try to develop a methodology for the optimization of dynamic systems. With optimization of a dynamic system is meant the *design* of a dynamic system subject to quality criteria determined by the purpose of the system. Optimization is part of the quality assurance of a dynamic system. After the design of a system has been completed, it has to be ascertained that the performance of the system under operational conditions is in accordance with the description of the quality of the system design. Real-time model validation or quality control is implemented by means of the DIA-procedure, which was discussed in the previous chapter. Finally the quality of the estimation result should be compared with the quality requirements. Quality assurance thus encompasses the steps of design, control, and validation with respect to the quality of the system, which are respectively related to the a priori, real-time, and a posteriori phases of the operation of a dynamic system. In this chapter we will focus on the design of dynamic systems and especially integrated navigation systems.

The concept of quality comprises precision and reliability. In the design phase of a system the precision and reliability requirements have to be reconciled with limiting conditions such as cost, available hardware, computer power, personnel, and time schedules. Our optimization procedure is limited to aspects of precision and reliability, because the other design aspects are too application dependent to be put in a general, practical framework.

Traditionally the performance of dynamic systems (and of integrated navigation systems in particular) is specified in terms of *precision*: how accurately can certain parameters (e.g. position, velocity) be estimated. The currently emerging demand for (real-time) quality control necessitates that also reliability is taken into account, i.e. the effect that possible model misspecifications have on the estimation results. The concept of reliability is closely related to the testing procedure that is implemented in the dynamic system, and has hitherto received little consideration. The measures of precision and reliability are mathematically tractable and are therefore very useful to

judge the performance of any system. The quality measures are independent of actual measurements and thus the quality of a dynamic system can indeed be analysed in the *design phase.*

The design procedure presented here can be considered as a generalization of the optimization technique for geodetic networks as has been developed by BAARDA [1968, 1973, 1977]. A first step towards a design procedure for dynamic systems was presented by TEUNISSEN [1990a, 1990b].

4.1.1 Overview of this Chapter

In Section 4.2 we will discuss a general framework for the design of dynamic systems and the concept of quality assurance. We will pay attention to quality criteria, quality measures, and design parameters of dynamic systems. The measures of precision and reliability are considered in Sections 4.3 and 4.4 respectively. The findings of Section 4.4 are directly related to the DIA testing procedure. In Section 4.5 we tentatively propose a design procedure for integrated navigation systems. In Section 4.6 we consider an example and compare our reliability description with a particular one found in the literature. Some concluding remarks are given in Section 4.7.

4.2 Design and Quality Assurance

In the introduction we have discussed the concept of quality assurance. In the design phase of a dynamic system one wants to assure that the results of the estimation procedure meet the preset quality requirements. We will not go into the specification of these requirements, and will assume quality requirements have been set. Once a particular design has been implemented and the system is operational, one has to assure that the quality of the estimation result is in accordance with the quality requirements set in the design phase. We will consider the following *quality criteria*:

precision What is the precision of the state estimators under the working hypothesis?

internal reliability

- **detectability** How well can certain model errors be detected and identified?
- **separability** How well can one make a distinction between different model errors?

external reliability

- **bias in state estimator** What is the impact of (undetected) model errors on the estimation results?
- **significance of bias** How significant are the biases in the state vector caused by (undetected) model errors?

The description of the precision is only valid under the null hypothesis. In Chapter 3 we have discussed the DIA procedure that is used, among other things, to validate this assumption. The reliability of the system depends on the tests that are executed with the DIA procedure (actually reliability is only defined if one tests for model misspecifications), and consequently the reliability of the system design has to be analysed, even if only precision requirements have to be met. The setup of the testing procedure should be derived from a design procedure in which the type, window length, and testing parameters of the tests are determined. The design procedure can then be regarded as an intrinsic part of the DIA procedure.

The design procedure will be based on *quality measures* for:

precision In general the precision of a dynamic system should be analysed by considering the complete covariance matrix of the state estimator covering all epochs. In the current chapter we will limit the analysis to (elements of) the covariance matrices of the predicted and filtered state, because we are primarily interested in the real-time quality of the dynamic system. The precision of the state vector elements related to position is visualised by point standard ellipses.

internal reliability Internal reliability is analysed using the so-called Minimal Detectable Biases (MDBs). The MDBs can be used as a measure of detectability and separability.

external reliability We analyse (functions of) biases in the filtered state estimator caused by models errors of the size of the MDBs. The direct analysis of the biases is quite laborious and therefore we primarily consider the significance of the bias in the state vector. A measure of significance is the Bias to Noise Ratio (BNR).

The assessment of the overall quality of a particular design is dependent on all quality criteria. Only if all quality criteria are taken into account simultaneously the quality of the estimators of a dynamic system can be assured. Depending on the application at hand the designer may decide that a certain quality criterium outweighs the others. If the quality requirements cannot be met, one can consider the following *design options* (or design parameters), that consist of changes in the:

functional model The functional model consists of the measurement model and the dynamic model. The measurement model may be changed by increasing the number of observations, introducing other types of measurement systems, changing measurement sampling rates, or by modification of the geometry in the measurement setup. The actual dynamics (or kinematics) of the system underlie the dynamic model, and hence this model cannot be changed at will. It seems natural, however, to choose the dynamic model as simple as possible on condition that the dynamics are properly described.

stochastic model The stochastic model is given by the covariance matrices of the observations and disturbances. Changes in the stochastic model of the observations

are directly related to changes in the precision of the measurements (by, for example, replacing a system with a similar system of a better precision). Furthermore a change in the sensor suite will result in a change of the stochastic model of the observations. The stochastic model of the disturbances strongly depends on the sophistication of the dynamic model and the (unmodelled) dynamics which the system is subjected to.

testing procedure In the design phase the parameters of the testing procedure only have an impact on the reliability description of the system. The testing parameters are the type of model errors one considers, the window lengths of the tests, and the size and power of the tests.

Changes in the functional and stochastic model have an impact on both the precision and reliability description of the system. It may be obvious that in practice certain model parameters cannot be changed (e.g. due to the limited availability of measurement systems). In the optimization procedure one should aim at a system design which meets, *but not exceeds*, the quality requirements. Furthermore the conclusion of the optimization procedure may be that it is impossible to design a system which meets the quality requirements under the given constraints.

The design procedure should finally provide the user with a *description* of the:

system model A description of the functional and stochastic model which meets the quality requirements.

quality of the system design A quantification of the precision and (internal and external) reliability of the system design.

testing procedure The parameters of the testing procedure which are to be implemented in the system, namely the type of model errors considered, the window lengths of the tests, and the size and power of the tests.

In the current context we focus on the description of the quality of the system and the design of a testing strategy. The results of the design procedure are evidently application dependent (our examples are related to precise positioning applications). The design procedure itself, however, is generally applicable.

Quality assurance is not realized by merely optimizing the design of the dynamic system. One will actually have to implement the proposed system model and equip the filter with the chosen testing strategy. During operation the quality of the system is monitored by the DIA-procedure. After the operation of the system has been completed one has to characterize the quality of the estimation result and has to demonstrate that the preset quality requirements have been met. Only if all steps have been completed the quality of the estimation result is assured. We assume the quality of the estimation result is described by the same quality measures which are used for the design of the system. As a consequence we will make no further distinction between the a priori and a posteriori quality assessment, and will only consider the design of dynamic systems.

It is very important to realize, however, that the quality of the design is *independent* of actual data, whereas the quality of the estimation result is *not.*

4.3 Precision

The best known and most widely used quality criterion in navigation is doubtless precision. The performance of navigation systems is at least always specified in terms of precision by their manufacturers. Generally the integrated navigation system has to meet certain well defined precision requirements imposed by the application at hand. The Kalman filter automatically provides the following (local) measures of precision, viz:

$$\begin{aligned} &\text{covariance matrix of the estimator of the predicted state at time } k\text{:} \quad P_{k|k-1} \\ &\text{covariance matrix of the estimator of the filtered state at time } k\text{:} \quad P_{k|k}\ . \end{aligned} \tag{4.1}$$

One can thus readily derive the precision (of functions) of estimators of the state vector. The covariance matrices (4.1) describe the random nature of the state estimators *under the null hypothesis.*

Precision is commonly described by the following measures:

- *Standard deviation* of elements of the state vector. If one considers the ith element of the filtered state vector, its standard deviation is given as $\sigma_i = \sqrt{(P_{k|k})_{ii}}$. Often one is also interested in the precision of certain linear functions of the state vector. If we consider the linear function $\hat{\underline{\theta}} = a^{\mathrm{T}}\hat{\underline{x}}_{k|k}$ of the filtered state vector (with $a \in \mathcal{R}^n$), the standard deviation of $\hat{\underline{\theta}}$ is given as $\sigma_{\hat{\theta}} = \sqrt{a^{\mathrm{T}} P_{k|k} a}$.

- *Point standard ellipse* (or ellipsoid) for the description of the precision of position for two- or three-dimensional applications. If we denote the part of the covariance matrix of the filtered state related to position as $P^{\mathrm{pos}}_{k|k}$, the direction and length of the principal axes of the point standard ellipsoid are given by the eigenvectors and square root of the eigenvalues of the matrix $P^{\mathrm{pos}}_{k|k}$ respectively.

The precision measures are dependent on the following design parameters:

- Stochastic model of the observables. It follows directly from the Kalman filter algorithm that the precision of the state estimator is a direct derivative of the covariance matrices of the observables $\underline{y}$ (R) and the disturbances $\underline{d}$ (Q), and a possible correlation between the observations and disturbances. Improving the precision of the observations and/or disturbances results in an enhanced precision of the (filtered and predicted) state estimators.

- Measurement model. The geometry of the measurement setup and the redundancy of the system have a direct impact on the precision of the state estimators. Increasing, for example, the redundancy (by, e.g., enlarging the number of observations or adding an additional measurement system) leads to a better precision.

- Dynamic model. Changes in the dynamic model (which generally also encompass changes in the stochastic model of the disturbances) have an impact on the precision, especially for the estimator of the predicted state.

It is difficult to quantify the improvement in precision caused by changes in the measurement and dynamic models.

4.3.1 Precision Criteria

The precision requirements for geodetic navigation applications are quite well documented and in Table 4.1 we provide an overview of precision requirements for some applications.

Application	Precision in RMS required for	
	position [m]	velocity [cm/s]
gravity (sea)	20	< 10
gravity (land, air)	1 (height)	< 0.01
gravity gradiometry (air)	20	< 10
relative geoid	1	10
3D-seismic (land,sea)	1-3	50
aeromagnetics	1	30
resource mapping (photogrammetry)		
1:50,000	2	100
1:20,000	0.5	25
1:10,000	0.1	5

Table 4.1: Precision requirements for geodetic and high precision navigation applications (from [SCHWARZ ET AL., 1989]).

In Table 4.1 the precision requirements are specified as RMS errors of the position and velocity states. If, for example, position in the plane is given in Easting (E) and Northing (N) coordinates, the RMS horizontal position error is defined as $\text{RMS} = \sqrt{(\sigma_E^2 + \sigma_N^2)/2}$. Because the objectives of dynamic surveys are usually well defined, the precision requirements for (geodetic) navigation applications can be clearly specified. The specification of precision requirements, however, is far from standardized. In geodetic practice usually standard deviations or criterion (covariance) matrices are used to quantify precision. In the navigation community measures such as RMS, DMRS (measures based on the covariance matrix), CEP (2-D), and SEP (3-D) (radial precision measures) are commonplace. A review of these measures and their relationship is given by MERTIKAS ET AL. [1985]. We propose to specify the precision requirements based on measures that can be derived from the covariance matrices of the state estimators such as standard deviations or point standard ellipsoids. One has to keep in mind that for the two-dimensional case the standard ellipse merely corresponds to a 39% confidence region and in the three-dimensional case the standard ellipsoid represents only a 20% confidence region.

4.3.2 Precision Testing

Given the criteria for precision we want to know if our design can meet the given requirements. It was seen that for precise positioning purposes the criteria are well known and therefore (the elements of) the covariance matrices of the state estimators can often be directly compared with the requirements. The analysis of precision in geodetic network design is largely based on the criterion matrix theory developed by BAARDA [1973]. For general purpose networks suitable criterion matrices have been developed. The variety of applications (and thus realizations) of integrated navigation systems limit the usefulness of a criterion theory for dynamic systems. We therefore propose to analyse the precision with measures that are based on the covariance matrix of the state estimator and can be directly compared with the requirements as specified in Table 4.1.

4.4 Reliability

In this section we deal with the reliability of dynamic systems. The section is meant to provide a general overview of the reliability aspects of dynamic system design. Reliability is discussed in [BAARDA, 1977; KOK, 1984] in the context of network design and for integrated navigation systems in [TEUNISSEN, 1990a]. Internal reliability describes the model misspecifications which can be detected by the statistical tests with a certain probability, that is internal reliability provides measures for the detectability of model errors. The separability between various alternative hypotheses is usually considered part of the internal reliability description as well. External reliability describes the influence of model misspecifications on the state estimators.

We begin this section on reliability with the concept of internal reliability and establish a link with the testing procedure discussed in Chapter 3. Internal reliability can be analysed by the minimal detectable bias (MDB), and we will use the MDB primarily as a measure of detectability. It will be indicated how the MDB can be used in the design procedure and how it is affected by various model and testing parameters. We then proceed with a discussion of external reliability and show how the MDBs are propagated as biases into the state estimators. A measure of significance of this bias is the bias to noise ratio (BNR) and we will show how the BNR can be used in the design procedure.

4.4.1 Internal Reliability

In general one will try to devise a testing procedure that gives a reasonable protection against type I (false alarm) and type II (missed detection) errors. Therefore one usually fixes the size and power of the test. In geodetic practice, however, one is more interested in the model error that can be detected with a certain probability than in the power of the test itself (cf. the discussion in Chapter 3). By fixing the size α and power γ of the one-dimensional test one fixes the non-centrality parameter λ_0. Assuming we use

the B-method of testing λ_0 is fixed for tests of arbitrary dimension. One can then solve the b-dimensional model error ∇ from the equation of the non-centrality parameter (cf. eq. 3.39):

$$\lambda_0 = \nabla^T C_v^T Q_v^{-1} C_v \nabla \ , \tag{4.2}$$

which can also be interpreted geometrically as (with metric Q_v^{-1}):

$$\lambda_0 = \|C_v \nabla\|^2 \ . \tag{4.3}$$

The $b \times 1$ vector ∇ is a measure of the bias one can detect with a prefixed probability γ_0, i.e. ∇ is a measure of *detectability*. The quadratic form (4.2) represents the equation for a b-dimensional (hyper-)ellipsoid. To obtain a convenient description we choose to parametrize the $b \times 1$ vector ∇ as (cf. [TEUNISSEN ET AL., 1987]):

$$\nabla = \|\nabla\| d \ , \tag{4.4}$$

where d is a $b \times 1$ unit vector. Inserting (4.4) in (4.2) it follows via

$$\lambda_0 = d^T C_v^T Q_v^{-1} C_v d \, \|\nabla\|^2 \ , \tag{4.5}$$

that

$$\nabla = \sqrt{\frac{\lambda_0}{d^T C_v^T Q_v^{-1} C_v d}} \ d \ . \tag{4.6}$$

If one lets d scan the unit (hyper-)sphere the $b \times 1$ vector ∇ describes the (hyper-)ellipsoidal boundary region of biases. Note that if the test is based on a window from time l to k, $C_v^T Q_v^{-1} C_v$ can be computed as $\sum_{i=l}^{k} C_{v_i}^T Q_{v_i}^{-1} C_{v_i}$. The biases related to the principal axes of the (hyper-)ellipsoid given by (4.2) can be computed as

$$\nabla_j = \sqrt{\frac{\lambda_0}{\lambda_j}} \ d_j \ , \tag{4.7}$$

for $j = 1, \ldots, b$, where λ_j and d_j are one of the b number of eigenvalues and associated normalized eigenvectors of the matrix $C_v^T Q_v^{-1} C_v$. The least detectable bias is connected with the smallest eigenvalue. If one considers one-dimensional hypotheses the matrix C_v reduces to a vector c_v and (4.6) reduces to

$$\nabla = \sqrt{\frac{\lambda_0}{c_v^T Q_v^{-1} c_v}} \ . \tag{4.8}$$

Up to this point we considered the non-centrality parameter of a test associated with its corresponding alternative hypothesis. It is, however, well known that the test statistics are mutually correlated and hence one can also consider the non-centrality parameter of tests which do not correspond with the actual model error. The non-centrality parameter then provides information on how model errors affect other test statistics. If we assume the true alternative hypothesis is parametrized as

$$H_A^{\text{true}} : \underline{v} \ \sim \ N(\overline{C}_v \nabla, Q_v) \ , \tag{4.9}$$

it follows from the definition of the test statistic

$$\underline{T} = \underline{v}^{\mathrm{T}} Q_v^{-1} C_v (C_v^{\mathrm{T}} Q_v^{-1} C_v)^{-1} C_v^{\mathrm{T}} Q_v^{-1} \underline{v} \ ,$$

that the non-centrality parameter is given as:

$$\lambda = \nabla^{\mathrm{T}} \overline{C}_v^{\mathrm{T}} Q_v^{-1} C_v (C_v^{\mathrm{T}} Q_v^{-1} C_v)^{-1} C_v^{\mathrm{T}} Q_v^{-1} \overline{C}_v \nabla \ . \tag{4.10}$$

Note that (4.10) indeed reduces to (4.2) if $\overline{C}_v = C_v$. The non-centrality parameter (4.10) can also be interpreted in geometric terms (cf. Fig. 4.1) as:

$$\lambda = \|P_{C_v} \overline{C}_v \nabla\|^2 \tag{4.11}$$

$$\lambda = \|\overline{C}_v \nabla\|^2 \cos^2 \phi \ , \tag{4.12}$$

where P_{C_v} is the orthogonal projector that projects onto the range space $R(C_v)$. The angle ϕ between $\overline{C}_v$ and the range space $R(C_v)$ is a measure of the *separability* between the alternative hypothesis associated with the executed test and the true alternative hypothesis. If the angle ϕ is small it is difficult to distinguish between H_A and H_A^{true}. From (4.3) and (4.12) it follows that $\lambda \leq \lambda_0$; in other words the detection power of all tests which do not correspond with H_A^{true} are smaller than that of the optimal test.

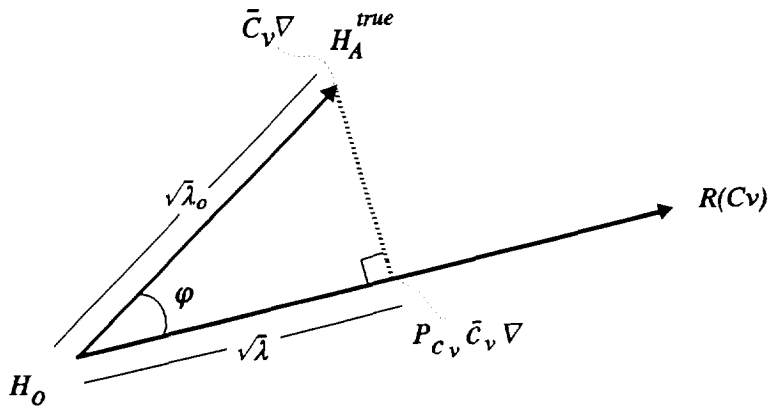


Figure 4.1: Predicted residual space with metric Q_v^{-1} and the two hypotheses H_A and H_A^{true}.

If one considers one-dimensional hypotheses the matrices $\overline{C}_v$ and C_v reduce to the vectors $\bar{c}_v$ and c_v respectively. The test associated with H_A^{true} has non-centrality parameter

$$\lambda_0 = \bar{c}_v^{\mathrm{T}} Q_v^{-1} \bar{c}_v \nabla^2 \tag{4.13}$$

whereas all other one-dimensional tests have non-centrality parameter (cf. eq. 4.10):

$$\lambda = \frac{(\bar{c}_v^{\mathrm{T}} Q_v^{-1} c_v)^2}{c_v^{\mathrm{T}} Q_v^{-1} c_v} \nabla^2 \ , \tag{4.14}$$

which using (4.12) can be written as:

$$\lambda = (\bar{c}_v^{\mathrm{T}} Q_v^{-1} \bar{c}_v) \nabla^2 \cos^2 \phi \ , \tag{4.15}$$

with

$$\cos^2 \phi = \frac{(\bar{c}_v^{\mathrm{T}} Q_v^{-1} c_v)^2}{(\bar{c}_v^{\mathrm{T}} Q_v^{-1} \bar{c}_v)(c_v^{\mathrm{T}} Q_v^{-1} c_v)} \ . \tag{4.16}$$

From (4.16) it follows that if one considers one-dimensional tests, $\cos^2 \phi$ represents the square of the correlation coefficient between the test statistics. The correspondence between $\cos^2 \phi$ and the correlation coefficient cannot be generalized in a straightforward manner to multi-dimensional test statistics. Let us (for the sake of completeness) consider two multi-dimensional alternative hypotheses (not necessarily of the same dimension) parametrized as

$$H_{A_i} : \underline{v} \sim N(C_{v_i} \nabla, Q_v) \ ; \ H_{A_j} : \underline{v} \sim N(C_{v_j} \nabla, Q_v) \ . \tag{4.17}$$

FÖRSTNER [1983] has shown that

$$\max \cos^2 \phi_{ij} = \max \lambda(M_{ij}), \tag{4.18}$$

with

$$M_{ij} = (C_{v_i}^{\mathrm{T}} Q_v^{-1} C_{v_j})(C_{v_j}^{\mathrm{T}} Q_v^{-1} C_{v_j})^{-1} (C_{v_j}^{\mathrm{T}} Q_v^{-1} C_{v_i})(C_{v_i}^{\mathrm{T}} Q_v^{-1} C_{v_i})^{-1} \ ,$$

where $\max \lambda(M_{ij})$ is the largest eigenvalue of M_{ij} and ϕ_{ij} is the angle between any pair of column vectors of which one is contained in the range space $R(C_{v_i})$ and one in the range space $R(C_{v_j})$ respectively.

4.4.2 The Minimal Detectable Bias

The Minimal Detectable Bias (MDB) is defined as the size of the model error that can be detected by a *one-dimensional* test. The non-centrality parameter has been fixed at λ_0. We start with the most general definition of the MDB and we consider the following two one-dimensional alternative hypotheses (not necessarily related to the same type of model error):

$$H_{A_1} : \underline{v} \sim N(c_{v_1} \nabla, Q_v) \ ; H_{A_2} : \underline{v} \sim N(c_{v_2} \nabla, Q_v) \ , \tag{4.19}$$

where the model misspecifications related to H_{A_1} and H_{A_2} start at l_1 and l_2 respectively. If we assume $H_A^{\mathrm{true}} = H_{A_2}$ the MDB related to the test associated with H_{A_1} at time k is (cf. eqs. 4.12, 4.14, and 4.16):

$$|\nabla_{l_2}^{l_1,k}| = \sqrt{\frac{\lambda_0}{(c_{v_2}^{\mathrm{T}} Q_v^{-1} c_{v_2}) \cos^2 \phi_{12}}} \tag{4.20}$$

$$= \sqrt{\frac{\lambda_0 (c_{v_1}^{\mathrm{T}} Q_v^{-1} c_{v_1})}{(c_{v_2}^{\mathrm{T}} Q_v^{-1} c_{v_1})^2}} \ . \tag{4.21}$$

The scalars $(c_{v_i}^{\mathrm{T}} Q_v^{-1} c_{v_j})$ can be computed as $(\sum_{r=\max(l_1,l_2)}^{k} c_{v_{i_r}}^{\mathrm{T}} Q_{v_r}^{-1} c_{v_{j_r}})$ with $i,j = 1,2$. The MDB associated with a test not corresponding to H_A^{true} is always larger than the MDB for the test corresponding to H_A^{true} (this is in agreement with the remarks made in connection with eq. 4.12). The smallest MDB at time k is obtained if H_{A_1} and H_{A_2} relate to the same model error and $l_1 = l_2$ (i.e. $H_{A_1} \equiv H_{A_2}$). This follows directly from (4.20) because only then $\cos^2 \phi_{12}$ is equal to one. The generality of the definitions (4.20) (or alternatively (4.21)) allows the use of the MDB as a measure of detectability and separability. An illustration of the use of the MDB as a measure of detectability and separability in analytical form is given by TEUNISSEN [1990a, 1990b].

We will use the MDB as a measure of detectability. Therefore we limit ourselves to the case that $l_1 = l_2 = l$ and we assume the test is performed for the actual model error. The MDB associated with a test at time k for a model misspecification with time of occurence l (denoted as $|\nabla^{l,k}|$) is then given as (cf. eq. 4.20):

$$\boxed{|\nabla^{l,k}| = \sqrt{\frac{\lambda_0}{\sum_{i=l}^{k} c_{v_i}^{\mathrm{T}} Q_{v_i}^{-1} c_{v_i}}}\,.} \tag{4.22}$$

Note that the MDBs are *not* dimensionless; the units of a particular MDB correspond with those of the related alternative hypothesis (the MDB of an outlier or slip in a range observation, for example, is specified in metres). The summation operation in the denominator of (4.22) causes the MDB to decrease with increasing window length $(k - l + 1)$ of the test. All subsequent derivations and computations will be based on (4.22). The separability of one-dimensional alternative hypotheses will be analysed using the correlation coefficient $\sqrt{\cos^2 \phi}$ (cf. eq. 4.16). If we consider the two hypotheses given by (4.19) the correlation coefficient can be computed as:

$$\rho_{12} = \frac{\sum_{j=\max(l_1,l_2)}^{k} c_{v_{1_j}}^{\mathrm{T}} Q_{v_j}^{-1} c_{v_{2_j}}}{(\sum_{j=l_1}^{k} c_{v_{1_j}}^{\mathrm{T}} Q_{v_j}^{-1} c_{v_{1_j}})^{1/2} (\sum_{j=l_2}^{k} c_{v_{2_j}}^{\mathrm{T}} Q_{v_j}^{-1} c_{v_{2_j}})^{1/2}}\,. \tag{4.23}$$

4.4.3 The Minimal Detectable Bias as a Design Tool

Minimal detectable biases are a convenient design tool because they can be easily interpreted. The designer of an integrated navigation system will generally have some knowledge of the type and magnitude of the model errors that are likely to occur. By means of the MDBs the designer can judge if these biases can really be detected by the proposed testing procedure. In the design procedure the MDB can be used to:

1. Verify the chosen testing strategy. The design of the integrated navigation system should be improved if the MDB computations reveal that certain biases are poorly detectable. If one knows from experience that a particular bias has little

impact on the estimation result and is not likely to occur, the corresponding test statistic may be deleted. More often, however, the effect of an actual bias may be detrimental to (part of) the estimation result. Then the design of the integrated navigation system has to be reconsidered (by, for example, adding additional sensors), or the testing parameters have to be adjusted (by, for example, increasing the window length of the tests).

2. Determine the delays and lags of the tests. The MDB provides a useful tool for the choice of the delays and lags of the tests for various types of alternative hypotheses (cf. Section 3.5.3). If, for example, the detection power of the local test is too small (i.e. a particular MDB is larger than the size of the likely model error), one may use tests with a certain delay of detection (from (4.22) follows that the MDB decreases with increasing window length). If the detection power of the local test is insufficient one could even consider not to execute the local test (one then introduces a certain lag). We still suggest to choose the lag of the tests equal to zero, because then real-time identification of model errors remains possible (a somewhat less likely model error could be larger than the MDB). The general idea of choosing an appropriate delay is illustrated in Fig. 4.2, where the delay of the test is determined by the point at which no significant decrease of the MDB occurs.

3. Compute measures of external reliability. In Sections 4.4.4 and 4.4.5 we will show how the bias in the state estimator due to an model error with the size of the MDB can be computed and analyzed in a straightforward manner.

The minimal detectable biases are at the base of the description of the internal and external reliability, and as such they are pivotal in reliability theory.

l ↓ \ k →	0	1	2	3	4	5
		0	1	2	3	4
			0	1	2	3
				0	1	2
					0	1
						0

(a)

l ↓ \ k →	8.3	5.9	5.5	5.4	5.4	5.4
		8.3	5.9	5.5	5.4	5.4
			8.3	5.9	5.5	5.4
				8.3	5.9	5.5
					8.3	5.9
						8.3

(b)

l ↓ \ k →	x	x	x	–	–	–
		x	x	x	–	–
			x	x	x	–
				x	x	x
					x	x
						x

(c)

Figure 4.2: Use of the MDB to determine the delay of the tests; (a) Delay $(k-l)$ in computing test statistics for $l \leq k$, with k time of testing and l time of occurence model misspecification; (b) MDB-matrix $|\nabla^{l,k}|$ for a hypothetical test; (c) Test statistics $t^{l,k}$ computed (x) with delay $k-l=2$ (or window length $k-l+1=3$).

The size of the MDB is influenced by the following design parameters:

- Measurement model. If more sensors are integrated into the navigation system the MDBs will generally decrease as the redundancy of the model is increased. How much the MDBs decrease depends mainly on the type, number and precision of the additional sensors. With each additional observation the number of MDBs to be evaluated increases.

- Measurement geometry. The geometry is an aspect in case land or space-based radiopositioning systems are part of the sensor set. In case, for example, only one (radio-)positioning system is used the size of the MDBs depends to a large extent on the transmitter geometry relative to the sensor position.

- Sample rate of sensors. Increasing the measurement sample rate results in a smaller contribution of the system noise between measurement updates and hence in smaller MDBs.

- Choice of the state-space model. If more sensors are integrated into the navigation system, the state space model can (or has to) be expanded by additional states (e.g. instrumental biases). The impact of changes in the dynamic model (usually accompanied by changes in the covariance matrix of the disturbances) on the MDBs is difficult to predict.

- Stochastic model of the observations. The covariance matrix of the observations (R) enters the definition of the MDBs directly via the covariance matrix of the predicted residuals (Q_{v_i}) and indirectly via the vectors c_{v_i}. Improvement of the measurement precision results in smaller MDBs.

- Stochastic model of the disturbances. The covariance matrix of the disturbances (Q) enters the covariance matrix of predicted residuals (Q_{v_i}) via the variance of the predicted observations and indirectly via the vectors c_{v_i}. Lower system noise leads to smaller MDBs.

- Filter concept. The data processing may be performed using one central filter or be based on a decentralized filter approach [CARLSON, 1988]. The data processing scheme has an impact on the testing strategy and thus on the MDBs.

- Testing parameters. The MDBs are a function of the non-centrality parameter λ_0, which in its turn is a monotonic decreasing function of the level of significance α_0 of the test (for a fixed power) and a monotonic increasing function of the power γ_0 (for a fixed level of significance) of the test. Increasing the window length $(k - l + 1)$ results in smaller MDBs.

For most practical applications, it is difficult to quantify the actual effect of the design parameters on the MDBs. For a simple case we can, however, demonstrate the properties of the MDB by an analytical example. We consider the following model with

one-dimensional state and observation vector (i.e. $m = 1$ and $n = 1$):

$$\begin{aligned} E\{\underline{d}_k\} &= x_k - x_{k-1} & D\{\underline{d}_k\} &= q\Delta T^2 \\ E\{\underline{y}_k\} &= x_k & D\{\underline{y}_k\} &= r\ , \end{aligned} \tag{4.24}$$

where ΔT is the (constant) sample interval. If we assume that the filter for model (4.24) is in steady state, it can be shown that the MDB associated with the local test for an outlier in an observation can be written as:

$$\nabla = \sqrt{\frac{\lambda_0[q\Delta T^2(1+\omega+\sqrt{1+2\omega})]}{2}} \quad \text{with} \quad \omega = \frac{2r}{q\Delta T^2}\ , \tag{4.25}$$

where $\omega > 0$ because r and q are variances. Equation (4.25) can be derived by solving for the steady state variance of the predicted state analogously to FRIEDLAND [1973], who considers a two-dimensional state space model. From (4.25) it can be seen that:

- The MDB increases if the system noise q increases.
- The MDB increases if the measurement noise r increases.
- The MDB increases if the sample interval ΔT increases.
- The MDB increases if the non-centrality parameter λ_0 increases.

These findings are in accordance with the general properties of the MDB listed above.

4.4.4 External Reliability

Undetected model errors have an impact on the state estimates. It is of interest to analyse how particular model errors are propagated as biases in the state vector or functions thereof. The effect of model errors on the filtered state follows directly from the filter algorithm. For a slip (of the size of the MDB) in observation i starting at time l ($c_j = c_i\ \forall\ j \geq l$) the bias in the filtered state at time k is

$$\nabla\hat{x}_{k|k} = \Phi_{k,k+1}\left(\sum_{j=l}^{k}\left\{\prod_{i=j+1}^{k}[\Phi_{i+1,i}(I - K_iA_i)]\right\}\Phi_{j+1,1}K_jc_j\right)\nabla_i^{l,k}, \tag{4.26}$$

where $\nabla_i^{l,k}$ is the MDB associated with the test statistic $\underline{t}^{l,k}$. The bias in the filtered state at time k due to an outlier in observation i (again of the size of the MDB) at time l is

$$\nabla\hat{x}_{k|k} = \Phi_{k,k+1}\left\{\prod_{i=l+1}^{k}[\Phi_{i+1,i}(I - K_iA_i)]\right\}\Phi_{l+1,1}K_lc_l\nabla_i^{l,k}. \tag{4.27}$$

Similar forms can be found for jumps and slips in the dynamic model. The biases can be computed recursively in the manner outlined in Section 3.5. The part of equation (4.26)

between parentheses, for example, corresponds to $X_{k+1,l}$ in (3.72), so that (4.26) (with eq. 3.84) can be written as:

$$\begin{aligned} \nabla\hat{\boldsymbol{x}}_{k|k} &= \Phi_{k,k+1} X_{k+1,l} \nabla_i^{l,k} \\ &= \overline{X}_{k,l} \nabla_i^{l,k} \ . \end{aligned} \tag{4.28}$$

The evaluation of the external reliability using the bias vectors $\nabla\hat{\boldsymbol{x}}_{k|k}$ is laborious because every alternative hypothesis results in such a (n-dimensional) vector. On the other hand the analysis of the vectors $\nabla\hat{\boldsymbol{x}}_{k|k}$ provides a direct insight into the impact of model errors on the state vector. If the filter under consideration is in steady state, the bias vectors $\nabla\hat{\boldsymbol{x}}_{k|k}$ are representative for the whole trajectory and consequently only a limited number of vectors has to be considered. The bias vectors $\nabla\hat{\boldsymbol{x}}_{k|k}$ may also serve, at a later stage, to explain the properties of derived external reliability measures such as the bias to noise ratio.

4.4.5 The Bias to Noise Ratio

Often not the bias vector $\nabla\hat{\boldsymbol{x}}_{k|k}$ itself, but its significance is analysed, because the analysis of the bias vectors is quite laborious. A measure of the significance of the bias in the state vector is the Bias to Noise Ratio (BNR):

$$\lambda_{\hat{x}} = \nabla\hat{\boldsymbol{x}}_{k|k}^{\mathrm{T}} P_{k|k}^{-1} \nabla\hat{\boldsymbol{x}}_{k|k} \ . \tag{4.29}$$

The BNR is a *scalar* measure of the significance of the bias in the state vector and is a *dimensionless* quantity. The BNR can geometrically be interpreted as the square of the length of the vector $\nabla\hat{\boldsymbol{x}}_{k|k}$ in the vector space $\mathcal{R}^n$ (with metric $P_{k|k}^{-1}$):

$$\lambda_{\hat{x}} = \|\nabla\hat{\boldsymbol{x}}_{k|k}\|^2 \ . \tag{4.30}$$

Sometimes one is only interested in the effect of a model error on particular functions of the state vector. One can represent these linear(-ized) functions of interest by a $r \times n$ matrix F^{T}. The BNR pertaining to these functions follows from (4.29) as:

$$\boxed{\lambda_{F^{\mathrm{T}}\hat{x}} = (F^{\mathrm{T}}\nabla\hat{\boldsymbol{x}}_{k|k})^{\mathrm{T}} (F^{\mathrm{T}} P_{k|k} F)^{-1} F^{\mathrm{T}} \nabla\hat{\boldsymbol{x}}_{k|k} \ .} \tag{4.31}$$

In case $F^{\mathrm{T}} = I_n$ equation (4.31) reduces to (4.29). Geometrically the BNR given by (4.31) can be interpreted as [TEUNISSEN,1990c]:

$$\lambda_{F^{\mathrm{T}}\hat{x}} = \|P_{(P_{k|k}F)} \nabla\hat{\boldsymbol{x}}_{k|k}\|^2 \ , \tag{4.32}$$

where $P_{(P_{k|k}F)}$ is the orthogonal projecter on the range space $R(P_{k|k}F)$ along $N(F^{\mathrm{T}})$. A comparison of (4.32) with (4.30) shows that the BNR associated with a subset of the

state vector is always smaller than the BNR related to the complete state vector, that is for any $r \times n$ matrix F^{T} of full rank

$$\lambda_{F^{\mathrm{T}}\hat{x}} \leq \lambda_{\hat{x}}. \tag{4.33}$$

In practice the matrix F^{T} is often chosen in such a way that certain elements of the state vector are selected. If, for example, one wants to investigate the significance of the bias in the first two elements of a 4-dimensional state vector, the 2×4 F^{T}-matrix is given as:

$$F^{\mathrm{T}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

The BNRs given by (4.29) and (4.31) are invariant under reparametrization of the state vector.

The interpretation of the BNR is simplified if we consider the bias of an arbitrary linear function $\nabla\hat{\theta} = a^{\mathrm{T}}(F^{\mathrm{T}}\nabla\hat{\boldsymbol{x}}_{k|k})$, with $a \in \mathcal{R}^r$. The function $\nabla\hat{\theta}$ can also be written as:

$$\nabla\hat{\theta} = a^{\mathrm{T}}(F^{\mathrm{T}}P_{k|k}F)(F^{\mathrm{T}}P_{k|k}F)^{-1}F^{\mathrm{T}}\nabla\hat{\boldsymbol{x}}_{k|k}\ .$$

Application of the cosine rule gives

$$\begin{aligned}
\nabla\hat{\theta} &= \|(F^{\mathrm{T}}P_{k|k}F)a\|\|F^{\mathrm{T}}\nabla\hat{\boldsymbol{x}}_{k|k}\|\cos((F^{\mathrm{T}}P_{k|k}F)a, F^{\mathrm{T}}\nabla\hat{\boldsymbol{x}}_{k|k}) \\
\nabla\hat{\theta} &\leq \sqrt{a^{\mathrm{T}}(F^{\mathrm{T}}P_{k|k}F)(F^{\mathrm{T}}P_{k|k}F)^{-1}(F^{\mathrm{T}}P_{k|k}F)a}\sqrt{\nabla\hat{\boldsymbol{x}}_{k|k}^{\mathrm{T}}F(F^{\mathrm{T}}P_{k|k}F)^{-1}F^{\mathrm{T}}\nabla\hat{\boldsymbol{x}}_{k|k}} \\
\nabla\hat{\theta} &\leq \sqrt{a^{\mathrm{T}}(F^{\mathrm{T}}P_{k|k}F)a}\sqrt{\nabla\hat{\boldsymbol{x}}_{k|k}^{\mathrm{T}}F(F^{\mathrm{T}}P_{k|k}F)^{-1}F^{\mathrm{T}}\nabla\hat{\boldsymbol{x}}_{k|k}} \\
\nabla\hat{\theta} &\leq \sqrt{\sigma^2_{\nabla\hat{\theta}}}\sqrt{\lambda_{F^{\mathrm{T}}\hat{x}}} \quad \text{(cf. eq. 4.31)}\ .
\end{aligned}$$

From this then immediately follows that:

$$\left|\frac{\nabla\hat{\theta}}{\sigma_{\nabla\hat{\theta}}}\right| \leq \sqrt{\lambda_{F^{\mathrm{T}}\hat{x}}}\ , \tag{4.34}$$

so that the BNR times the standard deviation provides a convenient upperbound for the bias $\nabla\hat{\theta}$ of an arbitrary linear function of the state vector, viz:

$$|\nabla\hat{\theta}| \leq \sqrt{\lambda_{F^{\mathrm{T}}\hat{x}}} \times \sigma_{\nabla\hat{\theta}}\ . \tag{4.35}$$

If only a single element of the state vector is considered, the BNR times the standard deviation provides the following upperbound:

$$|\nabla\hat{\boldsymbol{x}}_{k|k_i}| \leq \sigma_{\hat{x}_i} \times \sqrt{\lambda_{\hat{x}}}\ . \tag{4.36}$$

The inequalities (4.35) and (4.36) provide useful interpretations of the external reliability measures $\lambda_{\hat{x}}$ and $\lambda_{F^{\mathrm{T}}\hat{x}}$.

In case one considers multi-dimensional hypotheses the BNR is still a scalar quantity, but its magnitude depends on the direction vector d specified in (4.4) (cf. [KÖSTERS, 1992]). If we insert (4.6) in (4.28) with $\nabla_i^{l,k}$ replaced by ∇, the BNR (4.29) can be written as:

$$\lambda_{\hat{x}} = \lambda_0 \frac{d^{\mathrm{T}} \overline{X}_{k,l}^{\mathrm{T}} P_{k|k}^{-1} \overline{X}_{k,l} d}{d^{\mathrm{T}} C_v^{\mathrm{T}} Q_v^{-1} C_v d} . \tag{4.37}$$

In equation 4.37 $\lambda_{\hat{x}}/\lambda_0$ can be conceived as a Rayleigh quotient and thus the extrema of the BNR can be obtained by solving the generalized eigenvalue problem

$$(\overline{X}_{k,l}^{\mathrm{T}} P_{k|k}^{-1} \overline{X}_{k,l}) d = \lambda (C_v^{\mathrm{T}} Q_v^{-1} C_v) d \tag{4.38}$$

for the largest and smallest eigenvalue. The bias which has the largest influence on the parameters is in the direction of the eigenvector associated with the largest eigenvalue. The (normalized) eigenvectors associated with $\lambda_{\max}$ and $\lambda_{\min}$ do not necessarily correspond with the principal axes of the internal reliability ellipsoid obtained with (4.7).

In practice one generally analyses the square root of the BNR ($\sqrt{\lambda_{\hat{x}}}$ or $\sqrt{\lambda_{F^{\mathrm{T}}\hat{x}}}$) to facilitate a direct comparison with the bias $\nabla\hat{x}$, the MDBs, and the standard deviation of the (filtered) state (cf. equations 4.35 and 4.36).

4.4.6 The Bias to Noise Ratio as a Design Tool

The bias to noise ratio is an important design tool because it provides a measure of the significance of the bias in the state estimator caused by an model error of the size of the MDB. The interpretion of the BNR is difficult because the (dimensionless) measure consists of two elements, namely the bias $\nabla\hat{x}_{k|k}$ and the covariance matrix of the state vector. The interpretation is simplified if one uses the BNR as in the inequalities (4.35) and (4.36). In the design procedure the BNR can be used to:

1. Verify the testing strategy. If certain model errors result in large BNRs (and thus significant biases in the estimators), the design of the system should be reconsidered. Small BNRs on the other hand indicate that the system design may be relaxed (by, for example, lowering the redundancy of the system), but do not indicate that a particular test can be automatically deleted from the testing procedure, because if the likely model error is expected to be (much) larger than the MDB it can still cause a considerable bias in the state vector.

2. Determine the delays of the tests. The determination of the delays of the tests by means of the MDB has already been discussed. If the analysis of the MDBs indicates that a small delay of the test is sufficient, but the significance of the bias remains high (i.e. the BNR is large) with increasing delay, the delay of the test should be chosen larger than the analysis of the MDBs indicates. The analysis of the BNRs for the determination of the delays thus serves to substantiate the determination of the delays by means of the MDBs. The general idea is sketched

in Figure 4.3 (the notation of Fig. 4.2 is maintained) where we consider scenarios for an outlier in an observation causing an insignificant bias for large delays and a slip which causes a significant bias.

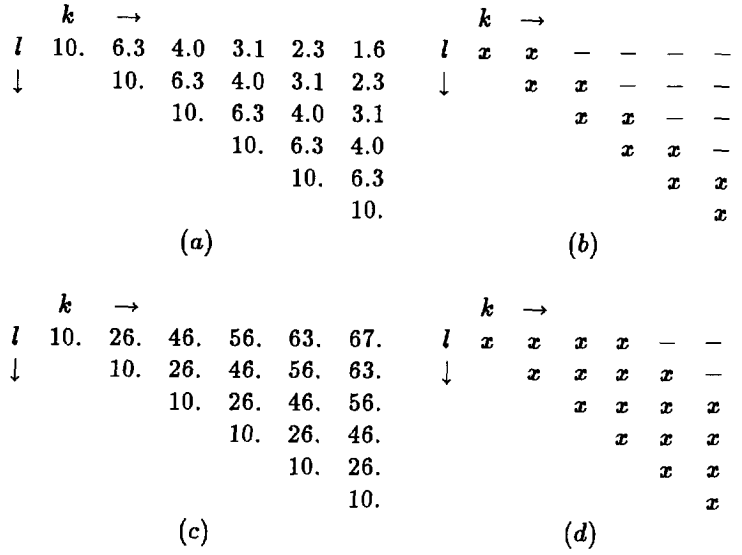


Figure 4.3: Use of the BNR to determine the delay of the tests; (a) BNR-matrix associated with a hypothetical outlier; (b) Test statistics computed (x) with delay $k - l = 1$; (c) BNR-matrix associated with a hypothetical slip; (d) Test statistics computed (x) with delay $k - l = 3$.

The BNR is influenced by the same design parameters as the MDB (cf. Section 4.4.3). Its response to design changes is difficult to predict as the BNR is a combined measure of the bias and precision of the state estimator. A design option that causes the bias to decrease and at the same time improves the precision will have little effect on the value of the BNR. The behaviour of the BNR can be predicted fairly well for the following design changes:

- Increasing the redundancy of the system (this is usually accomplished by using additional observations) results in lower BNRs. With each additional observation the number of BNRs to be evaluated increases.

- Increasing the non-centrality parameter λ_0 leads to larger values of the BNRs.

The size of the BNR is furthermore dependent on the subset or linear function (modelled with F^T) of the state vector one is interested in. The analysis of the external reliability for integrated navigation systems is usually limited to the state vector elements related to position and/or velocity. The BNR corresponding to this subset will always be

smaller than the BNR pertaining to the full state vector. The choice to which (linear(-ized) function of the) unknowns the BNR relates, should be given serious consideration.

The characteristics of the BNR can be nicely illustrated for the simple model given by (4.24). For the local test for an outlier in an observation it can be shown that the BNR can be written as:

$$\lambda_{\hat{x}} = \frac{\lambda_0(1+\sqrt{1+2\omega})}{\omega} \tag{4.39}$$

with $\omega = 2r/(q\Delta T^2)$ (cf. eq. 4.25). From (4.39) follows that $\lambda_{\hat{x}}$ is a monotonic decreasing function in ω. As a consequence it follows directly from this example that:

- The BNR increases if the system noise q increases.
- The BNR decreases if the measurement noise r increases.
- The BNR increases if the sample interval ΔT increases.
- The BNR increases if the non-centrality parameter λ_0 increases.

Unfortunately the properties of the BNR for dynamic systems in general cannot be illustrated so easily.

4.4.7 Choice of Conventional Alternative Hypotheses

The choice of the alternative hypotheses is application dependent. One may, however, consider whether the concept of the conventional alternative hypothesis can also be maintained for dynamic systems. BAARDA [1968] suggested the use of so-called conventional alternative hypotheses to arrive at a standard reliability description of networks. In practice the proposed conventional hypothesis (assuming an outlier in a single observation) is often the only specific alternative hypothesis used in the testing procedure. As a result the choice of which alternative hypotheses should actually be tested has received little attention (a notable exception being the field of deformation analysis). The growing importance of GPS (with its inherent cycle-slip problem) and integrated navigation systems requires that the specification of alternative hypotheses should seriously be reconsidered. If we limit ourselves to the slippage tests of the DIA procedure we see that basically two types of model errors can be conveniently modelled (viz. outliers and slips in the observables). Furthermore the DIA procedure requires that the window length of the test is specified.

Currently too little experience is available for the definition of conventional alternative hypotheses in the design of integrated navigation systems. The only universal model error (probably for all sensor types) is the outlier in a single observation. It seems, however, that this convention is rather limited, because then model errors related to 'soft' sensor failures, slips in the observations, and changes in the dynamic model are not taken into account. It might very well be possible that every *type* of sensor and dynamic model requires a separate convention. The generality of BAARDA'S proposal is then lost, but this is not asthonishing, considering the wealth of navigation sensors available.

4.4.8 Computational Aspects

In the previous sections we have provided an overview of the important reliability measures. If one looks at the equations that define the MDB and the BNR (we refer to (4.22) and (4.31) in particular), one sees that there exists a close relationship with the equations we found in the derivation of the testing procedure. The denominator of the MDB (4.22) is identical to the denominator of the test statistic corresponding to the MDB (3.68). Also the BNR (4.31) can be computed in a straightforward manner. The covariance matrix of the state estimator is provided by the Kalman filter. The bias vector $\nabla\hat{\boldsymbol{x}}_{k|k}$ can be computed recursively with (4.28), where the vectors (or matrices) $X_{k+1,l}$ are computed according to one of the schemes provided by (3.69) to (3.72). Consequently the quality measures can be computed by the same software that is used for the filter computations (in the software for the design and adjustment of geodetic networks this has long been exploited). The close relationship between design and filter computations allows that the reliability is even monitored in real-time.

4.4.9 Specification of Reliability Requirements

In practice reliability requirements for navigation systems are rarely specified explicitely. For many real-time precise positioning applications, however, it is nowadays required that some quality control procedure, usually a local testing procedure, is implemented (see, e.g., [NICOLAI, 1988]). The purpose of the testing procedure is the validation of the null hypothesis. If H_0 is valid, the precision description as provided by the filter is then used to quantify the quality of the system. A table with reliability requirements for geodetic navigation applications (similar to Table 4.1 but expressed in MDBs, BNRs, or (functions of) the bias vector $\nabla\hat{\boldsymbol{x}}$.) is not yet available. It is likely that reliability requirements will be specified in terms of external reliability measures (because one is primarily interested in the impact of model errors on the estimation *result*). The BNR would be a convenient measure as it can be more efficiently analysed than a set of bias vectors for a full range of alternative hypotheses and because it provides a link with the precision of the filtered state.

In the meantime one can proceed as follows. For most (integrated) navigation systems the type and magnitude (or range) of the model errors that are likely to occur are quite well known. Implicitely these likely model errors thereby provide the basis for an internal reliability requirement: the MDB should be smaller than the size of the model error. Requirements with respect to external reliability can often be specified in terms of functions of the bias vector. In 3-D seismic applications, for example, it is for instance very important that a seismic signal can be allocated to a certain bin (a bin represents a, say, 25m $\times$ 25m area on which the geophysical data processing is based). The maximally tolerated bias in position can then directly be derived from the size of the bin (note that such a requirement may result in much stricter quality requirements than the precision requirement). Based on the given bias, one can then derive (using inequality (4.35)) an upperbound for the BNR.

4.4.10 Summary

In this section we have discussed internal and external reliability (measures) and have commented on their use in navigation system design. Internal reliability is described with the MDB, which represents the size of the bias that can be detected by a certain test with a certain probability. We have indicated how the MDB, which can be easily interpreted, can be used to determine the window lengths of the tests. External reliability is described by (functions of) the bias vector $\nabla\hat{x}$ and the BNR, which represents the significance of the bias relative to the precision. The characteristics of the BNR allow a direct comparison of the reliability related to various hypotheses. The interpretation of the BNRs can be somewhat involved, but we have shown how the BNR is related to the precision and bias of the filtered state vector. The description of reliability is dependent on which alternative hypotheses are considered in the testing procedure. No convention concerning the choice of alternative hypotheses could be given yet.

4.5 A First Step towards a Design Procedure for Integrated Navigation Systems

In the following we propose a design strategy for integrated navigation systems based on the quality measures for precision and reliability derived in Sections 4.3 and 4.4.

The precision requirements for (geodetic) navigation applications are well known (cf. Table 4.1). Therefore we suggest to start the procedure by designing a system which meets the precision requirements. After this has been accomplished we can then check if the reliability of the system is sufficient to guarantee that the precision requirements are met under operational conditions. Although usually no reliability requirements are explicitely specified, the model errors which are likely to occur can be pinpointed rather well. Implicitely this knowledge provides us with requirements for the internal reliability. A first step is to validate if the size of the likely model errors is smaller than the size of the corresponding MDBs. If this is not the case the design of the system has to be improved until the internal reliability meets the suggested requirement. In a second step one can, based on the finally obtained MDBs, compute the BNRs for the various alternative hypotheses. Using inequality (4.35) one can compute an upperbound for the bias in the filtered state. If this bias does not exceed some preset threshold (usually one will, depending on the application, have some idea on the maximum allowable bias), the external reliability is sufficient. In the procedure one should give much consideration to the specification of possible model errors, because the choice of the alternative hypotheses determines the completeness of the reliability description.

The system design should not exceed the given precision (and possibly reliability) requirements too much. Furthermore it is likely that in practice the design options are limited by constraints on, e.g., budgets and available measurement systems. The result of the design procedure is a description of the functional and stochastic model of the integrated navigation system, a quantification of the quality of the system, and

a description of the proposed testing strategy (cf. our discussion in Section 4.2). The testing procedure really has to be implemented in the integrated navigation system in order to warrant the reliability of the system under real-time operation. Therefore we are of the opinion that the part of the design procedure which relates to reliability is an intrinsic part of the DIA procedure.

The design procedure for integrated navigation systems differs from the well-established design procedure for control networks. The difference is primarily due to the fact that the purpose of an integrated navigation system is usually well defined. The endeavour to arrive at a design which has a 'homogeneous' precision and reliability description therefore not necessarily applies to integrated navigation systems. The mere variety of alternative hypotheses one has to consider in the design phase of a navigation system, renders the realization of a 'homogeneous' reliability description unlikely.

4.6 Reliability and RAIM

In Section 4.4 we noted that in practice reliability requirements are not often explicitely specified. A notable exception is found in the civil aviation community, where much research has been performed concerning integrity. Integrity means that the navigator is given a timely warning if certain precision and reliability requirements are not met. By means of an example we will relate our reliability measures with the reliability requirements which can be derived from a particular integrity monitoring scheme, namely RAIM.

For aviation applications it is (legally) required that a system failure, causing a certain bias in (horizontal) position, is detected in real-time by a test with a certain level of significance and power. If this failure detection is accomplished by checking the self-consistency within a single subsystem, the testing procedure is generally denoted as Receiver Autonomous Integrity Monitoring (RAIM).

A test of a certain dimension can be characterized by fixing two out of three testing parameters, namely the size of the test α, the power of the test γ, and the non-centrality parameter λ. In the navigation literature the size and power of the test are usually denoted as the probability of false alarm (α) and the probability of detection (γ) (sometimes the probability of missed detection ($1-\gamma$) is used).

In the civil aviation community it is common practice to fix α and γ and to give an upper limit in terms of the position error for each phase of flight [RTCA, 1991]. This actually means that an upper limit is given for all three testing parameters. A bias which causes a certain position error should be detected by a test (which test is not explicitely specified, but note it could very well be the detection procedure discussed in Chapter 3) with at least the prespecified α and γ. If this requirement cannot be met, the test is declared unreliable. STURZA [1988], STURZA AND BROWN [1990], and BROWN ET AL. [1991] have performed design computations to establish under which conditions GPS (in pseudo-range mode) can meet aeronautical requirements (assuming only one pseudo range is affected by a model error). The investigations are

based on a least squares adjustment per fix. The example we consider is the phase of flight called 'enroute'. In Table 4.2 the RAIM requirements are given. The minimum

alarm limit	maximum allowable alarm rate	time to alarm	minimum detection probability
3704 [m] (2 [nmi])	0.002/hour	30 [sec]	0.999

Table 4.2: RAIM requirements for enroute navigation [RTCA,1991].

detection probability corresponds to a lower limit for γ. Following the interpretation of STURZA AND BROWN [1990] a maximal allowable alarm rate of 0.002/hour with a decision every 30 seconds (time to alarm) corresponds to a upper limit for α of 1.7E-5 ($= (0.002/3600) \times 30$). The alarm limit is the acceptable radial position error and is specified in terms of a function of the bias vector $\nabla\hat{x}$. For a given satellite configuration one can then derive the size of the bias in the measurement which will cause that the alarm limit is exceeded. Contrary to the B-method of testing the size of the (detection) test used in the RAIM methods is fixed for various degrees of freedom.

Using our approach the RAIM procedure could be implemented (and extended) as follows. One uses the overall model test with the specified size α for the particular phase of flight. If the local overall model test statistic is larger than the critical value ($\chi^2_\alpha(b,0)$) a model misspecification is detected and one can (if the redundancy b is larger than 1) identify the model error (although identification is not part of the RAIM requirements). Based on the α corresponding to the test with b degrees of freedom, one can compute α_0 for a one-dimensional test using the B-method. The non-centrality parameter λ_0 is then obtained from the inverse power function (the computations may seem rather complex, but can be easily tabulated, cf. Table 4.3). One can then compute the MDB and BNR for each satellite range. From Table 4.3 it follows that the non-centrality parameter and thus the MDB increase with increasing redundancy. This seems to contradict our findings of Section 4.4.3, but there the α_0 was kept fixed. Using inequality (4.35) one can then check if one stays within the alarm limit and declare the (RAIM) test unreliable if one does not stay within the alarm limit. One thus has available a methodology which does not only detect but also identifies model misspecifications and provides real-time quality assurance. Furthermore this RAIM

degrees of freedom b	α	$\chi^2_\alpha(b,0)$	γ	λ_0	α_0	$\chi^2_{\alpha_0}(1,0)$
1	1.7E-5	18.5	0.999	54.6	1.7E-5	18.5
2	1.7E-5	22.0	0.999	59.2	4.4E-6	21.2
3	1.7E-5	24.8	0.999	62.6	1.4E-6	23.3
4	1.7E-5	27.3	0.999	65.4	5.7E-7	25.0
5	1.7E-5	29.7	0.999	67.9	2.6E-7	26.5
6	1.7E-5	31.9	0.999	70.2	1.2E-7	27.9

Table 4.3: Testing parameters according to the B-method corresponding to the RAIM requirements for enroute navigation.

procedure can be easily extended to the case where the data processing is based on a Kalman filter.

4.7 Concluding Remarks

The design of dynamic systems is a part of the quality assurance of dynamic systems. Only if the full cycle of design, real-time quality control with the DIA procedure, and validation of the output has been completed, the quality of the process is assured. Because the output of a (navigation) system is dependent on actual data and environmental conditions, the quality of the output of the system may differ significantly from the quality of the system design.

In this chapter we limited ourselves to the design aspects precision and reliability. Although we paid relatively much attention to reliability, precision is just as an important aspect of quality. Even if only precision requirements have to be met, one still has to perform a reliability analysis. The quantification of the precision (which is valid under H_0) can only be guaranteed if the reliability of the system is sufficient.

We want to stress again that the reliability description is dependent on the model errors which are considered and therefore the choice of an appropriate set of alternative hypotheses is of crucial importance in the design of an integrated navigation system. The specification of alternative hypotheses is application dependent, but the measures of reliability can be used for the design of arbitrary dynamic systems (irrespective of the application).

Based on the measures of precision and reliability we tentatively proposed a design procedure for integrated navigation systems. In the next chapter we will apply the measures we derived for precision and reliability to a case study in which we consider the design of an integrated navigation system.

Chapter 5

Design Computations

5.1 Introduction

In this chapter we will consider the design of dynamic systems with respect to quality. The concept of quality and the measures by which it can be quantified have been discussed in Chapter 4. We will limit ourselves to the design of a simple linear system and an integrated (hydrographic) navigation system. Up to the present little experience is available concerning the use of statistical quality measures in the design of integrated navigation systems (contrary to the land surveying practice where quality measures haven proven to be invaluable design tools). One of the main objectives of this case study is therefore to analyse the characteristics of the quality measures under various model assumptions with emphasis on the reliability measures. Consequently this design study is primarily a comparative study. We will furthermore try to establish the suitability of the various design measures for the design of integrated navigation systems. Finally, this chapter may serve to provide a basis for a methodology for the design of a testing procedure for integrated systems. Based on the design study the window lengths of the tests have to be chosen.

We start the analysis with a thorough, though limited, design study of a simple, linear two-dimensional model. Such a model allows a first introduction to the use of the quality measures in system design and facilitates our understanding of the results related to the more sophisticated navigation systems. The design study of the navigation system is based on the setup of the NAVGRAV-survey, which was executed in 1986 [HAAGMANS ET AL., 1988]. Because the study is primarily directed towards the analysis of various quality measures in integrated navigation system design, it does not deal with the design of the actual navigation system (or rather hydrographic surveying system) itself. The design study is limited to statistical quality measures, namely precision and reliability. Other design parameters such as cost and available hardware are not considered. No a priori quality requirements have been specified, and as a consequence no optimization with respect to some preset quality measures is performed. Some results pertaining to the navigation system considered here have been reported

previously in [SALZMANN, 1990;1991].

As we are concerned with the *design* of (integrated navigation) systems, all the results presented in this chapter are independent of actual data (but, of course, still dependent on the chosen measurement scenario).

5.1.1 Overview of this Chapter

We start by briefly recapitulating the quality measures in Section 5.2. In Section 5.3 the simple linear model with which we will start our design computations is introduced. Design results for the linear model are discussed in Section 5.4. The setup of the integrated navigation system is given in Section 5.5 and the results pertaining to this system are discussed extensively in Section 5.6. Section 5.7 is devoted to design studies (especially with regard to reliability) performed by others. The chapter is concluded by summarizing the main findings in Section 5.8.

5.2 Quality Measures

We briefly repeat the measures which are used in the design studies and have been derived in Chapter 4. The analysis of reliability is limited to one-dimensional alternative hypotheses. The reliability measures are computed based on the assumption that the filter operates under the null hypothesis except for a *single* model error. The following quality measures are considered:

Precision

- Standard deviations of the filtered states.
- Point error ellipses.

Both precision measures are a function of the covariance matrix of the state vector.

Internal Reliability

- Minimal Detectable Biases (MDBs). The MDB is defined by (4.22) and represents the size of the the model error that can be detected by a particular one-dimensional test. The MDB is used as a measure of detectability.
- Correlation coefficient between hypotheses. The correlation coefficient (given by eq. 4.23) is a measure of the separability between two alternative hypotheses which are related to the same time of occurence l of the model error and the same time of computation k.

External Reliability

- Bias in the filtered state. The direct influence of model errors on the filtered state is given by the (elements of the) bias vector $\nabla\hat{x}_{k|k}$. Elements (of functions) of the bias vector are plotted in so-called response graphs.

- Bias to Noise Ratios (BNRs). The scalar, dimensionless BNR is defined by (4.29) and is a measure of the significance of the bias in the state vector. We will analyse the square root of the BNR to facilitate a direct comparison with (elements of) the bias vector $\nabla\hat{x}_{k|k}$, the MDBs, and the standard deviation of the predicted state.

The description of the reliability of a dynamic system is very dependent on the alternative hypotheses which are specified. The specification of alternative hypotheses is application dependent and is therefore discussed in conjunction with the linear and navigation models separately. To provide a link between the various reliability measures the flowchart given in Fig. 5.1 might be helpful. In Fig. 5.1 we consider model errors of outlier and slip type which occur at time l and are analyzed at time k. (The notation $\tilde{\nabla}\hat{x}_{i|i}$ indicates an intermediate result.)

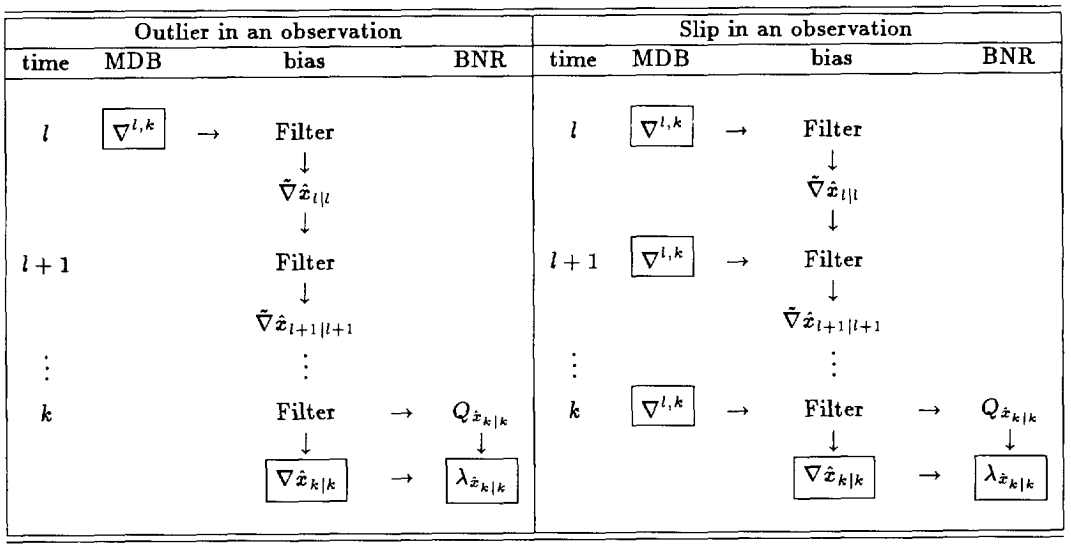


Figure 5.1: Interdependence of measures of internal and external reliability.

5.2.1 Testing parameters

For the design computations the level of significance is set at $\alpha_0 = 0.001$ for the one-dimensional tests. The power of the tests is $\gamma_0 = 0.80$. This results in a noncentrality parameter of $\lambda_0 = 17.07$. The choice of the values of the testing parameters corresponds to the choice generally made for network design. In land-surveying the values $\alpha_0 = 0.001$ and $\gamma_0 = 0.80$ are commonplace and are derived from an extensive measurement experience. It might be that in the future (navigation) applications require different values of the testing parameters. Note, however, that the magnitude of the testing parameters only has an impact on the size of the reliability measures.

5.3 Design Setup — Linear Models

We begin our design studies by considering a simple, linear model. We will extensively discuss the design of this model, which will serve to illustrate the properties of especially the reliability measures, because we think many aspects of the design procedure can be much better understood using a simple model. We will only analyse two linear models and therefore our analysis of these models only provides a limited insight on the impact of the design parameters on the quality measures.

The linear (two-dimensional) polynomial model is based on the following dynamic model:

$$E\{\underline{d}_k\} = \begin{pmatrix} x \\ v_x \end{pmatrix}_k - \begin{pmatrix} 1 & t_k - t_{k-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ v_x \end{pmatrix}_{k-1}, \tag{5.1}$$

with $D\{\underline{d}_k\} = Q_k$. The (linear) measurement model is given as:

$$E\{\underline{y}_k\} = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ v_x \end{pmatrix}_k, \tag{5.2}$$

with $D\{\underline{y}_k\} = R_k$. We assume a one-second sampling interval (i.e. $t_k - t_{k-1} = 1$). Furthermore we assume that $R_k = 1$ for all k. The two particular cases we consider are denoted as LM1 and LM2, and are only different as far the covariance matrix of the system noise (Q_k) is concerned:

$$Q_{\text{LM1}} = \begin{pmatrix} 0.333 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}; \quad Q_{\text{LM2}} = \begin{pmatrix} 0.00333 & 0.005 \\ 0.005 & 0.01 \end{pmatrix}. \tag{5.3}$$

This linear model can be interpreted as a navigation problem along a line, where the position (x) and velocity (v_x) along the line (expressed in metres and metres/second respectively) are the state vector elements and the position coordinate is the observable. The choice of system noise in (5.3) corresponds to standard deviations of the accelerations (which model the disturbances in model 5.1) of 1m/s^2 for model LM1 and 0.1m/s^2 for model LM2.

Based on the model description given by (5.1) to (5.3) two datasets with a duration of 100 seconds (this corresponds to 100 fixes) have been created for each model. The analysis of reliability is based on alternative hypotheses for outliers and slips in the observations.

5.4 Design Results — Linear Models

In this section we will successively analyse the precision, internal reliability (MDBs), and external reliability (using the bias vectors $\nabla\hat{x}_{k|k}$, and the BNR) pertaining to the models LM1 and LM2. For the design computations the initial state and its covariance matrix are chosen as:

$$x_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \quad Q_{x_0} = \begin{pmatrix} 10^7 & 0 \\ 0 & 10^7 \end{pmatrix}. \tag{5.4}$$

After 30 fixes the filters for models LM1 and LM2 are in steady state (i.e. the entries in the covariance matrices $P_{k|k}$ and $P_{k+1|k}$ remain constant).

EKSTRAND [1983] has derived an analytical solution for this two-dimensional model and gives closed form expressions for the elements of the covariance matrices of the estimators of the predicted and filtered state and the elements of the gain matrix. EKSTRAND'S results (which are identical to our numerical results once the filter has reached steady state) for models LM1 and LM2 are listed in Table 5.1.

	LM1	LM2
$P_{k\|k-1}$	$\begin{pmatrix} 3.11 & 1.76 \\ 1.76 & 2.03 \end{pmatrix}$	$\begin{pmatrix} 0.564 & 0.125 \\ 0.125 & 0.050 \end{pmatrix}$
$P_{k\|k}$	$\begin{pmatrix} 0.757 & 0.493 \\ 0.493 & 1.03 \end{pmatrix}$	$\begin{pmatrix} 0.361 & 0.080 \\ 0.080 & 0.040 \end{pmatrix}$
K_k	$\begin{pmatrix} 0.757 \\ 0.493 \end{pmatrix}$	$\begin{pmatrix} 0.361 \\ 0.08 \end{pmatrix}$

Table 5.1: Filter results (steady state solution) — cases LM1 and LM2.

5.4.1 Precision

The standard deviations of the filtered states are given in Table 5.2 and follow directly from Table 5.1. Clearly the standard deviations related to model LM2 are much smaller (in particular of the velocity state) than of those related to model LM1. It is well known that smaller system noise leads to better precision of the filtered state estimators.

LM1		LM2	
position	velocity	position	velocity
0.87	1.01	0.60	0.20

Table 5.2: Standard deviation of position [m] and velocity [m/s] estimators (steady state solution) — cases LM1 and LM2.

5.4.2 Internal Reliability

The Minimal Detectable Biases (MDBs) of the coordinate observables of models LM1 and LM2 are given in Table 5.3, which nicely illustrates the fact that the size of the MDBs decreases with increasing delay (or window length). The MDBs do not *significantly* decrease for delays larger than, say, three (except for the slip hypothesis in model LM2). This means that the contribution of the scalars c_{v_i} to the product $c_{v_i}^{\mathrm{T}} Q_{v_i}^{-1} c_{v_i}$ in (4.22) becomes negligible for large delays (Q_{v_i} is a constant because of the steady state

model	LM1		LM2	
delay	outlier	slip	outlier	slip
0	8.38		5.17	
1	5.23	8.13	4.73	4.51
2	5.20	7.50	4.53	4.42
3	5.17	7.30	4.44	4.42
4	5.14	7.28	4.41	4.39
5	5.14	7.28	4.40	4.31
6	5.14	7.28	4.40	4.22
7	5.14	7.28	4.40	4.13
8	5.14	7.28	4.40	4.05
9	5.14	7.28	4.39	4.00

Table 5.3: Minimal Detectable Biases [m] for outliers and slips in the coordinate observation at fix 90 ($l = 90$, $k - l = 0, \ldots, 9$) — cases LM1 and LM2.

condition). The definitions of the scalars c_{v_i} for outliers are:

$$\begin{aligned} c_{v_i} &= -A_i \Phi_{i,i-1} \left\{ \prod_{j=l+1}^{i-1} [(I - K_j A_j)\Phi_{j,j-1}] \right\} K_l c_y \quad i > l \\ c_{v_i} &= c_y \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i = l \end{aligned} \tag{5.5}$$

(cf. eq. 3.71), and for slips (cf. eq. 3.72):

$$\begin{aligned} c_{v_i} &= c_y - A_i \Phi_{i,i-1} \sum_{j=l+1}^{i} \left\{ \prod_{k=j}^{i-1} [(I - K_k A_k)\Phi_{k,k-1}] \right\} K_{j-1} c_y \quad i > l \\ c_{v_i} &= c_y \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i = l\,, \end{aligned} \tag{5.6}$$

where the term between braces should be taken as identical to the identity matrix if $j < i - 1$ in (5.5) or $k < i - 1$ in (5.6). After inserting the model parameters given by (5.1) and (5.2) and the results of Table 5.1 into (5.5) and (5.6) one obtains with $c_y = 1$ the values for the scalars c_{v_i} as given in Table 5.4. Indeed the entries of c_{v_i} rapidly decrease with increasing delay, except for the slip-type error in model LM2, which is in accordance with the gradual decrease of the MDBs for slips in model LM2 (cf. Table 5.3). Given that $Q_{v_i}^{-1} = 0.243$ for model LM1 and 0.639 for model LM2, the results of Table 5.4 inserted in (4.22) are identical to the MDBs given in Table 5.3.

From Table 5.3 it follows that a slip can be much better detected by a filter based on model LM2. This can be understood if one recalls that that a larger system noise (as in model LM1) results in a larger gain, and as a consequence the filter related to model LM1 will more easily absorb a slip in the observations.

Summarizing the analysis of the MDBs for models LM1 and LM2, it is concluded that the detectability of the model errors does not improve much with increasing delay. Furthermore the detectability is better for the model with the smaller system noise.

c_{v_i}	LM1		LM2	
delay	outlier	slip	outlier	slip
0	1.0		1.0	
1	−1.25	−0.25	−0.441	0.559
2	−0.181	−0.431	−0.322	0.237
3	0.168	−0.262	−0.222	0.015
4	0.170	−0.092	−0.141	−0.126
5	0.087	−0.005	−0.078	−0.204
6	0.024	0.019	−0.033	−0.237
7	−0.003	0.016	−0.018	−0.255
8	−0.008	0.008	0.018	−0.237
9	−0.005	0.003	0.029	−0.208

Table 5.4: Elements c_{v_i} for outliers and slips — cases LM1 and LM2.

5.4.3 External Reliability

We analyse the external reliability using so-called response graphs and Bias to Noise Ratios (BNRs). The response graphs serve to visualize the effect of the model errors on the state estimates and to explain the characteristics of the BNRs.

Response Graphs

Figs. 5.2 and 5.3 show the impact of an outlier of 1 metre at fix 30 and a slip of 1 metre on the filtered estimator for models LM1 and LM2. The filter quickly reacts to the model error. The instanteneous response at fix 30 corresponds to the values of the gain matrix given in Table. 5.1. It can be seen that the filters are underdamped.

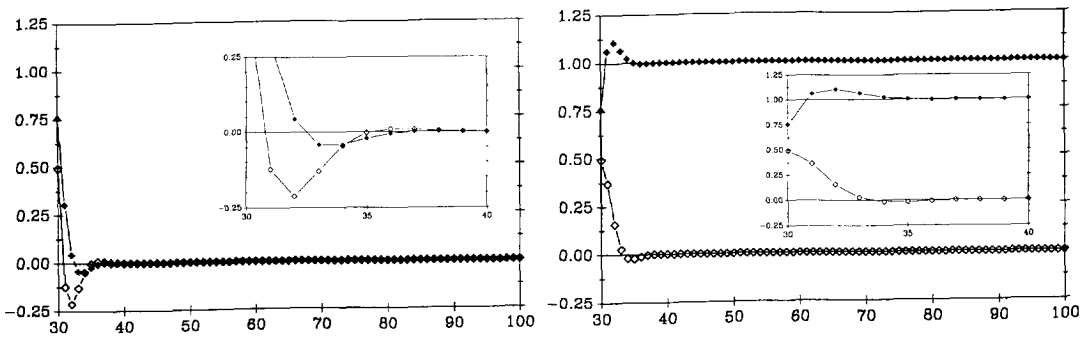


Figure 5.2: ◆ Position [m] and ◇ velocity response [m/s] vs. the fix number due to an outlier (left) and slip (right) of 1m in the coordinate observation — model LM1.

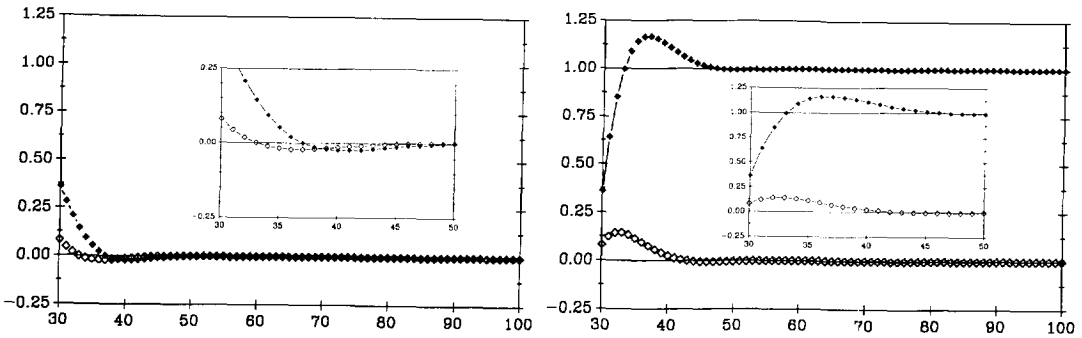


Figure 5.3: ◆ Position [m] and ◇ velocity response [m/s] vs. the fix number due to an outlier (left) and slip (right) of 1m in the coordinate observation — model LM2.

Bias to Noise Ratios

BNRs for models LM1 and LM2 are given in Table 5.5. The BNRs can be directly related to the reponse graphs in Fig. 5.2 because the filter is in steady state. The significance of the bias for outliers is seen to decrease very rapidly with increasing delay, which is in accordance with the response graphs given in Figs. 5.2 and 5.3. The bias for a slip remains significant with increasing delay. This phenomenon is easily understood if one recalls that the slip is fully absorbed in the position state estimate. The overshoot effect visible in Fig. 5.2 is also visible in the values of the BNRs, although part of the decrease of the BNRs for slips is due to the decrease of the bias in the velocity estimate. The values of the BNRs for model LM2 basically display the same behaviour as for model LM1, although the response is somewhat damped. This is a direct consequence of the smaller gain associated with model LM2.

model	LM1		LM2	
delay	outlier	slip	outlier	slip
0	7.29		3.10	
1	2.71	10.4	2.29	4.86
2	1.51	10.8	1.78	6.40
3	0.67	10.6	1.43	7.69
4	0.27	10.4	1.16	8.58
5	0.14	10.2	0.95	9.12
6	0.09	10.1	0.77	9.39
7	0.04	10.0	0.61	9.49
8	0.02	10.0	0.49	9.50
9	0.01	10.0	0.38	9.46

Table 5.5: Square root of the BNR for outliers and slips in observations at fix 90 ($l = 90$, $k - l = 0, \ldots, 9$) — cases LM1 and LM2.

We will now consider the BNR related to various subsets of the state vector, which can be specified by the matrix F^{T} in (4.31). We have computed BNRs related to both

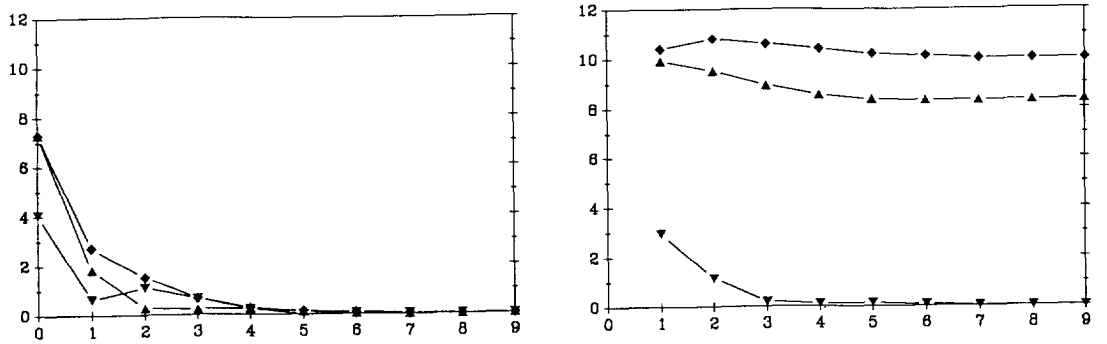


Figure 5.4: sqrt($\lambda_{\hat{x}}$) for outlier (left) and slip (right) in the coordinate observation vs. the delay in fixes; ◆ position and velocity; ▲ position; ▼ velocity — model LM1.

states (Table 5.5) and will now consider (for model LM1) BNRs related only to the position state ($F^{\mathrm{T}} = (1 \quad 0)$) or the velocity state ($F^{\mathrm{T}} = (0 \quad 1)$). The square roots of the various BNRs are plotted in Fig. 5.4. The significance of the bias caused by outliers decreases quickly for every subset, but the impact on the velocity is seen to increase in significance from delay 1 to 2 before it finally decreases (this behaviour is in accordance with the velocity response to an outlier in Fig. 5.2). Furthermore Fig. 5.4 shows that mainly the bias in the position state is significant after a slip in the coordinate observations has occurred (also this behaviour is confirmed by the reponse graph).

Summarizing, the analysis of the external reliability shows that the BNR is not necessarily a monotonic (decreasing) function of the delay (as are the MDBs). Moreover the magnitude of the BNRs is very dependent on the subset of the state vector one actually considers. The significance of the bias for model LM2 is comparable with that of model LM1.

5.4.4 Concluding Remarks

In the current section we analysed the design of the two simple linear models introduced in Section 5.3. The purpose of the rather limited analysis was to make the reader familiar with (some of) the properties of the quality measures. We illustrated some characteristics of the MDBs and compared these with results obtained analytically. The analysis of external reliability was somewhat complicated and use was made of response graphs to clarify the properties of the BNRs. After having analysed the precision, internal and external reliability for models LM1 and LM2, we find that the external reliability of both models is comparable, whereas the precision and internal reliability associated with model LM2 are clearly better. This observation demonstrates that the design of a system (w.r.t. quality) cannot be based on a single quality measure. Although we will see in later sections that some of the findings reported here will be substantiated, the findings of this section cannot be simply generalized to more sophisticated models.

5.5 Design Setup — Navigation Models

5.5.1 Design Configuration

We now consider the design of a hydrographic navigation system. We have chosen to use the configuration of the NAVGRAV-survey which was performed in the North Sea in 1986. The NAVGRAV survey was a combined navigation and sea-gravimetry experiment. Although a major objective of the navigation experiment was to obtain hands-on experience with dynamic GPS positioning, we will not consider the processing of GPS data in this case study. The design computations are performed on the basis of terrestrial radiopositioning systems and dead reckoning systems. The North Sea area is a good example of a region where integration can be considered as it is covered by various radiopositioning systems. Because we are primarily interested in the characteristics of the design parameters and want to establish the feasibility of the design procedure, we have not extensively modelled the actual sensors, but have used somewhat simplified measurement models, which, however, maintain the characteristics of the systems they represent.

We make use of two terrestrial radiopositioning systems, namely Syledis and Hyperfix. Syledis is a UHF system (and thus basically a line-of-sight sytem) which supports operation in range-range, pseudo-range, and so-called combined mode and as a consequence is by itself already a subject of detailed design computations. Hyperfix is a MF hyperbolic radio-positioning system which has replaced the similar Hifix system which was in use at the time of the NAVGRAV-survey. Hyperfix observables can be looked upon as range-differences. Usually these observables are expressed in [lanes] (on the baseline between two transmitter stations one lane corresponds to half a wavelength of the system). A pair of Hyperfix transmitters, between which the range-differences are measured, is called a pattern. For further details on these radiopositioning systems see, e.g., [BAKKER ET AL., 1989]. Syledis and Hyperfix constitute the 'absolute' positioning components of the system. These are supplemented by dead reckoning sensors. We integrate a (gyro)compass and a speed log (in bottom track mode). As we consider both 'absolute' and dead reckoning positioning sensors the results of this design study can be readily extended to other positioning systems.

Design computations are performed at two different locations (cf. Figure 5.5). The area with good Syledis coverage (denoted by 1 in Fig. 5.5) is primarily used to assess the impact of various model assumptions on the quality of the integrated system. The area with poor Syledis coverage (as Syledis is a line-of-sight system only two ranges can be observed in the area denoted by 2 in Fig. 5.5) is used to investigate the impact of various integration strategies on the quality of the positioning solution.

At both design locations a similar trajectory consisting of 100 position fixes at one second intervals and sailed at a constant velocity of 5m/s has been simulated (see Fig. 5.6). The curve has been introduced in the trajectory to detect possible direction dependent effects.

In Table 5.6 the coordinates of the starting points of the trajectories and the sta-

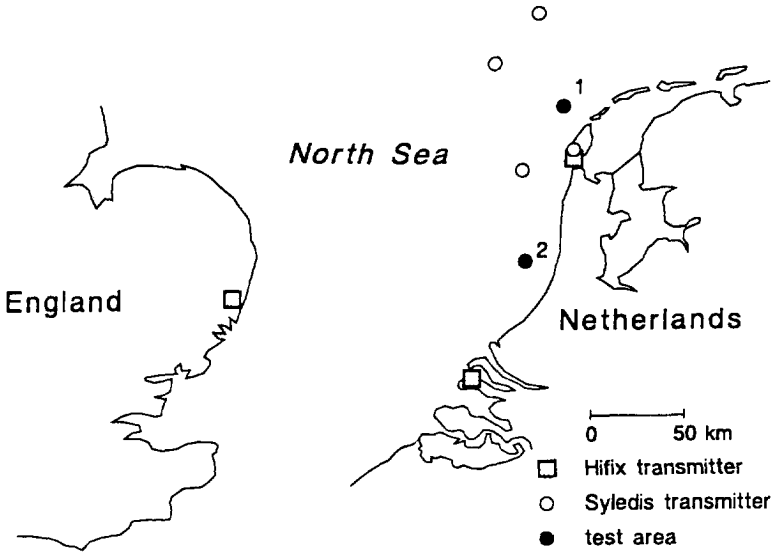


Figure 5.5: Design Area.

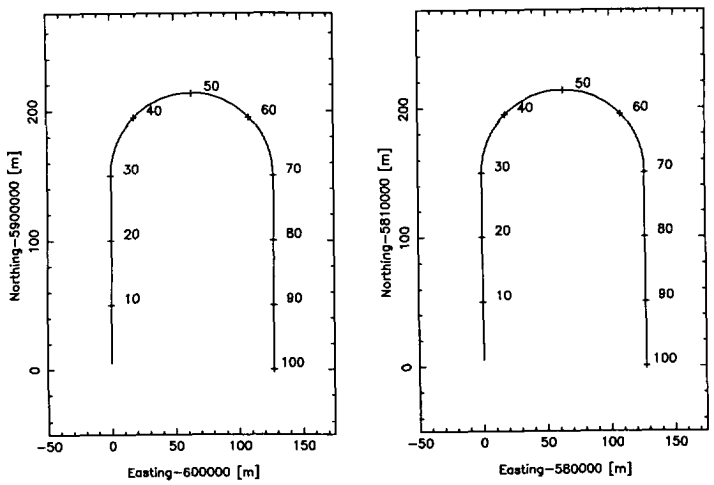


Figure 5.6: Simulated trajectories — area 1 (left); area 2 (right).

tion coordinates are given. For the design computations we use the Hyperfix patterns Renesse-Texel and Renesse-Pakefield.

5.5.2 Model Assumptions

The description of the measurement and dynamic models is based on plane geometry. The values given here are default values and any deviations from these values will be

	Easting [m]	Northing [m]	remarks
Syledis			
NAM K15	558200.30	5903510.80	only used in area 1
Union Q1A	577386.80	5864681.63	
Texel	618295.10	5874865.80	
L4A petro	572513.61	5953710.42	only used in area 1
Hyperfix			
Renesse	551173.93	5731889.84	
Pakefield	413865.59	5811444.80	
Texel	618383.20	5873936.01	
Starting point trajectories			
Area 1	600000.00	5900000.00	
Area 2	580000.00	5810000.00	

Table 5.6: Station coordinates and starting points of the trajectories.

explicitely indicated. The design computations are based on the linearized Kalman filter (LKF), and the measurement model is linearized with respect to the trajectories given in Fig. 5.6.

Measurement model We give the observables and their standard deviations (the observation equations are given in Appendix B).

- Syledis is used in range-range mode; the measurements have a standard deviation of 1.5m and are assumed to be mutually uncorrelated.
- We assume that the Hyperfix system has a wavelength of 140m. The hyperbolic observables have a standard deviation of 0.02lanes and a covariance of 0.0002lanes^2. For design area 2 this corresponds to standard deviations in the patterns Renesse-Texel and Renesse-Pakefield of 1.41m and 2.43m respectively and a covariance between the patterns of 1.71m^2.
- Gyro with a standard deviation of 2.0deg.
- Log in bottom track mode with a standard deviation of 0.2m/s.

The standard deviations of the gyro and log measurements have been chosen in such a way that at a velocity of 5m/s the precision of gyro and log is approximately similar (at 5m/s a standard deviation of 2deg in the gyro measurement for a one-second interval corresponds to a standard deviation of 0.175m/s of the velocity across track).

- The measurement sampling rate of all sensors is 1Hz.

Dynamic model The dynamic model is a constant velocity model with constant instrumental biases. The state model is of polynomial form. Given the state vector elements

- Position (Easting x_1, Northing x_2).
- Velocity (East x_3, North x_4).
- Gyro offset (if a gyro is included, x_5)
- Log bias (if log is included, x_6)

this results in the following linear dynamic model:

$$E\{\begin{pmatrix} \underline{d}_1 \\ \underline{d}_2 \\ \underline{d}_3 \\ \underline{d}_4 \\ \underline{d}_5 \\ \underline{d}_6 \end{pmatrix}_k\} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix}_k - \begin{pmatrix} 1 & 0 & \Delta T & 0 & 0 & 0 \\ 0 & 1 & 0 & \Delta T & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix}_{k-1}$$

with $\Delta T = t_k - t_{k-1}$. Note that although the trajectories used for the design computations contain a curve, accelerations in East and North direction are not included in the state vector.

The system noise related to the various subsets of the state vector is given as follows:

- For the position and velocity states the disturbance is modelled as (decoupled) acceleration in East and North direction with a standard deviation of $\sigma_a = 0.25\text{m/s}^2$. At 5m/s the 180° turn in the trajectory corresponds with an acceleration across track of 0.39m/s^2. As a rule of thumb one often chooses the standard deviation as one-half or one-third of the maximum of the expected disturbance; hence our choice is rather conservative. The level of the system noise cannot be varied at will, because the variance of the disturbances has to match the actual disturbances. Contrary to the case of the linear models this limits the range of variation of the system noise.
- The disturbance of the gyro offset is modelled as a drift rate with a standard deviation of $\sigma_d = 0.05\text{deg/s}$.
- The disturbance of the log bias is modelled as a bias drift with a standard deviation of $\sigma_b = 0.01\text{m/s}^2$.

Based on these values the covariance matrix of the disturbances reads:

$$D\{\begin{pmatrix} \underline{d}_1 \\ \underline{d}_2 \\ \underline{d}_3 \\ \underline{d}_4 \\ \underline{d}_5 \\ \underline{d}_6 \end{pmatrix}\} = \begin{pmatrix} \sigma_a^2 \begin{pmatrix} \frac{\Delta T^4}{3} & 0 & \frac{\Delta T^3}{2} & 0 \\ 0 & \frac{\Delta T^4}{3} & 0 & \frac{\Delta T^3}{2} \\ \frac{\Delta T^3}{2} & 0 & \Delta T^2 & 0 \\ 0 & \frac{\Delta T^3}{2} & 0 & \Delta T^2 \end{pmatrix} & \begin{matrix} 0 \\ \\ \\ \end{matrix} \\ \begin{matrix} 0 & \\ & \end{matrix} & \begin{matrix} \sigma_d^2 \Delta T^2 & 0 \\ 0 & \sigma_b^2 \Delta T^2 \end{matrix} \end{pmatrix}.$$

5.5.3 Design Cases — Navigation Models

The test cases considered are summarized in Tables 5.7 and 5.8 and unless indicated otherwise the model assumptions described in Section 5.5.2 apply. For the cases con-

			observations						notes
case	m	n	SR1	SR2	SR3	SR4	GYRO	LOG	
1a	6	6	x	x	x	x	x	x	
1b	4	4	x	x	x	x			1
1c	4	8	$T = 5$	$T = 5$	$T = 5$	$T = 5$			1,4
1d	6	7	PR1	PR2	PR3	PR4	x	x	2
1e	5	6	x	x	x		x	x	
1f	6	6	x	x	x	x	x	x	3
1g	6	6	x	x	x	x	$\sigma = 4$deg	$\sigma = 0.4$m/s	
1h	6	7	$\sigma = 2.5$m	$\sigma = 2.5$m	$\sigma = 2.5$m	$\sigma = 2.5$m	$T = 5$	x	4
1i	6	11	$T = 5$	$T = 5$	$T = 5$	$T = 5$	$T = 5$	x	4
1j	6	7	$\Delta t = 2$	$\Delta t = 2$	$\Delta t = 2$	$\Delta t = 2$	$T = 5$	x	4
1k	6	7	$\Delta t = 5$	$\Delta t = 5$	$\Delta t = 5$	$\Delta t = 5$	$T = 5$	x	4

m is the number of observations; n is the number of states;
x denotes available measurement; SR denotes Syledis range;
σ denotes standard deviation; PR denotes pseudo-range;
T correlation time in seconds; Δt measurement interval in seconds.

Note 1: no gyro-offset and log-bias states
Note 2: pseudo-ranges instead of ranges; clock bias additional state
s.d. clock bias drift 1.0d-8/s
Note 3: case with low system noise
s.d. acceleration in each coordinate direction 0.1m/s^2
Note 4: augmented state vector due to time-correlated measurements

Table 5.7: Design cases in area 1.

sidered in area 1 model 1a is used as a reference. Cases 1b and 1c allow a comparison with a non-integrated solution. Changes in geometry are studied using cases 1d and 1e where pseudo-ranges are used instead of ranges and only three ranges are measured respectively. Changes in the stochastic model of the disturbances (system noise) are studied in case 1f. The influence of the stochastic model of the observations on integrated system design is more closely studied in cases 1g, 1h, and 1i. Cases 1g and 1h are meant to assess the impact of precision variations between the positioning and dead reckoning sensors. In case 1i (and case 1c for the non-integrated system) the aspect of time-correlated measurements is investigated. Cases 1j and 1k finally serve to gain insight in the effect of different measurement update rates. Overall the design configurations listed in Table 5.7 allow an extensive analysis of the impact of various model assumptions.

In Table 5.8 the cases considered in design area 2 are given. Basically all possible scenarios of integration are considered from dead reckoning only (case 2d) to a fully integrated system with Syledis, Hyperfix, and dead reckoning (case 2g). In order not to obscure the effects of integration all other design parameters are left unchanged.

The analyses in area 1 and area 2 combined should yield an extensive insight in the

case	m	n	observations SR2	SR3	HY1	HY2	GYRO	LOG	notes
2a	2	4	x	x					1
2b	2	4			x	x			1
2c	4	4	x	x	x	x			1
2d	2	4					x	x	1
2e	4	6	x	x			x	x	
2f	4	6			x	x	x	x	
2g	6	6	x	x	x	x	x	x	
x denotes available measurement; SR denotes Syledis range; HY denotes Hyperfix pattern.									
Note 1: no gyro-offset and log-bias states									

Table 5.8: Design cases in area 2.

design of integrated navigation system for hydrographic purposes.

5.5.4 Processing of Time-Correlated Measurements

In some design cases we have to deal with exponentially time correlated measurements. Anticipating the analysis of the results, we indicate how the variance of the exponentially time-correlated noise is chosen and how it can be compared to the variance of uncorrelated observations. Exponentially time-correlated noise can be modelled as

$$\underline{e}_{k+1} = \alpha \underline{e}_k + \underline{n}_k, \tag{5.7}$$

where $\underline{e}_k$ is the measurement noise at time k, $\underline{n}_k$ is additional white noise, and $\alpha = \exp(-\Delta T/T)$ with $\Delta T = t_{k+1} - t_k$ and T the correlation time of the measurement noise. In the filter model the measurement noise is modelled as an additional state (one for every time-correlated measurement) and $\underline{n}$ as its corresponding system noise (cf. Section 2.6). The variance $D\{\underline{n}_k\}$ is chosen in such a way that

$$D\{\underline{e}_{k+1}\} = D\{\alpha \underline{e}_k + \underline{n}_k\} = R, \tag{5.8}$$

where R is the variance of the measurement in the uncorrelated case. If one assumes that $\underline{e}_k$ and $\underline{n}_k$ are uncorrelated, it follows that

$$D\{\underline{n}_k\} = (1 - \exp(\frac{-2\Delta T}{T}))R \;=\; (1 - \alpha^2)R. \tag{5.9}$$

Using (5.7) one finds for an arbitrary $\underline{e}_k$ that

$$\underline{e}_k = \alpha^{k-1}\underline{e}_1 + \sum_{i=1}^{k-1} \alpha^{i-1}\underline{n}_{k-i}\,. \tag{5.10}$$

For k large and $\alpha \in (0,1)$ the variance of $\underline{e}_k$ can be approximated by

$$D\{\underline{e}_k\} \approx \left(\frac{1}{1-\alpha}\right)^2 (1-\alpha^2)R \;=\; \left(\frac{1+\alpha}{1-\alpha}\right) R, \tag{5.11}$$

where use has been made of (5.9), the identity $\sum_{i=0}^{\infty} \alpha^i = \frac{1}{1-\alpha}$, and the fact that $\alpha^{k-1} \approx 0$. Based on (5.8) it follows from (5.11) that the processing of exponentially time-correlated measurements partly corresponds to the processing the data with a larger variance (note we do not consider the covariance between the various $\underline{e}_i$-terms).

5.6 Design Results — Navigation Models

In this section we will successively analyse the precision, the internal reliability, the external reliability, and all quality aspects combined for the navigation models introduced in Section 5.5. We have chosen not to strive for completeness, and limit ourselves to the highlights of the design studies. Firstly, so many variations in the design of a navigation system can be conceived that completeness can never be attained. Secondly, a presentation of all the results rather than the highlights would render this section unnecessarily long. To provide a reference one particular case is fully documented (case 1a in design area 1; cf. Table 5.7). We mainly try to provide an insight in the aspects that influence the quality measures in the design process and point out the effect that certain variations of design parameters have on relative behaviour of the quality measures. We have chosen to present most results based on a single design aspect and, if possible, in graphical form.

The analysis of internal and external reliability is limited to alternative hypotheses related to range and gyro observations, unmodelled acceleration along the sailing direction, and gyro drift. During a preliminary analysis of the results we found that the results pertaining to ranges and range-differences; gyro and log observations; (unmodelled) acceleration in along and across direction; and gyro offset and log bias were very similar. The reliability analyses for ranges, gyro, acceleration along track, and gyro drift therefore also provide insight into the reliability of range-differences, log, acceleration across track, and log bias respectively.

In the following we will often present results pertaining to fix 90 or 100. The graphs concerning MDBs (internal reliability) and (the square root of the) BNRs (external reliability) are plotted for model errors starting at fix 90 (i.e. $l = 90$). The value of the MDBs and BNRs is plotted with respect to the delay of the test. The hypotheses are considered for delays 0 to 9 (i.e. $k = 90, \ldots, 99$). Contrary to the situation of the linear models, the navigation models which incorporate dead reckoning sensors do not attain a steady state condition. The variations in the precision of the state estimators are, however, so small that an analysis at a particular fix is characteristic for the behaviour of a quality measure along the trajectory (except for the first 30 fixes, where the starting transient dominates the results). The variances of the initial state have been chosen as follows: $10^6 \mathrm{m}^2$ for the Easting and Northing coordinate; $10^4 (\mathrm{m/s})^2$ for the velocity in East and North direction; $1 \mathrm{rad}^2$ for the gyro offset; and $10 (\mathrm{m/s})^2$ for the log bias.

5.6.1 Precision

The results of the precision analysis for models 1a-1k are given in Table 5.9 and for models 2a-2g in Table 5.10 and Fig. 5.7. The precision of the filtered state vector elements is (after the initial transient of about 20-30 fixes) approximately constant for the whole trajectory, but the steady-state condition is not reached for the models with dead reckoning sensors (the precision of instrumental bias states keeps improving slowly). For model 2d (dead reckoning only) the standard deviation of the filtered coordinates is by definition identical to the initial standard deviation (cf. Table 5.10), as dead reckoning only provides relative position information.

We will, for ease of survey, discuss the impact of various design aspects on the precision of the filtered state estimators separately for the position components, the velocity components, and the instrumental biases. From Tables 5.9 and 5.10 (assuming that the measurement update interval is one second) one can, however, derive the rules of thumb that an upperbound for the precision of the filtered coordinates is given by the precision of the positioning system (Syledis, Hyperfix), and similarly that an upperbound for the precision of the filtered velocity components is provided by the measurement precision of the dead reckoning sensors. The precision is hardly influenced by the direction changes in the trajectory, because the precision of the gyro and log measurements closely corresponds.

case	Easting [m]	Northing [m]	velocity East [m/s]	velocity North [m/s]	gyro offset [deg]	log bias [m/s]	semi-major axis error ellipse [m]
1a	0.47	0.48	0.15	0.17	0.36	0.050	0.51
1b	0.79	0.74	0.40	0.39			0.85
1c	1.12	1.04	0.38	0.37			1.20
1d	0.54	0.51	0.15	0.17	0.36	0.051	0.59
1e	0.47	0.55	0.15	0.17	0.36	0.051	0.55
1f	0.45	0.46	0.12	0.13	0.35	0.049	0.49
1g	0.58	0.57	0.24	0.26	0.51	0.066	0.62
1h	0.82	0.67	0.16	0.17	0.71	0.053	0.84
1i	0.97	0.84	0.16	0.17	0.71	0.055	1.00
1j	0.70	0.58	0.16	0.17	0.73	0.052	0.72
	0.78	0.63	0.16	0.17	0.74	0.053	0.80
1k	0.86	0.72	0.16	0.17	0.75	0.054	0.89
	1.21	0.94	0.18	0.18	0.77	0.058	1.24

Table 5.9: Standard deviation of the filtered state estimators and the semi-major axis of the error ellipse at fix 100 (area 1). (For models 1j and 1k the best and worst situations are given.)

The precision of the estimators of the position components (i.e. Easting and Northing) is significantly influenced by the following:

case	Easting [m]	Northing [m]	velocity East [m/s]	velocity North [m/s]	gyro offset [deg]	log bias [m/s]	semi-major axis error ellipse [m]
2a	2.00	0.90	0.55	0.41			2.05
2b	1.21	0.86	0.45	0.39			1.40
2c	1.00	0.66	0.43	0.36			1.12
2d	n/a	n/a	0.15	0.17			n/a
2e	0.96	0.55	0.15	0.17	0.43	0.051	0.99
2f	0.64	0.51	0.15	0.17	0.38	0.050	0.74
2g	0.56	0.41	0.15	0.17	0.37	0.049	0.62

Table 5.10: Standard deviation of the filtered state estimators and the semi-major axis of the error ellipse at fix 100 (area 2).

- The level of integration of the navigation system. Especially the integration of dead reckoning sensors improves the precision of the filtered coordinates considerably (cf. Table 5.10 and Fig. 5.7). The better precision is largely due to the fact that the precision of the velocity components is improved by the dead reckoning observations. Consequently less uncertainty is propagated into the coordinate states at the time update of the filter. If more than one 'absolute' positioning system is used, the precision of the filtered coordinates is predominantly determined by the 'best' system (For instance, the integration of the Hyperfix with the Syledis measurements does not lead to major improvements (cf. 2c $\leftrightarrow$ 2b), because the former mainly determines the precision of the position fix).

- The standard deviations of the range and dead reckoning measurements. These directly influence the precision of the position (cf. 1h $\leftrightarrow$ 1a and 1g $\leftrightarrow$ 1a). A variation of the precision of the range measurements has the largest effect.

- The correlation in time of the range measurements. Exponentially time-correlated measurement noise results in larger standard deviations of the filtered coordinates (cf. 1c $\leftrightarrow$ 1b and 1i $\leftrightarrow$ 1a), which is in accordance with our discussion of the effect of processing time-correlated measurements in Section 5.5.4.

- A larger measurement update interval (or lower measurement rate) of the range measurements. This results in a higher standard deviation of the coordinates (cf. 1j/1k $\leftrightarrow$ 1a) because less observations are used.

The precision of the filtered coordinates is only moderately influenced by the following:

- A change in the variance matrix of the disturbances (acceleration along and across track) (cf. 1f $\leftrightarrow$ 1a.). For a given trajectory one can only marginally alter the variance matrix of the disturbances. It seems of little use to vary the level of system noise artificially in the design stage if only small changes in its magnitude

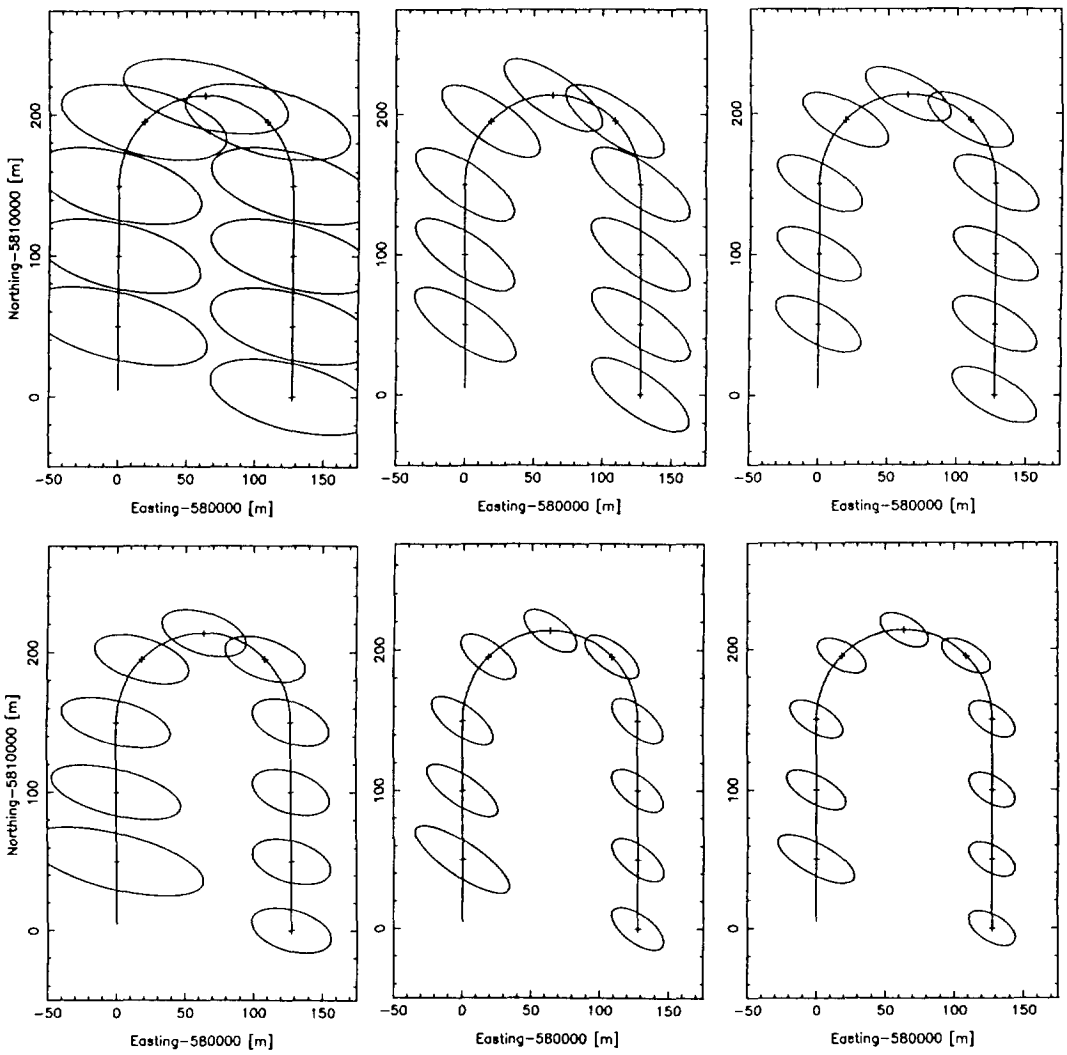


Figure 5.7: Point error ellipses (plotted every 10 fixes) for various levels of integration. Top: cases 2a (left), 2b (middle), 2c (right); Bottom: 2e (left), 2f (middle), 2g (right); Scale ellipses 30:1.

are likely. This does not alter the fact that the precision of the position states may be significantly affected if the system noise changes considerably (cf. the design study of the linear model in Section 5.4).

- Slight changes in the geometry of the configuration (cf. 1d/1e $\leftrightarrow$ 1a). If the redundancy is large enough deleting one range observation (1e) or processing the ranges as pseudo-ranges (1d) has only a small impact.

The precision of the filtered velocity states is significantly influenced by the following:

- The integration of dead reckoning sensors into the navigation system (cf. 1b $\leftrightarrow$ 1a and 2d-2g $\leftrightarrow$ 2a-2c). Dead reckoning sensors provide direct measurements of the velocity components and thereby provide an upperbound for the attainable precision of the filtered velocity states.

- The standard deviation of the dead reckoning observations (cf. 1g $\leftrightarrow$ 1a).

- A change of the variance matrix of the disturbances (in particular related to the accelerations, cf. 1f $\leftrightarrow$ 1a). In the time update of the filter the uncertainties in the accelerations are directly propagated into the covariances of the velocities, and therefore a lower system noise results in smaller standard deviations of the filtered velocities.

The precision of the filtered velocity states is only moderately influenced by the following:

- A variation in the standard deviation of the range measurements (cf. 1h $\leftrightarrow$ 1a). When dead reckoning sensors are part of the system, the precision of the filtered velocity states is mainly determined by the precision of the dead reckoning observations.

- Time correlated range measurements (cf. 1i $\leftrightarrow$ 1a).

- A change in the measurement update interval of the range measurements (cf. 1j/1k $\leftrightarrow$ 1a).

The precision of the filtered estimates of the instrumental states has not been extensively analysed, but from Table 5.9 one can see that the precision is primarily dependent on the following:

- The correlation in time of the dead reckoning sensors (we have only considered time-correlated gyro measurements). The standard deviation of the gyro offset increases for exponentially time-correlated gyro measurements (cf. 1i $\leftrightarrow$ 1a).

- The standard deviations of the dead reckoning observations (cf. 1g $\leftrightarrow$ 1a). The instrumental biases are an intrinsic part of the dynamic model of the dead reckoning sensors, and hence the standard deviations of the dead reckoning observations will directly influence the precision of the instrument states.

Summary of the Analysis of Precision

The precision of the complete filtered state of the integrated navigation system clearly benefits most of the integration of 'absolute' positioning systems and dead reckoning

systems. Furthermore, the precision obviously improves by using more precise observables. These findings comply with Section 4.3. Time-correlated measurements and large measurement intervals have a detrimental effect on the precision, the former because under our model assumptions exponentially time-correlated noise corresponds to processing the data with a larger observation noise (cf. Section 5.5.4), and the latter because lowering the sampling rate simply means that one uses less observations to estimate the state.

5.6.2 Internal Reliability

The aspect of internal reliability of the navigation models is analysed by investigating the detectability of model errors (making use of the MDBs (4.22)) and the separability between hypotheses by means of the correlation coefficient (4.23).

Minimal Detectable Biases

In the following we will successively consider the MDBs for outliers and slips in the ranges (which are also representative for range-differences), outliers and slips in the gyro measurements (which are also representative for log measurements), and slips in the state vector elements. We are not so much interested in the magnitude of the MDBs, but rather in their (relative) response to changes in the design parameters. We make a distinction between major and minor effects which influence the MDBs and will usually consider the MDBs of outliers and slips separately. As a reference we provide the results for case 1a in Tables 5.11 and 5.12.

	Outliers				Slips			
delay	range SR2 [m]	range SR3 [m]	gyro [deg]	log [m/s]	range SR2 [m]	range SR3 [m]	gyro [deg]	log [m/s]
0	6.59	6.48	15.87	1.46				
1	6.51	6.42	12.66	1.20	4.91	4.75	14.38	1.30
2	6.46	6.38	12.38	1.17	4.20	3.99	12.72	1.14
3	6.43	6.36	12.32	1.16	3.79	3.53	10.94	0.980
4	6.41	6.34	12.20	1.15	3.51	3.21	9.34	0.844
5	6.39	6.33	12.20	1.15	3.30	2.98	8.08	0.737
6	6.38	6.33	12.15	1.15	3.13	2.78	7.10	0.656
7	6.37	6.32	12.15	1.15	2.99	2.63	6.36	0.592
8	6.37	6.32	12.15	1.15	2.87	2.50	5.73	0.543
9	6.36	6.32	12.15	1.15	2.77	2.38	5.25	0.504

Table 5.11: Minimal Detectable Biases of outliers and slips in the observations at fix 90 ($l = 90, k - l = 0, \ldots, 9$); (SR2 and SR3 are the ranges with the largest and smallest MDBs respectively.) — case 1a.

delay	acceleration along track [m/s^2]	acceleration across track [m/s^2]	gyro drift [deg/s]	log bias drift [m/s^2]
1	0.904	0.869	9.34	0.841
2	0.688	0.670	6.53	0.576
3	0.574	0.562	4.72	0.414
4	0.502	0.494	3.48	0.305
5	0.451	0.446	2.61	0.231
6	0.413	0.409	2.02	0.180
7	0.384	0.380	1.59	0.144
8	0.360	0.357	1.28	0.117
9	0.340	0.337	1.06	0.098

Table 5.12: Minimal Detectable Biases of state related hypotheses at fix 90 ($l = 90$, $k - l = 0, \ldots, 9$) — case 1a.

MDBs of Ranges

The characteristics of the MDBs of ranges can be compared (to some extent) to those of the MDBs of the coordinate observations in the linear models, for both provide the 'absolute' position information in a constant velocity model (cf. Section 5.4). Indeed one finds that for the cases where no dead reckoning system is integrated (2a/2c) the characteristics of the MDBs of the ranges agree with those of the coordinate observations in the linear model. The agreement is lost, however, as soon as dead reckoning sensors are integrated (cf. Fig. 5.8). If dead reckoning sensors are integrated the MDBs related to outliers in the ranges do not significantly improve (i.e. get smaller) with increasing delay, whereas the improvement concerning slips is more pronounced (cf. Fig. 5.8 (2e/2g) and Table 5.11). We have found that the MDBs related to outliers in range observations are affected by the following:

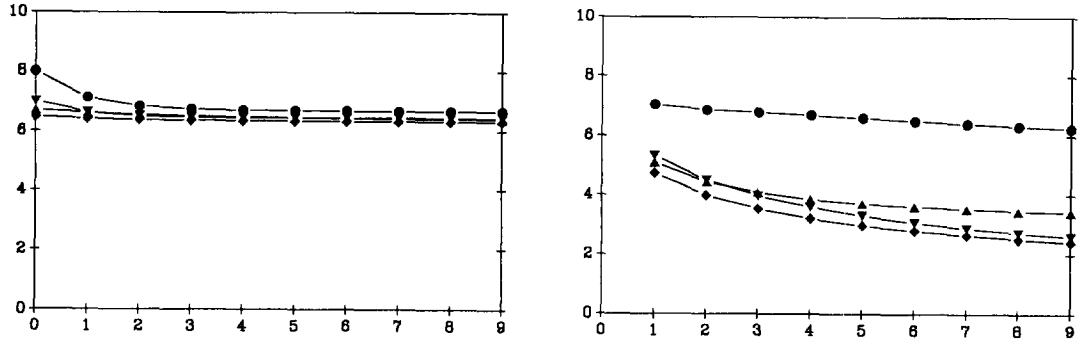


Figure 5.8: MDBs [m] concerning an outlier (left) and a slip (right) in Syledis range SR1 for various levels of integration; • 2a, ▼ 2c, ▲ 2e, ◆ 2g.

- The standard deviation of the observations. This was already observed in Section 4.4.3. The MDBs for the outliers in model 1h (σ_r = 2.5m) are, when compared to model 1a with σ_r = 1.5m, almost exactly 2.5/1.5 times as large as the MDBs of model 1a.

- Correlation in time of the observations (cf. Fig. 5.9). For outliers one sees a large improvement of the MDB at delay 1. Also the MDBs are smaller than those of the reference case (1a). The impact of seemingly larger observation noise (cf. Section 5.5.4) is more than compensated by the fact that the (observation) error states are part of the augmented state and are thereby linked in time, so that it is less likely that an outlier occurs unnoticed.

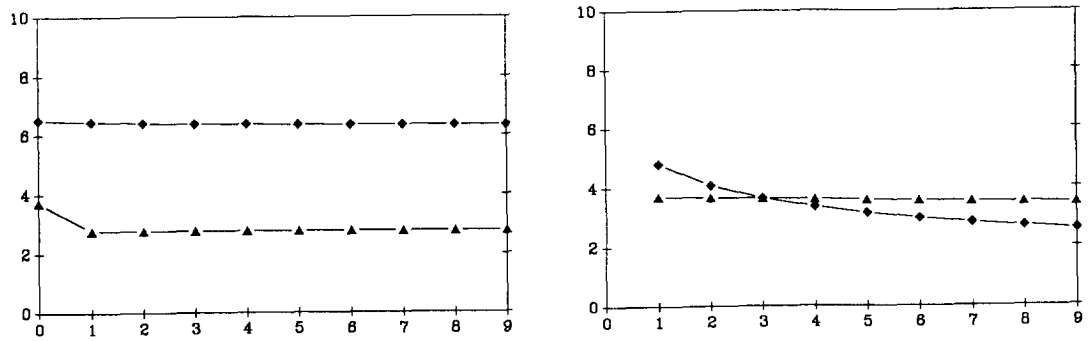


Figure 5.9: MDBs [m] concerning an outlier (left) and a slip (right) in Syledis range SR1; ▲ 1i, ◆ 1a.

The MDBs of outliers in the range observations are rather insensitive to the following:

- Integration with dead reckoning sensors. The assumption that integration always leads to *considerably* lower MDBs is apparently not true: MDBs of alternative hypotheses related to outliers in the ranges only slightly improve (cf. Fig. 5.8).

- A change in the variance matrix of the system noise. In practice the system noise for navigation systems cannot be chosen at will, but has to correspond with the actual disturbances, and thus varies within bounds. In Section 5.4 we saw, however, that for the linear model a major change in Q has an impact on the MDBs of the coordinate observations.

- Changes in geometry. If we use pseudo-ranges instead of ranges this leads to 10% larger MDBs. Three instead of four ranges only result in marginally larger MDBs, but a dead reckoning system integrated with four pseudoranges or three ranges (rather than four) is still highly redundant.

The MDBs concerning slips in the range observations become smaller with increasing delay (cf. Table 5.11). Integration with a second positioning system and/or dead

reckoning systems leads to considerably lower MDBs for slip-type hypotheses of ranges (cf. Fig. 5.8). Only for the case of exponentially time-correlated ranges (cf. Fig 5.9) one finds a completely different behaviour (the MDBs remain constant with increasing delay). Lower sample rates (cases 1j and 1k) cause a seemingly different behaviour of the MDBs (cf. Fig. 5.10), but one should take into account that a new range measurement is not available until delay 2 (case 1j) or delay 5 (case 1k).

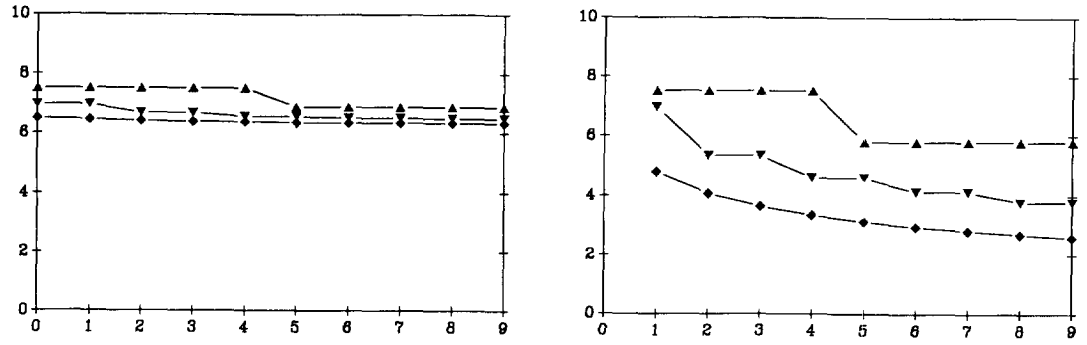


Figure 5.10: MDBs [m] concerning an outlier (left) and a slip (right) in Syledis range SR1 for different sample rates; ▲ 1k, ▼ 1j, ◆ 1a.

Summarizing, one can state that for ranges the MDBs decrease significantly with increasing delay for slips and hardly (except for delay 1) for outliers (only for exponentially time-correlated ranges we found a deviant behaviour). Therefore it is recommended to use different window sizes for tests related to outliers and slips in the ranges. It seems that tests for outliers should use a short window length (1 or 2), while for slips a somewhat larger window length (3 to 4) is appropriate. The magnitude of the MDB is primarily dependent on the standard deviation of the range measurement.

MDBs of Gyro Observations

The MDBs related to outliers of gyro observations generally show a major improvement going from delay 0 to delay 1 (cf. Fig. 5.11 and Table 5.11). For larger delays the improvement is insignificant. (At a speed of 5m/s an angle of 15deg in a one second interval corresponds to a velocity of 1.34m/s across track.) In Figs. 5.11 to 5.14 we have depicted the response of the MDBs of gyro observations to integration, a change in the covariance matrix of disturbances, a change in the variance of the observations, and exponentially time-correlated observations.

The MDBs related to outliers in the gyro observations are influenced by the following:

- A change in the variance matrix of the system noise (cf. Fig. 5.12). If the variance of the unmodelled accelerations is decreased the velocity at the next measurement update can be better predicted and smaller model errors in the gyro observations can be detected.

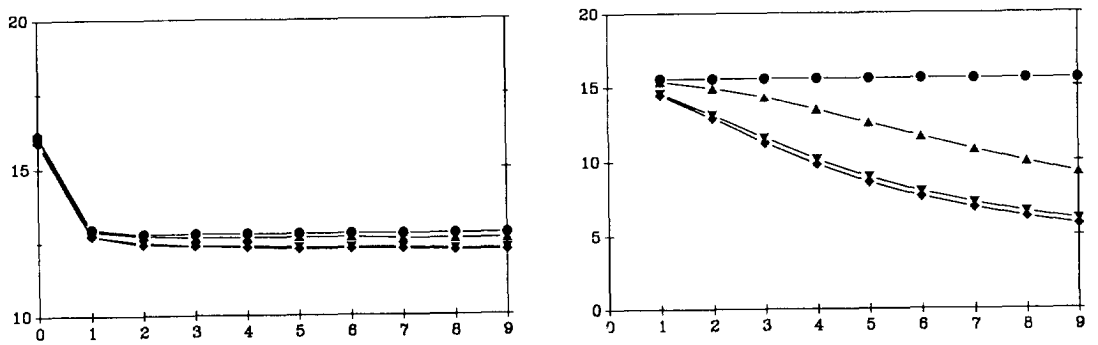


Figure 5.11: MDBs [deg] concerning an outlier (left) and a slip (right) in the gyro observation for various levels of integration; • 2d, ▲ 2e, ▼ 2f, ◆ 2g.

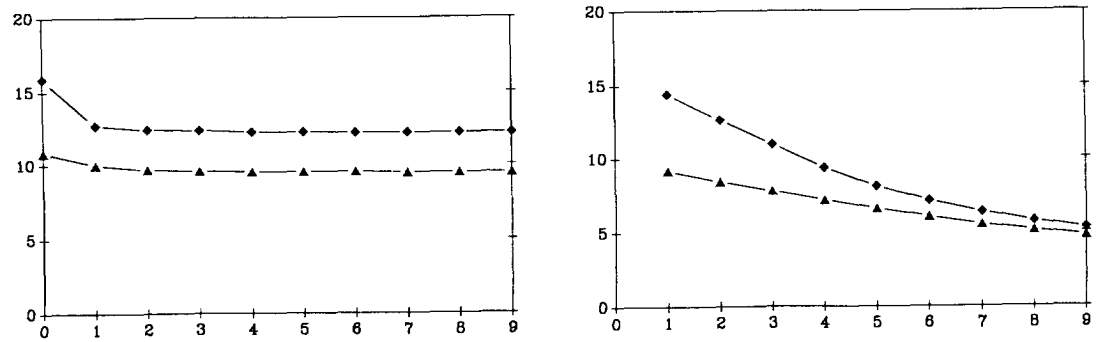


Figure 5.12: MDBs [deg] concerning an outlier (left) and a slip (right) in the gyro observation for varying system noise; ▲ 1f, ◆ 1a.

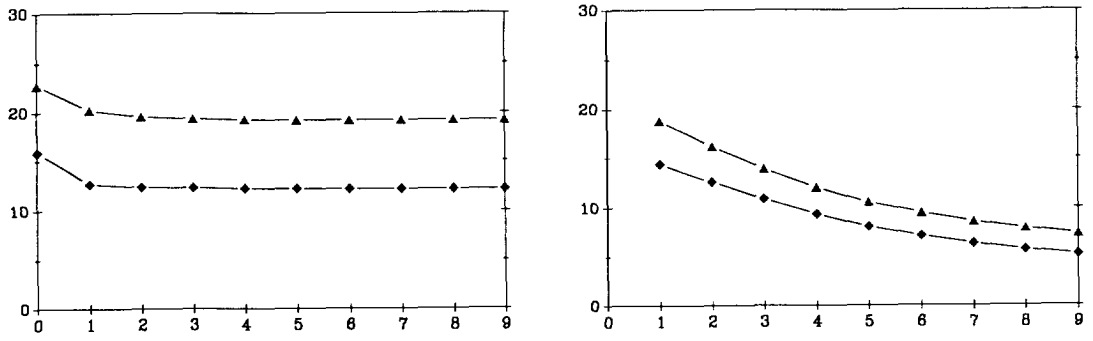


Figure 5.13: MDBs [deg] concerning an outlier (left) and a slip (right) in the gyro observation for varying measurement precision; ▲ 1g, ◆ 1a.

- The standard deviation of the gyro observations (cf. Fig. 5.13). A larger standard deviation (cf. case 1g where the standard deviation of the gyro observation is twice as large as for case 1a) results in larger MDBs, but the increase in size of

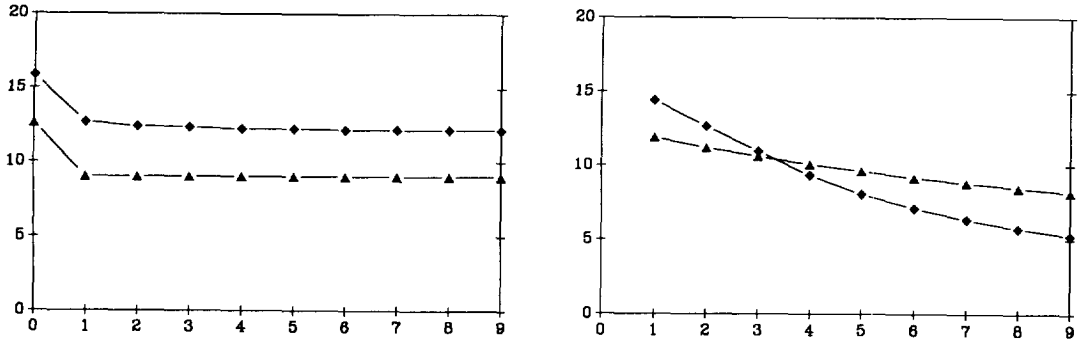


Figure 5.14: MDBs [deg] concerning an outlier (left) and a slip (right) in the gyro observation; ▲ 1i (*time-correlated observations*), ◆ 1a.

the MDB is not proportional to the change in standard deviation (actually it is somewhat less).

- Exponentially time-correlated gyro measurements (cf. Fig. 5.14, and the discussion in Section 5.5.4).

From Fig. 5.11 one sees that integration does not lead to smaller MDBs.

The MDBs of slips in the gyro observations decrease *significantly* with increasing delay. The size of the MDBs is directly related to the standard deviation of the observation (cf. Fig. 5.13). The improvement obtained going from delay 1 to 9 is approximately threefold. Note that if no positioning system is integrated the MDBs are constant with increasing delay (Fig. 5.11, case 2d). Integration clearly results in smaller MDBs. The MDBs of slips in the dead reckoning sensors are additionally influenced by the following:

- A change in the variance matrix of the system noise (cf. Fig. 5.12). Smaller system noise causes a proportional decrease of the MDBs with increasing delay (although the improvement is less pronounced for larger delays).

- Correlation in time of the gyro observations (cf. Fig. 5.14).

The improvement with increasing delay of the MDBs for slips is less pronounced for the cases with larger variances for the ranges (this also encompasses the case of exponentially time correlated ranges) and the cases with a reduced measurement rate of the ranges. This indicates that the range measurements significantly contribute to the detection of slips in the gyro observations.

Summarizing, one finds for tests related to model errors in the gyro observations that (compared to the local tests) the MDBs for outliers get smaller for tests with a delay of one, but then remain constant with increasing delay, while the MDBs for slips decrease with increasing delay. This indicates that window lengths of at least 2 should be used for tests for outliers, whereas the detection power of the tests for slips in the observation considerably benefits from larger window lengths. The size of the MDBs

is primarily dependent on the stochastic model of the observations, and somewhat less on the stochastic model of the disturbances.

MDBs of Unmodelled Acceleration and Instrumental Biases

The MDBs related to the model errors of unmodelled acceleration and instrumental biases are only briefly analysed. The MDBs related to the hypotheses concerning acceleration decrease with increasing window length (cf. Table 5.12). Where a gyrocompass and log are available these seem to provide an upper level for the size of the MDBs of the accelerations. The magnitude of the MDBs is affected by the system noise (cf. Fig. 5.15). For cases where no dead reckoning sensors are available the major improvement in the MDB is found in going from delay 1 to 2.

The MDBs for instrumental biases clearly show a spectacular improvement with increasing delays. Going from delay 1 to 9 a ninefold improvement can be seen. The MDBs of the hypotheses related to the instrumental biases are influenced by the following:

- A change in the variance matrix of the system noise.
- Sample interval of the range measurements. Larger sample intervals lead to considerably larger MDBs.
- The standard deviation of the dead reckoning measurements.

The impact of a change in system noise and measurement precision is depicted in Fig. 5.15.

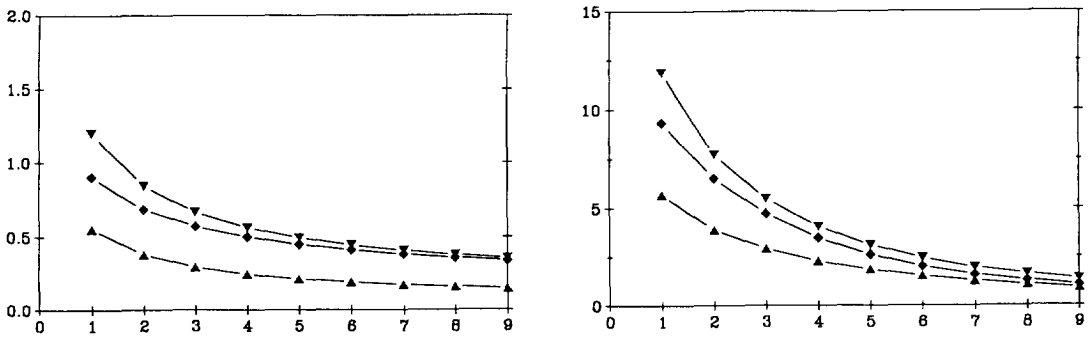


Figure 5.15: MDBs concerning acceleration along [m/s^2] (left) and gyro offset (right) [deg/s]; ▲ 1f, ▼ 1g, ◆ 1a.

Summarizing, it is found that the MDBs related to unmodelled accelerations and instrumental biases clearly benefit from increased window lengths.

Summary of Results of the Analysis of the MDBs

We have analysed the MDBs for outliers and slips in the range and gyro observations. The characteristics of the MDBs of the range and gyro observations are also valid for the MDBs of range-difference and log observations respectively. The analysis of the MDBs has shown that the MDBs always become smaller with increasing delay. At the same time it was seen that the decrease in size is not always significant. The magnitude of the MDBs related to model errors of the observations is directly related to the precision of the measurements. In design area 1 it is found that the MDBs related to the local tests (delay 0) of the ranges are (generally) in the range of 4-5 times the standard deviation of the range measurements. For the MDBs related to the local tests concerning the dead reckoning sensors the size is in the range of 5-7 times the standard deviation of the measurement. Integration mainly decreases the MDBs of the slip-type hypotheses. Overall we can tentatively conclude that for tests related to outliers small window lengths (1 or 2) are sufficient, whereas the detection power of tests for slips increases with increasing window length. The advice on the choice of window lengths is merely indicative, because if the MDBs associated with the local tests are smaller than the size of the likely model errors, local testing will probably be sufficient, whereas in other cases MDBs with a delay of, say, 10 might still be too large to detect a likely model error.

Correlation between Hypotheses

In choosing a testing strategy it is of importance to know how different alternative hypotheses are correlated. We have investigated the correlation of hypotheses referring to an identical delay $(k-l)$ and time of computation (k). It would also be interesting to compute the correlation between hypotheses related to different delays. One would expect, for example, a high correlation between hypotheses concerning slips in observations for consecutive delays. Unfortunately our software (based on the recursive Kalman filter) does not allow the computation of these correlation coefficients. We have analysed the correlation coefficient (4.23) for case 1a. In Figure 5.16 the results for delay 1 and 9 are given.

From Fig. 5.16 the following conclusions (which refer to case 1a) can be drawn:

- For all observations the correlation between the outlier and slip hypotheses for that particular channel slowly decrease with increasing delay.

- The correlation between hypotheses concerning a slip in gyro or log observations and the instrumental parameters gyro offset and log bias is very significant for all delays.

- For small delays there exists a significant correlation between hypotheses concerning unmodelled accelerations and instrumental biases and slips in dead reckoning sensors. The correlation between the acceleration hypotheses and slip hypotheses in the dead reckoning observations decreases very quickly with increasing delay,

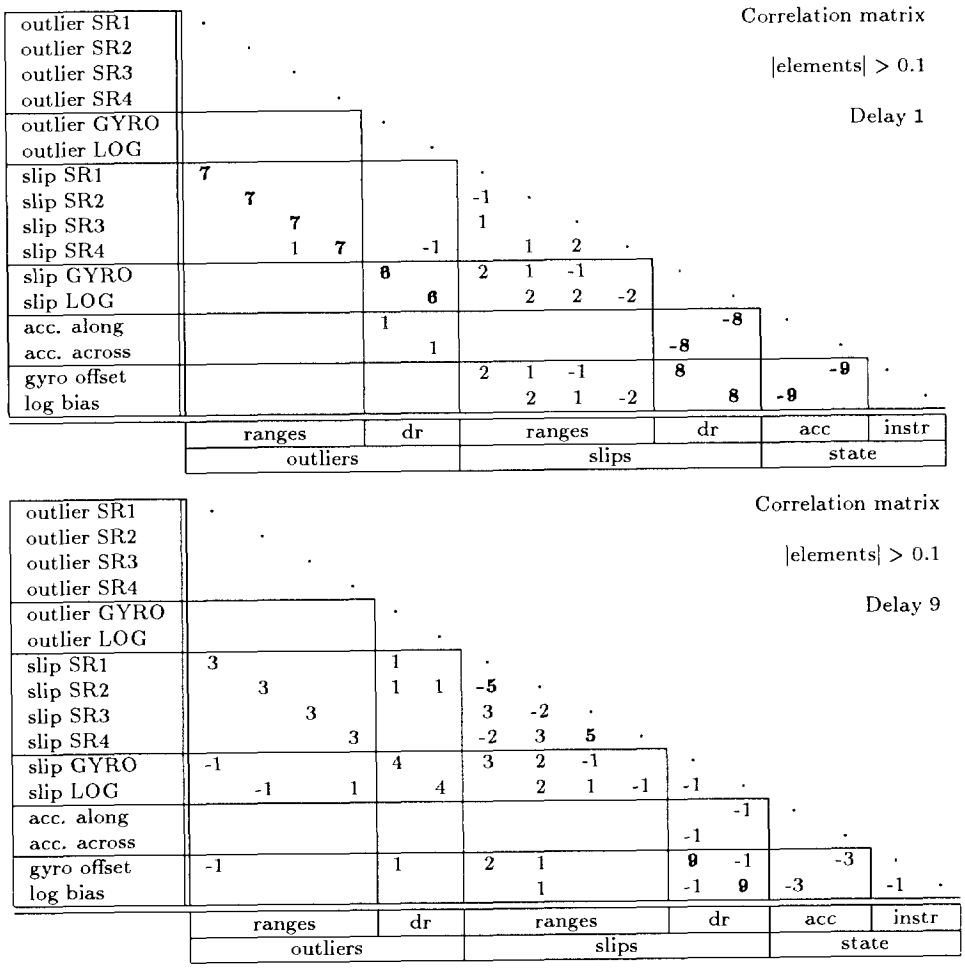


Figure 5.16: Correlation coefficients ($\times 10$ and rounded to the nearest integer) for delays 1 and 9 — case 1a

whereas the correlation between the acceleration hypotheses and the hypotheses concerning instrumental biases decreases gradually.

- The correlation between hypotheses concerning slips in the range observations and the dead reckoning observations tends to become larger with increasing delay.

The analysis of the correlation shows that for case 1a the correlation between a few hypotheses is significant. For small delays (0-2) outlier and slip-type hypotheses concerning the observations are strongly correlated and the correlation between the accelerations components and slips in the dead reckoning sensors and instrumental biases is significant. At a larger delay only the correlation between slips in the dead reckoning observations and the instrumental biases is still significant. The large correlation should be taken into account in the design of a testing procedure for an integrated system.

5.6.3 External Reliability

The impact of model errors on the filtered state is analysed using (elements of) the bias vector $\nabla\hat{\boldsymbol{x}}_{k|k}$ and the bias to noise ratio (BNR) (4.29). We start the analysis of external reliability by assessing the impact of model errors on the position and velocity estimates in so-called response graphs. Although the analysis of the response graphs is limited to a few cases, it provides a direct insight into the impact of model errors on the state estimates. Next we investigate the bias to noise ratios which are a measure of the significance of the biases $\nabla\hat{\boldsymbol{x}}_{k|k}$ with respect to the precision of the filtered estimates. If one assumes that the variance of the filtered states (the noise component of the BNR) is constant (which is more or less a valid assumption), the findings of the paragraph on response graphs can directly be used to explain the characteristics of the BNRs. In our analysis of the BNRs we are not so much interested in the magnitude of the BNRs, but rather in the effect various design options have on the characteristics of the BNRs.

Response Graphs

The response graphs are used to investigate (elements of) the bias vector $\nabla\hat{\boldsymbol{x}}_{k|k}$. If $\nabla E_{k|k}$ and $\nabla N_{k|k}$ denote the impact of a model error on the filtered Easting and Northing coordinate respectively, the impact of the model error on the position is defined as $(\nabla E^2_{k|k} + \nabla N^2_{k|k})^{1/2}$. In a similar way the impact on velocity is defined as $(\nabla V^2_{E\ k|k} + \nabla V^2_{N\ k|k})^{1/2}$. The response graphs directly visualize the effect of model errors on the state vector. A disadvantage of the evaluation of (elements of) $\nabla\hat{\boldsymbol{x}}_{k|k}$ is that many computations are required to obtain an overall insight into the external reliability of a particular filter model (because the analysis of each model error requires a separate filter run). Hence only a limited number of cases will be analysed.

The effect of model errors caused by outliers and slips in the observations has been computed for three cases, namely cases 1a, 1f (case with lower system noise), and 1i (case with exponentially time-correlated measurements). For this experiment error-free observations have been simulated for a straight line trajectory of 100 fixes with data available at one second intervals. The various model errors are introduced at fix 30. The size of the error corresponds with the MDB of the local test for the particular case considered at fix 30. Outliers and slips have been introduced in the first range (SR1), gyrocompass, and log. During a first evaluation of the results we found that the behaviour of model errors concerning gyro and log observations is largely similar. Therefore only the results pertaining to the gyro model errors are discussed.

In Figs. 5.17 to 5.22 the position and velocity response to model errors in range and gyro observations are given. Note that the amplitude of the response is different for the various cases, because the magnitudes of the MDBs are different for each case. It can be seen in Fig. 5.17 that the impact of an outlier in the range damps out fairly quickly, whereas the impact of an outlier in the gyro observation has a long-term effect on the position and velocity (cf. Fig. 5.18, where one can see that the filter displays an underdamped response to a gyro outlier). For an outlier the impact on the velocity is

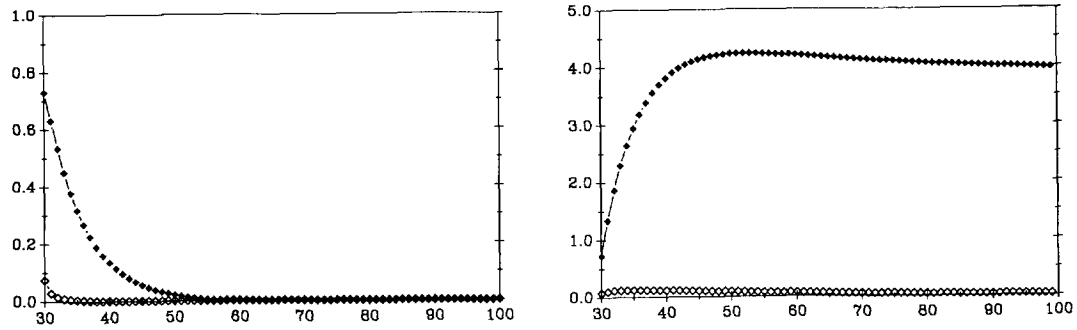


Figure 5.17: ◆ Position [m] and ◇ velocity [m/s] response due to an outlier (left) and slip (right) of 6.57m at fix 30 in range SR1 vs. the fix number — case 1a.

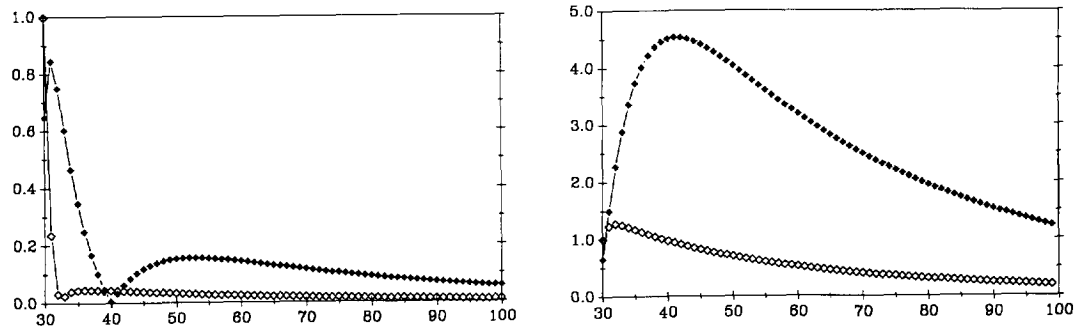


Figure 5.18: ◆ Position [m] and ◇ velocity [m/s] response due to an outlier (left) and slip (right) of 15.87deg at fix 30 in the gyro observation — case 1a.

largest at the time of occurence of the outlier. A slip in the range measurement (after an initial transient of approximately 15 fixes) causes an almost constant bias in the position estimate, while the impact on the velocity damps out slowly (cf. Fig. 5.17). A

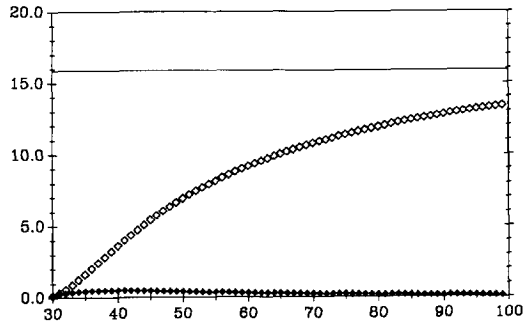


Figure 5.19: Estimate of gyro offset [deg] due to an outlier (◆) and slip (◇) of 15.87deg in the gyro observation at fix 30 — case 1a.

slip in the gyro observation, on the other hand, causes a slowly diminishing bias in the position and velocity estimates once a maximum has been reached. This behaviour is due to the fact that the model error is slowly absorbed in the gyro offset instrumental state, which is illustrated in Fig. 5.19. It is interesting to note that an error of the size of the MDB in either the range or gyro observation causes position biases of the same magnitude. The impact on the velocity estimate is, however, larger for model errors in the gyro observation.

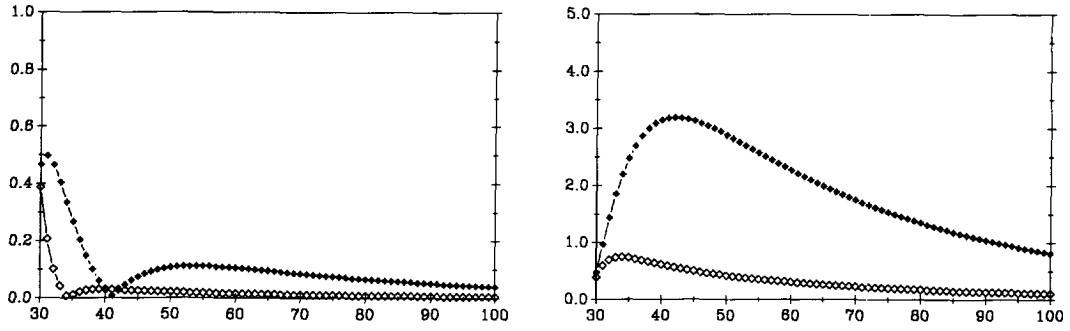


Figure 5.20: ◆ Position [m] and ◇ velocity [m/s] response due to an outlier (left) and slip (right) of 10.89deg at fix 30 in the gyro observation — case 1f.

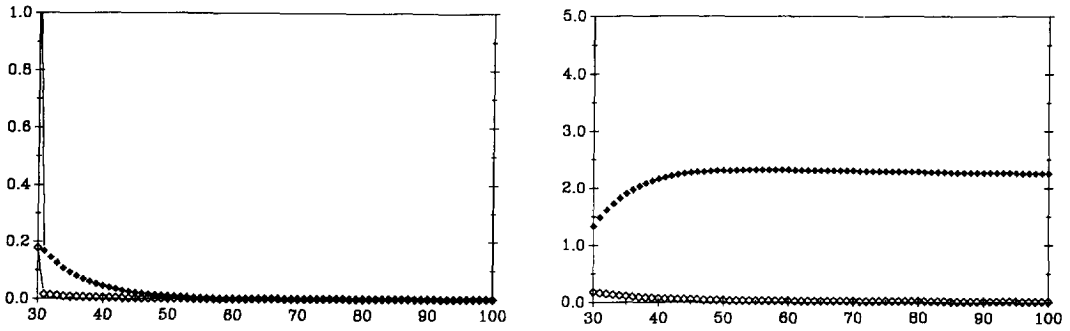


Figure 5.21: ◆ Position [m] and ◇ velocity [m/s] response due to an outlier (left) and slip (right) of 3.72m at fix 30 in range SR1 — case 1i.

For case 1f only the response graphs for an outlier and a slip in the gyro observation are given in Fig. 5.20, as the response for model errors in the range observations is almost identical to case 1a. It can be seen that lower system noise leads to a somewhat damped response for model errors in the gyro observations (compared to case 1a). The response of the model with exponentially time-correlated measurements (case 1i) is different as can be seen in Figs. 5.21 and 5.22. The response is nearly instantaneous, because in the augmented state model the measurements are noise-free (cf. the discussion in Section 2.6). The response to the gyro outlier is overdamped as opposed to the previous cases.

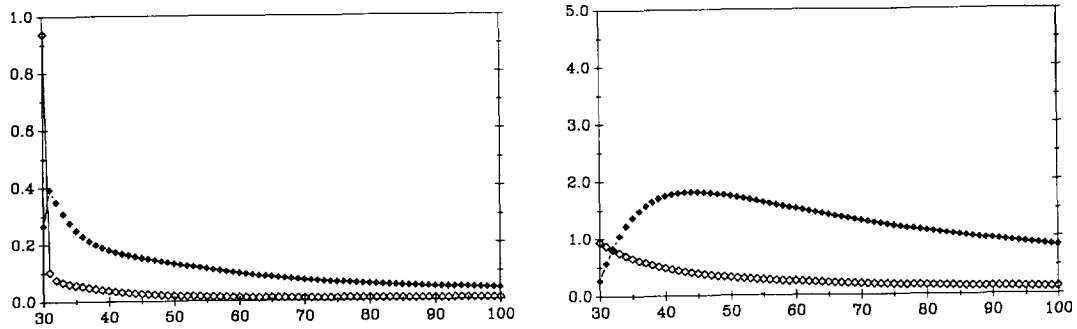


Figure 5.22: ◆ Position [m] and ◇ velocity [m/s] response due to an outlier (left) and slip (right) of 12.61deg at fix 30 in the gyro observation — case 1i.

The analysis of the response graphs shows that:

- A model error in the range observations mainly influences the position states.
- The effect of outliers in range observations on the position and velocity estimates is of short duration.
- Model errors of the gyro observations influence position and velocity.
- The impact of an outlier in the gyro observations is, although limited in magnitude, quite prolonged.
- Slip type model errors in gyro observations tend to be absorbed in instrumental bias states.
- Exponentially time-correlated measurement noise has a large influence on the position and velocity response.

Given these observations, we can now use the results regarding the response graphs to improve our understanding of the characteristics of the BNRs (a strategy which already proved very useful in Section 5.4, where we considered linear models).

Bias to Noise Ratios

In this subsection we will consider the BNRs for outliers and slips in the ranges and gyro observations. The results pertaining to the ranges and gyro observations are also representative for range-differences and log measurements respectively. The BNRs facilitate a direct comparison of the significance of the bias related to different alternative hypotheses and therefore the BNRs related to different hypotheses are discussed simultaneously. The analysis focusses on the comparison of the BNRs for different filter models. The analysis of BNRs is somewhat complicated by the fact that certain model errors may influence specific elements of the state vector. Here all BNRs refer to four

unknowns ($npd = 4$), viz. the position and velocity states. BNRs of other subsets of unknowns are discussed separately. In the graphs the square root of the BNRs is always plotted with respect to the delay of the test. The BNRs for the reference case 1a are tabulated in Tables 5.13 and 5.14.

	Outliers				Slips			
delay	range SR2	range SR3	gyro	log	range SR2	range SR3	gyro	log
0	1.49	1.26	6.63	5.81				
1	1.31	1.08	1.80	1.69	2.08	1.71	7.54	6.57
2	1.12	0.89	1.42	1.28	2.50	1.98	7.23	6.21
3	0.94	0.72	1.24	1.10	2.80	2.15	6.62	5.60
4	0.79	0.58	1.04	0.88	3.03	2.25	6.00	5.00
5	0.66	0.46	0.85	0.68	3.18	2.29	5.46	4.49
6	0.55	0.37	0.69	0.52	3.29	2.30	5.00	4.07
7	0.46	0.29	0.55	0.40	3.36	2.29	4.62	3.72
8	0.39	0.23	0.44	0.31	3.39	2.27	4.28	3.41
9	0.32	0.18	0.35	0.25	3.41	2.23	3.99	3.14

Table 5.13: Square root of Bias to Noise Ratios ($npd = 4$) of outliers and slips in the observations at fix 90 ($l = 90$, $k - l = 0, \ldots, 9$); (SR2 and SR3 are the ranges with the largest and smallest BNRs respectively.) — case 1a.

	acceleration		gyro	log bias
delay	along track	across track	drift	drift
1	2.35	2.02	10.15	7.54
2	1.89	1.65	11.96	8.23
3	1.59	1.40	12.41	8.19
4	1.38	1.24	12.04	7.76
5	1.24	1.12	11.27	7.22
6	1.13	1.03	10.44	6.67
7	1.05	0.95	9.64	6.18
8	0.98	0.90	8.92	5.75
9	0.93	0.85	8.29	5.38

Table 5.14: Square root BNRs ($npd = 4$) of state related hypotheses at fix 90 ($l = 90$, $k - l = 0, \ldots, 9$) — case 1a.

In Section 4.4.6 we observed that it is very difficult to predict the response of the BNRs to a change in the design parameters. We argued, however, that the BNRs would get smaller with increasing redundancy, which could be brought about by, for instance, integration. We also noted that increasing the number of sensors of the system indeed increases the redundancy of the system, but at the same time improves the precision of the filtered state estimates, and therefore the impact of integration on the BNRs is difficult to predict. It is expected, however, that adding observations of a type already

used in the system (e.g. using 3 instead of 2 ranges) will lead to smaller BNRs as the redundancy is improved more dramatically than the precision. With the analysis of (elements of) $\nabla \hat{x}_{k|k}$ we have some additional information available. We have seen that the impact of an outlier decreases quickly with increasing delay and hence it is expected that BNRs related to outliers in the observations will decrease with increasing delay as well. Slips in the observations have a large systematic effect on the filtered state and will result in large BNRs, which are not expected to decrease with increasing delay.

Considering case 1a for the moment (cf. Tables 5.13 and 5.14), we see indeed that the BNRs for outliers in all types of observations decrease quickly with increasing delay, as opposed to slips in the observations which lead to large (or even increasing) BNRs with increasing delays. In Table 5.13 we want to point out an interesting phenomenon. The BNRs related to the slips in the dead reckoning observations decrease from delay 2 onwards, while one would expect from the response graph (cf. Fig. 5.18) that the BNRs increase. The response graphs, however, represent the bias in the position and velocity states caused by an error of a specific size, which does not necessarily coincide with the MDB. From Table 5.11 it follows that the MDBs for slips in the dead reckoning observations rapidly decrease with increasing delay. The unexpected response of the BNR seen in Table 5.13 can thus be explained by the fact that the size of the MDBs becomes smaller more quickly than the bias in the state vector increases. Hopefully the reader will learn from this exposition that the analysis of the BNRs is not necessarily trivial. However, instead of going into much more detail and arriving at conclusions based on a single case, we consider some other important cases.

The impact of integration on the BNRs is depicted in Figs. 5.23 and 5.24, where BNRs related to outliers and slips in a range and gyro observation are considered. Also the BNRs of cases 1f (low system noise), 1g (large standard deviation of the dead reckoning observations), and 1i (exponentially time-correlated measurements) are given in Figs. 5.25 and 5.26. The BNRs of models 1f, 1g, and 1i should be compared with the BNRs of model 1a, which are depicted in Fig. 5.25. In Fig. 5.27 a comparison is made between cases 2f (Hyperfix, gyro, log) and 2g (Syledis, Hyperfix, gyro, log).

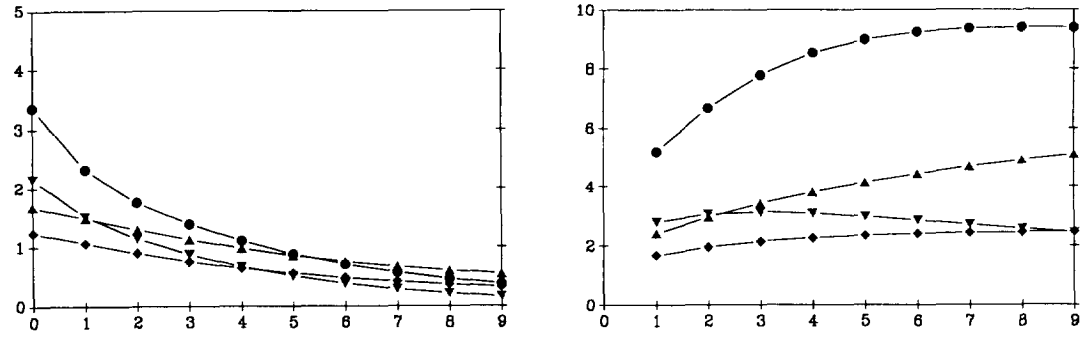


Figure 5.23: sqrt($\lambda_{\hat{x}}$) outlier (left) and slip (right) Syledis range vs. the delay; ● 2a; ▲ 2e, ▼ 2c, ◆ 2g.

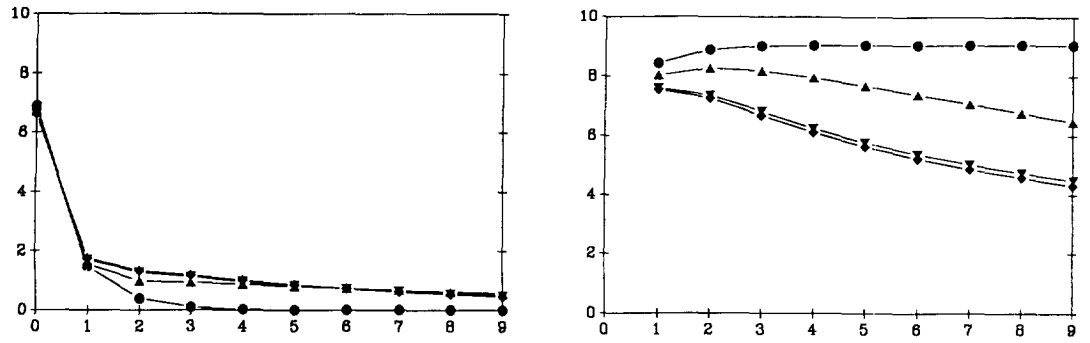


Figure 5.24: sqrt($\lambda_{\hat{x}}$) outlier (left) and slip (right) gyro; ● 2d; ▲ 2e, ▼ 2f, ◆ 2g.

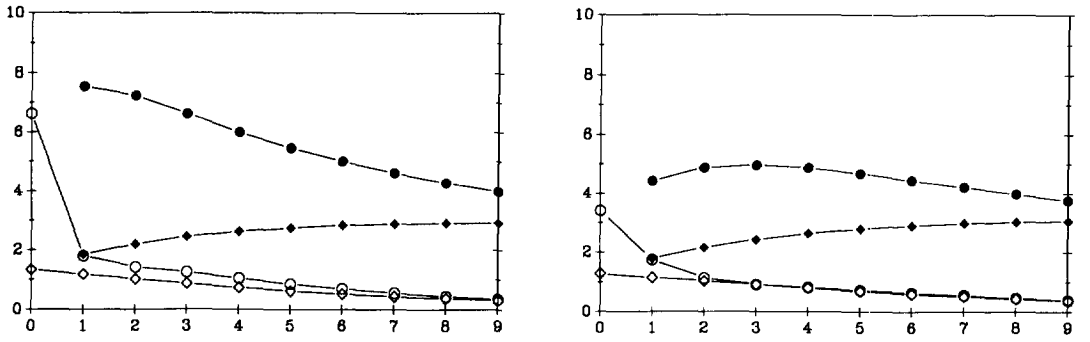


Figure 5.25: sqrt($\lambda_{\hat{x}}$) outlier range ◇, slip range ◆, outlier gyro ○, slip gyro ● — case 1a (left) and case 1f (right).

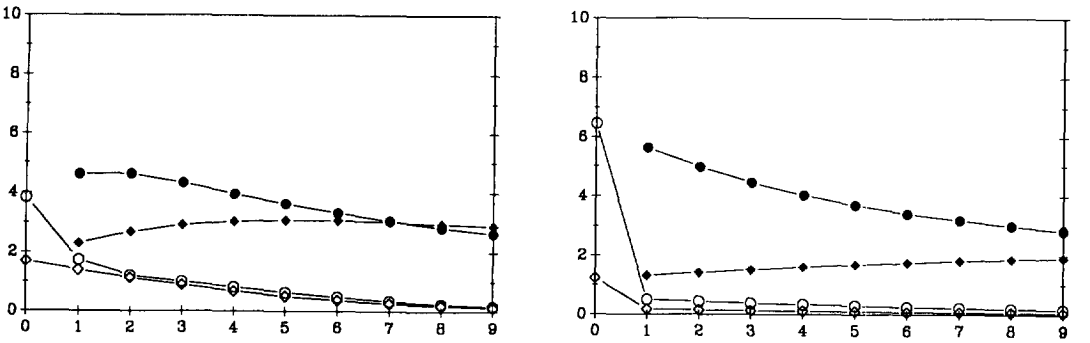


Figure 5.26: sqrt($\lambda_{\hat{x}}$) outlier range ◇, slip range ◆, outlier gyro ○, slip gyro ● — case 1g (left) and case 1i (right).

The analysis of BNRs for outliers and slips in the observations shows that (cf. Table 5.13 and Figs. 5.23 to 5.27):

- BNRs related to outliers are significant for delay 0 and decrease very quickly with increasing delay. In the response graphs the quick decrease of the bias in the state

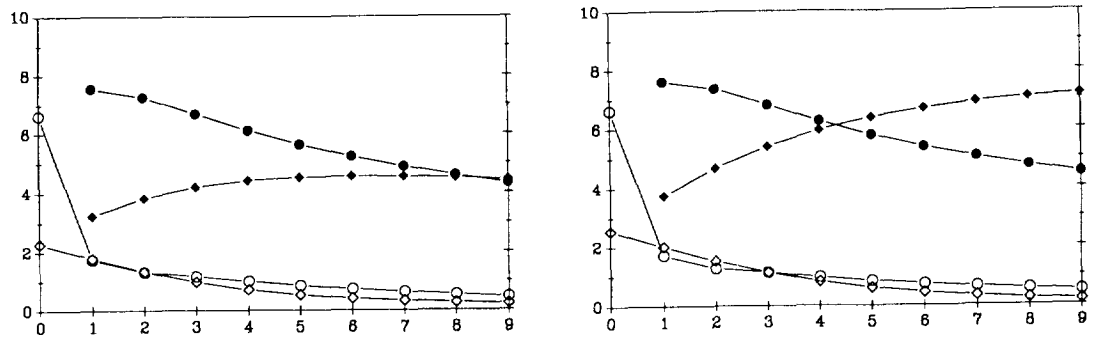


Figure 5.27: sqrt($\lambda_{\hat{x}}$) outlier range ◇, slip range ◆, outlier gyro ○, slip gyro ● —, case 2f (left) and case 2g (right).

vector is clearly illustrated.

- For delay 0 the BNRs of outliers in dead reckoning sensors are larger than the BNRs of ranges. This is caused by the fact that instantaneously a model error in a dead reckoning observation mainly influences the velocity states, which are determined more precisely than the position states (cf. Table 5.9).

- BNRs related to slips in observations are always larger than the BNRs related to outliers in the corresponding observations. Slips cause a permanent, significant bias in the state estimates and therefore BNRs related to slips remain large with increasing delay.

- BNRs related to slips in range observations have a tendency to grow with increasing delay, whereas BNRs related to slip in dead reckoning observations generally tend to decrease slowly with increasing delay. This corresponds with the results derived from the response graphs and the discussion concerning the results given in Table 5.13.

- Integration generally leads to smaller BNRs for slips in all observation types and outliers in range observations for small delays (cf. Figs. 5.23 and 5.24). This observation is, however, not always true (cf. Fig. 5.27). If integration of an additional system (in this case Hyperfix) also brings about a considerable improvement of the precision of the filtered states, the BNRs related to slips may even become larger.

- Smaller system noise (case 1f, Fig. 5.25) and larger standard deviations of the dead reckoning observations (case 1g, Fig. 5.26) lead to smaller BNRs of the dead reckoning observations (if compared with case 1a). For case 1f the decrease of the MDBs (which is associated with smaller BNRs, cf. Fig. 5.12) is more significant than the decrease of the standard deviations of the filtered state (which is associated with larger BNRs, cf. Table 5.9). For case 1g the increase of the

standard deviations of the filtered states is more pronounced than the increase of the MDBs (cf. Fig. 5.13 and Table 5.9).

It is also found that some design changes hardly affect the size of the BNRs:

- Larger standard deviations of the range measurements have little impact on the BNRs related to outliers and slips in the range observations, because simultaneously the biases in the position grow and the precision of the filtered position states decreases.

- Using pseudo-ranges instead of ranges has no significant effect on the BNRs of the (pseudo-)ranges.

Prediction of the behaviour of the BNRs for the cases with varying measurement update interval (cases 1j and 1k) is rather difficult. As the number of observations used reduces with increasing measurement interval it is expected that the BNRs will become larger. How much larger, however, is difficult to predict as also the precision of the filtered state will deteriorate. The effect of varying measurement interval on the BNRs for outliers and slips in range and gyro observations is illustrated in Figs. 5.28 and 5.29. One sees indeed that for outliers and slips of ranges the BNRs increase with increasing measurement interval. At the same time the BNRs related to the gyro measurements become smaller.

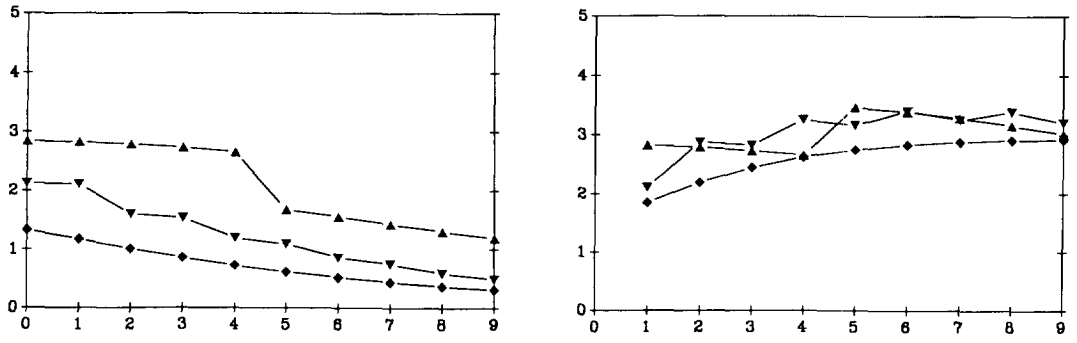


Figure 5.28: sqrt($\lambda_{\hat{x}}$) outlier (left) and slip (right) range; ◆ case 1a, ▼ case 1j, ▲ case 1k.

Up to this point the analysis of BNRs has been limited to the BNRs of outliers and slips in the observations. In Fig. 5.30 we briefly consider the model errors of unmodelled accelerations and instrumental biases. The BNRs related to unmodelled accelerations slowly decrease with increasing delay and the BNRs for the gyro offset are large compared with the BNRs encountered so far. Apparently instrumental biases have a significant impact on the filtered state, and therefore the part of the model related to the instrumental biases should be carefully designed.

Summarizing the results given in Figs. 5.23 to 5.27, it follows that the BNRs for outliers in the observations are significant for delay 0 only. For most cases it is found

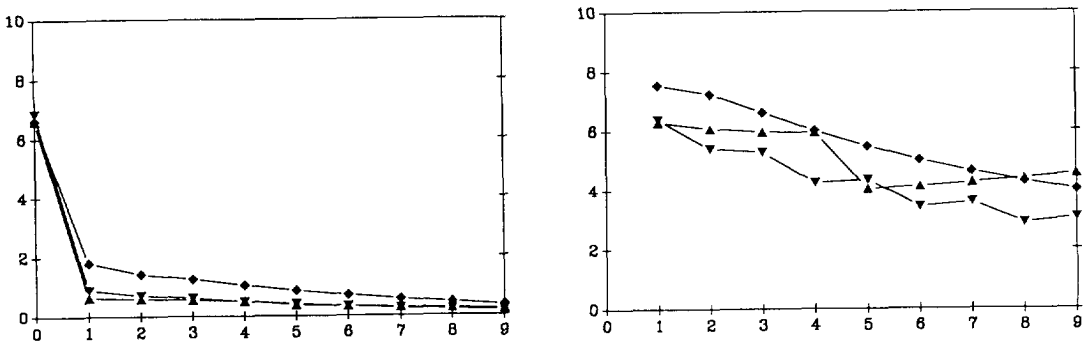


Figure 5.29: sqrt($\lambda_{\hat{x}}$) outlier (left) and slip (right) gyro; ◆ case 1a, ▼ case 1j, ▲ case 1k.

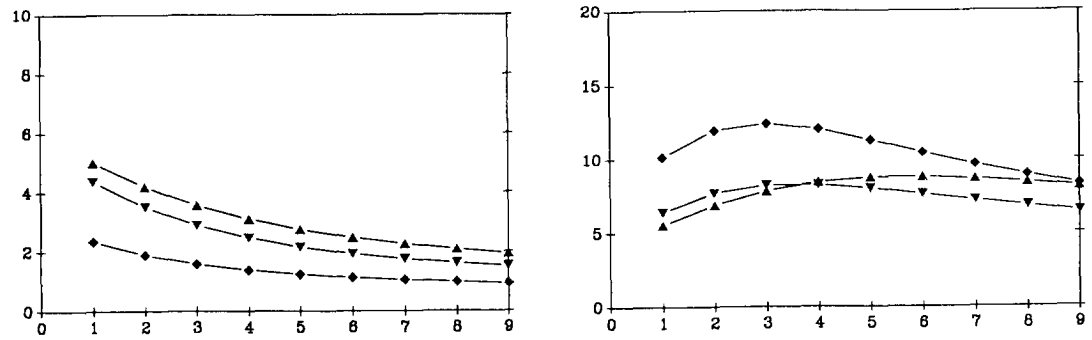


Figure 5.30: sqrt($\lambda_{\hat{x}}$) acceleration along (left) and gyro offset (right); ◆ case 1a, ▲ case 1f, ▼ case 1g.

that the BNRs for slips in observations remain large with increasing delay. In most cases slips in range observations lead to smaller BNRs than slips in gyro observations (for delays 0-9), which is caused by the fact that slips in the gyro (dead reckoning) observations have a larger impact on the precisely determined velocity states than the ranges. The analysis of the BNRs is greatly facilitated by the results obtained from the response graphs. The analysis of the BNRs indicates that for tests related to outliers small window lengths (1 or 2) suffice, as the impact of an undetected error rapidly decreases. Slip-type hypotheses on the other hand would require tests with large window lenghts.

BNRs of Subsets of the State

The BNRs do not necessarily refer to all state vector elements, but can also refer to particular functions of the state or a subset of the state vector (which is modelled by the matrix F^{T} in eq. 4.31). For case 1a we show how the BNR varies for different subsets of the state. We consider the cases $npd = 2, 4,$ and 6, where npd is the dimension of the subset (cf. Table 5.15).

	$npd = 2$	$npd = 4$	$npd = 6$
states considered	position	position velocity	position velocity instrumental parameters
F^{T}	$\begin{pmatrix} I_2 & 0_{2\times 4} \end{pmatrix}$	$\begin{pmatrix} I_4 & 0_{4\times 2} \end{pmatrix}$	I_6

Table 5.15: BNR analysis — Subsets of state vector considered ($0_{i\times j}$ denotes a $i \times j$-matrix with zero entries).

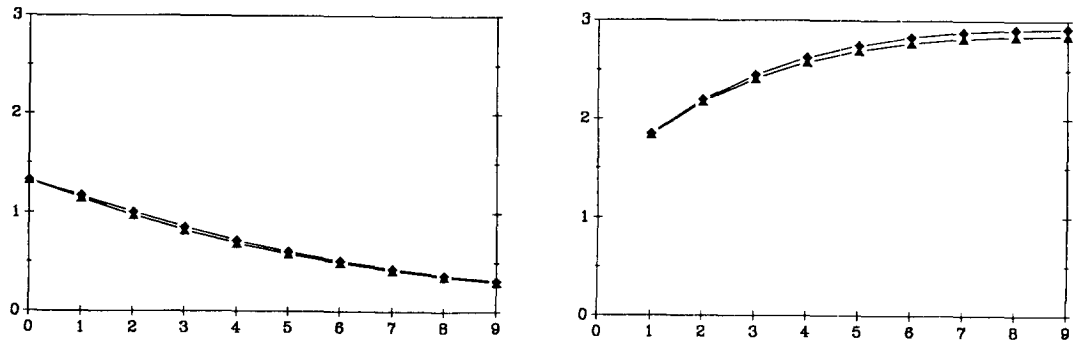


Figure 5.31: sqrt($\lambda_{\hat{x}}$) for outlier (left) and slip (right) in range with $npd = 2$(▲) and $npd = 4$(◆) — case 1a.

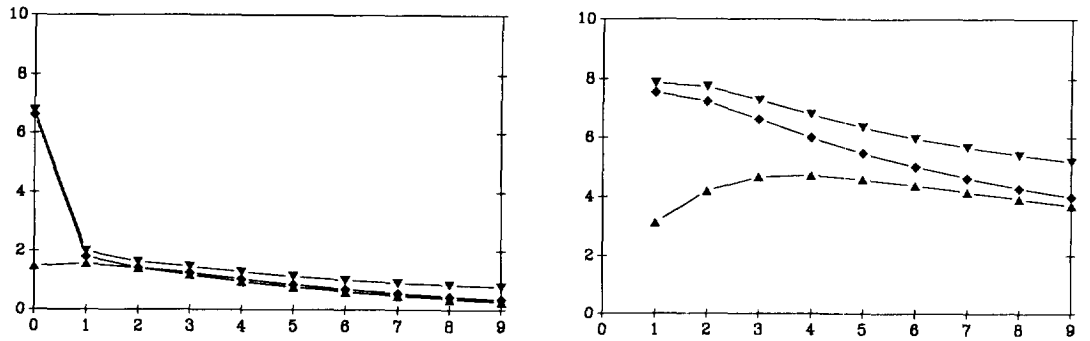


Figure 5.32: sqrt($\lambda_{\hat{x}}$) for outlier (left) and slip (right) in gyro with $npd = 2$(▲), $npd = 4$(◆) and $npd = 6$(▼) — case 1a.

The analysis of BNRs related to different subsets of the state vector reveals that:

- Model errors related to range measurements mainly affect the position unknowns (cf. Fig. 5.31, where it can be seen that the graphs of the BNRs almost coincide for $npd = 2$ and 4, which means that the bias in the velocity states (x_3, x_4) contributes little to the total significance of the bias). This result can also be

derived from the response plot in Fig. 5.17.

- Model errors related to gyro observations influence all states (see Fig. 5.32, in which it is clearly visible that $\sqrt{\lambda_{\hat{x}}}(npd = 6) > \sqrt{\lambda_{\hat{x}}}(npd = 4) > \sqrt{\lambda_{\hat{x}}}(npd = 2)$). This result also follows from Fig. 5.18. For outliers in the gyro observations the difference between the various BNRs is only significant the instant the outlier occurs, which can be explained by the fact that at the moment the outliers occurs primarily the more precisely determined velocity is affected.

Although not explicitely illustrated we also found that:

- Model errors concerning unmodelled accelerations strongly influence position and velocity unknowns ($\sqrt{\lambda_{\hat{x}}}(npd = 4) > \sqrt{\lambda_{\hat{x}}}(npd = 2)$), but hardly the instrumental biases ($\sqrt{\lambda_{\hat{x}}}(npd = 6) \approx \sqrt{\lambda_{\hat{x}}}(npd = 4)$).

- Model errors related to instrumental biases affect all unknowns.

Overall it follows that the magnitude of the BNR may vary considerably for different subsets of the state vector, especially for the hypotheses related to the gyro (dead reckoning) observations.

Summary of the Analysis of the BNRs

We have analysed the BNRs for outliers and slips in the range and gyro observations, and we have briefly considered the dependence of the BNRs on the choice of various subsets of the state. The characteristics of the BNRs of the range and gyro observations are also valid for the BNRs of range-difference and log observations respectively. The analysis of the BNRs was greatly facilitated by the results we obtained from the response plots, but the response plots alone are not sufficient to explain the behaviour of the BNRs. Unlike the MDB the BNR is not a monotonic (either decreasing or increasing) function of the delay. An advantage is that the BNRs concerning different types of hypotheses can be directly compared.

We have found that the significance of the biases in the state vector differs for outliers and slips. Generally outliers cause significant biases (associated with large BNRs) for a zero delay only. Slips, on the contrary, cause a permanent, significant bias with increasing delay. This brings us to the tentative conclusion that for outliers local tests suffice, while tests for slips should have larger window lengths. The analysis of the BNRs thus substantiates the testing strategy based on the analysis of the MDBs, and the strategy we proposed for the choice of the window lengths is Section 4.4.6 seems to be appropriate. The findings above do not automatically imply that one can discard tests of hypotheses for which the BNR is small. If the error that is likely to occur is much larger than the MDB, such an error can still cause a significant bias (the BNR is merely a measure of significance for a bias caused by a model error of the size of the MDB).

In the foregoing we did not discuss the absolute size of the BNRs, but limited ourselves to a comparison of BNRs. In geodetic control networks the rule of thumb is that an observation can be considered 'reliable' if the square root of the BNR is smaller than 10. For navigation systems such a limit has not been determined yet. For the time being it seems useful to judge the external reliability by using the inequality provided by (4.35). We found that the magnitude of the BNR is sensitive to changes in the stochastic model of the observables (especially exponentially time-correlated measurements) and the level of integration. The magnitude of the BNR of dead reckoning related hypotheses depends on which subset of the state vector is analysed. In practice the system designer should decide which subset of the state is of particular interest and should be used in the analysis of the BNRs.

5.6.4 Combined Analysis

Now we have studied the response of the various quality measures (for precision, internal and external reliability) to variations in the design parameters separately, we will briefly consider a combined analysis of all quality measures. Especially during the discussion of the BNRs it became obvious that the quality measures are strongly interrelated. This indicates that in general it is useless to optimize the design of a system with respect to a single quality measure, except, of course, in those cases where this is explicitly required. If one designs a system taking into account all aspects of quality simultaneously, one would like to obtain a system which meets the following three requirements:

1. The precision of the estimator of the filtered state is high (the standard deviations of the estimators are small).

2. The MDBs are smaller than the size of the model errors that are likely to occur, that is the model errors can readily be detected by the testing procedure.

3. The BNRs are small, that is the significance of the bias caused by a model error of the size of the MDB is small.

In practice it might be difficult to design a system that complies with all three design objectives. In the following we will illustrate that meeting one objective does not necessarily imply that the other two design goals are automatically fulfilled.

We consider the combined quality analysis for two alternative hypotheses, namely an outlier (delay 0) and a slip (with delay 9) in the Syledis range SR2 (cf. Table 5.16). The impact of these model errors on the Easting coordinate is investigated. In Table 5.16 the standard deviation of the Easting coordinate at fix 90, the MDBs and BNRs of the two alternative hypotheses, and the maximum bias in the Easting coordinate caused by a model error with the size of the MDB are given for the cases with ranges we analysed in design areas 1 and 2. The maximum bias in the Easting coordinate is computed from the inequality (cf. eq. 4.36)

$$|\nabla E| < \sqrt{\lambda_{\hat{x}}}\, \sigma_E \,. \tag{5.12}$$

The BNR refers to the position and velocity states ($npd = 4$) and thus (5.12) is clearly an upperbound.

case	σ_E [m]	outlier SR2 (delay 0) MDB [m]	$\sqrt{\lambda_{\hat{x}}}$ ($npd = 4$)	max$\lvert\nabla E\rvert$ [m]	slip SR2 (delay 9) MDB [m]	$\sqrt{\lambda_{\hat{x}}}$ ($npd = 4$)	max$\lvert\nabla E\rvert$ [m]
1a	0.47	6.59	1.49	0.70	2.77	3.41	1.60
1b	0.79	7.44	2.74	2.16	3.08	3.27	2.58
1c	1.12	4.09	2.27	2.54	3.70	2.24	2.51
1d	0.54	7.41	1.26	0.68	2.90	2.82	1.52
1e	0.48	6.64	1.58	0.76	2.99	3.99	1.92
1f	0.45	6.56	1.44	0.65	2.74	3.56	1.60
1g	0.58	6.80	1.86	1.08	2.92	3.38	1.96
1h	0.82	10.9	1.35	1.11	4.51	3.54	2.90
1i	0.97	3.72	1.18	1.14	3.50	1.96	1.90
1j	0.70	6.92	2.05	1.44	3.90	3.51	2.46
1k	0.86	7.46	2.77	2.38	5.89	3.21	2.76
2a	2.00	7.99	3.26	6.72	6.26	9.39	18.8
2c	1.00	6.99	2.16	2.16	2.60	2.45	2.45
2e	0.95	6.69	1.69	1.61	3.41	5.10	4.85
2g	0.56	6.47	1.24	0.69	2.42	2.47	1.38

Table 5.16: Combined analysis of the quality measures at fix 90 for hypotheses related to an outlier and a slip in range SR2.

The combined analysis in Table 5.16 shows that:

- Integration improves the overall quality of the system (cf. cases 2a $\leftrightarrow$ 2c $\leftrightarrow$ 2e $\leftrightarrow$ 2g, 1a $\leftrightarrow$ 1b, and 1c $\leftrightarrow$ 1i).
- The relative improvement with respect to one quality measure does not necessarily coincide with an improvement with regard to the other quality measures. From Table 5.16 one can see for instance that:
  - Small standard deviations of the filtered state do not always coincide with small MDBs (cf. 1a $\leftrightarrow$ 1i, 1b $\leftrightarrow$ 1c, and 2c $\leftrightarrow$ 2g). In Section 5.6.2 we saw that the size of the MDBs is related to the precision of the observations.
  - Large standard deviations can occur simultaneously with small BNRs (cf. 1h/1i $\leftrightarrow$ 1a). The BNR is a measure of significance of the bias relative to the precision of the filtered state, and hence the effect of larger variances of the filtered state is a decrease of the BNR (cf. Section 5.6.3).
  - Small BNRs can occur simultaneously with large MDBs (cf. 1h).

The results show that a quality assessment of an integrated navigation system should be based on a combination of quality measures. Reliance on a single quality measure

(e.g. precision) might result in a system which provides very precise but unreliable results.

5.7 Reliability Studies for Integrated Navigation Systems

In the literature one finds only a few references to design studies with are particularly devoted to reliability. In the following we briefly discuss three design studies in which reliability was explicitely considered.

An early investigation into reliability aspects of dynamic systems has been performed by BUENO ET AL. [1976], who have investigated detectability and separability issues of GLR-tests for a simple (two-dimensional) aerospace application. Their results are stated in terms of probabilities. Especially the results regarding separability are of interest, because they show that the separability of certain model errors may be small. Not surprisingly they find that the separability between a hypothesis concerning a slip in a particular state vector element and a slip in an observation of that state vector element is small. (Our analysis of the correlation between hypotheses in Section 5.6.2 yielded similar results). Their results are quite promising but are only available for a single design case.

Based on the quality measures given in Section 5.2, TIBERIUS [1991] has performed a design study for a vehicle location system. In his study it is shown that it is absolutely necessary to integrate positioning and dead-reckoning system(s) to meet the quality requirements for vehicle location systems. Apart from the measurement scenario and the implemented dynamic model, the quality measures are also shown to be trajectory and velocity dependent. These latter two phenomena were hardly seen in our (hydrographic) navigation system case study, as we worked in a much more benign dynamic environment than is encountered in a road-like situation with its constant accelerations. Furthermore TIBERIUS demonstrated numerically that the enlargement of the state by additional instrumental parameters does not have a significant impact on the reliability description. The additional parameters were, however, necessary to keep the estimators of the state vector unbiased.

LU AND LACHAPELLE [1990], based on the methodology developed at the Delft Geodetic Computing Centre, have considered the impact of model errors on the results of a GPS positioning system. Their results on external reliability are not based on a statistical concept like the MDB, and therefore their so-called Bias to Noise Ratios should rather be compared to our response plots. They find that outliers (in GPS phase measurements) mainly cause an instantaneous bias in the state, while (cycle) slips result in a permanent bias in the state vector. These findings agree with our analysis of the response plots in Section 5.6.3.

5.8 Concluding Remarks

In this chapter we have dealt with the design procedure for dynamic systems, and integrated navigation systems in particular. The purpose of the design study was to analyse the characteristics of the various quality measures under different model assumptions, to assess the usefulness of these quality measures for the design procedure of integrated navigation systems, and to provide insight into how the window lengths of the tests should be chosen.

The analysis of a simple, linear two-dimensional model proved very useful for a first characterization of the quality measures. We paid much attention to the analysis of the measures of internal and external reliability (MDBs and BNRs). Conclusions from the analysis of the various quality measures are the following:

- The precision of the filtered state is directly related to the precision of the measurements and can be improved by integration (cf. Section 5.6.1).

- The size of the MDB is primarily a function of the measurement precision. Integration improves the internal reliability and thus results in smaller MDBs. MDBs of alternative hypotheses related to slips in the observations and instrumental biases decrease with increasing delay. MDBs of hypotheses related to outliers only become smaller for the first delay (if they become smaller at all).

- The correlation between the test statistics is significant for certain pairs of alternative hypotheses.

- Outliers cause a bias of relatively short duration in the state vector, whereas slips cause a permanent bias in the filtered state. The analysis of the BNR can be quite involved, as the BNR is dependent on the MDB, the precision of the filtered state, and the delay of the test. Generally outliers only cause a significant bias the instant the error occurs, whereas slips result in permanent, significant biases. Integration generally leads to smaller BNRs.

The results concerning the precision and internal reliability correspond nicely with what one would expect intuitively from the results of conventional geodetic network design. The results concerning reliability are similar to those of GPS network design (see [KÖSTERS, 1992]). Overall the quality of the system is to a large extent dependent on the stochastic model of the observables. The stochastic model of the observations is largely dependent on the sensors used, while the stochastic model of the disturbances is dependent on the disturbances actually encountered. We also found that integration improves the overall quality of the system. Especially the integration of 'absolute' positioning systems and dead reckoning systems leads to large quality improvements. Fortunately the integration concept and the choice of the sensors can, to a large extent, be influenced by the designer (leaving cost limitations out of consideration). Taking the results of the design studies into account we think that if a design study indicates

that for a certain system the quality requirements cannot be met, enlarging the level of integration might be the best way to meet the requirements.

We have shown that in the design procedure all quality criteria should be taken into account simultaneously. An actual design study can very well be performed using the quality measures we discussed. The precision requirements for navigation systems are usually quite well known (cf. Section 4.3), and it can be easily verified whether the system meets these requirements. Also one generally has some knowledge of which model errors are likely to occur and what is the magnitude of these errors. If the MDBs are smaller than the size of the likely errors, the detectability of the model errors is sufficient. We could not provide, however, upperbounds for the BNRs. A rule of thumb (such as it exists for control networks) cannot be given yet for navigation systems.

The design study certainly provides guidance on the choice of the window lengths for the testing procedure. From the analysis of the MDBs and BNRs we conclude that for tests for outliers short window lengths (1 or 2) are sufficient, whereas tests for slips might require longer window lengths. Slips cause permanent, significant biases in the filtered state (which necessitates their detection and identification). At the same time slips can generally be better detected by tests with large window lengths. If, however, the size of the likely model error is much larger than the MDB, local tests are sufficient for the detection of slips.

Overall we think the findings of this chapter constitute a first step towards a full-scale design procedure for integrated navigation systems. Most conclusions are based on the navigation system introduced in Section 5.5. The correspondence between the results of the (limited) analysis of the linear model and the navigation models indicates that the conclusions drawn from these results can also be extended to other systems. Finally we think this study may provide a benchmark for future design studies.

Chapter 6

DIA Test Computations

6.1 Introduction

In the previous chapter the design of an integrated navigation system has been discussed. The reliability description of such a system was based on the assumption that the data processing is supported in real time by a testing strategy, viz. the DIA procedure. In this chapter we will more closely investigate the performance of the detection, identification and adaptation procedure. The analysis of the DIA procedure is performed step-wise. First we limit ourselves to the detection and identification steps of the DIA procedure and try to establish whether the actual detection and identification performance is in accordance with the results of the design procedure. Then we proceed with a performance study of the adaptation procedure. Adaptation in dynamic systems is a relatively new field of interest, and therefore it was deemed wise to follow a cautious approach where we limit ourselves to local adaptation (i.e. adaptation based on local identification tests). All our analyses are based on the simple linear model and navigation models introduced in Chapter 5. The linear model is mainly used to clarify the operation of the DIA procedure.

All analyses are based on simulated datasets. Because this study presents a first systematic investigation into the performance of the DIA procedure, it is felt that datasets with perfectly known (i.e. simulated) model errors should be used. If the method performs satisfactorily for simulated datasets it can be used for real datasets at a later stage. The processing of the data of the navigation models is based on the iterated extended Kalman filter (IEKF).

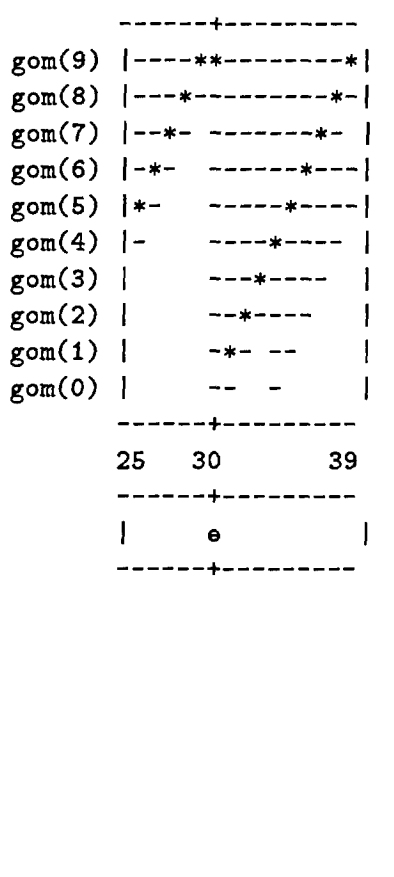
6.1.1 Overview of this Chapter

We first analyse the performance of the detection and identification part of the DIA-procedure. In Section 6.2 the detection and identification plots are introduced and explained. The performance of the DI-procedure is then analysed for a simple linear model and some of the navigation models in Sections 6.2.2 and 6.2.3 respectively. The adaptation procedure is investigated (using several adaptation strategies) for the simple

linear model and navigation models in Sections 6.3.1 and 6.3.2 and additionaly for the lane-slip problem of hyperbolic positioning in Section 6.3.3. In Section 6.4 the main findings of the DIA test computations are summarized.

6.2 Detection and Identification

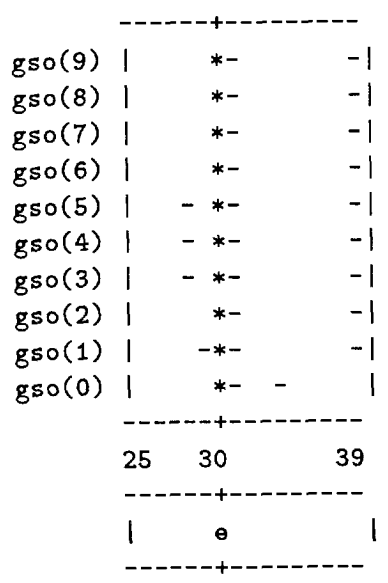
The results of the detection and identification procedure will be primarily presented in graphical form. A key to the so-called detection and identification plots, which are our primary analysis tools, is given in Figs. 6.1 and 6.2. The major difference in analysing the detection and identification plots is that the detection plots refer to the epoch of *computation*, whereas the identification plots refer to the epoch of *identification*.



The detection plots should be read as follows. On the horizontal axis the fix number is plotted (here fixes 25 to 39 are considered). On the vertical axis the detection test statistics are given (`gom(i)` denotes the global overall model test statistic with delay `i`; `gom(0)` corresponds with the local overall model test statistic). All test statistics that are larger than their critical value are indicated by a hyphen (-). The detection results refer to the epoch of computation. The largest detection test statistic for a particular epoch is denoted by an asterisk *. From the figure one can, for example, read that at fix 30 the detection test statistics for delays 0 (`gom(0)` plotted at fix 30 corresponds to $T^{l,k} = T^{30,30}$) to 9 (`gom(9)` plotted at fix 30 corresponds to $T^{21,30}$) are all larger than their critical value and that the largest test statistic is associated with the global overall model test statistic with delay 9 (`gom(9)` is tagged by an asterisk *). In the lower part of the figure it is indicated by an `e` where errors have been introduced in the simulated dataset.

Figure 6.1: Interpretation of the detection plots.

In Fig. 6.1 it can be seen that we also indicate the largest detection test statistic, but as discussed in Section 3.4 the overall model test statistics cannot be compared directly. In the detection plots the test statistic with the largest ratio between its size and critical value for a certain epoch is considered to be the 'largest'. The rationale for



The identification plots should be read as follows. On the horizontal axis the fix number is plotted (here fixes 25 to 39 are considered). On the vertical axis the identification test statistics are given (`gso(i)` denotes the global slippage test for outliers in the observations with delay `i`; `gso(0)` corresponds with the local slippage test). In subsequent plots `gss` refers to slips in the observations. All test statistics that are larger than their critical value are indicated by a hyphen (-). The identification results are adjusted for their delay. The largest identification test statistic for a particular epoch and type of hypothesis is denoted by a *. From the figure it follows for example that the global slippage tests for outliers is larger than its critical value for delay 0 at fix 30 (`gso(0)` plotted at fix 30 corresponds to $t^{l,k} = t^{30,30}$), for delay 1 at fix 31 (`gso(1)` plotted at fix 30 corresponds to $t^{30,31}$) a.s.o. Because the identification tests statistics are plotted adjusted for their delay one can see that the outlier simulated at fix 30 (indicated by `e`) is correctly identified.

Figure 6.2: Interpretation of the identification plots.

flagging the largest detection test statistic is to investigate whether the detection test statistic can be used to identify the time of occurence of the model error.

We will start the analysis of the performance of the detection and identification (DI) procedure with the simple linear, two-dimensional model introduced in Section 5.3. The results pertaining to this model will be discussed extensively, partly because the simple model allows us to give a full account of the results, and partly to illustrate the use of the analysis tools. Thereupon we will consider the navigation models of Section 5.5. A theoretically correct analysis of the performance of the DI procedure is based on the assumption that all datasets contain a single model error. We do not adapt for model errors at this point of our investigations and consequently the filtered estimators are biased (and also the predicted residuals on which all test statistics are based) after the occurrence of the model error. Nevertheless we have also chosen to analyse datasets containing multiple errors. In that case the detection and identification plots should rather be considered as 'sensitivity' plots, illustrating the response of the test statistics to a sequence of model errors.

6.2.1 Testing Parameters

For the DI computations the level of significance (α_0) and the power of the test (γ_0) are chosen identical to the values used in the design computations. For the one-dimensional tests we choose $\alpha_0 = 0.001$ and $\gamma_0 = 0.80$. The levels of significance of the multi-dimensional tests are determined by the B-method of testing.

6.2.2 Detection and Identification — Linear Models

In this section we consider the simple (two-dimensional) model introduced in Section 5.3 and described by (5.1) to (5.3) with starting values as given by (5.4). For models LM1 and LM2 four datasets each have been simulated. A description of the datasets is provided in Tables 6.1 and 6.2.

Dataset with a single outlier					
		MDBs outliers observations [m]			
fix	simulated	model LM1		model LM2	
number	error [m]	delay 0	delay 1	delay 0	delay 1
30	20.0	8.38	5.23	5.17	4.73

Dataset with multiple outliers					
		MDBs outliers observations [m]			
fix	simulated	model LM1		model LM2	
number	error [m]	delay 0	delay 1	delay 0	delay 1
10	20.0	8.38	5.23	5.29	4.79
20	18.0	8.38	5.23	5.17	4.73
30	16.0	8.38	5.23	5.17	4.73
40	14.0	8.38	5.23	5.17	4.73
50	12.0	8.38	5.23	5.17	4.73
60	10.0	8.38	5.23	5.17	4.73
70	8.0	8.38	5.23	5.17	4.73
80	6.0	8.38	5.23	5.17	4.73
90	4.0	8.38	5.23	5.17	4.73

Table 6.1: Description of the datasets with (an) outlier(s) and associated MDBs for models LM1 and LM2.

We will start the analysis by considering the dataset with the single outlier, then consider the datasets with slips and finally look at the dataset with multiple outliers. The size of the error in the second data set with slips and the size of some errors in the dataset with multiple outliers is chosen in such a way that the errors are approximately the size of the MDB.

The results pertaining to the dataset with a single outlier are given in Figs. 6.3 and 6.4. For convenience we have only plotted the time interval from fix 20 to fix 60. Figure 6.3 shows that the performance of the detection procedure is excellent. It can be seen that in this particular case the largest detection test statistic is related to the time of occurence of the model error (`gom(8)`, for example, is the largest detection test

Dataset with a single slip I					
		MDBs slips observations [m]			
fix	simulated	model LM1		model LM2	
number	error [m]	delay 1	delay 5	delay 1	delay 5
30-100	20.0	8.13	7.28	4.51	4.32
Dataset with a single slip II					
		MDBs slips observations [m]			
fix	simulated	model LM1		model LM2	
number	error [m]	delay 1	delay 5	delay 1	delay 5
30-100	7.0	8.13	7.28	4.51	4.32

Table 6.2: Description of the datasets with a slip and associated MDBs for models LM1 and LM2.

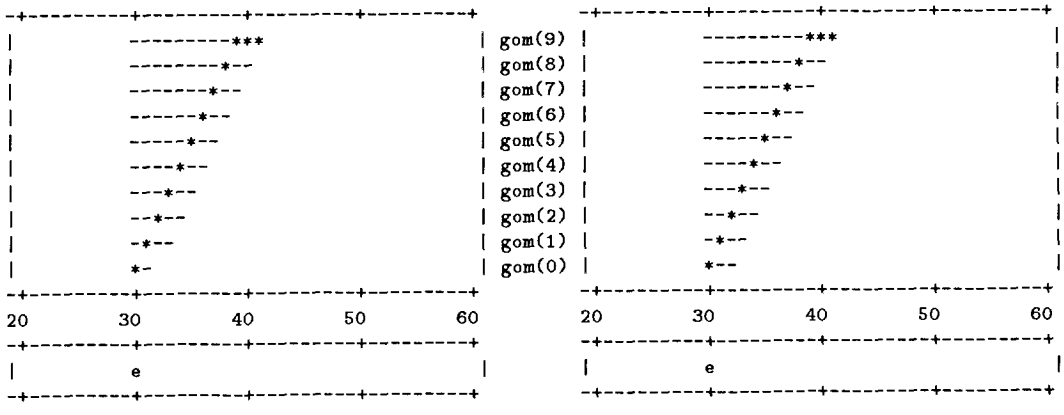


Figure 6.3: Detection Results for the dataset with an outlier at fix 30; model LM1 (left) and model LM2 (right).

statistic at $k = 38$ and thus gives reasons to believe that a model error occurred at $l = 30$). Although the detection test statistics are meant to test the overall validity of the null hypothesis, they seem, at least in this particular case, to be powerful enough to identify the time of occurrence of the model error. Fig. 6.3 also shows that every detection test statistic that incorporates the time of occurence of the model error is larger than its critical value (this results in the triangular shaped pattern). The reason that also test statistics which only refer to epochs after fix 30 are larger than their critical values (e.g. `gom(0)` at fix 31) is due to the fact that we do not adapt for the model error (and hence after fix 30 the predicted residuals are biased). Note, however, that the impact on the test statistics is quite shortlived, which is accordance with the response graphs for outliers (cf. Figs. 5.2 and 5.3).

In the identification plot (Fig. 6.4) an extension to the key given in Table 6.2 is introduced by also indicating (by means of an x) which identification test statistic is

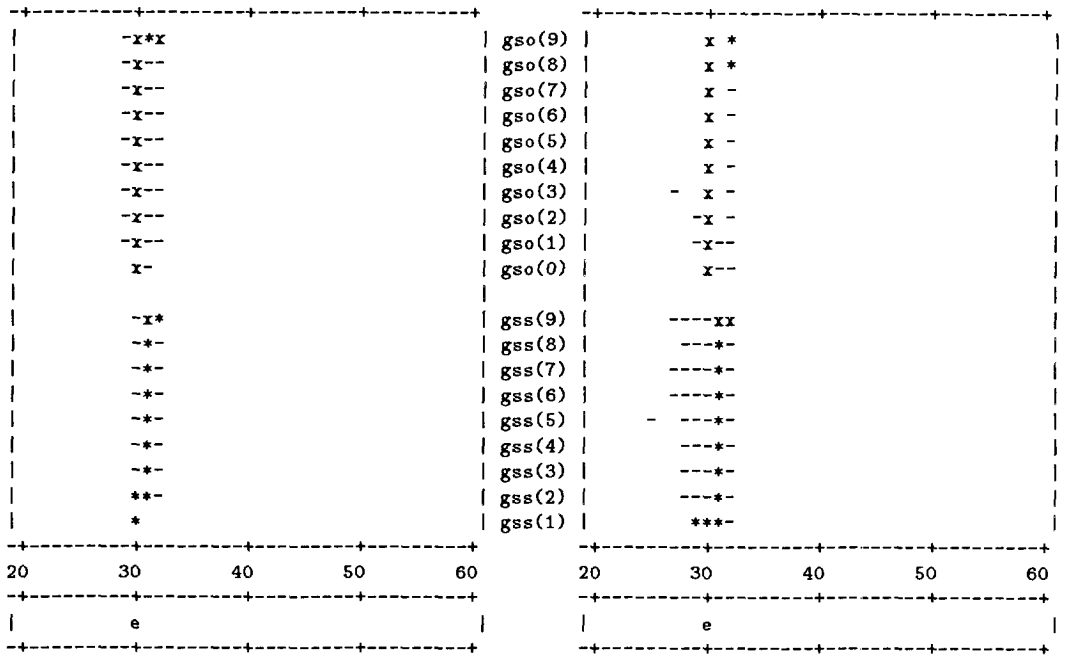


Figure 6.4: Identification Results for the dataset with an outlier at fix 30; model LM1 (left) and model LM2 (right). (An x indicates which identification test statistic is the largest for a particular epoch of computation.)

the largest for a particular epoch of computation. In Fig. 6.4 we differentiate between tests for outliers (denoted by `gso(.)`) and tests for slips (`gss(.)`). It can be seen, for example, that the tests for $l = 30$ with a delay of one $k - l = 1$ ($k = 31$), indicate that an outlier is more likely than a slip (note that at fix 30 `gso(1)>gss(1)`). The identification procedure operates correctly for both model LM1 and LM2. Especially in Fig. 6.4 it can be seen that the identification test statistics are correlated. The identification test statistic (`gso(.)`) is consistently (and correctly) the largest one until the end of the window at fix 39, but the synchronous slip test statistic and 'neighbouring' outlier and slip test statistics are also larger than their critical value.

We now turn to the dataset with the slip of 20m starting at fix 30 (from Table 6.2 it follows that the size of the slip is clearly larger than the MDB). Both the detection (Fig. 6.5) and identification (Fig. 6.6) plots show that the DI procedure works very well. In the identification plot one should note an interesting phenomenon, namely that the start of a slip is (usually) identified by the local slippage test (`gso(0)`), which might (depending on the type of alternative hypotheses specified) refer to an outlier rather than to a slip. This should be kept in mind if one wants to implement an adaptation procedure which is suited to handle slips and outliers simultaneously. The impact of the slip on the test statistics (especially for model LM2) is more pronounced than for

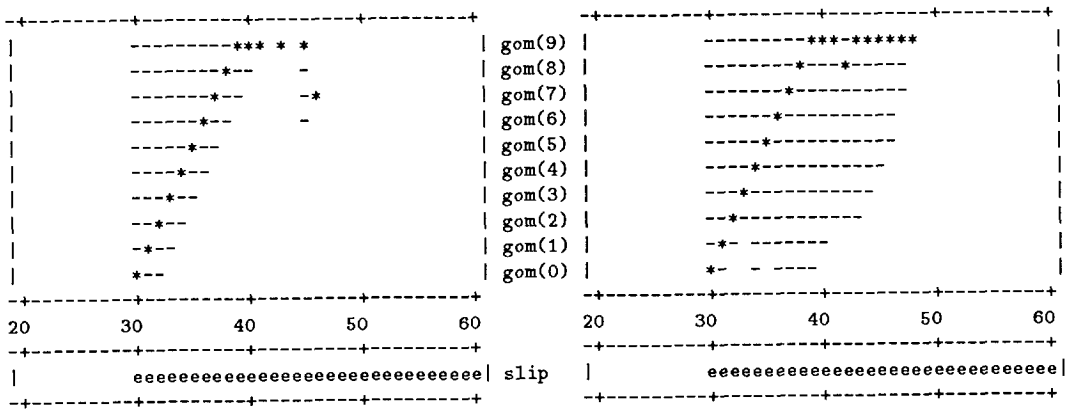


Figure 6.5: Detection Results for the dataset with a slip of 20m starting at fix 30; model LM1 (left) and model LM2 (right).

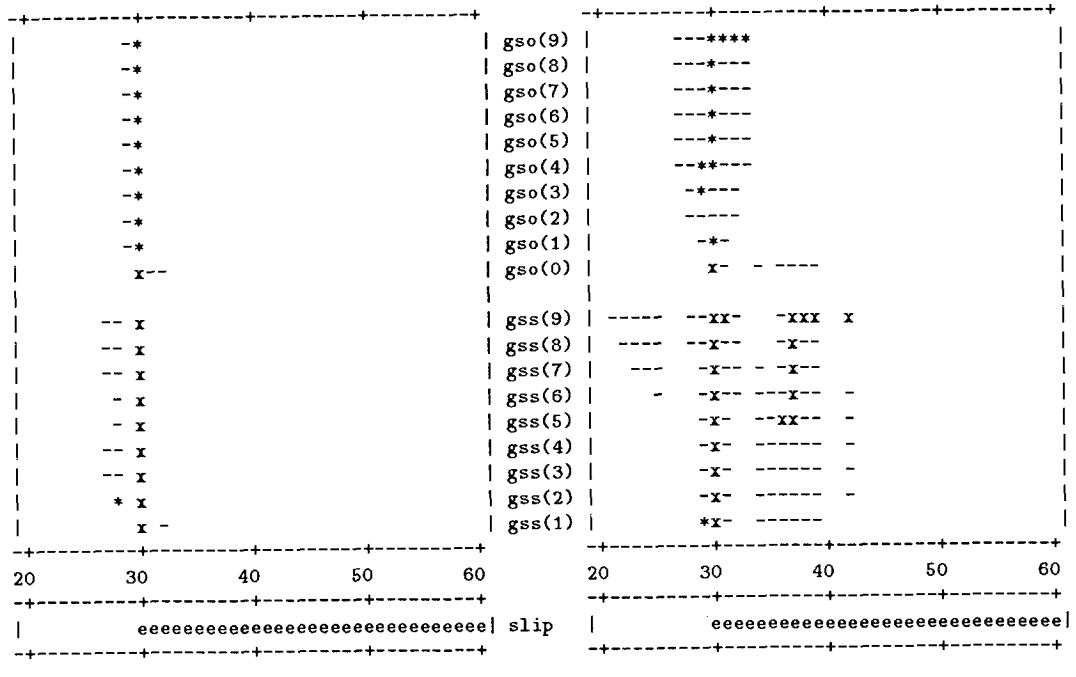


Figure 6.6: Identification Results for the dataset with a slip of 20m starting at fix 30; model LM1 (left) and model LM2 (right).

the outlier; this is in accordance with the response graphs for slips (cf. Figs. 5.2 and 5.4).

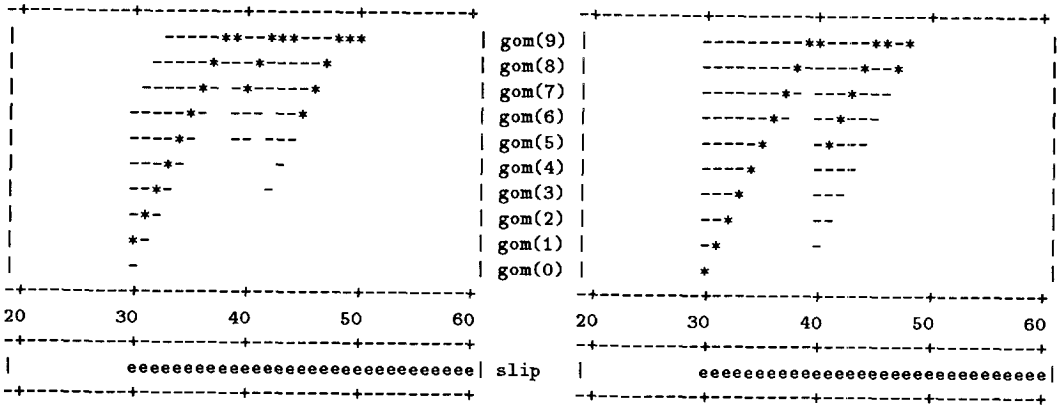


Figure 6.7: Detection Results for the dataset with a slip of 7m starting at fix 30; model LM1 (left) and model LM2 (right).

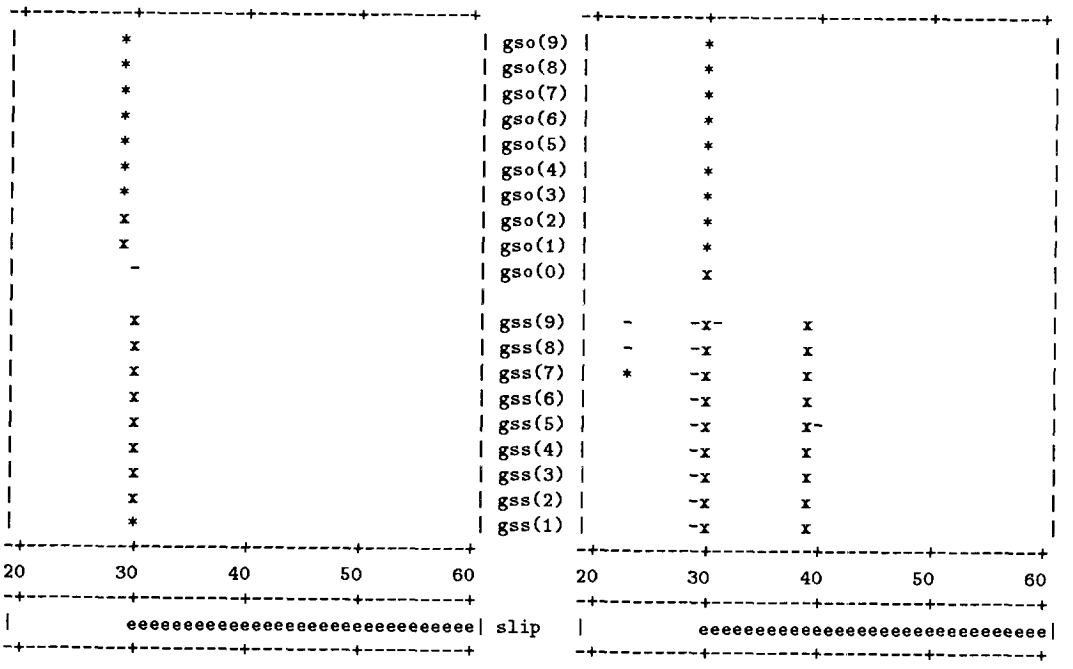


Figure 6.8: Identification Results for the dataset with slip of 7m starting at fix 30; model LM1 (left) and model LM2 (right).

If one considers the dataset with the slip of 7m (which is approximately the size of the MDB, cf. Table 6.2), the performance of the DI degrades somewhat when compared

to the case with the slip of 20m. Comparing Figs. 6.7 and 6.8 to Figs. 6.5 and 6.6 one sees that especially less identification test statistics are larger than their critical value. For model LM2 (where the size of the slip is larger than the MDB) the DI procedure works well once again, but for model LM1 the slip cannot so easily be identified. Also one should note that for model LM1 the largest detection test statistic no longer refers to the actual time of occurrence of the model error. This leads us to the conclusion that in general the detection test statistic may be a useful tool in identifying the time of occurence of the model error, but is not the most powerful tool to do so. The identification tests are able to pinpoint the exact time of occurence, albeit with a delay. It can be seen from Fig. 6.8 that the correct identification for model LM1 occurs with a delay of 2 (`gss(2)`). (At fixes 30 and 31 outliers with a delay of 1 and 2 (reffering to $l = 29$) are erroneously identified (cf. `gso(1)` and `gso(2)` in the upper left part of Fig. 6.8).) This case constitutes a nice illustration of the growing detection power with increasing delay.

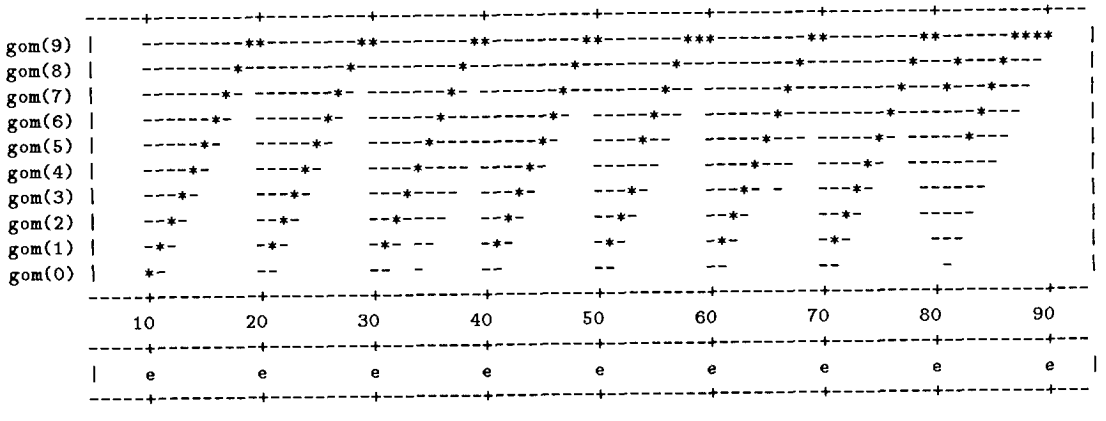


Figure 6.9: Response of detection test statistics to the dataset with multiple outliers — Model LM1.

We will end the analysis of the performance of the DI procedure by considering the response of the test statistics to a dataset with multiple outliers. Such an analysis may clarify how various model errors might cause interference between the test statistics. The results are provided in Figs. 6.9 to 6.12. For model LM1 it can be seen (cf. Figs. 6.9 and 6.10) that the outlier at fix 90 is not detected and identified, which is probably due to the fact that the simulated error is smaller than the MDB. Figure 6.10 furthermore shows that the identification fails for the local test (`gso(0)`) at fixes 30, 40, 50, 60, 70 (which is due to interference caused by the previous model error, because the interval between errors is identical to the largest windowlength of the identification tests). At fix 80 the failure of the local test is probably due to the small size of the model error. Apart from a misidentification at fix 49 (see `gso(1)`), it can be seen from Fig. 6.10 that the identification based on the global tests functions properly, even if the biases caused by the outliers are not accounted for.

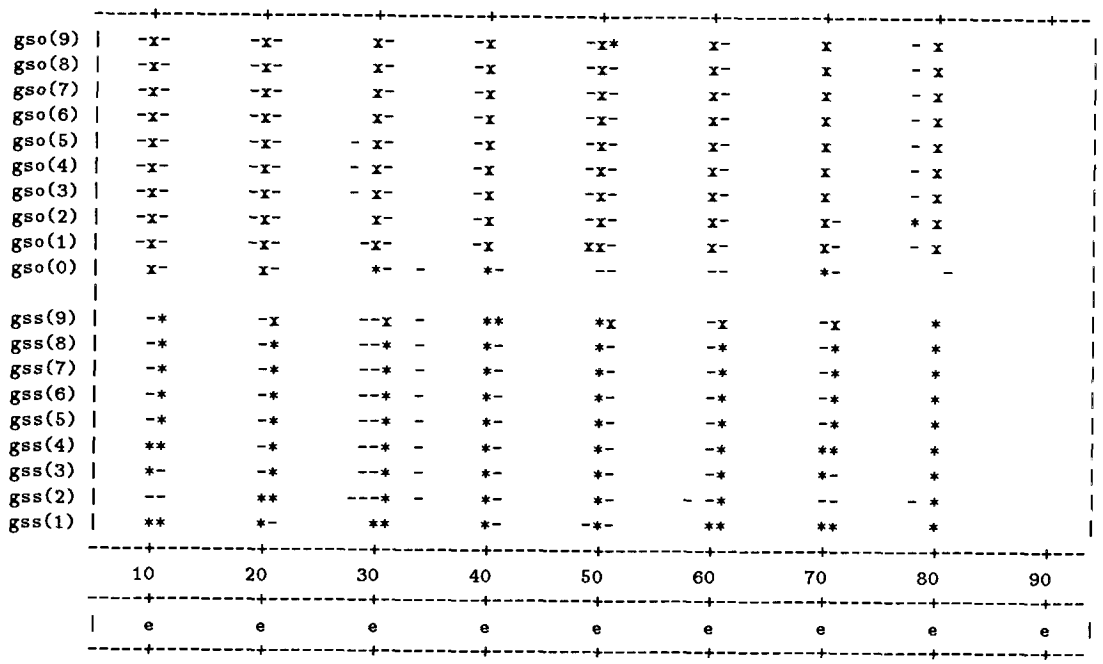


Figure 6.10: Response of identification test statistics to the dataset with multiple outliers — Model LM1.

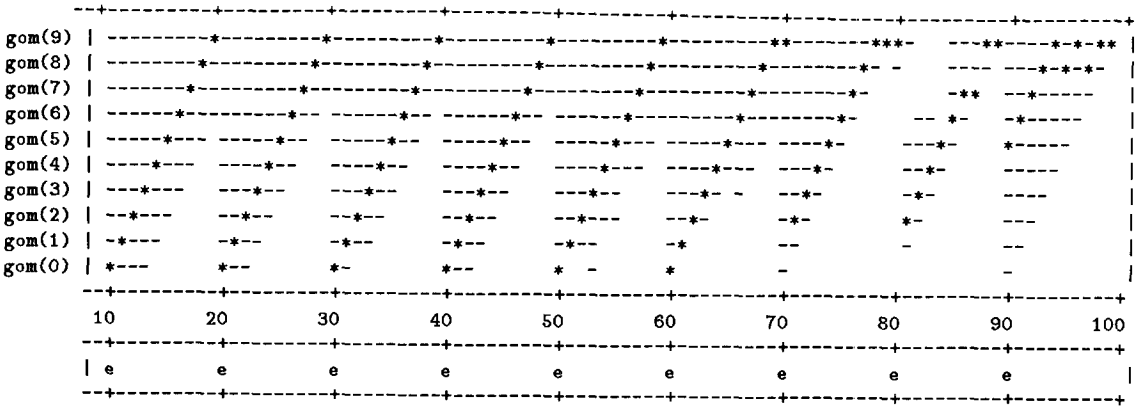


Figure 6.11: Response of detection test statistics to the dataset with multiple outliers — Model LM2.

Compared to the results of model LM1, the procedure performs better for model

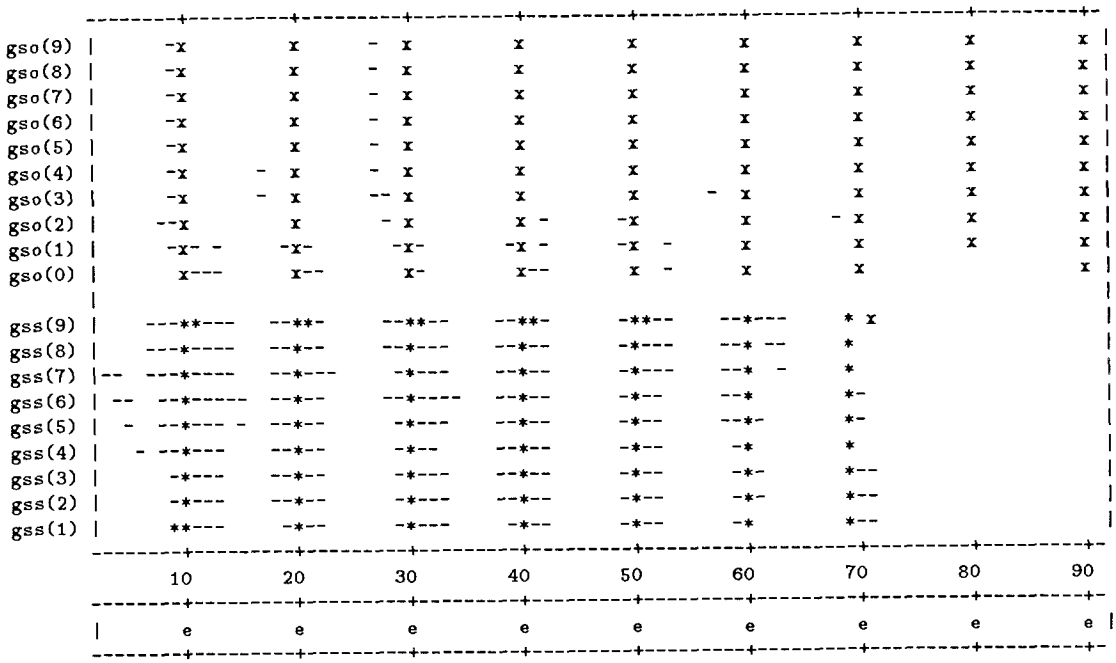


Figure 6.12: Response of identification test statistics to the dataset with multiple outliers — Model LM2.

LM2 (cf. Figs. 6.11 and 6.12). This is of course due to the fact that the MDBs associated with model LM2 are smaller than those of model LM1 (cf. Table 6.1). Except for fix 81 (at which the outlier is detected with a delay of one epoch (cf. Fig. 6.12)) the detection and identification procedure works perferctly using only the local tests. Overall it seems that the impact of the outliers vanishes so quickly that the DI results are not that much influenced.

Summarizing, the analysis of the performance of the detection and identification procedure for the linear models shows that:

- Outliers and slips are detected and identified correctly. If, however, the model error is smaller than the MDB, detection and identification, as expected, might fail. This confirms that the MDBs are a useful measure for the size of errors that can be identified.

- The identification test statistics for different types of alternative hypotheses (outliers and slips) and for various delays are correlated, but the largest test statistic does indicate the actual model error.

- In case no identification test statistics are computed, the 'largest' detection test statistic might be used to infer at which time the model error occurred. For this

purpose it is, however, not the most powerful test.

- The analysis of the dataset with the multiple outliers shows that the performance of the DI procedure might deteriorate if model errors occur (nearly) simultaneously.

6.2.3 Detection and Identification — Navigation Models

Now the performance of the detection and identification procedure has been extensively discussed for a simple linear model, we proceed with the navigation models which were introduced in Chapter 5. We will limit the analysis of the DI procedure to three different datasets. An extensive analysis of the navigation models revealed that one basically arrives at the same conclusions as have been obtained for the linear models. A major extension of the navigation models with respect to the linear cases is that one uses more than one measurement at every epoch. Therefore this section will primarily serve to demonstrate that the DI procedure indeed indicates the correct alternative hypothesis. In particular we consider models 1a (4 Syledis ranges, gyro, and log) and model 1b (4 Syledis ranges only), which have been specified in Section 5.5. All the datasets have been simulated without disturbances (i.e. $Q = 0_n$ was used in the simulations). Note, however, that the simulated trajectory contains a curve which, with our constant velocity models, can be considered as an unmodelled acceleration across track of 0.393m/s^2. The datasets considered are an errorless dataset for model 1a, a dataset with outliers for model 1a, and a dataset with outliers for model 1b. A description of the datasets with outliers is given in Table 6.3. All datasets have been processed with overall model tests and slippage tests with a lag 0 and a delay 9 (i.e. a window length of 10). We test for alternative hypotheses related to outliers and slips in the measurements, accelerations along and across track and gyro and log instrumental biases (the latter two only for model 1a).

fix number	observation	simulated error [m]	MDBs outliers observations [m] model 1a delay 0	model 1a delay 1	model 1b delay 0	model 1b delay 1
20	range 1	16.0	6.64	6.54	7.26	6.73
30	range 2	14.0	6.64	6.55	7.44	6.82
40	range 3	12.0	6.47	6.41	6.95	6.57
50	range 4	10.0	6.46	6.41	6.96	6.58
60	range 1	8.0	6.56	6.49	7.26	6.73
70	range 2	6.0	6.56	6.50	7.44	6.82
80	range 3	8.0	6.48	6.42	6.95	6.57
90	range 4	10.0	6.49	6.42	6.96	6.58

Table 6.3: Description of the datasets with outliers and associated MDBs for models 1a and 1b.

We start the analysis of the DI-procedure for the navigation models with the errorless dataset. Although no errors have been introduced in the dataset it is obvious that

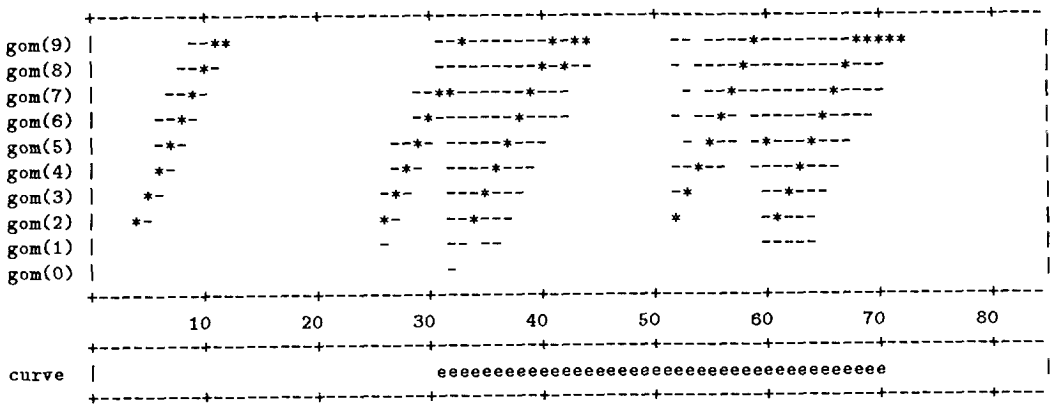


Figure 6.13: Detection results model 1a (errorless dataset).

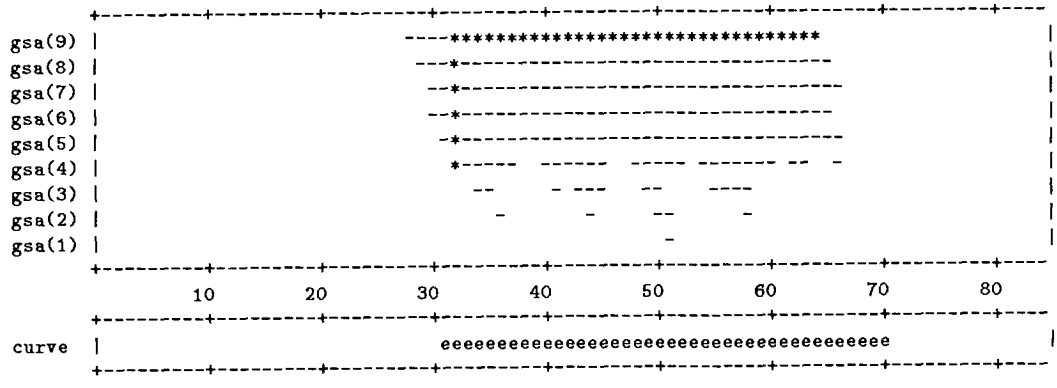
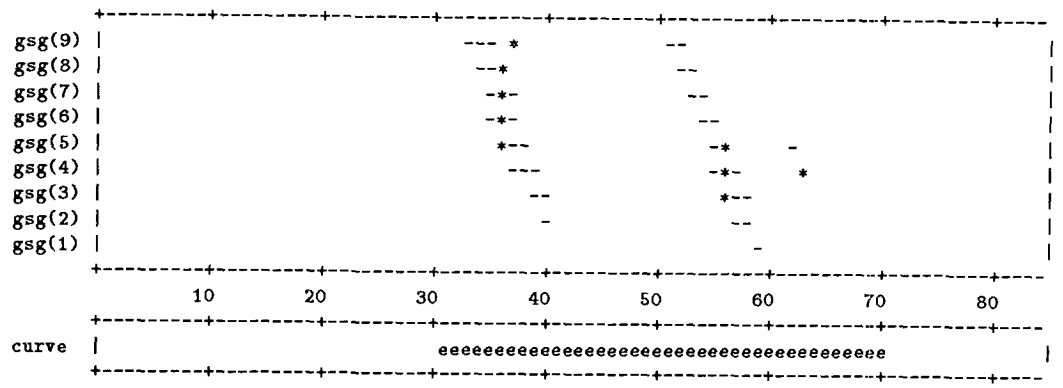


Figure 6.14: Identification results model 1a — acceleration across track (errorless dataset).

the turn in the trajectory constitutes a deviation from the assumed constant velocity model. In Fig. 6.13 the detection results are given. It can be seen that rather many detection test statistics exceed their critical value. In the fix interval 1 to 30 (in which no model misspecification is supposed to exist) the ratio of the test statistic and its critical value is at most 1.16. (In the fix interval 4 to 12 no simultaneous identification takes place; at fix 27 a nonexisting slip in a range is identified with a delay of three (the size of the particular identification test statistic is 1.08 times its critical value)). In the interval 30 to 70 (during the turn) this ratio reaches values up to 1.60. The curve seems to be detected. If one considers the range of possible alternative hypotheses the acceleration across track (denoted as `gsa(.)` in Fig. 6.14) is indeed identified starting at $l = 32$, but not by the test statistics with a small delay. The MDB for an accelera-


gsg(9)
gsg(8)
gsg(7)
gsg(6)
gsg(5)
gsg(4)
gsg(3)
gsg(2)
gsg(1)
10
20
30
40
50
60
70
80
curve
eeee

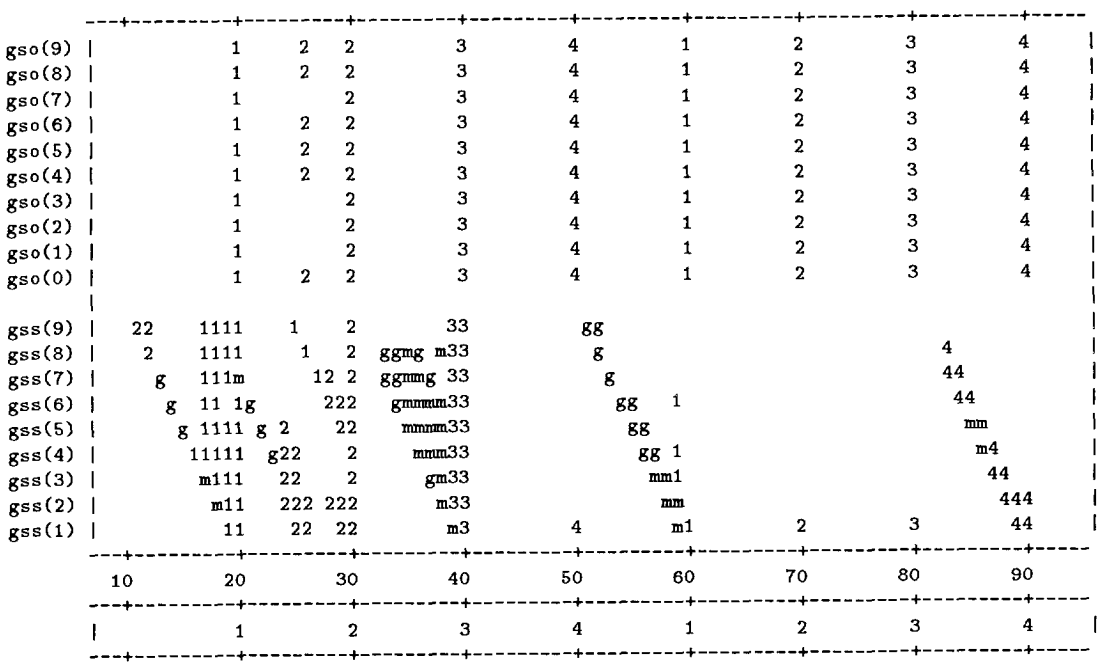


Figure 6.16: Identification results model 1a — slip (bottom) and outlier (top); 1,2,3,4 indicate ranges, g indicates gyro, and m means that more than one test statistic at a time is larger than its critical value.

fix 70) the local identification test statistic is powerful enough to identify the outlier in the range.

Also for model 1b the identification procedure performs very well. The simulated outlier is not identified at fixes 60 to 62 (where an acceleration across track is identified as the most likely hypothesis) and at fix 70 (where an nonexisting slip in range 1 is identified), but also for model 1b the range errors at fixes 60 and 70 are approximately the size of the MDBs (cf. Table 6.3). Furthermore the acceleration across track is also a model error in this case. The identification of an outlier at fix 14 (cf. Fig. 6.17) is not accompanied by a simultaneous detection and can thus be discarded.

The analysis of the performance of the detection and identification procedure for the models introduced in Chapter 5 shows that:

- Identification performs well, even if the errors in the datasets are not accounted for.

- If several model errors occur simultaneously (for example the unmodelled curve and an outlier) the identification test statistics tend to mask each other.

- The MDBs constitute a useful lower bound for the size of the errors which can be identified.

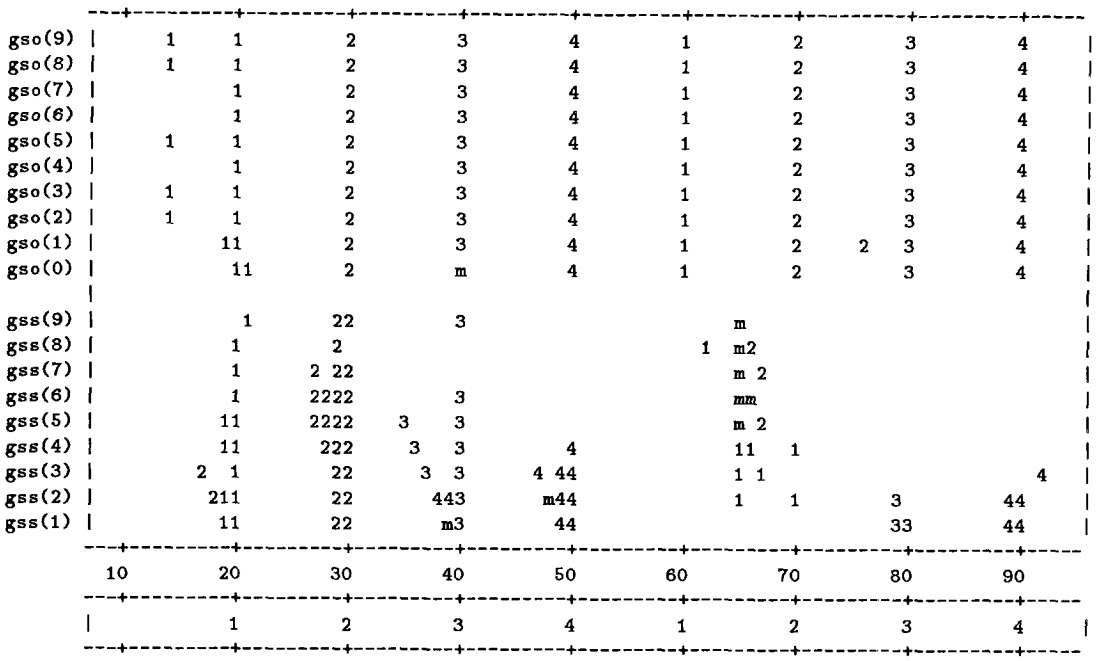


Figure 6.17: Identification results model 1b — slip (bottom) and outlier (top); 1,2,3,4 indicate ranges, m indicates more than one range test statistic is larger than its critical value.

6.2.4 Detection and Identification — Concluding Remarks

We have investigated the performance of the detection and identification (DI) procedure for a simple linear model and two navigation models. Because we do not adapt for the model errors at this stage, the performance of the DI-procedure can only be established exactly for cases with one model error, because after the occurrence of the first model error the filter solution is biased. We find, however, that the response of the DI test statistics, for errors far enough apart, largely corresponds to the cases where only one model error is studied at a time. The analysis of the DI procedure shows that:

- The MDB is a very useful measure of the detectability and model errors smaller than the MDB are indeed rarely identified. We have furthermore seen that a larger window length of the identification test statistics is benificial for the identification of small errors (at least for those hypotheses where the MDB decreases with increasing window length).

- Local testing is adequate for the detection and identification of large model errors.

- The detection procedure works well. If no explicit identification step follows the detection phase, the detection test statistic might be helpful for the determination of the time of occurence of the model error.

- The identification procedure performs excellently, although the identification test statistics are correlated (and thus more than one identification test statistic is larger than its critical value). Despite the correlation the largest identification test statistic indicates the actual model error.

- If errors occur (nearly) simultaneously the DI-procedure might have some problems to identify all errors correctly.

Considering the good performance of the DI procedure, one has to be aware of the fact that in practice only those model errors are identified for which a specific alternative hypothesis has been specified. A model error for which no matching alternative hypothesis is specified, probably results (due to the correlation between the test statistics) in the erroneous identification of some other model error. Thus the range of likely model errors has to be given careful consideration in the design phase.

6.3 Adaptation

After the performance of the detection and identification procedure has been analysed, we turn our attention to the most challenging part of real-time quality assurance in dynamic systems, namely adaptation. The objective of this section is to establish how well adaptation actually works.

In Section 3.6 we have shown that the exact adaptation procedure for outliers can be implemented by updating the estimator of the filtered state and its covariance matrix once, after which one can revert to the filter under the null hypothesis. This is not possible for slip-type errors and therefore we will consider several adaptation strategies for slips. Theoretically one should, from the time of identification onwards, continuously adapt the estimators obtained by the filter under H_0. To circumvent this permanent update it has been investigated if one obtains 'acceptable' results by various sub-optimal strategies. A drawback of these sub-optimal strategies is that they lack a theoretically rigorous basis. One of these strategies is denoted the semi-exact method and reverts to the filter under H_0 after a number of steps of the exact procedure. The so-called approximate method reverts to the filter under H_0 after the adaptation step immediately. A description of the implementation of the approximate and semi-exact adaptation procedures for slips is given in Appendix C.

In this section we limit ourselves to *local* adaptation for outliers and slips in the observations. This restriction is motivated by a number of reasons. Firstly and most importantly we have very little experience in the automatic adaptation for model errors in dynamic systems. Therefore it was considered wise to limit the investigations to those based on local identification. From the analysis of the DI procedure we concluded that local tests are adequate to detect and identify model errors, which are, say, two times

as large as the MDBs. Secondly the current software does not remove the bias from the test statistics once an model error has been identified and adapted for. To prevent misidentifications the window length of the tests is temporarily reset to one, which corresponds to local testing.

The adaptation computations are based on a level of significance of $\alpha_0 = 0.001$ for the one-dimensional tests and a power of the tests of $\gamma_0 = 0.80$ in the detection and identification procedure.

In this section we will investigate the adaptation procedure for in succesion the simple linear model, the two navigation models considered in Section 6.2.3, and the lane-slip problem in hyperbolic positioning. The simple linear model once more serves to introduce and explain the peculiarities and properties of the adaptation procedure, whereas the latter two cases correspond more or less to real world situations.

6.3.1 Adaptation — Linear Models

In this section we consider the performance of the adaptation procedure for the linear models LM1 and LM2. We consider three datasets of 100 fixes for each model, which are summarized in Table 6.4. The performance of the DI procedure for the datasets with multiple outliers and a single slip has been investigated in Section 6.2.2.

dataset with outliers		dataset with slips		dataset with single slip	
fix	size [m]	fix	size [m]	fix	size [m]
10	20.0				
20	18.0				
30	16.0	21-30	7.0	30-100	20.0
40	14.0				
50	12.0				
60	10.0				
70	8.0	61-70	20.0		
80	6.0				
90	4.0				

Table 6.4: Simulated errors in the datasets concerning the linear models LM1 and LM2.

We begin with the analysis of the dataset with outliers for which the results are presented in Table 6.5[1] and Fig 6.18. The adaptation procedure operates adequately for both models: all outliers are correctly identified and adapted for, except those at fixes 80 and 90 for model LM1 and the one at fix 80 for model LM2. The failure of the identification (and thus adaptation) procedure at fixes 80 and 90 is probably due to the fact that the simulated errors are not much larger than the MDBs, which corresponds to our findings in Section 6.2.2. For model LM1 the adaptation failure causes position biases of 5 to 13 metres in the fix interval 80-82. The damping effect of a lower system

[1] In the tables a + and a - indicate correct and incorrect adaptation respectively.

		Model LM1			Model LM2		
fix	true error [m]		error estimate $\hat{\nabla}$ [m]	σ_{∇} [m]		error estimate $\hat{\nabla}$ [m]	σ_{∇} [m]
10	20.0	+	18.96	2.03	+	21.73	1.25
20	18.0	+	18.03	2.03	+	17.27	1.25
30	16.0	+	13.36	2.03	+	15.67	1.25
40	14.0	+	15.56	2.03	+	13.27	1.25
50	12.0	+	14.37	2.03	+	11.96	1.25
60	10.0	+	9.56	2.03	+	8.51	1.25
70	8.0	+	9.44	2.03	+	7.09	1.25
80	6.0	−	−.−		−	−.−	
81		−	−7.13	2.03			
82		−	−14.13	3.24			
90	4.0	−	−.−		+	4.82	1.25

Table 6.5: Adaptation results for the dataset with outliers — adaptation for outliers; model LM1 (left), model LM2 (right).

noise on the state estimates is clearly visible if one compares the unadapted position estimates in Fig. 6.18.

Next we consider the dataset with a single slip. Such a dataset allows us to compare the performance of the exact adaptation procedure (where the adaptation step is carried out ad infinitum) with the approximate adaptation procedure for slips. The results for models LM1 and LM2 are given in Fig. 6.19. (For model LM1 the variance of the error estimator is constant after fix 57 and consequently the depicted result actually corresponds to the semi-exact adaptation procedure.) For model LM2 we additionaly consider the error estimate obtained through the exact adaptation procedure (Fig. 6.20 left) and the covariance between the estimators of the error and the adapted state (Fig. 6.20 right). The precision of the error estimator does not improve much with increasing window length (the same holds for the error estimate itself). This is in accordance with our findings in Section 5.4, where we saw that the MDBs for slips do not significantly decrease with increasing window length. From Fig 6.19 it follows that the performance of the approximate and exact adaptation procedure are comparable as far as the actual position estimates are concerned (which is due to the fact that one does not obtain a really better error estimator with increasing window length). From Fig. 6.19 it can be seen, however, that the variance of the position estimator obtained by the approximate procedure is too low (that is too optimistic); the variance given by the exact procedure is consistently larger. Furthermore Fig. 6.20 shows that (using the exact adaptation method) the covariance between the error estimator and the estimator of the adapted state ($P_{\hat{x}^a_{k|k}\hat{\nabla}^{l,k}}$) does not decrease with increasing window length and is not negligible. The approximate adaptation method is thus clearly not optimal from the point of view of the description of the precision of the state estimators in case of the linear models. In Fig. 6.19 one can see an interesting phenomenon in

the case of exact adaptation. The precision of the position estimator first deteriorates after adaptation and then (after some fixes) falls back gradually to a somewhat lower (though still elevated) level. (The same holds for the covariance between the estimators of the error and the position in Fig. 6.20). The variance of the adapted state estimator is primarily influenced by the behaviour of the response matrix $\overline{X}_{k,l}$ through $P^{a}_{k|k} = P^{0}_{k|k} + \overline{X}_{k,l} Q_{\hat{\nabla} l,k} \overline{X}^{T}_{k,l}$, because the variance of the error and position estimators is (more or less) constant. In Fig. 5.3 it can be seen that the impact of a slip of fixed size on the position estimate has identical characteristics (particularly the resemblance of Fig. 6.20 to Fig. 5.3 is striking).

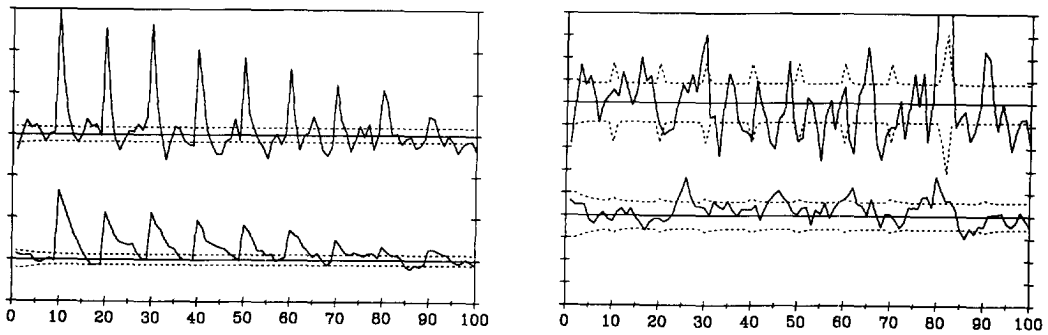


Figure 6.18: Datasets with outliers; Left: Position estimates $(+/-1\sigma)$ vs. the fix number without adaptation (LM1 top; LM2 bottom) with 1 tickmark corresponding to 5m; Right: Position estimates vs. the fix number with adaptation for outliers (LM1 top; LM2 bottom) with 1 tickmark corresponding to 1m.

The results of the adaptation procedure for the datasets with multiple slips for model LM2 are given in Figs. 6.21, where four scenarios are presented, namely no adaptation, adaptation for outliers, approximate adaptation for slips, and (semi-)exact adaptation for slips. (In this particular case the exact procedure was stopped if the decrease in variance of the error estimator between successive fixes was smaller than 1.0E-4 in order to enable identification of the end of the slip after 10 fixes.) Figure 6.21 shows that the strategy based on the adaptation for outliers is not satisfactory. The slip is conceived as sequence of outliers, a phenomenon which was already pointed out in our analysis of the identification procedure. After a number of adaptation steps (which in this particular case correspond to discarding the observation), the variance of the predicted state estimator (at fix 27) becomes so large that the next erroneous observation is not identified anymore and divergence occurs. The performance of the semi-exact and approximate adaptation procedure is comparable (except that the approximate method underestimates the variance of the adapted state).

Unfortunately the dataset with multiple slips cannot be processed satisfactorily for model LM1. Although the first slip is detected at fix 21, both the semi-exact and approximate adaptation procedure fail at the end of the slip at fix 30, after which divergence occurs. This can be easily understood if one considers the properties of the

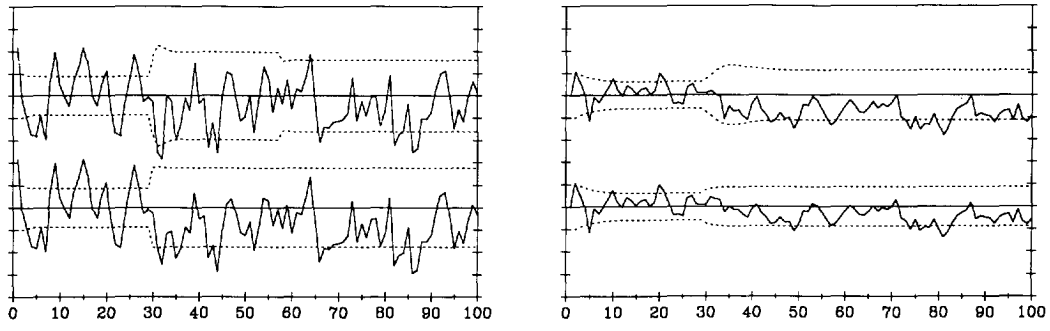


Figure 6.19: Datasets with a single slip; Position estimates $(+/-1\sigma)$ vs. the fix number for the exact (top) and approximate (bottom) solutions for models LM1 (left) and LM2 (right); 1 tickmark corresponds to 1m).

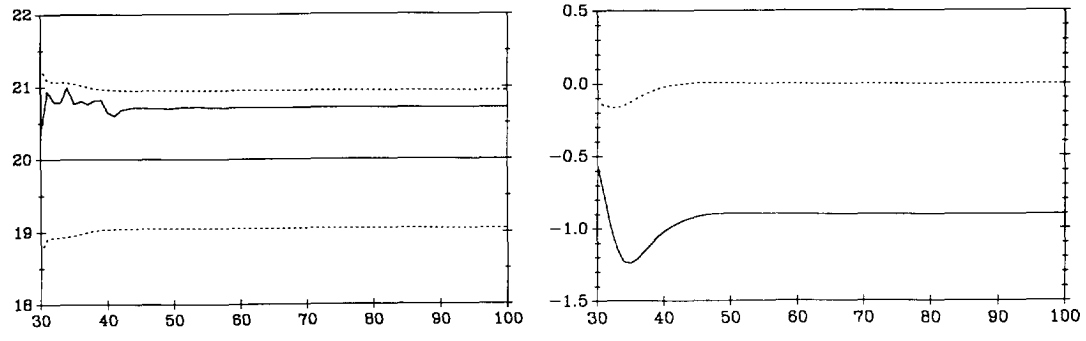


Figure 6.20: Left: Error estimate slip [m] for the dataset with a single slip $(+/-1\sigma$ relative to the simulated error of 20m); Right: Covariance between estimators of the error and position (solid line; $[m^2]$) and velocity (dashed line; $[m^2/s]$) — model LM2.

approximate adaptation procedure. After the adaptation at fix 21 the variance of the 'corrected' position observable is now 3.03m (cf. Table 6.5), which results in a MDB of 12.3m at fix 30. Clearly this MDB is much larger than the size of the jump of 7m at fix 30 and consequently it is not detected.

The analysis of the operation of the adaptation procedure for the linear models gives rise to the following (preliminary) conclusions:

- The adaptation procedure works better, if one only tests and adapts for those model errors which really occur. Otherwise slips are often identified as a sequence of outliers or some sequence of outliers may give rise to the identification of a non-existing slip. In such cases divergence may occur.

- The adaptation procedure works very well for outliers.

- The adaptation procedure works properly if a single slip is present in the data. For the slips in linear models considered in this section, the exact adaptation

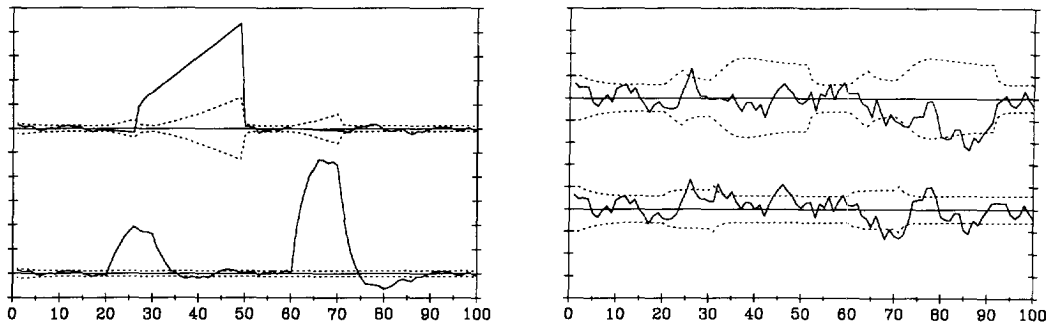


Figure 6.21: Dataset with multiple slips; Position estimates model LM2 $(+/-1\sigma)$; Left: no adaptation (bottom); adaptation for outliers (top) 1 tickmark corresponds to 5m; Right: approximate solution (bottom); semi-exact solution (top); 1 tickmark corresponds to 1m.

method does not provide better state estimates than the approximate procedure. The approximate method, however, suffers from the fact that it underestimates the variance of the adapted estimator and neglects the considerable covariance between the estimators of the error and adapted position.

- The precision of the error estimator does not significantly decrease with increasing window length.

- The performance of the adaptation procedure varies in case of multiple slips. For the model with low system noise (LM2) it performs rather well, but for the model with the larger system noise the performance is unsatisfactory.

6.3.2 Adaptation — Navigation Models

In Section 6.2.3 we discussed the performance of the detection and identification procedure for some of the navigation models introduced in Chapter 5. In this section we consider the performance of the adaptation procedure for the models 1a and 1b and datasets with outliers (cf. Table 6.3) and slips (cf. Table 6.7 furtheron). All datasets are based on the trajectory given in Section 5.5 and contain 100 fixes. Identification (and hence adaptation) is limited to local tests for either slips or outliers in the ranges and outliers in the dead reckoning observables.

We will first consider the performance of the adaptation procedure for the dataset with outliers. The adaptation procedure works very well as can be seen in Table 6.6 and Fig 6.22. At fix 70 the outlier is not identified for model 1b, but it follows from Table 6.3 that the size of the error is considerably smaller than the MDB (at fix 70 the LOM-test, however, is rejected). The estimation results (after adaptation) are presented in so-called 'filtered-true' difference plots, in which the difference between the estimated and true position is given vs. the fix number. The difference filtered-true

			model 1a			model 1b		
		true error		error estimate	σ_∇		error estimate	σ_∇
fix	range#	∇ [m]		$\hat{\nabla}$ [m]	[m]		$\hat{\nabla}$ [m]	[m]
20	1	16.0	+	16.92	1.61	+	16.69	1.76
30	2	14.0	+	13.86	1.61	+	15.25	1.80
40	3	12.0	+	11.92	1.57	+	13.18	1.68
50	4	10.0	+	7.91	1.56	+	12.07	1.68
60	1	8.0	+	8.28	1.59	+	6.71	1.76
70	2	6.0	+	5.96	1.59	−		
80	3	8.0	+	8.98	1.57	+	8.45	1.68
90	4	10.0	+	11.39	1.57	+	10.37	1.68

Table 6.6: Adaptation results for dataset with outliers; local adaptation for outliers — models 1a and 1b.

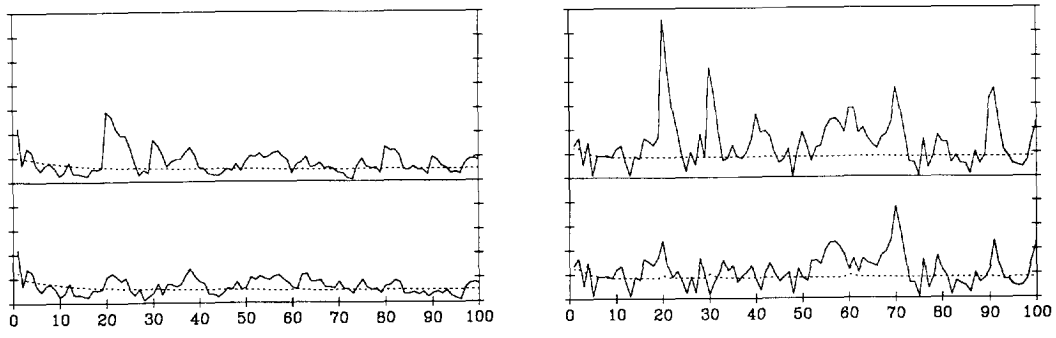


Figure 6.22: Position differences filtered-true [m] vs. the fix number; model 1a without adaptation (top left), model 1a with adaptation (bottom left); model 1b without adaptation (top right), model 1b with adaptation (bottom right); one tickmark corresponds to 1 metre.

is defined as $\sqrt{(\hat{E}-E_t)^2+(\hat{N}-N_t)^2}$ where $\hat{E}$ and E_t are the estimated and true (i.e. simulated) Easting and $\hat{N}$ and N_t are the estimated and true Northing respectively. In Fig. 6.22 the position differences for the filter without and with adaptation for outliers are depicted (in addition we have plotted the length of the semi-major axis of the point standard ellipse (represented by the short dashed line)). It can be seen that the outliers have a more profound effect on the coordinate estimators of model 1b (which lacks the damping effect of the dead reckoning observations). The non-identified outlier at fix 70 is clearly visible. Overall the adaptation procedure for outliers performs excellently.

Next we turn to the analysis of the datasets with slips in the range measurements (see Table 6.7 for a description of the datasets). We will extensively analyse the dataset with the single slip. We do not only consider how well the (position) state is estimated after adaptation (Fig. 6.23), but we also consider the error estimate $\hat{\nabla}^{l,k}$ and its associated standard deviation in Fig. 6.24. In Fig. 6.25 the precision of the (position)

estimator after adaptation is given and finally the covariance between the error estimator and the estimator of adapted state is depicted in Fig. 6.26.

dataset with a single slip			dataset with two slips		
fix interval	range#	size [m]	fix interval	range#	size [m]
26-100	1	20.0	26-100	1	20.0
			76-100	2	20.0

Table 6.7: Simulated errors in the datasets with slips concerning models 1a and 1b.

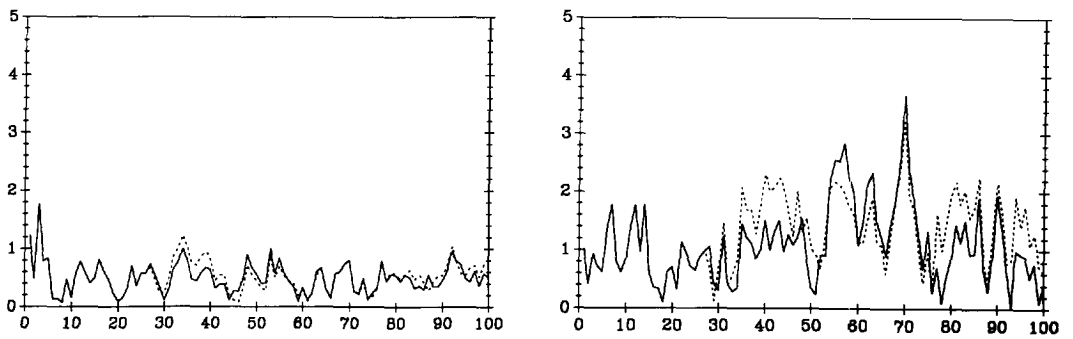


Figure 6.23: Position differences filtered-true [m] vs. the fix number — comparison of approximate (dashed line) and exact (solid line) adaptation methods; model 1a (left), model 1b (right).

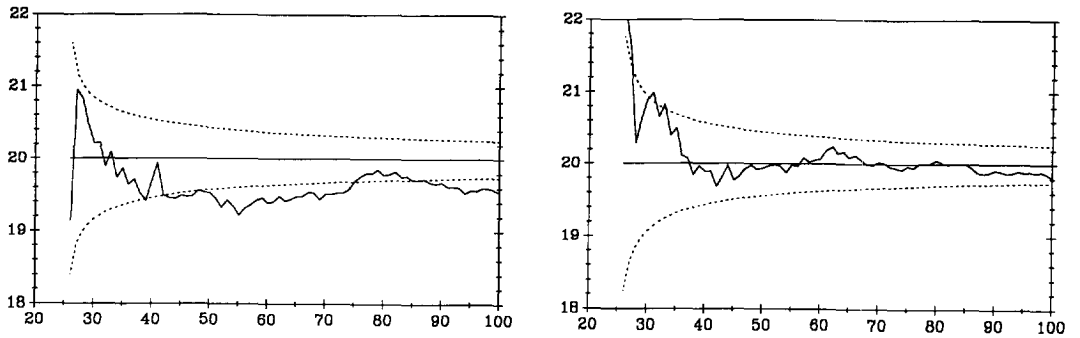


Figure 6.24: Error estimate slip [m] ($+/-1\sigma$ relative to the simulated error of 20m) vs. the fix number for the exact adaptation method; model 1a (left), model 1b (right).

Considering the filtered-true difference plots in Fig. 6.23, both the exact and approximate adaptation procedure operate well, although the exact method slightly outperforms the approximate method. It follows from Fig. 6.24 the precision of the error estimator improves with increasing window length (for the linear model the improvement with increasing window length was almost negligible after about 10 fixes). This

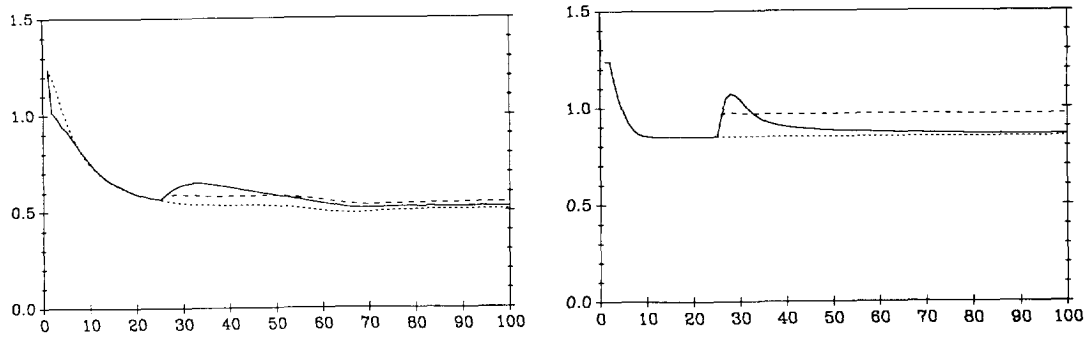


Figure 6.25: Length semi-major axis point standard ellipse [m] vs. the fix number — design (dotted line), filter with approximate (dashed line) and exact (solid line) adaptation; model 1a (left), model 1b (right).

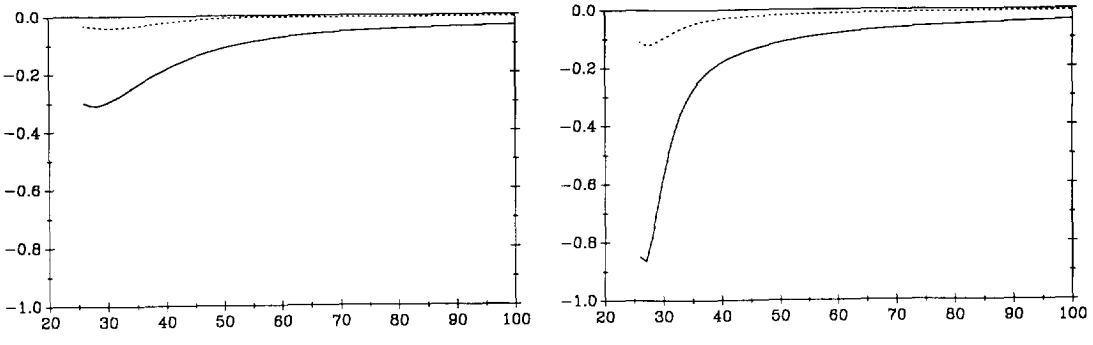


Figure 6.26: Covariance between error estimator and Easting (solid line) [m^2] and Northing (dashed line) [m^2] vs. the fix number; model 1a (left), model 1b (right).

improvement is probably due to the much higher redundancy in the navigation models. Besides the standard deviation of the error estimator for slips in the ranges does not benefit much from integration if one compares the ranges-only model (1b) to the integated model (1a). Apparently the contribution provided by the dead reckoning observables is 'local', providing primarily relative position information.

Contrary to our findings in the previous section the precision provided by the filter based on the approximate adaptation procedure is only too optimistic compared to the exact solution for a relatively short duration after the adaptation step (cf. Fig. 6.25). Due to the continuous improvement of the precision of the error estimator, the precision of the (coordinate) estimators based on the exact adaptation procedure drops below that of the approximate method. As a consequence the approximate adaptation method is not likely to cause divergence.

In Fig. 6.26 we see that the covariance between the error estimator and (elements of) the adapted state decreases with increasing window length, and more rapidly for model 1b. The difference in size of the covariance between models 1a and 1b is mainly

due to the difference in precision of the state estimator itself (In Fig. 6.25 it can be seen that the precision associated with model 1a is better than that for model 1b). The difference between the Northing and Easting component of the covariance is dependent on the relative receiver-transmitter geometry.

Overall the exact and approximate adaptation procedures perform adequately for the dataset with a single slip. If one operates a filter in conjunction with the exact adaptation procedure for slips, it can be seen (at least for models 1a and 1b) that after a reasonable number of fixes (say 50) the covariances between the error and state estimators and the variance of the error estimator become small. One can then consider to revert to the filter under H_0 combined with continuous corrections to the observations.

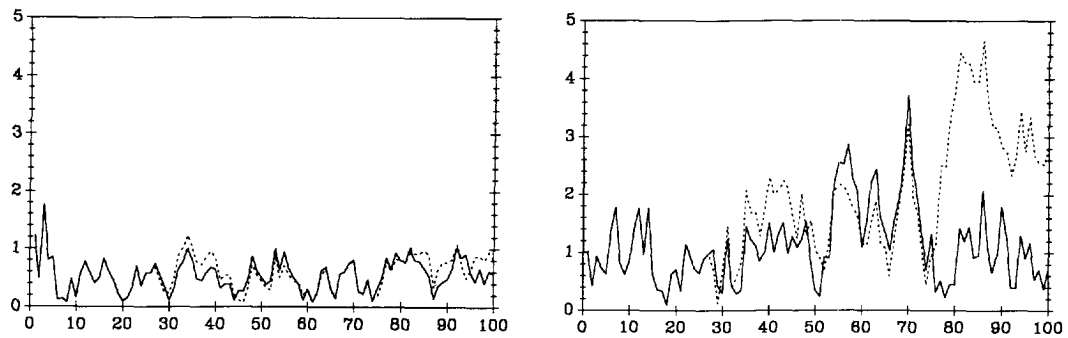


Figure 6.27: Position differences filtered-true [m] — comparison of approximate (dashed line) and semi-exact (solid line) adaptation methods; model 1a (left), model 1b (right).

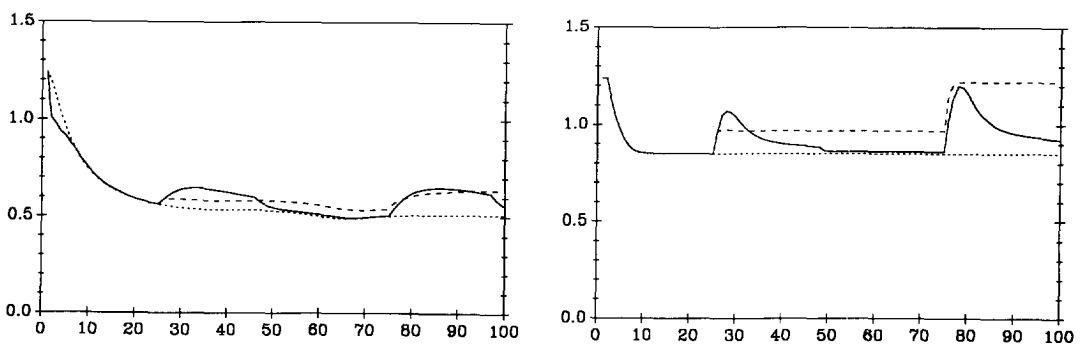


Figure 6.28: Length semi-major axis point standard ellipse [m] — design (dotted line), filter with approximate (dashed line) and semi-exact (solid line) adaptation; model 1a (left), model 1b (right).

The dataset with the multiple slips (cf. Table 6.7) has been processed using the approximate and the semi-exact adaptation procedure described in Appendix C. We had to invoke the semi-exact procedure (with a stop criterion of 0.01) in order to be able to identify the second slip. The estimation results are given in Figs. 6.27 and 6.28.

For models 1a and 1b the slips are correctly identified. Especially for model 1b the semi-exact method works better than the approximate adaptation procedure. Once more it can be seen that precision obtained by the approximate method is worse than that provided by the (semi-)exact method a number of fixes after the adaptation step.

Summarizing, the following conclusions can be drawn from the analysis of the navigation models:

- The adaptation procedures works properly for outliers. The MDB proves (once again) to be a useful design tool.
- In case of slip-type model errors the actual performance of the approximate and (semi-)exact adaptation procedure is comparable. The approximate method does not consistently underestimate the precision of the state estimators after adaptation.
- The precision of the error estimator decreases with increasing window length of the tests and consequently the covariance between the error estimator and the adapted state estimators decreases. This provides the opportunity to revert to the filter under H_0, even in case of slips.

6.3.3 Adaptation Procedure — Lane Slip Problem

In this section we perform a case study which is related to the lane slip problem in hyperbolic positioning (systems). We consider the lane slip problem for a number of reasons. A lane slip is an error that is much larger than the MDB and allows us to investigate the performance of the adaptation procedure for large size slips which also occur in practice (in our case a lane slip corresponds to an error (expressed as distance difference) of 140m). Secondly we want to investigate how the level of integration influences the performance of the adaptation procedure. For a precise analysis one would have to consider one lane slip at a time, but in this section we take a more ambitious approach. We have not only introduced two lane slips in the familiar dataset of 100 fixes, but we also consider the effect of a possible mismatch between the actual error and the error as specified by the alternative hypothesis. In this section we can therefore assess the performance of the adaptation procedure in the presence of very large errors and for model errors which do not exactly match the errors described the alternative hypotheses. The data will be processed with the semi-exact adaptation procedure, because more than one error is present at a time.

All datasets are situated in the design area 2 and are described in Table 6.8. In dataset II the second simulated slip does not fully correspond with the step-function characteristic which is assumed by the alternative hypothesis for slips.

The datasets given in Table 6.8 are processed with models 2b (2 hyperbolic patterns), 2c (2 hyperbolic patterns and 2 ranges), 2f (2 hyperbolic patterns and dead reckoning) and 2g (2 hyperbolic patterns, 2 ranges, and dead reckoning). At fix 21 the MDBs associated with the local tests for the hyperbolic pattern Renesse-Texel

Dataset I		
fix	observation	simulated error [lanes]
21-100	hyperbolic pattern Renesse-Texel	1.0
81-100	hyperbolic pattern Renesse-Pakefield	1.0
Dataset II		
fix	observation	simulated error [lanes]
21-100	hyperbolic pattern Renesse-Texel	1.0
80	hyperbolic pattern Renesse-Pakefield	0.5
81	hyperbolic pattern Renesse-Pakefield	1.0
82	hyperbolic pattern Renesse-Pakefield	1.25
83 - 100	hyperbolic pattern Renesse-Pakefield	1.0

Table 6.8: Description of the datasets with lane slips.

are 0.056λ, 0.049λ, 0.043λ, 0.042λ for models 2b, 2c, 2f, and 2g respectively (with $\lambda = 140$m). Similarly the MDBs at fix 81 for the pattern Renesse-Pakefield are 0.056λ (2b), 0.074λ (2c), 0.066λ (2f), and 0.066λ (2g). We have chosen to present the performance analysis of the adaptation procedure by means of plots of the estimated trajectory. These plots are somewhat less detailed than the difference plots used in the previous section, but provide a better overview of the overall performance of the various adaptation strategies. We look at the estimated trajectory provided by a filter without adaptation, and filters based on the approximate and semi-exact adaptation procedures. The left-hand plots in Figs. 6.29, 6.30, 6.32, and 6.34 depict the true trajectory and the trajectory estimated by the filter (for dataset I) without adaptation. In the plots in the middle and on the right-hand side of these figures the dashed-dotted line indicates the estimated track obtained by means of the approximate adaptation procedure and the solid line indicates the track obtained with the semi-exact adaptation procedure with a stop criterion of 0.01 (for case 2b we have used 0.001). For cases 2c and 2f we also consider the solution obtained by deleting (instead of adapting for) erroneous data (cf. Figs. 6.31 and 6.33). During the processing we limit ourselves to local slippage tests for slips for the Hyperfix measurements and outliers in the ranges and dead reckoning observations.

We begin our analysis by considering the results for model 2b in Fig 6.29. During the processing of both datasets a (non-existent) error was identified at fix 49 (approximately in the middle of the curve). Only the (semi-)exact adaptation procedure recovers from this misidentification and gives a well estimated track for dataset I. The start of the second slip in dataset II cannot be handled properly by either the approximate or the semi-exact adaptation procedure. The exact procedure cannot track the change in the size of the error, whereas the approximate method fails to identify the decrease of the error from fix 82 to 83 (after a number of adaptation steps the size of the error has become too small in relation to the MDB). The results of both adaptation strategies are somewhat disappointing. This comes as no surpise if one realizes that actually

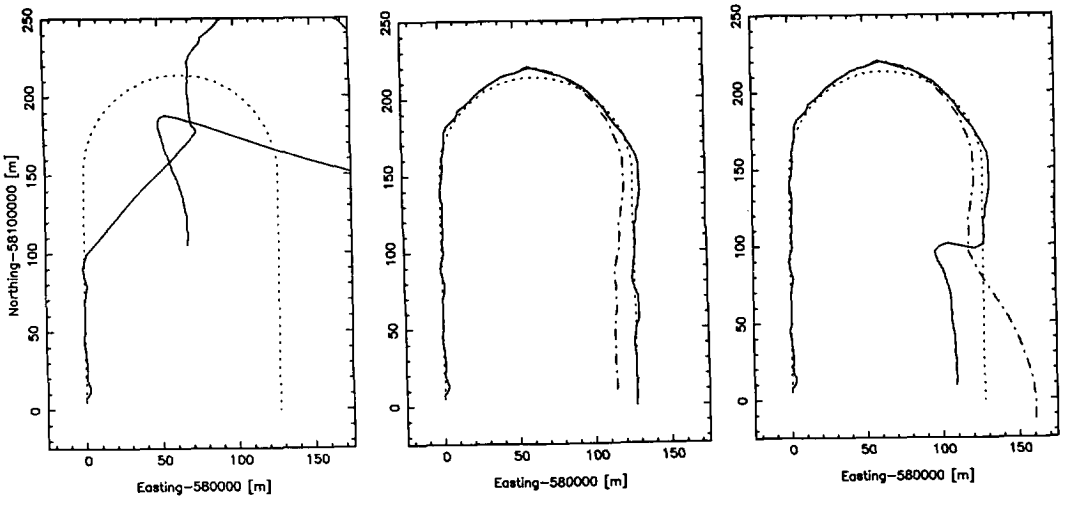


Figure 6.29: Lane slip situation — Case 2b; Estimated track without adaptation (left), for dataset I (middle), and dataset II (right).

50% of the data are contaminated by a large error! A scenario which is based on the elimination of faulty data is useless for model 2b as we only have two observations available per fix (after fix 81 we would have to depend completely on the dynamic model).

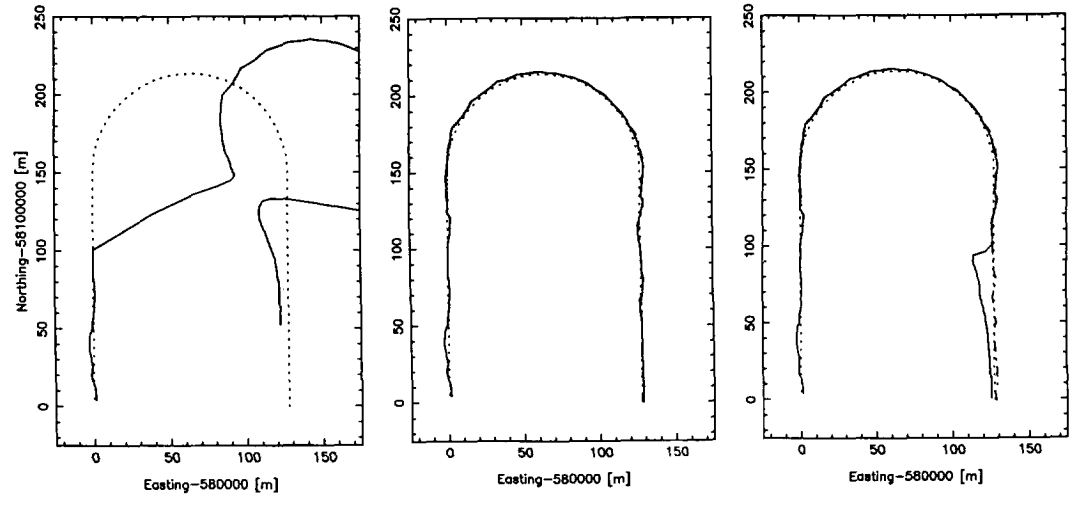


Figure 6.30: Lane slip situation — Case 2c; Estimated track without adaptation (left), for dataset I (middle), and dataset II (right).

The results pertaining to model 2c are given in Figs. 6.30 and 6.31. Compared to case 2b it can be seen that the integration of the additional ranges leads to a much improved result (certainly for dataset I). The semi-exact adaptation procedure applied

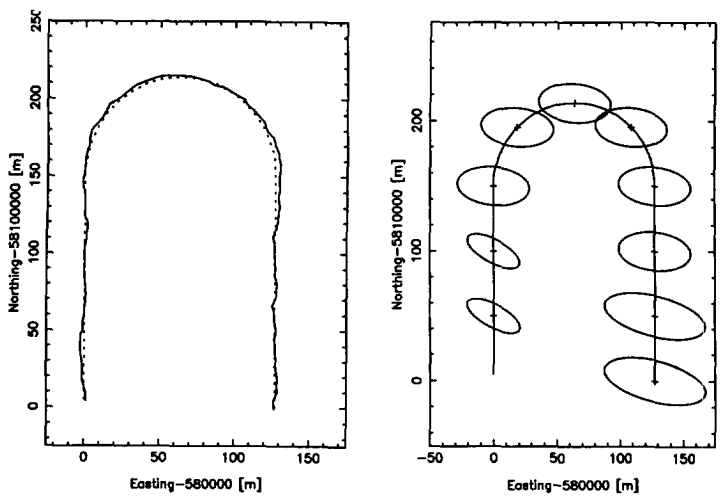


Figure 6.31: Lane slip situation — Case 2c; Estimated track after elimination of erroneous data (left), Error ellipses (scale 20:1) after elimination of erroneous data (right).

to dataset II has problems with recovering the exact size of the second slip once it has settled in. The approximate procedure recovers the varying error quite well, which is due to the fact that it operates on a fix by fix basis. The scenario in which the erroneous data are rejected (Fig. 6.31) results in very good track estimates, but the precision clearly suffers from the elimination of the data (after fix 81 the design results for model 2a (ranges only) apply).

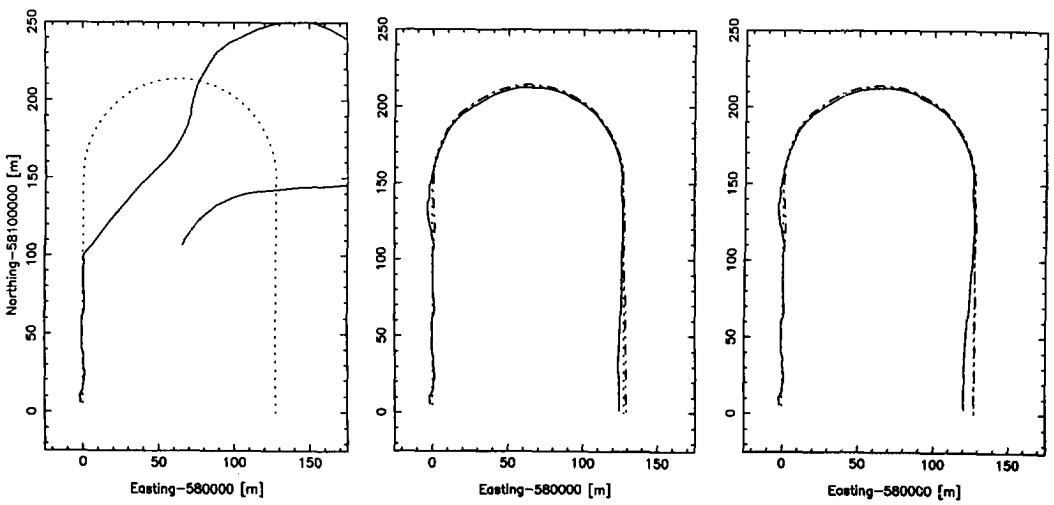


Figure 6.32: Lane slip situation — Case 2f; Estimated track without adaptation (left), for dataset I (middle), and dataset II (right).

The results for model 2f (given in Figs. 6.32 and 6.33) show that integrating model

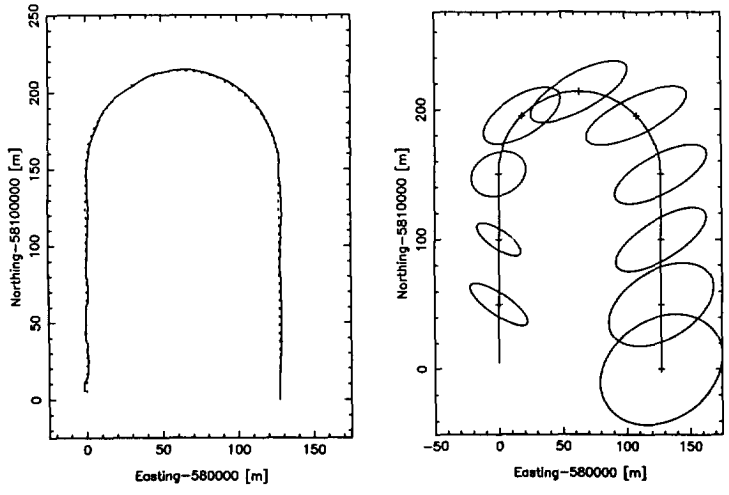


Figure 6.33: Lane slip situation — Case 2f; Estimated track after elimination of erroneous data (left), Error ellipses (scale 20:1) after elimination of erroneous data (right).

2b with dead reckoning observations also leads to an enhanced track estimate. The approximate adaptation procedure slightly outperforms the semi-exact adaptation procedure. After identification of a slip the filter under (semi-)exact adaptation procedure keeps operating under H_0, and because a lane slip is a large error, one actually estimates significant instrumental biases under H_0 (leading in turn to slightly biased solutions). These biases only slowly diminish because of the small system noise associated with the instrumental biases. Furthermore the approximate method manages to track the varying error size in dataset II better. With the scenario based on discarding of data (cf. Fig 6.33) one obtains excellent position estimates, but after two elimination steps one is left with a navigation system based on dead reckoning only (this is clearly illustrated by the growing size of the error ellipses).

Compared to model 2f the performance of the adaptation procedure for model 2g does not improve much as can be seen in Fig. 6.34. Actually the observations made with regard to model 2f remain valid for model 2g.

Summarizing, the analysis of the performance of the adaptation strategies for the lane slip problem shows that:

- Overall the performance of the approximate and (semi-)exact adaptation procedures is comparable. In case the error does not fully match the specified alternative hypothesis the approximate procedure seems to be the better alternative.

- The strategy based on the discarding of erroneous data yields good position estimates, but suffers from the fact that the redundancy of the system decreases and as a consequence the quality of the system rapidly degrades. Besides the elimination of data cannot be carried out ad infinitum; at a certain point there will be no data left.

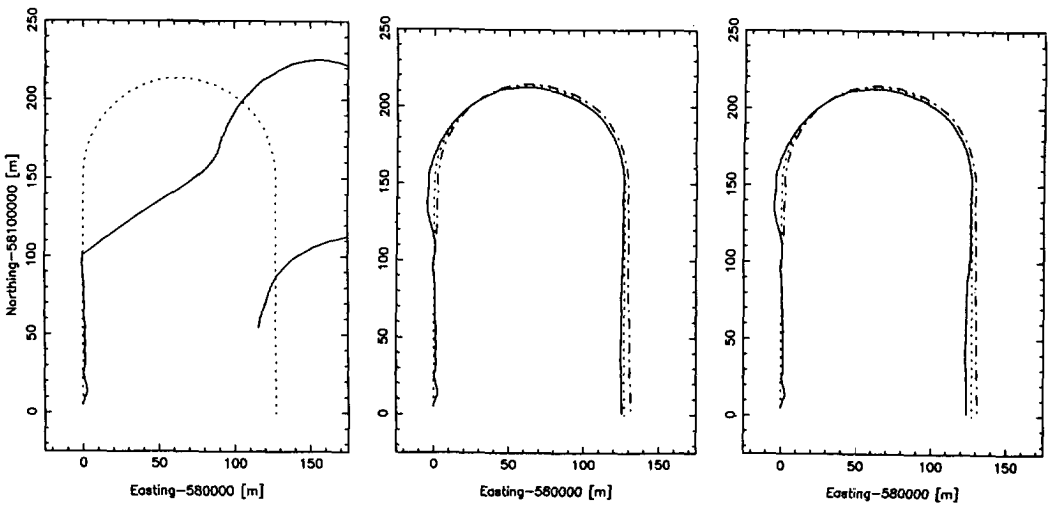


Figure 6.34: Lane slip situation — Case 2g; Estimated track without adaptation (left), for dataset I (middle), and dataset II (right).

- System integration improves the performance of the adaptation procedure. The solutions of models 2c, 2f, and 2g are clearly better than those of model 2b.

6.3.4 Adaptation — Concluding Remarks

In the previous subsections we have extensively investigated the performance of the adaptation procedure based on local tests only. Adaptation has been considered for a simple linear model, some of the navigation models, and the lane slip problem of hyperbolic positioning. We have compared the performance of the adaptation procedure for outliers (which is exact by definition) and slips. For slip-type model errors we have compared the performance of the (semi-)exact and an approximate adaptation procedure (both are described in Appendix C). We have found that:

- The adaptation procedure works very well for outliers.

- For slip-type errors the adaptation procedure correctly estimates the model error. The variance of the error estimator obtained by the exact procedure only improves with increasing window length for the navigation type models. The covariance between the error estimator and the (adapted) state estimator (which is negelected in the approximate method) remains significant for the linear models. For the navigation models the covariance diminishes with increasing delay.

- Adaptation for slips works reasonably well, both with the (semi-)exact and approximate adaptation procedure. The difference between the approximate and (semi-)exact adaptation procedures is small if judged by the state estimates. For the linear model the approximate method, however, underestimates the variance of the adapted state, which might cause filter divergence at a later stage. For the

navigation models the results of the exact procedure indicate one might revert to the filter operating under H_0 after a sufficient number of epochs.

- Although it seems tempting to eliminate faulty data rather than to adapt for the underlying model error, one has to keep in mind that by doing so the quality of the curtailed system rapidly deteriorates.

- System integration generally improves the performance of the adaptation procedure.

- One has to take care that the alternative hypotheses are properly specified. A slip may otherwise be treated as an endless sequence of outliers, possibly leading to filter divergence. If the characteristics of a slip do not fully correspond to a step function, the adaptation procedure for slips may have problems in estimating the model error correctly.

The findings above indicate that (any) adaptation procedure should be implemented with care. If one knows, however, which model errors are likely to occur and that the matching identification test statistics are powerful enough, the adaptation procedure will operate adequately.

6.4 Concluding Remarks and Recommendations

The analysis of the DIA procedure has been based on simulated datasets. This has allowed us to pinpoint specific strengths and pitfalls of the procedure. The detection and identification part have been analysed separately from the adaptation part of the procedure. The main findings are summarized in Sections 6.2.4 and 6.3.4.

The detection and identification procedure works well. The analysis of the DI procedure is hardly hampered by the fact that model errors have an impact on the state estimates and hence on the subsequent detection and identification of model errors. The concept of the MDB has been proven to be very useful, as MDBs really are the size of the biases that can be detected (and identified) with a certain probability.

The analysis of the adaptation procedure has been somewhat more involved and was limited to local adaptation. For practical reasons we have investigated several strategies to deal with slips. We found that the adaptation procedure is very suited to handle outliers, whereas slip-type model errors require more consideration.

Our findings are substantiated by the implementations of the DIA-procedure for kinematic GPS applications by TALBOT [1991] and LU AND LACHAPELLE [1992]. TALBOT has successfully applied the (local) detection and identification steps for detecting and isolating cycle slips in double difference phase data which were used in an experiment to obtain positions of a moving train. LU AND LACHAPELLE implemented the DIA-procedure, based on local tests, for a kinematic GPS positioning system, and obtained good results for cases where one model error was present at a time.

6.4.1 Recommendations

The currently implemented DIA procedure is not suited to handle several model errors simultaneously. It is therefore advised to develop an implementation that can handle multiple model errors at the same epoch. This requires that for each model error the bias in the test statistics and predicted residuals is successively removed.

Chapter 7

Conclusions

We have discussed extensively aspects of procedures for the data processing, testing, and design of dynamic systems. The testing and design procedures can be used for dynamic systems in general, but we have especially considered integrated navigation systems. Our main objective has been to provide a unified framework for the procedures of data processing, testing, and design for such systems.

The method of least squares provides us with the framework for the data processing procedure, namely the Kalman filter. Many results known from Kalman filter theory can be directly derived using the least squares approach. The batch type formulation of the filtering/smoothing problem using condition equations, instead of observation equations, greatly facilitates the derivation of results. The least squares approach moreover has deepened our understanding of the filtering algorithms for the case with alternative noise models (Section 2.6) and the case with non-linear measurement and dynamic models (Section 2.7).

In Chapter 3 we have derived a testing procedure for slip-type model misspecifications in the functional model. The theory of hypothesis testing in linear models provides us with the framework to develop our model validation and adaptation techniques. The testing procedure is based on generalized likelihood ratio tests, which in our case are uniformly most powerful invariant tests. The testing procedure consists of three steps, namely detection, identification, and adaptation (DIA), where the detection and identification steps are coupled via the B-method of testing. The suitability of the B-method for testing procedures in dynamic systems will require some further investigations. We make a distinction between local and global testing in that for the latter case the testing procedure covers several epochs at a time. The (local) detection and identification steps of the DIA procedure correspond to those of the testing procedures for geodetic networks. The DIA procedure is formulated recursively and consequently it can be efficiently implemented parallel to the filter. Much attention has been devoted to the adaptation procedure, because real-time estimation also requires real-time recovery from errors. The exact adaptation equations have been derived (which are based on a continuous updating of the state estimator and actually correspond to filtering under the alternative hypothesis), but we have also seen that in case of outliers one can, after

adaptation, switch back to the filter operating under the null hypothesis. In the case of adaptation for slips this is only possible for one particular case, namely the partially constant state space model. The derivation of the DIA procedure is based on the assumption that a single model error occurs. For model errors sufficiently separated in time this assumption is of little consequence for the performance of the procedure, but the implementation of the DIA procedure would benefit from strategies for handling model errors that occur very close together.

The testing procedure is an intrinsic part of the quality assurance cycle of a-priori, real-time, and a-posteriori model validation. To guarantee that the real-time DIA procedure works well, a system design should precede the implementation of any system. In Chapter 4 we have discussed the design with respect to quality criteria, namely precision and reliability. In the design phase one has to ascertain whether the model can meet precision and reliability requirements. Precision describes the quality of the system under the null hypothesis and is directly available from the Kalman filter. Reliability can be considered as the quality under the alternative hypotheses, and describes the sensitivity of the estimation results to undetected model errors. Optimization of the parameters of the testing procedure can be achieved by optimizing the reliability of the system. Likely model errors are specified in terms of alternative hypotheses. The specification of the alternative hypotheses is application dependent and is one of the most difficult tasks in system design. Based on the quality measures for precision and reliability (MDBs and BNRs, which are measures for the detectability of a model error and the significance of such a model error respectively) we have tentatively proposed a design procedure for dynamic systems. The proposal is somewhat hampered by the fact that we cannot provide explicit requirements for the reliability of integrated navigation systems, but we also have indicated how this drawback can be circumvented. A final proposal for a design procedure for geodetic navigation systems still needs to be developed, but this can be achieved with the tools presented in this report.

In Chapters 5 and 6 we have considered the application of the design and DIA procedures respectively. All results are based on a simulation study, but as this study is the first systematic study into the design and testing procedures based on our methodology, we think such a limitation can be justified. The findings of our design study closely correspond to what one would expect intuitively, but the use of measures for precision and reliability (MDBs, BNRs) additionally provides us with the opportunity to quantify the quality of a system and to compare various designs. A proper design should take precision, internal and external reliability into account simultaneously; optimization with respect to a single quality criterium does not automatically result in a system with a good overall quality. The precision of the state estimators depends mainly on the precision of the observables and can be improved by increasing the level of integration of the system. The design study has revealed that for reliability one has to make a distinction between model errors of outlier- and slip-type. The detectability of slips benefits from larger window lengths of the tests, whereas the detectability of outliers does not. Moreover the biases due to outliers are generally only significant instantaneously, whereas slips cause a constant bias in the state estimators. The anal-

ysis of the significance of the biases is performed by the BNRs, which allow a direct comparison between various alternative hypotheses. The BNRs are not always easy to evaluate, but the so-called response graphs considerably facilitate the evaluation of the BNRs. The separability between some alternative hypotheses is poor. Overall we have shown that the design measures are indeed useful tools to design a system with respect to quality. Based on a design study recommendations can be given on the choice of the window lengths of the tests. The design study also indicates that the design of a navigation system is not necessarily a trivial task. An issue that still needs to be resolved is the choice of the testing parameters, namely the level of significance and the power of the tests, but the magnitude of these parameters depends on the actual sensor suite of the system and cannot be derived from simulations.

The DIA procedure has been thoroughly verified in a simulation study where the detection and identification steps were analyzed separately from the adaptation step. The analysis of the detection and identification procedure has shown that the MDBs are indeed a very useful measure of detectability. Moreover it has been found that when a model error occurs many test statistics are rejected, but the largest one always indicates the correct alternative hypothesis. Errors approximately twice as large as the MDB are always detected and identified by the local tests. This has been one of the reasons we have considered adaptation based on local tests only. We deemed it wise to start with a simple adaptation strategy as no previous experience with the implementation of adaptation procedures was available. The adaptation procedure works very well for outliers and cases where a single slip is present in the dataset. For cases with multiple slips we have compared the exact procedure with a semi-exact and an approximate method (which reverts to the filter under the null hypothesis directly after the adaptation step but neglects the correlation between the estimators of the adapted state and the error). The correct specification of the alternative hypotheses is very important. Due to its local character the approximate method is the least sensitive to a misspecification of the alternative hypothesis. The analysis of the adaptation results for slips indicates that in some cases one might revert to the filter under the null hypothesis without making too large an approximation error, because the correlation one neglects when reverting to the filter under the null hypothesis decreases with increasing window length for many cases.

In summary, the framework of the least-squares adjustment and hypothesis testing in linear models allows surveyors to obtain a better understanding of the data processing and model validation techniques for dynamic systems. The examples of the testing and design procedures for dynamic systems have shown that these can, to a large extent, be considered as extensions of the procedures for the testing and design of geodetic networks.

References

ARNOLD, S.F. (1981). *The Theory of Linear Models and Multivariate Analysis.* John Wiley, New York.

BAARDA, W. (1967). Statistical concepts in geodesy, Netherlands Geodetic Commission, *Publ. on Geodesy, New Series*, Vol.2(4), Delft.

BAARDA, W. (1968). A Testing procedure for use in geodetic networks, Netherlands Geodetic Commission, *Publ. on Geodesy, New Series*, Vol.2(5), Delft.

BAARDA, W. (1973). S-Transformations and criterion matrices, Netherlands Geodetic Commission, *Publ. on Geodesy, New Series*, Vol.5(1), Delft.

BAARDA, W. (1977). Measures for the accuracy of geodetic networks, *Proceedings of the international symposium on optimization of design and computation of control networks*, Sopron, Hungary, pp.419-436.

BAKKER, G., J.C. DE MUNCK, G.L. STRANG VAN HEES (1989). *Radio Positioning at Sea.* Delft University Press, Delft.

BASSEVILLE, M. (1988). Detecting changes in signals and systems — a survey, *Automatica*, Vol.24(3), pp.309-326.

BIERMAN, G.J. (1977). *Factorization Methods for Discrete Sequential Estimation* (Vol.128, Series Mathematics in Science and Engineering). Academic Press, Orlando.

BORUTTA, H. (1988). Robuste Schätzverfahren für geodätische Anwendungen, Heft 33, *Schriftenreihe des Studiengangs Vermessungswesen*, Universität der Bundeswehr, München.

BROWN, R.G. AND P.Y. HWANG (1987). GPS failure detection by autonomous means within the cockpit, *Navigation (ION)*, Vol.33(4), pp.335-353.

BROWN, R.G., G.Y. CHIN, AND J.H. KRAEMER (1991). Update of GPS integrity requirements of the RTCA MOPS, *Proceedings ION-GPS-91*, pp.761-772.

BRYSON, A.E. AND L.J. HENRIKSON (1968). Estimation using sampled data containing sequentially correlated noise, *Journal Spacecraft and Rockets*, Vol.5(6), pp.662-665.

BUENO, R., E.Y. CHOW, K.P. DUNN, S.B. GERSHWIN, A.S. WILLSKY (1976). Status report on the generalized likelihood ratio failure detection technique, with application to the F-8 aircraft, *Proceedings 1976 IEEE Conference on Decision and Control*, pp.38-47.

CARLSON, N.A. (1988). Federated filter for fault-tolerant integrated navigation systems, *Proceedings of the 1988 IEEE PLANS*, Orlando, FL, Nov 29 - Dec 2, pp.110-119. IEEE Press, New York.

CARLSON, N.A. (1990). Federated square root filter for decentralized parallel processes, *IEEE Trans. on Aerospace and Electronic Systems*, Vol.26(3), pp.517-525.

CASELLA, G. AND R.L. BERGER (1990). *Statistical Inference.* Wadsworth & Brooks Cole, Pacific Grove, CA.

CHANG, C.B. AND K.P. DUNN (1979). On GLR detection and estimation of unexpected inputs in linear discrete systems, *IEEE Trans. on Automatic Control*, Vol.24(3), pp.491-501.

CHIN, L. (1979). Advances in adaptive filtering, In: Leondes, C.T. (ed.), *Control and Dynamic Systems*, Vol.15, Academic Press, New York.

CHIN, L. (1983). Advances in computational efficiencies of linear filtering, In: Leondes, C.T. (ed.), *Control and Dynamic Systems*, Vol.19, pp.125-193, Academic Press, New York.

DECARLO, R.A. (1989). *Linear Systems, A State Variable Approach with Numerical Implementation.* Prentice-Hall, Englewood Cliffs, NJ.

EKSTRAND, B. (1983). Analytical steady state solution for a Kalman tracking filter, *IEEE Trans. on Aerospace and Electronic Systems*, Vol.19(6), pp. 815-819.

FÖRSTNER, W. (1983). Reliability and discernability of extended Gauss-Markov models, In: Ackermann (ed.), Seminar on mathematical models of geodetic/photogrammetric point determination with regard to outliers and systematic errors, *DGK* A No.98, pp.79-103.

FRIEDLAND, B. (1969). Treatment of bias in recursive filtering, *IEEE Trans. on Automatic Control*, Vol.14(4), pp. 359-367.

FRIEDLAND, B. (1973). Optimum steady-state position and velocity estimation using noisy sampled position data, *IEEE Trans. on Aerospace and Electronic Systems*, Vol.9(6), pp. 906-911.

FRIEDLAND, B. (1983). Separated-bias estimation and some applications, In: Leondes, C.T. (ed.), *Control and Dynamic Systems*, Vol.20, pp.1-45, Academic Press, New York.

GELB, A. (ED.) (1974). *Applied Optimal Estimation.* MIT-Press, Cambridge, MA.

GOODWIN, G.C. AND R.L. PAYNE (1977). *Dynamic System Identification: Experiment Design and Data Analysis* (Vol.136, Series Mathematics in Science and Engineering). Academic Press, Orlando.

HAAGMANS, R., G.J. HUSTI, P. PLUGERS, J.H.M. SMIT, G.L. STRANG VAN HEES (1988). NAVGRAV Navigation and Gravimetric Experiment at the North Sea, Netherlands Geodetic Commission, *Publ. on Geodesy, New Series*, Vol.32, Delft.

HAMPEL, F.R., E.M. RONCHETTI, P.J. ROUSSEEUW, W.A. STAHEL (1986). *Robust Statistics — The Approach Based on Influence Functions.* Wiley, New York.

IGNAGNI, M.B. (1981). An alternate derivation and extension of Friedland's two-stage Kalman estimator, *IEEE Trans. on Automatic Control*, Vol.26(3), pp.746-750.

IGNAGNI, M.B. (1990). Separate-bias Kalman estimator with bias state noise, *IEEE Trans. on Automatic Control*, Vol.35(3), pp.338-341.

JAZWINSKI, A.H. (1970). *Stochastic Processes and Filtering Theory* (Vol.64, Series Mathematics in Science and Engineering). Academic Press, New York.

KAILATH, T. (1968). An innovations approach to least-squares estimation part I: Linear filtering in additive white noise, *IEEE Trans. on Automatic Control*, Vol.13(6), pp.646-655.

KALMAN, R.E. (1960). A new approach to linear filtering and prediction problems, *ASME Journal of Basic Engineering*, Vol. 82D, pp.35-45.

KAYTON, M. (ED.) (1990). *Navigation — Land, Sea, Air and Space.* IEEE Press, New York.

KERR, T.H. (1987). Decentralized filtering and redundancy management for multisensor navigation, *IEEE Trans. on Aerospace and Electronic Systems*, Vol.23(1), pp.83-119.

KOCH, K.R. (1988). *Parameter Estimation and Hypothesis Testing in Linear Models.* Springer Verlag, Berlin.

KOK, J.J. (1984). On data snooping and multiple outlier testing, *NOAA Technical Report* NOS NGS 30, NOAA, Rockville, Md.

KÖSTERS, A.J.M. (1992). Some aspects of a 3-dimensional reference system for surveying in the Netherlands — Quality analysis of GPS phase observations, Report 92.1, Section Mathemathical and Physical Geodesy, Department of Geodetic Engineering, Delft University of Technology.

KREBS, V. (1980). *Nichtlineare Filterung.* Oldenbourg Verlag, München.

LOOMIS, P.V.W., N. CARLSON, M.P. BERARDUCCI (1988). Common Kalman filter: fault-tolerant navigation for next generation aircraft, *Proceedings of the National Technical Meeting*, Santa Barbara, CA, Jan. 26-29, pp.38-45, Institute of Navigation, Washington, D.C.

LU, G., AND G. LACHAPELLE (1990). Reliability analysis applied to kinematic GPS position and velocity estimation, *Proceedings of IAG Symposium 107 on Kinematic Systems in Geodesy, Surveying, and Remote Sensing*, Banff, Canada, 10-13 September, pp.273-284. Springer Verlag, New York.

LU, G., AND G. LACHAPELLE (1992). Statistical quality control for kinematic GPS positioning, *Manuscripta Geodaetica*, Vol.17(5), pp.270-281.

MASRELIEZ, C.J. AND R.D. MARTIN (1977). Robust Bayesian estimation for the linear model and robustifying the Kalman filter, *IEEE Trans. on Automatic Control*, Vol.22(3), pp. 361-371.

MAYBECK, P.S. (1979). *Stochastic Models, Estimation, and Control*, Vol.1 (Vol.141-1, Series Mathematics in Science and Engineering). Academic Press, New York.

MAYBECK, P.S. (1982). *Stochastic Models, Estimation, and Control*, Vol.2 (Vol.141-2, Series Mathematics in Science and Engineering). Academic Press, New York.

MEDITCH, J.S. (1969). *Stochastic Optimal Linear Estimation and Control.* McGraw-Hill, New York.

MEHRA, R.K. AND J. PESCHON (1971). An innovations approach to fault detection and diagnosis in dynamic systems, *Automatica*, Vol.7, pp.637-640.

MELSA, J.L. AND D.L. COHN (1978). *Decision and Estimation Theory.* McGraw-Hill, New York.

MERTIKAS, S., D. WELLS, P. LEENHOUTS (1985). Treatment of navigational accuracies: proposals for the future, *Navigation (ION)*, Vol.32(1), pp.68-84.

NICOLAI, R. (1988) Statistics in positioning quality control, *Proceedings HYDRO88*, pp.97-106, Special Publication No.23, The Hydrographic Society, London.

PEÑA, D. AND I. GUTTMAN (1988). Bayesian approach to robustifying the Kalman filter, Chapter 9 in: Spall, J.C. (ed.), *Bayesian Analysis of Time Series and Dynamic Models*, Marcel Dekker, New York.

PEÑA, D. AND I. GUTTMAN (1989). Optimal collapsing of mixture distributions in robust recursive estimation, *Communications in Statistics*, Vol.18(3), pp.817-833.

RTCA (RADIO TECHNICAL COMMISSION FOR AERONAUTICS) (1991). *Minimal Operational Performance Standards for Airborne Supplemental Navigation Equipment using GPS*, RTCA DO-208, July, RTCA, Washington D.C.

SALZMANN, M.A. (1988). The impact of system integration on position estimation — two case studies, Technical Memorandum 23, Department of Surveying Engineering, University of New Brunswick.

SALZMANN, M.A. (1990). MDB: A Design tool for integrated navigation systems, *Proceedings of IAG Symposium 107 on Kinematic Systems in Geodesy, Surveying and Remote Sensing*, Banff, Sept.10-13, pp.218-227. Springer Verlag, New York.

SALZMANN, M.A. (1991). An assessment of external reliability in the design of integrated navigation systems, Paper presented at the XXth IUGG assembly, Vienna, Aug. 11-24.

SCHWARZ, K.P., M.E. CANNON, R.V.C. WONG (1989). A comparison of GPS kinematic models for the determination of position and velocity along a trajectory, *Manuscripta Geodaetica*, Vol.14(5), pp.345-353.

SORENSON, H.W. (1970a). Least-squares estimation: from Gauss to Kalman, *IEEE Spectrum*, Vol.7, pp.63-68.

SORENSON, H.W. (1970b). Comparison of Kalman, Bayesian and Maximum Likelihood Estimation Techniques, In: Leondes, C.T. (ed.), Theory and Applications of Kalman Filtering, *Agardograph* 139.

STURZA, M. (1988). Navigation system integrity monitoring using redundant measurements, *Proceedings of the National Technical Meeting*, Santa Barbara, CA, Jan. 26-29, pp.72-79, Institute of Navigation, Washington, D.C.

STURZA, M.A. AND A.K. BROWN (1990). Comparison of fixed and variable threshold RAIM algorithms, *Proceedings of ION GPS-90*, Colorado Springs, Sept.19-21, pp.437-443, Institute of Navigation, Washington, D.C.

SWERLING, P. (1971). Modern state estimation methods from the viewpoint of the method of least squares, *IEEE Trans. on Automatic Control*, Vol.16(6),pp.707-719.

TALBOT, N.C. (1991). High precision real-time GPS positioning concepts: modeling and results, *Navigation (ION)*, Vol.38(2), pp.147-161.

TEUNISSEN, P.J.G. (1985). The geometry of geodetic inverse linear mapping and non-linear adjustment, Netherlands Geodetic Commission, *Publ. on Geodesy, New Series*, Vol.8(1), Delft.

TEUNISSEN, P.J.G (1989a). Nonlinear inversion of geodetic and geophysical data: diagnosing nonlinearity, In: Brunner, F.K. and C. Rizos (eds.), *Developments in Four-Dimensional Geodesy*, LNES No.29, Springer Verlag Berlin.

TEUNISSEN, P.J.G (1989b). Estimation in nonlinear models, Proceedings 2nd Hotine-Marussi symposium, Pisa, June 4-7.

TEUNISSEN, P.J.G. (1990a). Quality control in integrated navigation systems, *Proceedings of the 1990 IEEE PLANS*, Las Vegas, Mar 20-23, pp.158-165. IEEE Press, New York.

TEUNISSEN, P.J.G. (1990b). An integrity and quality control procedure for use in multisensor integration, *Proceedings of ION GPS-90*, Colorado Springs, Sept. 19-21, pp.513-522, Institute of Navigation, Washington, D.C.

TEUNISSEN, P.J.G. (1990c). Some aspects of real-time model validation techniques for use in integrated systems, *Proceedings IAG Symposium 107 on Kinematic Systems in Geodesy, Surveying and Remote Sensing*, Banff, September 10-13, pp.191-200, Springer Verlag, New York.

TEUNISSEN, P.J.G. (1991). On the local convergence of the iterated extended Kalman filter, Contributions to geodetic theory and methodology, XXth general assembly of the IUGG/IAG section IV, Vienna, Aug.11-24, pp.177-184.

TEUNISSEN, P.J.G. (1992). Een voorbeeld van de DIA procedure toegepast op het partially constant state space model. Unpublished manuscript, Delft Geodetic Computing Centre.

TEUNISSEN, P.J.G., M.A. SALZMANN, H.M. DE HEUS (1987). Over het aansluiten van puntenvelden (3); Kwaliteitsaspekten van de aansluiting, *NGT Geodesia* No.87-9, pp.327-335.

TEUNISSEN, P.J.G. AND M.A. SALZMANN (1988). Performance analysis of Kalman filters, Report 88.2, Section Mathematical and Physical Geodesy, Department of Geodetic Engineering, Delft University of Technology, Delft.

TEUNISSEN, P.J.G. AND E.H. KNICKMEYER (1988). Non-linearity and least squares, *CISM Journal ACSGC*, Vol.42(4), pp.321-330.

TEUNISSEN, P.J.G. AND M.A. SALZMANN (1989). A recursive slippage test for use in state-space filtering, *Manuscripta Geodaetica*, Vol.14(6), pp.383-390.

TIBERIUS, C.C.J.M. (1991). Quality control and integration aspects of vehicle location systems, Thesis, Department of Geodetic Engineering, Delft University of Technology (also available as Publications of the Delft Geodetic Computing Centre No.1).

WILLSKY, A.S. (1976). A survey of design methods for failure detection in dynamic systems, *Automatica*, Vol. 12, pp.601-611.

WILLSKY, A.S. (1986). Detection of abrupt changes in dynamic systems, In: Basseville, M. and A. Beneviste (eds.), *Detection of Abrupt Changes in Signals and Dynamical Systems*, pp.27-49, LNCIS No.77, Springer Verlag, Berlin.

WILLSKY, A.S. AND H.L. JONES (1976). A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems, *IEEE Trans. on Automatic Control*, Vol.21(2), pp.108-112.

Appendix A

The Predicted Residual

A.1 The Predicted Residual

To render this report self-contained we will prove in the following that, if the filter operates at an optimum, the predicted residuals are uncorrelated from one epoch to the next. The proof is taken from [TEUNISSEN AND SALZMANN, 1989] (for an alternative proof see KAILATH [1968]). We will proof that

$$E\{\underline{v}_k \underline{v}_l^{\mathrm{T}}\} = 0, \text{ for } k \neq l \, . \tag{A.1}$$

We restrict ourselves to the case $l < k$. The case $l > k$ follows on the basis of symmetry. From

$$\underline{v}_k = \underline{y}_k - A_k \underline{\hat{x}}_{k|k-1} \tag{A.2}$$

it follows that

$$E\{\underline{v}_k \underline{v}_l^{\mathrm{T}}\} = Q_{y_k y_l} - Q_{y_k \hat{x}_{l|l-1}} A_l^{\mathrm{T}} - A_k Q_{\hat{x}_{k|k-1} y_l} + A_k Q_{\hat{x}_{k|k-1} \hat{x}_{l|l-1}} A_l^{\mathrm{T}} \, . \tag{A.3}$$

By definition the first two terms on the right hand side of (A.3) vanish for the case $l < k$. Thus

$$E\{\underline{v}_k \underline{v}_l^{\mathrm{T}}\} = -A_k Q_{\hat{x}_{k|k-1} y_l} + A_k Q_{\hat{x}_{k|k-1} \hat{x}_{l|l-1}} A_l^{\mathrm{T}} \, . \tag{A.4}$$

The two covariance matrices on the right hand side of (A.4) can be computed once the relation between $\hat{x}_{k|k-1}$ and the original input data of the Kalman filter is established. From a combination of the time and measurement update equations of the Kalman filter it follows

$$\underline{\hat{x}}_{i|i-1} = \Phi_{i,i-1}(I - K_{i-1} A_{i-1}) \underline{\hat{x}}_{i-1|i-2} + \underline{d}_i + \Phi_{i,i-1} K_{i-1} \underline{y}_{i-1} \, . \tag{A.5}$$

If we let i run from $l+1$ to k we get by combining the equations

$$\begin{aligned}\underline{\hat{x}}_{k|k-1} &= \prod_{i=l}^{k-1}[\Phi_{i+1,i}(I-K_iA_i)]\underline{\hat{x}}_{l|l-1} + \\ &\sum_{j=l}^{k-1}\left\{\prod_{i=j+1}^{k-1}[\Phi_{i+1,i}(I-K_iA_i)]\right\}\underline{d}_{j+1} + \\ &\sum_{j=l}^{k-1}\left\{\prod_{i=j+1}^{k-1}[\Phi_{i+1,i}(I-K_iA_i)]\right\}\Phi_{j+1,j}K_j\underline{y}_j \ .\end{aligned} \tag{A.6}$$

Note that for $j=k-1$ the product in the last two terms of (A.6) reduces to the identity matrix I. From (A.6) it follows that

$$Q_{\hat{x}_{k|k-1}y_l} = \{\prod_{i=l+1}^{k-1}[\Phi_{i+1,i}(I-K_iA_i)]\}\ \Phi_{l+1,l}K_lR_l \tag{A.7}$$

and

$$Q_{\hat{x}_{k|k-1}\hat{x}_{l|l-1}}A_l^{\mathrm{T}} = \{\prod_{i=l+1}^{k-1}[\Phi_{i+1,i}(I-K_iA_i)]\}\Phi_{l+1,l}(I-K_lA_l)P_{l|l-1}A_l^{\mathrm{T}} \ . \tag{A.8}$$

What remains to be shown is that the difference of eqs. A.7 and A.8 vanishes, or that

$$K_lR_l - (I-K_lA_l)P_{l|l-1}A_l^{\mathrm{T}} = 0 \ , \tag{A.9}$$

which is easily verified upon substitution of $K_l = P_{l|l-1}A_l^{\mathrm{T}}(R_l + A_lP_{l|l-1}A_l^{\mathrm{T}})^{-1}$. This concludes the proof of (A.1). This leaves us with the well-known result that for a batch of predicted residuals in the time interval $[l,\ldots,k]$, denoted as $\underline{v} = (\underline{v}_l^{\mathrm{T}}, \underline{v}_{l+1}^{\mathrm{T}}, \ldots, \underline{v}_k^{\mathrm{T}})^{\mathrm{T}}$, the covariance matrix Q_v is a block diagonal matrix $Q_v = \mathrm{diag}(Q_{v_l}, \ldots, Q_{v_k})$.

Appendix B

Observation Equations

B.1 Observation Equations

In the following we will describe the observation equations and their linearized forms which have been used in the computations of Chapters 5 and 6. The equations are taken from [SALZMANN, 1988].

Unknowns are indicated as x_i, where i indicates the number of the unknown. Observables considered are ranges ($\underline{r}_{is}$), hyperbolic lines of position (which can also be considered as distance-differences) ($\underline{r}_{irs}$), gyro readings ($\underline{\phi}$) and log measurements ($\underline{v}$). Approximate values are indicated with the superscript 0. For each observable we will give the non-linear observation equation, the computed observation and the linearized observation equation. (Known) Station coordinates for stations r and s are indicated as x_r, y_r (denoting Easting and Northing) and x_s, y_s respectively.

Unknowns

In the equations the following unknowns are used. For both the gyro and log one

unknown	type	unit
x_1	Easting	metres
x_2	Northing	metres
x_3	Velocity East	metres/second
x_4	Velocity North	metres/second
x_5	Gyro offset	radians
x_6	Log Bias	metres/second

Table B.1: Definition of unknowns.

additional instrumental state has been included in the model, which means that for each dead reckoning sensor all error sources are lumped into a single additive parameter. The gyro offset is the sum of the gyro instrument errors and the drift (the difference between the ship's heading and course made good due to wind, waves, and current).

The log bias (for a log operating in bottom track mode) contains instrumental errors only. We assume that the ship sails at a constant speed. Therefore the log bias can be included as an additive instrumental bias and the inclusion of a non-additive scale factor is not required.

Observation Equations

A range from the ship's position i to a station s is given in [metres].

- Observation equation

$$E\{\underline{r}_{is}\} = \sqrt{(x_s - x_1)^2 + (y_s - x_2)^2}$$

- Computed observation

$$r^0_{is} = \sqrt{(x_s - x^0_1)^2 + (y_s - x^0_2)^2}$$

- Linearized observation equation

$$E\{\underline{r}_{is}\} - r^0_{is} = \frac{x^0_1 - x_s}{r^0_{is}}\Delta x_1 + \frac{x^0_2 - y_s}{r^0_{is}}\Delta x_2$$

A hyperbolic line of position at the ship's position i w.r.t. master station r and secondary station s is generally expressed in [lanes]. The computations are based on the somewhat simplified observable:

$$E\{\underline{r}_{irs}\} = \frac{r_{rs} + r_{ir} - r_{is}}{\lambda}$$

where λ is the wavelength of the system in [metres] and the definition of the observable is chosen in such a way that only positive lane counts are possible. Multiplication of the observable in [lanes] by its wavelength yields an observable which is defined as a distance difference in [metres].

- Observation equation

$$E\{\underline{r}_{irs}\} = r_{rs} + \sqrt{(x_r - x_1)^2 + (y_r - x_2)^2} - \sqrt{(x_s - x_1)^2 + (y_s - x_2)^2}$$

- Computed observation

$$r^0_{irs} = r_{rs} + \sqrt{(x_r - x^0_1)^2 + (y_r - x^0_2)^2} - \sqrt{(x_s - x^0_1)^2 + (y_s - x^0_2)^2}$$

- Linearized observation equation

$$E\{\underline{r}_{irs}\} - r^0_{irs} = \left(\frac{x^0_1 - x_r}{r^0_{ir}} - \frac{x^0_1 - x_s}{r^0_{is}}\right)\Delta x_1 + \left(\frac{x^0_2 - y_r}{r^0_{ir}} - \frac{x^0_2 - y_s}{r^0_{is}}\right)\Delta x_2$$

A gyro reading is assumed to be in [radians].

- Observation equation

$$E\{\underline{\phi}\} = \arctan\frac{x_3}{x_4} - x_5$$

- Computed observation

$$\phi^0 = \arctan\frac{x_3^0}{x_4^0} - x_5^0$$

- Linearized observation equation

$$E\{\underline{\phi}\} - \phi^0 = \frac{x_4^0}{x_3^{0^2} + x_4^{0^2}}\Delta x_3 - \frac{x_3^0}{x_3^{0^2} + x_4^{0^2}}\Delta x_4 - \Delta x_5$$

A log measurement (with the log assumed to be in bottom track mode) is given in [metres/second].

- Observation equation

$$E\{\underline{v}\} = x_3 \sin\phi + x_4 \cos\phi - x_6$$

- Computed observation

$$v^0 = x_3^0 \sin\phi^0 + x_4^0 \cos\phi^0 - x_6^0$$

- Linearized observation equation

$$\begin{aligned} E\{\underline{v}\} - v^0 \;=\; & (\sin\phi^0 + a_2 x_3^0 \cos\phi^0 - a_2 x_4^0 \sin\phi^0)\Delta x_3 + \\ & (\cos\phi^0 + a_1 x_4^0 \sin\phi^0 - a_1 x_3^0 \cos\phi^0)\Delta x_4 + \\ & (x_4^0 \sin\phi^0 - x_3^0 \cos\phi^0)\Delta x_5 - \Delta x_6 \end{aligned}$$

where

$$a_1 = \frac{x_3^0}{x_3^{0^2} + x_4^{0^2}} \quad \text{and} \quad a_2 = \frac{x_4^0}{x_3^{0^2} + x_4^{0^2}} \,.$$

Appendix C

Software Implementation of the Adaptation Procedure

C.1 Adaptation Procedure — Software Implementation

In Chapter 3 we have derived the adaptation procedure. With eqs. 3.86 and 3.87 (which are repeated here for easy reference) we have available the exact adaptation procedure

$$\underline{\hat{x}}^{\mathrm{a}}_{k|k} = \underline{\hat{x}}^{0}_{k|k} - \overline{X}_{k,l}\underline{\hat{\nabla}}^{l,k}, \tag{C.1}$$

$$P^{\mathrm{a}}_{k|k} = P^{0}_{k|k} + \overline{X}_{k,l} Q_{\hat{\nabla}^{l,k}} \overline{X}^{\mathrm{T}}_{k,l} \tag{C.2}$$

for a model error ocurring at time l and adaptation at time k. We also have shown in Section 3.6.3 that for outliers performing a single adaptation step and reverting to the filter under H_0 is equivalent to the exact procedure given by (C.1) and (C.2). Except for the special case of the partially constant state space model this is not possible for slips. In the following we briefly describe the adaptation procedures for slips that have been implemented in the software.

For slip type model errors the software supports a so-called 'approximate' and the exact adaptation procedure. The main motivation for the implementation of the approximate method is its simplicity. It lacks, however, a sound theoretical foundation.

Adaptation is only performed whenever a model error is simultaneously detected *and* identified. Identification is limited to those alternative hypotheses which are explicitely specified by the user. At any epoch the software can only execute a single adaptation step, since no iterative identification scheme has been implemented in the software.

C.1.1 The Approximate Adaptation Procedure for Slips

After a slip-type model error has been detected and identified at time k, the estimator of the filtered state and its covariance matrix are adapted according to (C.1) and (C.2). The correlation between the estimators of the adapted state and the error is neglected.

Then the window length of all tests is temporarily reset to zero, in order to prevent any erroneous identifications due to smearing. Finally the estimated error is stored in a so-called 'slip-buffer' and observations from time k onwards are corrected for the slip. The impact of this correction on the covariance matrix of the observables is also taken into account. A flowchart of the approximate procedure is given in Fig. C.1.

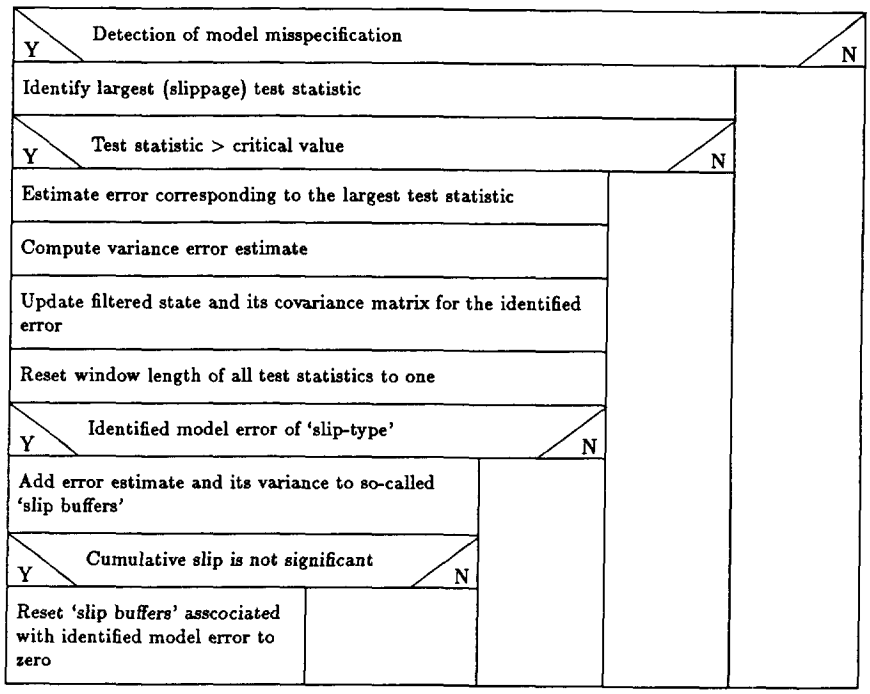


Figure C.1: The Approximate Adaptation Algorithm implemented in the software.

For an estimated slip in observation y_i the 'slip buffers' of the observations and the corresponding covariance matrix are updated as follows:

$$y_i^{\text{slip}} = y_i^{\text{slip}} + \hat{\nabla}^{l,k} \quad , \quad R_{i,i}^{\text{slip}} = R_{i,i}^{\text{slip}} + Q_{\hat{\nabla}^{l,k}} \; . \tag{C.3}$$

Similarly the 'slip buffer' of the disturbances and their covariance matrix for a slip in the dynamic model are updated as:

$$d^{\text{slip}} = d^{\text{slip}} + c_x \hat{\nabla}^{l,k} \quad , \quad Q^{\text{slip}} = Q^{\text{slip}} + c_x Q_{\hat{\nabla}^{l,k}} c_x^{\text{T}} \; , \tag{C.4}$$

where c_x models the impact of a slip of the dynamic model on the vector of predicted residuals. After adaptation for a 'slip'-type hypothesis the significance of the remaining slip in the updated 'slip-buffer' is checked. For an adapted slip in observation channel i it is checked if

$$\frac{|y_i^{\text{slip}}|}{\sqrt{Q_\nabla + Q_\nabla^{\text{old}}}} < \text{critical value slip test statistic.} \tag{C.5}$$

If the cumulative slip is found to be non-significant the buffers for observation channel i are reset as follows:

$$y_i^{\text{slip}} = 0.0 \quad , \quad R_{i,i}^{\text{slip}} = 0.0 \; . \tag{C.6}$$

The situation for an adapted slip in the state vector is a bit more complicated as the vector c_x often consists of more than one non-zero element. Therefore the procedure as sketched for adapted slips in observations is also applied, but the significance test is based on the element of d^{slip} related to the first element of the c_x vector.

If the remaining slip in the observations is found to be significant the measurement update of the filter is performed with the following adapted observation vector and covariance matrix (cf. eq. 3.104):

$$y^{\text{a}} = y - y^{\text{slip}} \quad , \quad R^{\text{a}} = R + R^{\text{slip}}.$$

For a significant slip in the dynamic model the time update is performed with the following adapted disturbance vector and associated covariance matrix:

$$d^{\text{a}} = -d^{\text{slip}} \quad , \quad Q^{\text{a}} = Q + Q^{\text{slip}}.$$

C.1.2 The Exact and Semi-Exact Adaptation Procedures for Slips

The exact adaptation procedure is implemented using (C.1) and (C.2) continuously after detection and identification at time k. During the phase of continuous adaptation the results of the testing procedure (i.e. detection and identification) are suppressed, because all test statistics are likely to be contaminated by the slip-type model error (i.e. suffer from the smearing effect). To counterbalance the deactivation of the detection and identification procedure a stop criterion for the exact procedure can be specified by the user. Once the decrease in the variance of the error estimator ($Q_{\hat{\nabla}l,i+1} - Q_{\hat{\nabla}l,i}$ (for $i > k$)) falls below a specified threshold, the software switches to the approximate adaptation procedure, updates the 'slip buffers', and proceeds as described in the previous subsection. After the switch to the approximate method this method is no longer exact and is therefore called semi-exact. Another obvious strategy to stop the exact adaptation procedure would be to terminate the continuous adaptation once the variance of the error estimate falls below a particular threshold and the covariance between the error estimator and the estimator of the adapted state is negligible. Because this strategy did not perform well for the simple linear models we investigated, it has not been implemented in the software, but it is a viable (and theoretically sound) alternative.