



KTH Electrical Engineering

Terrain Navigation for Underwater Vehicles

Ingemar Nygren

TRITA-S3-SB-0571

ISSN 1103-8039

ISRN KTH/SB/R--05/71--SE

Doctoral Thesis in Signal Processing
Stockholm, Sweden 2005

Signal Processing Group
Department of Signals, Sensors and Systems
School of Electrical Engineering
Royal Institute of Technology (KTH)
SE-100 44 Stockholm, Sweden
Tel. +46 8 790 6000, Fax. +46 8 790 7260
<http://www.s3.kth.se>

The work presented in this thesis is funded by

Swedish Defence Materiel Administration
SE-115 88 STOCKHOLM
SWEDEN

Akademisk avhandling som med tillstånd av Kungliga Tekniska Högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen onsdagen den 14 december 2005 kl 13.00 i sal F3, Kungliga Tekniska Högskolan, Lindstedtsvägen 26, Stockholm.

Copyright © Ingemar Nygren, 2005

Tryck: Universitetsservice US AB

Abstract

In this thesis a terrain positioning method for underwater vehicles called the correlation method is presented. Using the method the vehicle can determine its absolute position with the help of a sonar and a map of the bottom topography. The thesis is focused towards underwater positioning but most of the material is directly applicable to flying vehicles as well.

The positioning of surface vehicles has been revolutionized by the global positioning system (GPS). However, since the GPS signal does not penetrate into the sea water volume, underwater vehicles still have to use the inertial navigation system (INS) for navigation. Terrain positioning is therefore a serious alternative to GPS for underwater vehicles for zeroing out the INS error in military applications.

The thesis begins with a review of different estimation methods as Bayesian and extended Kalman filter methods that have been used for terrain navigation. Some other methods that may be used as the unscented Kalman filter or solving the Fokker-Planck equation using finite element methods are also discussed.

The correlation method is then described and the well known problem with multiple terrain positions is discussed. It is shown that the risk of false positions decreases exponentially with the number of measurement beams. A simple hypothesis test of false peaks is presented. It is also shown that the likelihood function for the position under weak assumptions converges to a Gaussian probability density function when the number of measuring beams tends to infinity.

The Cramér-Rao lower bound on the position error covariance is determined and it is shown that the proposed method achieves this bound asymptotically. The problem with measurement bias causing position bias is discussed and a simple method for removing

the measurement bias is presented.

By adjusting the footprint of the measuring sonar beams to the bottom topography a large increase in accuracy and robustness can be achieved in many bottom areas. This matter is discussed and a systematic theory about how to choose way-points is developed.

Three sea-trials have been conducted to verify the characteristics of the method and some results from the last one in October 2002 are presented. The sea-trials verify to a very high degree the theory presented. Finally the method is briefly discussed under the assumption that the bottom topography can be described by an autoregressive stochastic process.

Acknowledgments

First of all I want to thank Professor Björn Ottersten for letting me complete my research work on terrain navigation at the Signal Processing Lab at the School of Electrical Engineering. I am also grateful to my supervisor Docent Magnus Jansson at the same department for all the effort he has put into the work of proof reading the papers, articles and this thesis and for the discussions we have had during my time at the Signal Processing Lab.

This thesis had not come about without the financial support of the Swedish Defence Materiel Administration (FMV), which is gratefully acknowledged. I am especially grateful to Cdr. Stefan Ahlberg, MSc., Swedish navy, Mr. Carl-Johan Andersson, MSc., and Mr. Filip Traugott, MSc., both at FMV, not only for the financial support but also for their genuine interest in underwater navigation. Thanks goes also to Capt. Håkan Larsson, Ret., former head of the Torpedo Department of the FMV, for the support of the idea when it first came about and to Professor Ilkka Karasalo for his participation in the reference group for the project.

I am also grateful for valuable ideas from the submarines Cdr. Magnus Odéen and Lt. Cdr. Christer Brandt in the Swedish navy. The idea of using a 3D sonar for terrain positioning came originally from Cdr. M. Odéen.

My thanks goes also to all the people that have been involved in the sea-trials during the years and have made them a success.

Contents

1	Introduction	1
1.1	Introduction	1
1.2	Contributions	2
1.2.1	Published results	6
2	Terrain navigation	9
2.1	Terrain navigation methods in general	9
2.1.1	Introduction	9
2.1.2	The Bayesian approach	11
2.1.3	The extended Kalman filter from a Bayesian perspective	16
2.1.4	TERCOM	20
2.1.5	The mass point filter	21
2.1.6	The particle filter	21
2.1.7	The unscented transform	25
2.1.8	The stochastic differential equation	26
2.2	Terrain navigation methods for underwater vehicles	30
2.2.1	Correlation navigation with a bathymetric sonar	31
3	The positioning method	37
3.1	Introduction	37
3.2	The positioning method	40
3.3	The Bayesian and the maximum likelihood method (ML)	44
3.3.1	An illustrating example	44
3.3.2	The linear fusing method	47
3.3.3	Multiple likelihood peaks	49
3.4	Concluding remarks about the basic terrain navigation method	53
3.5	A way of implementing the approximate method	54
3.6	Comparison between the single beam missile method and the	

	correlation method	56
3.7	Determining the measurement bias	61
3.8	Incorporation of sediment data in the correlation	63
3.8.1	Introduction	63
3.8.2	Incorporation of sediment data	64
3.9	Incorporation of side-scan data in the correlation	68
3.10	Incorporation of external sound sources in the correlation	69
4	Characteristics of the likelihood function	73
4.1	Introduction	73
4.2	False peaks	74
4.2.1	The probability for false position	74
4.2.2	The amplitude of false peaks	77
4.2.3	False peaks, hypothesis testing	78
4.3	The shape of the likelihood function	80
4.3.1	The convergence of the likelihood curve	82
4.3.2	The convergence of the N -normalized likelihood curve	84
4.3.3	The convergence of the likelihood curve in the frequency space	84
4.4	The linear Kalman filter	87
4.4.1	The conditional mean	87
4.4.2	The linear Kalman filter	88
5	The Cramér-Rao lower bound	93
5.1	Introduction	93
5.2	The CRLB	95
5.2.1	The scalar CRLB	95
5.2.2	The vector CRLB	102
5.3	Additive Gaussian noise	104
5.3.1	The scalar case, Gaussian noise	104
5.3.2	The vector case, Gaussian noise	104
5.4	The bias expression	106
5.4.1	The linear case	106
5.4.2	The nonlinear case	107
5.5	Some examples of the CRLB in terrain navigation	110
5.5.1	The posterior FIM	110
5.5.2	Two measures of the posterior accuracy	112
5.5.3	Constant measurement covariance matrix, no measurement bias	113
5.5.4	Profile matching, constant measurement covariance matrix	113
5.5.5	Regular matching, measurement bias and constant covariance matrix	114

5.5.6	CRLB in the case of uncompensated sound speed gradients	116
6	The characteristics of the position estimate	123
6.1	Introduction	123
6.2	Characteristics of the positioning estimate	124
7	Optimal beam pattern and way-point selection	137
7.1	Introduction	137
7.2	Optimal footprint beam pattern	137
7.3	How to choose way-points	143
7.3.1	An example of how to choose a way-point	147
8	Sea-trial October 2002	153
8.1	Introduction	153
8.2	Sonar and position error causes	153
8.2.1	Introduction	153
8.2.2	Bathymetric sonars	154
8.2.3	Measurement errors	158
8.3	Sea-trial October 2002, simulation	160
8.3.1	Introduction	160
8.3.2	Simulation of the positioning	162
8.4	Sea-trial October 2002, results	166
8.4.1	Way-point 1	166
8.4.2	Way-point 2	168
8.4.3	Way-point 2B	170
8.4.4	Way-point 3	172
8.4.5	Way-point 7	174
8.5	Conclusions	176
9	The autoregressive bottom model	177
9.1	Introduction	177
9.2	AR bottom profiles	178
9.2.1	AR(1) bottom profile	178
9.2.2	AR(2) bottom profile	180
9.2.3	Determining the expected correlation function	180
9.3	The covariance matrix	181
9.4	Decomposition of the covariance matrix	184
9.4.1	Choleskey decomposition of the covariance matrix	185
9.5	The mean and the variance of the correlation sum	188
9.6	The probability density function of the correlation sum	191
9.7	The probability density function for the position	193
9.8	Approximative determination of the probability density function for the correlation sum	197
9.9	The Cramér-Rao lower bound	199

10	Summary	203
10.1	Summary	203
10.1.1	The problem	203
10.1.2	Solutions and conclusions	204
10.2	Further research	205
Appendix A		207
A.1	Inertial navigation	207
A.1.1	The principle of inertial navigation (INS)	207
A.1.2	The position error for INS	209
A.2	The hardware navigator	212
A.2.1	The basic principle	213
A.2.2	The principle of implementation	215
A.2.3	The practical implementation in an FPGA	217
Appendix B		221
B.1	Introduction	221
B.2	Reconstruction of signals from samples	222
B.3	Maximum a posterior interpolation	227
B.4	The least squares autoregressive interpolation	229
B.5	Interpolation in the frequency domain	235
Notational conventions		237
List of acronyms		243
Bibliography		245



Chapter 1

Introduction

1.1 Introduction

In this thesis a terrain positioning method for underwater vehicles called the correlation method is presented. Using the method the vehicle can determine its absolute position with the help of a sonar and a map of the bottom topography. The thesis is focused towards underwater positioning but most of the material is directly applicable to flying vehicles as well.

The positioning of surface vehicles has been revolutionized by the global positioning system (GPS). However, since the GPS signal does not penetrate into the sea water volume, underwater vehicles still have to use the inertial navigation system (INS) for navigation. A great problem with inertial navigation systems is the drift of the gyros, i.e., the position error will grow exponentially with time. This means that the underwater vehicle has to break the surface regularly to zero out the position error by reading its GPS position. This is undesired in military applications since it creates a risk of being revealed and the GPS signal may not be available or may even be jammed. Terrain positioning is therefore a serious

alternative to GPS for zeroing out the positioning error of the INS. Certainly the sonar emits a potentially revealing signal but the short duration of the sonar pulse makes the risk of revealing very low.

During the last five years, three large sea-trials have been conducted to support the presented theory. The test result has been very good and this thesis can also be seen as a presentation of the theoretical basis for the trials.

Compared to other terrain navigation methods, including the traditional methods for arial terrain navigation, the benefits of this method are

- Robustness
- High accuracy
- Minimal need of maps

In an attempt to make the thesis easy to read even for persons without immediate knowledge in estimation theory the method has been presented with numerous figures and often with an intuitive approach but hopefully the cogency of the thesis has not suffered from this.

1.2 Contributions

Optimal methods, in a minimum mean square error (MMSE) sense, for terrain navigation has been around for some time but as shown in this thesis the performance can be considerably increased if the measurement of the terrain topography is made in several points simultaneously. The thesis describes the proposed method, its performance and presents briefly the result from one of the sea-trials that have been conducted for verification of the theoretical findings.

This thesis is organized in the following chapters besides this introduction chapter.

Chapter 2 Terrain navigation methods

In this chapter the Bayesian approach to terrain navigation is outlined and the extended Kalman filter in a Bayesian perspective is presented. The extended Kalman filter (EKF) has been used in terrain navigation since long ago and the extended Kalman filter is here presented in a Bayesian setting.

The TERCOM terrain navigation method, which seems to be the first terrain

navigation method to be used, is described together with some other methods. The mass-point and particle filter, which have been used for flying vehicles with good results, are also briefly described.

A navigation filter based on the unscented transform (UT) is briefly sketched. This is a non-optimal filtering method that seems to fit the terrain navigation problem well even if no one yet has presented the method in a terrain navigation example. This goes also for the next method discussed. The amazing increase in computational power lately means that filters based on solving the Fokker-Planck equation are highly interesting and such filters may be a strong competitor to the mass-point filter. A brief review of such a filter is presented. Compared to the UT-filter this is an optimal method as the mass point filter is.

Finally this chapter ends with a review of some terrain navigation methods for underwater vehicles that have been presented at conferences and in journal papers during the last years.

Chapter 3 The positioning method

This chapter describes the background for the proposed terrain navigation method as well as the benefit of using it. It also touches briefly on the problem of false correlation peaks due to terrain repeatability.

The computational demand for the correlation method is substantial and an implementation method to handle this problem is presented.

The difference between measuring the terrain topography as in traditional terrain navigation for flying vehicles and as in the proposed method is discussed and it is shown that the proposed measuring method is superior with regard to accuracy.

A great problem in terrain navigation is position bias due to measurement bias and a simple method to eliminate it is presented.

The chapter ends with a discussion of how the position estimate can be improved by using bottom sediment information and/or data from side-scan sonars as well as from external sound sources.

Chapter 4 The likelihood function and the Kalman filter

An often quoted problem with the correlation methods is the problem of false

peaks due to terrain repeatability, i.e., positions which give high correlations but still do not correspond to the true position. In the chapter it is shown that the risk of false peaks decreases exponentially with the number of measurement beams. A large number of measuring beams will thus eliminate the problem. A simple hypothesis test of false peaks is also presented. With the test false peaks can to some degree be detected and a coarse “false peaks filter” can be constructed.

As described in Chapter 2 the Bayesian approach means that the prior PDF for the vehicle position is multiplied with the likelihood function from the measurement. This means that if the prior PDF is Gaussian and the likelihood function also is Gaussian the posterior PDF will be Gaussian. In the chapter it is shown that the likelihood function converges under weak assumptions to a Gaussian PDF when the number of measuring beams tends to infinity.

A consequence of this is that a linear Kalman filter can be used for fusing the position measurement with the prior position information. This simplifies greatly the calculation of the posterior position PDF since an analytic expression can be used instead of having to rely on numerical methods. This also means that the robustness with regards to filter divergence is greatly improved and of course the computational errors of the numerical methods are avoided.

Chapter 5 The Cramér-Rao lower bound

It is of course not satisfactory to simply have a position from a navigational system. Information about the accuracy of the position is also needed so we know to what degree the position figure can be trusted. However, in many cases it is not an easy task to give an exact figure of the actual accuracy for a position estimation method and it is in this context the Cramér-Rao lower bound (CRLB) shall be seen. The CRLB gives the lowest possible position error variance, given the prerequisites, for the position estimate regardless of the actual estimation formula. By comparing the accuracy we achieve in simulation with the CRLB we can judge whether the estimation method has potentially good accuracy compared to other methods. Thus the CRLB plays an important role in judging the accuracy of the positioning method proposed in this thesis.

The chapter starts with a presentation of the original proof that H. Cramér gave of the CRLB theorem in 1945 which is not so often quoted nowadays. Then some special formulas for the terrain navigation problem are presented. As examples of this are cases with additional measurement noise. Another example is the CRLB dependence on the bending of the measurement beams due to uncompensated

temperature gradients in the sea.

Chapter 6 The characteristics of the position estimate

This chapter discusses the characteristics of the position estimate and the position bias that may be the consequence of measurement bias, i.e., the case when the measured depth in a position differs from the depth according to the map in the same point due to a fix measurement error often caused by an incorrect height of the actual sea level.

Chapter 7 Optimal beampattern and way-point selection

By adjusting the footprint of the measuring sonar beams to the bottom topography a large increase in accuracy and robustness can be achieved in many bottom areas. This matter is discussed in the first part of the chapter. Then a systematic theory about how to choose way-points in a way that minimizes the probability of failure for the positioning is developed. The chapter ends with an example of how to choose a way-point.

Chapter 8 Sonars and the sea-trial October 2002

The chapter starts with a brief review of sonars and their measurement errors but the main content of the chapter is a review of a sea-trial. Three sea-trials have been conducted to verify the characteristics of the method. Some results from the latest test in October 2002 are presented in the chapter. The sea-trials verify to a very high degree the theory presented in the thesis.

Chapter 9 Terrain navigation for autoregressive bottom models

This chapter is freestanding from the previous chapters. It presents the correlation method under the assumption that the bottom topography can be described by an autoregressive stochastic process.

If the bottom topography could be described in statistical terms, for example as a stationary stochastic process, qualitative and quantitative judgements could be drawn about the positioning accuracy and adherent matters. This would be of great advantage, but up to now the available models covering larger bottom areas are very complex and the relevance of the models might be questioned. However, it turns out that very simple autoregressive models can be of value when describing local bottom characteristics. The chapter discusses and draws conclusions from such models.

Appendix A

The appendix discusses very briefly the inertial navigation system (INS). This part may be of interest for readers unfamiliar with inertial navigation systems. The chapter also briefly reviews a master thesis about implementing the calculation of the likelihood function in a field programmable array (FPGA). This is of great interest if the positioning method is to be used for flying vehicles since the performance of such an implementation makes the method also attractive in such applications. For flying vehicles a radar with a narrow scanning beam or a scanning laser beam may be used.

Appendix B

The positioning method requires a considerable amount of interpolation in the underwater map to which many well known interpolation methods can be used. However, the interpolation error will change the bottom spectrum by introducing high frequency components which may influence the position accuracy. An interpolation method that does not change the spectrum is thus of interest and the chapter discusses such probabilistic methods after a short introduction about reconstruction of signals from samples of the signal. The probabilistic methods have a remarkable performance if the bottom characteristics are stationary and known which, however, seldom is the case.

1.2.1 Published results

The major results in this thesis have been published in the following papers and journal articles.

I. Nygren, *Terrängnavigering för undervattensfarkoster. Resultat av fältförsök 1998* (Terrain Navigation of Underwater Vehicles. Results from Sea-Trial 1998), Report FOA-R—99-01298-313—SE, December 1999.

I. Nygren, *A Method for Terrain Positioning of an AUV*, Proceedings, 12th International Symposium UUST01, Durham, New Hampshire, USA, 2001.

I. Nygren, *A Method for Terrain Navigation of an AUV*, Conference proceedings, MTS/IEEE Oceans 2001.

I. Nygren, *Recursive Terrain Navigation, Application of the Correlator Method*, Report FOI-R—0764—SE, Sweden, August 2002.

C-J. Andersson, I. Nygren, *A Method for Terrain Positioning of Underwater Vehicles*, Proceedings, Undersea Defence Technology (UDT) Europe 2003.

I. Nygren, M. Jansson, *Recursive Terrain Navigation with the Correlator Method for High Position Accuracy*, Proceedings, 13th International Symposium UUST03, Durham, New Hampshire, USA, 2003.

I. Nygren, M. Jansson, *Robust Terrain Navigation with the Correlator Method for High Position Accuracy*, Conference proceedings, MTS/IEEE Oceans 2003.

I. Nygren, M. Jansson, *Terrain Navigation Using the Correlator Method*, Conference proceedings, IEEE Position Location And Navigation Symposium, Monterey, Cal., USA, April 27-29, 2004.

I. Nygren, M. Jansson, *Terrain Navigation for Underwater Vehicles Using the Correlation Method*, IEEE Journal of Oceanic Engineering, July 2004, Vol. 29, No. 3.

J. Carlström, I. Nygren, *Terrain Navigation of the Swedish AUV62F*, Proceedings, 14th International Symposium UUST05, Durham, New Hampshire, USA, 2005.

I. Nygren, M. Jansson, *A Terrain Navigation Method for UAVs and AUVs Based on Correlation*, IEEE Transactions on Aerospace and Electronic Systems, to be published.





Chapter 2

Terrain navigation methods

2.1 Terrain navigation methods in general

2.1.1 Introduction

Terrain navigation has up to now mostly been used for flying vehicles and has during the last decennium become an accepted method to improve and aid inertial navigation systems. The first studies and tests were done in the late fifties but most of the development work was done during the seventies. Lately the use has been substantially increased due to the increased availability of high accuracy digital terrain maps as a result of charting by satellites.

Examples of commercial terrain navigation methods within the flying community are TERCOM [GO80], TERPROM [AR88], TERNAV [SK85, PUJP00], SITAN [HOAN85], BITAN [CYYT92], LATAN [LA88] and others. TERCOM was the first method to be used. The principle for all methods is to measure the terrain profile along a flight passage as in Figure 2.1 and to compare it with a digital map and by that establish a position. The comparison can be done as a batch operation as in TERCOM or as a recursive operation as in TERPROM or SITAN.

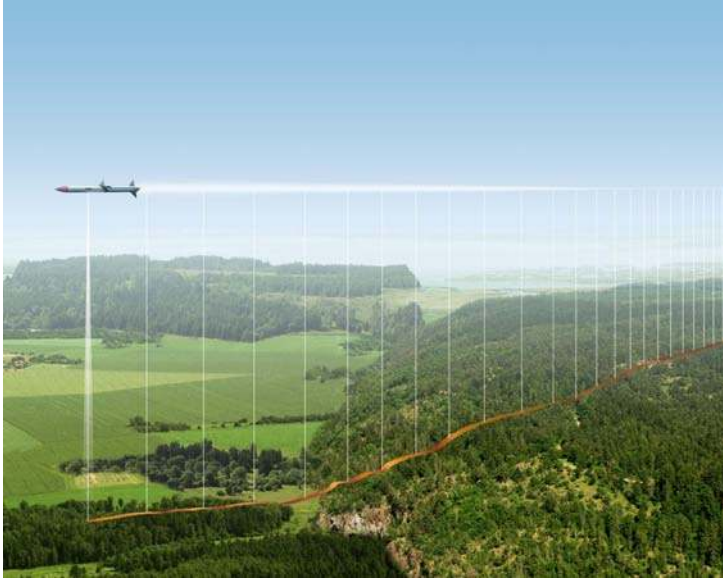


Figure 2.1: The principle of a terrain navigating missile.

The ground distance between the measurement points is in the range 30 – 150 meter and the sampling of the height is usually initiated by a clock. The radar altitude meter has a beam width of about 50° to make banking of the vehicle possible without losing ground contact.

The navigation method has two phases, one initial phase and one tracking phase. During the initial phase the starting position is determined with an accuracy of at least 100 meter. If the starting point is established by terrain reference it is usually done by collecting a batch of height measurements and matching them against the map by correlation. The search area in the matching process is determined by the accuracy of the inertial navigation system and can be in the range of 15 km x 15 km. When the starting position is established and the tracking phase can begin the determination of the position is done recursively at each new height measurement and often it is done by an extended Kalman filter (EKF). Lately also sequential Monte Carlo filters (SMC) as particle filters have been used [BE99, MM01, KA05].

For flying vehicles the vertical positions are of great interest besides the horizontal position. For underwater vehicles the vertical position can easily and accurately be determined by other means and the presentation will therefore be focused on methods to determine the horizontal position. We will assume that the movement of the vehicle in the horizontal plane is governed by the first of the following two equations. The height measurement is described by the second equation.

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \mathbf{u}_t + \mathbf{v}_t \quad t=0,1,2,\dots \quad (2.1)$$

$$y_t = h(\mathbf{x}_t) + e_t \quad (2.2)$$

The vector \mathbf{x}_t is the position of the vehicle in the horizontal plane and the vector \mathbf{u}_t is the displacement from the earlier position and the vector \mathbf{v}_t is the uncertainty in the displacement. The displacement can be thought of as coming from the INS-system or by integrating the equations of the vehicle movement. The measured height y_t is the height according to the map in the position \mathbf{x}_t with the addition of a measurement error e_t . The error sequences \mathbf{v}_t and e_t are considered to be white with a zero mean and independent of each other.

2.1.2. The Bayesian approach to terrain navigation

The Bayesian method for recursive estimation is particularly suitable for nonlinear problems which not easily can be linearized allowing an EKF-filter to be used [BUSE71]. For applications of the Bayesian approach in terrain navigation, see [RU85, BE99, METR02, KA05]. The Bayesian approach means that Bayes formula is used to incorporate the measurement data into the estimation and in most cases this also means that a numerical method has to be used to establish a position. This can be computationally demanding. With low order of the state vector the method is easily illustrated which can give extra insight into the estimation problem. The idea behind the method is the concept of a likelihood function [HAFO04]. The likelihood for the random variable \mathbf{Y} to have the value \mathbf{y} when the random variable \mathbf{X} has the value \mathbf{x} is in the discrete case

$$L(\mathbf{x}; \mathbf{y}) = \Pr\{\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x}\} \quad \text{for } \mathbf{x} \in \text{state space of } \mathbf{x} \quad (2.3)$$

where $\Pr\{\cdot\}$ denotes the probability.

In the continuous case the likelihood $L(\mathbf{x}; \mathbf{y})$ is defined correspondingly as the value of the probability density function (PDF) for \mathbf{y} at position \mathbf{x} . Since the value of \mathbf{y} is given by the measurement, the likelihood function is often considered to be a function of \mathbf{x} . For a random \mathbf{y} the likelihood function is a random function. The likelihood function can be seen as a measurement of how likely it is to have the measurement \mathbf{y} at position \mathbf{x} in state space. If several measurements from independent sensors are used the likelihood functions for the different measurements can be multiplied by each other. This makes it easy to incorporate measurements from different sources.

Figure 2.2 shows an underwater vehicle navigating according to the method developed and proposed in this thesis. The vehicle measures the bottom topography over a large bottom area with a large number of sonar beams at each sampling event. The assumption for the evolution of the state space is

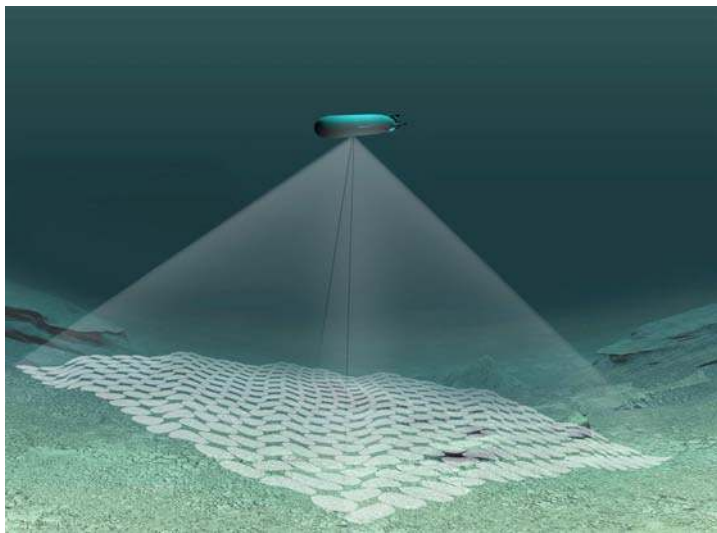


Figure 2.2: A vehicle measuring the bottom topography simultaneously with several sonar beams.

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \mathbf{u}_t + \mathbf{v}_t \quad t=0,1,2,\dots \quad (2.4)$$

$$\mathbb{Y}_t = \mathbb{H}(\mathbf{x}_t) + \mathbb{E}_t \quad (2.5)$$

where \mathbf{x}_t is the position in the horizontal plane at time t and \mathbf{u}_t - the distance between the positions - is provided by the INS system. The measured depths to the bottom are collected in the matrix \mathbb{Y}_t and the matrix $\mathbb{H}(\mathbf{x}_t)$ collects the depths according to the map if we are in position \mathbf{x}_t . The error in the INS system is \mathbf{v}_t and \mathbb{E}_t is the error in the depth measurement (possibly including map errors, interpolation errors etc.). The number of measurement points (sonar beams) is N . In order to simplify the discussion we vectorize (2.5)

$$\mathbf{y}_t = \mathbf{h}(\mathbf{x}_t) + \mathbf{e}_t \quad (2.5b)$$

and we will for simplicity assume that the measurement errors are independent white Gaussian sequences. If the errors are correlated with a known covariance matrix, this information can easily be incorporated (see Chapter 5).

It is well known that the optimal estimate in the linear and nonlinear cases is given by the conditional mean in the sense that it minimizes any scalar-value monotonically increasing cost function of the conditional mean square error [TS02]. The conditional probability density function (PDF) is given by [JA70]

$$p(\mathbf{x} | \mathbf{y}) = \frac{p(\mathbf{y} | \mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} = \frac{L(\mathbf{x}; \mathbf{y})p(\mathbf{x})}{\int_{R^2} L(\mathbf{x}; \mathbf{y})p(\mathbf{x})d\mathbf{x}} \quad (2.6)$$

where $\int_{R^2} L(\mathbf{x}; \mathbf{y})p(\mathbf{x})d\mathbf{x}$ can be interpreted as a normalizing constant. The PDF $p(\mathbf{y}|\mathbf{x})$ is called the likelihood function and the most common notation is $L(\mathbf{x};\mathbf{y})$ since the likelihood function is often thought on as a function of \mathbf{x} . We refer to $p(\mathbf{x} | \mathbf{y})$ as the posterior PDF, while $p(\mathbf{x})$ is referred to as the prior PDF.

The recursive procedure means that we will start from a given prior PDF that we will propagate according to the movement of the vehicle and the uncertainty

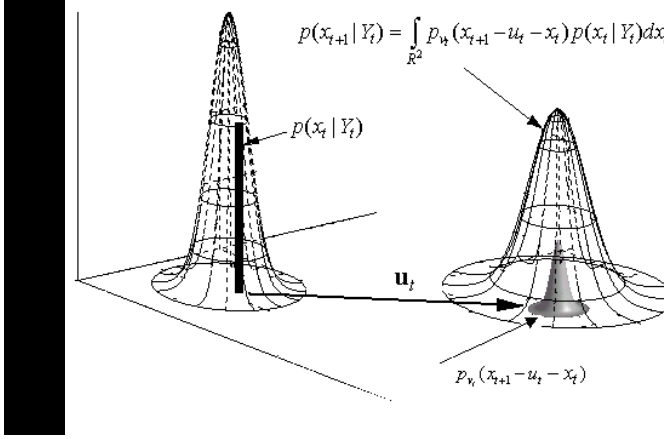


Figure 2.3: Propagation of the PDF for the vehicle position.

of the movement. The movement can, for example, be determined from an INS system that has a specified uncertainty. We will denote the propagated, but not measurement updated, PDF as $p^-(\mathbf{x})$. The next step in the recursive procedure is to do the measurement update of $p^-(\mathbf{x})$ by the multiplication of $p^-(\mathbf{x})$ and $L(\mathbf{y} | \mathbf{x})$ in order to arrive at the posterior PDF. In the following section we will look at this in some more details.

A. Propagation of the PDF for the vehicle position

Equation (2.4) describes how the position changes between two sampling events. Figure 2.3 illustrates this. The left PDF (the previous posterior PDF) has a bearing upon the vehicle's position at sampling time t and the black pile is the relative frequency of the number of realizations which have ended in point \mathbf{x}_t , i.e. $p(\mathbf{x}_t | \mathbf{Y}_t)$. The notation $p(\cdot | \mathbf{Y}_t)$ indicates that the PDF is based on all measurements up to and including the measurement at time t . The realizations are moved the distance \mathbf{u}_t but due to the uncertainty \mathbf{v}_t we will have a spread in the new position. We multiply this new PDF - the small one in the figure - by the relative frequency $p(\mathbf{x}_t | \mathbf{Y}_t)$ and total the contributions of

all points in the left PDF by an integral to get the position PDF at time $t+1$. We can express the propagation as

$$p(\mathbf{x}_{t+1}|\mathbf{Y}_t) = \int_{R^2} p_{\mathbf{v}_t}(\mathbf{x}_{t+1} - \mathbf{u}_t - \mathbf{x}_t) p(\mathbf{x}_t|\mathbf{Y}_t) d\mathbf{x}_t \quad (2.7)$$

where $p_{\mathbf{v}_t}(\cdot)$ is the PDF for the error term \mathbf{v}_t . This is the Chapman-Kolmogorovs equation for the propagation of a Markov process [JA70]. See also Section 2.1.8 about how to calculate the integral expression when \mathbf{u}_t is not just a simple fix translation.

B. The measurement update

The measurement is made according to (2.5b). We assume the errors in the depth measurements to be Gaussian. Therefore

$$L(\mathbf{x}_t; \mathbf{y}_t) = \frac{1}{\sqrt{(2\pi)^N \det(\mathbf{C}_e)}} \exp\left(-\frac{1}{2}(\mathbf{y}_t - \mathbf{h}(\mathbf{x}_t))^T \mathbf{C}_e^{-1}(\mathbf{y}_t - \mathbf{h}(\mathbf{x}_t))\right) \quad (2.8)$$

where \mathbf{C}_e is the measurement error covariance matrix. The function $L(\mathbf{x}_t; \mathbf{y}_t)$ is our likelihood function since, for a given position \mathbf{x}_t , it gives the likelihood of having the measurement value \mathbf{y}_t . The assumption that the errors of different beams are uncorrelated makes the covariance matrix \mathbf{C}_e diagonal and, if the measurement variance is the same in all beams, equal to $\sigma_e^2 \mathbf{I}$. Therefore the likelihood function will be

$$L(\mathbf{x}_t; \mathbf{y}_t) = \frac{1}{\sqrt{(2\pi\sigma_e^2)^N}} \exp\left(-\frac{1}{2\sigma_e^2} \sum_{k=1}^N (y_{t,k} - h_k(\mathbf{x}_t))^2\right) \quad (2.9)$$

where $y_{t,k}$ and $h_k(\mathbf{x}_t)$ are the k th components of the vectors \mathbf{y}_t and $\mathbf{h}(\mathbf{x}_t)$, respectively.

The next step in the recursion is to fuse the likelihood function for the measurement with the propagated PDF of the position. The non-normalized posterior PDF is obtained as

$$p(\mathbf{x}_{t+1} | \mathbf{Y}_{t+1}) \sim L(\mathbf{x}_{t+1}; \mathbf{y}_{t+1})p(\mathbf{x}_{t+1} | \mathbf{Y}_t) \quad (2.10)$$

If the time between the measurements is large the variance in the propagated PDF, $p^-(\mathbf{x})$, may increase so much that it will not noticeably improve the estimation of the position based on the likelihood function. Hence, when the prior PDF has a large variance the estimation problem becomes a maximum likelihood (ML) estimation problem. The measurement update by Bayes formula is further discussed in Section 3.3.

Bayesian estimation [HAFO04] is used in many scientific disciplines besides engineering. In [RO01] and [BS02] the subject is treated in a more mathematical way. More easy to read are [TP01], [HL01], [WL90] and [AS72] and a tutorial book is [SIL96]. A comparison between different estimation methods for nonlinear problems can be found in [FD05].

2.1.3. The extended Kalman filter from a Bayesian perspective

It is often advantageous to think about a filter as consisting of a propagation stage followed by a measurement update stage. In the extended Kalman filter (EKF) the nonlinear equations for movement and measurement are linearized in order to make a linear Kalman filter feasible [TS02, KSH00]. In studying the filtering problem from a Bayesian perspective we look at the PDFs of the involved stochastic variables whereby the filter procedure becomes quite natural.

The first equation (2.4) which describes the vehicle track with its stochastic fluctuations, often called the process equation, is already in a linear form. We will assume that the vehicle position at time $t = 0$ is Gaussian distributed and it will be that also at time $t = 1$ provided \mathbf{v}_t is Gaussian. The mean and covariance are calculated from the PDF given by (2.7). See also Chapter 4 for explicit expressions.

The measurement equation (2.5b) is nonlinear and is therefore linearized around the predicted position. The Taylor expansion of the terrain surface $h(\mathbf{x}_t)$ gives

$$h(\mathbf{x}_t) = h(\hat{\mathbf{x}}_{t|t-1}) + \mathbf{G}_{t|t-1}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) + (\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})^T \mathbf{H}_{t|t-1}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) + \dots \quad (2.11)$$

where $\mathbf{G}_{t|t-1}$ is the gradient vector and $\mathbf{H}_{t|t-1}$ is the Hessian matrix in the predicted

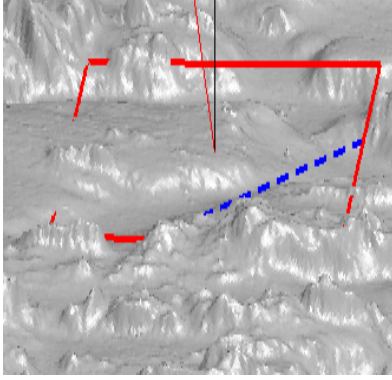


Figure 2.4: A tangent plane is placed in the predicted position.

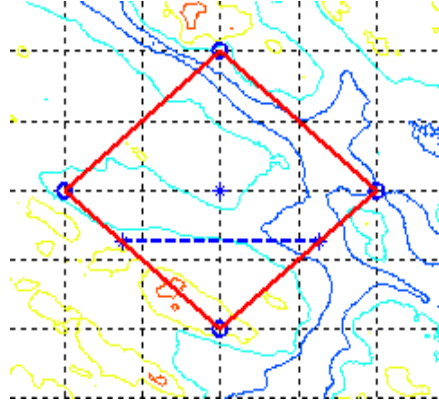


Figure 2.5: The tangent plane with the locus for all depths equal the measured depth (the dashed line).

position. If we truncate the series after the second term the equation can be illustrated by the tangent plane in the predicted position. The linearization means that a tangent plane to the surface is placed in the predicted position, see Figure 2.4. The likelihood function will be

$$\begin{aligned}
 L(\mathbf{x}_t; y_t) &= \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left(-\frac{1}{2\sigma_e^2} (y_t - h(\mathbf{x}_t))^2\right) = \\
 &\approx \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left(-\frac{1}{2\sigma_e^2} (y_t - h(\hat{\mathbf{x}}_{t|t-1}) - \mathbf{G}_{t|t-1}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}))^2\right)
 \end{aligned} \tag{2.12}$$

In the Figures 2.4 and 2.5 the locus for the positions with the same depths as the measurement is shown as a dashed line. The dashed line is also a symmetry line for the PDF for the measurement error e_t .

The PDF for the predicted position and the PDF for the measurement are illustrated in Figure 2.6 while Figure 2.7 shows the PDF after the measurement update. The update is made by point wise multiplications of the PDFs and will give a Gaussian PDF if both involved PDFs are Gaussian, i.e., the errors \mathbf{v}_t and e_t are Gaussian distributed as well as the starting PDF.

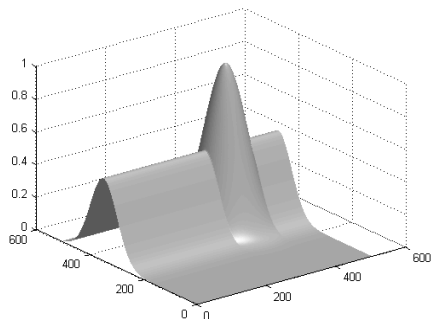


Figure 2.6: The prior PDF and the PDF for the measurement.

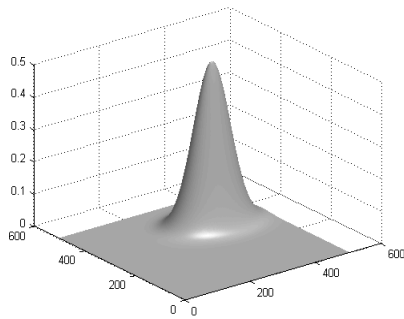


Figure 2.7: The posterior PDF, i.e. the prior PDF is updated by pointwise multiplication with the measurement PDF.

As can be seen the result will to a high degree depend on how well the predicted position is in agreement with the true position and how well the linearization plane describes the actual terrain around the predicted position. If the predicted position is close to the true position a better linearization can often be obtained by adjusting the plane to more than just one terrain point. The procedure is called stochastic linearization and means that the tangent plane is determined from several points around the predicted position by the least squares method. Considerable improvement of the accuracy of the position by using this procedure is reported [YCH91].

Another step to improve the result is to adapt the following iterative procedure which is easy to implement. After the posterior PDF has been calculated as indicated above a new linearization is made around the obtained position and a new posterior is calculated and so on. Only a few recursive steps are most often needed for obtaining a good position. The procedure is called the iterative EKF [TS02].

An improvement in the EKF-filter can be expected if also the second order term is included in the Taylor expansion of the terrain topography [TS02]. In this case we have a second order surface passing through the predicted point and it is likely that this surface will give a better fit than the tangent plane. The terrain equation will in this case be

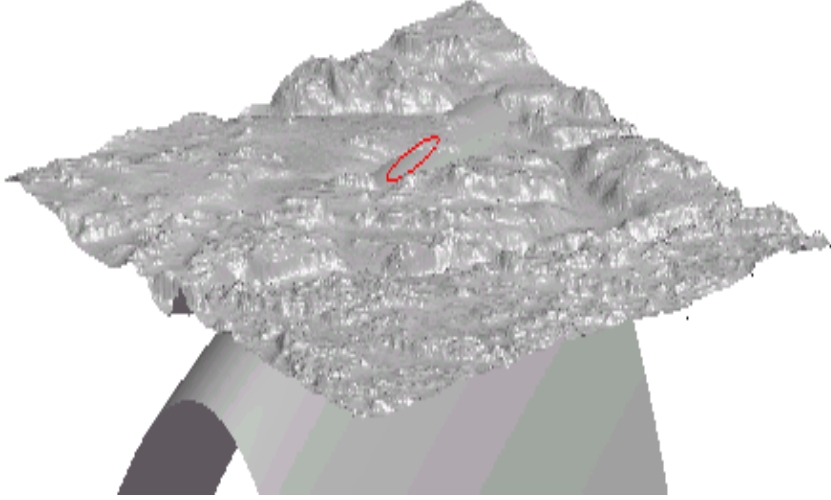


Figure 2.8: A second order surface fitted to terrain surface at the predicted point by a least squares approach.

$$h(\mathbf{x}_t) = h(\hat{\mathbf{x}}_{t|t-1}) + \mathbf{G}_{t|t-1}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) + (\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})^T \mathbf{H}_{t|t-1}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) + \dots \quad (2.13)$$

and the likelihood function

$$\begin{aligned} L(\mathbf{x}_t; y_t) &= \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left(-\frac{1}{2\sigma_e^2} (y_t - h(\hat{\mathbf{x}}_t))^2\right) = \\ &\approx \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left(-\frac{1}{2\sigma_e^2} (y_t - h(\hat{\mathbf{x}}_{t|t-1}) - \mathbf{G}_{t|t-1}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) \right. \\ &\quad \left. - (\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})^T \mathbf{H}_{t|t-1}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}))^2\right) \end{aligned} \quad (2.14)$$

The drawback of the second order filter is besides considerably more complex calculations that the posterior function will no more be Gaussian even if the measurement error and the prior are Gaussian. The iterated filter is often used but the use of the second order filter seems to be rare.

As can be understood the EKF filter is likely to diverge if the predicted position is bad or if the linearization does not fit the terrain well. A practical way to handle some of the divergence problems is to have several filters running in parallel. If it can be assumed that the positions of the filters are distributed

around the true position a better estimated position would be to take the mean of the individual filters as the predicted position and linearize around that position. As the new covariance the mean of the covariances of the individual filters can be taken. Should any of the filters show divergence it will be closed and a new filter will be started up based on the estimated position and its error covariance.

The terrain navigation method SITAN (Sandia Inertial Terrain Aided Navigation) as it is described in [HOAN85] is almost equivalent to the described EKF method above even if the filter in that paper is presented by equations. In [HOAN85] parallel filter structures and stochastic linearization is also discussed. A multiple model approach is discussed in [META83]. The tracking phase of the TERPROM is also almost equivalent to the described EKF method.

2.1.4 TERCOM

TERCOM (Terrain Contour Matching) was the first terrain navigation method to be used. Measured height data along a straight line flight path were collected in a vector which then was correlated with a digital map over the terrain. The relative distance between the measurement positions was determined by the INS-system or by integrating the equations for the vehicle movement. This means that the distance and orientation between the measurement footprint points will be afflicted with stochastic errors which will decrease the accuracy of the positioning.

When a new height measurement is available the oldest measurement is thrown away, cf. a FIFO-register. The early days requirement for a straight line vehicle course, due to the computation burden in calculation the orientation of the positions, is nowadays removed. The accuracy of the TERCOM method is in the range 30 – 100 meter for hilly terrain [YCH91].

In the original TERCOM method the positioning was done in special navigation areas, way-points, with hilly terrain for which accurate maps were available. The vehicle moved between the way-points by help of the INS-system.

A similar correlation method is HELI/SITAN [HO90, HO91] also developed at the Sandia Laboratories.

2.1.5 The mass point filter

The mass point filter is a numerical method for solving the Bayesian filter problem in an asymptotically optimal way. The state space is divided into a grid and the continuous PDF is replaced by corresponding probability masses in the grid points. A prerequisite is that the state space is of low order otherwise the computational burden will be prohibitive since quadrature in high dimensions has to be done.

The fix grid size can cause problems in areas with high PDF gradients so it is desirable with a small grid size to have a good agreement between the discrete and continuous PDFs. However, this can lead to increased computational burden since a coarser grid could be sufficient in large areas of state space. A way to handle this is to have a variable grid size determined by the gradient of the PDF but of course this will increase the complexity of the calculations. As will be seen later the involved PDFs are not particularly suited for the numerical calculations in the mass point filter (Figure 3.3 and Figure 3.7). The method, with some variations, is described in [BUSE71, BE99].

One of the advantages with the method is the optimality in the minimum mean square sense but the numerical solutions lead of course to numerical errors.

2.1.6 The particle filter

The particle filter [AMGC02, BE99, GGBFJKN02, MLG99, DFG01, KA05], is also a numerical method where a large number of samples of the state vector is generated according to its PDF. Each sample is called a particle and the procedure means that the density of particles in a position corresponds to the value of the PDF at that position. The particle filter is asymptotically optimal in a minimum mean square sense.

The Figure 2.9 shows a PDF and Figure 2.10 and 2.11 show the corresponding particle cloud. The particle density in a point in Figure 2.10 corresponds to the value of the PDF in that point. In the simplest form of the particle filters, the Bootstrap filter, the algorithm can be described as follows [BE98, FG00]. The propagation of the posterior PDF is done by propagation of each particle according to the propagation equation including the error term which should be simulated with correct error PDF. Se Figure 2.11 and Figure 2.12.

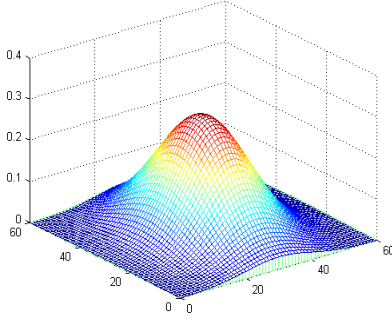


Figure 2.9: A PDF for generating the particle cloud in Figure 2.10.

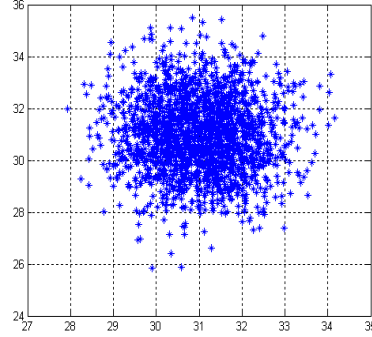


Figure 2.10: The particle cloud corresponding to the PDF in Figure 2.9.

The measurement update is then done by calculating the normalized likelihood value for each particle given the measurement y_t . If M is the number of particles and $p_{e_i}(\cdot)$ is the PDF for the measurement error then the normalized likelihood for the particle i is [BE98]

$$w_i = \frac{p_{e_i}(y_t - h(x_t^i))}{\sum_{j=1}^M p_{e_i}(y_t - h(x_t^j))} \quad (2.16)$$

The numerator can be seen as the likelihood for receiving the measurement y_t at the position x_t^i . The position estimate is

$$\hat{x}_t = \sum_{i=1}^M w_i x_t^i \quad (2.17)$$

i.e., the mean.

The normalized likelihood function will be the PDF from which a new cloud of particles is generated. The actual generation of new particles is a resampling with replacement [FG00].

The number of particles needed to describe the PDFs of a problem is often quite large 50000 – 100000 [GGBFJKN02] and by that the computations will

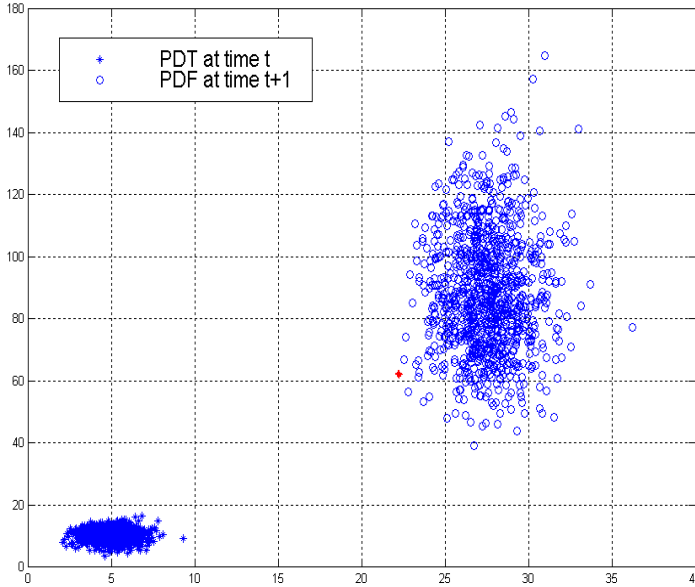


Figure 2.11: Down to the left is the initial particle cloud and up to the right are the propagated particles, see also Figure 2.12.

be demanding. A practical problem that is often encountered is that the particles tends to stick together in a single point after a number of iterations. This will happen if the error terms are small and a fix for that problem is to add some extra noise to the real noise.

A well working resampling algorithm is thus crucial for the method to avoid particle sticking. A practical way to circumvent the problem can be to project the result from the propagation phase on a known PDF family, i.e., mainly the exponential family [SY98, AS01, BR96] or another family that reasonable well can model the PDF described by the particle cloud. By this an analytic expression can be found for the posterior PDF from which a new particle cloud is generated. The procedure is of course an approximation of the propagated PDF and the performance may differ depending on the application and choice of PDF family to project on.

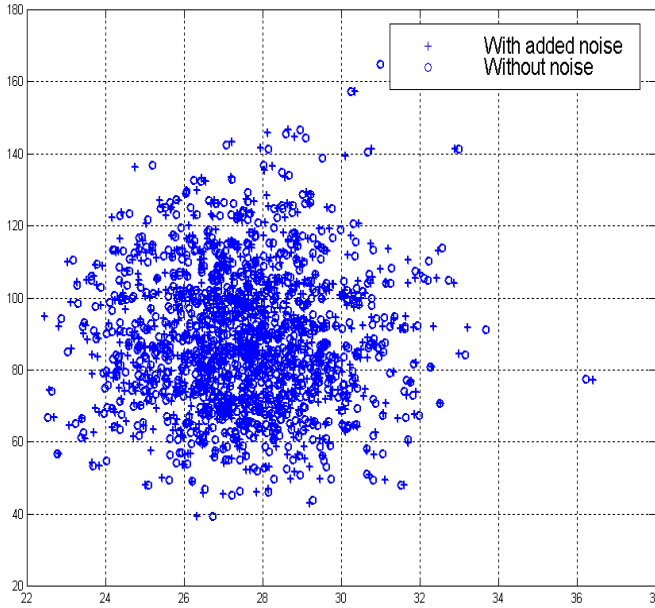


Figure 2.12: A close up of the propagated particles showing the particles with and without noise error.

The particle filter method is asymptotically optimal but it should be noted that methods that interpolate in the map introduces interpolation errors. In rough hilly terrain linear interpolation, if used, can severely decrease the performance of single beam methods.

It can often be that some parts of a system with advantage can be described by a linear Kalman filter and for other parts of a system a particle filter description is of advantage. A method to implement this is a procedure called Rao-Blackwellization [DFG01].

A comparison between correlation, mass-point and particle filters is given in [MWT02, METR02].

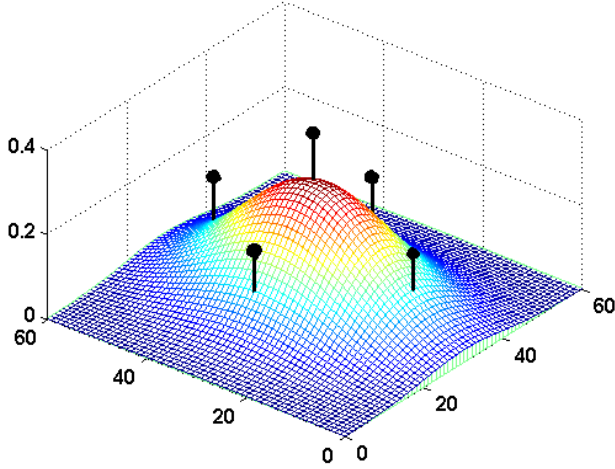


Figure 2.13: The figure illustrates how the PDF is replaced by “Sigma points” having the same mean and covariance as the PDF.

2.1.7 The unscented transform

The filter based on the unscented transform (UT) is somewhat improperly called the unscented Kalman filter (UKF). The basic idea is to approximate the PDF with a small number of probability mass points [JU97, JUUH04, MDFW00, HA01]. The performance of the filter should preferably be compared with the extended Kalman filter. The method means that the PDF of the position, in this case, is replaced by 5 discrete probability mass points (sigma points), one in the mean and the other symmetrically situated around the mean. Figure 2.13 illustrates the procedure.

The coordinates and the probability weights (masses) are in the simplest form of the transform determined in the following way. Let the PDF that should be approximated have the mean $\bar{\mathbf{x}}$ and the covariance \mathbf{P}_x . The coordinates and weights for the sigma points are then given by Table 2.1 where n_x is the dimension of the state vector and the number of sigma points is $2n_x+1$. κ is a scale factor and $\left(\sqrt{(n_x + \kappa)\mathbf{P}_x}\right)_i$ means the i th column in that matrix. The square root can for example be calculated by the Cholesky decomposition. The scalar scaling factor

Table 2.1: Coordinates and weights for the sigma points.

Coordinates	Weight (Probability mass)
Central point $\mathcal{X}_0 = \bar{\mathbf{x}}$	$W_i = \kappa / (n_x + \kappa)$
Odd points $\mathcal{X}_i = \bar{\mathbf{x}} + \left(\sqrt{(n_x + \kappa) \mathbf{P}_x} \right)_i$	$W_i = \kappa / \{2(n_x + \kappa)\}$
Even points $\mathcal{X}_i = \bar{\mathbf{x}} - \left(\sqrt{(n_x + \kappa) \mathbf{P}_x} \right)_i$ $i=1,2,\dots,2n_x$	$W_i = \kappa / \{2(n_x + \kappa)\}$

κ governs the spread of the sigma points. A suitable value is often $\kappa=2$ if the PDF is Gaussian like.

Each sigma point is propagated according to the propagation equation excluding the noise term and the mean and the covariance is calculated for the propagated sigma points and the covariance for the noise term is added as in the linear Kalman filter. The posterior PDF is then calculated by multiplying the propagated PDF by the likelihood function. The theory of UT filters is still under development and more elaborate filters are around but the filter outlined above would probably be sufficient for terrain navigation. However, no application of the filter in terrain navigation is yet known.

The unscented transform can be seen as approximating the real PDF with a Gaussian PDF with the same mean and covariance as the true PDF. In the cases that the real PDF is close to Gaussian it may be an acceptable approximation. The propagated PDF will also be Gaussian and the measurement update is just multiplying a Gaussian PDF with the likelihood function which is easily done. After normalization of the obtained posterior the mean and covariance are calculated and a new set of sigma points is determined for propagation. The mean is taken as the position estimate.

2.1.8 The stochastic differential equation

The motion of the vehicle and the measurement update can be modeled by stochastic differential equations and by solving the equations the evolution of

the position PDF in time and space can be determined and by that the position of the vehicle. The stochastic differential equation corresponding to (2.1) is [KL01, MI00, BR96]

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t)dt + \mathbf{G}(\mathbf{x}, t)d\boldsymbol{\beta}_t \quad (2.18)$$

where

$$E\{d\boldsymbol{\beta}_t d\boldsymbol{\beta}_t^T\} = \mathbf{Q}dt \quad (2.19)$$

and describes the evolution of the stochastic variable \mathbf{x} in the case of additive noise. In the equation $\mathbf{f}(\mathbf{x}, t)$ and $\mathbf{G}(\mathbf{x}, t)$ are deterministic vector valued functions and $d\boldsymbol{\beta}_t$ is an increment of the Brownian movement, i.e., an independent Gaussian distributed increment. The differential equation describes the propagation phase in terrain navigation quite well. The function $\mathbf{f}(\mathbf{x}, t)$ models the convection or transportation and the function $\mathbf{G}(\mathbf{x}, t)$ models the diffusion rate. An engineering approach to stochastic differential equation methods can be found in [MA82a, MA82b].

The weak solutions of (2.18) is the Fokker-Planck's partial differential equation (FPE) for the PDF of the stochastic variable \mathbf{x} [GA04, RI84]. In the two dimensional case we have

$$\frac{\partial p(\mathbf{x}, t)}{\partial t} = -\sum_{i=1}^2 \frac{\partial(p(\mathbf{x}, t)f_i(\mathbf{x}))}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^2 \frac{\partial^2 (p(\mathbf{x}, t)(\mathbf{G}(\mathbf{x})\mathbf{Q}\mathbf{G}(\mathbf{x})^T)_{i,j})}{\partial x_i \partial x_j} \quad (2.20)$$

where $p(\mathbf{x}, t)$ is the PDF for the variable \mathbf{x} at time t . The Fokker-Planck equation is of the parabolic type and is well known in physics and is easy to solve numerically [SM03, GOOR92, KICH91] if the dimension of the state vector is low as in our case. The earlier mentioned methods, masspoint filter, particle filter and the unscented Kalman filter can be thought of as numerical methods to solve the FPE. As an example of that which relates to the particle filter is the "molecular dynamics method" from the seventies. This also indicates that the mentioned methods suffer from certain numerical errors due to the chosen method.

In terrain navigation where the displacement of the position between two sampling instances is given by the inertial navigation system the function \mathbf{f} is a constant independent of the position. The function \mathbf{G} can also be considered constant, i.e., $\mathbf{G}(\mathbf{x}, t) = \mathbf{G}$. The Fokker-Planck equation can now be written as

$$\frac{\partial p(\mathbf{x}, t)}{\partial t} = -\sum_{i=1}^2 \mu_i \frac{\partial p(\mathbf{x}, t)}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^2 s_{i,j} \frac{\partial^2 p(\mathbf{x}, t)}{\partial x_i \partial x_j} \quad (2.21)$$

where we also have introduced the notation

$$s_{i,j} = (\mathbf{G}(\mathbf{x}) \mathbf{Q} \mathbf{G}(\mathbf{x})^T)_{i,j} \quad (2.22)$$

and

$$\boldsymbol{\mu} = \mathbf{f} \quad (2.23)$$

If we also adopt the notation common for partial differential equations in physics [CO04] we can write (2.21) as

$$\frac{\partial p}{\partial t} = -\nabla(\boldsymbol{\mu} p) + \nabla \bullet \left(\frac{1}{2} \mathbf{S} \nabla p \right) \quad (2.24)$$

where \mathbf{S} is defined by (2.22) and \bullet is the divergence operator and the del-operator is defined as

$$\nabla = \left[\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2} \right] \quad (2.25)$$

If we now introduce the transformation

$$\mathbf{y} = \mathbf{x} - \boldsymbol{\mu} t \quad (2.26)$$

we can write (2.24) as

$$\frac{\partial p}{\partial t} = \frac{1}{2} \nabla \bullet (\mathbf{S} \nabla p) \quad (2.27)$$

See also [ST97] which discusses transformations of different forms of the FPE to the simple basic form (2.27).

This type of partial differential equation is common in heat and diffusion problems and can be solved readily by most PDE-solvers if the dimension of the state

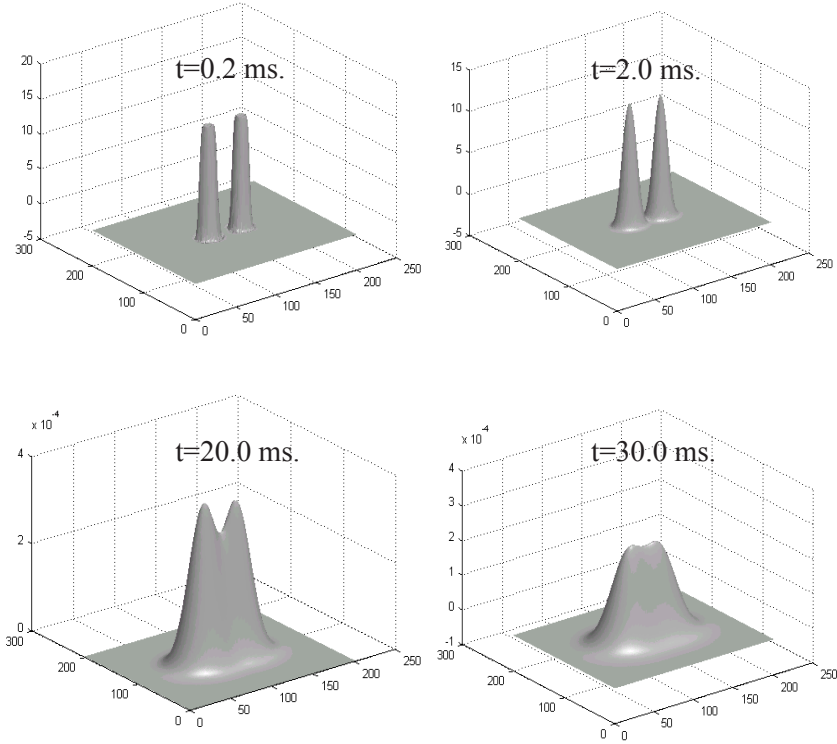


Figure 2.14: The figure shows how the solution to the Fokker-Planck differential equation varies with time, i.e., the propagation of a multimodal PDF with time. The time events is in scaled time and $S=I$.

variable is low, i.e., ≤ 3 . The dominating method to solve the parabolic type of equations seems to be the finite element method (FEM) [JO87, TH97]. The boundary conditions for a PDF can reasonable well be approximated by the FE method without increasing the computational burden which is of great advantage. Another advantage with the FE method is that it can handle situations where the PDF behaves very differently in different parts of the domain [DABJ74]. The computational power of today is sufficient to solve the FPE in real time using the FE method without problems for slowly moving vehicles as submarines and AUVs in the two dimensional state vector case. The Figure 2.14 shows the evolution of (2.27) for different times of a PDF for $S=I$. The PDF at time $t=0$ consists of two uniform cylindrical peaks. From the figure it

can be seen that the influence of the diffusion term leads to the convergence towards a Gaussian PDF if the noise is Gaussian. The influence of the diffusion term at different times depends of course on the magnitude of the diffusion function $G(\mathbf{x}, t)$. The graphs in Figure 2.14 are calculated by the PDE-solver in [CO04].

Equation (2.18) describes the evolution of the state vector in time and a similar equation can be formulated for the evolution of the continuous measurement vector in time. The equations fused together leads to the Kushner equation [MA82b, JA70, KU67] which gives the continuous measurement updated state vector in time. In the case of discrete measurements the measurement update may be done by using Bayes formula, i.e., (2.10).

2.2 Terrain navigation methods for underwater vehicles

Terrain navigation has been used for underwater vehicles to a much lesser extent than for flying vehicles. One of the main reasons for that is certainly the lack of reliable and accurate underwater maps of a quality which would allow the methods mentioned in Section 2.1 to be used.

The lack of maps of sufficient accuracy has lead to the development of other methods which rely on measurements of distances to known underwater sound sources. An example of such methods is hyperbolic acoustic navigation which relies on acoustic sources which synchronically transmit a coded sound signal at certain points of time and the receiver measures the time difference for the arrival of the signals to the vehicle. Other systems rely on measuring bearing and distance to one ore more sound sources. Often the sound sources are of transponder type, i.e., the sound transmission is initiated by a start pulse, “wake up signal,” from the receiver. The accuracy is relatively high, say 10 meter, if the sound propagation conditions are good. In areas characterized by multipath propagation, eg., archipelago areas, the methods will work less satisfactory.

Ground and flying vehicles use today to a high degree the global positioning system (GPS) to determine the position. However, the high frequency GPS signal in the 1.5 GHz radar band does not penetrate into the sea due to the high attenuation caused by the sea water on the weak satellite signal. Therefore systems have been developed with the GPS antenna or the whole receiver

contained in a buoy floating on the sea surface with a transponder system hanging below. The system can be deployed from air and the principle of operation is that the transponder transmits the GPS position of the buoy on request.

However, during the last decennium some terrain navigation systems for underwater vehicles have been developed and some of them have also been tested in the real world. Below is a review of some of them.

2.2.1 Correlation navigation with a bathymetric sonar

In [BER93] a correlation method is presented which uses a bathymetric sonar with 60 measuring beams, a multibeam sonar. The correlation against the underwater map gives rise to a large number of possible positions with about the same correlation strength. See also Chapter 3 which shows how the number of possible positions varies with the number of beam planes in the correlation. Therefore, the author discusses the possibility to use probabilistic data association (PDA) methods [BSFO88] to handle the uncertainty of the true position caused by the large number of correlation peaks. He discusses mainly “nearest neighbour,” “track splitting” and “Probabilistic Data Association Filter” (PDAF).

The data association methods have been developed to discriminate target echoes on a radar screen from so called clutter which is echoes from weather phenomena, from terrain, buildings and random measurement noise.

A prerequisite for the probabilistic data methods is that the clutter is random in both time and space and intensity. It is doubtful if this requirement is fulfilled in the terrain navigation case since the correlation result for repeated measurements over the same bottom location under the same conditions will give almost the same result in the number of correlation peaks and their locations. This is different to a radar which follows the target where the target background differs all the time and the number of clutter echoes and their locations and strengths also varies from radar scan to radar scan.

The author of [BER93] chooses, however, not to use the PDA-methods without discussion but adapts the concept of a “validation gate,” that is an area around the predicted position which a priori cover a certain percent of probability that the true position is within that area. A figure of 87 % is mentioned. The area is an ellipse in the case of a Gaussian prior and the actual correlation is done in the grid points in a gridded rectangle circumscribing the ellipse.

The author also defines, without more exact discussion of its shape, a matching strength function which, with the notation used in this thesis is,

$$f_p(\mathbf{x}) = \exp\left(-\frac{1}{2\sigma^2} \left[\sum_{j=0}^{59} \frac{|y_j - h_j(\mathbf{x})|}{k_j} \right]^2\right) \quad (2.28)$$

where y_j and $h_j(\mathbf{x})$ are the measured depth respectively the depth according to the map for beam j in position \mathbf{x} . The factor k_j takes the different measurement accuracies for different sonar beams into account. The standard deviation σ is assumed to be $\sigma=1.0$ which is a reasonable value. We see that this measurement strength function differs from the usual maximum likelihood function in that the sum of the absolute values is squared instead of the individual terms in the sum. This will graphically show sharper correlation peaks but it does however not improve the position accuracy. The position error covariance at time t is

$$\mathbf{S}(t) = E \left\{ (\mathbf{x}(t) - \hat{\mathbf{x}}(t))(\mathbf{x}(t) - \hat{\mathbf{x}}(t))^T \right\} \quad (2.29)$$

and the author defines the normalized matching strength function to be the PDF for the position error and the covariance of that PDF to be $\mathbf{S}(t)$. Thus the mean and covariance are calculated from the matching strength function but the position estimate is taken as the position for the maximum of the function. The covariance $\mathbf{S}(t)$ is used in the linear Kalman filter for determining the vehicles track. The covariance $\mathbf{S}(t)$ is also used to determine the validation gap.

However, it may be discussed if this is proper approximation of the position error covariance. The position error seems to be more a function of the local properties of the function around the maximum peak than the properties of the other smaller matching peaks in the neighbourhood of the maximum peak.

In the abstract to the thesis [BER93] the author mentions that the navigation method, as it is presented, is not particularly robust. However, if the vehicle speed is also measured and used in the Kalman filter for the vehicle position a substantial increase in robustness and accuracy is achieved. This indicates that the lack of robustness and accuracy is caused by an improper prior PDF, i.e., the validation gap is miss-placed.

Another thesis is [MA97] which is a simulation study of a single beam method

which to some extents is similar to the TERCOM method. See also [MAST97]. However, the positions for high correlation in the validation zone are fused together by the probabilistic PDFAI approach [DOBS90]. This algorithm is a further development of the PDAF algorithm to the extent that it also takes the strength of the correlation into account. In the correlation the mean absolute difference (MAD) algorithm is used and the number of measurements is 50 and the distance between the measurements is 5 meter which also is the grid size of the map. The simulation study shows very small position errors. In the MAD-algorithm the sum of the absolute beam errors between the measurement and the map is used instead of the sum of the squares of the errors. This measurement method was often used in early terrain correlation methods, for example in TERCOM, due to the computational load in calculating the correlation sum. It has also the benefit not to enhance large measurement errors (outliers) as the squaring of the errors does.

In the paper [SOA99] a method is described which has its roots in the analysis of images and uses a multibeam sonar to sample depth data. From the depth data a local map is first constructed and used for matching with the underwater map, the reference map. The matching is done with differential map attributes and by that it is very sensitive to noise in the map and therefore both maps are carefully filtered in order to reduce the noise. The filtering is done with an anisotropic diffusion operator, cf. diffusion in Chapter 2.1.8. The map attributes should be invariant to rotation and translation.

The next step in the process is to determine characteristic points in the reference map and the local map since those are the points that will be matched against each other. Characteristic points can, for example, be borders between different terrain shapes. To each characteristic point an attribute vector, whose elements are the attributes in that point, is attached. Examples of attributes are gradients, curvature, Laplacian and the absolute depth if this is available.

The matching is done between the characteristic points by calculating the Mahalanobis distance, based on the attribute vector, between the points. The covariance of the attribute vector is considered known. If the distance is below a certain threshold value the characteristic points are considered to refer to the same points.

The matching process gives a large number of possible positions for the vehicle and to choose among those an extended Kalman filter is used. The state

vector in the filter consists not only of the positions but also the rotations of the characteristic points in the reference map. The measurement vector consists of the corresponding rotated and translated points in the local map.

In the paper [SCA99] another correlation method using a bathymetry sonar is described but unlike the previous method no local map is constructed. The bottom profile along a line is matched directly against the reference map. For every possible location, (i, j) , in the map and for every beam k in the beam plane the absolute difference between the measured depth and the depth according to the map is calculated

$$\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K |h(i, j) - y_k| p_{ijk} \quad (2.30)$$

The horizontal uncertainty in the map and in the footprint of the measurement beams is given by the factor p_{ijk} and without motivation these errors are considered Gaussian. Besides this stochastic approach a Fuzzy approach is also discussed.

In the paper a simulation study with a real underwater map using a Kalman filter is described and figures of the radial position error is given. The stochastic approach gives radial errors of ~110 meter and the fuzzy approach ~80 meter.

The terrain navigation method described in [BMB00] consists of three parts; terrain matching, state estimator and a slant range corrector. The terrain matching is a maximum likelihood estimation using not only the horizontal positions but also the depth position. Nothing is mentioned how to calculate the position error covariance matrix. In the state estimator the position from the terrain matching is fused with the position from dead reckoning of the vehicle position. The slant range corrector is a measurement of the distance and the vertical angle to an external sound source (transponder) with a known position. The information in the along direction of the vehicle is only used. The result from the simulation study shows very small errors (< 1 meter).

In [MM01, JMHP04] no new methods are presented but instead interesting results from terrain navigation as a daily method for INS aiding of the underwater vehicle Hugin which has a multibeam bathymetric sonar. They have used the TERCOM method with the calculation of the position error covariance as described in [BER93] but also the particle filter and the mass point filter have

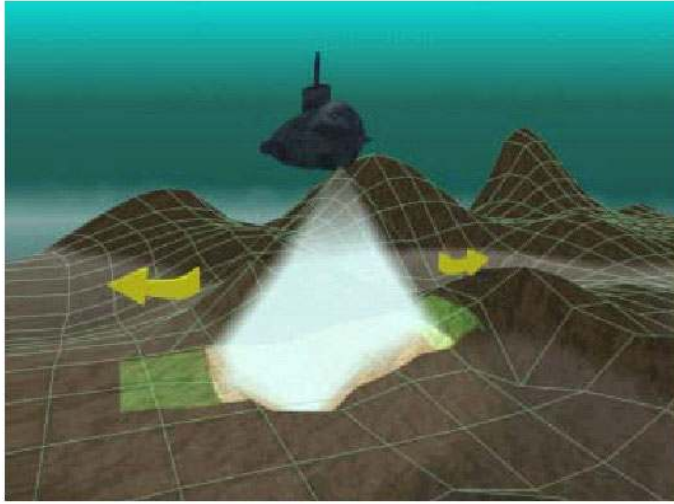


Figure 2.15: The figure shows how the Kongsberg Defence & Aerospace AS, Division for Naval Systems, in Norway envisions terrain navigation for submarines (Courtesy of Kongsberg Defence & Aerospace AS).

been tested.

Figure 2.15 from [FJ03] illustrates how the Kongsberg Company in Norway envisions terrain navigation for submarines. The method they propose is basically a result from the research work published in this thesis.





Chapter 3

The positioning method

3.1 Introduction

The global positioning system (GPS) has implied a revolution within land, air and sea navigation during the last decades. Unfortunately this does not include navigation of underwater vehicles as submarines or autonomous underwater vehicles since the GPS signal does not penetrate into the sea. Underwater vehicles have therefore still to rely on dead reckoning, gyro compass or inertial navigation, see Appendix A.

Inertial navigation has the great advantage to be independent of external signals and is by that very attractive in military applications where the threat of jamming is a great problem. However, inertial navigation has a great problem in the gyro drift and biases in the gyros and accelerometers. Inertial navigation for submarines where low drift system is an absolute requirement is very costly and the system has to be re-initiated regularly with an accurate position. This can of course be done by a differential GPS position but there is always a risk of being detected or that the GPS signal is jammed. Terrain navigation is a possibility for an underwater vehicle to establish a position fix without breaking the sea surface.

Terrain navigation and inertial navigation techniques can be said to be a successful couple since they are complementary with regard to the position errors and they both share the characteristic of being resistant to jamming. Hence they together make a system which is negligible revealing.

Terrain navigation has up to today not been used to any larger extent for underwater vehicles but that is certainly due to lack of high quality underwater maps. However, the GPS system together with the multibeam bathymetric sonar has also brought a revolution into mapping of the underwater world. The cost today to chart an underwater area is only a fraction of what it was some decades ago and the underwater maps are of a quality that was not reached before. However, the extent of the mapped areas are still very limited and will so be for a long time ahead. Remember that more than 70 % of the earth surface is sea area.

Some terrain navigation attempts [NY99, BER93, JMHP04, SCA99] have however been done with multibeam sonars and methods similar to the TERCOM method which was early implemented in the Tomahawk cruise missiles [MUOT82]. One of the problems that has to be handled if the positioning is to be successful is the problem with multiple correlation peaks.

Figure 3.1 shows a typical matching result when a beam plane of 60 sonar beams are matched against an underwater map of size 5 km x 5 km. The peaks in the figure correspond to positions where the measured profile well agrees with

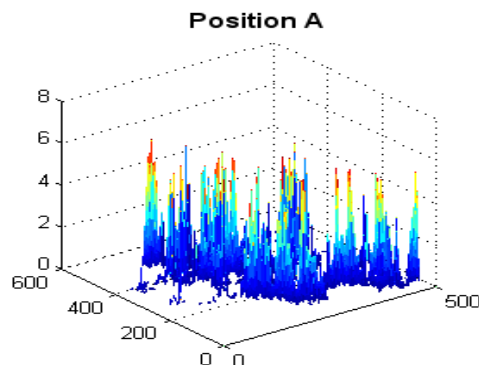


Figure 3.1: Correlation peaks when matching with one beam plane.

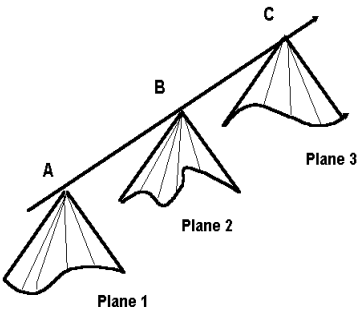


Figure 3.2: The principle for assembling several beam planes from a bathymetric sonar to form a 3D measurement of the bottom topography.

the map. As can be seen there are many likely positions for the vehicle. However, if the matching is done simultaneously with several beam planes as Figure 3.2 indicates the result will be very different as can be seen in Figures 3.3 and 3.4. The distance between the beamplanes is 10 - 20 meters. As will be shown later the number of likely positions will decrease exponentially with the number of measurement beams.

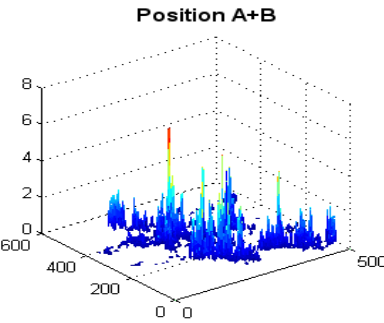


Figure 3.3: Correlation peaks when using two beamplanes.

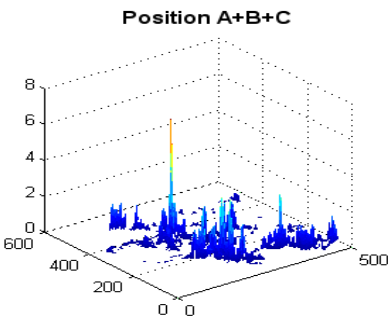


Figure 3.4: Correlation peaks when using three beamplanes.

The measurement of the bottom topography, according to the proposed method, should be done by a true 3D sonar. In some cases the 3D picture of the bottom may be constructed by assembling beam planes from a bathymetric sonar. In this case, however, there will be an error in the distance between the beam planes and the angle between the planes and the performance will therefore be degraded. These errors will however be very small if the approximate method is applied to a submarine. A submarine has a very large mass (~1600 ton or more) so the changes in speed or course during the sampling time of the bottom profile will be very small if the submarine is held on a steady course. This has also been the experience of the sea-trials that have been conducted with surface ships.

The problem with lack of high quality underwater maps can to some extent be solved by the same method that was used, and still is in use, in the early days of cruise missiles. The INS is only updated in special areas called navigation cells which are chartered and have good properties for terrain navigation. The navigation technique is therefore especially valuable for planned vehicle operations of considerable duration such as reconnaissance operations.

3.2 The positioning method

Navigation according to the proposed method is therefore that the vehicle moves between so-called “navigation cells,” as shown in Figure 3.5, with the help of the conventional navigation system such as INS, or dead reckon/Doppler-log. When the vehicle is, according to its navigation system, within the navigation cell it determines its position by measuring the bottom 3D topography with a large number of sonar beams over a large bottom area as in Figure 3.6. The measured area typically has the size of 40 m x 40 m to 300 m x 300 m depending on the actual depth. The bottom topography is then matched with the underwater map by comparing the measurement with the map at every point in the map.

The proposed navigation method has some important characteristics [NY01a, NY01b, NY02, NY99, NYJA04a, NYJA04b, NYJA04c, JCN05]:

- Robustness with respect to the navigation filter divergence

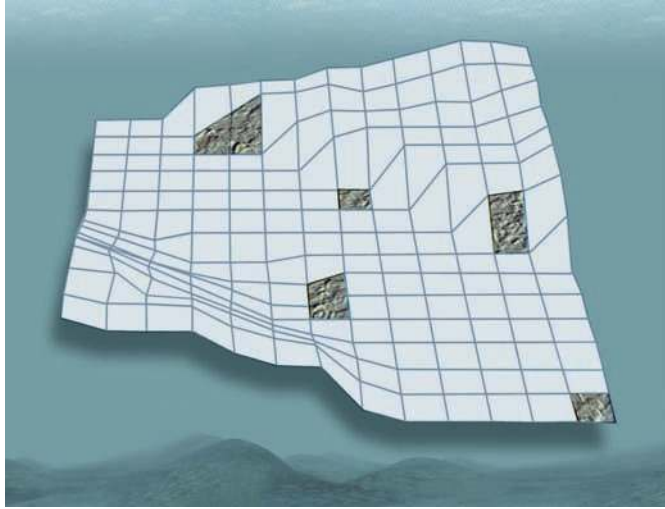


Figure 3.5: A map with indicated navigation cells.

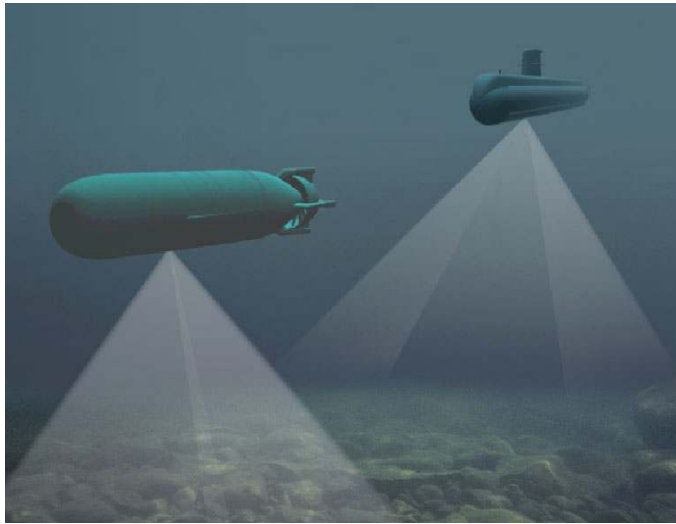


Figure 3.6: The bottom topography measurement with the correlation method. The vehicles are measuring the bottom topography by several sonar beams. Likely positions are positions with the same bottom topography.

- Low requirements on availability and quality of underwater maps
- Very low risk of revealing when a 3D-sonar is used since the duration of the sonar ping is ~ 1 ms.
- Very high accuracy of the position
- Ability to determine own bearing
- High ability to determine the position in flat bottomed areas
- Ability to do own mapping of the bottom topography.

Other important characteristics are:

- Very low initialization requirements
- The system can be used by a stationary vehicle
- Very short time for a position fix from a non-initiated system

Most of these characteristics will be shown to be true in the coming chapters.

The vehicle position is thus established by computing the correlation sum for every point in the map and taking the position with the minimum correlation sum as the true position. The correlation sum in a map position is

$$T(\mathbf{x}_t) = \sum_{k=1}^N (y_{t,k} - h_k(\mathbf{x}_t))^2 \quad (3.1)$$

That is, calculating the sum of the squared differences of the measured depth $y_{t,k}$ at sonar beam k with the corresponding map depth $h_k(\mathbf{x}_t)$ at beam k at position \mathbf{x}_t . The measurement is done at time t . The number of measurement beams is N . By letting the depth values be deviations from the respective mean values, the bottom profiles are compared and the influence of the actual sea level is eliminated. However, it should be noted that the absolute depth value occasionally may be needed for a unique position.

We know from Chapter 2 that the PDF for the position is the result of multiplication of the prior PDF and a likelihood function

$$p(\mathbf{x}_{t+1} | \mathbf{Y}_{t+1}) \sim L(\mathbf{x}_{t+1}; \mathbf{y}_{t+1})p(\mathbf{x}_{t+1} | \mathbf{Y}_t) \quad (3.2)$$

It is therefore of interest and quite illustrative to look at the shape of the likelihood function calculated according to (2.8) and the posterior PDF at different numbers of measurement beams to see the great difference between using only

one beam as in single beam methods, the traditional method, and using the many beam method as proposed in this thesis.

Figure 3.7 shows the non-normalized likelihood function for a measurement with only one beam, which is the usual way that the measurement is done in terrain navigation. Figure 3.8 shows the likelihood function corresponding to a measurement with 9×9 beams. The size of the map is 4×4 km in both cases. It can be seen from the figures that the larger number of beams drastically reduces the number of likely positions. It will be shown in Chapter 4 that the secondary false correlation peaks are due to terrain repeatability and that they will decrease exponentially with the number of measurement beams, i.e., the size of the

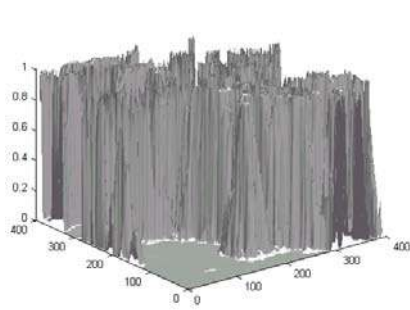


Figure 3.7: Likelihood function when using one measurement beam.

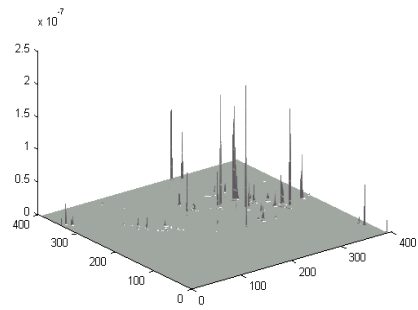


Figure 3.8: Likelihood function when using 9×9 measurement beams.

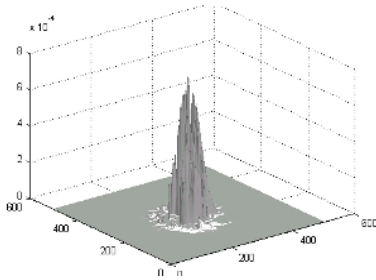


Figure 3.9: The measurement updated prior for one measurement beam.

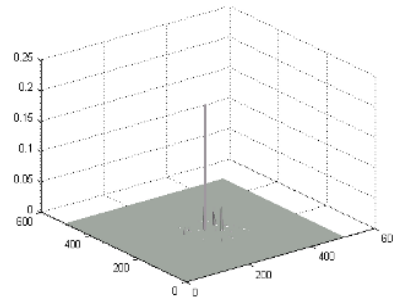


Figure 3.10: The measurement updated prior for 9×9 measurement beams.

correlation area properly sampled. The reason for naming them to be false is that they disappear when N tends to infinity and there will only be one peak left, the true position.

3.3 The Bayesian and the maximum likelihood method (ML)

3.3.1 An illustrating example

An underwater vehicle is assumed to move from a position A to a position B by the help of an inertial navigation system (INS). When the vehicle assumes it is in position B it measures the depth for aiding in the determination of the position. Figure 3.11 shows the bottom profile at position B.

By assuming error distributions of the position errors of the INS-system and the depth measurement we can simulate the journey from A to B by a Monte Carlo experiment. Figures 3.12 and 3.13 show the joint PDF of position and depth measurement from an experiment with Gaussian error distributions. In the contour plot Figure 3.13 the line $p(x, y=y_m)$ represents the outcomes of the experiment that have the measured depth y_m , i.e., $p(x|y=y_m)p(y=y_m)$. The likelihood function at position x is the relative frequency of the outcomes at position x that has the measured depth y_m , i.e., $p(x, y=y_m)/p(x)$.

Let us now extend the Monte Carlo experiment by also examining the position accuracy. The likelihood function (2.8) can also be seen as the likelihood function in maximum likelihood estimation (ML). Figure 3.14 shows the outcome of this part of the maximum likelihood experiment. The likelihood function is calcu-

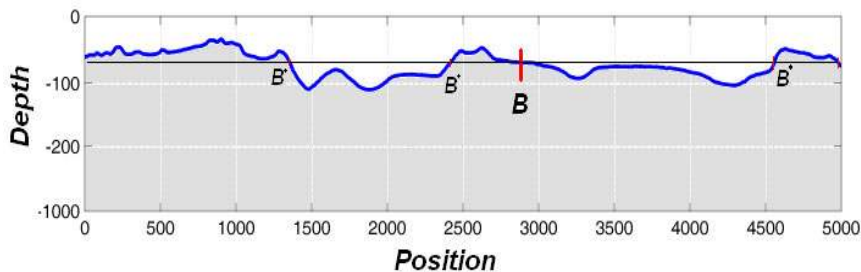


Figure 3.11: The bottom profile around position B. Alternative possible positions according to the measurement are indicated.

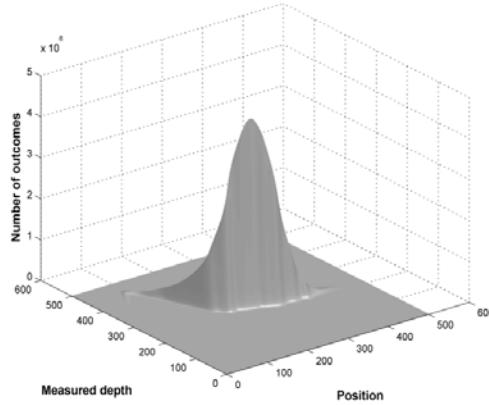


Figure 3.12: Illustration of the joint probability density function, centered around position B, of the position and the measured depth.

lated using $N=51$ measuring beams instead of 1 beam as in the previous experiment. The likelihood function will differ slightly from realization to realization and in the figure the mean of all realizations is shown. From the figure we clearly see that the maximum likelihood estimation gives a much better accuracy than

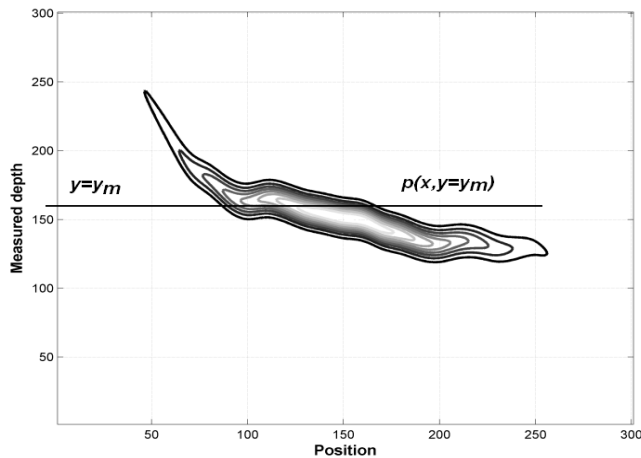


Figure 3.13: A contour plot of the joint density function $p(x, y)$ of position and depth in Figure 3.12.

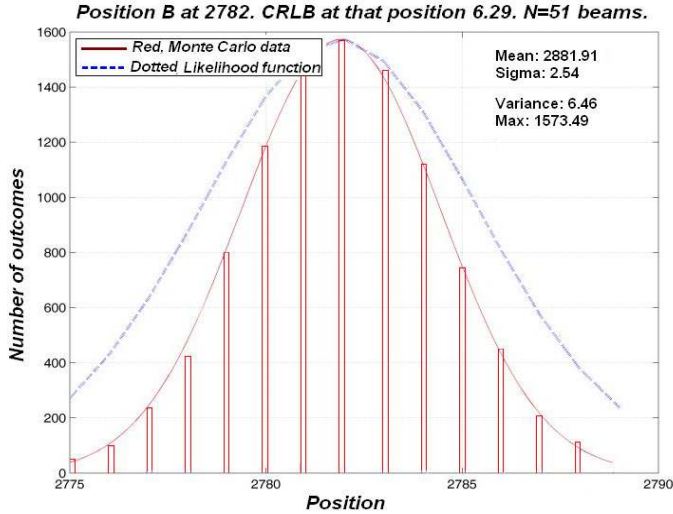


Figure 3.14: A histogram for the outcomes of the Monte Carlo positioning experiment. The dotted curve is the mean of the likelihood functions for each outcome. The variance calculated from the mean likelihood function is 10.7.

indicated by the mean likelihood function which is natural since the horizontal position of each individual likelihood function varies according to the positions given by the histogram. The Cramér-Rao lower bound, see Chapter 5, gives 6.29 as the lowest error variance we can achieve and we reached 6.46. The error distribution is asymptotically Gaussian, that is when the number of trials tends to infinity it will become Gaussian. We conclude from the figure that we are close to the asymptotic distribution.

Instead of the Bayesian estimation thinking we can perceive the situation as having two measurements of the position at B, one from the INS-system and one from the maximum likelihood estimation. A way of fusing the measurements together is to use the linear fusing method discussed in Section 3.3.2 which will give a minimum variance estimate of the position.

Since the linear fusing method will give the smallest position error variance, we will use this method in the practical estimation of the position. The involved distributions are also close to Gaussian due to the large number of measuring

beams we are typically using (~ 400).

We may interpret the situation as follows. When we are in position A all we can say about our future position at B is contained in the joint PDF of the position and an assumed future measurement. However, when we are in position B our position is no longer stochastic but deterministic but unknown by us. Even if we do not know the position ourselves others may know it and by that it is strictly not stochastic any more. We are thus in a position to estimate the position by linearly fusing the positions given by the INS-system and the maximum likelihood estimation method.

3.3.2 The linear fusing method

Linear fusing is a way to present the Kalman equations when several sensor data are fused together into a single measurement. The method means that a new estimate is formed by a linear combination of independent unbiased estimates. The linear fusing method is equivalent to the linear Kalman filter. In the case we have two estimates of the position we calculate the fused estimate of $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ as [KSH00, CANY05]

$$\boldsymbol{\mu} = E\{\hat{\mathbf{x}}\} = \mathbf{A}_1 E\{\hat{\mathbf{x}}_1\} + \mathbf{A}_2 E\{\hat{\mathbf{x}}_2\} = (\mathbf{A}_1 + \mathbf{A}_2)\boldsymbol{\mu} \quad (3.3)$$

The new estimate should be unbiased

$$E\{\hat{\mathbf{x}}\} = \mathbf{A}_1 E\{\hat{\mathbf{x}}_1\} + \mathbf{A}_2 E\{\hat{\mathbf{x}}_2\} \quad (3.4)$$

which gives

$$\mathbf{A}_1 + \mathbf{A}_2 = \mathbf{I} \quad (3.5)$$

The covariance of the fused estimate is

$$\mathbf{P}_{\hat{\mathbf{x}}\hat{\mathbf{x}}} = \mathbf{A}_1 \mathbf{P}_{\hat{\mathbf{x}}_1\hat{\mathbf{x}}_1} \mathbf{A}_1^T + \mathbf{A}_2 \mathbf{P}_{\hat{\mathbf{x}}_2\hat{\mathbf{x}}_2} \mathbf{A}_2^T = \mathbf{A}_1 \mathbf{P}_{\hat{\mathbf{x}}_1\hat{\mathbf{x}}_1} \mathbf{A}_1^T + (\mathbf{I} - \mathbf{A}_1) \mathbf{P}_{\hat{\mathbf{x}}_2\hat{\mathbf{x}}_2} (\mathbf{I} - \mathbf{A}_1)^T \quad (3.6)$$

which is quadratic in \mathbf{A}_1 . Completing the square gives the minimum trace of the covariance matrix for

$$\mathbf{A}_1 = \mathbf{P}_{\hat{\mathbf{x}}_2 \hat{\mathbf{x}}_2} \left[\mathbf{P}_{\hat{\mathbf{x}}_1 \hat{\mathbf{x}}_1} + \mathbf{P}_{\hat{\mathbf{x}}_2 \hat{\mathbf{x}}_2} \right]^{-1} \quad (3.7)$$

$$\mathbf{A}_2 = \mathbf{P}_{\hat{\mathbf{x}}_1 \hat{\mathbf{x}}_1} \left[\mathbf{P}_{\hat{\mathbf{x}}_1 \hat{\mathbf{x}}_1} + \mathbf{P}_{\hat{\mathbf{x}}_2 \hat{\mathbf{x}}_2} \right]^{-1} \quad (3.8)$$

and the minimum covariance of the fused estimator as

$$\mathbf{P}_{\hat{\mathbf{x}}\hat{\mathbf{x}}}^{-1} = \mathbf{P}_{\hat{\mathbf{x}}_1 \hat{\mathbf{x}}_1}^{-1} + \mathbf{P}_{\hat{\mathbf{x}}_2 \hat{\mathbf{x}}_2}^{-1} \quad (3.9)$$

and the fused estimate as

$$\hat{\mathbf{x}} = \mathbf{P}_{\hat{\mathbf{x}}_2 \hat{\mathbf{x}}_2} \left[\mathbf{P}_{\hat{\mathbf{x}}_1 \hat{\mathbf{x}}_1} + \mathbf{P}_{\hat{\mathbf{x}}_2 \hat{\mathbf{x}}_2} \right]^{-1} \hat{\mathbf{x}}_1 + \mathbf{P}_{\hat{\mathbf{x}}_1 \hat{\mathbf{x}}_1} \left[\mathbf{P}_{\hat{\mathbf{x}}_1 \hat{\mathbf{x}}_1} + \mathbf{P}_{\hat{\mathbf{x}}_2 \hat{\mathbf{x}}_2} \right]^{-1} \hat{\mathbf{x}}_2 \quad (3.10)$$

By recursion the expressions in (3.9) and (3.10) can easily be generalized to M measurements

$$\mathbf{P}_{\hat{\mathbf{x}}\hat{\mathbf{x}}}^{-1} = \sum_{i=1}^M \mathbf{P}_{\hat{\mathbf{x}}_i \hat{\mathbf{x}}_i}^{-1} \quad (3.11)$$

$$\hat{\mathbf{x}} = \mathbf{P}_{\hat{\mathbf{x}}\hat{\mathbf{x}}} \sum_{i=1}^M \mathbf{P}_{\hat{\mathbf{x}}_i \hat{\mathbf{x}}_i}^{-1} \hat{\mathbf{x}}_i \quad (3.12)$$

Assume now, to begin with, that the frequency distributions p_1 and p_2 of $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$, respectively, are Gaussian with dimension n . We multiply the PDF's

$$\begin{aligned} p(\mathbf{x}) &= p_1(\mathbf{x}) p_2(\mathbf{x}) \\ &= \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{P}_{\hat{\mathbf{x}}_1 \hat{\mathbf{x}}_1})}} \exp\left(-\frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}}_1)^T \mathbf{P}_{\hat{\mathbf{x}}_1 \hat{\mathbf{x}}_1}^{-1} (\mathbf{x} - \hat{\mathbf{x}}_1)\right) \\ &\quad \cdot \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{P}_{\hat{\mathbf{x}}_2 \hat{\mathbf{x}}_2})}} \exp\left(-\frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}}_2)^T \mathbf{P}_{\hat{\mathbf{x}}_2 \hat{\mathbf{x}}_2}^{-1} (\mathbf{x} - \hat{\mathbf{x}}_2)\right) \end{aligned} \quad (3.13)$$

and it is well known [AN79, MA82a] that this function, when normalized, is a Gaussian PDF with a covariance and a mean of

$$\mathbf{P}_{\hat{\mathbf{x}}\hat{\mathbf{x}}}^{-1} = \sum_{i=1}^{M=2} \mathbf{P}_{\hat{\mathbf{x}}_i\hat{\mathbf{x}}_i}^{-1} \quad (3.14)$$

$$\hat{\mathbf{x}} = \mathbf{P}_{\hat{\mathbf{x}}\hat{\mathbf{x}}} \sum_{i=1}^{M=2} \mathbf{P}_{\hat{\mathbf{x}}_i\hat{\mathbf{x}}_i}^{-1} \hat{\mathbf{x}}_i \quad (3.15)$$

The linear fusion can thus also be seen as a multiplication of the PDFs if the involved PDFs are Gaussian.

However, since no explicit expression is known for the variance of the maximum likelihood estimate (MLE) we have to resort to the Cramér-Rao expression for the variance. This will in many cases give an accurate approximation of the variance of the maximum likelihood estimate since the convergence towards the Cramér-Rao limit is mostly quite fast [AZ96, KA93]. The matter will be further touched on in Chapter 6 and 8 where some graphs will show the rapid convergence in our case.

3.3.3 Multiple likelihood peaks

The difference between the Bayesian method and the linear fusing of the prior PDF and the maximum likelihood estimate is illustrated in Figure 3.15 where a step in the Bayesian recursion is illustrated. To the left in the figure is the previous estimated PDF for the position and to the right is that PDF propagated and it is our estimated position PDF before our measurement of the position. Now assume that our measurement gives a bimodal likelihood function due to terrain repeatability. The right peak of the likelihood function has the highest amplitude and from our knowledge of the measurement error we conclude that the highest peak indicates our position with high probability. Still, after fusing the likelihood function and the prior position PDF the left peak will give the maximum posterior position. The mean position would be between the peaks but close to the left peak. The Bayesian approach will thus give different positions depending on if the mean or the maximum a posteriori (MAP) position is used as the estimated position. It may also give a position which has a low probability according to the measurement. The maximum likelihood estimation, which is asymptotically unbiased, will choose the higher peak as the true one and the linear fusing method will thus fuse this peak with the prior position PDF. This estimated position will thus differ from the one estimated by the Bayesian approach.

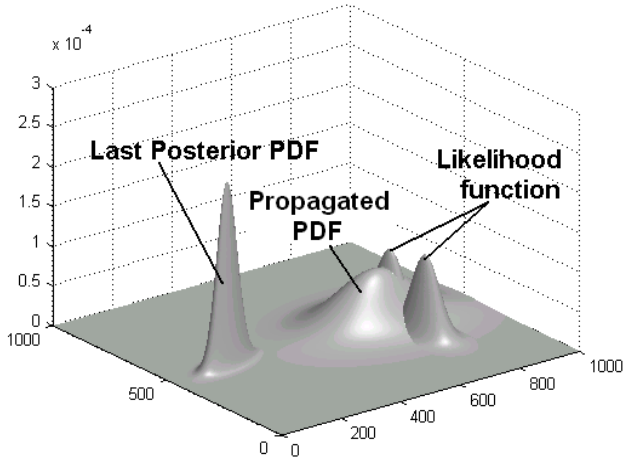


Figure 3.15: Illustration of the prior position PDF and a bi-modal likelihood function.

In Chapter 4 it will be shown that not only the probability of wrong position due to false peaks will decrease exponentially but also that the amplitudes of the false peaks will decrease exponentially with the number of measurement beams.

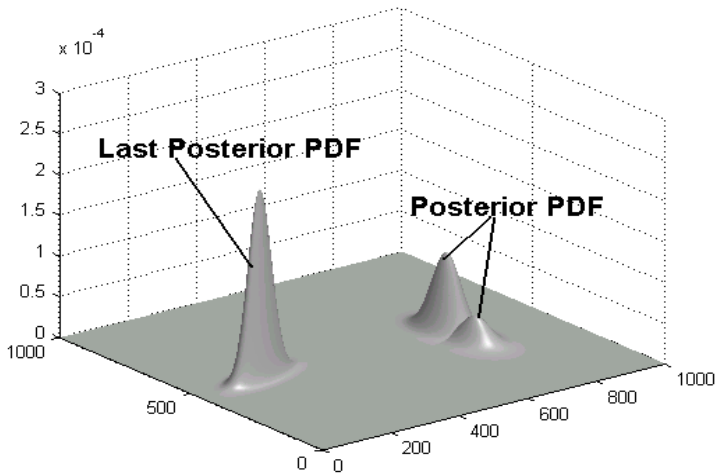


Figure 3.16: Illustration of the posterior (measurement updated prior) position PDF .

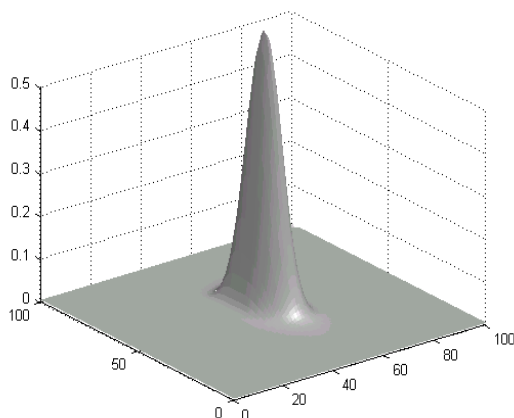


Figure 3.17: A typical N -normalized likelihood function from a sea-trial (400 beams).

Figures 3.9 and 3.10 show the non-normalized measurement-updated PDF for a real map. A comparison between figures shows that the variance is much smaller when correlation with many measurement beams is used.

If we look closer at the correlation peak in Figure 3.10 we will find that it is approximately a Gaussian curve. Figure 3.17 shows the N -normalized likelihood function from a positioning in a sea-trial and the Gaussian shape is clearly visible. The curve in Figure 3.17 is N -normalized, i.e., if the curve is self-multiplied N times (the number of beams) we will have the true likelihood function. It will be shown in Chapter 4 that the likelihood function converges to a Gaussian function when the number of measurement beams tends to infinity. It will also be shown that this Gaussian curve, if normalized to have unit volume, will asymptotically coincide with the PDF of the estimation error in the maximum likelihood estimation. That is, the unit normalized likelihood function when the number of measurement beams is large will also be the PDF for the position error.

If the time between the measurements is large the variance in the propagated PDF, $p^-(\mathbf{x})$, may increase so much that it will not noticeably improve the estimation of the position based on the likelihood function. A typical RMS-error of the prior PDF in submarine applications may be ~ 3000 meter while maximum likelihood estimation gives an RMS-error of, say, less than a meter.

Hence, when the prior PDF has a large variance, the estimation problem becomes a maximum likelihood estimation problem without prior PDF.

A problem both in the Bayesian method and the linear fusing method is the problem with multiple peaks, false peaks, due to terrain repeatability since the likelihood function (2.8) is not a one to one function, see Figure 3.11. Figure 3.18 shows a terrain profile and the vehicle measures the depth y when it is in the position x_0 . The PDF for the measurement error is assumed to be symmetric with cut tails. Positions outside the range of the PDF can not be explained by the measurement error and the likelihood function for the true position which is the one marked with x_0 in the figure.

Now, as can be seen from Figure 3.18 there are other positions that will give the same depth measurement but of course the vehicle can only be at one place at a time. As common in the Bayesian method to use the mean as the probable position may lead to an estimated position between the peaks of the posterior PDF with no likelihood. This is not acceptable from a physical standpoint.

Therefore all likely positions according to the likelihood function have to be evaluated one by one. A technique for handling such situations is track splitting which is frequently used in tracking of flying objects [BSFO88]. In this technique parallel tracks are formed and as soon as a track is found false it is terminated. However, this is not as easy as expected and does not give an answer in real time about which peak represents the true position which may be

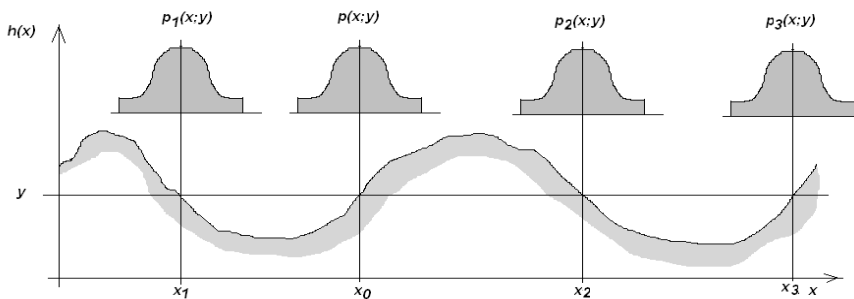


Figure 3.18: The figure illustrates how false likelihood peaks arise due to terrain repeatability. True position is x_0 .

needed by the guidance system because its action may depend on the estimated position.

In practice the individual PDFs for the measured position candidates may not be so well separated as assumed above. However, we can draw the important conclusion that in Bayesian terrain navigation the maximum a posteriori (MAP) position estimate is to prefer instead of the mean estimate when the posteriori PDF has multiple peaks [GDW00]. Then we note that the linear fused estimate is not so influenced by false peaks as the Bayesian MAP position if the peaks are not so well separated. We also note that a position variance calculation based on the variance of the Bayesian posterior PDF is influenced by false peaks. False peaks will also give a position bias if the position is calculated as the mean of the Bayesian posterior. It is only when we have a large number of measurement beams that the two methods of calculation of the position and position error will coincide. A broad discussion about the role of the mean-measure and other measures in Bayesian inference can be found in [HAFO04]. In [CA88] multiple peaks in maximum likelihood estimation are discussed.

3.4 Concluding remarks about the basic terrain navigation method

Determining the position of underwater vehicles using the proposed correlation method has been studied over the past years. The correlation method has proven to be robust, highly precise and minimally revealing. It is also easily implemented and demands a minimum of underwater maps.

The proposed method do only terrain positioning in the navigation cells but, of course, the method is also suitable for continuous navigation if the required maps are available.

The terrain positioning can easily be aided by side-scan and sediments maps or by directions to external sound sources since likelihood functions from independent measurements are simply fused by multiplications if the processes are Markov which they often can be considered to be.

3.5 A way of implementing the approximate method

When an approximate 3D picture of the bottom is to be assembled from several beam planes it is also possible to correlate each beam plane against the underwater map separately at each ping and then join them together. This will shorten the time delay for a position estimate after the last ping. Figure 3.19 shows the beam footprint pattern from 5 beam planes which are moved around, as a packet, in the map to find the position where it fits the best. The best fit is considered to be the position where the correlation function

$$T(\mathbf{x}, d) = \sum_{i=1}^N (y_i - h_i(\mathbf{x}) - d)^2 \quad (3.16)$$

has its minimum. In the expression y_i is the measured depth at beam i and $h_i(\mathbf{x})$ is the corresponding depth according to the map in position \mathbf{x} and d is the difference between the actual sea level and the reference level of the map (the measurement bias). The total number of beams is N .

Usually the correlation function (3.16) is calculated for all beam planes at once but the correlation function can also be written as

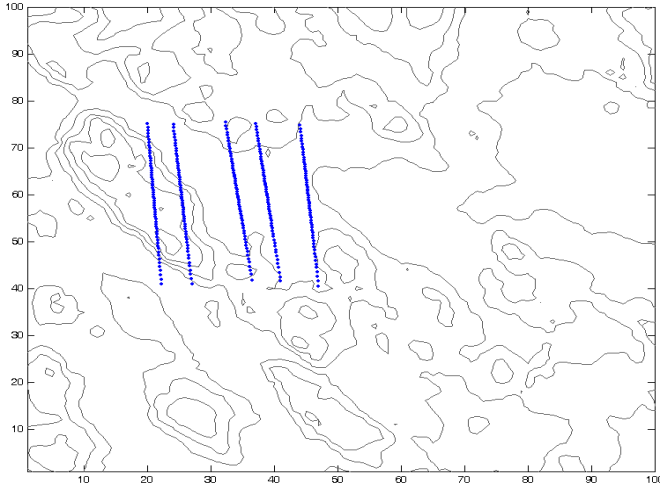


Figure 3.19: Beam footprint pattern from five beamplanes to be moved around in the map.

$$\begin{aligned}
T(\mathbf{x}, d) = & \underbrace{\sum_{i=1}^{N_1} (y_{1,i} - h_{1,i}(\mathbf{x}) - d)^2}_{\text{Beam-plane 1}} + \underbrace{\sum_{i=1}^{N_2} (y_{2,i} - h_{2,i}(\mathbf{x}) - d)^2}_{\text{Beam-plane 2}} + \dots \\
& + \underbrace{\sum_{i=1}^{N_M} (y_{M,i} - h_{M,i}(\mathbf{x}) - d)^2}_{\text{Beam-plane M}}
\end{aligned} \tag{3.17}$$

where the correlation sum for each beam plane is calculated separately with the same reference point and after the final measurement (ping) they are summed up. The calculation for the separate beam planes must use the same reference point so the calculation is equivalent to matching all beam planes simultaneously.

The measurement bias shall be determined when all the final measurements are done in order to have the best accuracy. Since the partial correlation sum is

$$T_k(\mathbf{x}, d) = \sum_{i=1}^{N_k} (y_{k,i} - h_{k,i}(\mathbf{x}) - d)^2 = \sum_{i=1}^{N_k} (y_{k,i} - h_{k,i}(\mathbf{x}))^2 - 2d \sum_{i=1}^{N_k} (y_{k,i} - h_{k,i}(\mathbf{x})) + N_k d^2 \tag{3.18}$$

the sum $\sum_{i=1}^{N_k} (y_{k,i} - h_{k,i}(\mathbf{x}))$ has thus also to be calculated. When all measure-

ments are done the estimated position can be determined by minimizing $T(\mathbf{x}, d)$ by some numerical procedure. This 3D minimization problem can, however, be simplified by first minimizing $T(\mathbf{x}, d)$ with respect to d and substituting the expression for d into (3.18) making it a 2D minimization problem. Then the estimated position is the \mathbf{x} that minimizes

$$\sum_{k=1}^M \sum_{i=1}^{N_k} (y_{k,i} - h_{k,i}(\mathbf{x}))^2 - \frac{1}{N} \left[\sum_{k=1}^M \sum_{i=1}^{N_k} (y_{k,i} - h_{k,i}(\mathbf{x})) \right]^2 \tag{3.19}$$

The estimated measurement bias is then calculated as (see also Chapter 5)

$$\hat{d} = \frac{1}{N} \sum_{i=1}^N (y_i - h_i(\hat{\mathbf{x}})) \tag{3.20}$$

Using this implementation method the total computational load for a position estimate is divided into parts which are computed sequentially at the ping events. The time between the pings (measurements) is often quite large (~ 10 s)

so there will be sufficient time for the calculations even if a general computer is used for the calculation. Another advantage is that the position accuracy can be monitored as a function of the number of measurements (pings) by calculation of the Hessian in the position of maximum of the likelihood function. This is of interest since assembling of beam planes for forming a 3D picture of the bottom is just an approximation. When assembling beam planes there will always be a stochastic error in the distance and angle between the beam planes. The more planes that are added the larger will the error between the first and the last plane be. It is thus of interest to use as few beamplanes as possible. In the sea-trial that has been conducted, with the equipment that has been used, the number of beam planes should not be more than 11 if the distance between the planes is 25 meter. An increase above that will in most cases not improve the position accuracy. By studying the Hessian after each measurement it is possible to interrupt the collection of beam planes when the increase in accuracy falls below some value.

3.6 Comparison between the single beam missile method and the correlation method

The main purpose of the correlation method in recursive applications is to make a linear Kalman filter feasible, to achieve high position accuracy and robustness. Therefore the method, even in a recursive application with dense sampling, always uses a large number of measuring beams. However, even for an equal total number of measurement beams, the correlation method will always give a smaller position error compared to the traditional terrain navigation method with a single measuring beam. This because the distances between the footprints of the measurement beams are random variables in the single beam methods but fix known values in the correlation method.

The implication of this can easily be illustrated in the one-dimensional case. If we look at a single beam method as a batch method we collect N depth measurements. Then we perform a correlation against the map and we will have a situation as depicted in Figure 3.20. In the correlator method we have fixed known distances between the nominal sampling points which are indicated by circles in the figure. In the single beam methods we have errors in determining our position and these errors are indicated by ϵ_k . These position errors will cause errors in the map heights determination.

A first order approximation will give us the likelihood function

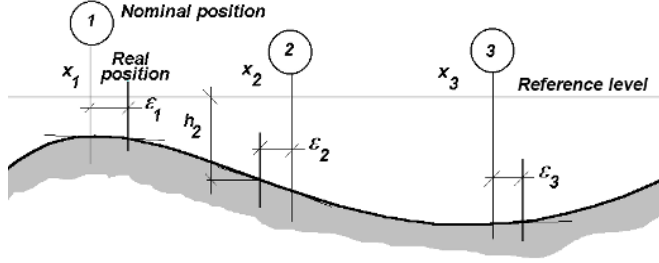


Figure 3.20: The figure illustrates that an offset from the nominal sampling positions, caused by the uncertainty in the position, causes an additive height error.

$$L(x; y) = \frac{1}{\sqrt{(2\pi\sigma_e^2)^N}} \exp\left(-\frac{1}{2\sigma_e^2} \sum_{k=1}^N (y_k - h(x_k) - \frac{\partial h(x_k)}{\partial x} \epsilon_k)^2\right) \quad (3.21)$$

where x_k is the believed position at time instant k and ϵ_k is the random position error from that position. We can see this as having an additive white measurement noise with the total covariance matrix

$$\mathbf{C} = \text{diag}(\mu\sigma_{\epsilon,i}^2 + \sigma_e^2)_{i=1 \dots N} \quad (3.22)$$

where the terrain constant μ is the expected value of the square of the terrain gradient

$$\mu = E\left\{\left[\frac{\partial h}{\partial x}\right]^2\right\} \quad (3.23)$$

The terrain constant μ is assumed to be the same in all points and it is a characteristic of the actual terrain. Flat terrain will have a small μ -value and more hilly terrain will have a higher μ -value. Figure 3.21 shows the one-sided spatial power spectral density, $S_h(\omega)$, for a one dimensional terrain profile. For a stochastic variable $h(x)$ the expected value of the square of a derivative of the variable is [NR75]

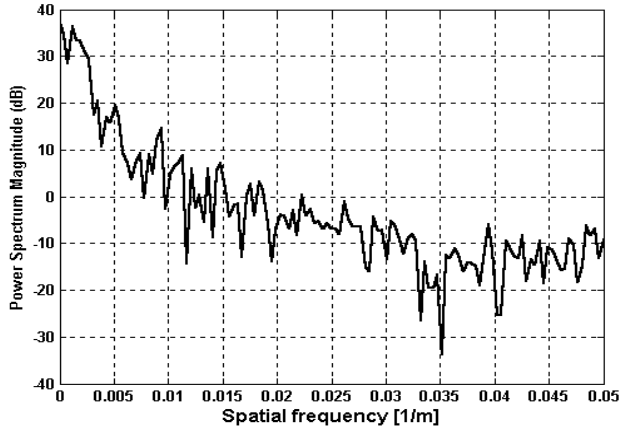


Figure 3.21: The figure shows the spatial one-sided power spectral density for a terrain profile.

$$\mu(n) = E \left\{ \left[\frac{\partial^n h}{\partial x^n} \right]^2 \right\} = \int_0^\infty \omega^{2n} S_h(\omega) d\omega \quad (3.24)$$



Figure 3.22: The position error variance as a function of the position error variance in sampling of the terrain topography. Synthetic $AR(1)$ terrain with a pole in 0.98.

The spectrum of a terrain profile can often be assumed to have an exponential form $S_h(\omega) = S_0 \exp(-\omega / \omega_0)$ where ω_0 is a characteristic of the actual terrain. In such a case we have

$$\mu = 2S_0\omega_0^3. \quad (3.25)$$

We note the heavy dependence on the high frequency content in the terrain profile. The assumption of the same μ -value for all sampling points will give the following simple expression for position the error covariance matrix in the two dimensional case

$$\mathbf{R}_{Single\ beam} = (\sigma_e^2 + \mu\sigma_\varepsilon^2) \begin{bmatrix} \sum_{i=1}^N \left[\frac{\partial h_i}{\partial x_1} \right]^2 & \sum_{i=1}^N \frac{\partial h_i}{\partial x_1} \frac{\partial h_i}{\partial x_2} \\ \sum_{i=1}^N \frac{\partial h_i}{\partial x_2} \frac{\partial h_i}{\partial x_1} & \sum_{i=1}^N \left[\frac{\partial h_i}{\partial x_2} \right]^2 \end{bmatrix}^{-1} \quad (3.26)$$

or

$$\mathbf{R}_{Single\ beam} = (1 + \mu \frac{\sigma_\varepsilon^2}{\sigma_e^2}) \mathbf{R}_{ETL} \quad (3.27)$$

where \mathbf{R}_{ETL} is the position error covariance for the proposed method. Chapter 5 and 6 will go into the details of the expression. The orders of the two terms in the first factor of (3.26) are about the same in flat areas so even for the same number of measuring beams the correlator method clearly gives smaller position errors. Figure 3.22 shows how the resulting position error varies with the position error in sampling the terrain topography. The figure shows the result from a one dimensional Monte Carlo simulation with a synthetic $AR(1)$ terrain topography.

Figure 3.22 seems to indicate that the relation should be more quadratic instead of linear but in (3.21) we have only included the linear term, adding another term in the Taylor expansion would add a quadratic term as well. The expression also explains why the performance of the missile method in rough terrain is lower than expected from the terrain roughness only. The expression (3.26), given the assumptions made, is formally equivalent to the solution of the algebraic Riccati equation for the estimation error in linear Kalman filtering.

The difference in accuracy between the two methods can also be illustrated in the 1D-case if we make some approximations. It follows from the central limit theorem that the expected function of the likelihood function is a Gaussian

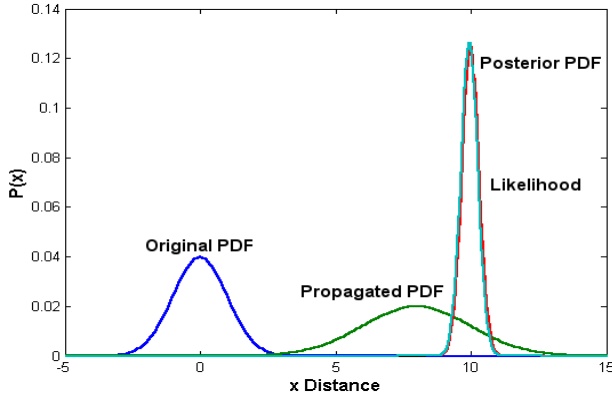


Figure 3.23: Illustration of a recursive step under the simplified assumptions given in the text. The posterior PDF and the likelihood function almost coincide.

function. If we therefore replace the likelihood function, when we measure with only one beam, with its expected Gaussian function and assume that the vehicle position at the sampling event has a Gaussian error, we can directly apply the equations for the linear Kalman filter (see Chapter 4). We also assume that the likelihood function and the position error PDF are the same at all sampling events. The starting PDF is assumed to have a variance of σ_0^2 and it is assumed that during the propagation it will broaden to $\alpha\sigma_0^2$ where $\alpha > 1$. The likelihood function, normalized to a PDF, is assumed to have a variance of σ_L^2 . The posterior PDF will therefore have the following variance according to the “parallel resistor formulae” for adding variances

$$\frac{1}{\sigma^2} = \frac{1}{\alpha\sigma_0^2} + \frac{1}{\sigma_L^2} \quad (3.28)$$

If we now simplify our discussion by assuming that the propagation constant and likelihood function will be the same in all following steps we will come to the following result

$$\frac{1}{\sigma^2} = \frac{1}{\alpha^N \sigma_0^2} + \frac{1}{\sigma_L^2} \left[\frac{1}{\alpha^N} + \frac{1}{\alpha^{N-1}} + \dots + 1 \right] \quad (3.29)$$

For infinitely many steps we will have

$$\frac{1}{\sigma^2} = \frac{1}{\sigma_L^2} \frac{1}{1 - 1/\alpha} \quad (3.30)$$

The procedure is illustrated in Figure 3.23. For large N the starting distribution has no influence on the estimation and for large propagation errors the estimation is only dependent on the last likelihood function. Now, it must be noticed that replacing the likelihood function with its estimated function is a rather cruel action.

3.7 Determining the measurement bias

Position bias can be caused by many sources but caused by measurement bias it is seldom any problem since in many cases the errors in the height measurements will give a diagonal position error covariance matrix $\mathbf{C}_e = \sigma_e^2 \mathbf{I}$. In this case the bias is easily removed by reducing the measured heights and the map heights by their means before the correlation process. Figure 3.24 shows how measurement bias occurs when the sea level differs from the reference level of the map.

The upper figure in Figure 3.25 shows the measured profile and the map profile

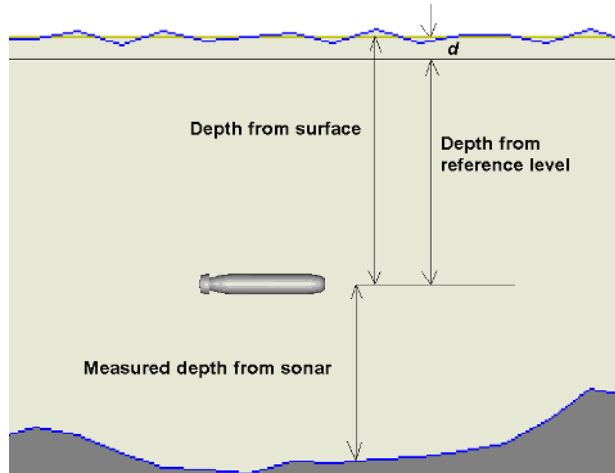


Figure 3.24: The vehicle measures the depth from the actual sea level. The depth in the underwater map is from the reference level.

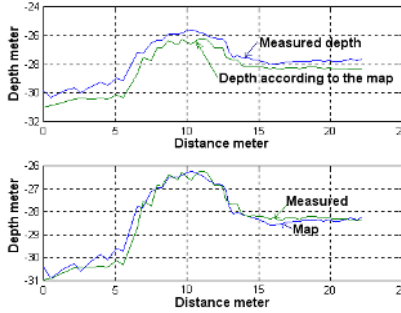


Figure 3.25: Illustration of measured depth and depth according to the underwater map. Sea-trial 1998.

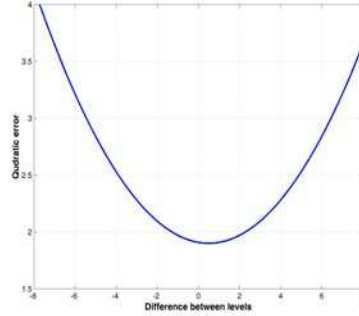


Figure 3.26: The correlation function in the true position as a function of the measurement bias.

from a sea trial performed in 1998. As can be seen there is a measurement bias. By sliding the measured profile downwards as in the lower figure in Figure 3.25 the correlation error sum will decrease as shown in Figure 3.26 to a minimum before rising. The measurement bias is taken as the displacement which gives the minimum correlation error sum. Formally we can determine the minimum as follows.

Assume that we have a measurement bias of d in each measurement beam. We consider d to be an unknown constant. Our correlation function in the true position \mathbf{x}_0 , when our measurement is \mathbf{y} , can then be written as

$$T(\mathbf{x}_0) = (\mathbf{y} - \mathbf{h}(\mathbf{x}_0) - d\mathbf{u})^T \mathbf{C}_e^{-1} (\mathbf{y} - \mathbf{h}(\mathbf{x}_0) - d\mathbf{u}) \quad (3.31)$$

The column vector \mathbf{u} has ones as elements. $T(\mathbf{x}_0)$ is a function of the bias d and taking the first derivative with respect to d and equalling it to zero gives

$$\mathbf{u}^T \mathbf{C}_e^{-1} (\mathbf{y} - \mathbf{h}(\mathbf{x}_0) - d\mathbf{u}) = 0$$

or

$$\hat{d} = [\mathbf{u}^T \mathbf{C}_e^{-1} \mathbf{u}]^{-1} \mathbf{u}^T \mathbf{C}_e^{-1} (\mathbf{y} - \mathbf{h}(\mathbf{x}_0)) \quad (3.32)$$

In the case we have

$$C_e = \text{diag}(\sigma_{e,1}^2, \dots, \sigma_{e,N}^2)$$

$$\hat{d} = \left[\sum_{i=1}^N \frac{1}{\sigma_{e,i}^2} y_i - \sum_{i=1}^N \frac{1}{\sigma_{e,i}^2} h_i(\mathbf{x}_0) \right] \bigg/ \sum_{i=1}^N \frac{1}{\sigma_{e,i}^2} \quad (3.33)$$

and if we have the same measurement uncertainty in all beams

$$\hat{d} = \frac{1}{N} \sum_{i=1}^N y_i - \frac{1}{N} \sum_{i=1}^N h_i(\mathbf{x}_0) = \bar{y} - \bar{h}(\mathbf{x}_0) \quad (3.34)$$

where the scalars \bar{y} and \bar{h} are the mean values of the elements in respective vector. The true position may be situated between the steps in the matching process and in this case the above estimation of the measurement bias will not be correct. However, the error due to that may often be neglected.

3.8 Incorporation of sediment data in the correlation

3.8.1 Introduction

Even if the proposed correlation method increases the navigation performance substantially in flat bottom areas compared to single beam methods there will of course be a limit when the method no longer is sufficient for establishing a position. One way to push this limit is to also include the sediment profile or side-scan pictures in determining the position.

Whether our estimation is to be considered as Bayesian or a linear fusing process it is easy to include disparate independent measurements of the state of a Markov process since for a Markov process we have the conditional PDF

$$p(x | y_1, y_2, \dots, y_N) = p(x | y_1) p(x | y_2) \dots p(x | y_N) \quad (3.35)$$

and thus for the likelihood function

$$L(x; y_1, y_2, \dots, y_N) = L_1(x; y_1) L_2(x; y_2) \dots L_N(x; y_N) \quad (3.36)$$

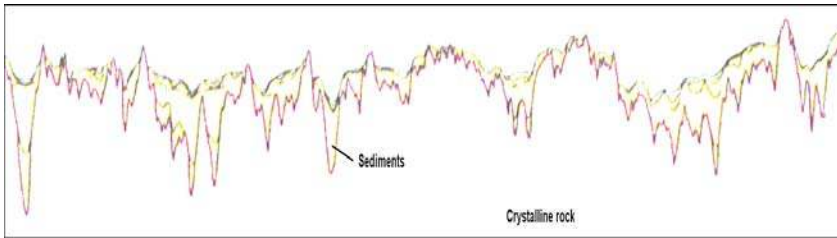


Figure 3.27: The sediment profile along a line in the test area during the October 2002 sea-trial.

Not only physical measurements but also subjective information can be used for formulating a likelihood function. If the measurements are not independent that has to be taken into account.

3.8.2 Incorporation of sediment data

Flat bottom areas are often characterized by soft sediment layers and sediment rocks (here taken together as sediment layers) on a more hilly rough crystalline underground of primary rock or moraine ridges, or similar objects, from the last glacial period. The ridges may have been created at the edge of the ice cap or created by ice river deltas. The Baltic sea, for which the navigation method has



Figure 3.28: An example of a parametric sonar manufactured by Innomar, Germany. Secondary frequency 4...12 kHz and penetration depth down to 50 meter. Size 250 x 350 x 80 mm.

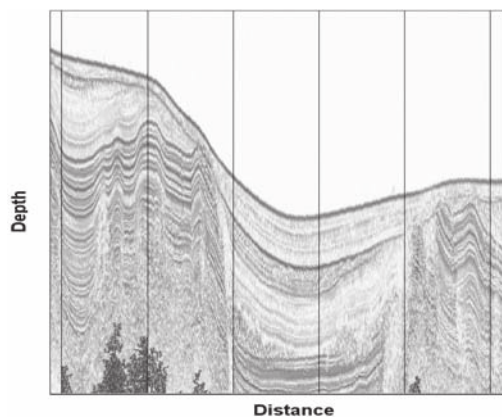


Figure 3.29: Sediment profile at way-point 2C in the October 2002 sea-trial. Note that the vertical scale is exaggerated, see Figure 3.30.

been developed and tested in, has also during the millenniums undergone many different courses of events as sea and inland sea to what it is today as a brackish sea. The size has also varied through the millenniums. During the different time periods different kinds of sediments have been deposited on the bottom and by that created a profile structure which varies with the location (Figure 3.27). A general description about the sediment structures in the Swedish coastal waters can be found in [CKN92].

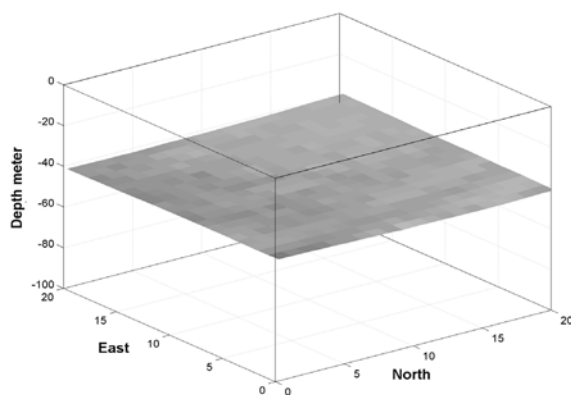


Figure 3.30: The picture shows an area, 200 meter x 200 meter, around way-point BP2C in the October 2002 sea-trial.

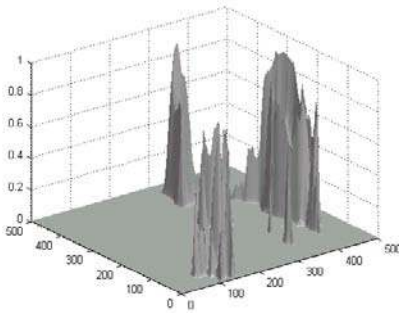


Figure 3.31: The likelihood function for 5 beam planes at way-point BP2C. Horizontal scale in 10th of meter.

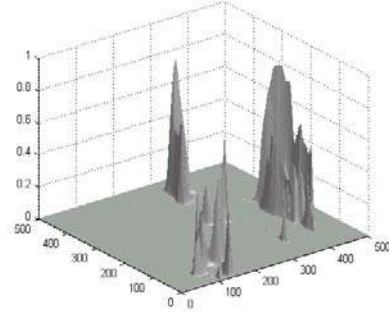


Figure 3.32: The likelihood function for 11 beam planes at way-point BP2C. Horizontal scale in 10th of meter.

For measuring the sediment profile a sub bottom profiler is used, (Figure 3.28) which transmits a sound pulse of a few kHz down into the sediments and records the returning reflections from the different layers. Figure 3.29 shows the sediment profile at the way-point BP2C in the October 2002 sea-trial (Chapter 9). The picture is created from repeated sound pulses transmitted about 2 meter apart. Each pulse gives a vertical line in the diagram. The received return signal is low-pass filtered to enhance the horizontal structure in the picture. This will also mean that the appearance of the picture will depend on the actual filter. The beamwidth of the receiver is about 19°.

With a sub bottom profiler the sediments down to several hundreds of meters can be inspected which is sufficient for the conditions in the Baltic sea even if the sediments layers in the south Baltic can be very thick. South of the so called Tornquist line the thickness is several kilometers.

When using the sediment profile for military navigation purposes one needs a transmitter with a very narrow sound beam, i.e., a parametric sonar, to keep the revealing distance short. This can also be done by lowering the transmitter to near the bottom. However, one must be aware of that the reflecting energy will spread spherically from the sediment layers.

To study the possibilities to use the sediment layers for navigation a way-point in the October sea-trial was placed in a very flat area as shown in Figure 3.30,

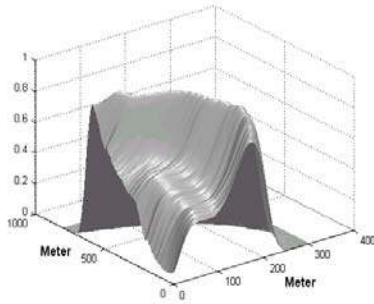


Figure 3.33: Likelihood function around the true position based on the topography at way-point BP2C.

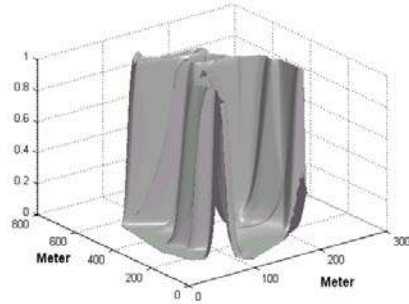


Figure 3.34: Likelihood function based on the first sediment layer at way-point BP2C.

where a unique terrain positioning based on the topography only was believed to be difficult.

The terrain positioning based on the topography gave about the same result for all five positionings at the way-point. Figures 3.31 and 3.32 show the likelihood functions when using 5 and 11 beam planes and we see from the figures that we have false peaks. The result is somewhat better for 11 beam planes which gives

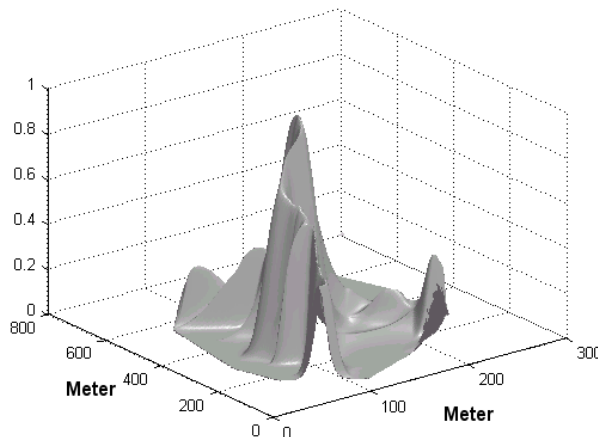


Figure 3.35: Resulting likelihood function. The position RMS-error dropped from 57 to 15 meter in the east direction and from 143 to 21 meter in the north direction.

a correct position but a large position error. If a true 3D sonar had been used the likelihood function would probably have been more distinct. The true peak is the one to the far right.

Suppose now that we use the distance to the first reflecting layer in Figure 3.29. We measure, in this example, with only one sub bottom profiler beam (normally several sub profiler beams would be used) and we assume that we have the same accuracy as before, i.e., 0.5 meter RMS. The likelihood functions for the topographic measurement and the sediment measurement can be seen in Figure 3.33 and 3.34.

As can be seen from Figures 3.27, 3.29 and, 3.34 the sediment layer has a more broken surface than the bottom surface. This is common in the archipelago areas of the Baltic. Multiplying the two likelihood functions will give the resulting likelihood function (Figure 3.35). As can be seen in this case the position error is considerably smaller and the likelihood function more distinct.

To conclude the use of sediment data to improve the positioning. Underlying sediment layers are naturally more broken than the current bottom surface and if used for positioning they can considerably improve the position accuracy in flat bottom areas.

The best positioning result in flat terrain areas will be achieved if a true 3D sonar is used due to the higher accuracy of the distance between the individual beam footprints and if the 3D sonar also have a sub bottom profiler function. A parametric sub bottom profiler will give very narrow beams at the low frequencies used and it is ideal for military use. The sonars may transmit simultaneously but the duration of the sub profiler pulse will be considerable longer.

3.9 Incorporation of side-scan data in the correlation

To improve the positioning by using side-scan pictures might not be as easy as to use sediment data. There are several reasons for that. First of all much of the information in a side-scan picture is from the light and shades from the height variations of the bottom. This information is, however, already in the usual 3D sonar data set. This information also decreases as the bottom becomes flatter, i.e., when the extra information is most needed. Secondly, the information from

the bottom texture which is not in the usual 3D sonar data set varies with time and may not be appropriate after some years. Thirdly, a side-scan picture depends very much on the side scan sonar attitude and distance to the bottom which will make the matching difficult and not as easy as for the topographic height information.

Figure 3.36 shows two side-scan pictures over the flat bottom area for which the sediment profile was used for the positioning. The pictures are taken only a few hours apart. From the pictures we see that there is some texture information that might be used. A drawback for military use is the long time it takes to collect the side scan data and the revealing distance is also much longer than for a bathymetric sonar.

3.10 Incorporation of external sound sources in the correlation

Submarines with their long flank arrays can determine the bearing to sound sources with high accuracy. In the case the position of the sound source is known this information can be used to formulate a likelihood function which in some cases, i.e., in flat areas, can improve the position estimate based on terrain positioning and the inertial navigation system.

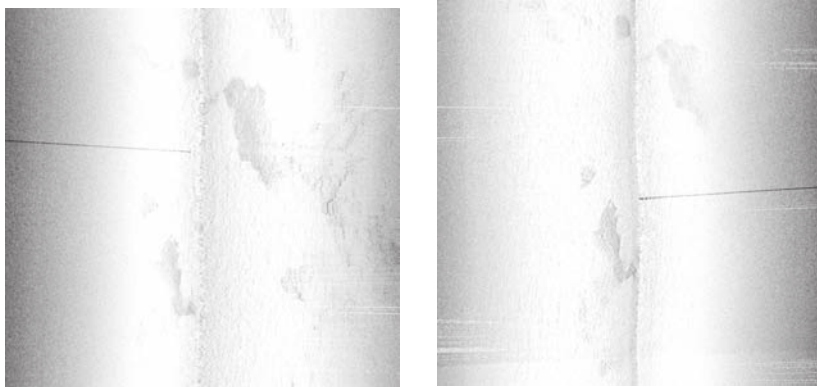


Figure 3.36: Side-scan pictures from two passages over way-point 2C.

Suppose the external sound source is at the position (x_{ss}, y_{ss}) as shown in Figures 3.37 and 3.38. While the vehicle's position is at (x, y) and the measured bearing to the sound source is θ_0 . If we assume that the measurement error is Gaussian distributed we will have the following likelihood function

$$L_{ES}(x, y) = \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \exp\left(\frac{-1}{2\sigma_\theta^2} \left(\theta_0 - \arctan \frac{x_{ss} - x}{y_{ss} - y}\right)^2\right) \quad (3.37)$$

Where σ_θ^2 is the uncertainty in the angle measurement. The Figures 3.39, 3.40 and 3.41 show an example of a prior PDF, the likelihood function and the posterior PDF. The likelihood function calculated from the external sound source will not be Gaussian in a Cartesian coordinate system.

The likelihood function along the arc segment with radius R , see Figure 3.38, and centered around the point (x, y) is

$$L_{ES}(\Delta s) \propto \exp\left(-\frac{\Delta s^2}{2R^2\sigma_\theta^2}\right) \quad (3.38)$$

Where Δs is the distance along the circular arc, i.e., $\Delta s = R\Delta\theta$ and R and $\Delta\theta$ are defined in Figure 3.38. For large R this will approximately also be the likelihood function along the tangent in the point (x, y) . The improvement of the accuracy by measuring the bearing to the external sound source is determined by calculating the Fisher information matrix

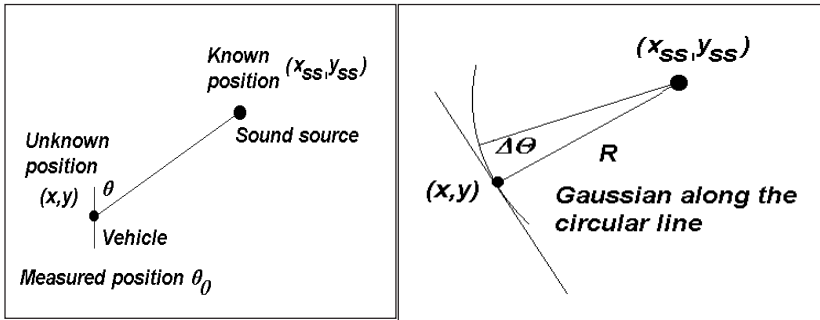


Figure 3.37: External sound source and own vehicle measuring the bearing to the source.

Figure 3.38: Definitions of variables.

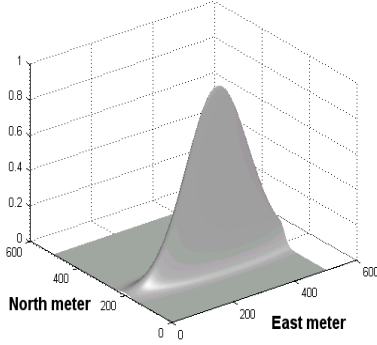


Figure 3.39: The prior position PDF.

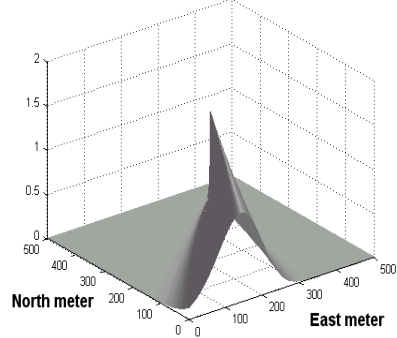


Figure 3.40: The likelihood function for the bearing measurement.

$$\begin{aligned}
 \mathbf{J} &= -E\left\{\frac{\partial^2[\log L(\mathbf{x}; \mathbf{y})]}{\partial \mathbf{x} \partial \mathbf{x}^T}\right\} = -E\left\{\frac{\partial^2[\log(L_{TP}(\mathbf{x}; \mathbf{y})L_{ES}(\mathbf{x}; \theta))]}{\partial \mathbf{x} \partial \mathbf{x}^T}\right\} = \\
 &= -E\left\{\frac{\partial^2[\log L_{TP}(\mathbf{x}; \mathbf{y}) + \log L_{ES}(\mathbf{x}; \theta)]}{\partial \mathbf{x} \partial \mathbf{x}^T}\right\} = \mathbf{J}_{TN} + \mathbf{J}_{ES}
 \end{aligned} \tag{3.39}$$

where subindex *TP* refers to terrain navigation and *ES* to the external sound source. The approximation for large distances R will give the inverse information

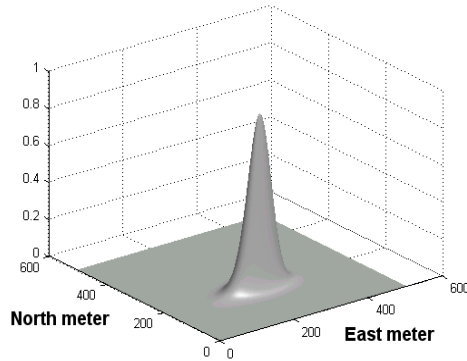


Figure 3.41: The posterior PDF for the position.

matrix in the principal directions of the likelihood function for the external sound source as

$$\mathbf{J}_{ES}^{-1} = \begin{bmatrix} \sigma_{\theta}^2 R^2 & 0 \\ 0 & \infty \end{bmatrix} \quad (3.40)$$

By repeating the procedure recursively with the same sound source an improvement in the position estimate can be achieved. Figure 3.42 illustrates the procedure. Up to the left in the figure is the PDF for position of the vehicle after the terrain measurement update in position **A**. Across to the right is the PDF for the bearing measurement and the two PDFs are fused together by multiplying the PDFs as described earlier. This fused PDF is then propagated as described in Chapter 2 to time t_2 and is indicated to the left at position **B**. At time instant t_2 a new bearing measurement is done which is fused with the previous propagated PDF and the resulting PDF is then propagated and so on. As understood this scheme can be applied to aiding from several external sound sources.

In the cases the distance to the sound source is large the PDF from the bearing measurement can be approximated to be Gaussian and again we can use a linear Kalman filter.

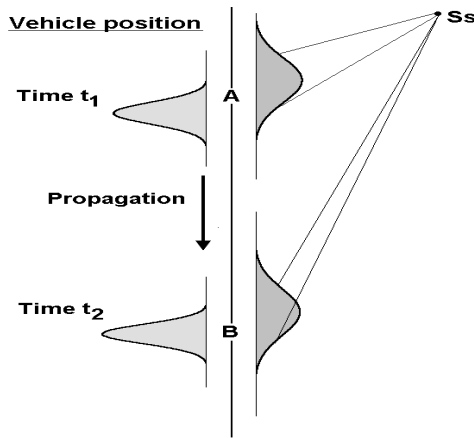


Figure 3.42: Recursively determining of bearing to external sound source.



Chapter 4

The likelihood function and the Kalman filter

4.1 Introduction

The problem with false correlation peaks has already been touched on in Chapter 3 where the Figures 3.7 and 3.8 show the large difference between using one or many measurement beams. The false peaks are due to terrain repeatability and of course the larger the measured bottom area is, the more unique the measured area will be and thereby fewer false peaks. The traditional way of selecting among the correlation peaks is to fuse with the prior position PDF or to use a validation gap. But even so, the false peaks may influence the estimated position and the position error covariance if the Bayesian approach to terrain positioning is used. It is thus important to reduce the number of false peaks and we have already seen that increasing the number of measurement beams is an effective method for achieving that.

In studying the problem of how false correlation peaks depends on the number of measuring beams two approaches are selected. In the first one we calculate how the probability for detecting a false position due to terrain repeatability depends on the number of measurement beams, i.e., the size of the correlation area. In the second approach we study how the amplitudes of the correlation

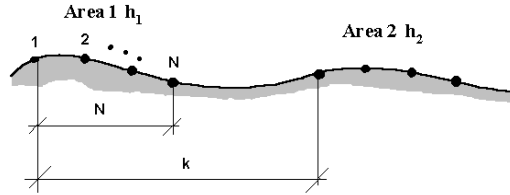


Figure 4.1: A bottom profile section.

peaks vary with the number of measurement beams. In both case it turns out that the false peaks will disappear as the number of beams increases. A method of how to test if a correlation peak corresponds the true position or not is also introduced.

Later in the chapter the shape of the likelihood function is discussed. It is shown that if a large number of measuring beams are used the likelihood function will be close to Gaussian, bias free, and coincide with the PDF for the position error if it is estimated by the maximum likelihood method.

Since the normalized likelihood function and the position error is Gaussian, a linear Kalman filter based on the normalized likelihood function can be used to estimate the position.

4.2 False peaks

4.2.1 The probability for false position

Let us for simplicity study the problem with false positions in one dimension [NY01a, NY01b]. Assume that we have a stochastic bottom profile as in Figure 4.1. The sonar, which has N measurement beams, has measured the profile in area 1 and it is found to be

$$\mathbf{y} = \mathbf{h}_1 + \mathbf{e} \quad (4.1)$$

where \mathbf{e} is the measurement error. Now we are comparing this measurement with the underwater map in area 2 which has the bottom profile \mathbf{h}_2 . In the following, we will assume that the elements in the vectors \mathbf{h}_1 and \mathbf{h}_2 are correlated Gaussian sequences. Let

$$\Delta \mathbf{h} = \mathbf{h}_1 - \mathbf{h}_2 \quad (4.2)$$

Now we want to find the probability that the absolute value of the elements in the vector $|\Delta \mathbf{h}| < \varepsilon_h$ since this means that the two bottom profiles are very close to each other. The notation $|\Delta \mathbf{h}| < \varepsilon_h$ means that the absolute value of every element in the vector $\Delta \mathbf{h}$ should be smaller than ε_h . Since we assumed \mathbf{h}_1 and \mathbf{h}_2 to be Gaussian, we have

$$p(\Delta \mathbf{h}) = \frac{1}{(2\pi)^{N/2} \sqrt{\det(\mathbf{C})}} \exp\left(-\frac{1}{2} \Delta \mathbf{h}^T \mathbf{C}^{-1} \Delta \mathbf{h}\right) \quad (4.3)$$

where \mathbf{C} is the covariance matrix of $\Delta \mathbf{h}$. The simplest way to calculate the probability that the two bottom profiles are close to each other may be to first decorrelate the vector $\Delta \mathbf{h}$ by a mode decomposition of the covariance matrix

$$\mathbf{C} = \mathbf{U} \mathbf{L} \mathbf{U}^T \quad (4.4)$$

where \mathbf{U} is the eigenvector matrix and \mathbf{L} is the eigenvalue matrix. Making the substitution $\Delta \mathbf{h} = \mathbf{U} \mathbf{y}$ makes it possible to compute an upper bound of the probability that the difference in depth profiles of the two locations is less than a small number.

$$\begin{aligned} P(|\Delta \mathbf{h}| \leq \varepsilon_h) &\leq P(|\Delta \mathbf{y}| \leq \varepsilon) \\ &= \int_{-\varepsilon}^{\varepsilon} \cdots \int_{-\varepsilon}^{\varepsilon} \prod_{i=1}^N \frac{1}{\sqrt{2\pi\lambda_i}} \exp\left(-\frac{y_i^2}{2\lambda_i}\right) dy_1 \cdots dy_N = \prod_{i=1}^N \int_{-\varepsilon}^{\varepsilon} \frac{1}{\sqrt{2\pi\lambda_i}} \exp\left(-\frac{y_i^2}{2\lambda_i}\right) dy_i \end{aligned} \quad (4.5)$$

The hypercube ε encloses the hypercube ε_h . The covariance matrix is assumed to be positive definite, i.e., all eigenvalues $\lambda_i > 0$, so each factor

$$\int_{-\varepsilon}^{\varepsilon} \frac{1}{\sqrt{2\pi\lambda_i}} \exp\left(-\frac{y_i^2}{2\lambda_i}\right) dy_i < 1 \quad (4.6)$$

if $\varepsilon > 0$. The probability that the terrain repeats itself will thus decrease exponentially with N .

It is easy to show that the above conclusion is valid also for non-Gaussian correlated sequences \mathbf{h} since any PDF can be approximated as closely as desired by a sum of Gaussian PDFs $p_m(y)$, [AS72, AN79]. Equation (4.5) can therefore be written as

$$\begin{aligned} P(|\Delta \mathbf{h}| \leq \varepsilon_h) &\leq P(|\mathbf{y}| \leq \varepsilon) \\ &= \int_{-\varepsilon}^{\varepsilon} \cdots \int_{-\varepsilon}^{\varepsilon} \sum_{m=1}^M p_m(y) d\mathbf{y} = \sum_{m=1}^M \int_{-\varepsilon}^{\varepsilon} \cdots \int_{-\varepsilon}^{\varepsilon} p_m(y) d\mathbf{y} \end{aligned} \quad (4.7)$$

For every fix M the probability will exponentially approach zero when N increases above all bounds.

Figure 4.2 illustrates the exponential decrease in probability that the terrain repeats itself, according to (4.5), for two synthetic bottom profiles generated by

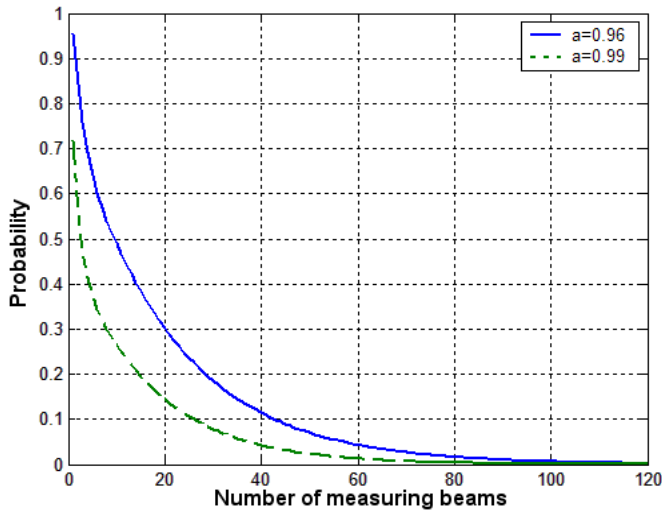


Figure 4.2: Probability that the bottom repeats itself.

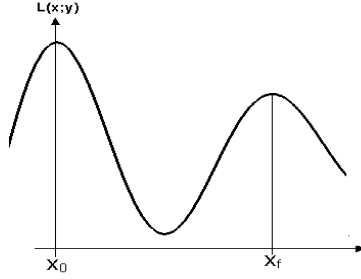


Figure 4.3: A likelihood function with a secondary peak at position x_f caused by terrain repeatability. True position at x_0 .

passing white noise through an $AR(1)$ -filter with a pole in $a = 0.96$ and $a = 0.99$, respectively.

4.2.2 The amplitude of the false peaks

In the previous section the probability of false peaks was studied. Here instead the amplitudes of the false peaks are studied. Let us formulate the ratio between the ordinates of the likelihood function in the false position x_f and in the true position x_0 when the measurement is taken at the true position x_0 , see Figure 4.3

$$\begin{aligned}
 q &= \frac{L(x_f; y(x_0))}{L(x_0; y(x_0))} = \frac{\exp(-\frac{1}{2\sigma_e^2} \sum_{k=1}^N (y_k(x_0) - h_k(x_f))^2)}{\exp(-\frac{1}{2\sigma_e^2} \sum_{k=1}^N (y_k(x_0) - h_k(x_0))^2)} \\
 &= \exp(-\frac{1}{2\sigma_e^2} \sum_{k=1}^N (y_k(x_0) - h_k(x_f))^2 - e_k^2) \\
 &= \exp(-\frac{1}{2\sigma_e^2} \sum_{k=1}^N (h_k(x_0) - h_k(x_f))^2 + 2e_k(h_k(x_0) - h_k(x_f))) \\
 &\rightarrow \exp(-\frac{1}{2\sigma_e^2} \sum_{k=1}^N (h_k(x_0) - h_k(x_f))^2)
 \end{aligned}$$

for large N .

(4.8)

Now, the only way for the ratio q not to approach zero when N tends to infinity

is that there is a symmetry in the terrain which extends into infinity. This is not the case for a natural terrain and we conclude that secondary peaks will disappear as the number of beams increase.

An increase in the number of measurement beams not only decreases the number of false likelihood positions but it also makes the peak of the likelihood function much sharper that is the position accuracy increases as will be shown in Chapter 5.

We conclude from the above that false positions will not be a problem if a large number of measuring beams is used. This has also been the experience from our sea trials. When it occasionally still happens in flat bottom areas the simplest way to eliminate the problem is to increase the size of the total sonar footprint since the probability of a false position decays exponentially with the size of that area.

4.2.3 False peaks, hypothesis testing

In Chapter 3, Figure 3.15, a positioning example was shown where multiple likelihood peaks were assumed to have occurred. Multiple likelihood peaks appear if we have too few measurement beams in relation to the height variations in the terrain. We can distinguish between two cases. The first case occurs if we are using very few beams in a rough terrain and the second case occurs if we have many measurement beams but the terrain is very flat. In the last case the likelihood function will be close to Gaussian.

Assume that we have two likelihood peaks which are close to each other in magnitude but the one with index 1 is slightly higher than the one with index 2. Since peak no. 1 is the highest we will assume that it shows the true position. Our question is now if this is just due to the randomness of the measurement error or is this position really the true position. The correlation function at the true position and at the false position is respectively

$$T(\mathbf{x}_1) = (\mathbf{y} - \mathbf{h}_1)^T (\mathbf{y} - \mathbf{h}_1) = \mathbf{e}^T \mathbf{e} \quad (4.9)$$

and

$$\begin{aligned} T(\mathbf{x}_2) &= (\mathbf{y} - \mathbf{h}_2)^T (\mathbf{y} - \mathbf{h}_2) = (\mathbf{y} - \mathbf{h}_1 + \Delta \mathbf{h})^T (\mathbf{y} - \mathbf{h}_1 + \Delta \mathbf{h}) \\ &= \mathbf{e}^T \mathbf{e} + \Delta \mathbf{h}^T \Delta \mathbf{h} + 2\Delta \mathbf{h}^T \mathbf{e} \end{aligned} \quad (4.10)$$

Now, define the stochastic variable ξ as

$$\xi = \Delta \mathbf{h}^T \Delta \mathbf{h} + 2\Delta \mathbf{h}^T \mathbf{e} \quad (4.11)$$

and it is clear that depending on the sign of ξ either of the positions will have the highest magnitude of the likelihood function. Since we know the positions, we also know the vector $\Delta \mathbf{h}$. For large N the stochastic variable ξ will mostly be positive which means that position 2 will have a larger correlation sum and it is unlikely that it is the true position. When $\xi < 0$ position 2 will have the smallest correlation sum and by that be the most likely positions, i.e. the hypothesis fails. Thus, when

$$\Delta \mathbf{h}^T \Delta \mathbf{h} < -2\Delta \mathbf{h}^T \mathbf{e} \quad (4.12)$$

position 2 will be the likely to some extent. The stochastic variable to the right is a sum of Gaussian zero mean variables and thus a zero mean Gaussian variable. The variance of the individual variable in the sum is $4\Delta h_i^2 \sigma_e^2$ since we have assumed the measurement error variance in all beams to be σ_e^2 . The variance for the right hand side in (4.12) will be $4\sigma_e^2 \sum_{i=1}^N (\Delta h_i)^2$. We can illustrate expression (4.12) by Figure 4.4 where the shaded area corresponds to the probability

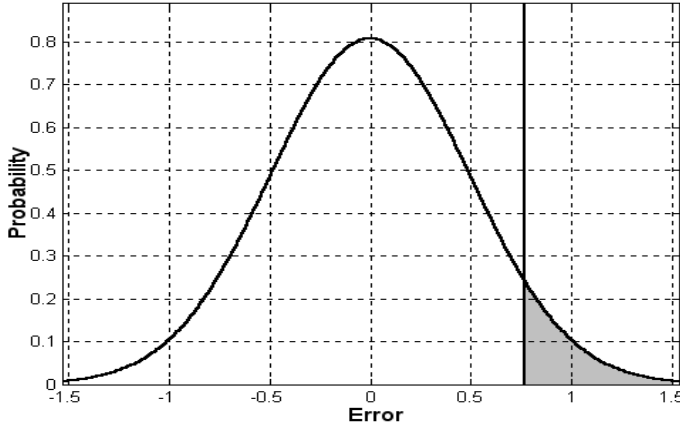


Figure 4.4: The PDF for the right hand side of (4.12). The vertical line corresponds to the left hand side of (4.12). The shaded area corresponds to the probability that position 2 is the true position.

that position 2 is the true position. When the difference in terrain profiles between the two positions will tend to zero the probability will tend to 0.5.

The Gaussian assumption about the measurement error is very convenient to use but the real PDF for the measurement error will have the tails cut. That means that it is possible to tell if the difference between the measured profile and the map profile can be explained by the measurement error or not. If it can not be explained the position is truly false. Since the right hand side in (4.12) is a sum of scaled stochastic variables from the same zero mean measurement error PDF the criterion is as easy to use as in the case of the Gaussian assumption of the measurement error. By applying the criterion (4.12) to all points in the map and setting the likelihood function to zero for the positions where the criterion is not fulfilled the map can in some sense be cleaned from false positions.

4.3 The shape of the likelihood function

We have seen in the previous chapter, Figure 3.17 that the likelihood function resembles a Gaussian curve. It is also true that the likelihood function converges under mild conditions towards a Gaussian curve when the number of measuring beams increases, see for example [CA88, KH49, PE95, SESI93, VA98]. This can also be shown plausible by a Taylor expansion of the likelihood function around the point of its maximum and by assuming certain conditions for the derivatives of the likelihood function. Let us look at the one-dimensional case and let the Taylor expansion be around $x = 0$

$$L(x; y) = C \cdot \exp\left(-\frac{1}{2\sigma_e^2} \left[\sum_{k=1}^N \left(\frac{\partial h_k}{\partial x} \right)^2 x^2 + \sum_{k=1}^N \left(\frac{\partial^2 h_k}{\partial x^2} \right)^2 x^4 H(x) \right] \right) \quad (4.13)$$

where the derivatives are taken at $x = 0$. The function $H(x)$ is the remainder term and is assumed to be bounded for all x . The constant C is a normalizing constant making $\int_{-\infty}^{\infty} L(x; y) dx = 1$. When developing the expansion the measurement error has been assumed to be a white sequence and the terms that tend to zero when N increases have been set to zero. We can also write the equation as

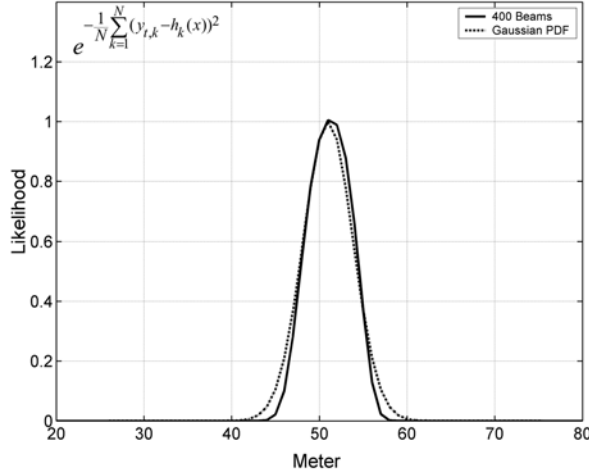


Figure 4.5: Comparison between the likelihood function and the Gaussian PDF. Data from WP3 in the October 2002 sea-trial.

$$L(x; y) = C \cdot \exp\left(-\frac{1}{2\sigma_e^2 / \sum_{k=1}^N \left(\frac{\partial h_k}{\partial x}\right)^2} x^2 \left[1 + \sum_{k=1}^N \left(\frac{\partial^2 h_k}{\partial x^2}\right)^2 / \sum_{k=1}^N \left(\frac{\partial h_k}{\partial x}\right)^2 x^2 H(x)\right]\right) \quad (4.14)$$

For small x we can neglect the second term within the brackets in comparison to

1 assuming that $\sum_{k=1}^N \left(\frac{\partial^2 h_k}{\partial x^2}\right)^2 / \sum_{k=1}^N \left(\frac{\partial h_k}{\partial x}\right)^2$ is bounded. Left is a Gaussian curve with the variance

$$\sigma^2 = \sigma_e^2 / \sum_{k=1}^N \left(\frac{\partial h_k}{\partial x}\right)^2 \quad (4.15)$$

which corresponds exactly to the variance given by the CRLB in Chapter 6. This is important since it means that the normalized likelihood function is the same as the position error PDF. From Figure 4.5 which is from a sea-trial it can be seen that the likelihood function agrees well with a Gaussian curve. Another proof of the asymptotic convergence towards a Gaussian curve is given below.

4.3.1 The convergence of the likelihood curve

We will assume that the terrain profile $\mathbf{h}(\mathbf{x})$ is continuous with continuous higher derivatives. When measuring the terrain profile we will have a certain error, \mathbf{e} , which we will assume Gaussian distributed. The number of measurement beams is N . The measured profile can thus be written $\mathbf{y}(\mathbf{x}_0) = \mathbf{h}(\mathbf{x}_0) + \mathbf{e}$, where \mathbf{x}_0 is the true position.

The likelihood function can in the one dimensional case be written as

$$L_N(x; y(x_0)) = C \cdot \exp \left[-\frac{1}{2\sigma_e^2} \sum_{i=1}^N [y_i(x_0) - h_i(x)]^2 \right] \quad (4.16)$$

where C is a constant. We know that $L_N(x; y(x_0))$ converges to a Dirac pulse when $N \rightarrow \infty$ but we want to show that $L_N(x; y(x_0))$ also tends to a Gaussian curve when $N \rightarrow \infty$. Besides our previous assumptions about the measurement noise we assume that for a given fix N the likelihood function and its derivatives to be continuous so a Taylor expansion of the likelihood function around $x_0=0$ is permissible. Accordingly

$$L_N(x; y(x_0)) = C \cdot \exp \left[-\frac{1}{2\sigma_e^2} \left[\sum_{i=1}^N \left[\frac{\partial h_i(x_0)}{\partial x} \right]^2 x^2 + \frac{1}{4} \sum_{i=1}^N \left[\frac{\partial^2 h_i(x_0)}{\partial^2 x} \right]^2 x^4 + \dots \right. \right. \\ \left. \left. + \frac{1}{(k!)^2} \sum_{i=1}^N \left[\frac{\partial^k h_i(x_0)}{\partial^k x} \right]^2 x^{2k} + \dots \right] \right] \quad (4.17)$$

or

$$L_N(x; y(x_0)) = C \cdot \exp \left[-\frac{x^2}{2\sigma_e^2 / \sum_{i=1}^N \left[\frac{\partial h_i(x_0)}{\partial x} \right]^2} \left[1 + \frac{1}{4} \sum_{i=1}^N \left[\frac{\partial^2 h_i(x_0)}{\partial^2 x} \right]^2 / \sum_{i=1}^N \left[\frac{\partial h_i(x_0)}{\partial x} \right]^2 x^2 + \dots \right. \right. \\ \left. \left. + \frac{1}{(k!)^2} \sum_{i=1}^N \left[\frac{\partial^k h_i(x_0)}{\partial^k x} \right]^2 / \sum_{i=1}^N \left[\frac{\partial h_i(x_0)}{\partial x} \right]^2 x^{2k-2} + \dots \right] \right] \quad (4.18)$$

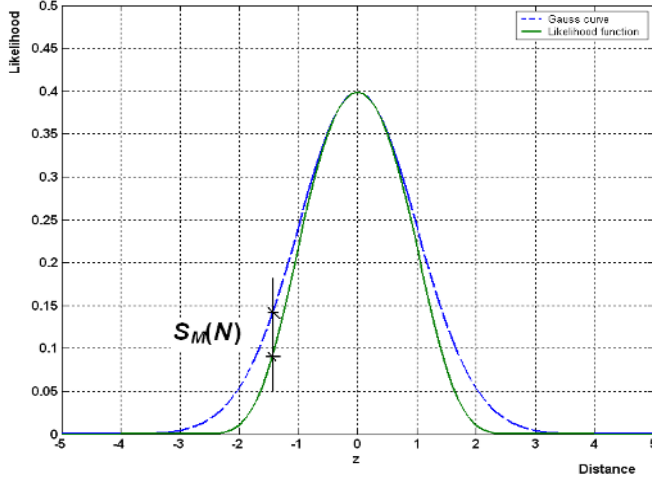


Figure 4.6: Illustration of the difference between the likelihood function and the Gaussian function.

We recognize that the likelihood function $L_N(x; y(x_0))$ is not uniformly convergent and we therefore introduce the variable z through the transformation $x = z\sigma_N$ where

$$\sigma_N^2 = \sigma_e^2 / \sum_{i=1}^N \left[\frac{\partial h_i}{\partial x} \right]^2 \quad (4.19)$$

which will give us a uniformly convergent series, see Figure 4.6. The likelihood function can now be written

$$\begin{aligned} L_N(z; y(x_0)) &= C \cdot \exp \left[-\frac{z^2}{2} \left[1 + \frac{\sigma_e^2}{4} \sum_{i=1}^N \left[\frac{\partial h_i^2(x_0)}{\partial^2 x} \right] / \left[\sum_{i=1}^N \left[\frac{\partial h_i(x_0)}{\partial x} \right]^2 \right]^2 z^2 + \dots \right. \right. \\ &\quad \left. \left. + \frac{\sigma_e^{2k}}{(k!)^2} \sum_{i=1}^N \left[\frac{\partial h_i^k(x_0)}{\partial^k x} \right] / \left[\sum_{i=1}^N \left[\frac{\partial h_i(x_0)}{\partial x} \right]^2 \right]^{2k-1} z^{2k-2} + \dots \right] \right] \\ &= C \cdot \exp \left[-\frac{z^2}{2} [1 + S_\infty(N)] \right] \end{aligned} \quad (4.20)$$

where we introduced the partial sum $S_M(N)$ of M terms as

$$S_M(N) = \sum_{k=2}^M \frac{\sigma_e^{2k}}{(k!)^2} \sum_{i=1}^N \left[\frac{\partial h_i^k(x_0)}{\partial^k x} \right]^2 / \left[\sum_{i=1}^N \left[\frac{\partial h_i(x_0)}{\partial x} \right]^2 \right]^{2k-1} z^{2k-2} \quad (4.21)$$

Now, since every term in the sum is strictly monotonically decreasing with M and can be made as small as desired by choosing a sufficiently large N , the partial sum $S_M(N)$ can be made arbitrarily small for any choice of M . This means that $S_M(N) \rightarrow 0$ when $N \rightarrow \infty$ independent of M , [RÄWE95, p. 181], and that the likelihood function converges to a Gaussian curve in the transformed space. Since the transformation does not affect the vertical scale, this will also be true in the non-transformed space.

4.3.2 The convergence of the N -normalized likelihood curve

It is often of interest to look at a likelihood expression where the exponent has been normalized with the number of beams since the true likelihood function in most cases will just be a very narrow peak when plotted due to the rapid variance decrease with the number of beams. The N -normalized likelihood function is

$$L_{NNorm}(x; y(x_0)) = \exp \left[-\frac{1}{2\sigma_e^2} \frac{1}{N} \sum_{i=1}^N [y_i(x_0) - h_i(x)]^2 \right] \quad (4.22)$$

Since the N -normalized likelihood function is just the N th root of the true likelihood function it will also be a Gaussian PDF if it is normalized to have the volume one. The expected value of the exponent in (4.22) in the true position is $-1/2$. Often one plots the N -normalized likelihood function under the assumption of $\sigma_e^2 = 1$. The N -normalized likelihood functions for different N in a certain position will in most cases be close to each other.

4.3.3 The convergence of the likelihood curve in frequency space

It is also possible to come to the same conclusion as above by assuming that the bottom was created by a band limited stationary stochastic process with a certain autocorrelation function. Let us look at the one-dimensional case and let us

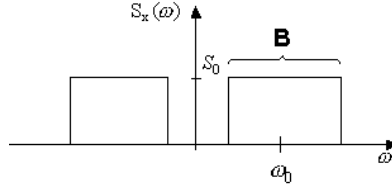


Figure 4.7: Power spectrum for a band-limited process.

assume that the bottom topography was created by a stationary stochastic process $h(x)$. Our measurement of the vector $\mathbf{h}(x)$ is \mathbf{y} , i.e. $\mathbf{y}_t = \mathbf{h}(x_t) + \mathbf{e}_t$ where \mathbf{e}_t is the Gaussian measurement error. The dimension of the vectors is N .

We have

$$\begin{aligned}
 E\left\{\frac{T(x)}{2N\sigma_e^2}\right\} &= E\left\{\frac{1}{2N\sigma_e^2} \sum_{k=1}^N (y_{t,k} - h_k(x))^2\right\} \\
 &= \frac{1}{2\sigma_e^2} [2\sigma_h^2 + \sigma_e^2 - 2r_h(\tau)] \\
 &= \kappa(1 - \bar{r}_h(\tau)) + \frac{1}{2}
 \end{aligned} \tag{4.23}$$

where the bottom signal to noise ratio is $\kappa = \sigma_h^2 / \sigma_e^2$, $\tau = x_t - x$ and $\bar{r}_h(\cdot)$ is the normalized autocorrelation function for the bottom profile.

As an example assume that the bottom is created by a band limited stochastic process, see Figure 4.7. The band limited assumption is a very mild assumption and always fulfilled for real terrains and it may therefore be proper to say that this discussion is valid for any physical terrain. The autocorrelation function is

$$r_h(\tau) = \frac{4S_0}{\tau} \sin\left(\frac{B}{2}\tau\right) \cos(\omega_0\tau) \tag{4.24}$$

After Taylor expansion we have the normalized autocorrelation function

$$\bar{r}_h(\tau) = 1 - \left[\frac{B^2}{24} + \frac{\omega_0^2}{2} \right] \tau^2 + O(\tau^4) \tag{4.25}$$

The remainder term is small in the neighbourhood of $\tau=0$ which is the surrounding we are interested in. Disregarding the remainder term we have

$$E\left\{\frac{T(x)}{2N\sigma_e^2}\right\} = \kappa \left[\frac{B^2}{24} + \frac{\omega_0^2}{2} \right] \tau^2 + \frac{1}{2} = \frac{\tau^2}{\sigma_{BL}^2} + \frac{1}{2} \quad (4.26)$$

where

$$\sigma_{BL}^2 = \frac{1}{\kappa \left[\frac{B^2}{24} + \frac{\omega_0^2}{2} \right]} \quad (4.27)$$

A well known approximation formula which is valid with good accuracy in our case is

$$f(x) = E\left\{\exp\left(-\frac{T(x)}{2N\sigma_e^2}\right)\right\} \approx \exp\left(-E\left\{\frac{T(x)}{2N\sigma_e^2}\right\}\right) \propto \exp\left(-\frac{\tau^2}{\sigma_{BL}^2}\right) \quad (4.28)$$

In Figure 4.8 we compare this approximation with a Gaussian curve and we see that the larger the κ -value is, the more the likelihood curve tends towards a Gaussian function. In areas of interest for terrain navigation, the terrain variations

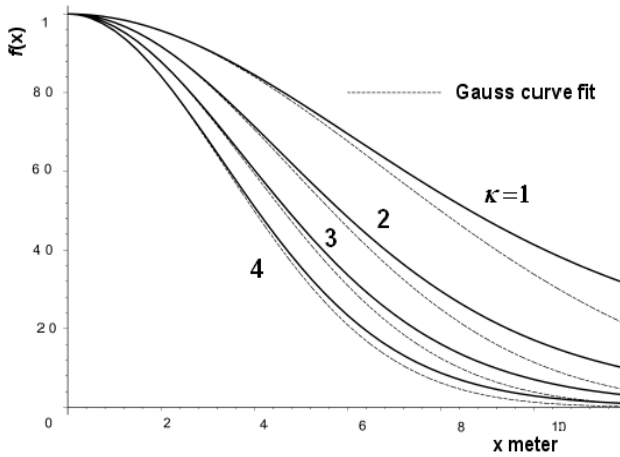


Figure 4.8: Comparison between the expected likelihood functions for a stochastic band-limited bottom profile and Gaussian curves.

are usually much larger than the ones in Figure 4.8. We observe here that large κ values also mean large bottom gradients. Equation (4.27) can be used as an indicator of the suitability for terrain positioning.

4.4 The linear Kalman filter

The linear Kalman filter equation can be derived in many ways; Stochastic differential equations as in [KL01], least squares as in [STBO97], minimum mean squares as in [BRHW97], by the orthogonality principle as in [KSH00] to mention a few.

An approach that is suitable in the context of this thesis may be an approach based on Bayes theorem [BRHW97]. Let the vehicle position and our measurement be given by the equations

$$\mathbf{x}(k+1) = \mathbf{F}(k)\mathbf{x}(k) + \mathbf{v}(k) \quad (4.29)$$

$$\mathbf{y}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{e}(k) \quad (4.30)$$

where $\mathbf{v}(k)$ and $\mathbf{e}(k)$ are white, independent, Gaussian sequences with covariances $\mathbf{Q}(k)$ and $\mathbf{R}(k)$.

4.4.1. The conditional mean

Let $\hat{\mathbf{x}} = \mathbf{g}(\mathbf{y}(k), \mathbf{y}(k-1), \dots, \mathbf{y}(0)) = \mathbf{g}(\mathbf{y}^*)$ be our estimate of the position based on all available measurements. We can have different criteria for the optimality of the estimate. One criteria is the minimum mean square estimate, that is, the estimator that makes the mean square error (mse) of the estimate as small as possible. The expression for the mse is

$$\begin{aligned} E\{(\mathbf{x} - \hat{\mathbf{x}})^T (\mathbf{x} - \hat{\mathbf{x}}) | \mathbf{y}^*\} &= E\{\mathbf{x}^T \mathbf{x} - \mathbf{x}^T \hat{\mathbf{x}} - \hat{\mathbf{x}}^T \mathbf{x} + \hat{\mathbf{x}}^T \hat{\mathbf{x}} | \mathbf{y}^*\} = \\ E\{\mathbf{x}^T \mathbf{x} | \mathbf{y}^*\} &+ E\{\mathbf{x}^T | \mathbf{y}^*\} \hat{\mathbf{x}} + \hat{\mathbf{x}}^T E\{\mathbf{x} | \mathbf{y}^*\} + \hat{\mathbf{x}}^T \hat{\mathbf{x}} \end{aligned} \quad (4.31)$$

since $\hat{\mathbf{x}}$ is a deterministic function of \mathbf{y}^* . Completing the square will give

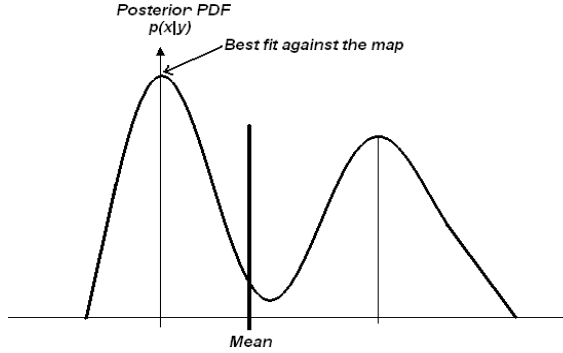


Figure 4.9: Bimodal posterior position PDF.

$$\begin{aligned}
 E\{(\mathbf{x} - \hat{\mathbf{x}})^T (\mathbf{x} - \hat{\mathbf{x}}) | \mathbf{y}^*\} = \\
 E\{\mathbf{x}^T \mathbf{x} | \mathbf{y}^*\} + [\hat{\mathbf{x}} - E\{\mathbf{x} | \mathbf{y}^*\}]^T [\hat{\mathbf{x}} - E\{\mathbf{x} | \mathbf{y}^*\}] - E\{\mathbf{x}^T | \mathbf{y}^*\} E\{\mathbf{x} | \mathbf{y}^*\} \quad (4.32)
 \end{aligned}$$

which will have a minimum for

$$\hat{\mathbf{x}} = E\{\mathbf{x} | \mathbf{y}^*\} \quad (4.33)$$

Thus, the conditional mean is the estimator that minimizes the mse of the estimate in both linear and nonlinear estimation problems. Still, the conditional mean may not be the best criteria for deciding upon the position in terrain navigation. Suppose we have a posterior position PDF as indicated in Figure 4.9 then the mean would give a position between the peaks which is an unlikely position. The only case when the maximum a posteriori or maximum likelihood estimate coincide with the minimum mean square estimate is when the posterior PDF is symmetric and unimodal [JA70, TS02].

4.4.2. The linear Kalman filter

We will assume that the starting position of the vehicle is Gaussian distributed $\mathcal{N}(\mathbf{x}_0, \mathbf{P}_0)$. Then the position, just before measurement update, will be

$$\mathbf{x}_0^- = \mathbf{F}_k \mathbf{x}_0 \quad (4.34)$$

where we have simplified the time step notation and will thus also be Gaussian distributed. For the time step k we will have

$$p(\mathbf{x}) \sim \mathcal{N}(\mathbf{x}_k^-, \mathbf{P}_k^-) = \mathcal{N}(\mathbf{F}_k \mathbf{x}_k, \mathbf{F}_k \mathbf{P}_{k-1} \mathbf{F}_k^T + \mathbf{Q}_k) \quad (4.35)$$

Our measurement \mathbf{y} is also assumed to be Gaussian distributed. From (4.30) we have for the prediction

$$p(\mathbf{y}) \sim \mathcal{N}(\mathbf{H}_k \mathbf{x}_k^-, \mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k) \quad (4.36)$$

The conditional density of \mathbf{y}_k given \mathbf{x}_k is

$$p(\mathbf{y}_k | \mathbf{x}_k) \sim \mathcal{N}(\mathbf{H}_k \mathbf{x}_k, \mathbf{R}_k) \quad (4.37)$$

Now we apply the theorem of Bayes to the Gaussian densities (4.35), (4.37)

$$p(\mathbf{x} | \mathbf{y}) = \frac{p(\mathbf{y} | \mathbf{x}) p(\mathbf{x})}{p(\mathbf{y})} \quad (4.38)$$

The calculation of (4.38) can be simplified by observing that if two Gaussian PDFs, $\mathcal{N}(\boldsymbol{\mu}_1, \mathbf{C}_1)$ and $\mathcal{N}(\boldsymbol{\mu}_2, \mathbf{C}_2)$ are multiplied by each other the result, after normalization is $\mathcal{N}([\mathbf{C}_1^{-1} + \mathbf{C}_2^{-1}]^{-1} [\mathbf{C}_1^{-1} \boldsymbol{\mu}_1 + \mathbf{C}_2^{-1} \boldsymbol{\mu}_2], [\mathbf{C}_1^{-1} + \mathbf{C}_2^{-1}]^{-1})$, i.e., the covariances are added together by the parallel resistor formula and the new mean is a weighted, by the covariances, sum of the individual means.

Otherwise, inserting (4.35), (4.36) and (4.37) into (4.38) will after some calculations, give

$$p(\mathbf{x} | \mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}_x, \mathbf{C}_x) \quad (4.39)$$

where

$$\begin{aligned} \boldsymbol{\mu}_x &= \mathbf{x}_k^- + \mathbf{P}_k^- \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k)^{-1} (\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^-) \\ &= \mathbf{x}_k^- + \mathbf{P}_k^- \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k)^{-1} (\mathbf{y}_k - \mathbf{y}_k^-) \end{aligned} \quad (4.40)$$

and

$$\mathbf{C}_x = \left[\left[\mathbf{P}_k^- \right]^{-1} + \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k \right]^{-1} \quad (4.41)$$

We can also view the mean in the measurement space

$$\boldsymbol{\mu}_y = \mathbf{y}_k^- + \mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k)^{-1} (\mathbf{y}_k - \mathbf{y}_k^-) \quad (4.42)$$

from which we clearly see that the new position is a weighted mean of the propagated position in measurement space and the measured position. A maybe simpler approach would be to directly transform the measurement uncertainty covariance to state space and add it to the \mathbf{P}_k^- by the parallel resistor formula.

Assume that a Gaussian stochastic vector \mathbf{x} has the density function

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \det(\mathbf{C}_x)^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m}_x)^T \mathbf{C}_x^{-1} (\mathbf{x} - \mathbf{m}_x)\right) \quad (4.43)$$

Let now a transformation be $\mathbf{y} = \mathbf{A}\mathbf{x}$. It is then easy to show that the new distribution is also a Gaussian density whos mean and covariance are given by

$$\mathbf{m}_y = \mathbf{A}\mathbf{m}_x \quad (4.44)$$

$$\mathbf{C}_y = \mathbf{A}\mathbf{C}_x\mathbf{A}^T \quad (4.45)$$

We note that \mathbf{C}_y does not depend on \mathbf{x} . Since the new density is Gaussian we can directly write the PDF

$$p(\mathbf{y}) = \frac{1}{(2\pi)^{m/2} \det(\mathbf{C}_y)^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{m}_y)^T \mathbf{C}_y^{-1} (\mathbf{y} - \mathbf{m}_y)\right) \quad (4.46)$$

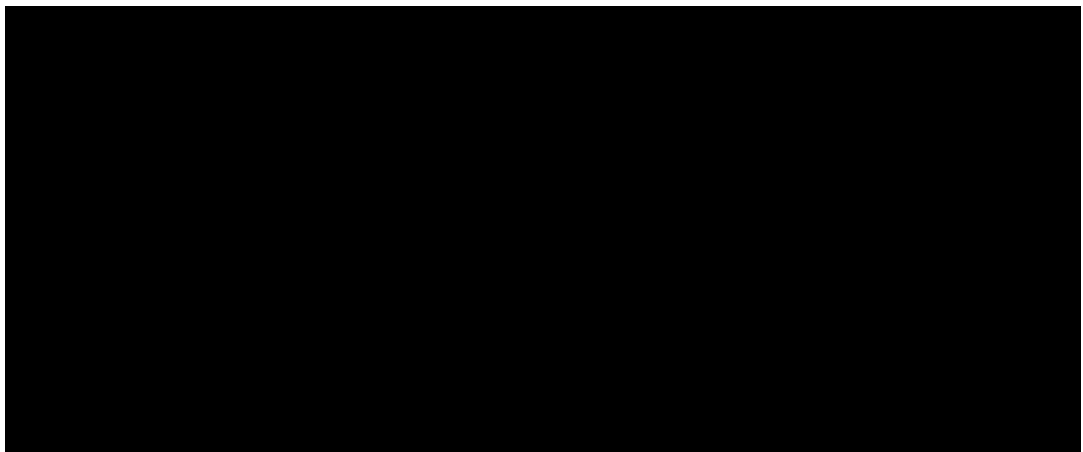
If we now partly re-substitute \mathbf{x} we will have

$$\begin{aligned} p(\mathbf{x}) &= \frac{1}{(2\pi)^{m/2} \det(\mathbf{C}_y)^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{m}_x)^T \mathbf{C}_y^{-1} (\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{m}_x)\right) \\ &= \frac{1}{(2\pi)^{m/2} \det(\mathbf{C}_y)^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m}_x)^T \mathbf{A}^T \mathbf{C}_y^{-1} \mathbf{A} (\mathbf{x} - \mathbf{m}_x)\right) \end{aligned} \quad (4.47)$$

The inverted covariance expression in (4.47) does not depend on \mathbf{x} . We note that $\mathbf{A}^T \mathbf{C}_y^{-1} \mathbf{A}$ is similar to

$$\mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k \quad (4.48)$$

in (4.41) which represents the measurement noise and this is how the measurement noise maps into the state space for a given transformation \mathbf{H}_k .





Chapter 5

The Cramér-Rao lower bound

5.1. Introduction

The Cramér-Rao lower bound (CRLB) gives a lower limit of the accuracy that can be achieved in estimation of parameters and plays an important role in establishing the accuracy of the navigation method presented in this thesis. The chapter will therefore start with a general review of the CRLB followed by an application to the terrain navigation problem.

We are using estimators to estimate the characteristics of a stochastic variable as the mean, the variance or some other quantity characterizing the variable. We make the estimation by taking samples of the stochastic variable and from the samples we calculate a value of the quantity we want to estimate. In most cases the estimator is just a formula giving us a value, an estimate, of the characteristic of the variable when we plug in the sample values.

When we take a new sample and calculate a new estimate we usually get a different value and by that we understand that our estimate also is a stochastic variable. If the estimator gives a large spread in its values, when we successively use it, the estimator has a large variance. Of course we want an estimator with a small variance and a mean close to the true value so we can be sure that the

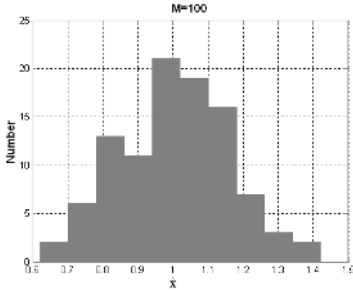


Figure 5.1: Histogram for the estimator, 100 trials.

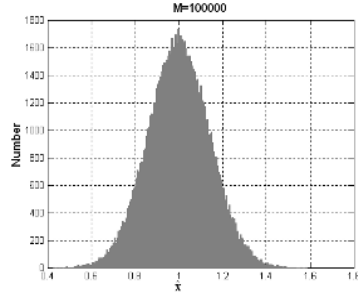


Figure 5.2: Histogram for the estimator, 100 000 trials.

estimate is close to the true value for the characteristic we are estimating. Thus the estimate shall be unbiased and have a small variance.

The CRLB will give us some guidance in evaluating the performance of an estimator since it is the lowest variance an unbiased estimator can have if we exclude so called super-efficient estimators which are of no practical importance in our case [SO16]. If the performance of our estimator is unbiased and is close to the CRLB our estimator is a good one, if we are far from the CRLB the estimator is poor and we may perhaps find a better estimator. If the variance of our unbiased estimates attains the CRLB the estimator is called efficient. An estimator is asymptotically efficient if it reaches the CRLB and is/or becomes unbiased when the sample size tends to infinity. This will be the case for our estimation of the position from terrain matching.

If we estimate a quantity x from the set of independent samples y_1, y_2, \dots, y_N by the formula $\hat{x} = \hat{x}(y_1, y_2, \dots, y_N)$ repeatably based on new sample sets each time we can arrange the estimates in a histogram as in Figure 5.1 and 5.2. We will typically find that the standard deviation of the histogram will become smaller if the estimator uses a larger sample set, i.e., a larger N .

If we increase the number of estimation trials to infinity the normalized histogram will converge to the probability density function for the estimator and the variance of the histogram is the variance of our estimator. If our estimates have no bias and this variance equals the CRLB we know we can not find an estimator with a

smaller variance. If our variance is greater than the CRLB we might find, but it is not for sure, an estimator formula which will give us an estimate with smaller variance.

There are several ways to derive the CRLB expression, [RA45, CR46, LJ99, NA69, KA93, VT68, SC91] to mention only a few places in the engineering literature where the derivation can be found. It is common nowadays in the derivation to assume that the estimator is unbiased since it leads to a short derivation. This is inline with the derivation presented in [RA45]. The derivation of the CRLB which will be reviewed here follows instead the derivation in [CR46], which do not assume that the estimator is unbiased (bias will frequently occur in terrain navigation). Some comments and illustrations have been added to, hopefully, make the derivation easier to follow. The Cramér derivation also discusses an interesting invariance property for transformations of the sample PDF. The review of the Cramér derivation may also be interesting from a historical point of view.

5.2 The CRLB

5.2.1 The scalar CRLB

The derivation starts by finding an expression of the mean square error based on an assumed known estimator $\hat{x} = \hat{x}(y_1, y_2, \dots, y_N)$ of x . Our observation set of x is y_1, y_2, \dots, y_N which can be thought of as been generated by

$$y_i = f(x, e_i) \quad (5.1)$$

where e_i is an error with some known PDF. This means that the joint PDF $p(y_1, y_2, \dots, y_N | x)$ is assumed known.

We will assume that our estimator has a bias defined by the expression

$$E\{\hat{x} | x\} \triangleq \psi(x) = x + b(x) \quad (5.2)$$

There can be several reasons behind the bias $b(x)$ but in terrain navigation such a position bias will often occur if we have a fix depth measurement error. Due to the bottom slope this will cause a position error that varies with the position. The mean of the estimator \hat{x} is by definition

$$\psi(x) = \int_{\hat{x}} \hat{x} p(\hat{x} | x) d\hat{x} \quad (5.3)$$

where $p(\hat{x} | x)$ is the PDF for our estimator (cf. Figure 5.2). Since $p(\hat{x} | x)$ is a PDF we have

$$\int_{\hat{x}} p(\hat{x} | x) d\hat{x} = 1 \quad (5.4)$$

The left hand side of this expression is a function of our true value x . Assuming it is admissible to differentiate under the integral sign we have

$$\int_{\hat{x}} \frac{\partial p(\hat{x} | x)}{\partial x} d\hat{x} = 0 \quad (5.5)$$

We now multiply both sides with the constant x and, since the integration is with respect to \hat{x} , we can move x under the integral sign

$$\int_{\hat{x}} x \frac{\partial p(\hat{x} | x)}{\partial x} d\hat{x} = 0 \quad (5.6)$$

Now, differentiating (5.3) with respect to x gives

$$\frac{\partial \psi(x)}{\partial x} = \int_{\hat{x}} \hat{x} \frac{\partial p(\hat{x} | x)}{\partial x} d\hat{x} \quad (5.7)$$

If we now subtract (5.6) from (5.7) we have

$$\frac{\partial \psi(x)}{\partial x} = \int_{\hat{x}} (\hat{x} - x) \frac{\partial p(\hat{x} | x)}{\partial x} d\hat{x} \quad (5.8)$$

We want to use Schwartz inequality for integrals and we therefore rewrite this expression as

$$\frac{\partial \psi(x)}{\partial x} = \int_{\hat{x}} (\hat{x} - x) \sqrt{p(\hat{x} | x)} \frac{1}{\sqrt{p(\hat{x} | x)}} \frac{\partial p(\hat{x} | x)}{\partial x} d\hat{x} \quad (5.9)$$

If we now take the absolute value of both sides and then apply the Schwartz inequality we have

$$\left[\frac{\partial \psi(x)}{\partial x} \right]^2 \leq \underbrace{\left[\int_{\hat{x}} (\hat{x} - x)^2 p(\hat{x} | x) d\hat{x} \right]}_{E\{(\hat{x} - x)^2\}} \left[\int_{\hat{x}} \left[\frac{\partial p(\hat{x} | x)}{\partial x} \right]^2 \frac{1}{p(\hat{x} | x)} d\hat{x} \right] \quad (5.10)$$

or

$$E\{(\hat{x} - x)^2\} \geq \frac{\left[\frac{\partial \psi(x)}{\partial x} \right]^2}{\int_{\hat{x}} \left[\frac{\partial p(\hat{x} | x)}{\partial x} \right]^2 \frac{1}{p(\hat{x} | x)} d\hat{x}} \quad (5.11)$$

By that we have a lower bound on the mean square error for our actual choice of estimator.

Next we will look at a transformation invariance property of the denominator in (5.11). Let $p(y | x)$ be the PDF of our observation vector $y = [y_1, \dots, y_N]^T$ and let us transform our observation vector by the one to one transformation $z = z(y)$, $y = f(z)$ and $p(z | x)$ is the PDF of the vector $z = [z_1, \dots, z_N]^T$.

We then have

$$p(z | x) = p(y | x) \det(J) \quad (5.12)$$

where $\det(J)$ is the Jacobian of the transformation $y = f(z)$ which does not depend on x . We also have $dz \det(J) = dy$. It then follows that

$$\int_z \left[\frac{\partial p(z | x)}{\partial x} \right]^2 \frac{1}{p(z | x)} dz = \int_y \left[\frac{\partial p(y | x)}{\partial x} \right]^2 \frac{1}{p(y | x)} dy \quad (5.13)$$

i.e., the value of the integral is independent of the transformation. This is quite natural since a transformation means stretching or compressing the horizontal axis and since the area below the PDF is always one, the value of the frequency function has to change accordingly, i.e., the $\det(J)$ factor in numerator and denominator cancels out. Remember now that our discussion so far deals with a supposed known estimator. The next step in the derivation is therefore to find a suitable transformation which will allow us to say that our findings also will be valid for any estimator, especially for the one that gives the lowest variance.

An illustration of how to choose this transformation may be of value. Assume therefore that we want to estimate the mean of a PDF and our sample size is, say, two samples, $y = [y_1, y_2]^T$, with equal means and variances for y_i . We choose our estimator to be $\hat{x} = (y_1 + y_2)/2$. The contours of the PDF for the sample vector are illustrated in Figure 5.3 together with the lines C_i for y_1 and y_2 which will give the same estimate of x .

A suitable transformation is $z(y) = [\hat{x}(y), \xi(y)]^T$ where the dimension of the vector ξ is $N-1$ and the dimension of the vector y is N . The requirement on ξ in the transformation is that the transformation should be 1:1 and the derivatives

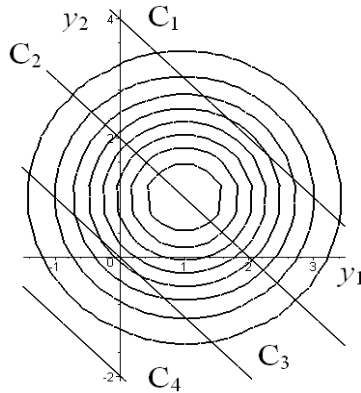


Figure 5.3: The contour lines for the sample PDF with shown hyper surfaces giving the same estimate of the mean.

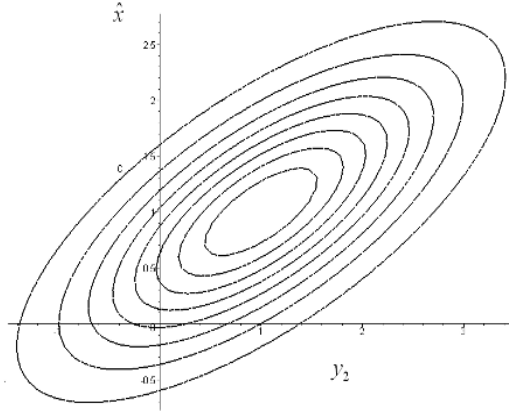


Figure 5.4: The contour lines for the transformed sample PDF in Figure 5.3.

$\partial z_i / \partial \xi_k$ should exist [CR46]. The transformation means that the slanted lines C_i in Figure 5.3 will then be horizontal in a graph of the transformed variable, see Figure 5.4, and this makes it easy to split up (5.13) by Bayes theorem. We can achieve this transformation by for every point along a line C_i map the PDF-value at this point at the vertical coordinate C_i and horizontal coordinate y_2 in the transformed PDF. The next step now would be to use the conditional identity $p(\alpha, \beta) = p(\alpha|\beta)p(\beta)$. This will give us

$$p(z | x) = p(\hat{x}, \xi | x) = p(\hat{x} | x) p(\xi | \hat{x}, x) \quad (5.14)$$

If we differentiate this expression we will have

$$\frac{\partial p(z | x)}{\partial x} = \frac{\partial p(\hat{x} | x)}{\partial x} p(\xi | \hat{x}, x) + p(\hat{x} | x) \frac{\partial p(\xi | \hat{x}, x)}{\partial x} \quad (5.15)$$

Substituting this into (5.13) gives

$$\begin{aligned}
\int_y \left[\frac{\partial p(y|x)}{\partial x} \right]^2 \frac{1}{p(y|x)} dy &= \int_z \left[\frac{\partial p(z|x)}{\partial x} \right]^2 \frac{1}{p(z|x)} dz \\
&= \int_z \left[\frac{\partial p(\hat{x}|x)}{\partial x} p(\xi|\hat{x},x) + p(\hat{x}|x) \frac{\partial p(\xi|\hat{x},x)}{\partial x} \right]^2 \frac{1}{p(z|x)} dz \\
&= \int_{\hat{x}, \xi} \left[\frac{\partial p(\hat{x}|x)}{\partial x} \right]^2 \frac{p(\xi|\hat{x},x)}{p(\hat{x}|x)} d\hat{x} d\xi + 2 \int_{\hat{x}, \xi} \frac{\partial p(\xi|\hat{x},x)}{\partial x} \frac{\partial p(\hat{x}|x)}{\partial x} d\hat{x} d\xi \quad (5.16) \\
&\quad + \int_{\hat{x}, \xi} \left[\frac{\partial p(\xi|\hat{x},x)}{\partial x} \right]^2 \frac{p(\hat{x}|x)}{p(\xi|\hat{x},x)} d\hat{x} d\xi
\end{aligned}$$

The middle term can be written as the product of two integrals and since

$$\int_{\xi} \frac{\partial p(\xi|\hat{x},x)}{\partial x} d\xi = 0 \quad \text{due to} \quad \int_{\xi} p(\xi|\hat{x},x) d\xi = 1$$

the middle term will be zero. The last integral will always be greater or equal to zero since the differential is in square and the involved PDFs are always positive. We also note that the first integral can be written as the product of two integrals, one with the value 1. We therefore have

$$\int_y \left[\frac{\partial p(y|x)}{\partial x} \right]^2 \frac{1}{p(y|x)} dy \geq \int_{\hat{x}} \left[\frac{\partial p(\hat{x}|x)}{\partial x} \right]^2 \frac{1}{p(\hat{x}|x)} d\hat{x} \quad (5.17)$$

This relates the PDF for our particular estimator with the PDF for the sample vector. We can therefore write expression (5.11) as

$$E\{(\hat{x} - x)^2\} \geq \frac{\left[\frac{\partial \psi(x)}{\partial x} \right]^2}{\int_y \left[\frac{\partial p(y|x)}{\partial x} \right]^2 \frac{1}{p(y|x)} dy} \quad (5.18)$$

which is the scalar parameter CRLB when our sample PDF is $p(y|x)$ and we note that this expression does not depend on our choice of estimator. The mean

square error is always greater or equal to the right hand side of the expression independent of our choice of estimator. If we do not have bias, i.e., $\psi(x) = x$, this will also be the variance of the estimator. Expression (5.18) can also be written

$$E\{(\hat{x} - x)^2\} \geq \frac{\left[\frac{\partial \psi(x)}{\partial x}\right]^2}{E\left\{\left[\frac{\partial \log(p(y|x))}{\partial x}\right]^2\right\}} \quad (5.19)$$

or

$$E\{(\hat{x} - x)^2\} \geq -\frac{\left[\frac{\partial \psi(x)}{\partial x}\right]^2}{E\left\{\frac{\partial^2 \log(p(y|x))}{\partial x^2}\right\}} \quad (5.20)$$

We can also derive a lower bound on the variance. If we in (5.6) multiply with $E\{\hat{x}|x\} = x + b(x)$ instead of x we will have

$$\frac{\partial \psi(x)}{\partial x} = \int_{\hat{x}} (\hat{x} - E\{\hat{x}|x\}) \frac{\partial p(\hat{x}|x)}{\partial x} d\hat{x} \quad (5.21)$$

instead of (5.8) and

$$\left[\frac{\partial \psi(x)}{\partial x}\right]^2 \leq \underbrace{\left[\int_{\hat{x}} (\hat{x} - E\{\hat{x}|x\})^2 p(\hat{x}|x) d\hat{x}\right]}_{\text{Var}\{\hat{x}\}} \left[\int_{\hat{x}} \left[\frac{\partial p(\hat{x}|x)}{\partial x}\right]^2 \frac{1}{p(\hat{x}|x)} d\hat{x}\right] \quad (5.22)$$

or

$$\text{Var}\{\hat{x}\} \geq \frac{\left[\frac{\partial \psi(x)}{\partial x}\right]^2}{E\left\{\left[\frac{\partial \log(p(y|x))}{\partial x}\right]^2\right\}} \quad (5.23)$$

5.2.2 The vector parameter CRLB

A simple approach would be to note that in the derivation of the scalar CRLB we have not specifically used the information that the parameter was a scalar. The derivation would formally been the same if the scalar parameter x instead had been a parameter vector \mathbf{x} . Therefore the CRLB for a biased estimator is

$$\begin{aligned} E\{(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^T\} &\geq -\left[\frac{\partial \boldsymbol{\Psi}(\mathbf{x})}{\partial \mathbf{x}}\right]^T \left[E\left\{\frac{\partial^2 \log(p(\mathbf{y} | \mathbf{x}))}{\partial \mathbf{x} \partial \mathbf{x}^T}\right\}\right]^{-1} \frac{\partial \boldsymbol{\Psi}(\mathbf{x})}{\partial \mathbf{x}} \\ &= -\left[\mathbf{I} + \frac{\partial \mathbf{b}(\mathbf{x})}{\partial \mathbf{x}}\right]^T \left[E\left\{\frac{\partial^2 \log(p(\mathbf{y} | \mathbf{x}))}{\partial \mathbf{x} \partial \mathbf{x}^T}\right\}\right]^{-1} \left[\mathbf{I} + \frac{\partial \mathbf{b}(\mathbf{x})}{\partial \mathbf{x}}\right] \end{aligned} \quad (5.24)$$

The inequality means that if the matrix to the left is subtracted with the matrix to the right the resulting matrix is positive semidefinite. The quantity

$$\mathbf{J} = -E\left\{\frac{\partial^2 \log(p(\mathbf{y} | \mathbf{x}))}{\partial \mathbf{x} \partial \mathbf{x}^T}\right\} \quad (5.25)$$

is called the Fisher's information matrix (FIM).

In the case we have bias the relation between the mean square error and the covariance is

$$\begin{aligned} Mse\{\hat{\mathbf{x}}\} &= E\{(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^T\} \\ &= E\{[(\hat{\mathbf{x}} - E\{\hat{\mathbf{x}}\}) + (E\{\hat{\mathbf{x}}\} - \mathbf{x})][(\hat{\mathbf{x}} - E\{\hat{\mathbf{x}}\}) + (E\{\hat{\mathbf{x}}\} - \mathbf{x})]^T\} \\ &= Cov\{\hat{\mathbf{x}}\} + \mathbf{b}(\mathbf{x})\mathbf{b}(\mathbf{x})^T \end{aligned} \quad (5.26)$$

If the estimator is unbiased then the Mse and Cov are identical and we have

$$Cov\{\hat{\mathbf{x}}\} = E\{(\hat{\mathbf{x}} - E\{\hat{\mathbf{x}}\})(\hat{\mathbf{x}} - E\{\hat{\mathbf{x}}\})^T\} \geq -\left[E\left\{\frac{\partial^2 \log(p(\mathbf{y} | \mathbf{x}))}{\partial \mathbf{x} \partial \mathbf{x}^T}\right\}\right]^{-1} \quad (5.27)$$

Figure 5.5 illustrates the relations between the estimators.

If we know both the bias and the mean square error, we can calculate the covariance from $Mse(\hat{\mathbf{x}} | \mathbf{x}) = Cov\{\hat{\mathbf{x}} | \mathbf{x}\} + \mathbf{b}(\mathbf{x})\mathbf{b}(\mathbf{x})^T$, in other cases we have to resort to Monte Carlo methods to determine $E\{\hat{\mathbf{x}}\}$ and $Cov(\hat{\mathbf{x}})$ besides $Mse(\hat{\mathbf{x}})$. Generally we have the inequality

$$Mse\{\hat{\mathbf{x}}\} \geq Cov\{\hat{\mathbf{x}}\} \geq -\left[\mathbf{I} + \frac{\partial \mathbf{b}(\mathbf{x})}{\partial \mathbf{x}}\right]^T \left[E \left\{ \frac{\partial^2 \log(p(\mathbf{y} | \mathbf{x}))}{\partial \mathbf{x} \partial \mathbf{x}^T} \right\} \right]^{-1} \left[\mathbf{I} + \frac{\partial \mathbf{b}(\mathbf{x})}{\partial \mathbf{x}} \right] \quad (5.28)$$

The mean square error for a biased estimator in the case the CRLB is attained is

$$Mse\{\hat{\mathbf{x}}\} = -\left[\mathbf{I} + \frac{\partial \mathbf{b}(\mathbf{x})}{\partial \mathbf{x}}\right]^T \left[E \left\{ \frac{\partial^2 \log(p(\mathbf{y} | \mathbf{x}))}{\partial \mathbf{x} \partial \mathbf{x}^T} \right\} \right]^{-1} \left[\mathbf{I} + \frac{\partial \mathbf{b}(\mathbf{x})}{\partial \mathbf{x}} \right] + \mathbf{b}(\mathbf{x})\mathbf{b}(\mathbf{x})^T \quad (5.29)$$

and the covariance

$$Cov\{\hat{\mathbf{x}}\} = -\left[\mathbf{I} + \frac{\partial \mathbf{b}(\mathbf{x})}{\partial \mathbf{x}}\right]^T \left[E \left\{ \frac{\partial^2 \log(p(\mathbf{y} | \mathbf{x}))}{\partial \mathbf{x} \partial \mathbf{x}^T} \right\} \right]^{-1} \left[\mathbf{I} + \frac{\partial \mathbf{b}(\mathbf{x})}{\partial \mathbf{x}} \right] \quad (5.30)$$

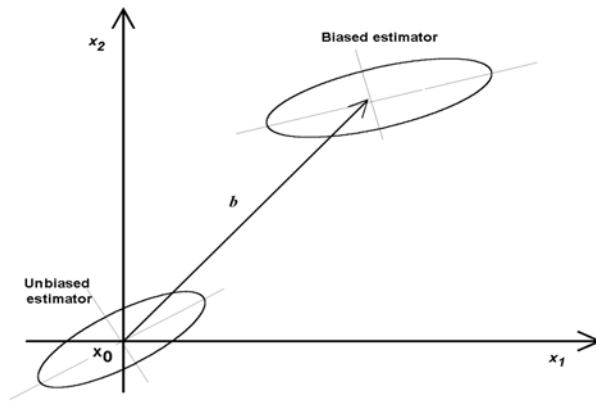


Figure 5.5: Illustration of the PDF for a biased and an unbiased estimator.

5.3 Additive Gaussian noise

5.3.1 The scalar case

In the derivation of the scalar parameter CRLB we did not make any assumptions about the noise or the structure of the observations y_i . Often the noise can be characterized as additive and our observations are $y_i = h_i(x) + e_i, i=1,2,3 \dots N$ or in vector form

$$\mathbf{y} = \mathbf{h}(x) + \mathbf{e} \quad (5.31)$$

Assuming that the noise is Gaussian with zero mean with a covariance matrix \mathbf{C} , the joint PDF of the observations is

$$p(\mathbf{y} | x) = \frac{1}{(2\pi)^{N/2} \sqrt{\det(\mathbf{C})}} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{h}(x))^T \mathbf{C}^{-1}(\mathbf{y} - \mathbf{h}(x))\right) \quad (5.32)$$

Differentiating with respect to x and taking the expectation yields

$$E \left\{ \left[\frac{\partial \log p(\mathbf{y} | x)}{\partial x} \right]^2 \right\} = \left[\frac{\partial \mathbf{h}(x)}{\partial x} \right]^T \mathbf{C}^{-1} \left[\frac{\partial \mathbf{h}(x)}{\partial x} \right] \quad (5.33)$$

and, hence, according to (5.23)

$$\text{Var} \{\hat{x}\} \geq \frac{\left[\frac{\partial \psi(x)}{\partial x} \right]^2}{\left[\frac{\partial \mathbf{h}(x)}{\partial x} \right]^T \mathbf{C}^{-1} \left[\frac{\partial \mathbf{h}(x)}{\partial x} \right]} \quad (5.34)$$

5.3.2 The vector case

Assume that the observation vector is Gaussian $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \mathbf{C}(\boldsymbol{\theta}))$ where the mean vector is $N \times 1$ and the covariance matrix is $N \times N$. Both depend on the

unknown parameter vector $\boldsymbol{\theta}$ which we want to estimate. The PDF for the observation vector is thus

$$p(\mathbf{y} | \boldsymbol{\theta}) = \frac{1}{(2\pi)^{N/2} \sqrt{\det(\mathbf{C}(\boldsymbol{\theta}))}} \exp \left[-\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\theta}))^T \mathbf{C}^{-1}(\boldsymbol{\theta}) (\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\theta})) \right] \quad (5.35)$$

Taking the first derivative yields

$$\frac{\partial \log p(\mathbf{y} | \boldsymbol{\theta})}{\partial \theta_k} = -\frac{1}{2} \frac{\partial \log \det[\mathbf{C}(\boldsymbol{\theta})]}{\partial \theta_k} - \frac{1}{2} \frac{\partial}{\partial \theta_k} \left[(\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\theta}))^T \mathbf{C}^{-1}(\boldsymbol{\theta}) (\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\theta})) \right] \quad (5.36)$$

Now it can be shown, see for example [KA93, SC91], that the two terms in (5.36) can be written

$$\frac{\partial \log \det[\mathbf{C}(\boldsymbol{\theta})]}{\partial \theta_k} = \text{Tr} \left[\mathbf{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_k} \right] \quad (5.37)$$

and

$$\begin{aligned} & \frac{\partial}{\partial \theta_k} \left[(\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\theta}))^T \mathbf{C}^{-1}(\boldsymbol{\theta}) (\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\theta})) \right] \\ &= -2 \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})^T}{\partial \theta_k} \mathbf{C}^{-1}(\boldsymbol{\theta}) (\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\theta})) + (\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\theta}))^T \frac{\partial \mathbf{C}^{-1}(\boldsymbol{\theta})}{\partial \theta_k} (\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\theta})) \end{aligned} \quad (5.38)$$

where the derivatives $\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_i}$ and $\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_i}$ mean the first derivative of all elements

in the vector and matrix respectively with respect to the parameter θ_i . The next step is to evaluate the elements in the Fisher matrix (5.25). After some manipulations, see [VT02, KA93, SC91], the result is

$$[\mathbf{J}(\boldsymbol{\theta})]_{i,j} = \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})^T}{\partial \theta_i} \mathbf{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_j} + \frac{1}{2} \text{Tr} \left[\mathbf{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_j} \right] \quad (5.39)$$

The scalar parameter case follows also directly from (5.39)

$$J(\theta) = \frac{\partial \boldsymbol{\mu}(\theta)^T}{\partial \theta} \mathbf{C}^{-1}(\theta) \frac{\partial \boldsymbol{\mu}(\theta)}{\partial \theta} + \frac{1}{2} \text{Tr} \left[\left[\mathbf{C}^{-1}(\theta) \frac{\partial \mathbf{C}(\theta)}{\partial \theta} \right]^2 \right] \quad (5.40)$$

and in the case the covariance matrix does not depend on the parameter

$$J(\theta) = \frac{\partial \boldsymbol{\mu}(\theta)^T}{\partial \theta} \mathbf{C}^{-1} \frac{\partial \boldsymbol{\mu}(\theta)}{\partial \theta} \quad (5.41)$$

5.4 The bias expression

We have previously assumed that the bias may be expressed as $E\{\hat{\mathbf{x}} | \mathbf{x}\} = \mathbf{x} + \mathbf{b}(\mathbf{x})$, that is the expected value of the estimate of \mathbf{x} depends in a nonlinear fashion of the true value of \mathbf{x} . In most cases the bias is introduced in the sampling process and we will here analyze two cases more in detail. We will restrict ourself to the case of estimating the mean.

5.4.1 The linear case

Let the stochastic variable ξ have the PDF $f_{\xi}(x)$ with the mean x . We can construct the following unbiased estimator of the mean based on N samples of the stochastic variable ξ

$$\hat{x}_{Unb} = \frac{1}{N} \sum_{i=1}^N \xi_i \quad (5.42)$$

Now, assume that our samples, due to an imperfect sampling device, instead are $\eta_i = \alpha + \beta \xi_i$ and that we estimate the mean from those samples. This will give us a biased estimator

$$\hat{x}_{Bias} = \frac{1}{N} \sum_{i=1}^N \eta_i = \frac{1}{N} \sum_{i=1}^N (\alpha + \beta \xi_i) = \alpha + \frac{\beta}{N} \sum_{i=1}^N \xi_i = \alpha + \beta \hat{x}_{Unb} \quad (5.43)$$

Thus the biased estimate of the mean is an affine transformation of the unbiased estimate. We can determine the mean of the biased estimate

$$E\{\hat{x}_{Bias}\} = E\{\alpha + \beta \hat{x}_{Unb}\} = \alpha + \beta E\{\hat{x}_{Unb}\} = \alpha + \beta x \quad (5.44)$$

This can also be written

$$E\{\hat{x}_{Bias}\} = x + \alpha + (\beta - 1)x \quad (5.45)$$

This means that the bias expression in (5.2) in the case of an affine transformation of the samples is

$$b(x) = \alpha + (\beta - 1)x \quad (5.46)$$

The variance of the biased estimator is (5.23) if the bound is obtained

$$Var\{\hat{x}_{Bias}\} = (1 + \frac{\partial b(x)}{\partial x})^2 Var\{\hat{x}_{Unb}\} = (1 + \beta - 1)^2 Var\{\hat{x}_{Unb}\} = \beta^2 Var\{\hat{x}_{Unb}\} \quad (5.47)$$

This can also be shown directly by

$$Var\{\hat{x}_{Bias}\} = Var\{\alpha\} + Var\{\beta \hat{x}_{Unb}\} = \beta^2 Var\{\hat{x}_{Unb}\} \quad (5.48)$$

We can also study the PDF for the stochastic variable with bias. We have

$$f_{\eta}(x) = \frac{\partial}{\partial x} P\{\eta \leq x\} = \frac{\partial}{\partial x} P\{\xi \leq -\frac{\alpha}{\beta} + \frac{1}{\beta}x\} = \frac{1}{\beta} f_{\xi}(-\frac{\alpha}{\beta} + \frac{1}{\beta}x) \quad (5.49)$$

That means that the PDF for the stochastic variable with bias is offset by $-\alpha/\beta$ and that the horizontal scale is changed to $1/\beta$ but the vertical scale is also changed the same amount so the area below the PDF remains one.

5.4.2 The nonlinear case

Let us now study a nonlinear transformation, $y=g(x)$, of the stochastic variable ξ , see Figure 5.6

$$\eta = g(\xi) \quad (5.50)$$

Looking at Figure 5.6 it is obvious that it is possible to numerically calculate the bias as a function of the true value x if $g(\cdot)$ is known. From the figure we also see that depending on the curvature of $g(\cdot)$ the variance will increase or decrease. Let us examine the situation more closely.

Assume that the true value of the parameter estimated by ξ is $x=0$ and that $E\{\xi\}=x$ and do a Taylor expansion of the function $g(\cdot)$ around that point. We assume that the Taylor expansion is convergent, which is reasonable if both the biased and the unbiased estimate of x exist. The estimator will be

$$\begin{aligned}\hat{x}_{Bias} &= \frac{1}{N} \sum_{i=1}^N \eta_i = \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^{\infty} \frac{\xi_i^k}{k!} \frac{\partial^k g}{\partial x^k} = g_0 + g_1 \frac{1}{N} \sum_{i=1}^N \xi_i + \frac{1}{N} \sum_{i=1}^N \sum_{k=2}^{\infty} g_k \frac{\xi_i^k}{k!} \\ &= g_0 + g_1 \hat{x}_{Unb} + \frac{1}{N} \sum_{i=1}^N \sum_{k=2}^{\infty} g_k \frac{\xi_i^k}{k!}\end{aligned}\quad (5.51)$$

where g_k are the derivatives of g in the Taylor expansion of $y=g(x)$ at $x=0$. We note here that different sequences of ξ_1, \dots, ξ_N may result in the

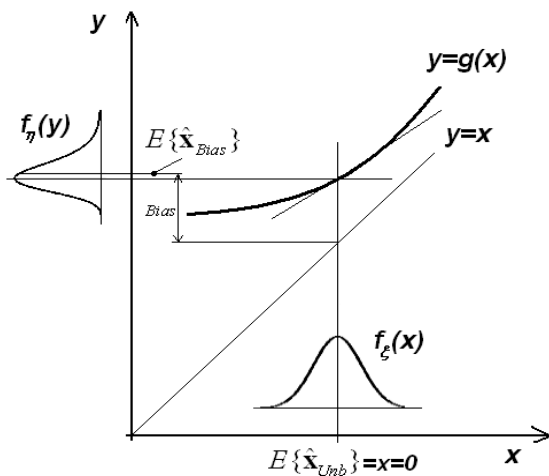


Figure 5.6: Illustration of the nonlinear transformation of the stochastic variable ξ . The PDFs in the figure refer to the estimates of the mean. True value of the parameter estimated by ξ is $x=0$.

same \hat{x}_{Unb} but different third term. After taking the expectation of both sides we have

$$E\{\hat{x}_{Bias}\} = g_0 + g_1 x + \frac{1}{N} \sum_{i=1}^N \sum_{k=2}^{\infty} g_k \frac{E\{\xi_i^k\}}{k!} = g_0 + \frac{1}{N} \sum_{i=1}^N \sum_{k=2}^{\infty} g_k \frac{E\{\xi_i^k\}}{k!} \quad (5.52)$$

since $x=0$. The last double sum is a function of the point x of the Taylor expansion, in this case $x=0$. An analytic expression of the expectation $E\{\xi_i^k\}$ is easily found for a Gaussian distributed ξ_i with zero mean

$$E\{\xi_i^k\} = 1 \cdot 3 \cdot 5 \cdots (k-1) Var\{\xi_i\} = (k-1)!! Var\{\xi_i\} \quad (5.53)$$

Often only a few terms in the Taylor expansion, i.e., the inner sum is needed since the tails of the PDF of ξ are cut due to physical reasons. This means that the bias expression in (5.52) can be calculated for different points x of the Taylor expansion and the bias as a function of x

$$B(x) = x + b(x) \quad (5.54)$$

can be calculated. A linear approximation of the bias function $B(x)$ in a point x is

$$E\{\hat{x}_{Bias}\} = \alpha + \beta x \quad (5.55)$$

We have also from (5.51)

$$\begin{aligned} Var\{\hat{x}_{Bias}\} &= Var\left\{\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^{\infty} g_k \frac{\xi_i^k}{k!}\right\} = \frac{1}{N} Var\left\{\sum_{k=1}^{\infty} g_k \frac{\xi^k}{k!}\right\} \\ &\leq \frac{\max(g_k)^2}{N} \sum_{k=1}^{\infty} \frac{(2k-1)!!}{(k!)^2} \sigma_{\xi}^{2k} \end{aligned} \quad (5.56)$$

where σ_{ξ}^2 is the variance of ξ .

5.5 Some examples of the CRLB in terrain navigation

The CRLB for a position estimate is not only of interest in our Kalman filter but also when deciding upon way-points. In the so called posterior CRLB defined in [VT68] a CRLB based on the mean of the FIM matrix which takes the information from the prior position PDF into account is calculated. Since this bound gives what we at best can expect in accuracy at a way-point we will discuss it first before developing some general expressions for the CRLB when we have a Gaussian measurement noise. We will also develop an expression for the position error covariance when we have bending of the measurement beams due to temperature gradients in the water

In the following we are assuming that our prior position PDF has its mean in the way-point and that our estimates are the position x_1 and x_2 in East and North directions, respectively, and in some cases the measurement bias d . The numbers of parameters to estimate are thus two or three. The number of measurement beams is N and the measurement accuracy is assumed to be known. The measurement noise is assumed to be additive and Gaussian.

5.5.1 The posterior FIM

In calculation of the posterior CRLB defined in [VT68] the improvement in accuracy by the prior is taken into account in the following way. The posterior CRLB is the inverse of the expected posterior Fisher matrix. The posterior FIM is

$$\begin{aligned} \mathbf{J} &= -E\left\{\frac{\partial^2 [\log p(\mathbf{x}, \mathbf{y})]}{\partial \mathbf{x} \partial \mathbf{x}^T}\right\} = -E\left\{\frac{\partial^2 [\log(p(\mathbf{y} | \mathbf{x})p(\mathbf{x}))]}{\partial \mathbf{x} \partial \mathbf{x}^T}\right\} \\ &= -E\left\{\frac{\partial^2 [\log p(\mathbf{y} | \mathbf{x})]}{\partial \mathbf{x} \partial \mathbf{x}^T}\right\} - E\left\{\frac{\partial^2 [\log(p(\mathbf{x}))]}{\partial \mathbf{x} \partial \mathbf{x}^T}\right\} = \mathbf{J}_{L,J} + \mathbf{J}_p \end{aligned} \quad (5.57)$$

The equation (5.57) gives a “global bound” [VT02] which is independent of the actual position since the expectation in (5.57) is with respect to both \mathbf{x} and \mathbf{y} .

The first term can be written as

$$\begin{aligned}
\mathbf{J}_{L,J} &= -E\left\{\frac{\partial^2[\log p(\mathbf{y} | \mathbf{x})]}{\partial \mathbf{x} \partial \mathbf{x}^T}\right\} = -\iint_{\mathbf{x}, \mathbf{y}} \frac{\partial^2[\log p(\mathbf{y} | \mathbf{x})]}{\partial \mathbf{x} \partial \mathbf{x}^T} p(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} \\
&= -\int_{\mathbf{x}} \left\{ \int_{\mathbf{y}} \frac{\partial^2[\log p(\mathbf{y} | \mathbf{x})]}{\partial \mathbf{x} \partial \mathbf{x}^T} p(\mathbf{y} | \mathbf{x}) d\mathbf{y} \right\} p(\mathbf{x}) d\mathbf{x} \\
&= -\int_{\mathbf{x}} E_{\mathbf{y}} \left\{ \frac{\partial^2[\log p(\mathbf{y} | \mathbf{x})]}{\partial \mathbf{x} \partial \mathbf{x}^T} \right\} p(\mathbf{x}) d\mathbf{x}
\end{aligned} \tag{5.58}$$

Similarly, the second term in (5.57) can be expressed as

$$\begin{aligned}
\mathbf{J}_p &= -E\left\{\frac{\partial^2[\log p(\mathbf{x})]}{\partial \mathbf{x} \partial \mathbf{x}^T}\right\} = -\iint_{\mathbf{x}, \mathbf{y}} \frac{\partial^2[\log p(\mathbf{x})]}{\partial \mathbf{x} \partial \mathbf{x}^T} p(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} \\
&= -\int_{\mathbf{x}} \frac{\partial^2[\log(p(\mathbf{x}))]}{\partial \mathbf{x} \partial \mathbf{x}^T} p(\mathbf{x}) d\mathbf{x} \left\{ \int_{\mathbf{y}} p(\mathbf{y} | \mathbf{x}) d\mathbf{y} \right\} \\
&= -\int_{\mathbf{x}} \frac{\partial^2[\log(p(\mathbf{x}))]}{\partial \mathbf{x} \partial \mathbf{x}^T} p(\mathbf{x}) d\mathbf{x}
\end{aligned} \tag{5.59}$$

which is the expectation of the FIM for the prior and represents the improvement due to the prior knowledge.

Sometimes we will be interested in the FIM for a certain position \mathbf{x} . Therefore, omitting the expectation over \mathbf{x} , we have

$$\begin{aligned}
\mathbf{J} &= -E_{\mathbf{y}} \left\{ \frac{\partial^2[\log L(\mathbf{x}; \mathbf{y}) p(\mathbf{x})]}{\partial \mathbf{x} \partial \mathbf{x}^T} \right\} \\
&= -E_{\mathbf{y}} \left\{ \frac{\partial^2 \log L(\mathbf{x}; \mathbf{y})}{\partial \mathbf{x} \partial \mathbf{x}^T} + \frac{\partial^2 \log p(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^T} \right\} = \mathbf{J}_L + \mathbf{J}_p
\end{aligned} \tag{5.60}$$

The notation $L(\mathbf{x}; \mathbf{y})$ for the likelihood function has been used in (5.60). This bound is thus a function of the true position \mathbf{x} .

5.5.2 Two measures of the posterior accuracy

Planning of vehicle routes means that way-points have to be selected and it is of great interest to know what accuracy that can be expected when positioning in the way-point areas.

In many interesting cases, discussed in Chapter 6, our estimate and our true position will be close to each other and our accuracy will almost reach the CRLB. By calculating the expected value of the posterior CRLB, based on (5.60), we will therefore have a measure of what accuracy we can expect to be used in our Kalman filter for an actual choice of a way-point

$$E\{\mathbf{R}_p\} = \int_{\mathbf{x}} \mathbf{R}_p(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \quad (5.61)$$

where $\mathbf{R}_p(\mathbf{x})$ is the local posterior CRLB given by the inverse of (5.60) and the integration is over the area where $\mathbf{R}_p(\mathbf{x})$ is defined. The CRLB based on (5.57) will give a lower bound for (5.61), cf. the relation between the harmonic and the arithmetic means.

The accuracy of the position based on the prior PDF is often much less than the accuracy based on the terrain profile measurement which means that the improvement by the prior can be neglected. In this case the measure will be

$$E\{\mathbf{R}\} = \int_{\mathbf{x}} \mathbf{R}(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \quad (5.62)$$

where $\mathbf{R}(\mathbf{x})$ is the usual position error covariance matrix, i.e., the CRLB based on (5.60) where the improvement by the prior has been neglected. The integration is over the area where $\mathbf{R}(\mathbf{x})$ is defined.

The reason for introducing the measures (5.61, 5.62) instead of using the posterior CRLB based on (5.57) is the following. Firstly it mimics our actual use of the CRLB in the Kalman filter. Secondly the bottom area covered by the prior PDF may contain flat parts where the local FIM (5.60) is almost zero. However, calculating the expected value as in (5.57), the global FIM, may give an expected value that indicates good navigation properties which obviously may not be the case. In fact, most of the area can be useless for terrain navigation purposes but the global FIM indicates otherwise.

In general we can measure our position by several means in addition to the

terrain positioning which is indicated in Chapter 3. Assuming that the measurements are independent our likelihood function can be written as

$$L(\mathbf{x}; \mathbf{y}) = \prod_{k=1}^M L_k(\mathbf{x}; \mathbf{y}_k) \quad (5.63)$$

and the Fisher's information matrix is

$$\mathbf{J} = \sum_{k=1}^M \mathbf{J}_{k,L} + \mathbf{J}_p \quad (5.64)$$

5.5.3 Constant measurement covariance matrix, no measurement bias

In Chapter 5.3.2 the Fisher matrix was determined for the general case of a process with a Gaussian observation vector with the probability density $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \mathbf{C}(\boldsymbol{\theta}))$. Using these findings for the observation vector $\mathbf{y} = \mathbf{h}(\mathbf{x}) + \mathbf{e}$, i.e., where \mathbf{e} is Gaussian and we have no measurement bias equation (5.39) gives directly

$$[\mathbf{J}(\boldsymbol{\theta})]_{i,j} = \frac{\partial \mathbf{h}(\mathbf{x})^T}{\partial x_i} \mathbf{C}^{-1} \frac{\partial \mathbf{h}(\mathbf{x})}{\partial x_j} \quad \text{where } i,j=1,2 \quad (5.65)$$

If we assume the measurement beams to be independent the covariance matrix will be diagonal. In the case the measurement accuracy is the same in all beams equation (5.65) simplifies to

$$[\mathbf{J}(\boldsymbol{\theta})]_{i,j} = \frac{1}{\sigma_e^2} \frac{\partial \mathbf{h}(\mathbf{x})^T}{\partial x_i} \frac{\partial \mathbf{h}(\mathbf{x})}{\partial x_j} = \frac{1}{\sigma_e^2} \sum_{k=1}^N \frac{\partial h_k(\mathbf{x})}{\partial x_i} \frac{\partial h_k(\mathbf{x})}{\partial x_j} \quad (5.66)$$

where σ_e^2 is the measurement error variance.

5.5.4 Profile matching, constant measurement covariance matrix

One way to reduce the measurement bias caused by the variation of the sea surface level, variation in sound speed and instrument biases is to match the measured bottom profile and the profile according to the map, i.e., the sonar measurement of the bottom is reduced by the mean of the measurement and the corresponding map depths are reduced by their mean. These vectors are also given by pre-multiplying the vectors with the absolute depth values by a projection matrix

$$\mathbf{\Pi} = \mathbf{I}_N - \mathbf{1}_{N,N} / N \quad (5.67)$$

where \mathbf{I}_N is the identity matrix and $\mathbf{1}_{N,N}$ is a matrix with ones as elements, both of dimension $N \times N$. We now want to use (5.39) directly and we therefore observe that our observation vector now is $\mathbf{\Pi y}$. Our data model is therefore $\mathbf{\Pi y} = \mathbf{\Pi h(x)} + \mathbf{\epsilon}$. Note that reducing the measurement by the mean introduces a correlation between the individual stochastic beam errors. The covariance matrix $\mathbf{C}_e = \sigma_e^2 (\mathbf{I}_N - \mathbf{1}_{N,N}/N) = \sigma_e^2 \mathbf{\Pi}$ for zero mean Gaussian measurements with variance σ_e^2 . Since the covariance matrix is singular (5.39) gives after a simple limiting calculation

$$[\mathbf{J(x)}]_{i,j} = \frac{1}{\sigma_e^2} \frac{\partial [\mathbf{\Pi h(x)}]^T}{\partial x_i} \frac{\partial [\mathbf{\Pi h(x)}]}{\partial x_j} = \frac{1}{\sigma_e^2} \frac{\partial \mathbf{h(x)}^T}{\partial x_i} \mathbf{\Pi} \frac{\partial \mathbf{h(x)}}{\partial x_j} \quad (5.68)$$

5.5.5 Regular matching, measurement bias and constant measurement covariance matrix

We can apply (5.39) to the case when the measurement bias is explicitly modeled, i.e., $\mathbf{y} = \mathbf{h(x)} + d\mathbf{u} + \mathbf{e}$. The stochastic measurement error is assumed to be zero mean Gaussian with the covariance \mathbf{C} and d is the bias and the vector \mathbf{u} is a N dimensional vector of ones. As before if $\mathbf{x} = [x_1, x_2, d]^T$

$$[\mathbf{J(x)}]_{i,j} = \frac{\partial \mathbf{\mu(x)}^T}{\partial x_i} \mathbf{C}^{-1} \frac{\partial \mathbf{\mu(x)}}{\partial x_j} \text{ where } i,j=1,2,3 \quad (5.69)$$

The mean of the observation vector is

$$\mathbf{\mu(x)} = \mathbf{h(x)} + d\mathbf{u} \quad (5.70)$$

where d is the measurement bias, equal in all beams. Collect the derivatives into the matrix

$$[\mathbf{G}, \mathbf{u}] = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} & 1 \\ \frac{\partial h_2}{\partial x_1} & \frac{\partial h_2}{\partial x_2} & 1 \\ \vdots & \vdots & \vdots \\ \frac{\partial h_N}{\partial x_1} & \frac{\partial h_N}{\partial x_2} & 1 \end{bmatrix} \quad (5.71)$$

then

$$\mathbf{J} = \begin{bmatrix} \mathbf{G}^T \\ \mathbf{u}^T \end{bmatrix} \mathbf{C}^{-1} \begin{bmatrix} \mathbf{G} & \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{G}^T \mathbf{C}^{-1} \mathbf{G} & \mathbf{G}^T \mathbf{C}^{-1} \mathbf{u} \\ \mathbf{u}^T \mathbf{C}^{-1} \mathbf{G} & \mathbf{u}^T \mathbf{C}^{-1} \mathbf{u} \end{bmatrix} \quad (5.72)$$

Using the inversion formulae [LÜ96]

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{B}\mathbf{D}^{-1}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{bmatrix} \quad (5.73)$$

we have the block matrix

$$\begin{aligned} [\mathbf{J}^{-1}]_{1,1} &= [\mathbf{G}^T \mathbf{C}^{-1} \mathbf{G} - \mathbf{G}^T \mathbf{C}^{-1} \mathbf{u} (\mathbf{u}^T \mathbf{C}^{-1} \mathbf{u})^{-1} \mathbf{u}^T \mathbf{C}^{-1} \mathbf{G}]^{-1} \\ &= [\mathbf{G}^T \mathbf{C}^{-1/2} [\mathbf{I} - \mathbf{C}^{-1/2} \mathbf{u} (\mathbf{u}^T \mathbf{C}^{-1} \mathbf{u})^{-1} \mathbf{u}^T \mathbf{C}^{-1/2}] \mathbf{C}^{-1/2} \mathbf{G}]^{-1} \end{aligned} \quad (5.74)$$

In the case the covariance matrix is $\mathbf{C} = \sigma_e^2 \mathbf{I}_N$ this expression transforms to (5.68). When estimating d the inverse Fisher matrix element will be

$$[\mathbf{J}^{-1}]_{2,2} = [\mathbf{u}^T \mathbf{C}^{-1} \mathbf{u} - \mathbf{u}^T \mathbf{C}^{-1} \mathbf{G} (\mathbf{G}^T \mathbf{C}^{-1} \mathbf{G})^{-1} \mathbf{G}^T \mathbf{C}^{-1} \mathbf{u}]^{-1} \quad (5.75)$$

In the case the covariance matrix is diagonal with equal elements we have

$$\begin{aligned} [\mathbf{J}^{-1}]_{2,2} &= \frac{\sigma_e^2}{N} \left[1 - \frac{1}{N} \mathbf{u}^T \mathbf{G} (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{u} \right]^{-1} \\ &= \frac{\sigma_e^2}{N} \left[1 - \frac{1}{N} \left[\sum_{k=1}^N \frac{\partial h(\mathbf{x})}{\partial x_1} \quad \sum_{k=1}^N \frac{\partial h(\mathbf{x})}{\partial x_2} \right] \right. \\ &\quad \cdot \left. \begin{bmatrix} \sum_{k=1}^N \left[\frac{\partial h(\mathbf{x})}{\partial x_1} \right]^2 & \sum_{k=1}^N \frac{\partial h(\mathbf{x})}{\partial x_1} \frac{\partial h(\mathbf{x})}{\partial x_2} \\ \sum_{k=1}^N \frac{\partial h(\mathbf{x})}{\partial x_2} \frac{\partial h(\mathbf{x})}{\partial x_1} & \sum_{k=1}^N \left[\frac{\partial h(\mathbf{x})}{\partial x_2} \right]^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{k=1}^N \frac{\partial h(\mathbf{x})}{\partial x_1} \\ \sum_{k=1}^N \frac{\partial h(\mathbf{x})}{\partial x_2} \end{bmatrix} \right]^{-1} \end{aligned} \quad (5.76)$$

which gives the CRLB for the variance in the estimation of d when simultaneously estimating the position. In the case the off diagonal elements in the matrix within the square brackets are zero we have

$$[\mathbf{J}^{-1}]_{2,2} = \frac{\sigma_e^2}{N} \left[1 - \frac{1}{N} \frac{\left[\sum_{k=1}^N \frac{\partial h(\mathbf{x})}{\partial x_1} \right]^2}{\sum_{k=1}^N \left[\frac{\partial h(\mathbf{x})}{\partial x_1} \right]^2} - \frac{1}{N} \frac{\left[\sum_{k=1}^N \frac{\partial h(\mathbf{x})}{\partial x_2} \right]^2}{\sum_{k=1}^N \left[\frac{\partial h(\mathbf{x})}{\partial x_2} \right]^2} \right] \rightarrow \frac{\sigma_e^2}{N} \quad (5.77)$$

for large N .

5.5.6 CRLB in the case of uncompensated sound speed gradients

The measurement of depth by a sonar is in fact a measurement of the time for a sound reflection of the sonar beam against the bottom. Therefore the depth values will depend on the actual sound speed in the water. Sound speed gradients, which are common, will bend the sonar beam and by that the sound has to travel a shorter or longer path than else. The sonar software will always take the actual sound speed and speed gradients into account and try to compensate for the speed influence but the compensation will in most cases not be fully complete and thus an additional measurement error will occur.

The influence of speed gradients on the measured depth is illustrated in Figure 5.7. Assume the sonar beam has an out going angle of β relative the vertical line and is bent along a circular line as if the speed gradient is constant. It will then hit the bottom in point A. If the sonar does not have any compensation of the speed gradient it will wrongly assume that the beam hits the bottom at point B which corresponds to the travelling time of the sound pulse and by that we will have an error in both the vertical and horizontal direction.

If we know the speed gradient we are able to determine the exact path for the beam by simple means, i.e., by ray tracing programs. The radius of curvature for a sound beam is in every point determined by the speed gradient at that point and in the case the gradient is constant the radius of curvature is constant, i.e., the beam will follow a circular path [BU91, UR83].

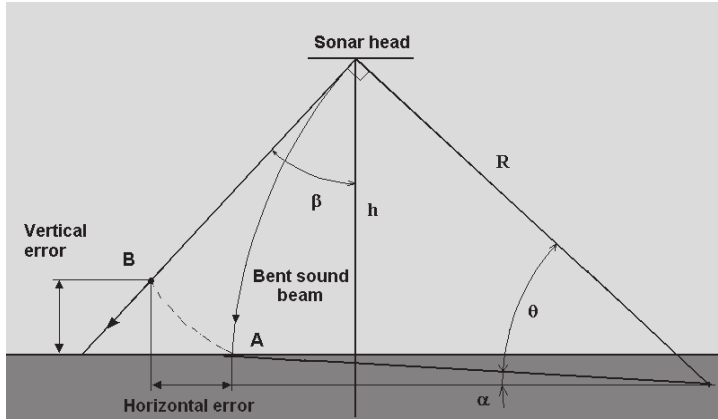


Figure 5.7: The sound beam is bent in the water due to temperature gradients.

With reference to Figure 5.7 we have

$$\sin \alpha = \sin \beta - \frac{h}{R} \quad (5.78)$$

and

$$\theta = \beta - \arcsin\left(\sin \beta - \frac{h}{R}\right) \quad (5.79)$$

and the length of the arc

$$l = \theta R \quad (5.80)$$

The depth error is thus

$$d_{err} = h \left[1 - \frac{R}{h} \cos \beta \left(\beta - \arcsin\left(\sin \beta - \frac{h}{R}\right) \right) \right] \quad (5.81)$$

and the horizontal error

$$h_{err} = (h - d_{err}) \tan \beta - R(\cos \alpha - \cos \beta) \quad (5.82)$$

We also have [BU91, UR83]

$$R = -\frac{c}{g \cos(\pi/2 - \beta)} \quad (5.83)$$

where c is the speed of sound and g is the vertical sound speed gradient.

A reasonable assumption about the temperature gradient in summer in the Baltic Sea is a linear varying water temperature from 14°C at the surface to 4°C at the bottom at 100 meter which gives a sound speed according to Figure 5.8. This will give a vertical sound speed gradient as in Figure 5.9. Assuming a vertical speed gradient which is the mean between the gradient at the surface and the gradient at the bottom we can calculate the depth and horizontal error according to (5.78 - 5.83) and the result is shown in Figure 5.10.

From the figure we see that the vertical error for sonar opening angles below $\sim 40^\circ$ is limited to ~ 1 meter for an uncompensated sonar. Today, sonars often have opening angles of 130° , corresponding to 65° in the figure. The conclusion is that if the vertical temperature gradient is known then the bending of the sound beam can be accurately compensated. This is also always done in

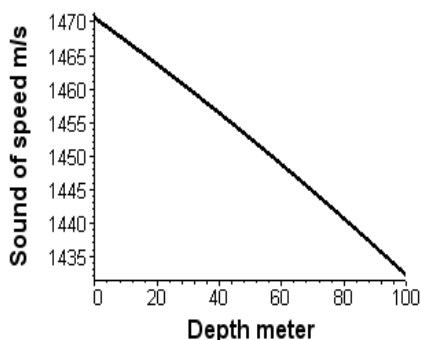


Figure 5.8: Sound speed as a function of the depth at a linear varying temperature from 14°C at the surface to 4°C at the bottom.

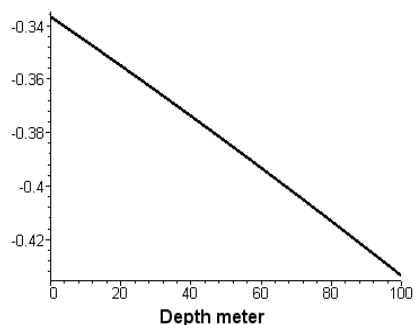


Figure 5.9: Sound speed gradient as a function of the depth at a linear varying temperature from 14°C at the surface to 4°C at the bottom.

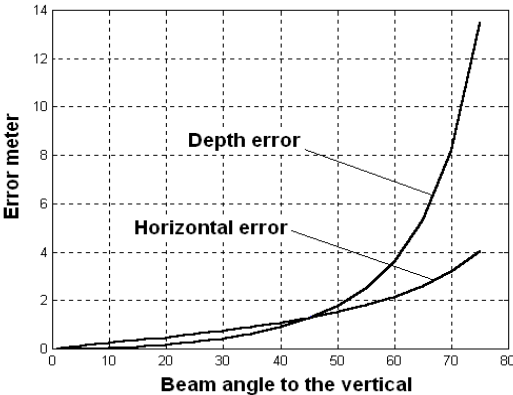


Figure 5.10: Depth and horizontal error due to uncompensated temperature gradient.

bathymetric measurements. The method used for this is a more sophisticated version of the theory in (5.78) - (5.83), see [BU91, UR83]. Still a certain small remaining stochastic error remains after the compensation and this section will deal with its effect on the lower bound off the position accuracy, i.e., the CRLB.

The vertical and horizontal errors for the different beams can be calculated according to (5.78 - 5.83), see Figures 5.11 and 5.12. More interesting is however

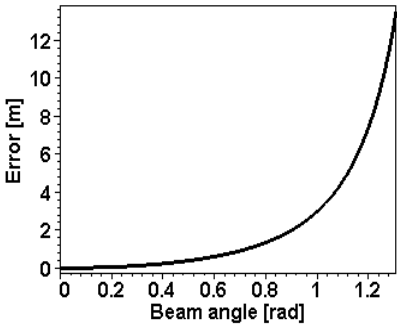


Figure 5.11: Depth error for different beams.

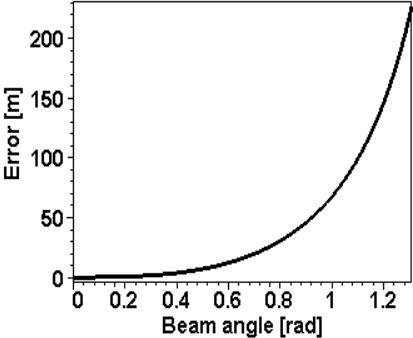


Figure 5.12: Horizontal error for different beams.

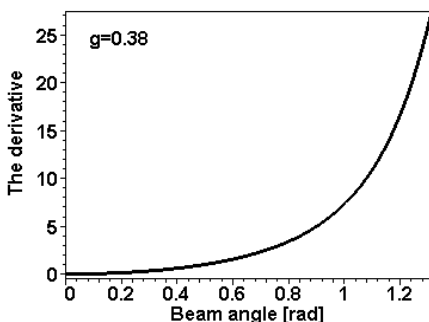


Figure 5.13: The derivative of the depth error with respect to the speed gradient for different beams.

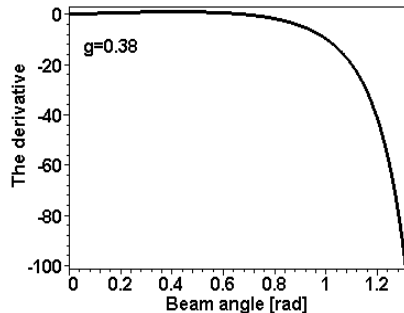


Figure 5.14: The derivative of the horizontal error with respect to the speed gradient for different beams.

the variation with the error in the sound speed gradient because that will give us the depth error for a certain gradient error. The derivatives of the sound speed gradient, g , can be seen in Figure 5.13 and 5.14.

As can be seen from the figures we have a strong correlation between the errors in the beams due to uncompensated speed gradients. After adding the depth error due to uncompensated speed gradient to previous measurement error we have for beam no. i

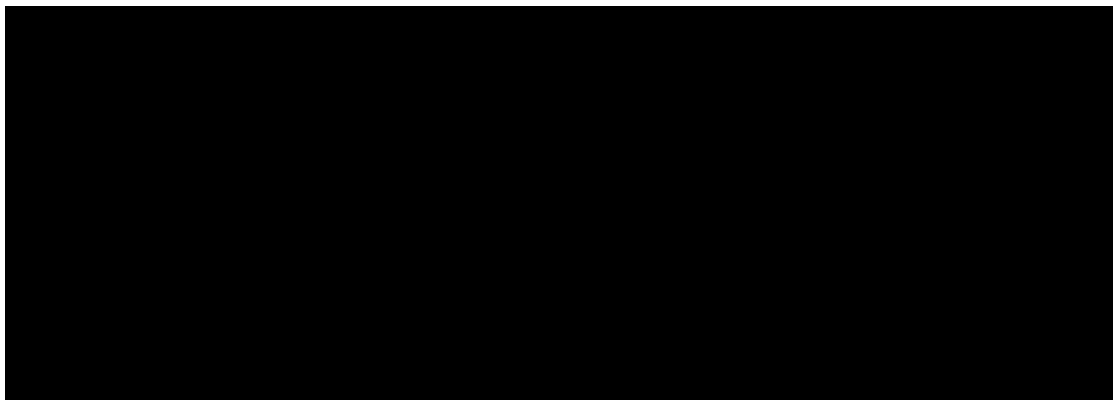
$$e_i = \varepsilon_i + \left[\frac{\partial h_i}{\partial x} \frac{\partial h_{err,i}}{\partial g} + \frac{\partial d_{err,i}}{\partial g} \right] \Delta g \quad (5.84)$$

where Δg is the gradient error. The derivative $\frac{\partial h_i}{\partial x}$ refers to the gradient of the bottom where beam i hits the bottom.

The elements in the measurement error covariance matrix for calculating the Cramér-Rao lower bound are therefore

$$\sigma_{e,i,j}^2 = \sigma_{\varepsilon,i,j}^2 + \left[\frac{\partial h_i}{\partial x} \frac{\partial h_{err,i}}{\partial g} \frac{\partial h_j}{\partial x} \frac{\partial h_{err,j}}{\partial g} + \frac{\partial d_{err,i}}{\partial g} \frac{\partial d_{err,j}}{\partial g} \right] \sigma_g^2 \quad (5.85)$$

where $\sigma_{\varepsilon,i,j}^2$ is the usual white random measurement error used in previous chapters. With reasonable assumptions about the size of the uncompensated sound speed gradient it seems that the error due to error in measuring the sound speed gradient will cause a position error almost a magnitude greater than the usual sonar measurement error.





Chapter 6

The characteristics of the position estimate

6.1 Introduction

From Chapter 4 we know that the normalized likelihood function becomes close to a Gaussian PDF when the number of measuring beams grows and it was also shown that false correlation peaks disappear. If we look at the position estimate given by the position of the maximum of the likelihood function we find it to be the maximum likelihood (ML) estimate and that the normalized likelihood function is also the PDF for the estimated position. In the case the prior position PDF, the INS position PDF, has a large error covariance, which is the case if long time has past since the last INS update, the prior will only to a negligible amount improve the ML estimated position. Since ML estimation is one of the most investigated estimation methods, there are many results that we can benefit from in establishing the accuracy of our position estimate.

6.2. Characteristics of the position estimate

We know from the theory of ML estimation that, under weak assumptions, the ML-estimate [AOA02, KA93, VT68]

- is asymptotically unbiased
- asymptotically reaches the Cramér-Rao lower bound (CRLB)
- is asymptotically Gaussian distributed

The asymptotic covariance of the ML-estimate $\hat{\mathbf{x}}(\mathbf{y})$ can therefore be expressed as

$$\mathbf{R} = E\{(\hat{\mathbf{x}}(\mathbf{y}) - \mathbf{x})(\hat{\mathbf{x}}(\mathbf{y}) - \mathbf{x})^T\} = \mathbf{J}^{-1} \quad (6.1)$$

where \mathbf{J} is Fisher's information matrix (FIM).

Let the gradient of the terrain topography in the sampling points be given by

$$\mathbf{G} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \dots & \frac{\partial h_N}{\partial x_1} \\ \frac{\partial h_1}{\partial x_2} & \dots & \frac{\partial h_N}{\partial x_2} \end{bmatrix} \quad (6.2)$$

where x_1, x_2 refer to the east and north directions, respectively, then the Cramér-Rao matrix can be shown to be (see Chapter 5)

$$\mathbf{R}_{ETL} = [\mathbf{G}\mathbf{C}^{-1}\mathbf{G}^T]^{-1} \quad (6.3)$$

where \mathbf{C} is the symmetric covariance matrix for the measurement beam errors and the sub index *ETL* indicates exact terrain reference level, i.e. no measurement bias. In the case the covariance matrix is diagonal with the same error variance in all beams i.e.,

$$\mathbf{C} = \sigma_e^2 \mathbf{I}_N \quad (6.4)$$

then

$$\mathbf{R}_{ETL} = \sigma_e^2 [\mathbf{G}\mathbf{G}^T]^{-1} = \sigma_e^2 \begin{bmatrix} \sum_{i=1}^N \left[\frac{\partial h_i}{\partial x_1} \right]^2 & \sum_{i=1}^N \frac{\partial h_i}{\partial x_1} \frac{\partial h_i}{\partial x_2} \\ \sum_{i=1}^N \frac{\partial h_i}{\partial x_2} \frac{\partial h_i}{\partial x_1} & \sum_{i=1}^N \left[\frac{\partial h_i}{\partial x_2} \right]^2 \end{bmatrix}^{-1} \quad (6.5)$$

where index i refers to the derivative at beam i .

Our aim is now to use the asymptotic position error covariance matrix, i.e., \mathbf{R}_{ETL} , in the linear Kalman filter as the true covariance matrix but this is of course only acceptable if we are close to the asymptotic covariance matrix which often will be the case. Another requirement is that our estimation error is small since the covariance matrix should be calculated in the true position. The latter will be shown to be true in Chapter 8. Figure 6.1 shows the matrix element (1,1) in \mathbf{R}_{ETL} against the number of measuring beams in a positioning in a sea-trial (see Chapter 8). As can be seen the convergence to the asymptotic value is fast, even at 15 - 20 beams we are already close to the asymptotic value. Figure 6.2 shows the convergence towards the Gaussian curve for different numbers of beams and beam planes. Even here 15 beams would be acceptable. Of course it may be cases when the Gaussian approximation may seem not so acceptable as a whole but the important thing is that the likelihood function at its maximum point can

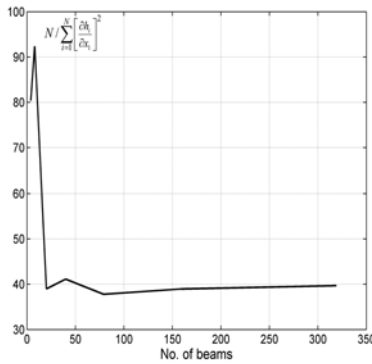


Figure 6.1: The convergence of the Cramér-Rao bound as a function of the number of measurement beams. Data from way-point 3 in the October 2002 sea-trial.

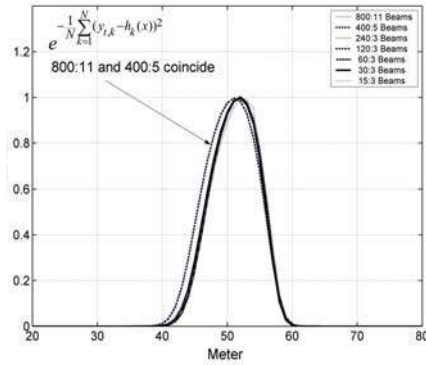


Figure 6.2: Comparison between the N normalized correlation functions for different numbers of the measuring beams. Data from way-point 3 in the October 2002 sea-trial.

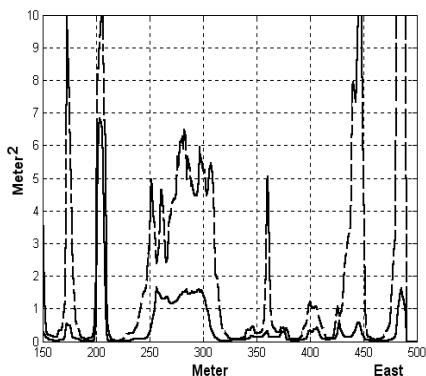


Figure 6.3: The largest and the smallest eigenvalues of the covariance matrix \mathbf{R} along a track in the map, $\sigma_e^2 = 1$, $N=100$ beams.

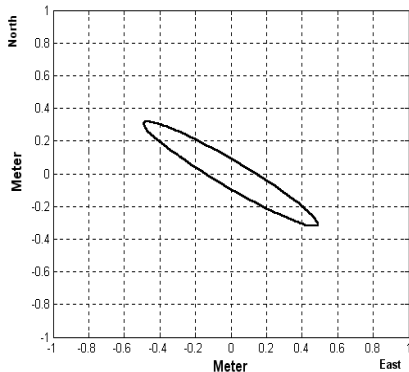


Figure 6.4: The covariance matrix at WP 3 in the sea trial. To every point in the map an uncertainty ellipse can be assigned.

be well approximated with a second order curve. The shape of the likelihood curve far from the maximum point is less important, see also the discussion in Section 3.3.3. Figure 6.3 shows the largest and smallest eigenvalue along a track in an underwater map and Figure 6.4 shows the shape of the position uncertainty ellipse at way-point WP 3 in the October 2002 sea-trial.

The expression (6.5) requires that we know the exact reference level of the sea surface (ETL) which may not be the case since the level varies with time due to many reasons as wind, wind directions, high and low pressure and the moon cycle etc.. It is easy to show that a bias in the depth measurements in general will cause a displacement of the maximum of the likelihood function (see Figure 6.5) for small N . It will be shown that large N (large correlation area) will make the estimate unbiased so the more measurement beams we have the more robust against position bias will our position estimate be.

Let us in order to more specifically understand the consequence of the measurement bias factorize the likelihood function for the one dimensional case. Assume the measured data have a bias that we are unaware of, i.e., the data are generated as $y_i(x) = h_i(x) + d + e_i$ where e_i is white noise, d is the measurement bias and that our measurement of N samples is at position x_0 , the true position. We can rewrite the likelihood function in the following way in the one dimensional case (ignoring the normalizing constant)

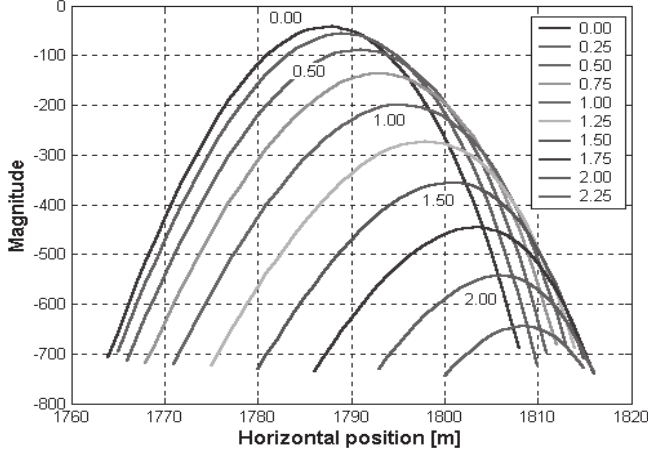


Figure 6.5: The log likelihood function at different measurement bias in meter for a relatively flat area. The number of beams $N = 100$, variance in measurement noise 0.25 m^2 . The figure is based on a real map.

$$\begin{aligned}
 L(x; y(x_0), d, N) &= \exp\left(-\frac{1}{2\sigma_e^2} \sum_{i=1}^N ((y_i(x_0) - h_i(x) - d) + d)^2\right) \\
 &= \exp\left(-\frac{1}{2\sigma_e^2} (Nd^2 + 2d \sum_{i=1}^N (y_i(x_0) - h_i(x) - d))\right) \\
 &\quad \cdot \exp\left(-\frac{1}{2\sigma_e^2} \sum_{i=1}^N (y_i(x_0) - h_i(x) - d)^2\right) \\
 &= L_h(x; y(x_0), d, N) L_{UB}(x; y(x_0), N)
 \end{aligned} \tag{6.6}$$

We see that the factor $L_{UB}(x; y(x_0), N)$ does not depend on d but is the likelihood function if we had no measurement bias. The first factor can be further factorized

$$\begin{aligned}
 L_h(x; y(x_0), d, N) &= \exp\left(\frac{Nd\bar{h}(x)}{\sigma_e^2}\right) \cdot \exp\left(-\frac{Nd(\bar{y}(x_0) - d/2)}{\sigma_e^2}\right) \\
 &= L_{h\bar{h}}(x; \bar{h}(x), d, N) L_{h\bar{y}}(x; \bar{y}(x_0), d, N)
 \end{aligned} \tag{6.7}$$

where $\bar{y}(x_0)$ is the mean of the measurements at position x_0 and $\bar{h}(x)$ the mean of depths at position x thus indicating that the bias depends on the slope of the bottom. It is easy to determine the displacement in position caused by the measurement bias d in an actual case given the values of d and N . We have

$$\begin{aligned} b(x_0 | y(x_0), d, N) &= \hat{x}_B(y(x_0)) - \hat{x}_{UB}(y(x_0)) \\ &= \arg \max_x (L(x | y(x_0), d, N)) \\ &\quad - \arg \max_x (L(x | y(x_0), d, N) / L_h(x | y(x_0), d, N)) \end{aligned} \quad (6.8)$$

which is easily computed. An approximate solution can also be found, when the unbiased likelihood function can be approximated by a Gaussian curve in the following way. Replace in (6.6) L_{UB} by its Gaussian curve equivalent assuming the maximum point is at $x = x_0$

$$L_{UBG} = \exp\left(-\frac{(x - x_0)^2}{2\sigma_e^2 / \sum_{i=1}^N \left[\frac{\partial h_i(x_0)}{\partial x}\right]^2}\right) \quad (6.9)$$

and use that

$$\frac{\partial \log L(x)}{\partial x} = \frac{\partial \log L_{UBG}(x)}{\partial x} + \frac{\partial \log L_h(x)}{\partial x} = 0 \quad (6.10)$$

at the maximum point. The result is

$$b(x_0 | d, N) = d \frac{\sum_{i=1}^N \frac{\partial h_i(x_0)}{\partial x}}{\sum_{i=1}^N \left[\frac{\partial h_i(x_0)}{\partial x}\right]^2} \quad (6.11)$$

or to show the dependence of the mean bottom slope

$$b(x_0 | d, N) \propto d \frac{\partial \bar{h}(x_0)}{\partial x} \quad (6.12)$$

This can be seen as an amplification factor of the measurement bias. Now, (6.7)

can also be written

$$\begin{aligned}
 L_h(x; y(x_0), d, N) \\
 &= \exp\left(-\frac{1}{2\sigma_e^2} \left(Nd^2 + 2d \sum_{i=1}^N (y_i(x_0) - h_i(x) - d)\right)\right) \\
 &\approx \exp\left(-\frac{1}{2\sigma_e^2} \left(Nd^2 + 2d(x_0 - x) \sum_{i=1}^N \frac{\partial h_i(x_0)}{\partial x}\right)\right)
 \end{aligned} \tag{6.13}$$

for large N and x close to x_0 . The width of the peak of the likelihood function with no measurement bias will decrease inversely with N so if we look at the neighbourhood of the peak, the first term of the exponent will dominate which means that the unbiased likelihood function is multiplied by a constant, that is, the likelihood function with measurement bias will for large N give a position estimate with no offset from the true position and a covariance as (6.5).

The bias problem can also be analyzed somewhat differently (cf. Section 5.4). A result from large sample theory [BS02] is that if a stochastic variable is asymptotically normal distributed $\mathbf{x} \stackrel{a}{\sim} \mathcal{N}(\mathbf{x}_0, \mathbf{R})$, as in our case, and undergoes a one to one transformation $\mathbf{z} = \mathbf{B}(\mathbf{x})$ then the transformed variable is also asymptotically normal distributed

$$\mathbf{z} \stackrel{a}{\sim} \mathcal{N}(\mathbf{z}_0 = \mathbf{B}(\mathbf{x}_0), \left[\frac{\partial \mathbf{B}(\mathbf{x}_0)}{\partial \mathbf{x}}\right]^T \mathbf{R} \frac{\partial \mathbf{B}(\mathbf{x}_0)}{\partial \mathbf{x}}) \tag{6.14}$$

A requirement is that the largest eigenvalue of $\mathbf{R} \rightarrow 0$ when the number of samples (beams) $N \rightarrow \infty$.

In the bias discussion in Section 5.4 the mean was assumed to be the position estimate. However we can deduct a similar expression when we are assuming the position of the maximum of the likelihood function to be the measured position. Let us again look at the one dimensional case and let us assume that the measured height in each beam can be described by

$$y_i(x) = h_i(x) + d + e_i \tag{6.15}$$

The ML position estimate is given by the solution of the equation

$$\begin{aligned}
 f(x) &= \frac{\partial}{\partial x} (\log L(x; y(x_0))) \\
 &= \frac{1}{\sigma_e^2} \left[\sum_{i=1}^N [h_i(x_0) - h_i(x)] \frac{\partial h_i(x)}{\partial x} + \sum_{i=1}^N (d + e_i) \frac{\partial h_i(x)}{\partial x} \right] = 0
 \end{aligned} \tag{6.16}$$

The function $f(x)$ is almost linear around $x=x_0$ for moderate N . By Taylor expanding $f(x)$ around $x=x_0$ the following expression for the position error b can be found

$$b \approx \frac{\sum_{i=1}^N (d + e_i) \frac{\partial h_i(x_0)}{\partial x}}{\sum_{i=1}^N \left[\frac{\partial h_i(x_0)}{\partial x} \right]^2 - \sum_{i=1}^N (d + e_i) \frac{\partial^2 h_i(x_0)}{\partial x^2}} \tag{6.17}$$

Thus the position error can be expressed as the ratio of two stochastic variables. The variance of the denominator

$$\sigma_e^2 \sum_{i=1}^N \left[\frac{\partial^2 h_i(x_0)}{\partial x^2} \right]^2 \tag{6.18}$$

is typically much smaller than the variance of the numerator

$$\sigma_e^2 \sum_{i=1}^N \left[\frac{\partial h_i(x_0)}{\partial x} \right]^2 \tag{6.19}$$

This can be seen by looking at their expected values. For a stochastic variable $h(x)$ the expected value of the square of a derivative of the variable is [NE75]

$$\mu(n) = E \left\{ \left[\frac{\partial^n h}{\partial x^n} \right]^2 \right\} = \int_0^\infty \omega^{2^n} S_h(\omega) d\omega \tag{6.20}$$

The spectrum of a terrain profile can often be assumed to have an exponential form $S_h(\omega) = S_0 \exp(-\omega / \omega_0)$ where ω_0 is a characteristic of the actual terrain.

In such a case we have $\mu(1) = 2S_0\omega_0^3$ and $\mu(2) = 24S_0\omega_0^5$. The value of ω_0 is in most cases much less than 0.02 which shows that the variance of the denominator is much less than the variance of the numerator and thereby the approximations below are justified (cf. Figure 6.6).

The variance of the denominator is much smaller than the mean, thus the posi-

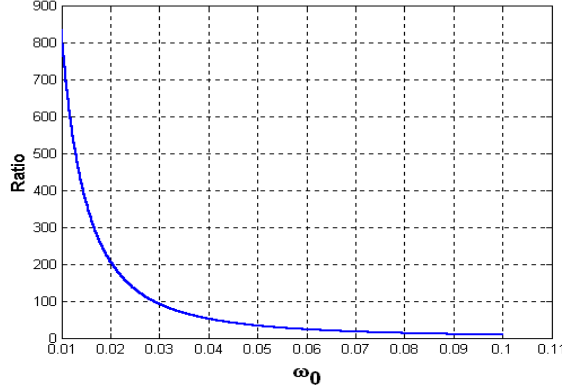


Figure 6.6: The ratio between $\mu(1)$ and $\mu(2)$ as a function of the fall off constant ω_0 .

tion error will be close to Gaussian distributed. Disregarding the stochastic part of the denominator the variance will be

$$\begin{aligned}
 Var(b) &\approx \frac{\sigma_e^2}{\sum_{i=1}^N \left[\frac{\partial h_i(x_0)}{\partial x} \right]^2} \frac{1}{\left[1 - d \sum_{i=1}^N \frac{\partial^2 h_i(x_0)}{\partial x^2} / \sum_{i=1}^N \left[\frac{\partial h_i(x_0)}{\partial x} \right]^2 \right]^2} \\
 &\approx \frac{\sigma_e^2}{\sum_{i=1}^N \left[\frac{\partial h_i(x_0)}{\partial x} \right]^2} \left[1 + d \sum_{i=1}^N \frac{\partial^2 h_i(x_0)}{\partial x^2} / \sum_{i=1}^N \left[\frac{\partial h_i(x_0)}{\partial x} \right]^2 \right]^2 \quad (6.21)
 \end{aligned}$$

See Figure 6.7 which compares (6.17) with a Monte Carlo simulation of the position bias.

An accurate approximation of the bias for larger N is

$$E\{b\} = \bar{b} \approx d \frac{\sum_{i=1}^N \left[\frac{\partial h_i(x_0)}{\partial x} \right]}{\sum_{i=1}^N \left[\frac{\partial h_i(x_0)}{\partial x} \right]^2} \quad (6.22)$$

and approximately we have

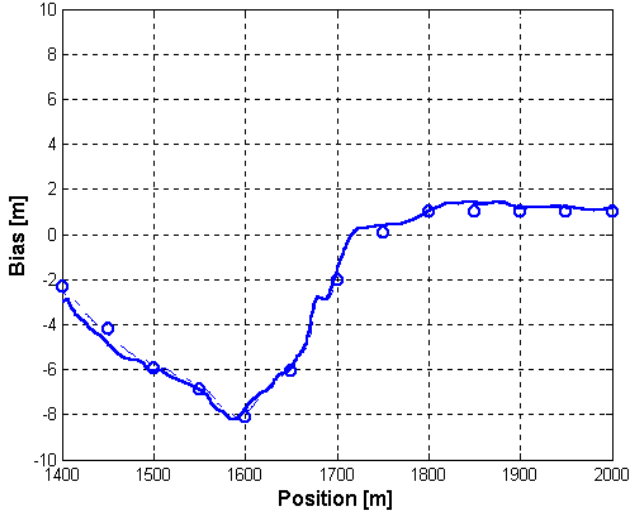


Figure 6.7: Position bias in a real map calculated according to (6.17) and (6.22). The circles represent the result from Monte Carlo simulations. The solid line refer to (6.17) and the dashed line to (6.22). Number of beams is $N=400$, measurement bias $d=-2$ m and $\sigma_e=0.5$ m.

$$\frac{\partial \bar{b}}{\partial x} \approx d \sum_{i=1}^N \frac{\partial^2 h_i(x_0)}{\partial x^2} \bigg/ \sum_{i=1}^N \left[\frac{\partial h_i(x_0)}{\partial x} \right]^2 \quad (6.23)$$

That is, according to (6.21)

$$Var(b) \approx \sigma_e^2 \left[1 + \frac{\partial \bar{b}}{\partial x} \right]^2 \bigg/ \sum_{i=1}^N \left[\frac{\partial h_i(x_0)}{\partial x} \right]^2 \quad (6.24)$$

See also Figure 6.8.

The above results can easily be generalized to the two dimensional case. For moderate N and small and equal measurement bias d in all beams equation (6.22) can thus be approximated by

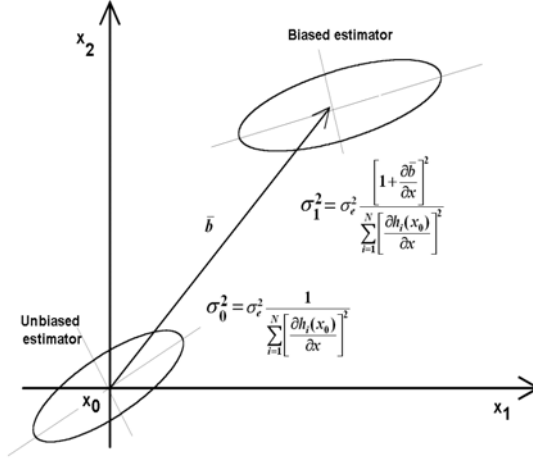


Figure 6.8: The uncertainty ellipses for an unbiased and a biased position estimator. The Cramér-Rao variance expressions refer to the one dimensional case. The origin of the coordinate system is in position \mathbf{x}_0 .

$$\mathbf{b}(\mathbf{x}_0 | d, N) = d \cdot N \cdot \mathbf{R}_{ETL}^{-1} \bar{\mathbf{G}}(\mathbf{x}_0) \quad (6.25)$$

where

$$\bar{\mathbf{G}}(\mathbf{x}) = \frac{1}{N} \begin{bmatrix} \sum_{i=1}^N \left[\frac{\partial h_i}{\partial x_1} \right] \\ \sum_{i=1}^N \left[\frac{\partial h_i}{\partial x_2} \right] \end{bmatrix} \quad (6.26)$$

That is, the horizontal position bias is proportional to the mean gradient $\bar{\mathbf{G}}(\mathbf{x}_0)$ of the terrain which would be expected. For large N the mean gradient will tend to zero and making the estimate unbiased.

Thus, in the case we have a measurement bias the variance will be given as shown by (6.24) which in the two dimensional case can be written as

$$\mathbf{R}_{Bias} \approx \frac{\partial \mathbf{B}(\mathbf{x}_0)^T}{\partial \mathbf{x}} \mathbf{R}_{ETL} \frac{\partial \mathbf{B}(\mathbf{x}_0)}{\partial \mathbf{x}} \quad (6.27)$$

where $\mathbf{B}(\mathbf{x}) = \mathbf{x} + \mathbf{b}(\mathbf{x})$.

In the estimation of the position the existence of measurement bias can of course be taken into account from the very beginning. This has been shown in Chapter 5. Often, however, we can approximate the measurement bias to be the difference between the means of the measured depths and corresponding map depths as pointed out earlier. The main difference between this approximation and the more correct method (3.20) in Chapter 3 is that the first assumes that the likelihood maximum occur at a step point, which also should be the true position, in the matching which may not be the case. With the number of measurement beams we are using the difference is negligible.

The reduction of the measurement and map vector with the means can be described as pre-multiplying the profile vectors $\mathbf{y}(\mathbf{x}_0)$ and $\mathbf{h}(\mathbf{x})$ by the projection matrix which will subtract respective mean from the vectors. The projection matrix is

$$\mathbf{\Pi} = \mathbf{I}_N - \mathbf{1}_{N,N}/N \quad (6.28)$$

where \mathbf{I}_N is the identity matrix and $\mathbf{1}_{N,N}$ is a matrix with ones as elements. Straightforward calculations of the second derivative, see Chapter 5, of the log likelihood function for this case give the CRLB matrix (cf. (6.5)).

$$\mathbf{R}_{PM} = \sigma_e^2 [\mathbf{G}\mathbf{\Pi}\mathbf{G}^T]^{-1} \quad (6.29)$$

When profile matching gives a unique position and we have a large number of beams expressions (6.5) and (6.29) will be close to each other.

In the general case when we have a measurement bias and a full measurement error covariance matrix the CRLB for the position error is as shown in Chapter 5

$$\mathbf{R} = [\mathbf{G}\mathbf{C}^{-1/2} [\mathbf{I} - \mathbf{C}^{-1/2} \mathbf{u}(\mathbf{u}^T \mathbf{C}^{-1} \mathbf{u})^{-1} \mathbf{u}^T \mathbf{C}^{-1/2}] \mathbf{C}^{-1/2} \mathbf{G}^T]^{-1} \quad (6.30)$$

where \mathbf{u} is a column vector of N ones and \mathbf{C} the position error covariance matrix

and \mathbf{G} defined by (6.2). Since the elements in the covariance matrix of the position error correspond to the sum of the terrain gradients in the measuring footprint points, this also means that the method with several measuring beams can be used in much flatter terrain areas than the measuring with only one beam since adding of beams will compensate for the smaller terrain gradients.

The position error covariance matrix for the estimate can thus be calculated in real time or in advance from the terrain map. We see from (6.5) that the error covariance is approximately inversely proportional to the number of beams. In the case we know the autocorrelation function for the bottom profile the estimated value of the position error covariance matrix is just a function of N .





Chapter 7

Optimal beampattern and way-point selection

7.1 Introduction

The first part of this chapter will discuss what is called optimal footprint beam pattern before turning to the problem of choosing way-points. We will restrict our discussion to rectangular beam patterns since these naturally arise from bathymetric sonars with a linear or square sensor array which is the common array shape. It turns out that the position accuracy could in some cases be considerably increased by adjusting the footprint of the beam pattern to the structure of the bottom terrain which is a simple thing to do if a true 3D bathymetric sonar is used. The second part of the chapter will discuss how to choose way-points. Since it is possible to determine the position accuracy directly from the underwater map, this information can be used in planning vehicle missions to secure that good position fixes will be achieved.

7.2 Optimal footprint beam pattern

The achievable position accuracy is given by the Cramér-Rao lower bound expression (6.1). It is not obvious from (6.1) that the actual accuracy strongly

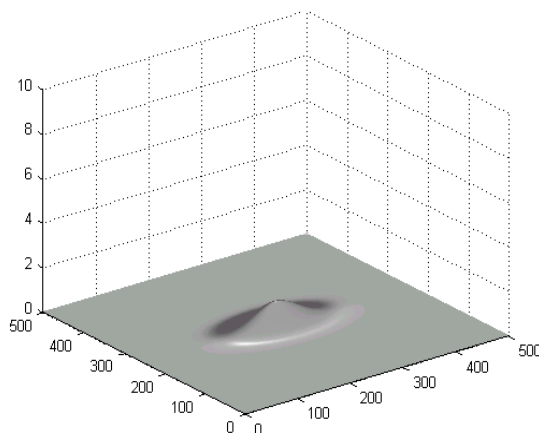


Figure 7.1: Flat terrain with a Gaussian hill. The true position is in the centre of the hill.

depends on the beam footprint array pattern in many cases. A simple example will show this.

Assume that we have a non-circular Gaussian hill as in Figure 7.1 and our true position at the sampling event is at the top of the hill. Correlation is then done by an array of 20×20 beams and by an array of 10×40 beams. The same number of beams is thus used in both cases. The contour plots for the likelihood functions in the two cases are shown in Figure 7.2 and 7.3. The contour of the likelihood function will be elliptical for a square beam pattern. The semiaxis of the uncertainty ellipse in the center of the Gaussian hill are given by the eigenvectors and eigenvalues of the Cramér-Rao matrix in that point. By comparing Figure 7.2 with Figure 7.3 we see a great improvement in accuracy in the worst direction at the expense of the accuracy in the best direction.

Figure 7.4 and 7.5 show a real example from a sea-trial when the way-point is situated on the slope of an underwater ridge. In Figure 7.4 the beam array pattern is 10×10 in a rectangular grid of 10 meter. The angle between the grid and the ridge is approximately 30 degrees. In Figure 7.5 the beam pattern is 50×2 beams and the pattern is rotated 30 degrees to align the long side with the ridge.

Figure 7.6 shows the magnitude of the largest semiaxis for different footprint

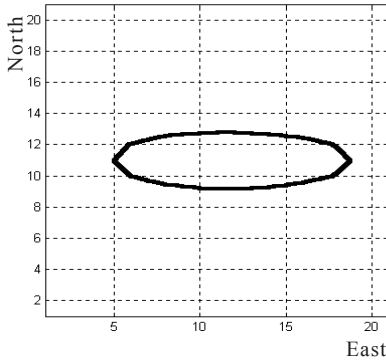


Figure 7.2: Uncertainty ellipse for an array pattern of 20 x 20 beams at 80% height of the maximum of likelihood function.

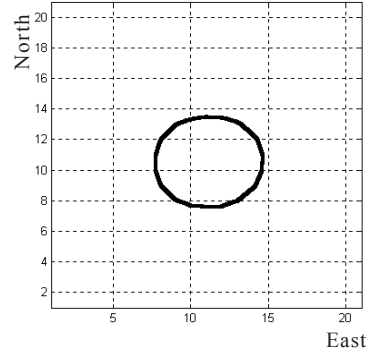


Figure 7.3: Uncertainty ellipse for an array pattern of 10 x 40 beams at 80% height of the maximum of likelihood function.

patterns, from 1x256 to 16x16, when the beampattern is rotated 360 degrees. For a quadratic beam pattern there is only small variation in the magnitude of the largest semiaxis but for the 2 x 128 beam pattern we have a radically smaller largest semiaxis for a rotation angle of 150 degrees.

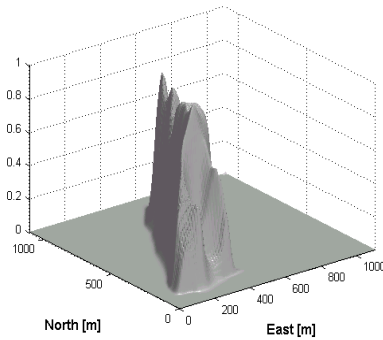


Figure 7.4: The likelihood function for a way-point situated at the slope of a ridge.

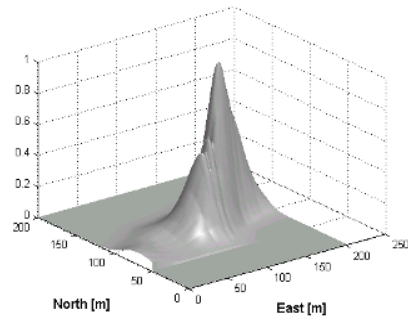


Figure 7.5: The likelihood function when the beam pattern has been adjusted and rotated. Note the difference in horizontal scale compared to Figure 7.4.

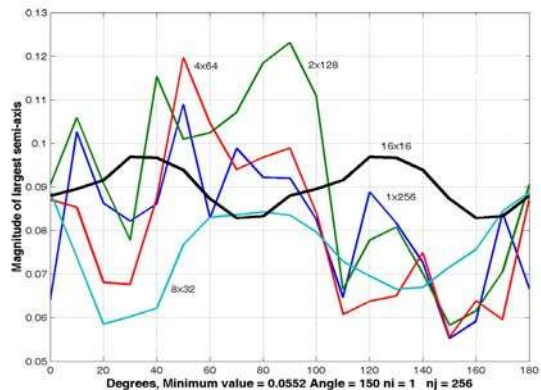


Figure 7.6: The figure shows an example of how the magnitude of the largest semiaxis varies with the beam pattern for a position in an underwater map.

The discussion above indicates that a higher accuracy in the worst direction can be achieved in many cases by adjusting the beam footprint pattern shape after the expected bottom appearance at the way-point. This can be done in the following way for a measuring system which only uses beamforming in the receiver stage. The data from the physical array are first beamformed into a square beam pattern and the eigenvalues of the Cramér-Rao matrix in that point are calculated at the maximum point. A new beam pattern based on these eigenvalues is then laid down and a new likelihood function is calculated using the

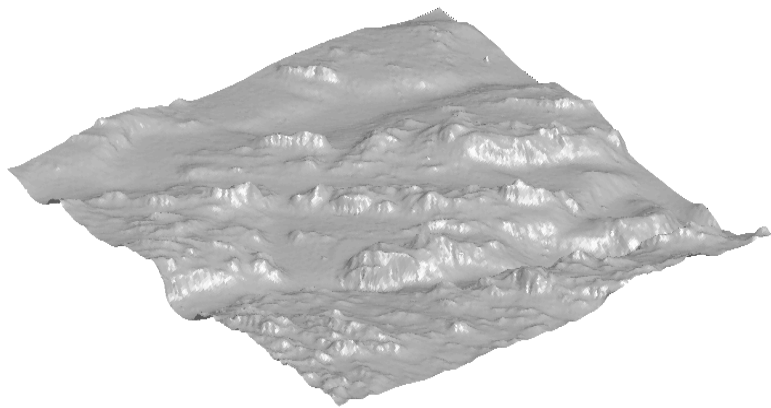


Figure 7.7: A real underwater map of size 3 km x 3 km. The vertical scale is approximately 25 times enlarged.

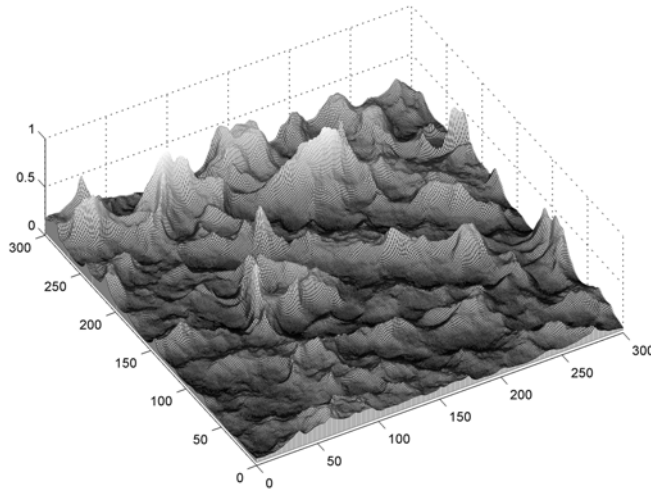


Figure 7.8: The map shows the magnitude of the largest semiaxis for a fixed beam footprint pattern of 16×16 beams for a real map.

same data. There will of course be many places where an approximation with a Gaussian hill is not so good but an optimal beam footprint pattern can always be found by adjusting the side relation and the rotation of the beam footprint pattern (see Figure 7.6).

Now we will look at a real map, see Figure 7.7, to see what this will mean in reality. Figure 7.8 shows a map of the largest semiaxis for a fixed beam footprint pattern of 16×16 beams for this map. Figure 7.9 shows the same map but now with the optimal beampattern, i.e., the beam pattern varies from 2×128 to 16×16 and is rotated to achieve the smallest largest semiaxis of the uncertainty ellipse. As can be seen the topography is now more chiselled and the magnitude is much lower in many areas. Note the difference in vertical scale compared to Figure 7.8. The figures are based on the same number of beams. Figure 7.10 shows the percentage of the total map area which has the largest semiaxis below a certain value. For example, if a magnitude of the largest semiaxis of 0.1 is acceptable then the optimal technique would more than double the map area where that accuracy, or better, could be achieved. As can be seen from the figures there is great improvement in certain areas and less in other areas.

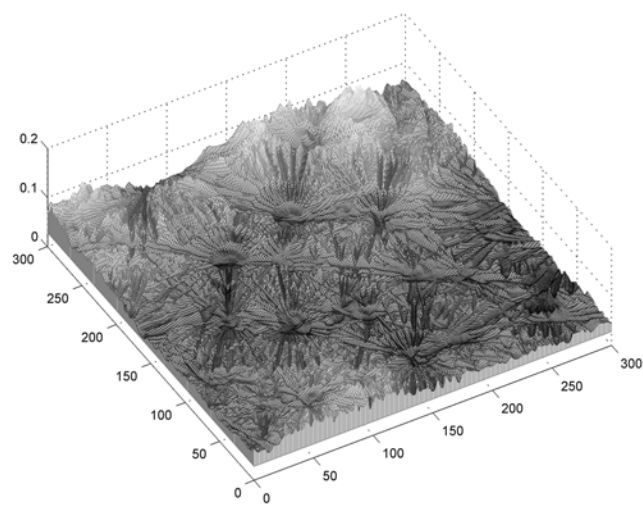


Figure 7.9: The map shows the magnitude of the largest semi-axis for the same map as used in Figure 7.8 but now with the optimal beam pattern technique used.

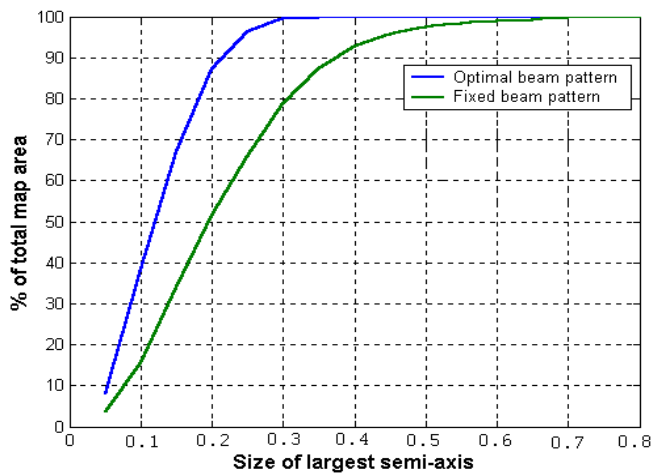


Figure 7.10: The graph shows the percentage of the map area which has the largest semi-axis below a certain value. For example if the largest semi-axis should be below 0.1 the optimal beam footprint technique would more than double the map area where this is the case.

The information in the maps in Figure 7.8 and 7.9 can also be used in the planning of the mission for the underwater vehicle to secure good position fixes.

7.3 How to choose way-points

Way-points and tracks shall of course be placed where the terrain has large relative height variations but since often maps are non-homogeneous in terrain characteristics a more sophisticated method of selecting way-point areas is needed and the findings in the previous section will be of great help for that.

The Figures 7.8 and 7.9 show the magnitude of the largest semiaxis of the uncertainty ellipse for a fixed beam pattern and for an optimal beam pattern. Small magnitudes indicate small position error due to random measurement errors and hence good terrain for terrain navigation. Large magnitudes mean poor conditions for terrain navigation. Consequently way-points shall be placed in low areas of these maps and tracks should be laid down to follow valleys or low areas in the maps in case of continuous terrain navigation.

The graph in Figure 7.11 shows the optimal beam pattern overlaid the contour plot for the map in Figure 7.7. The rectangles symbolize the beam footprint pattern. If we look closer at a part of that map, Figure 7.12, we have Figure 7.13 showing a map of the smallest largest semiaxis in the uncertainty ellipse while Figure 7.14 shows the optimal beam pattern overlaid a contour plot of Figure 7.12. From Figure 7.13 and Figure 7.14 we can see that we will have the highest accuracy if we do our positioning in the valley shown in Figure 7.13. This is also the valley in Figure 7.12 but this should not be taken as a rule. A smaller beam pattern would change the situation.

A reasonable first criteria for positioning failure is that the largest semiaxis, for the selected beam pattern falls above some critical value, a_{cr} . In Figure 7.15 the shaded area shows such a situation for the map in Figure 7.7. We define a likelihood function L_{wp} by assigning to it the value 0 when the largest eigenvalue exceeds the criterion (the shaded area in the figure) and the value 1 otherwise (the white area). The percentage of localizations that will succeed given the criterion a_{cr} is then given by

$$F(\mathbf{x}_{wp} | a_{cr}) = \int_{R^2} p_{wp}^-(\mathbf{x}) L_{wp}(\mathbf{x} | a_{cr}) d\mathbf{x} \quad (7.1)$$

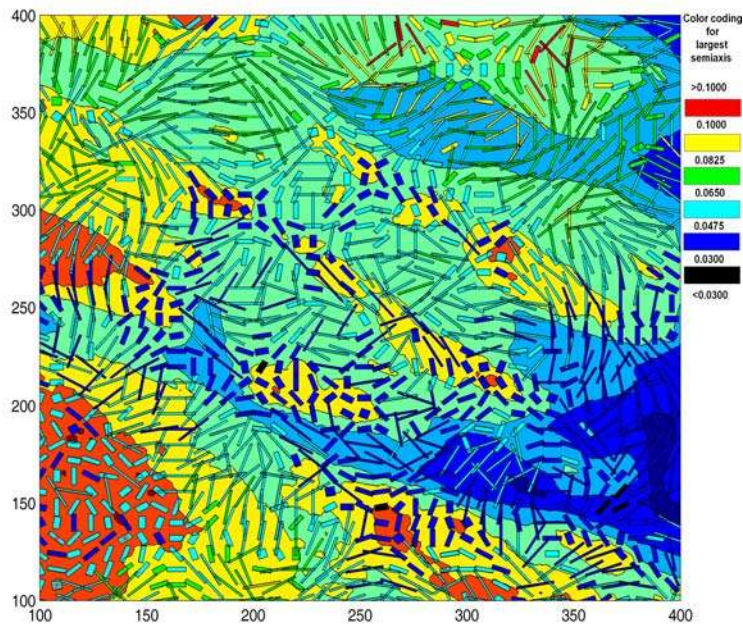


Figure 7.11: Optimal beam pattern overlaid the contour plot of a map. Darker beam patterns are good areas for terrain positioning. Light (red) areas are bad for terrain positioning. The scale of the vertical and horizontal axis and size of semiaxis is in *10th* of meters.

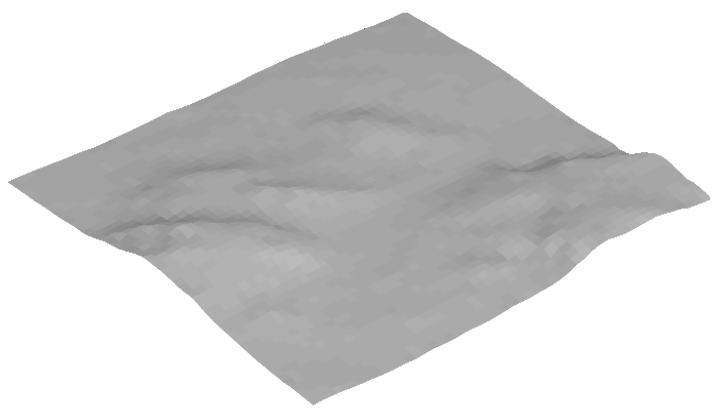


Figure 7.12: Part of the map shown in Figure 7.7. The vertical scale is enlarged approximately 50 times.

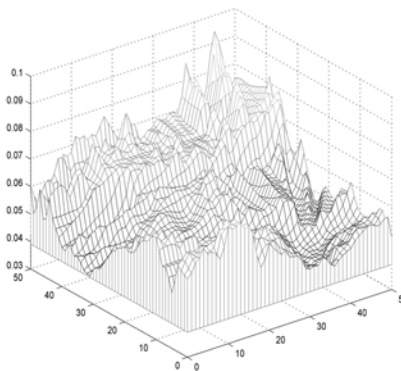


Figure 7.13: The map over the largest semiaxis in the uncertainty ellipse when the optimal beam pattern has been used, cf. Figure 7.12. The scale of the vertical and horizontal axis is in 10th of meters.

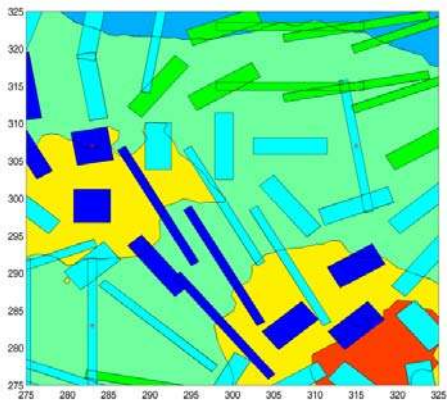


Figure 7.14: The optimal beam pattern overlaid the contour map for Figure 7.12 for determination of suitable way-points. The scale of the vertical and horizontal axis is in 10th of meters.

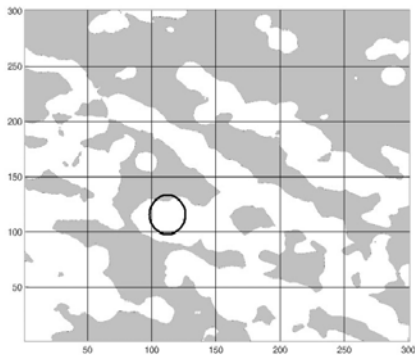


Figure 7.15: White areas are areas where the largest semiaxis of the uncertainty ellipse is smaller than a certain threshold value, i.e., 0.15.

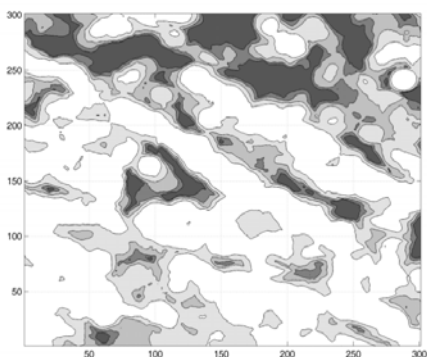


Figure 7.16: The value of the largest semi axis for the landscape in Figure 7.7. Contour lines are for 0.1, 0.15, 0.20, 0.25.

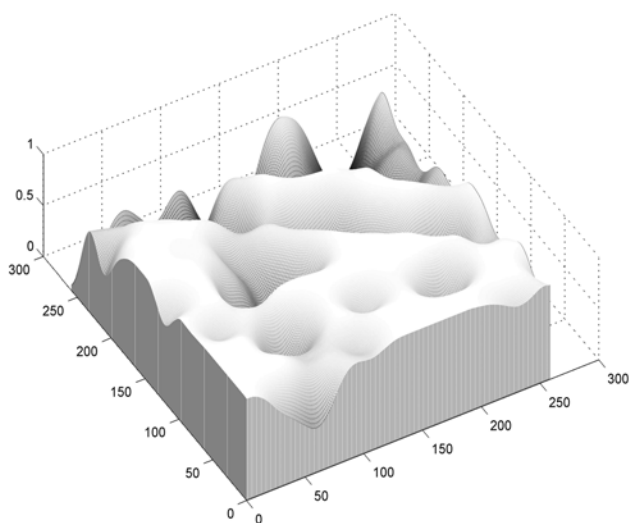


Figure 7.17: The diagram shows the probability for success for way-points in the map in Figure 7.7.

Here $p_{wp}^-(\mathbf{x})$ is the INS-system's PDF with centre in the way-point \mathbf{x}_{wp} , the prior PDF.

This is illustrated in Figure 7.15 by a possible way-point in the centre of the circle, where the size of the circle symbolizes the $1-\sigma$ INS error. Figure 7.16 shows the largest semiaxis for the landscape in Figure 7.7. If the actual arrival at the way-point is in the shaded area then positioning is impossible, the error is too large. On the other hand, if it is in the white area a position fix is possible. In such a way, for every point in the map, the percentage of success can be calculated according to (7.1), see Figures 7.17 and 7.18. The choice of a_{cr} will depend on the accuracy needed when other sources of errors have been taken into consideration (see Chapter 8).

7.3.1 An example of how to choose a way-point

Our goal is to decide upon a way-point in a given area for which we have an underwater map. Our sonar equipment is a bathymetric sonar with known characteristics as number of beams, beam angles, measurement accuracy and so forth. We will also assume that the number of measurement beams is so large that the Gaussian approximation is reasonable. The positioning result will be used for updating an INS-system. The steps for establishing the way point will be as follows.

The first step will be to calculate the position error covariance matrix for every point in the map. According to (6.1) in Chapter 6 the error covariance matrix with no beam weighting is

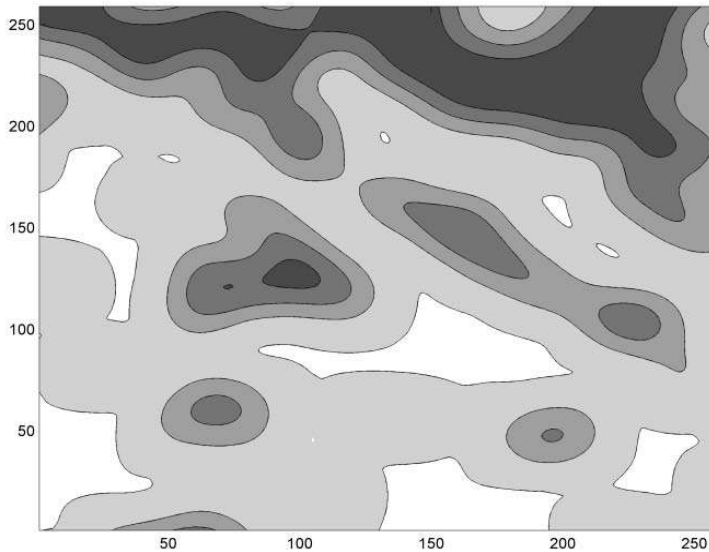


Figure 7.18: Contour map of the diagram in Figure 7.17. In the dark shaded areas the probability of success in positioning is less than 25% and in the light shaded areas the probability is almost 100 %. The step in shading is 25%. The prior PDF from the INS is assumed to be 100 meter RMS. The scale of the vertical and horizontal axes is in 10th of meters.

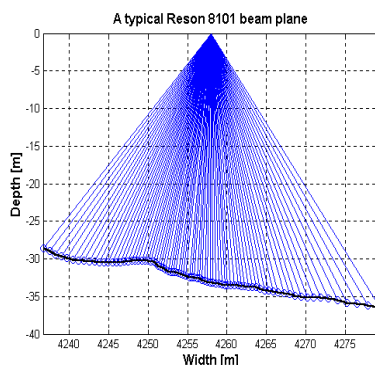


Figure 7.19: A typical beam plane.

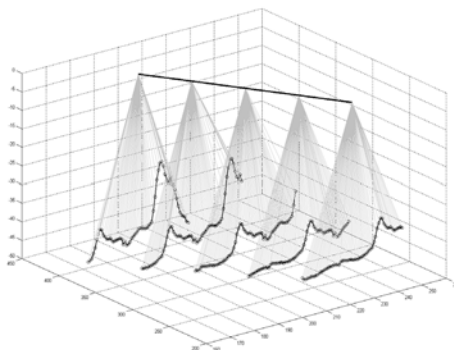


Figure 7.20: An assembled beam package of five beam planes.

$$\mathbf{R}_{ETL} = \sigma_e^2 \left[\mathbf{G} \mathbf{G}^T \right]^{-1} = \sigma_e^2 \begin{bmatrix} \sum_{i=1}^N \left[\frac{\partial h_i}{\partial x_1} \right]^2 & \sum_{i=1}^N \frac{\partial h_i}{\partial x_1} \frac{\partial h_i}{\partial x_2} \\ \sum_{i=1}^N \frac{\partial h_i}{\partial x_2} \frac{\partial h_i}{\partial x_1} & \sum_{i=1}^N \left[\frac{\partial h_i}{\partial x_2} \right]^2 \end{bmatrix}^{-1} \quad (7.2)$$

If we have a bathymetric sonar with only one beamplane we may have to assemble a number of beamplanes to measure the bottom 3D topography. A good choice of the number of beam planes by experience may be 5 with a distance between the planes of 20 - 25 meter.

When calculating the uncertainty ellipse account shall preferably be given to the actual sonar and the real footprint of the sonar against the bottom, see Figures 7.19 and 7.20. By that, the enlargement of the total footprint with increased depth will automatically be taken into account. This is important since a larger footprint coverage will mean a smaller positioning error. A way of determining the intersection of the individual beams with the bottom is to first approximate the bottom with triangles with their vertices in the grid points of the map. The intersection of the beams and the planes determined by the triangles is then calculated and by that the positions of the beam footprints and the depth for the beams are known, Figure 7.21.

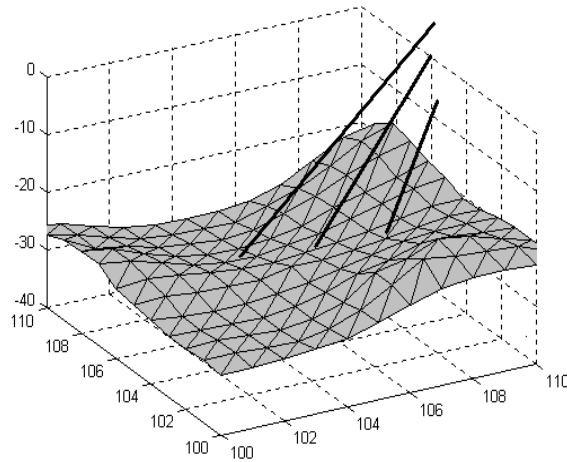


Figure 7.21: The individual beam footprint is calculated by first finding the intersection between the bottom and the beam package of five beam planes.

The chosen number of beam planes shall then be assembled with a fix distance and angle between the beam planes. The beam plane assembly is then moved in steps over the map and the depth measurements are calculated from a sonar model which can be more or less complex, i.e., taking temperature gradients and other things into account. A more approximate solution using a beam pattern with a fixed distance between the individual beam footprints, see Figure 7.22, is often of interest since it is much less computationally demanding and still gives a good view of the good areas for terrain navigation. When those are found then a more accurate computational method can be used. In both ways interpolation of the map depth has to be used. In relatively flat areas a linear interpolation method can be used but in more rough hilly terrains a more accurate method must be used (see Appendix B).

As an alternative to interpolation a sonar map method can be used, that is a local map of the sonar measurement with the same grid size as the underwater map is produced and this map is correlated with the underwater map in the gridpoints. The main complication with this method appears when the correlations shall be done with different rotation angles of the beam pattern.

The next step in the process is to calculate, in every point of the map, the

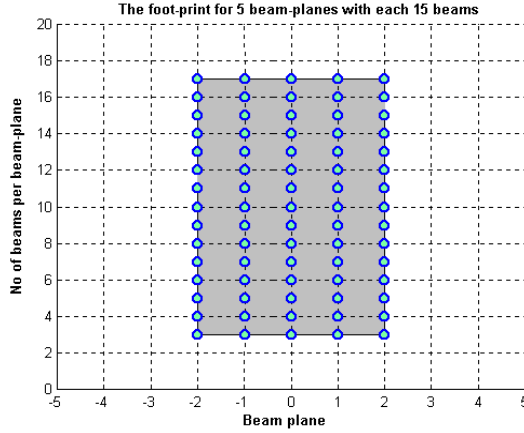


Figure 7.22: Footprint of the beam plane assembly for a flat bottom.

eigenvalues of the position error covariances. This will give the major and minor semiaxes of the uncertainty ellipse and the eigenvector gives the direction of the semiaxis and by that the orientation of the likelihood function is known. In the direction of the eigenvector corresponding to the largest eigenvalue we have the poorest accuracy and in the direction of the eigenvector corresponding to the smallest eigenvalue we have the best accuracy. By now a graph like Figure 7.13 can be drawn for the largest eigenvalue (semiaxis for the uncertainty ellipse) and it is in the low areas the best places for way-points are.

The covariance error matrix (7.2), corresponds to the likelihood function

$$L(\mathbf{x}_t; \mathbf{y}_t) = \frac{1}{\sqrt{(2\pi\sigma_e^2)^N}} \exp\left(-\frac{1}{2\sigma_e^2} \sum_{k=1}^N (y_{t,k} - h_k(\mathbf{x}_t))^2\right) \quad (7.3)$$

If the more complex likelihood function with beam weighting is used

$$L(\mathbf{x}_t; \mathbf{y}_t) = \frac{1}{\sqrt{(2\pi)^N \det(\mathbf{C}_e)}} \exp\left(-\frac{1}{2} (\mathbf{y}_t - \mathbf{h}(\mathbf{x}_t))^T \mathbf{C}_e^{-1} (\mathbf{y}_t - \mathbf{h}(\mathbf{x}_t))\right) \quad (7.4)$$

the corresponding error covariance matrix should be used in the calculations, see Chapter 5.

Now, for updating our IN-system with an accurate position for eliminating the drift error we can only accept terrain positions with a largest semiaxis smaller than a certain value so we produce a new map, the success map, of the area where the pixels are 1 if we meet the criteria or 0 if we fail, cf. Figure 7.15. This new map will now be the basis for the calculation if we will fail or succeed in establishing a terrain position with required accuracy. When the vehicle is approaching the way-point it will only occasionally hit the way-point so the actual vehicle positions, when the vehicle thinks she is at the way-point, will be spread according to the accuracy of the navigation system. Sometimes the vehicle will end up in the white areas and sometimes in the black areas.

The success rate is then calculated for every assumed way-point in the map according to (7.1) which gives the probability of success when approaching a way-point at position \mathbf{x}_{wp} . A graph of the function is shown in Figure 7.18. The final step is then to select an actual point in the light shaded areas in that figure.





Chapter 8

Sonars and the sea-trial October 2002

8.1 Introduction

This chapter starts with a discussion of sonars suitable for terrain navigation as an introduction to the presentation of some results from one of the sea-trials that have been conducted to verify the theory presented earlier in this thesis.

8.2 Sonars and position error causes

8.2.1 Introduction

The same type of sonars that are used for charting of ship routes can be used for terrain navigation. These sonars are called bathymetric sonars and have a fan-shaped beamplane with ~60 - 250 beams with beamwidths ~0.5 - 1.5 degrees.

Often such sonars work in the frequency range 100 - 200 kHz to have a reasonable range in depth. The lower the frequency is the lower will the attenuation of the sonar pulse be and by that an increased range for the sonar.

Large depths mean that the beam footprint will be large so a small beam angle is essential for mapping smaller objects at large depths.

There are many types of errors encountered in terrain navigation besides the random measurement error previously discussed. To name a few of them we have errors caused by:

- Ship movement
Pitch, roll and yaw and with surface ships also heave.
- The mounting of the sonarhead
Reflection
Incorrect assumptions about the center of vehicle movement
- Characteristics of the sonar
Beamwidth
Number of beams and spacing
Accuracy in depth measurement
Power
Algorithm for bottom detection
- Measurement errors caused by water characteristics
Insufficient compensation of temperature gradients
Reflections of the outer beams due to stratifying layers.
- Measurement errors caused by askew reflections of sonarbeams
Due to bottom slant
Due to rocks
Due to objects in the water column, i.e., fish schools

8.2.2 Bathymetric sonars

Figures 8.1 and 8.2 show two types of bathymetric sonars that have been used for terrain navigation. Figure 8.1 shows the Reson SeaBat 9001. The cylinder on the main body is the transmitting unit consisting of a cylindrical pile of discs of a piezoelectric ceramic material. To each disc a phased sonar pulse is sent and by that a fan shaped beam is formed as sketched in Figure 8.3. The physical sound wave is transmitted from the cylindrical surfaces of the ceramic discs.



Figure 8.1: The Reson SeaBat 9001 sonar used in the sea-trial 1998.



Figure 8.2: The Reson SeaBat 8101 sonar used in the sea-trial 2002.

The half cylindrical element at the front end of the main body is the receiver. Also the receiver consists of piezoelectric ceramic elements now in the form of wedge-shaped plates. When the return pulse hits the ends of the plates an electric signal is created. The signals from the elements are after amplification beamformed into 60 beams in a fan shaped plane, perpendicular to the transmitting beam and by that the elements only catches the return pulse from the cross-over by the fan shaped transmitting and receiving beams. The frequency is 455 kHz and the beamwidth is 1.5×1.5 degrees with spacing 1.5 degrees. The physical size of the main body is 473 x 190 x 126 mm.

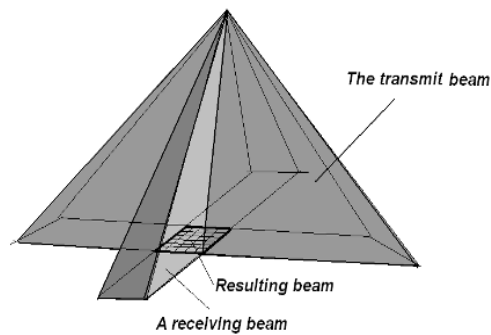


Figure 8.3: The principle to obtain a measurement beam in the Reson SeaBat 9001 and 8101 sonars as cross sections between the transmitting and receiving fan shaped beams.

It is also possible to steer the transmitting fan shaped beam in an angle to a vertical line as in Figure 8.4. By firing 5 successive pings with the beam plane at an angle of, say, -45° , -27° , 0° , 27° and 45° degrees a 3D bottom picture corresponding to Figure 7.22 can be achieved. The time between the individual pings is as short as possible and the result is compensated for the distance travelled between the pings. There will be an increased sensitivity for uncompensated temperature gradients in the outer beamsplanes since they are not vertical any longer. On the other hand the error in distance and angle between the beam planes will be smaller than the arrangement according to Figure 7.20.

The Reson SeaBat 8101 sonar in Figure 8.2 was used in the sea-trial described later in this chapter. A glimpse of the cylindrical transmitting element can be found at the left end of the body. The receiving element is the dark cylindrical strip on the main body between the mounting details. The transmitting and receiving beams are formed in the same manner as for the SeaBat 9001 sonar.

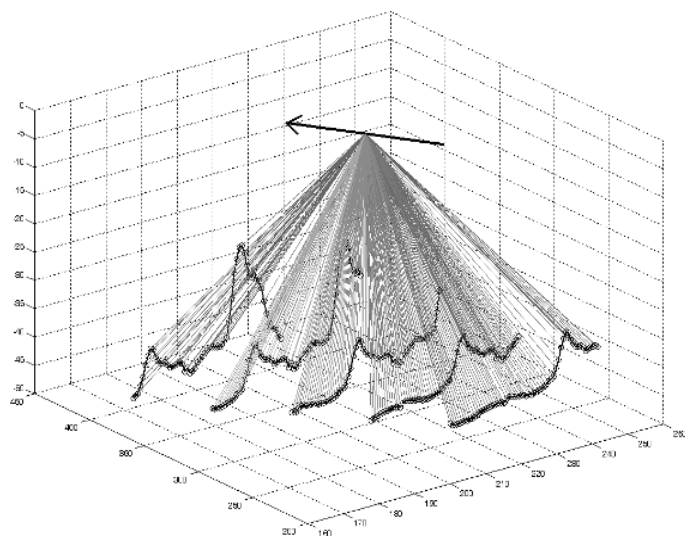


Figure 8.4: A sonar firing five transmit beams at different angles to the vertical line, cf. Figure 7.22.

The frequency is 240 kHz and there are 101 beams of width 1.5×1.5 degrees and with a spacing of 1.5 degrees.

Another sonar that has been used in the tests is the OmniTech EchoScope, Figure 8.5, which is a true 3D-sonar with a 40×40 receiver array forming 4096 receiver beams of 1.5×1.5 degrees with a beam spacing of 1.5 degrees. The receiver array, the shaded square, is made of a piezoelectric polymer and the three transmitters are made up of ceramics. The cylindrical element above and the two elements below the receiver array are the transmitters. The sonar can transmit three different CW frequencies and the curvature of the transmitting element determines the beam width of the transmitting pulse. Figure 8.6 shows a bottom picture taken by this sonar. Figure 8.7 shows the footprint of the beams. The ideal terrain navigation sonar is a true 3D sonar, due to the fixed distance between the beams, with a few hundred beams and a short single transmitting pulse if covertness is required.

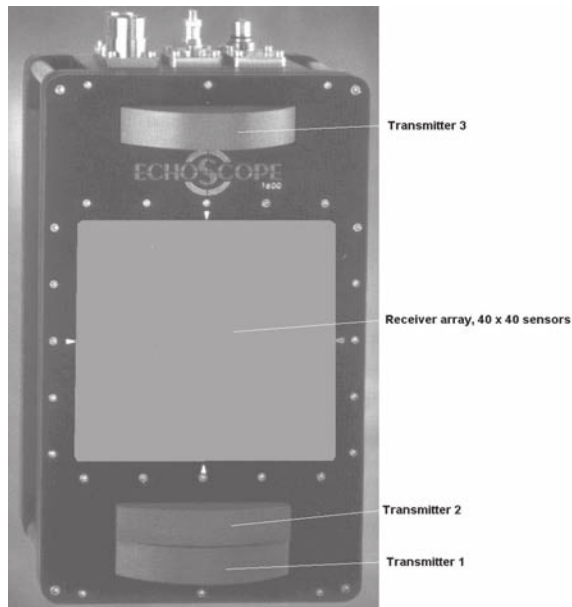


Figure 8.5: The OmniTech, EchoScope sonar is a true 3D sonar with a 40×40 receiving array of a piezoelectric polymer, PVDF. The different transmitters have different frequencies, 150, 300 and 600 kHz. The curvature of the transmitters forms the transmitting beam shape. The sonar was used in the sea-trial 2000.

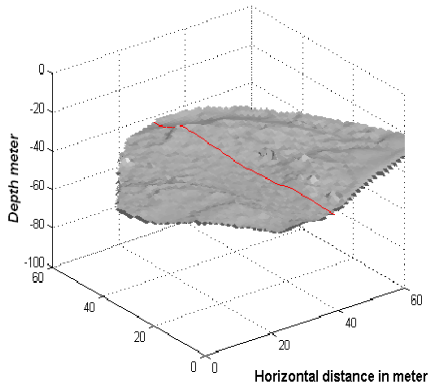


Figure 8.6: A bottom picture from the EchoScope sonar.

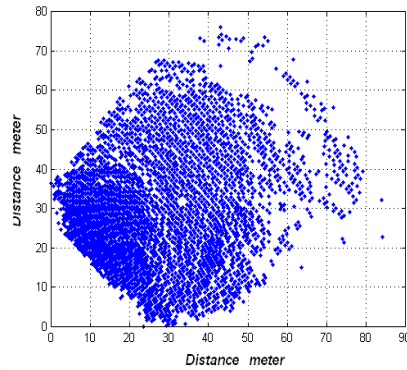


Figure 8.7: The footprints from the measurement beams for the EchoScope sonar.

8.2.3 Measurement errors

When measuring the terrain profile consideration has to be taken to several types of measurement errors with varying degree of seriousness.

Errors due to the ship movement

The usual mounting of the sonarhead is flush to the ship body. To be able to compensate for the ship movement an inertial measurement unit (IMU) is mounted as close as possible to the sonar head. However it will not be possible to completely compensate for the ship movement and thus some errors will remain.

An earlier study by a sonar manufacturer gave the following measurement error for a case with a 50 meter depth and a total sonar opening beam angle of 90 degrees, 60 beams with beamwidth 1.5 degrees.

Accurate sensorsuite

Uncompensated pitch and roll 0.05 degrees.

Uncompensated heave 0.05 meter.

Depth error in individual beams at 95% confidence interval 0.25 meter.

Typical sensorsuite

Uncompensated pitch and roll 0.1 degrees.

Uncompensated heave 0.05 meter .

Depth error in individual beams at 95% confidence interval 0.33 meter.

The sonar mounting

The transmitter of the sonar sound pulse is often of the omni directional type which means that reflections from the vehicle body may appear.

The characteristics of the sonar

The depth resolution of the sonar depends on its bandwidth and pulse shape and is in most cases of secondary importance (<0.05 m) for the depth error. The beamwidth can, however, in certain cases be of importance. Bathymetry sonars often have a beamwidth of 1.5×1.5 degrees which means that the footprint of the beam is approximately 1.3×1.3 meter at 50 meter depth. Since the sonar in most cases will take the smallest depth within the footprint as the depth, an error may result at high slant angles, especially at larger depths.

However, the greatest error will in most cases be caused by different detections of the bottom line. Depending on the sonar frequency the sonar beam will penetrate differently into the bottom sediment. The depth of the soft sediment layer can be several meters. A low frequency sonar will penetrate deep into the bottom sediment while a high frequency sonar will penetrate only a little. Sometimes, when using profile matching, this will not change the bottom profile in other cases, i.e., large stone blocks in the sediment, the profile will change.

The upper part of the bottom sediments will have a very high content of water which is successively decreased with increasing depth. This will cause the return echo to be diffuse and depending on the bottom detection algorithm different sonars will give different depth measures. This can be a problem for the beams at high angle of incidence since they travel long into the sediments. The harder the bottom is the less is the problem.

Sound bending due to temperature gradients and stratifying layers

The depth shown by the sonar is calculated from the travelling time of the sound pulse. This means that if the path of the beam is longer due to the beam bending a larger depth than the real depth is shown. If the temperature gradient is known the sonar software will compensate for this. However, in shallow

water the temperature gradient will also depend on the horizontal position. One reason for that is underwater currents.

In order to minimize the problem with beam bending the sonar beams shall be as perpendicular to the bottom as possible and the weighting of the outer beams shall be less than the central beams in the matching process.

Measurement error due to slant reflections, fishes and similar things

Frequently depth errors will occur due to non-reflecting beams. That means that the side lobes of the beam will catch a return pulse from some other direction than expected and show a wrong depth. Especially the outer beams are exposed to this at high bottom slant angles.

Fish and objects in the water column may give large depth errors but those errors are often easy to correct by removing the depth values without changing the others by filtering. The traditional way when charting is to remove these types of errors manually by an operator.

The accuracy of the map

The underwater maps produced by the Maritime authority is produced for merchant shipping and this may influence the use of them for terrain navigation. Sea bottom areas are most often charted by low frequency sonars. Depths given are conservative i.e., the actual depth may be larger.

The official depth accuracy given is not better than 1.5 meter. In reality it might be better. The horizontal position error of the depth given is in the range of 1 meter.

8.3 Sea-trial October 2002, Simulation

8.3.1 Introduction

To verify the characteristics of the proposed navigation method three large sea-trials have been conducted with similar results. Some of the results from the last sea-trial in October 2002 will be presented here. The sea-trial was conducted in the way that a track of 65 km around an island was divided by a number of way-points as shown in Figure 8.8. At every way-point it was tested if it was possible to get a position with the suggested navigation method. The way-points were selected from the underwater map by the eye to have good

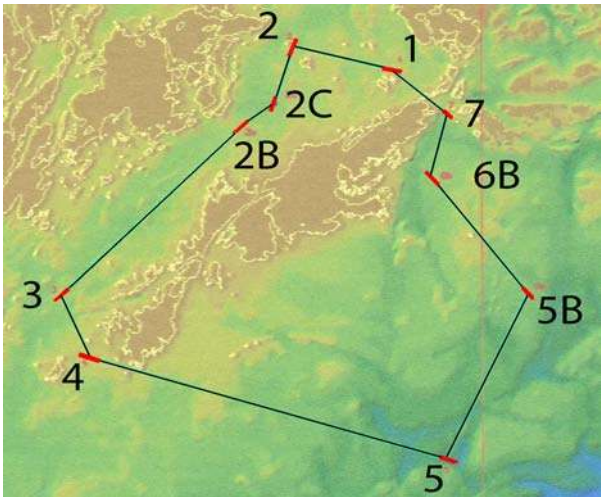


Figure 8.8: The 65 km long test track with the way-points marked.



Figure 8.9: Lowering the sonar through the moon-pool. In the middle of the picture is the IMU for measuring the sonar attitude.

characteristics for terrain navigation. At that time the theory of how to select way-points, Chapter 7, was not developed. Intensive Monte Carlo studies to confirm the suitability of the way-points were conducted. One of the way-points was deliberately placed in a very flat area to examine the performance of the method in such an area.

Practically the test was conducted by mounting the sonar, RESON 8101, at one end of a long sturdy tube which was lowered into the water through the moon-pool in the rear of the ship, see Figure 8.2 and 8.9. The used sonar has 101 beams 1.5° apart in a fan-shape with a beam width of $1.5^\circ \times 1.5^\circ$.

8.3.2 Simulation of the positioning

To investigate the suitability of the way-points for terrain navigation Monte Carlo simulations were conducted. A 3D sonar was simulated by assembling 5 beam planes of 60 beams 1.5° apart in a fan-shape with a beam width of $1.5^\circ \times 1.5^\circ$. This corresponds to the SeaBat 9001 sonar. The number of beams was thus lower than used in the sea-trial. The distance between the beam planes was 15 meter. The beam width has been considered negligible. A measurement error of 0.5 meter RMS was assumed based on earlier sea-trials.

When the vehicle moves from one way-point to another way-point with the help of the IN-system it will not hit the way-point exactly due to the INS error. The INS error consists of different random errors which can not be calculated in advance. Some of the more important errors will increase with at least the third power of time which means that the position error at the way-point largely depends on the travelling time since the last reset of the IN-system.

In the simulations the position error at the arrival to a way-point is assumed to have an error of 150 meter RMS. This means that 68% of all arrivals will be within a circle with a radius of 150 meter. At way-point BP7, which is close to BP6B a standard deviation of 30 meter has been used. The reason for placing BP6B so close to BP7, see Figure 8.8, is that the terrain at BP7 is not so good for terrain positioning (the depth is very small) so a small INS position error is needed for an unambiguous position at BP7. The short distance between BP6B and BP7 means that false correlation peaks can be excluded. The search area for a terrain position has been the whole underwater map, i.e., 5000 x 5000 meter for all way-points. In the simulation a bearing error of 1.5° RMS of the beam pattern and the journey between the way-points has also been used, see Figure 8.10. The number of trials at each way-point is 1000.

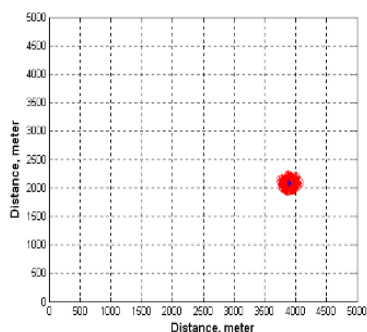


Figure 8.10: The spread in position at arrival at way-point BP3.

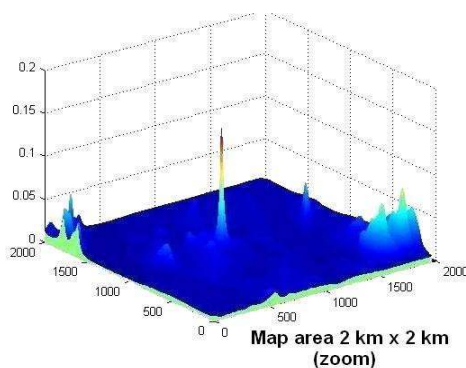


Figure 8.11: The correlation result at way-point BP3.

An example of simulation result at way-point BP3

The depth at way-point BP3 is 38 meter. The Figure 8.10 shows the spread in position at arrival at the way-point and Figure 8.11 shows the inverse of the correlation sum $T(x)$, see Chapter 3. From Figure 8.11 it can be seen that there are no false peaks of importance.

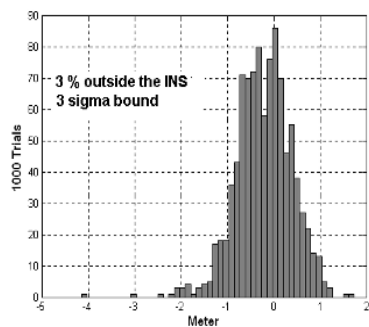


Figure 8.12: Position error in the x-direction at BP3.

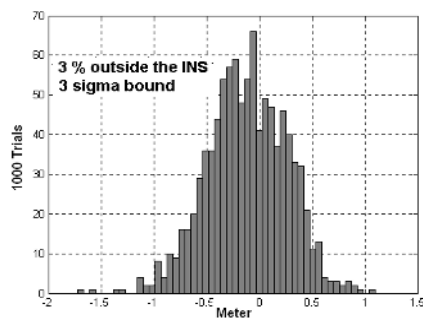


Figure 8.13: Position error in the y-direction at BP3.

The result from the simulation is that at this way-point of all 1000 trials only 33 trials gave a position with larger error than $3\sigma = 450$ meter. This may be considered as a large number but is probably explained by the simulated random bearing error and if so does not depend on the terrain positioning. If we are considering these to be outliers the distribution of the position error is as in Figure 8.12 and 8.13.

From the histograms we see that they remind of Gaussian distributions. In Chapter 4 it was shown that for large number of measurement beams the position error distribution will approximately be Gaussian.

The step in the matching procedure was 1 meter. The mean of the position error was 0.7 meter and the standard deviation was RMS 0.45 meter. Note that the real world position error also include errors due to other reasons than the correlation and that these will add to the correlation error given here.

Other simulation results

Way-point BP1

The depth at way-point BP1 is 58 meter. The mean of the position error was 0.61 meter and the standard deviation was RMS 0.36 meter.

Way-point BP2

The depth at way-point BP2 is 39 meter. The mean of the position error was 0.56 meter and the standard deviation was RMS 0.33 meter. In this case 20 positions outside the 3σ -limit were obtained.

Way-point BP4

The depth at way-point BP4 is 31 meter. The mean of the position error was 0.51 meter and the standard deviation was RMS 0.29 meter. In this case 20 positions outside the 3σ -limit were obtained, see Figure 8.14 and 8.15.

Way-point BP5

The depth at way-point BP5 is 101 meter. The mean of the position error was 0.60 meter and the standard deviation was RMS 0.38 meter, no positions outside the 3σ -limit were obtained, see Figure 8.16 and 8.17.

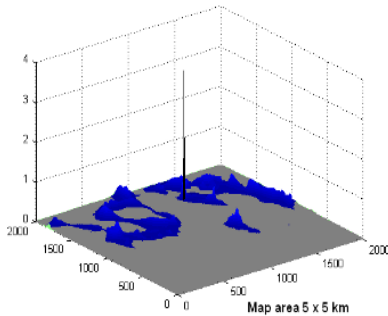


Figure 8.14: The correlation peak for way-point BP4, 5 beam planes. Map area 2 x 2 km.

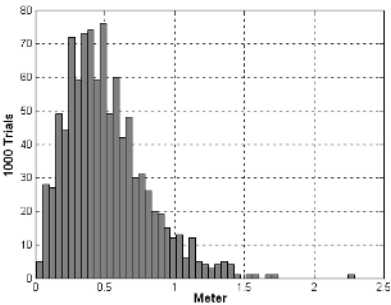


Figure 8.15: Radial position error at way-point BP4.

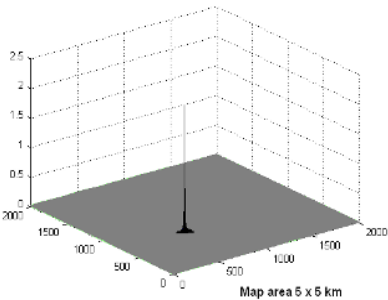


Figure 8.16: The correlation peak for way-point BP5, 5 beam planes. Map area 2 x 2 km.

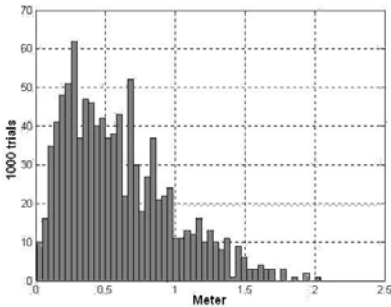


Figure 8.17: Radial position error at way-point BP5.

Way-point BP5B

The depth at way-point BP5B is 59 meter. The mean of the position error was 0.49 meter and the standard deviation was RMS 0.28 meter, 5 positions outside the 3σ -limit were obtained.

Way-point BP6B

The depth at way-point BP6B is 39 meter. The mean of the position error was 0.54 meter and the standard deviation was RMS 0.31 meter, 8 positions outside the 3σ -limit were obtained.

Way-point BP7

The depth at way-point BP7 is 10 meter. The mean of the position error was 1.0 meter and the standard deviation was RMS 0.7 meter, 145 positions outside the 3σ -limit were obtained.

8.4 Sea-trial October 2002, results

Due to bad weather conditions the way-points 4, 5, 5B and 6B could not be approached. A reasonable assumption based on the simulations is that the positioning result at these way-points would have been as good as at the other way-points. The depth at these way-points would have given a large bottom coverage for the sonar and the terrain has also large variations at the way-points. Below are some brief correlation results from the other way-points. In the tables the notation 5/75 means 5 beamplanes 75 pings apart. Length and width refer to the correlation area and outliers are removed. The position reference is a DGPS measure with a believed accuracy of 3 meter RMS.

8.4.1 Way-point 1

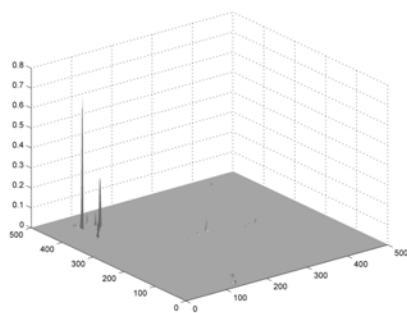


Figure 8.18. Likelihood function. Map area 5 km by 5 km, 5 beam planes. Map area 5 x 5 km. BP1

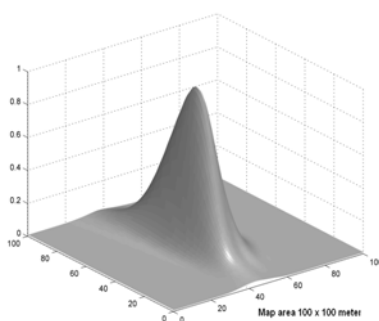


Figure 8.19: The N -normalized likelihood function. Horizontal scale in meter. BP1.

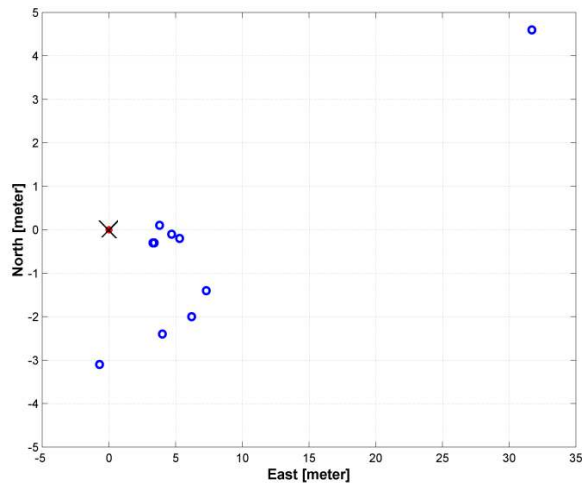


Figure 8.20: Positions at the way-point BP1. The cross marks the way-point.

Table 8.1. Position error at way-point BP1

Ping No.	No./Dist. beam planes	depth m	x error m	y error m	length m	width m
1730	5/75	-16.0	0.1	3.8	106	82
1150	5/75	-19.0	-2.0	6.2	160	93
550	5/75	-16.0	-0.3	3.3	153	84
1100	5/75	-19.0	-2.4	4.0	141	92
1700	5/75	-16.0	-0.1	4.7	110	75
1080	5/75	-19.0	-1.4	7.3	170	89
600	5/75	-15.9	-0.3	3.4	106	59
1350	5/75	-18.8	-3.1	-0.7	119	92
1280	5/75	-15.9	-0.2	5.3	163	84
820	5/75	-18.8	4.6	31.7	164	91

Radial position error 4.7 m with respect to origo. Radial position error 2.0 m with respect to the mean error.

8.4.2 Way-point 2

Way-point 2 is situated in a flat area. The Figures 8.21 and 8.22 show the likelihood function for 5 beam planes. The Figures 8.23 and 8.24 show the likelihood function for 11 beam planes, almost doubling the correlation area.

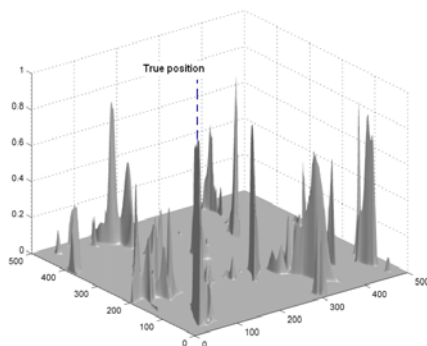


Figure 8.21: Likelihood function, 5 beam planes. Map area 5 x 5 km . BP2.

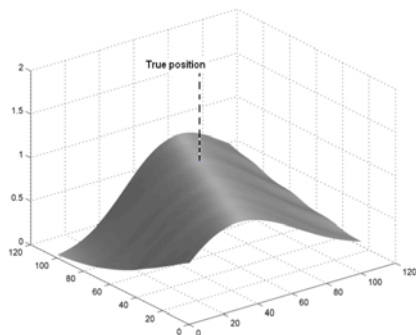


Figure 8.22: The N -normalized likelihood function. Horizontal scale in meter, 5 beam planes.

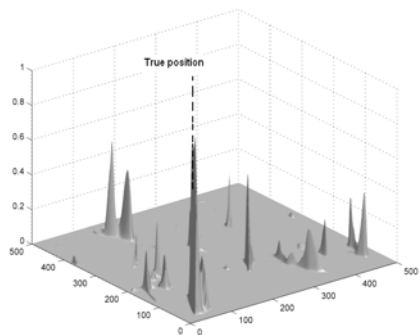


Figure 8.23: Likelihood function, 11 beam planes. Map area 5 x 5 km . BP2. Note the decrease in false peaks.

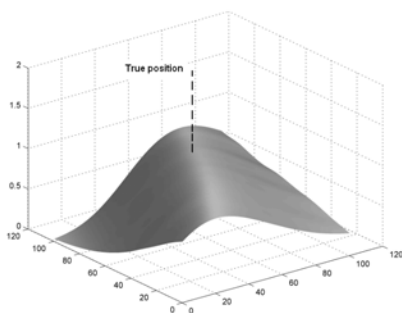


Figure 8.24: The N -normalized likelihood function. Horizontal scale in meter, 11 beam planes.

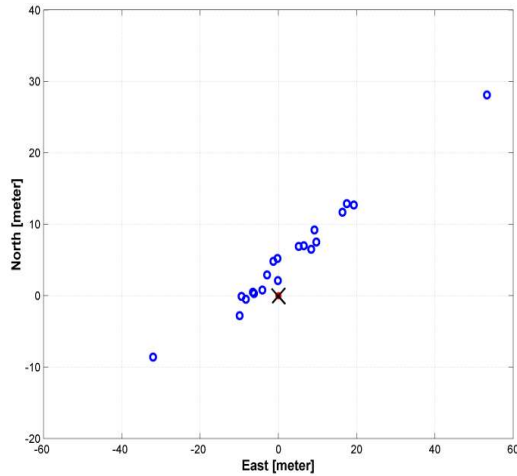


Figure 8.25: Positions at the way-point 2. The cross marks the way-point. The way-point is situated at the slant of a ridge. There is good accuracy across the ridge but worse along the ridge.

Table 8.2. Position error at way-point BP2

Ping No.	No./Dist. beam planes	depth m	x error m	y error m	length m	width m
1300	5/75	-38.4	7.0	6.5	101	109
1500	5/75	-38.4	-8.6	-32.0	106	109
1500	11/75	-38.4	0.5	-6.5	264	109
1700	5/75	-38.4	-2.8	-9.9	109	109
1300	5/75	-37.7	-0.5	-8.3	101	107
1500	5/ 75	-37.5	28.1	53.3	102	107
1500	11/75	-37.5	2.1	-0.1	253	107
1700	5/75	-37.4	12.7	19.3	100	106
1300	5/ 75	-37.1	6.9	5.2	95	105
1500	5/75	-37.4	-0.1	-9.4	98	106
1500	11/75	-37.4	9.2	9.2	243	106
1700	5/75	-37.5	5.2	0.2	98	106
1300	5/75	-37.4	0.8	-4.1	97	106

Table 8.2. *Cont. Position error at way-point BP2*

Ping No.	No./Dist. beam planes	depth m	x error m	y error m	length m	width m
1500	5/75	-37.4	0.3	-6.2	105	106
1500	11/75	-37.4	6.5	8.4	260	106
1700	5/75	-37.4	7.5	9.7	111	106
1300	5/75	-37.2	11.7	16.4	109	106
1500	11/75	-37.6	12.9	17.5	270	107
1700	5/75	-37.7	4.8	-1.3	109	107

Radial position error 10.0 m with respect to origo. Radial position error 9.2 m with respect to the mean error.

8.4.3 Way-point 2B

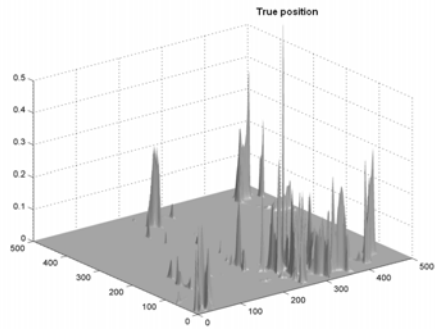


Figure 8.26: Likelihood function, 5 beam planes. Map area 5 x 5 km. BP2B.

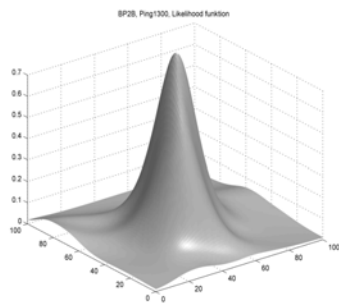


Figure 8.27: The N -normalized likelihood function. Horizontal scale in meter, 5 beam planes. BP2B.

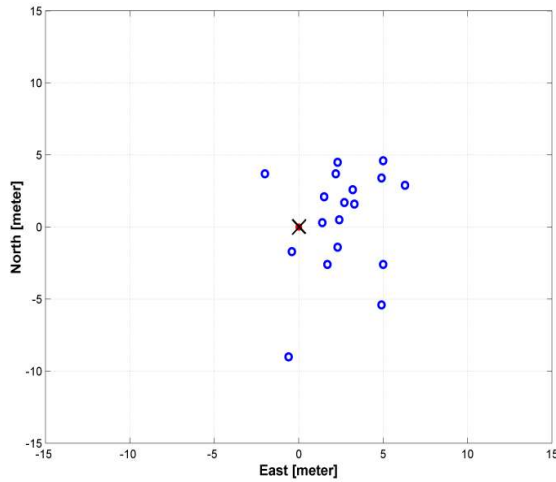


Figure 8.28: Positions at the way-point 2B. The cross marks the way-point.

Table 8.3. Position error at way-point BP2B

Ping No.	No./Dist. beam planes	depth m	x error m	y error m	length m	width m
1300	5/ 75	-30.3	3.7	2.2	0	0
1300	5/75	-26.5	3.7	-2.0	0	0
1500	11/75	-26.5	0.3	1.4	0	0
1700	5/ 75	-27.6	-2.6	1.7	0	0
1300	5/ 75	-26.1	201.0	442.5	0	0
1500	5/75	-28.9	2.9	6.3	0	0
1500	11/75	-28.9	2.1	1.5	0	0
1700	5/ 75	-30.8	2.6	3.2	0	0
1300	5/ 75	-29.9	4.6	5.0	0	0
1500	5/ 75	-25.9	3.4	4.9	0	0
1500	11/75	-25.9	1.7	2.7	0	0
1700	5/ 75	-27.7	-5.4	4.9	0	0
1300	5/75	-30.4	1.6	3.3	0	0
1500	5/75	-32.4	0.5	2.4	0	0
1500	11/75	-32.4	-1.4	2.3	0	0
1700	5/75	-33.6	-2.6	5.0	0	0

Table 8.3. *Cont. Position error at way-point BP2B*

Ping No.	No./Dist. beam planes	depth m	x error m	y error m	length m	width m
1300	5/75	-28.5	4.5	2.3	131	85
1500	5/75	-26.0	330.9	-69.0	120	78
1500	11/75	-26.0	-1.7	-0.4	313	78
1700	5/75	-27.6	-9.0	-0.6	121	83

Note: Missing length and width data are marked with 0.

8.4.4 Way-point 3

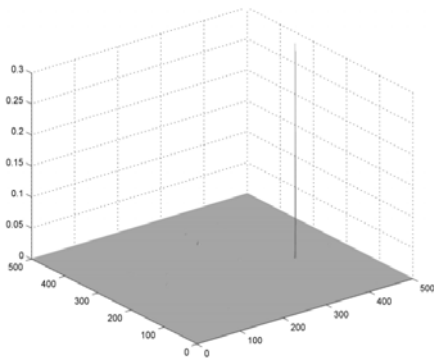


Figure 8.29: BP3, likelihood function, 5 beam planes. Map area 5 x 5 km.

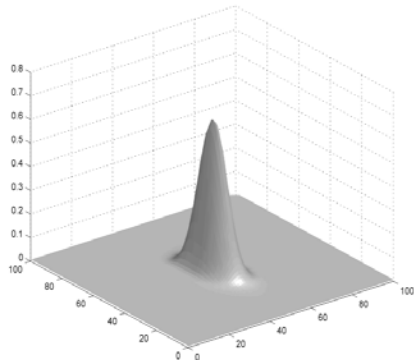


Figure 8.30: BP3, the *N*-normalized likelihood function. Horizontal scale in meter, 5 beam planes.

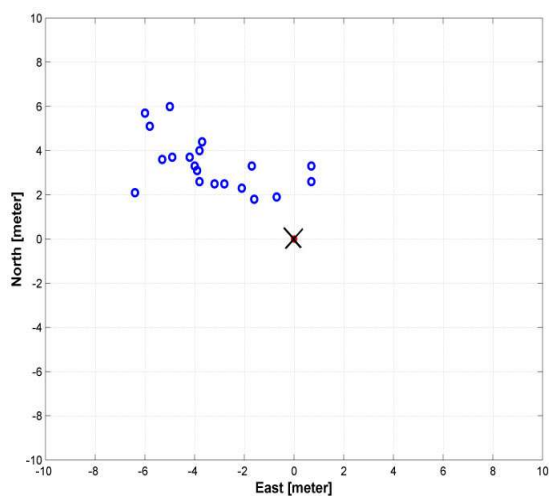


Figure 8.31: Positions at the way-point 3. The cross marks the way-point.

Table 8.4. Position error at BP3

Ping No.	No./Dist. beam planes	depth m	x error m	y error m	length m	width m
1300	5/75	-26.7	2.1	-6.4	121	100
1500	5/75	-32.6	5.1	-5.8	117	100
1500	11/75	-32.6	4.0	-3.8	296	100
1700	5/75	-34.1	3.7	-4.9	116	104
1300	5/75	-24.4	2.3	-2.1	125	80
1500	5/75	-28.0	2.6	0.7	121	99
1500	11/75	-28.0	1.9	-0.7	306	99
1700	5/75	-36.5	3.3	0.7	120	105
1300	5/75	-24.1	4.4	-3.7	121	82
1500	5/75	-32.3	3.3	-4.0	119	104
1500	11/75	-32.3	3.1	-3.9	303	104
1700	5/75	-31.6	5.7	-6.0	123	98
1300	5/75	-26.0	1.8	-1.6	116	85
1500	5/75	-28.6	2.5	-2.8	117	94
1500	11/75	-28.6	2.6	-3.8	295	94
1700	5/75	-34.5	3.3	-1.7	121	102
1300	5/75	-25.7	2.5	-3.2	126	87

Table 8.4. *Cont. Position error at BP3*

Ping No.	No./Dist. beam planes	depth m	x error m	y error m	length m	width m
1500	5/75	-32.3	3.6	-5.3	125	103
1500	11/75	-32.3	3.7	-4.2	315	103
1700	5/75	-33.6	6.0	-5.0	126	100

Radial position error 5.0 m with respect to origo. Radial position error 2.0 m with respect to the mean.

8.4.5 Way-point 7

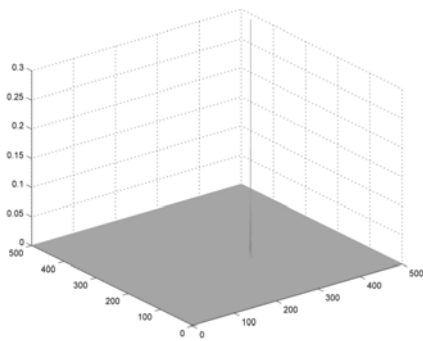


Figure 8.32: BP7, likelihood function, 5 beam planes. Map area 5 x 5 km.

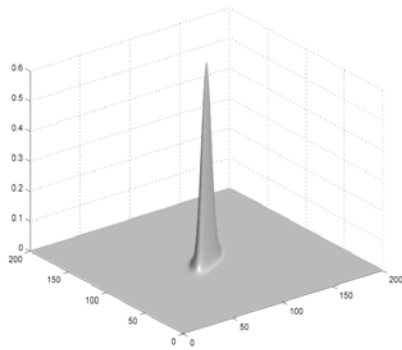


Figure 8.33: BP7, the *N*-normalized likelihood function. Horizontal scale in meter, 5 beam planes.

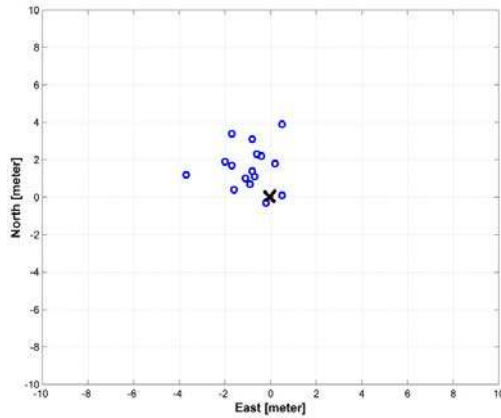


Figure 8.34: Positions at the way-point 7. The cross marks the way-point.

Table 8.5. Position error at way-point BP7

Ping No.	No./Dist. beam planes	depth m	x error m	y error m	length m	width m
304	5/75	-41.4	0.4	-1.6	195	166
500	5/75	-20.6	1.9	-2.0	215	105
500	11/75	-20.6	3.1	-0.8	512	105
600	5/75	-30.9	1.0	-1.1	216	144
200	5/75	-33.8	3.9	0.5	247	172
300	5/75	-36.1	0.7	-0.9	238	172
400	5/75	-21.6	3.4	-1.7	245	75
200	5/75	-36.1	1.1	0.7	318	229
300	5/75	-22.1	1.2	-3.7	314	88
500	5/75	-32.3	1.4	-0.8	299	201
200	5/75	-47.2	1.8	0.2	253	243
300	5/75	-33.8	0.1	0.5	255	217
600	5/75	-22.6	1.7	-1.7	255	100
300	5/75	-21.1	2.2	-0.4	261	129
400	5/75	-36.6	-0.3	-0.2	267	176
500	5/75	-32.6	2.3	-0.6	275	216

Radial position error 2.2 m with respect to origo. Radial position error 1.3 m with respect to the mean.

8.5 Conclusions

The horizontal position error in the simulation was less than 1 meter RMS while the corresponding error in the sea-trial was a few meters. This might be seen as a large deviation but since the DGPS device only had an accuracy of about 3 meter RMS we could not expect an error smaller than that.

Compared to other terrain navigation methods a radial position error of 3 meter is extremely good (note that this figure includes the DGPS error) since the position error for the methods mentioned in Chapter 2 is in the range of 30 meter.

We see from the figures showing the estimated positions at the way-points that there is a position bias. Possible reasons for this bias have been carefully studied and it is believed that it is caused by bias in the map and/or bias in the DGPS position due to multipath in the satellite signal [PE04].

The method also shows very good performance in very flat areas as shown in way-point 2 which is considered by the eye to be completely flat. The importance of a large correlation area is also shown for that way-point.

It is also shown that for manned underwater vehicles the quality of the positioning, which is extremely important, could easily be determined by looking at the likelihood function. However the “variance” of the likelihood function may not always be a good measurement of the quality of the positioning due to terrain repeatability. It may be better to always use the Hessian of the log likelihood function at its maximum point.

A likelihood function as in Figure 8.21 shows that the positioning needs improvement either by increasing the correlation area or adding some external measurements as mentioned in Chapter 3. Is the likelihood function as in Figure 8.18, 8.26, 8.29, 8.32 it is almost sure that the achieved position is the correct position, this can be confirmed by the hypothesis testing method described in Section 4.2.3.

The Gaussian shape of the likelihood function when a large number of measurement beams are used is demonstrated for several way-points.



Chapter 9

Terrain navigation for autoregressive bottom models

9.1 Introduction

With the uprising of the computer game industry a great number of synthetic terrain models have become available, some are very advanced and some are very simple. Also within the military sector, simulation of war theatres has become common which has lead to the development of advanced terrain models. For terrain navigation, simple synthetic terrain models have been used for a long time to evaluate performance of navigation methods [BAAB84] with varying degree of success.

If in the same way the bottom topography could be described in statistical terms, e.g., as a stationary stochastic process, qualitative and quantitative judgements could be drawn about the positioning accuracy and adherent matters directly from the model. This would be of great advantage, but up to now the relevance of the models, covering large areas, for that purpose might be questioned. However, it seems at least possible that local models can be of value.

One of the simplest models one can think of is to generate a random matrix (in image theory called random field), symbolizing the terrain, and make 2D-filtering to give the landscape the requested correlation between locations [JA89]. However, a simple 1D terrain model obtained by filtering white noise through a first order filter may often be sufficient, that is an autoregressive process of first order, $AR(1)$, [BRDA91, TS02].

This chapter will therefore study the terrain navigation problem when the bottom profile can be approximated by a low order autoregressive process. A higher order process model will adjust better to the real world but that also means more parameters to be identified. The chapter is to a high degree freestanding from the previous chapters.

9.2 AR -bottom profiles

In this section the AR -assumption will be studied in some more details for the one dimensional bottom model.

9.2.1 $AR(1)$ bottom profile

Figure 9.1 shows a bottom profile along a 5 km straight line in an underwater map from the Stockholm archipelago and Figure 9.2 shows the autocorrelation

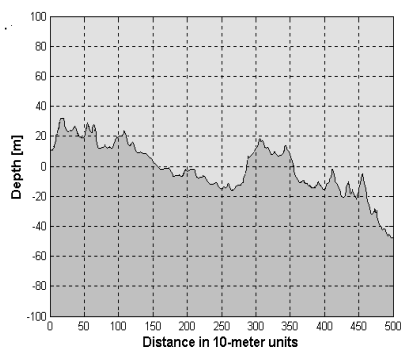


Figure 9.1: A typical zero mean bottom profile.

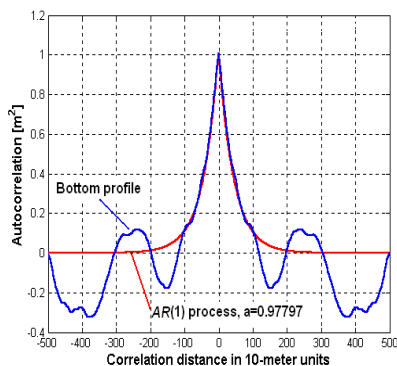


Figure 9.2: An adjusted $AR(1)$ process to the autocorrelation curve for the bottom in Figure 9.1

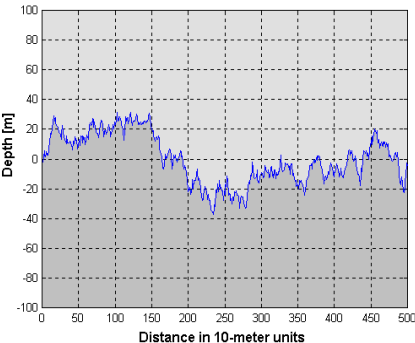


Figure 9.3: Synthetic $AR(1)$ bottom profile.

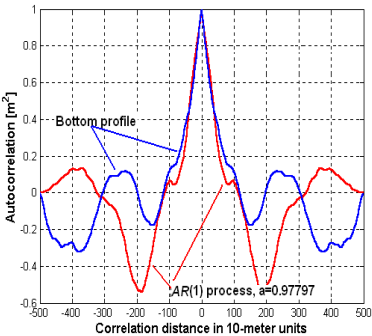


Figure 9.4: Autocorrelation curve for the $AR(1)$ process in Figure 9.3 and the autocorrelation curve for Figure 9.1.

of the bottom profile along with the autocorrelation curve for an $AR(1)$ process with the pole in $a=0.97797$. A realization of an $AR(1)$ -bottom profile with $a=0.97699$ can be seen in Figure 9.3.

As can be seen from Figure 9.3 there is a high similarity between the real bottom profile and the $AR(1)$ -profile for the lower frequencies. The large side-lobes in Figure 9.2 and 9.4 are because no windowing has been used for depressing the side-lobes. Figure 9.5 shows some typical $AR(1)$ profiles for the pole in the range $0.4 < a < 0.9999$.

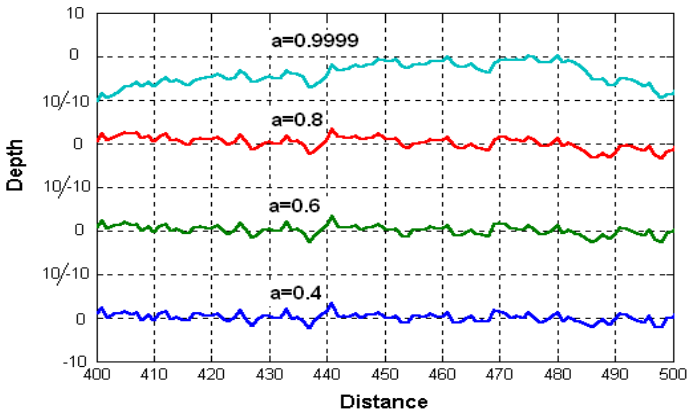


Figure 9.5: Some typical $AR(1)$ profiles for the pole in the range $0.4 < a < 0.9999$. The generating noise is the same for all profiles.

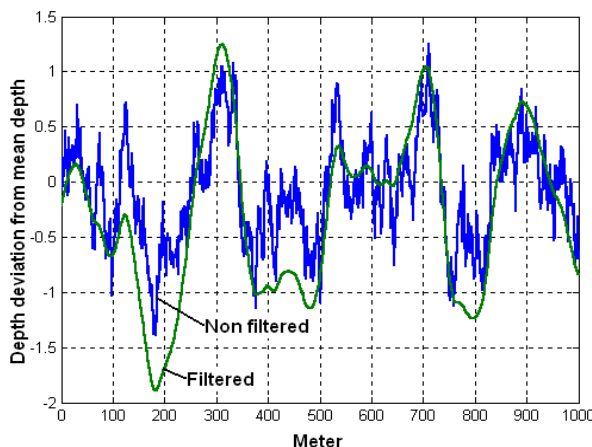


Figure 9.6: A filtered $AR(1)$ process for removing high frequencies.

9.2.2 $AR(2)$ bottom profile

Figure 9.6 shows an $AR(1)$ bottom and the same bottom profile filtered through a first order filter for removing high frequencies and by that receiving a more real bottom like profile. Filtering an $AR(1)$ process with an $AR(1)$ filter gives an $AR(2)$ process. It is natural that an $AR(2)$ - process will better describe the real bottom since it has more parameters. Another advantage relative the first order process is that the autocorrelation function for a continuous stochastic process can be differentiated in all points.

9.2.3 Determining the expected correlation function

In order to determine an estimate of the position we will later in this chapter calculate the correlation function

$$T(x) = [\mathbf{y} - \mathbf{h}(x)]^T [\mathbf{y} - \mathbf{h}(x)] \quad (9.1)$$

where \mathbf{y} is the measured bottom profile and $\mathbf{h}(x)$ is the profile according to the underwater map in position x . Let x be the true position which we measure and compare it with the bottom profile at the distance k from x . The measured profile is $\mathbf{y} = \mathbf{h}(x) + \mathbf{e}$ where \mathbf{e} is the measurement error.

We would like to investigate if the correlation function (9.1) may be expressed in statistical terms of the bottom profile. Therefore we will determine the expected value of $T(x+k)$ where k is the sampel index corresponding to some sampling distance. To simplify the analysis we will assume that x is a scalar and the stochastic process $\mathbf{h}(x)$ is stationary and uncorrelated with \mathbf{e} . The length of the vectors \mathbf{y} and \mathbf{e} is N . We also assume that the elements in \mathbf{e} are white Gaussian noise. We have

$$\begin{aligned}
 E\{T(x+k)\} &= E\left\{\sum_{j=0}^{N-1} (y_j - h_j(x+k))^2\right\} = E\left\{\sum_{j=0}^{N-1} (h_j(x) + e_j - h_{j+k}(x))^2\right\} \\
 &= 2N\sigma_h^2 - 2\sum_{j=0}^{N-1} \underbrace{E\{h_j(x)h_{j+k}(x)\}}_{r_h(k)} + N\sigma_e^2 \\
 &= N(2\sigma_h^2 + \sigma_e^2) - 2Nr_h(k)
 \end{aligned} \tag{9.2}$$

where we have introduced $\sigma_h^2 = \frac{1}{N} E\left\{\sum_{j=0}^{N-1} (h_j(x))^2\right\}$ and $\sigma_e^2 = \frac{1}{N} E\left\{\sum_{j=0}^{N-1} e_j^2\right\}$.

We can thus through the autocorrelation function, $r_h(k)$, get an understanding of the shape of the correlation function $T(x)$.

9.3 The covariance matrix

In the previous section we found an expression for the expected value of the correlation function. We are equally interested in finding an expression for the covariance of the difference between the bottom profile at two different locations. Assume therefore that the bottom profile, \mathbf{h} , is generated by a stochastic $AR(1)$ -process achieved by filtering a white Gaussian noise sequence $\mathbf{v}(k)$, which is

$\mathcal{N}(0, \sigma_0^2)$, through the first order filter $H(z) = \frac{1}{1-az^{-1}}$ where we will assume

that the pole is in the range $0 < a < 1$.

In Figure 9.7 two positions along the bottom profile have been marked, Area 1 with bottom profile \mathbf{h}_1 and Area 2 with bottom profile \mathbf{h}_2 . The vectors \mathbf{h}_1 and \mathbf{h}_2 are sections of the bottom profile, \mathbf{h} . The dimension of \mathbf{h}_1 and \mathbf{h}_2 is N .

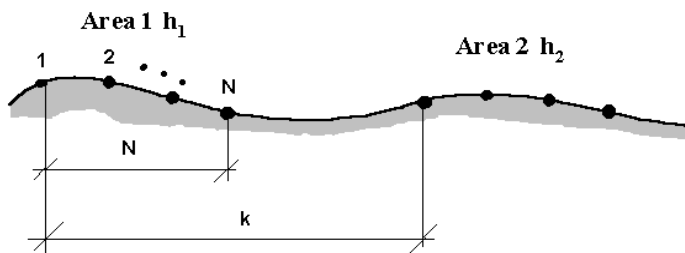


Figure 9.7: Synthetic bottom profile which is assumed to be generated by an $AR(1)$ process.

Let the position \mathbf{h}_1 correspond to the profile measured by the sonar while position \mathbf{h}_2 is a position situated k units from position \mathbf{h}_1 , which we match the sonar measurement with. The elements of the vector \mathbf{h}_1 are the depth values from the underwater map. The elements in the vectors \mathbf{h}_1 and \mathbf{h}_2 are $\mathcal{N}(0, \sigma_{ar1}^2)$ where $\sigma_{ar1}^2 = \sigma_0^2 / (1 - a^2)$. The autocorrelation function for the process that generates \mathbf{h} is $E\{h_i(x)h_{i+k}(x)\} = \sigma_{ar1}^2 a^{-|k|}$ if k is the distance between the elements of \mathbf{h} .

We will to begin with assume that the bottom profile is measured without errors. The correlation means that we compare the bottom profiles at Area 1 and Area 2 by forming the difference between the profiles. Thus, we have the difference $\Delta h = h_{2,i} - h_{1,i} = h_{k0+k+i} - h_{k0+i}$ where index i refers to the i th element in \mathbf{h}_1 and \mathbf{h}_2 and index $k0$ refers to the starting point of the vector \mathbf{h}_1 .

For an element in the vector $\Delta \mathbf{h} = \mathbf{h}_2 - \mathbf{h}_1$ we have

$$E(\Delta h_i) = 0 \quad (9.3)$$

$$\begin{aligned} Var(\Delta h_i) &= E(\Delta h_i^2) - [E(\Delta h_i)]^2 \\ &= E((h_{1,i} - h_{2,i})^2) = E(h_{1,i}^2) + E(h_{2,i}^2) - 2E(h_{1,i}h_{2,i}) = 2\sigma_{ar1}^2(1 - a^{|k|}) \end{aligned} \quad (9.4)$$

The elements in the vector $\Delta \mathbf{h}$ are thus $\mathcal{N}(0, \sigma^2)$ where

$$\sigma^2 = 2\sigma_{ar1}^2(1 - a^{|k|}) = \frac{2\sigma_0^2}{1 - a^2}(1 - a^{|k|}), \text{ they are also correlated with each other. If}$$

the distance between the positions i and j is m then we have the following expression for the covariance

$$\begin{aligned}\rho(m) &= E(\Delta h_i \Delta h_j) = E\{(h_{1,i} - h_{2,i})(h_{1,j} - h_{2,j})\} \\ &= E(h_{1,i} h_{1,j}) + E(h_{2,i} h_{2,j}) - E(h_{1,i} h_{2,j}) - E(h_{2,i} h_{1,j})\end{aligned}\quad (9.5)$$

We arrive at the following expression for the elements in the covariance for an $AR(1)$ -process with the pole in a and where m is the distance between elements in the $\Delta \mathbf{h}$ vector and k is the displacement between \mathbf{h}_1 and \mathbf{h}_2

$$\rho(m) = \left(2a^{|m|} - a^{|k+m|} - a^{|k-m|}\right) \sigma_{ar1}^2, \quad |m| \leq N-1, \quad \forall k \quad (9.6)$$

By that the covariance matrix for the bottom profile difference in the case of no measurement error is determined. The Figures 9.8 och 9.9 show $\rho(m)$ as a function of k at two different values of the pole a .

We note also that the covariance matrix is given directly by the expression

$$\begin{aligned}\mathbf{C}_{\Delta \mathbf{h}} &= E\{\Delta \mathbf{h} \Delta \mathbf{h}^T\} = E\{(\mathbf{h}_2 - \mathbf{h}_1)(\mathbf{h}_2 - \mathbf{h}_1)^T\} \\ &= 2\mathbf{C}_h - E\{(\mathbf{h}_2 \mathbf{h}_1^T) - E\{\mathbf{h}_1 \mathbf{h}_2^T\}\end{aligned}\quad (9.7)$$

where $\mathbf{C}_h = E\{\mathbf{h}_1 \mathbf{h}_1^T\} = E\{\mathbf{h}_2 \mathbf{h}_2^T\}$ which for large displacements k simplifies to

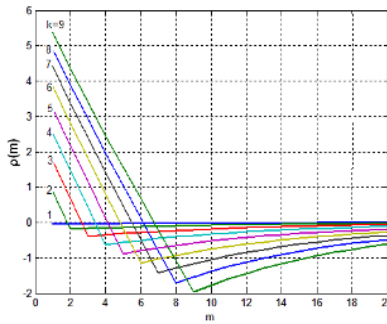


Figure 9.8: The elements in the covariance matrix for $a=0.90$.

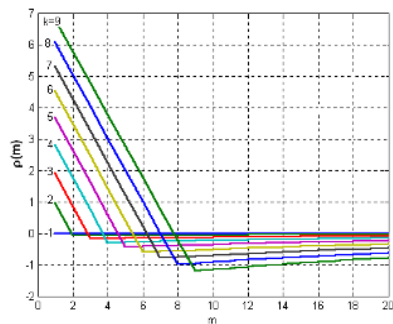


Figure 9.9: The elements in the covariance matrix for $a=0.96$.

$$\mathbf{C}_{\Delta\mathbf{h}} = 2\mathbf{C}_{\mathbf{h}} \quad (9.8)$$

since for large k \mathbf{h}_1 and \mathbf{h}_2 will be uncorrelated.

9.4 Decomposition of the covariance matrix

The measured bottom profile at the true position, x , can be written as $\mathbf{y}=\mathbf{h}_1+\mathbf{e}$ where \mathbf{e} is the white measurement noise and the correlation sum is

$$T(x) = (\mathbf{y} - \mathbf{h}_1)^T (\mathbf{y} - \mathbf{h}_1) = \mathbf{e}^T \mathbf{e} = \sum_{i=1}^N e_i^2 \quad (9.9)$$

which will be χ_N^2 -distributed if e_i is Gaussian $\mathcal{N}(0,1)$.

At the position \mathbf{h}_2 we have a similar but not equal bottom profile $\mathbf{h}_2=\mathbf{h}_1+\Delta\mathbf{h}$, where $\Delta\mathbf{h}$ is Gaussian and correlated as described before. The correlation sum will be

$$T(x+k) = (\mathbf{y} - \mathbf{h}_2)^T (\mathbf{y} - \mathbf{h}_2) = \mathbf{e}^T \mathbf{e} - 2\mathbf{e}^T \Delta\mathbf{h} + \Delta\mathbf{h}^T \Delta\mathbf{h} \rightarrow \mathbf{e}^T \mathbf{e} + \Delta\mathbf{h}^T \Delta\mathbf{h} \quad (9.10)$$

for large N and

$$E\{T\} = E\{(\mathbf{y} - \mathbf{h}_2)^T (\mathbf{y} - \mathbf{h}_2)\} = E\{\mathbf{e}^T \mathbf{e}\} + E\{\Delta\mathbf{h}^T \Delta\mathbf{h}\} \quad (9.11)$$

The last term in (9.10) can be seen to have a kind of a χ_N^2 looking distribution by first decorrelating the stochastic variables Δh_i in the $\Delta\mathbf{h}$ vector by a Cholesky decomposition, $\mathbf{C}_{\Delta\mathbf{h}} = \mathbf{L}\mathbf{D}\mathbf{L}^T$, or a modal decomposition of the covariance matrix, $\mathbf{C}_{\Delta\mathbf{h}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$. The difference being that the modal decomposition conserves the scale of the axes during the rotation of the axes which the Cholesky decomposition does not. The modal decomposition of the correlation matrix has been discussed in Chapter 4 and we will here discuss only the Cholesky decomposition.

The vectors \mathbf{h}_1 and \mathbf{h}_2 are zero mean Gaussian stochastic variables and $\Delta\mathbf{h}=\mathbf{h}_2-\mathbf{h}_1$ is the difference between the two profiles. The probability density function for $\Delta\mathbf{h}$ is

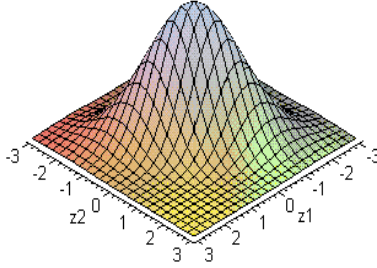


Figure 9.10: The probability density function when $N=2$.

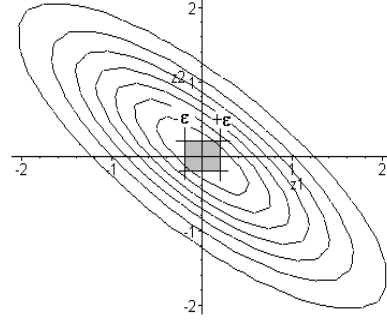


Figure 9.11: Contour lines for the PDF in Figure 9.10.

$$p(\Delta \mathbf{h}) = \frac{1}{(2\pi)^{N/2} \sqrt{\det(\mathbf{C}_{\Delta \mathbf{h}})}} \exp\left(-\frac{1}{2} \Delta \mathbf{h}^T \mathbf{C}_{\Delta \mathbf{h}}^{-1} \Delta \mathbf{h}\right) \quad (9.12)$$

where $\mathbf{C}_{\Delta \mathbf{h}}$ is the covariance matrix for $\Delta \mathbf{h}$. The Figures 9.10 and 9.11 illustrate the probability density function for the case $N=2$.

In the Figure 9.11 the area where $|\Delta \mathbf{h}| \leq \epsilon$ is marked. The volume below the probability density function in the shaded area corresponds to the probability that $|\Delta \mathbf{h}| \leq \epsilon$, i.e., the absolute value of the depth difference in each sonar beam does not exceed $\epsilon = [\epsilon_1, \dots, \epsilon_N]^T$.

9.4.1 Cholesky decomposition of the covariance matrix

Cholesky decomposition and inversion of the covariance matrix, $\mathbf{C}_{\Delta \mathbf{h}} = \mathbf{L} \mathbf{D} \mathbf{L}^T$, give

$$\Delta \mathbf{h}^T \mathbf{C}_{\Delta \mathbf{h}}^{-1} \Delta \mathbf{h} = \Delta \mathbf{h}^T (\mathbf{L}^T)^{-1} \mathbf{D}^{-1} \mathbf{L}^{-1} \Delta \mathbf{h} = \mathbf{y}^T \mathbf{D}^{-1} \mathbf{y} \quad (9.13)$$

where $\mathbf{y} = \mathbf{L}^{-1} \Delta \mathbf{h}$.

We also have that $\det(\mathbf{C}_{\Delta\mathbf{h}}) = \prod_{i=1}^N d_{i,i}$, i.e., the product of the diagonal elements in \mathbf{D} since the \mathbf{L} matrix is a lower triangular matrix with 1 as diagonal elements. If we call the diagonal elements of \mathbf{D} for $d_{i,i} = \sigma_i^2$ and the elements in the vector \mathbf{y} for y_i then we can write

$$\begin{aligned} p(\Delta\mathbf{h}) &= \frac{1}{(2\pi)^{N/2} \sqrt{\sigma_1^2 \dots \sigma_N^2}} \exp\left(-\frac{1}{2} \sum_{i=1}^N \frac{1}{\sigma_i^2} y_i^2\right) \\ &= \frac{1}{(2\pi)^{N/2} \sqrt{\sigma_1^2 \dots \sigma_N^2}} \exp\left(\frac{-y_1^2}{2\sigma_1^2}\right) \dots \exp\left(\frac{-y_N^2}{2\sigma_N^2}\right) \end{aligned} \quad (9.14)$$

or as a product of probability density functions

$$p(y; \sigma_i^2) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{y^2}{2\sigma_i^2}\right) \quad (9.15)$$

By that we also have conducted a rotation and scale change of the coordinate system. Therefore we also have to change the integration area when solving the integral for the probability for a position in error, $\Pr(|\Delta\mathbf{h}| \leq \varepsilon)$. The variance σ_i depends on k but for large k values, see Figure 9.12, all diagonal values, except the first, are approximately the same. We can therefore approximately write

$$p(\Delta\mathbf{h}) \approx p(y; \sigma_1^2) \prod_{i=2}^N p(y; \sigma_2^2) \quad (9.16)$$

with equality for large k and where

$$p(y; \sigma_1^2) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{y^2}{2\sigma_1^2}\right) \quad (9.17)$$

and

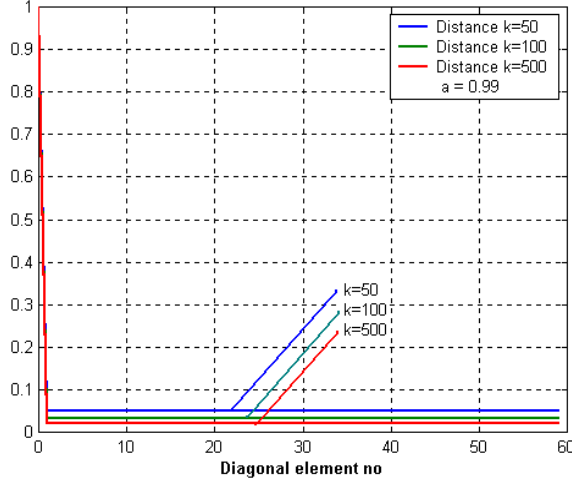


Figure 9.12: The diagonal elements, σ_i^2 , for the Cholesky decomposition of the covariance matrix.

$$p(y; \sigma_2^2) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{y^2}{2\sigma_2^2}\right) \quad (9.18)$$

The probability that the difference between two bottom profiles is less than ε is then approximately

$$\Pr(|\Delta \mathbf{h}| \leq \varepsilon) = \int_{-\varepsilon_1}^{\varepsilon_1} p_1(y; \sigma_1^2) dy \cdot \prod_{i=2}^N \int_{-\varepsilon_i}^{\varepsilon_i} p(y; \sigma_2^2) dy \quad (9.19)$$

A similar expression was derived in Chapter 4. The original integration limits will change with the transformation. For large N the integrand in the second product factor will tend to zero since the value of the integral is < 1 . The size of the elements in the diagonal matrix can be seen in Figure 9.12. For large k values $d_{ii} = 1 - a^2$ when $i > 1$ which follows from (9.8). We can see that the curves for $k = 50$ and $k = 100$ have a similar shape as the curve for $k = 500$. For $k = 500$ and $a = 0.99$ we have $d_{60,60} = 0.020032$ compared to $(1 - a^2) = 0.0199$.

9.5 The mean and variance of the correlation sum

Let the mode decomposition of the correlation matrix $\mathbf{C} = E\{(\mathbf{y} - \mathbf{h}(x))(\mathbf{y} - \mathbf{h}(x))^T\} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ then the correlation sum can be written as [CT92, KA93]

$$\begin{aligned} T(x) &= (\mathbf{y} - \mathbf{h}(x))^T (\mathbf{y} - \mathbf{h}(x)) = (\mathbf{y} - \mathbf{h}(x))^T \mathbf{V}\mathbf{V}^T (\mathbf{y} - \mathbf{h}(x)) \\ &= (\mathbf{V}^T (\mathbf{y} - \mathbf{h}(x)))^T (\mathbf{V}^T (\mathbf{y} - \mathbf{h}(x))) = \sum_{i=1}^N \alpha_i z_i^2(x) \end{aligned} \quad (9.20)$$

where $z_i(x)$ is $\mathcal{N}(0,1)$ and

$$\alpha_i = \sigma_{\Delta h, i}^2 + \sigma_e^2 \quad (9.21)$$

where σ_e^2 is the measurement noise variance and $\sigma_{\Delta h, i}^2 = \lambda_{\Delta h, i}$ is an eigenvalue of

$\mathbf{C}_{\Delta h}$, see Section 9.3. The stochastic variable $T(x)$ thus has a χ_N^2 looking distribution. The Figures 9.13 - 9.16 show some of the largest eigenvalues for $N=20, 120$ and $a=0.96, 0.99$ for different displacements k .

The distribution also has the following characteristics [NY01a]

$$E\left\{\frac{T}{N}\right\} = \sigma_{\Delta h}^2 + \sigma_e^2 \quad (9.22)$$

where $\sigma_{\Delta h}^2 = \frac{1}{N} \sum_{i=1}^N \sigma_{\Delta h, i}^2$ and

$$Var\left\{\frac{T}{N}\right\} = \frac{2\|\mathbf{C}\|_F^2}{N^2} \quad (9.23)$$

where $\|\mathbf{C}\|_F$ is the Frobenius norm of the covariance matrix. In the case the bottom profile is an $AR(1)$ process and the displacement k is large

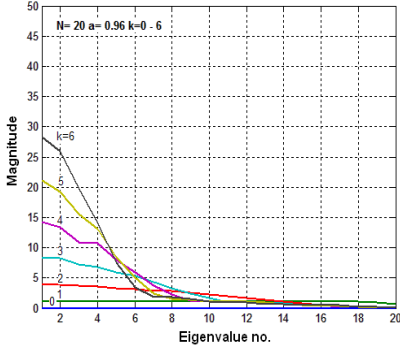


Figure 9.13: Eigenvalues for $N=20$ and $a=0.96$ for different displacements k .

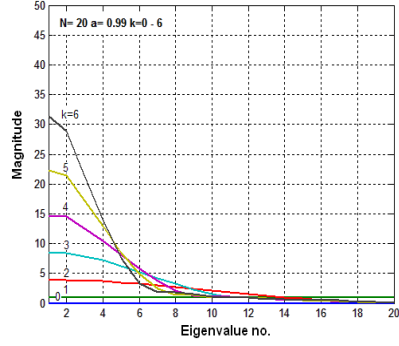


Figure 9.14: Eigenvalues for $N=20$ and $a=0.99$ for different displacements k .

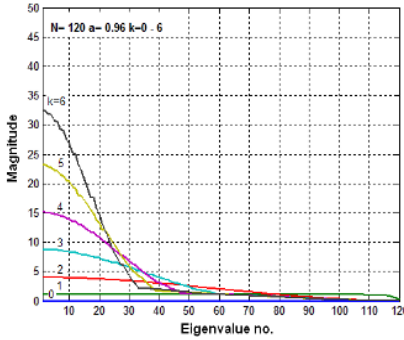


Figure 9.15: Eigenvalues for $N=120$ and $a=0.96$ for different displacements k .

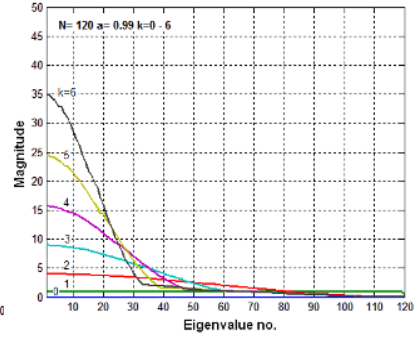


Figure 9.16: Eigenvalues for $N=120$ and $a=0.99$ for different displacements k .

$$\text{Var} \left\{ \frac{T}{N} \right\} = \frac{2\sigma_e^4}{N} + \frac{4\sigma_e^2\sigma_0^2}{N} + \frac{4\sigma_0^4}{N^2} \left[2a^2 \frac{a^{2N}-1}{(a^2-1)^2} - N \frac{(a^2+1)}{(a^2-1)} \right] \quad (9.24)$$

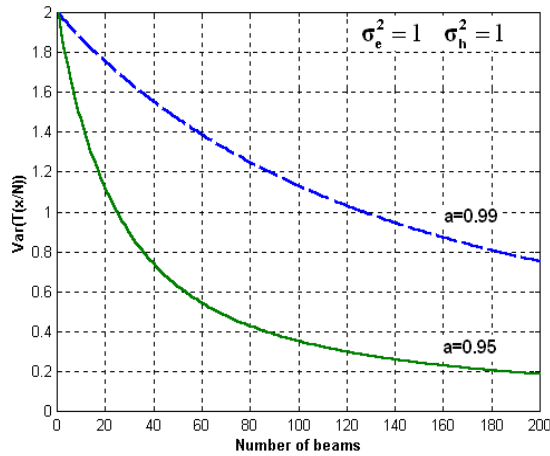


Figure 9.17: The variance of the correlation sum as a function of the number of measurement beams, N , at two different pole placements.

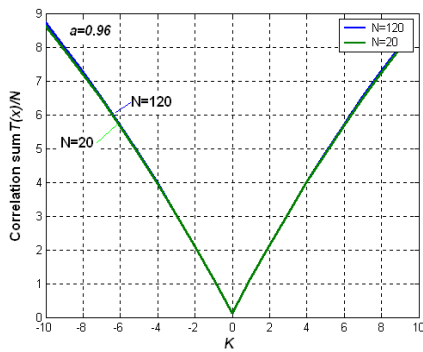


Figure 9.18: The expected value of $T(x)/N$ at $a=0.96$ for different displacements k .

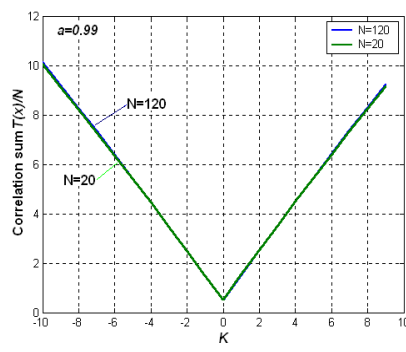


Figure 9.19: The expected value of $T(x)/N$ at $a=0.99$ for different displacements k .

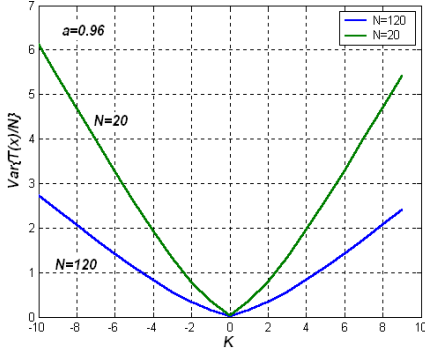


Figure 9.20: The variance of $T(x)/N$ at $a=0.96$ for different displacements k .

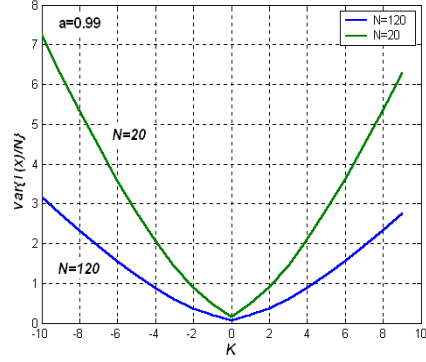


Figure 9.21: The variance of $T(x)/N$ at $a=0.99$ for different displacements k .

where a is the pole in the $AR(1)$ expression for the bottom profile, see Section 9.3. Figure 9.17 shows the variance for two different pole placements as a function of the number of beams. Figures 9.18 – 9.21 show how the expected value of the correlation sum and variance differ with the distance k between \mathbf{h}_1 and \mathbf{h}_2 .

9.6 The probability density function of the correlation sum

A simple way to determine the probability density function is to use generating functions. The characteristic equation for a χ_1^2 distributed stochastic variable

is $\phi_{\chi_1^2} = \frac{1}{\sqrt{1-2j\omega}}$ and therefore the characteristic equation for $T(x)$ will be

[NKB94, KA93]

$$\phi_T(\omega) = \prod_{i=1}^N \frac{1}{\sqrt{1-2j\alpha_i\omega}} \quad (9.25)$$

The inverse Fourier transform is

$$p_T(t) = \begin{cases} \int_{-\infty}^{\infty} \left(\prod_{i=1}^N \frac{1}{\sqrt{1-2j\alpha_i\omega}} \right) e^{-j\omega t} \cdot \frac{d\omega}{2\pi} & t \geq 0 \\ 0 & t < 0 \end{cases} \quad (9.26)$$

For every displacement k between the bottom profile and the measured profile we will have different covariance matrices where the elements $\rho(m)$ are calculated according to (9.6). The eigenvalues of the covariance matrix will differ, as can be

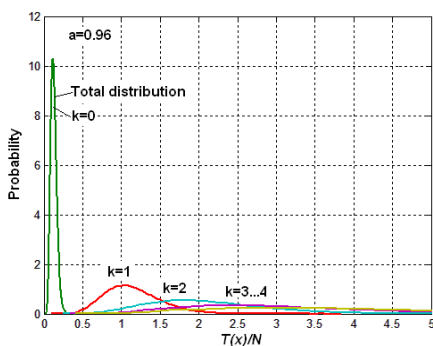


Figure 9.22: The probability density function for $T(x)/N$ while $N=20$, $a=0.96$ for different displacements k .

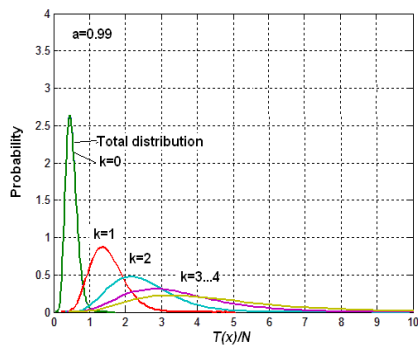


Figure 9.23: The probability density function for $T(x)/N$ while $N=20$, $a=0.99$ for different displacements k .

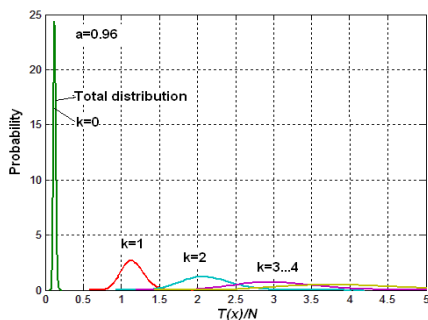


Figure 9.24: The probability density function for $T(x)/N$ while $N=120$, $a=0.96$ for different displacements k .

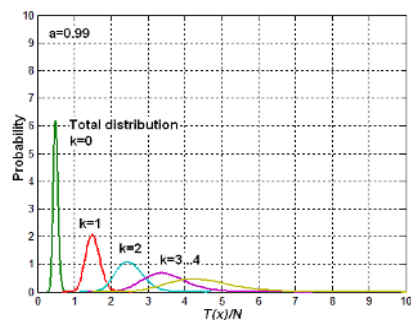


Figure 9.25: The probability density function for $T(x)/N$ while $N=120$, $a=0.99$ for different displacements k .

seen in Figures 9.13 – 9.16, depending on the displacement k and therefore we will have different probability density functions for the correlation sum $T(x+k)$.

The position estimate is determined from the position with smallest value of the correlation sum. The Figures 9.22 - 9.25 show, from a Monte Carlo trial, the probability density function for different k values but also the density function for the smallest correlation sum among all k values (the density functions are overlapping). As can be seen from the figures the density function for the smallest correlation sum is almost identical with the density function for $k=0$. This can especially be seen in Figures 9.24 and 9.25 where the overlap of the density functions between $k=0$ and $k=1$ is very small.

From the figures we see that a smaller value of a , i.e., a more hillier terrain, gives a better distinction between the density functions for the correlation sum and by that the risk of larger errors in the position decreases. We also see that a larger number of measurement beams has the same effect. We also note the well known fact that the χ_N^2 distribution asymptotically becomes normal when N increases. Since the variance also decreases as $1/N$, the likelihood function becomes close to normal which already has been proved in Chapter 4.

9.7 The probability density function for the position

If the joint probability density functions are known for M different displacements between \mathbf{h}_1 and \mathbf{h}_2 the density function for the smallest correlation sum within this truncated PDF can be determined.

First, in order to have an analytic expression, we calculate the probability that a displacement of k will give the smallest correlation sum. We will have

$$\begin{aligned} \Pr(T_{|k|=i} \leq T_{|k|=1..i-1,i+1..M}) \\ = \int_0^\infty \int_{T_i}^\infty \dots \int_{T_i}^\infty p(T_1, \dots, T_M) dT_1 \dots dT_{i-1} dT_{i+1} \dots dT_M dT_i \end{aligned} \quad (9.27)$$

where $p(T_1 \dots T_M)$ is the joint probability density.

To be able to determine the joint distribution we must know the correlation between the stochastic variables described by the particular density functions for different k , cf. Figures 9.22 – 9.25 and Figures 9.26 – 9.29 .

From the Figures 9.22 – 9.25 we can draw the conclusion that the smallest value of the correlation sum almost solely depends on the density function for $k=0$ and $k=1$. The correlation for the corresponding stochastic variables is shown in the Figures 9.26 – 9.29. From these figures we see that the major axes of the PDFs for $k>1$ are approximately oriented in the same direction. That means that a simultaneous diagonalization procedure may be applied [CT92] and thus making the correlation sums independent of each other. However, the most important distributions for determining the minimum of the correlation sum are those for $k=0$ and $k=1$ and we may only take those into consideration. The joint distribution of $k=0$ and $k=1$ are originally not truly independent because they have the same measurement noise but can, without approximation, be decorrelated by a simultaneous mode decomposition. A hillier terrain will make the stochastic variables $T(x)$ and $T(x+1)$ less dependent of each other. In the case we can consider the correlation sums independent after a simultaneous diagonalization procedure, exact or approximate, the integral (9.27) can be decomposed as

$$\begin{aligned}
 & \Pr(T_{|k|=i} \leq T_{|k|=1..i-1, i+1..M}) = \\
 & = \int_{-0}^{\infty} p_{T_i} \cdot \left\{ \int_{T_i}^{\infty} p_{T_1} \cdot \left\{ \int_{T_i}^{\infty} \dots \left\{ \int_{T_i}^{\infty} p_{T_{i-1}} \cdot \left\{ \int_{T_i}^{\infty} p_{T_{i+1}} \dots \right. \right. \right. \right. \\
 & \quad \left. \left. \left. \left. \int_{T_M}^{\infty} p_{T_M}(T_M) dT_M \right\} \dots dT_{i+1} \right\} dT_{i-1} \right\} \dots dT_1 \right\} dT_i
 \end{aligned} \tag{9.28}$$

The probability density functions for the distribution $p(T(x), T(x+1))$ and the number of trials which will give no position error, $T(x) < T(x+1)$, corresponds to the volume below the joint distribution above the 45 ° line, see Figures 9.30 and 9.31. The volume below the 45 ° line gives a position error corresponding to one grid distance.

If the stochastic variables $T(x)$ and $T(x+1)$ were independent and the cumulative frequency distribution is denoted by $F(\cdot)$ then the cumulative frequency

distribution for the maximum $T(x)$ is [PP02]

$$F_{T(x)} = 1 - [1 - F_{T(x|k=0)}][1 - F_{T(x|k=1)}] \quad (9.29)$$

The assumption of independence gives the expression (9.29) which seems to be quite accurate in this special case. The Figures 9.32 and 9.33 show a comparison between the approximation and the result from a Monte Carlo simulation. The figures show the distribution for the minimum correlation sum

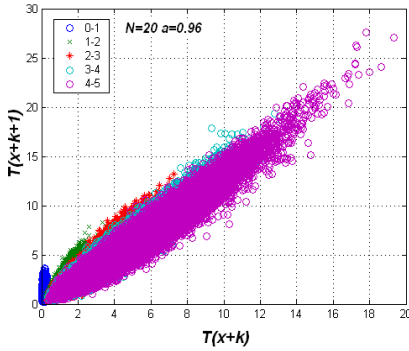


Figure 9.26: Correlation between $T(x+k)$ and $T(x+k+1)$ for $N=20$, $\alpha=0.96$ and $k=0,1,\dots,4$.

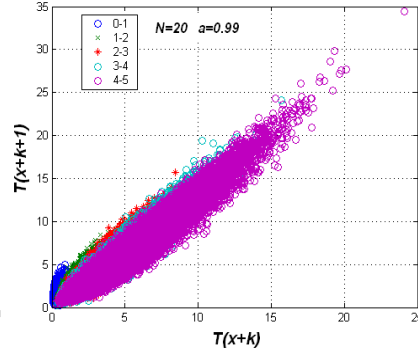


Figure 9.27: Correlation between $T(x+k)$ and $T(x+k+1)$ for $N=20$, $\alpha=0.99$ and $k=0,1,\dots,4$.

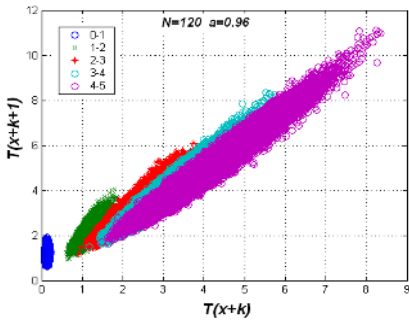


Figure 9.28: Correlation between $T(x+k)$ and $T(x+k+1)$ for $N=120$, $\alpha=0.96$ and $k=0,1,\dots,4$.

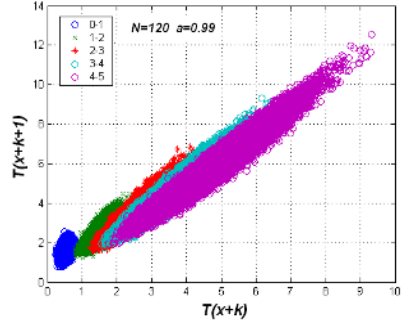


Figure 9.29: Correlation between $T(x+k)$ and $T(x+k+1)$ for $N=120$, $\alpha=0.99$ and $k=0,1,\dots,4$.

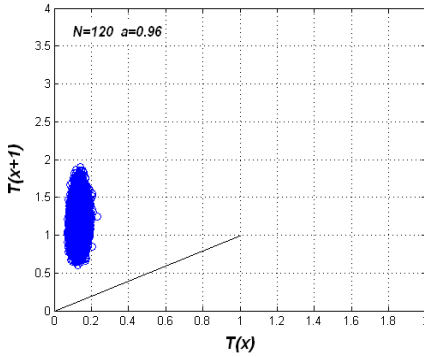


Figure 9.30: Correlation between $T(x)$, $T(x+1)$ for $N=120$, $a=0.96$.

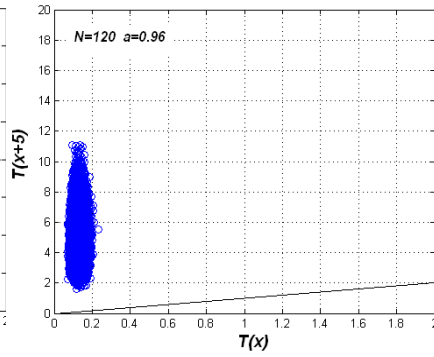


Figure 9.31: Correlation between $T(x)$, $T(x+5)$ for $N=120$, $a=0.96$.

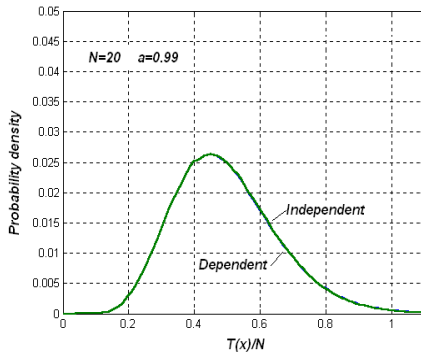


Figure 9.32: The probability density function for $T(x)/N$ for dependent and independent variables. $N=20$, $a=0.99$.

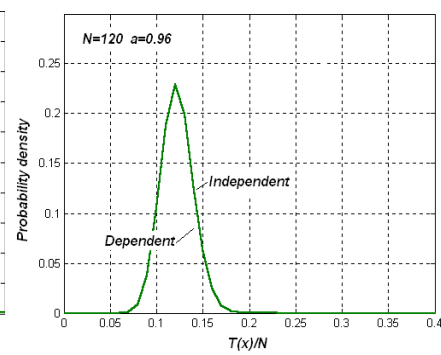


Figure 9.32: The probability density function for $T(x)/N$ for dependent and independent variables. $N=120$, $a=0.96$.

if the stochastic variables $T(x)$ and $T(x+1)$ is dependent or assumed independent As can be seen the difference is very small.

9.8 Approximate determination of the probability density function for the correlation sum

To determine the probability density function for $T(x)$ according to (9.26), i.e.,

$$p_T(t) = \begin{cases} \int_{-\infty}^{\infty} \left(\prod_{i=1}^N \frac{1}{\sqrt{1-2j\alpha_i\omega}} \right) \cdot e^{-j\omega t} \cdot \frac{d\omega}{2\pi} & t \geq 0 \\ 0 & t < 0 \end{cases} \quad (9.30)$$

demands a large computational work since it has to be determined numerically and an approximation is therefore of interest. From the Figures 9.22 – 9.25 we see the typical χ^2 shape. Let us therefore assume, as an approximation, that this is the case and let us determine the χ^2 distribution which has the same mean and variance as $p_T(\cdot)$, which we already know from (9.22) and (9.23).

$$E\left\{\frac{T}{N}\right\} = \frac{Tr(\mathbf{C})}{N} \quad (9.31)$$

$$Var\left\{\frac{T}{N}\right\} = \frac{2\|\mathbf{C}\|_F^2}{N^2} \quad (9.32)$$

For a χ_v^2 distributed stochastic variable, ξ , we have

$$\begin{aligned} E(\xi) &= v \\ Var(\xi) &= 2v \end{aligned} \quad (9.33)$$

where v is the degree of freedom in the distribution. This will give the density function for $T(x)$ approximately the same density as the stochastic variable

$$T(x) \sim \beta \chi_M^2 \text{ where [BSFO88]}$$

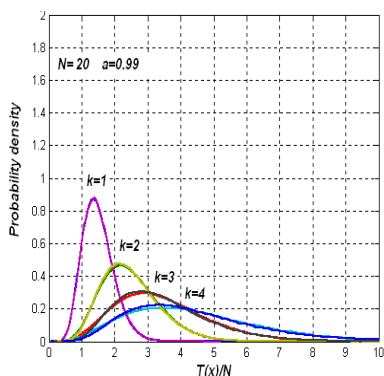


Figure 9.33: Approximative PDF compared to Monte Carlo simulated PDF for $T(x)/N$, $N=20$, $a=0.99$. The approximative PDF is slightly below the Monte Carlo PDF.

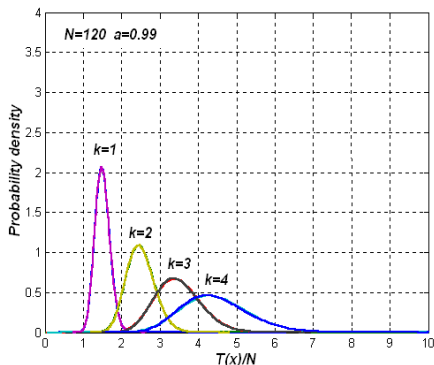


Figure 9.34: Approximative PDF compared to Monte Carlo simulated PDF for $T(x)/N$, $N=120$, $a=0.99$. The approximative PDF is slightly below the Monte Carlo PDF.

$$\beta = \frac{\|\mathbf{C}\|_F^2}{Tr(\mathbf{C})}$$

$$M = \frac{Tr(\mathbf{C})^2}{\|\mathbf{C}\|_F^2} \quad (9.34)$$

Figures 9.33 and 9.34 show the approximated probability density functions compared to those determined by Monte Carlo techniques and as can be seen there is a good agreement between the above approximation and the true PDFs.

The density function for $k=0$ is a true χ^2 density function and does not have to be approximated.

9.9 Cramér-Rao Lower Bound

The Cramér-Rao lower bound has already been discussed in Chapter 5. The presentation here will therefore be focused on the *AR*-assumption of the bottom topography.

Let the scalar parameter x denote our unknown position and $y(x)$ our measurement of the bottom profile and let the depth profile according to the map be denoted by $h(x)$. Then the estimate of the position has the bound on the variance

$$Var(\hat{x}) = E\{[\hat{x}(y) - x]^2\} \geq \left[E\left\{ \frac{\partial \log p(y|x)}{\partial x} \right\}^2 \right]^{-1} = \left[-E\left\{ \frac{\partial^2 \log p(y|x)}{\partial x^2} \right\} \right]^{-1} \quad (9.35)$$

if the position estimate is unbiased. The probability density function $p(y|x)$ is the density function for the observation vector in the true position x .

For the one dimensional case, in the case we have a Gaussian measurement error, this will give us

$$\frac{\partial \log p(y|x)}{\partial x} = -\frac{1}{2\sigma_e^2} \frac{\partial}{\partial x} \left(\sum_{k=1}^N (y_k - h_k(x))^2 \right) = -\frac{1}{2\sigma_e^2} \frac{\partial T(x)}{\partial x} \quad (9.36)$$

which gives after the expectation has been taken

$$Var(\hat{x}) = \sigma_e^2 \left[\sum_{i=1}^N \left[\frac{\partial h_i}{\partial x} \right]^2 \right]^{-1} \quad (9.37)$$

since for a Gaussian measurement error the position error variance will attain the CRLB when N tends to infinity, see also Chapter 5. Figure 9.35 shows how the CRLB for an *AR*(1)-process varies with the number of measurement points. The graph is determined by Monte-Carlo simulations. In general the CRLB is given by the inverse of the Fisher matrix \mathbf{J} calculated in the true position x which in this case is a scalar

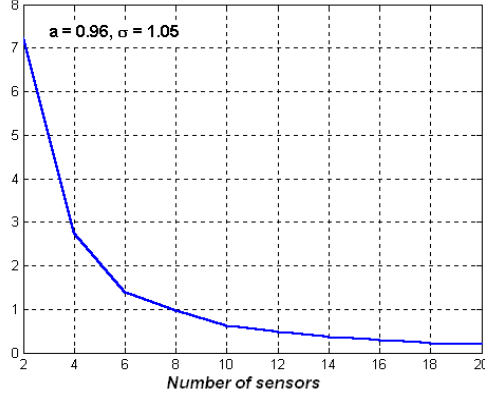


Figure 9.35: CRLB as a function of the number of measurement points N .

$$J = \frac{1}{2\sigma_e^2} E \left\{ \frac{\partial^2 T(x)}{\partial x^2} \right\} \quad (9.38)$$

We can estimate the derivatives in (9.37) and (9.38) under the assumption of an $AR(1)$ -process. Let h denote the map depth and rewrite the state equation in the following way

$$h_i = ah_{i-1} + v_i = -h_{i-1} + h_{i-1} + ah_{i-1} + v_i \quad (9.39)$$

or $h_i - h_{i-1} = (a-1)h_{i-1} + v_i$ which will give

$$E\{[h_i - h_{i-1}]^2\} = E[(a-1)h_{i-1} + v_i]^2 = (a-1)^2 E\{h_{i-1}^2\} + \sigma_v^2 \quad (9.40)$$

$$\text{For a stationary process we have } E\{h_i^2\} = E\{h_{i-1}^2\} = \sigma_h^2 \quad (9.41)$$

Further the state equation gives

$$\sigma_h^2 = a^2 \sigma_h^2 + \sigma_v^2 \quad (9.42)$$

or

$$\sigma_v^2 = (1 - a^2) \sigma_h^2 \quad (9.43)$$

Thus,

$$E\{[h_i - h_{i-1}]^2\} = (a - 1)^2 \sigma_h^2 + (1 - a^2) \sigma_h^2 = 2(1 - a) \sigma_h^2 \quad (9.44)$$

or alternatively

$$E\{[h_i - h_{i-1}]^2\} = \frac{2\sigma_v^2}{1 + a} \quad (9.45)$$

$$\begin{aligned} J &= \frac{1}{2\sigma_e^2} E \left\{ \frac{\partial^2 T(x)}{\partial x^2} \right\} = \frac{1}{\sigma_e^2} \sum_{i=1}^N E \left\{ \left[\frac{\partial h(x)}{\partial x} \right]^2 \right\} \\ &= \frac{1}{\sigma_e^2} \sum_{i=1}^N \frac{2\sigma_v^2}{1 + a} = \frac{2N\sigma_v^2}{\sigma_e^2(1 + a)} \end{aligned} \quad (9.46)$$

where we have substituted the terrain height derivative with the discrete derivative (9.45). The sampling interval is thus one grid distance.

For an $AR(1)$ generated bottom profile with the pole a near 1 the variance for the position estimate will be

$$Var(\hat{x}) = \frac{\sigma_e^2(1 + a)}{2N\sigma_v^2} \approx \frac{\sigma_e^2}{N\sigma_v^2} \quad (9.47)$$

An alternative expression is given if we substitute the terrain height derivative with the discrete derivative (9.44)

$$Var(\hat{x}) = \frac{\sigma_e^2}{2N(1 - a)\sigma_h^2} \quad (9.48)$$

where $\sigma_h^2 \approx \frac{1}{N} \sum_{i=1}^N (h_i - \bar{h})^2$ can be estimated directly from the map. For a specific value of the variance the σ_e^2 must be known or estimated by some other

means. This means that not only the roughness, σ_h^2/σ_e^2 but also the pole placement a affects the position accuracy.

The estimation of the bottom height variance can also be made from simplified spectral assumptions of the bottom profile since the autocorrelation or spectral density function can be seen as a measure how fast the terrain profile changes. We have as before (Chapter 6.2)

$$E\{h^2\} = \int_{-\infty}^{\infty} S_h(\omega) d\omega \quad (9.49)$$

see also Chapter 4.3.3 and Chapter 9.2.



Chapter 10

Summary

10.1 Summary

10.1.1 The problem

In the thesis a method for determining the position of underwater vehicles has been presented which to a large extent solves the difficult problem of aiding the inertial navigation system without breaking the surface for obtaining position fixes from GPS or radio navigation systems.

Terrain navigation systems with single beam methods for aiding the inertial navigation systems in flying vehicles have been around since the fifties. Also within the underwater community the single beam methods have been tested since long ago but the result has been discouraging independently of what estimation method has been used. The method has consequently not had any practical use in the past. The main reasons may be lack of underwater maps of sufficient accuracy and lack of robustness. The methods have also required the underwater terrain to have large height variations. By nature the underwater

terrain has smaller height variations than land terrain since the height variations in the sea have been smoothed during the milleniums due to sediment deposits. Other possible reasons may be that the methods require accurate starting positions to work. For submarines it is also important that the positions must be determined even if the submarine does not move in the water volume. Frequent pinging with the sonar is also not acceptable since it will reveal the presence of the submarine.

The aiding problem is also more difficult for underwater vehicles for the following reason. The estimated position is a fused position from the inertial navigation system and the terrain positioning system. The accuracy is thus dependent on the accuracy of the inertial navigation system. Now, the position error of the inertial navigation increases with the third power of the time [LA92] due to bias error in accelerometers and gyros. That means that the error from the inertial navigation system, for a given distance, will be much greater for the slow speed underwater vehicles compared to the fast moving flying vehicles. The prior position PDF for an underwater vehicle will thus have considerable greater variance than for a flying vehicle even if they use the same inertial system.

10.1.2 Solutions and conclusions

The presented positioning method has been tested in three larger sea-trials and has shown to be very accurate and robust with regards to different types of measurement errors and map errors. The robustness is a very important characteristics because even if a method sometimes works excellent it is useless if you do not know when this is the case. Occasional measurement errors due to objects in the water volume or due to interpolation errors of the depth or due to unmapped boulders should not cause the navigation system to loose its position.

The presented method does not need to know the starting position with any greater accuracy. It also works in very flat areas which means that the method can be used in much larger areas than the single beam methods. The method works equally well for moving or not moving vehicles and it uses only a very short sonar ping for determining the position which means that the risk of revealing is low. The method also lowers the requirement on the inertial navigation system compared to single beam methods for a specified required accuracy.

The drawback of the method is that it requires a true 3D sonar if all good characteristics should be fully achieved. However, this is a disadvantage which is quickly disappearing with all the new 3D-sonars that are introduced into the market with steady decreasing prices.

10.2 Further research

The main drawbacks of terrain navigation for underwater vehicles are lack of underwater maps and flat bottom areas. The lack of bottom terrain maps can be taken care of by further charting but doing something about the flat bottom areas seems in most cases impossible. The terrain navigation method examined in this thesis extends the use of terrain navigation in flat bottom areas considerably but still there is a lower limit of the needed bottom roughness to obtain a unique position. However, the needed roughness is always in relation to the measurement error so by decreasing the measurement error the area useful for positioning may be increased. A prerequisite is, of course, that the underwater map has sufficient accuracy so the map is not setting the bound for the positioning accuracy.

It has been the experience during the sea-trials that the largest measurement error in many cases is due to incomplete compensation of the sound speed gradient in the water although the sonar manufacturers procedures for compensation have been used. A future research area could therefore be to look into the possibilities to adaptively adjust the compensation of the speed gradients after each terrain profile measurement. Such a method would also make the cumbersome sonar calibration procedure easier.

Other sonar measurement methods, i.e., interferometric measurement methods instead of beam forming methods are another area for future research.

The sonars that have been used in the sea-trials mentioned in the thesis all have been “off the shelf type” sonars from the manufacturers and they have all worked well. However, it seems likely that implementing the measurement concept shown in Chapter 8, Figure 8.4, in a standard bathymetric sonar could give a highly cost effective navigation sonar with high performance. The research field would here be to find suitable signal forms for such a sonar.

The terrain navigation method in the thesis has not been tested for flying vehicles but it seems likely that it would solve two difficult problems that are encountered in current methods. The first problem is the establishing of the terrain profile in forest areas where the radar measures the distance to the tree tops instead of measuring the distance to the true ground. Using narrow radar beams for the terrain measurement would mean that some beams always will reach the true ground and by that the true terrain profile can be established. The second problem, in single beam methods, is the estimation of the measurement bias. This is today mostly estimated as if it was a constant measurement error but in reality it can rapidly change from one measurement to another. The first of these problems may be an area for further research while the second is solved in the thesis.



Appendix A

Inertial navigation systems and the hardware navigator

A.1 Inertial navigation

When the vehicle is moving from navigation cell A to cell B, see Figure 3.5, it is vital that the vehicle ends up being in cell B so it can do the correlation with the underwater map and determine its position otherwise it will lose its position. The question is thus, how large does the navigation cell B have to be? To answer this question a brief review of the principles for inertial navigation will be given.

A.1.1 The principle of an inertial navigation system

A gyroscope, or gyro for short, is a device that keeps its orientation relative the far stars if it is without losses. They can thus be used as pointers of the axis of a coordinate system. A triad of gyros and accelerometers as in Figure A.1, can therefore, if attached to a vehicle, measure the displacement of a vehicle by integrating the accelerations twice in the directions pointed out by the gyros.

There are basically two types of inertial navigation systems. In the first and the original system, the platform system, the sensor triad is mounted on a platform

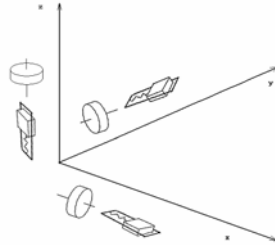


Figure A.1: The basic principle of an inertial navigation system. The accelerations in the directions pointed out by the gyros are integrated twice.

which is held free of rotation relative to the far stars by the output from the gyros. This type of system is still the most accurate system but it is costly and requires relative large mounting space in the vehicle.

In the strapdown IN system, which is becoming more and more used, the gyros and the accelerometers, the sensor triad, are rigidly mounted directly on the vehicle and by that the triad measures the acceleration and rotation of the vehicle in vehicle coordinates which then have to be transformed to a more suitable coordinate system. By the increasing computational capacity of computers the strapdown system has become more and more accurate and has gained acceptance. Compared to a platform system the gyros have to measure the vehicle rotation and not just keep the platform non-rotating so a higher requirement is placed on the gyros. An advantage of the strapdown system, besides its lower requirement of space, is that it is more mechanically robust than a platform system.

Now, gyros have losses which means that they will slowly drift away from the initial directions. They also have bias, i.e., an output signal in spite of no rotation. They also have scale errors. Accelerometers have bias errors and scale errors. Besides these types of errors, there are the random "white" type errors in the output signals. The result of all this is that the position given by an IN system will have an error which is increasing exponentially with time so the system has to be frequently reset by known position fixes.

The most cumbersome type of errors are the bias errors because they vary slowly with time. This type of errors is also called "day-to-day" errors. The analysis in the next paragraph does only consider this type of error because

they determine the time for resetting of the IN-system. When analyzing the system performance the bias errors are considered being random zero mean stochastic variables with a given variance.

There are many different principles for designing gyros. The first ones were based on a mechanically rotating mass. Today resonant laser ring gyros (RLG) and fiber optic gyros (FOG) are frequently used in high performance systems. Both types rely on the Sagnac effect. A 1 *Nautical Mile/hour* system is typically based on laser ring gyros. Accelerometers are based on measuring the force on a small mass, directly or indirectly as measuring the frequency change of vibrations due to that force. A review of different components in IN-systems can be found in [LA92].

A.1.2 The position error for the INS

The accuracy of IN-systems has to a great extent been treated in the literature. The study below is in large part taken from [TIWE97] but could also be found in [BRW97], see also [MIMI93, BI99, JE01]. One assumption for the simplified treatment is that the vehicle is moving in a small area, say 50 x 50 km with low speed which will be the case for underwater vehicles.

The errors in an IN-system can roughly be categorized in bias error, random walk errors and other errors. The bias errors, are stochastic errors that are left after the best nulling of the errors. The random walk errors are random errors that vary more often and are easier to remove using the vehicle dynamics (Kalman filtering).

The bias errors and the errors from initializing and setting of the IN-system will determine the size of the positioning area over which the likelihood function should be calculated.

We will in order to simplify the analysis assume that the vehicle has a northern heading and a speed of 5 knots. The position errors in the northern and eastern directions can be assumed to be independent of each other. A model for studying the influence of gyro and accelerometer bias can be formulated, see Figure A.2. The model will be the same for both platform and strapped down systems. The model is valid for movements within a limited area, for example, between navigation cells.

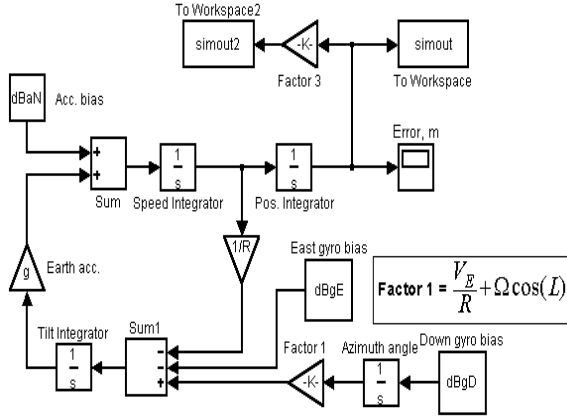


Figure A.2: Simulink model for the Northern channel for studying the influence of gyro and accelerometer bias. The Eastern channel has a similar model.

The navigation principle is that the accelerometers in the IN-system are integrated twice in the directions pointed out by the gyros. It is customary to look at the errors instead of the positions and the Simulink schematic refers to the errors.

In Figure A.2 dBaN means the accelerometer bias in the northern direction and dBgE and dBgD means the gyrobias in the eastern and in the down direction and L is the latitude. The performance of the IN-system is highly dependent on the size of the bias in the accelerometers and in the gyros. The differential equation for the model in Figure A.2 can be solved and we will have the following analytical expression for the 1 σ -limit of the position error [JE01]

$$e(t) = \frac{gB_G t}{\omega_s^2} \left(1 - \frac{\sin(\omega_s t)}{\omega_s t}\right) + B_A \frac{1 - \cos(\omega_s t)}{\omega_s^2} \quad (\text{A1})$$

where B_G and B_A are the RMS-values for the gyro and accelerometer bias, g is the earth acceleration and ω_s is the Schuler frequency with a period time of 84.4 minutes.

If B_G and B_A refer to independent stochastic variables with variances $\sigma_{B_G}^2$ and $\sigma_{B_A}^2$, respectively, the standard deviation will be

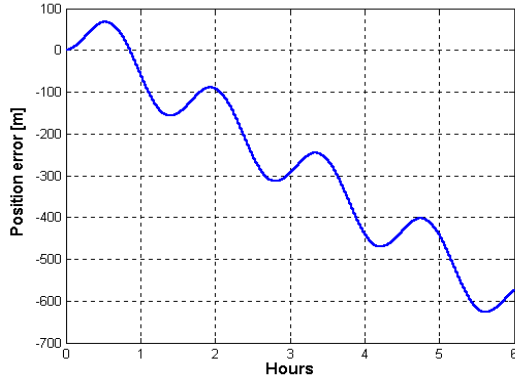


Figure A.3: Total position error for a submarine IN-system due to accelerometer- och gyrobias according to the model in Figure A.2.

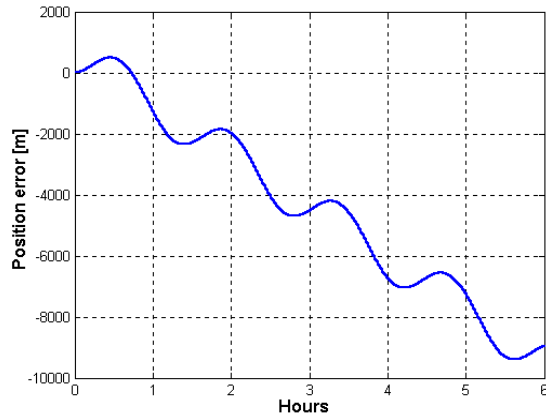


Figure A.4: Total position error for a good IN-system due to accelerometer- och gyrobias according to the model in Figure A.2.

$$\sigma(t) = \sqrt{\left[\frac{gt}{\omega_s^2} \left(1 - \frac{\sin(\omega_s t)}{\omega_s t} \right) \right]^2 \sigma_{B_G}^2 + \left[\frac{1 - \cos(\omega_s t)}{\omega_s^2} \right]^2 \sigma_{B_A}^2} \quad (\text{A2})$$

For a vehicle with submarine performance we will have a considerably smaller error than for a good ordinary IN-system in the first place due to a lower gyro bias. Figures A.3 och A.4 show typical position errors for a good ordinary system, a so called 1 *Nautical mile /h* system and a good submarine system.

A.2 The hardware navigator

The advantages with the correlation method is won at the expense of the computing time due to the larger number of measurement beams. Figure A.5 shows how the computing time for the likelihood function depends on the number of measurement beams for one grid point. The number given in the diagram should be multiplied with the total number of gridpoints. Here, the correlation is assumed to be done with a local map with the same grid so no interpolation between grid points is assumed.

The time for computing the likelihood function for $9 \times 9 = 81$ beams will be about 40 times longer than that using only one beam. That means computing times in the range of 4 s for a map with 501×501 grid points which also means that the vehicle travelling with 4 knots will have a position answer not until the vehicle has travelled ~ 100 meter from the actual position. If, in addition, the bearing is

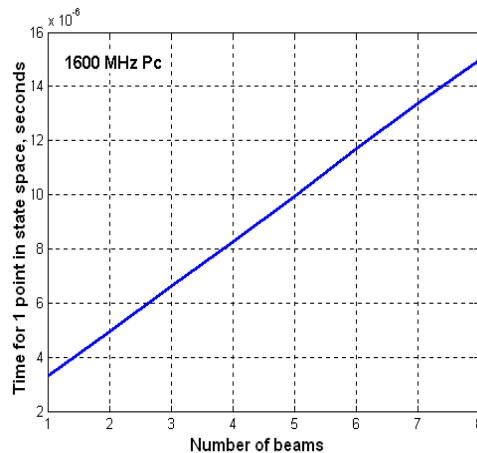


Figure A.5: The time for computing a likelihood value in state space for different numbers of measurement beams.

to be determined by doing several positionings with a rotated map it will be further delayed. If interpolation is needed the computing time will be considerably longer.

However, it turns out that the calculation of the quadratic sum is easily mechanized so the complete correlation of a whole map can be done in a few hundreds of a second or less with modern computer technology.

A.2.1 The basic principle

A way of reducing the computing time is to implement the correlation algorithm in a FPGA (Field Programmable Gate Array) and two ways of doing that is described here.

Figure A.6 shows a part of a 5 x 5 km underwater map with a local sonar map inserted. The sonar map contains gridded depth data from the sonar. In the figure the sonar map has only 3 x 3 nodes usually it is much larger. The sonar map is moved in steps of one node distance horizontally and vertically until a comparison has been made with all nodes in the map. A sonar map is thus created at each sonar ping and correlated against the larger underwater map to determine the coordinates for the best fit.

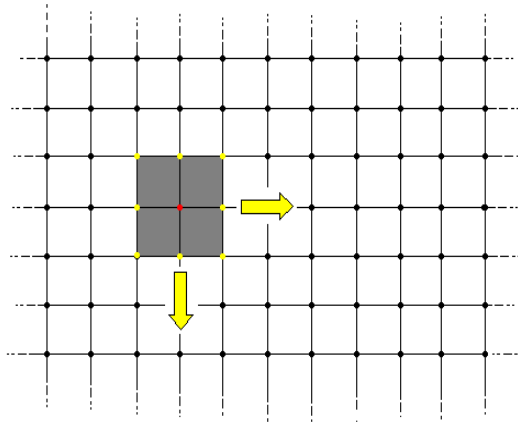


Figure A.6: Map grid with a local sonar map inserted. The arrows indicate the movement of the sonar map.

For each node in the larger map, which can be reached by the sonarmap's reference point, we calculate

$$V(i, j) = \sum_{(r,s) \in \text{sonar map}} |d_{i+r,j+s}^{tm} - d_{i+r,j+s}^{sm}| \quad (\text{A.3})$$

where d is the mean compensated depth value in the node. Thus, $V(i, j)$ is the sum of the absolute differences of the mean compensated depth values in the terrain map (tm) and the sonar map (sm) and (i, j) are the coordinates of the reference point.

The reason for looking at the absolute values (MAD - Mean Absolute Distance) is that this algorithm is easier to implement in the FPGA chip than the square of the depth errors. The reason for looking at the differences from the mean depth is due to the problem of establishing a common reference altitude for the terrain map and the sonarmap, see Chapter 3.

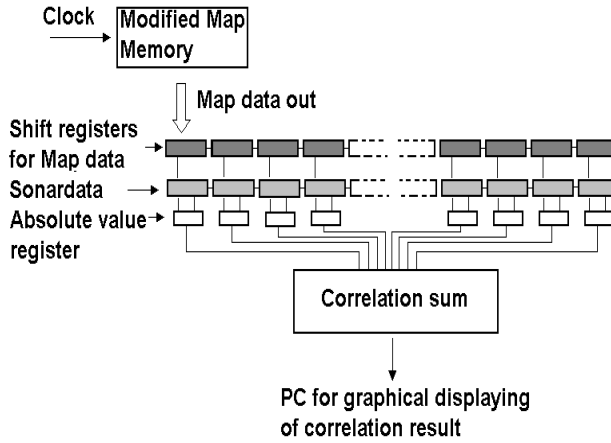


Figure A.7: One principle for a correlator.

In normal cases the terrain map will have 501 x 501 nodes and the sonar map 10 x 10 nodes. The distance between the nodes is 10 meters.

A.2.2 The principle of implementation

One principle for a correlator is shown in Figure A.7. The correlator consists of two lines of shift registers, one is for the terrain map data and the other is for the sonar map data. The number of shift registers in the lines is 100 if the size of the sonar map is 10 x 10 nodes. The length of the shift registers is 16 bits.

Mean value compensated gridded depth data from the sonar are shifted in from the left in the lower line and will be fixed there during the whole correlation process. The modified map memory has mean value compensated depth data organized so that the correct data will always be shifted out at each clock pulse.

It turns out that in different proposals the memory access time will always be the bottleneck. A recursive solution where some of the shift registers are replaced by a recursive adding loop, Figure A.8, can therefore be used without hampering the performance of the correlator. Such a solution will greatly reduce the number of gates needed in the implementation.

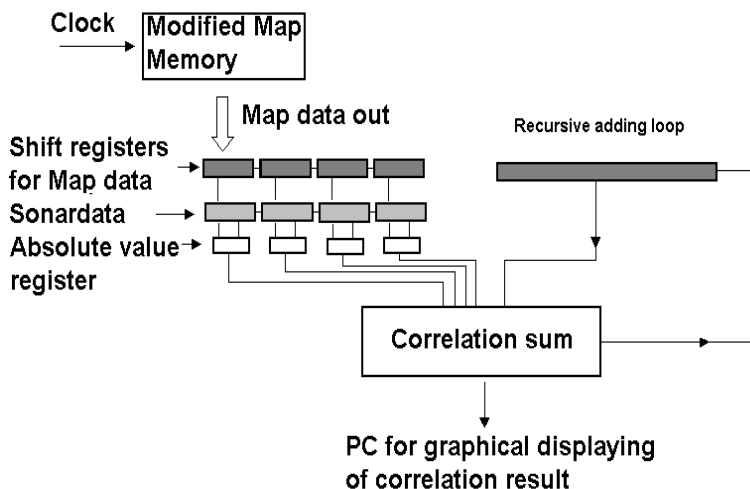


Figure A.8: Recursive addition instead of shift registers.

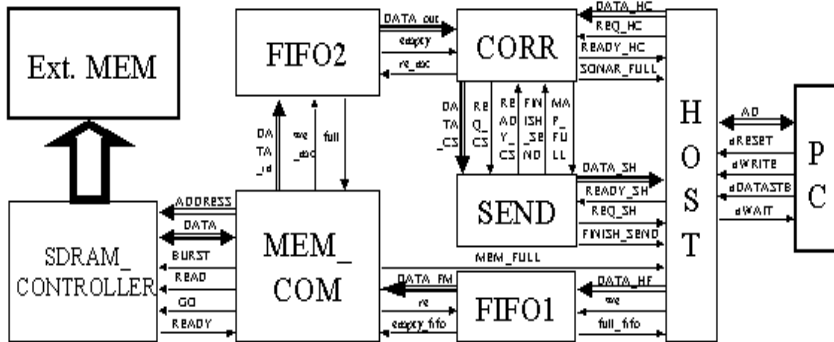


Figure A.9: Block schematic of the implemented terrain correlator. From [ANTR02].



Figure A.10: Photo of the completed prototype correlator. Natural size.

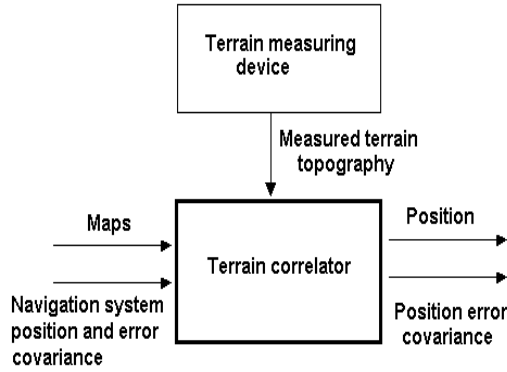


Figure A.11: Schematic of the terrain navigator.

A2.3 The practical implementation in an FPGA

A hardware correlator has been implemented as a master thesis work according to the principle with recursive addition [ANTR02], see also [TALL03]. A block schematic of the solution can be seen in Figure A.9 and the completed correlator in Figure A.10.

The terrain navigator in Figure A.11 is built around the hardware correlator in Figure A.10. The input to the navigator is, besides maps, the current position and its error and the measured terrain topography. Compared to the INS system bias error the position error from the terrain correlation is small. The main reason for the input of the current position error is therefore to restrict the search area in the correlation process in order to minimize the probability of false positions. The search area is typically a rectangle circumscribing the $3\text{-}\sigma$ uncertainty ellipse. Output from the navigator is besides the position the calculated position error variance. Output from the navigator can also be bearing, bearing error, speed and speed error. However, since these quantities often easily can be obtained from other sources this is assumed in the following discussion.

Figure A.12 shows a simplified schematic of the north channel for an INS-system showing the position error due to biases in the accelerometers and down and east gyros [TIWE97, MIMI93, BI99, JE01]. The figure shows symbolically three possible circle marked points of integration. Integrating at point A, feed-forward

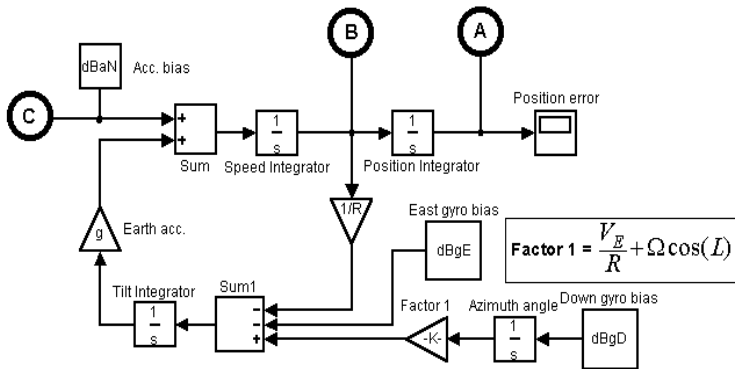


Figure A.12: Simplified schematic of the North Channel of an INS-system with marked places for possible integration of the hardware correlator.

integration, means that only the positions are fused together as described in Chapter 4. Integrating at point B, feed-back integration, means that we are resetting the INS-system error states for position and speed. Integrating at point C means that we are nulling out the biases in accelerometers and the gyros. This is called tight integration and may give the best result since it will remove the cause to

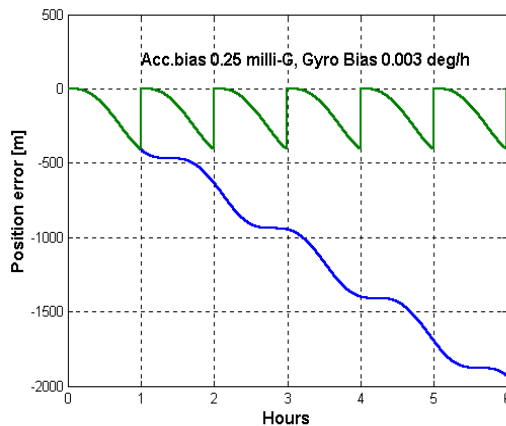


Figure A.13: The 1σ position error for a high performance INS-system with and without resetting by the correlator (speed and position).

the INS-drift. What we actually do is estimating the bias components in the Kalman filter and then subtract them from the inputs of the accelerometers and gyros.

Figure A.13 shows the result from a simulation of the INS-position error according to the schematic in Figure A.12. As expected the resetting gives an upper bound on the INS position error. The INS-error after nulling the bias-errors is extremely sensitive to the quality in the estimate of the bias term so it should not be performed before the Kalman filter has been settled. The integration of the navigator output resembles to a high degree the integration of a GPS-system to an INS-system and has been exhaustively studied in the literature.

The conclusion of this section is that it in most cases will be sufficient with a simple feedback of position/speed, i.e., the alternatives A and B.



Appendix B

Interpolation in the map

B.1 Introduction

Figure B.1 shows a typical footprint from the measurement beams against the bottom in the case a 3D topographic bottom picture is assembled from five beam planes. A picture of the footprint of a true 3D-sonar can be seen in Figure 8.7. In order to calculate the likelihood function the depths at the individual footprint

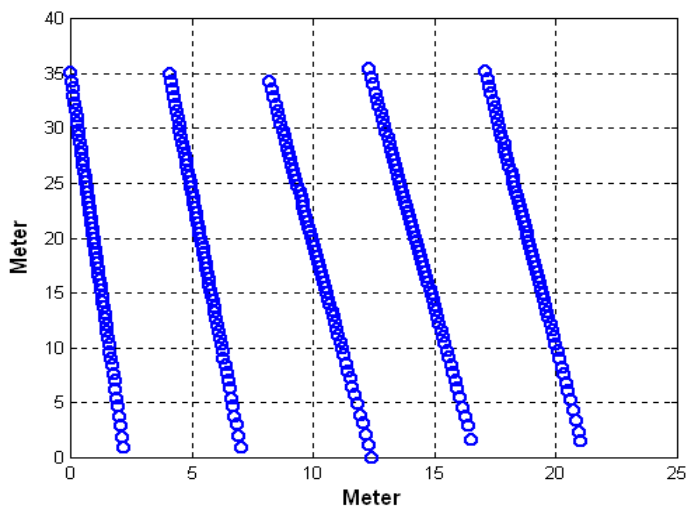


Figure B.1: The footprint from 5 beam planes.

points have to be determined. This can be done by signal reconstruction, if the bottom profile has been properly sampled, or by interpolation.

B.2 Reconstruction of signals from samples

We will restrict the discussion here to one dimensional signals. A discussion of sampling and reconstruction of multidimensional signals can be found in [BP71]. It is well known that when sampling a signal the sampling frequency has to be more than twice the highest frequency in the signal if the signal is to be uniquely recovered from the samples [HN28]. This is easy to understand if it is observed that a finite time signal can be uniquely represented by a Fourier series. To be able to uniquely reconstruct the highest frequency in the Fourier series the distance between the samples must be less than half the period time of this signal component, Figure B.2.

If the signal contains frequencies higher than half the sampling frequency the sampled signal spectra will wrongly have increased amplitudes for frequencies less than half the sampling frequency. This because the high frequency components will mistakenly be assumed to belong to the lower frequency components. This phenomenon is called signal aliasing. To avoid signal aliasing due to high frequency noise the signal is often low pass filtered to remove all frequencies above half the sampling frequency. This makes it possible to uniquely reconstruct the filtered signal but of course not the original signal.

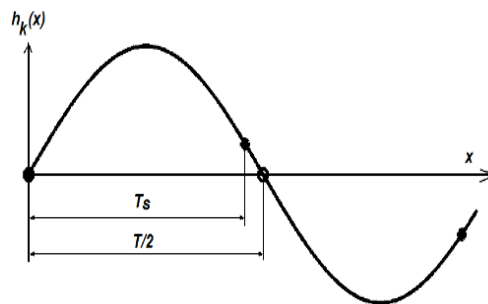


Figure B.2: If the spatial sampling distance T_s is less than half the period of the sinusoid it can be uniquely reconstructed.

It is easy to formally analyze the aliasing phenomena if the sampled signal can be assumed to be the product of the continuous signal and an periodic impulse-train, the sampling function [OW97, SH90, PK92]. Let the continuous signal be $h(x)$ and the sampled signal $h_s(x)$ then

$$h_s(x) = h(x)f(x) \quad (\text{B.1})$$

if $f(x)$ is the impulse train

$$f(x) = \sum_{n=-\infty}^{\infty} \delta(x - nT_s) \quad (\text{B.2})$$

where $\delta(\cdot)$ is the Dirac function and T_s the sampling interval and it has been observed that the integral of the Dirac functions represents the actual signal values at the sampling event.

Since multiplication in the time domain is equivalent to convolution in the frequency domain the spectrum of the sampled signal will be [OW97, SH90]

$$S_{hs}(\omega) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} S_h(\omega - n\omega_s) \quad (\text{B.3})$$

where S_h is the spectrum for the continuous signal, the continuous bottom profile, and $\omega_s = 2\pi/T_s$ is the sampling frequency. The spectrum for the sampled

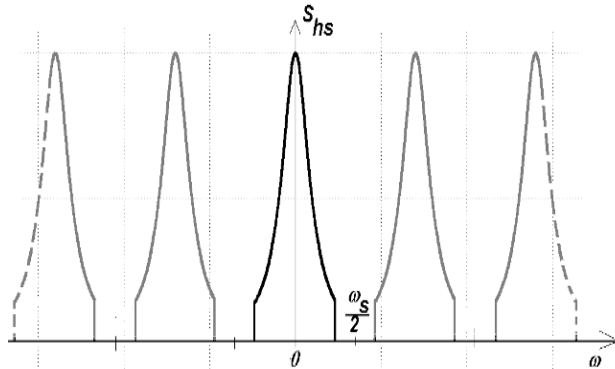


Figure B.3: The power spectrum for the sampled signal $h_s(x)$ of a band limited continuous signal $h(x)$.

bottom profile can thus be seen as a superposition of the shifted and scaled band limited base spectrum S_h , see Figure B.3.

The power spectral density for the sampled $AR(1)$ bottom profile is

$$|S_{hs}(j\omega)|^2 = \frac{aT_s}{2} \frac{1 - \exp(-2aT_s)}{1 + \exp(-2aT_s) - 2\exp(-aT_s)\cos(\omega T_s)} \quad (\text{B.4})$$

where T_s is the sampling period. The spectrum repeats itself for $|\omega| > \pi/T_s$. Figure B.4 shows the aliasing effect for different aT_s for the $AR(1)$ process. All curves show aliasing since the $AR(1)$ process is not band limited but the highest sampling rate $aT_s=0.5$ gives only a small amount of aliasing.

Underwater maps are mainly produced for the safety of shipping and that means that in shallow waters if it is a risk of going ashore the sampling is dense but in other areas there are long distances between samples. Naturally the distance between samples is determined by the speed of the ship doing the charting, the depth to the bottom and the pulse repeating frequency of the sonar. The speed of the sound pulse is ~ 1450 m/s which means a turn around time of the sonar pulse of ~ 0.2 s in shallow waters and ~ 2 s in deep waters. With a ship speed of 5 m/s (10 knots) this means a spatial distance of 1 – 10 meter between samples.

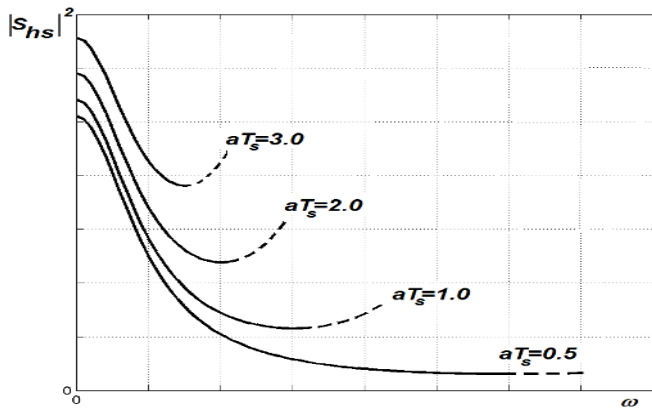


Figure B.4: Aliasing effect for different sampling rates for an $AR(1)$ process.

In the case of a multibeam sonar this will be the distance between beam planes. Usually a multibeam sonar has a spacing of 1.5° between beams which means that the distance between the samples along the beamplane will be 4 – 40 meter in the center.

From the discussion above it is clear that the spectrum of an underwater map always will have some aliasing. It is also clear that the degree of aliasing will vary between bottom areas depending on how dense the sampling once was done. The consequence of this is that the bottom profile can not be reconstructed from the samples using the reconstruction formula

$$h(x) = \sum_{n=-\infty}^{\infty} h(nT_s) \frac{\sin(\frac{\omega_s}{2}(x - nT_s))}{\frac{\omega_s}{2}(x - nT_s)} \quad (\text{B.5})$$

That is, we will never know the exact shape of the bottom profile between the sample points. To reconstruct a signal from its samples, where it is theoretically possible, we have to sum the slowly converging infinite series (B.5) which means in practice that we have to truncate it after a fairly high number of terms. It is, however, interesting to note that (B.5) can be thought of as an expansion of the bottom profile as

$$h(x) = \sum_{n=-\infty}^{\infty} c_n \phi_n(x) \quad (\text{B.6})$$

where c_n in this special case is the sampled bottom depth and ϕ_n is the sinc function. A deeper analysis [OJ72] shows that there are many other choices of function classes and coefficients possible for a true reconstruction. This, together with the discussion above, has lead to the use of interpolation instead of reconstruction to determine the depth values between the depth samples. The following chapters will discuss some interpolation methods that preserve the spectrum of the bottom.

The underwater map often has a grid size of 10 meter but it can also be 5 or 1 meter so it will be necessary to interpolate in the map to determine the actual depth values at the footprint points. Deterministic methods as Lagrange, Hermite or spline interpolation [BJ74, MU99], well known since long ago, can be used but depending on the bottom topography the result may differ and perhaps not be satisfactory. In certain areas one method will work well but in other areas not so well. However, the difference is often not so large that it will be crucial for the

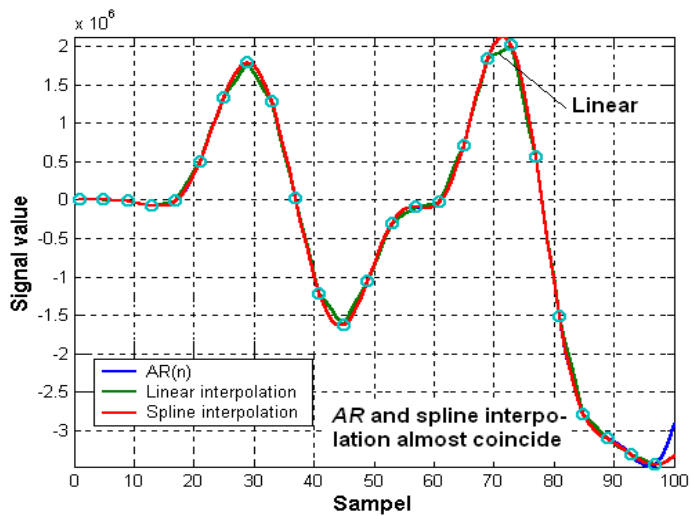


Figure B.5: An $AR(20)$ signal and the signal interpolated by the linear and the cubic spline methods at certain points.

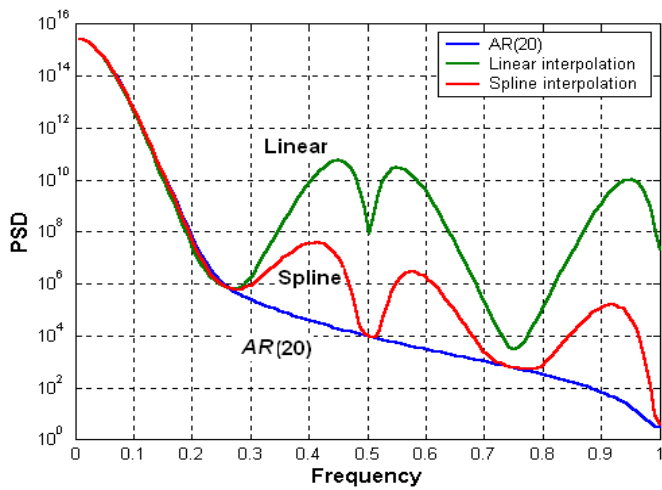


Figure B.6: The power spectrum for the signals shown in Figure B.5. Note the clear difference in spectra in spite of the similarity between the AR curve and the spline curve in Figure B.5.

result when using many measurement beams.

A consequence of the deterministic methods is that they change the statistical properties of the bottom topography, especially they change the power spectrum at higher frequencies which will impact the position accuracy. Figure B.5 shows an $AR(20)$ signal, a linearly interpolated signal and cubic spline interpolated signal. The corresponding power spectra can be seen in Figure B.6 and as can be seen there is a difference at higher frequencies.

Now, the expected value of the square of the gradients, which determines the expected value of the accuracy that can be calculated from (see Chapter 6)

$$\mu(n) = E \left\{ \left[\frac{\partial^n h}{\partial x^n} \right]^2 \right\} = \int_0^\infty \omega^{2n} S_h(\omega) d\omega \quad (\text{B.7})$$

when $n=1$. This will thus give us the Fischer matrix (FIM), i.e., what we can expect it to be in the terrain considered. We use the Fischer matrix to calculate the position error covariance matrix. Therefore it would be desirable to have an interpolation method that does not change the statistical properties of the bottom topography, i.e., the power spectra. In the following three sections different probabilistic approaches will be introductorily discussed.

B.3 Maximum a posteriori interpolation

Let us look at a case with a band of missing samples since it clearly demonstrates the power of the method. The interpolation case for a finer grid follows immediately. Assume the length of the signal is N and that a band of M samples is missing, Figure B.7. We arrange the signal vector in known and unknown values as follows

$$\begin{aligned} \mathbf{x} &= \begin{bmatrix} \mathbf{x}_{Kn_1} \\ \mathbf{x}_{Un} \\ \mathbf{x}_{Kn_2} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{Kn_1} \\ 0 \\ \mathbf{x}_{Kn_2} \end{bmatrix} + \begin{bmatrix} 0 \\ \mathbf{x}_{Un} \\ 0 \end{bmatrix} \\ &= \mathbf{K}\mathbf{x}_{Kn} + \mathbf{U}\mathbf{x}_{Un} \end{aligned} \quad (\text{B.8})$$

where the matrices \mathbf{K} and \mathbf{U} only rearrange the signal values.

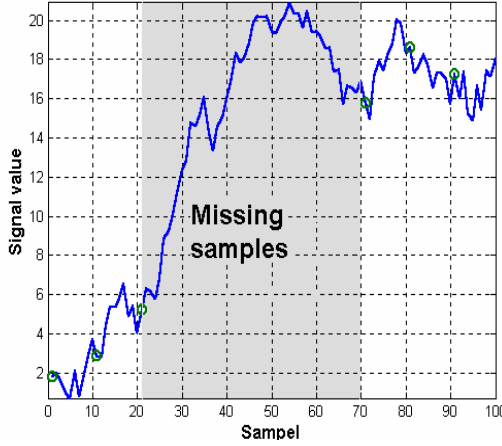


Figure B.7: A signal with a band of missing samples, the shaded area.

The probability density function for the signal can be written as

$$p_{\mathbf{x}}(\mathbf{x}_{Kn}, \mathbf{x}_{Un}) = p_{\mathbf{x}}(\mathbf{x}_{Un} | \mathbf{x}_{Kn}) p(\mathbf{x}_{Kn}) \quad (\text{B.9})$$

or

$$p_{\mathbf{x}}(\mathbf{x}_{Un} | \mathbf{x}_{Kn}) = p_{\mathbf{x}}(\mathbf{x} = \mathbf{K}\mathbf{x}_{Kn} + \mathbf{U}\mathbf{x}_{Un}) / p_{\mathbf{x}}(\mathbf{x}_{Kn}) \quad (\text{B.10})$$

where $p_{\mathbf{x}}(\cdot)$ is assumed to be known and $p_{\mathbf{x}}(\mathbf{x}_{Kn})$ is a constant. We now search for the \mathbf{x}_{Un} that maximizes the PDF

$$\hat{\mathbf{x}}_{Un}^{MAP} = \arg \max_{\mathbf{x}_{Un}} p_{\mathbf{x}}(\mathbf{x} = \mathbf{K}\mathbf{x}_{Kn} + \mathbf{U}\mathbf{x}_{Un}) \quad (\text{B.11})$$

In the case we have a Gaussian signal with known covariance matrix $\Sigma_{\mathbf{xx}}$

$$p_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2\pi)^{N/2} \det(\Sigma_{\mathbf{xx}})^{1/2}} \exp\left(-\frac{1}{2} \mathbf{x}^T \Sigma_{\mathbf{xx}}^{-1} \mathbf{x}\right) \quad (\text{B.12})$$

Our estimate of the missing samples is

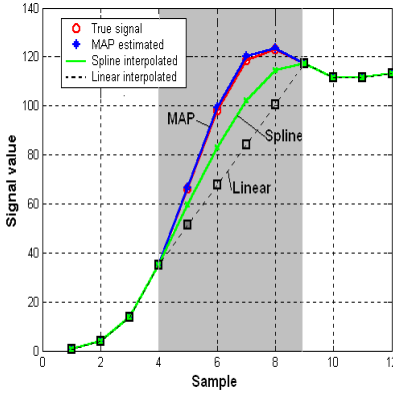


Figure B.8: MAP interpolation compared to linear and spline interpolation

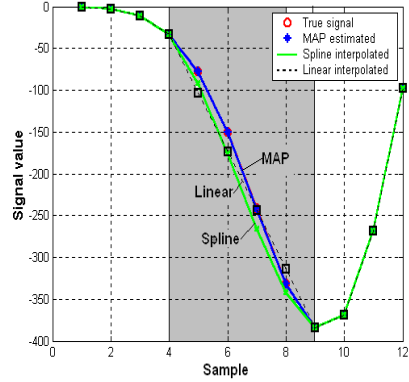


Figure B.9: MAP interpolation compared to linear and spline interpolation

$$\hat{\mathbf{x}}_{Un}^{MAP} = -(\mathbf{U}^T \Sigma_{xx}^{-1} \mathbf{U})^{-1} \Sigma_{xx}^{-1} \mathbf{K} \mathbf{x}_{Kn} \quad (\text{B.13})$$

Figures B.8 and B.9 show two examples of MAP-interpolation of a Gaussian signal with a missing block of samples.

B.4 The least square error autoregressive interpolation

The basic principle here is to model the signal as an autoregressive process and identify the signal parameters and then calculate the missing samples by using the identified *AR*-process. In practice the problem is solved in two steps:

1. Assume some values for the missing samples and then determine the signal parameters.
2. Calculate the values for the missing samples by using the *AR*-expression.

Possibly go back to the first step for a refinement.

$$\begin{aligned}
 & \begin{bmatrix} e(P) \\ e(P+1) \\ \vdots \\ e(k-1) \end{bmatrix} = \begin{bmatrix} x(P) \\ x(P+1) \\ \vdots \\ x(k-1) \end{bmatrix} \begin{bmatrix} x(P-1) & x(P-2) & \dots & x(0) \\ x(P) & x(P-1) & \dots & x(1) \\ \vdots & \vdots & \ddots & \vdots \\ x(k-2) & x(k-3) & \dots & x(k-P-1) \end{bmatrix} \\
 & \begin{bmatrix} e(k) \\ e(k+1) \\ e(k+2) \\ \vdots \\ e(k+M+P-2) \end{bmatrix} = \begin{bmatrix} x_{u_n}(k) \\ x_{u_n}(k+1) \\ x_{u_n}(k+2) \\ \vdots \\ x(k+M+P-2) \end{bmatrix} \begin{bmatrix} x(k-1) & x(k-2) & \dots & x(k-P) \\ x_{u_n}(k) & x(k-1) & \dots & x(k-P+1) \\ x_{u_n}(k+1) & x_{u_n}(k) & \dots & x(k-P+2) \\ \vdots & \vdots & \ddots & \vdots \\ x(k+M+P-3) & x(k+M+P-4) & \dots & x_{u_n}(k+M-2) \end{bmatrix} \\
 & \begin{bmatrix} e(k+M+P-1) \\ e(k+M+P) \\ e(k+M+P+1) \\ \vdots \\ e(N-1) \end{bmatrix} = \begin{bmatrix} x(k+M+P-1) \\ x(k+M+P) \\ x(k+M+P+1) \\ \vdots \\ x(N-1) \end{bmatrix} \begin{bmatrix} x(k+M+P-2) & x(k+M+P-3) & \dots & x_{u_n}(k+M-1) \\ x(k+M+P-1) & x(k+M+P-2) & \dots & x(k+M) \\ x(k+M+P) & x(k+M+P-1) & \dots & x(k+M+1) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-2) & x(N-3) & \dots & x(N-P-1) \end{bmatrix} \\
 & \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_P \end{bmatrix}
 \end{aligned}
 \tag{B.14}$$

Figure B.10: The system of equations for the *AR*-function.

We assume that the signal can be modeled as an autoregressive process, i.e.,

$$x(m) = \sum_{k=1}^P a_k x(m-k) + e(m) \quad (\text{B.15})$$

We rewrite the system of equations as (see Figure B.10)

$$\mathbf{e}(\mathbf{x}_{Um}, \mathbf{a}) = \mathbf{x} - \mathbf{X}\mathbf{a} \quad (\text{B.16})$$

Now we form the least squares criterion

$$\mathbf{e}^T \mathbf{e} = \mathbf{x}^T \mathbf{x} + \mathbf{a}^T \mathbf{X}^T \mathbf{X} \mathbf{a} - 2\mathbf{a}^T \mathbf{X}^T \mathbf{x}$$

which we minimize by the mentioned two stage algorithm above. As an example of how to proceed in the second step consider the system.

$$x_4 = a_1 x_3 + a_2 x_2 + a_3 x_1 + e_4 \quad (\text{B.17})$$

$$\begin{aligned} x_5 &= a_1 x_4 + a_2 x_3 + a_3 x_2 + e_5 \\ x_6 &= a_1 x_5 + a_2 x_4 + a_3 x_3 + e_6 \\ x_7 &= a_1 x_6 + a_2 x_5 + a_3 x_4 + e_7 \\ x_8 &= a_1 x_7 + a_2 x_6 + a_3 x_5 + e_8 \end{aligned}$$

$$x_9 = a_1 x_8 + a_2 x_7 + a_3 x_6 + e_9$$

$$x_{10} = a_1 x_9 + a_2 x_8 + a_3 x_7 + e_{10}$$

$$x_{11} = a_1 x_{10} + a_2 x_9 + a_3 x_8 + e_{11}$$

where x_5 to x_8 are our unknowns. We can rewrite the system as

$$\begin{bmatrix} e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \\ e_{10} \\ e_{11} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ -a_1 & 1 & 0 & 0 \\ -a_2 & -a_1 & 1 & 0 \\ -a_3 & -a_2 & -a_1 & 0 \\ 0 & -a_3 & -a_2 & -a_1 \\ 0 & 0 & -a_3 & -a_2 \\ 0 & 0 & 0 & -a_3 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix}}_{\text{Unknown}} + \underbrace{\begin{bmatrix} -a_3 & -a_2 & 0 & 0 \\ 0 & -a_3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -a_1 & 0 \\ 0 & 0 & -a_2 & -a_1 \end{bmatrix}}_{\mathbf{B}} \underbrace{\begin{bmatrix} x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix}}_{\text{Known}} \quad (\text{B.18})$$

This can be written as

$$\mathbf{e} = \mathbf{A}\mathbf{x}_{Un} + \mathbf{B}\mathbf{x}_{Kn} \quad (\text{B.19})$$

and

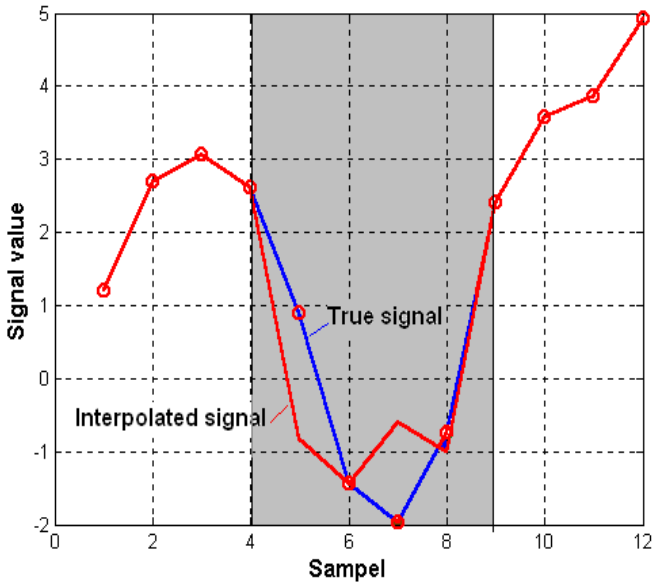


Figure B.11: An $AR(4)$ signal and the interpolated signal with the least squares error autoregressive interpolation.

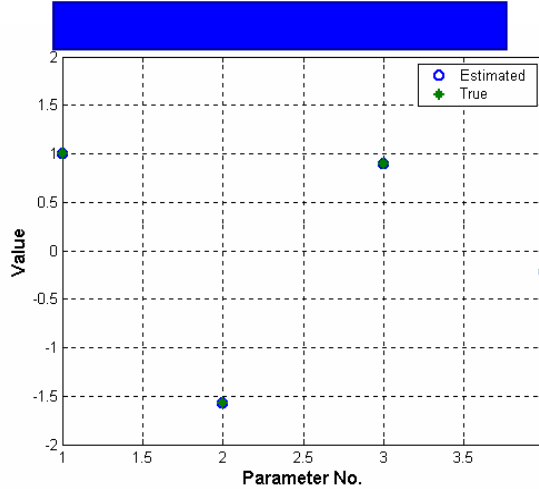


Figure B.12: The true and the estimated parameters of the $AR(4)$ signal.

$$\begin{aligned} \mathbf{e}^T \mathbf{e} &= (\mathbf{A}\mathbf{x}_{Un} + \mathbf{B}\mathbf{x}_{Kn})^T (\mathbf{A}\mathbf{x}_{Un} + \mathbf{B}\mathbf{x}_{Kn}) = \\ &= \mathbf{x}_{Un}^T \mathbf{A}^T \mathbf{A} \mathbf{x}_{Un} + 2\mathbf{x}_{Kn}^T \mathbf{B}^T \mathbf{A} \mathbf{x}_{Un} + \mathbf{x}_{Kn}^T \mathbf{B}^T \mathbf{B} \mathbf{x}_{Kn} \end{aligned} \quad (\text{B.20})$$

Taking the derivative gives

$$\frac{\partial \mathbf{e}^T \mathbf{e}}{\partial \mathbf{x}_{Un}} = 2\mathbf{A}^T \mathbf{A} \mathbf{x}_{Un} + 2\mathbf{A}^T \mathbf{B} \mathbf{x}_{Kn} \quad (\text{B.21})$$

and setting it to zero gives

$$\hat{\mathbf{x}}_{Un} = -[\mathbf{A}^T \mathbf{A}]^{-1} \mathbf{A}^T \mathbf{B} \mathbf{x}_{Kn} \quad (\text{B.22})$$

That is, we have minimized the prediction error [KSH00, KA93].

As an example an $AR(4)$ process was generated and a block of samples was considered unknown, Figure B.11. In the first step the signal parameters were estimated disregarding the unknown samples [LJ91, LJ99, SÖST89], see Figure

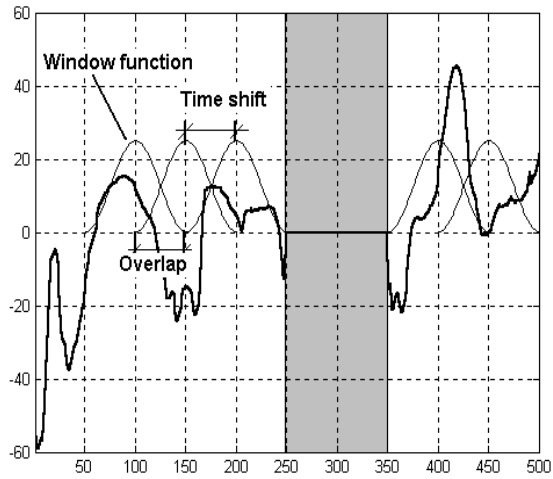


Figure B.13: A bottom profile signal with a missing block of samples. The window function is faintly outlined.

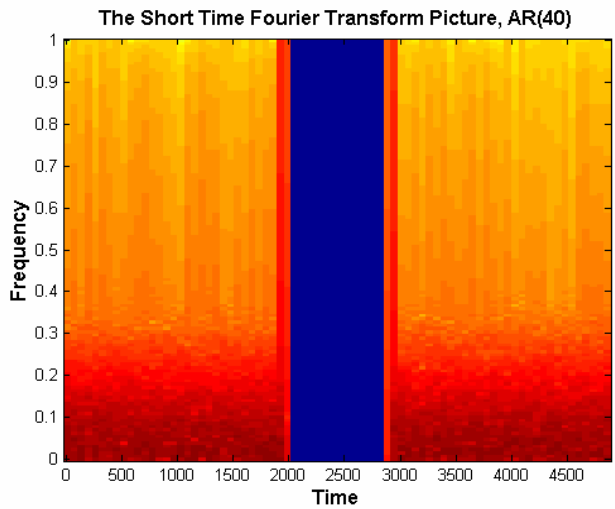


Figure B.14: The output of the Spectgram function in Matlab Signal Processing Toolbox.

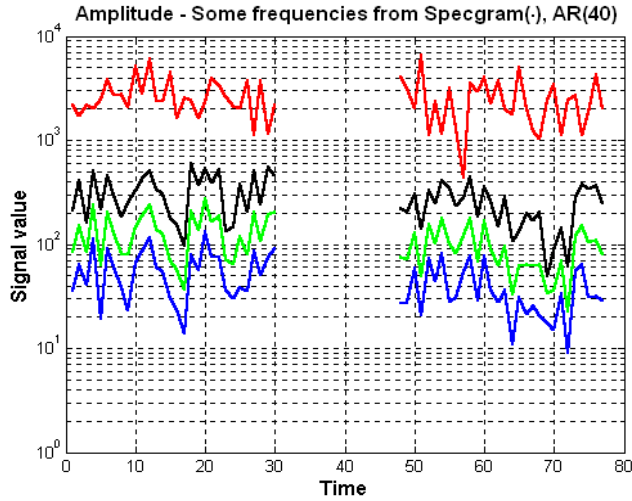


Figure B.15: Some of the frequency lines constituting the specgram in Figure B.14.

B.12. In the second step the estimated parameters were used to calculate the unknown samples.

B.5 Interpolation in the frequency domain

A third method for interpolation is to interpolate in the frequency domain. As before, the method will be illustrated in the case of a missing block of samples, Figure B.13.

The short time Fourier transform decomposes the signal in time and frequency. It is

$$X(k, m) = \sum_{i=1}^{N-1} w(i)x(m(N-D)+i)e^{-j\frac{2\pi ik}{N}} \quad (\text{B.23})$$

where N is block length, D = block overlap, k =frequency, m = block number, $w(i)$ =window. The transform can be calculated by Matlab's Specgram function

[KSL93], see Figure B.14. The different frequency signals shown in Figure B.15 can be interpolated by any of the previously mentioned methods and the missing samples in the time domain are recovered by the inverse transform.

B.6 Summary

This introductory study of the probabilistic interpolation methods shows, in some cases, excellent performance of the methods compared to the deterministic methods, see Figure B.8. The probabilistic methods would be the given choice if the signal characteristics of the bottom profile would be known at the interpolation point. This is, however, seldom the case. Instead the characteristics have to be estimated from a few bottom depth samples which will degrade the performance considerably. Therefore, in many cases, we will have to resort to the deterministic methods which we by experience know will have sufficient performance.



Notational Conventions

Symbols

Throughout the thesis uppercase boldface letters denote matrices, lowercase bold letters denote column vectors and italics denote scalars.

Notational distinctions between the symbol for a stochastic variable and its outcome is only done if the meaning is not clear from the context. If a notational distinction is necessary Greek letters and uppercase letters are used to denote the stochastic variable.

$\sim f$ Notation for distributed as f .

$\stackrel{a}{\sim} f$ Notation for asymptotically distributed as f .

$\hat{\mathbf{x}}$ The $\hat{}$ notation means estimated value or estimator depending on context.

$\mathcal{N}(\mu, \sigma^2)$ The Gaussian probability density function with the mean μ and variance σ^2 . The standard deviation $\sigma > 0$.

$\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ The vector Gaussian probability density function with the mean $\boldsymbol{\mu}$ and covariance $\mathbf{C} > 0$, i.e., positive definite.

$\arg \max_{\boldsymbol{\theta}} f(\boldsymbol{\theta})$	Denotes the value of $\boldsymbol{\theta}$ that maximizes $f(\boldsymbol{\theta})$.
$\mathbf{x}_t, \mathbf{A}_t$	Short for $\mathbf{x}(t)$ or $\mathbf{A}(t)$, ie., the vector \mathbf{x} or matrix \mathbf{A} is a function of time.
\mathbb{Y}_t	Alternative notation for matrices.
$\mathbf{1}_{N,N}$	Short for a matrix of dimension $N \times N$ whose elements are ones.
$\mathbf{I}_{N,N}$	The identity matrix of dimension $N \times N$ which has ones in the diagonal and zeros elsewhere.
$\boldsymbol{\Pi}$	The projection matrix $\boldsymbol{\Pi} = \mathbf{I}_N - \mathbf{1}_N / N$. The projection matrix is idempotent, ie., $\boldsymbol{\Pi} \cdot \boldsymbol{\Pi} = \boldsymbol{\Pi}$.
$\mathbf{x}^T, \mathbf{A}^T$	Superindex T denotes transpose.
\mathbf{A}^{-1}	The inverse of the square matrix \mathbf{A} .
$\det(\mathbf{A})$	The determinant of the square matrix \mathbf{A} .
$\text{Tr}(\mathbf{A})$	The trace of the square matrix \mathbf{A} .
$h(\mathbf{x})$	A scalar valued function of the vector \mathbf{x} .
$\mathbf{h}(\mathbf{x})$	A vector valued function of the vector \mathbf{x} .
$p(\mathbf{x})$	The probability density function.
$f(\mathbf{x})$	Alternative notation for the probability density function.
$p(\mathbf{x})$	The prior probability density function before measurement update.
$p(\mathbf{x} \mathbf{y})$	The conditional density function given the event \mathbf{y} .
$L(\mathbf{x}; \mathbf{y})$	The likelihood function as a function of \mathbf{x} given the event \mathbf{y} .
$\Pr\{\}$	Probability
$E\{\}$	Expectation.

$$\frac{\partial f}{\partial \mathbf{x}}$$

A column vector whose elements are the partial derivatives of the scalar function $f(\mathbf{x})$ with respect to the elements in the vector \mathbf{x} . The vector is also called the gradient vector of $f(\mathbf{x})$.

$$\frac{\partial f}{\partial \mathbf{x}^T}$$

A row vector whose elements are the derivatives of the scalar function f with respect to the elements in the vector

$$\mathbf{x}, \text{ i.e. } \frac{\partial f}{\partial \mathbf{x}^T} = \left[\frac{\partial f}{\partial \mathbf{x}} \right]^T.$$

$$\nabla_{\mathbf{x}} f = \frac{\partial f}{\partial \mathbf{x}}$$

A notation of the gradient often used in physics and considered to be a row vector.

$$\Delta_{\mathbf{x}} f$$

The Laplacian, a scalar function which is the sum of the second partial derivatives of the elements in the

vector \mathbf{x} of dimension m , $\sum_{i=1}^m \frac{\partial^2 f}{\partial x_i^2}$.

$$\Delta_{\mathbf{x}} f = \nabla_{\mathbf{x}} \nabla_{\mathbf{x}} f = \nabla_{\mathbf{x}}^2 f.$$

$$\frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{x}^T}$$

The Hessian matrix which has the elements $\frac{\partial^2 f}{\partial x_i \partial x_j}$ where i is the row index and j the column index.

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}^T}$$

The Jacobian matrix which has the elements $\frac{\partial y_i}{\partial x_j}$

where i is the row index and j the column index. The

matrix $\frac{\partial \mathbf{y}^T}{\partial \mathbf{x}} = \left[\frac{\partial \mathbf{y}}{\partial \mathbf{x}^T} \right]^T$ is called the vector gradient

matrix and has the elements $\frac{\partial y_j}{\partial x_i}$.

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$$

The gradient matrix. The gradient matrix is the transpose

of the Jacobian matrix. Also denoted $\frac{\partial \mathbf{y}^T}{\partial \mathbf{x}}$ in the literature.

$$\left| \frac{\partial \mathbf{y}}{\partial \mathbf{x}^T} \right|$$

The Jacobian or Jacobian determinant. $\left| \frac{\partial \mathbf{y}}{\partial \mathbf{x}^T} \right| = \left| \frac{\partial \mathbf{y}^T}{\partial \mathbf{x}} \right|$.

The Jacobian is the determinant of the Jacobian matrix, i.e., $\det\left(\frac{\partial \mathbf{y}}{\partial \mathbf{x}^T}\right)$.

$$\frac{\partial \mathbf{A}}{\partial x_k}$$

A matrix with the elements $\frac{\partial a_{ij}}{\partial x_k}$ where a_{ij} is an element of the matrix \mathbf{A} and i is the row index and j is the column index.

$$\frac{\partial \mathbf{y}(\mathbf{x})}{\partial x_i}$$

A column vector whos elements are the derivatives of the components of the vector \mathbf{y} with respect to

x_i , i.e., the j :th component is $\frac{\partial y_j(\mathbf{x})}{\partial x_i}$.

$$\frac{\partial \mathbf{C}(\mathbf{x})}{\partial x_i}$$

A matrix whose elements are the derivatives of the elements in the matrix \mathbf{C} , i.e.,

$$\left[\frac{\partial \mathbf{C}(\mathbf{x})}{\partial x_i} \right]_{r,s} = \frac{\partial [\mathbf{C}(\mathbf{x})]_{r,s}}{\partial x_i}.$$





List of acronyms

Below is a list of acronyms that appear in the thesis.

2D	Two dimensional
3D	Three dimensional
AR	Autoregressive
BITAN	Beijing University of Aeronautics and Astronautics inertial terrain-aided navigation
COV	Covariance
CRLB	Cramér-Rao lower bound
CW	Continuous waveform
EKF	Extended Kalman filter
FE	Finite element
FEM	Finite element method
FP	Fokker-Planck
FPE	Fokker-Planck partial differential equation
FIFO	First in, first out
FIM	Fisher information matrix
FOG	Fiber optic gyro
FPGA	Field programmable gate array
Gaussian	Normally distributed
GPS	Global positioning system
IMU	Inertial measurement unit

IN	Inertial navigation
INS	Inertial navigation system
LATAN	Low altitude terrain avoidance and navigation
MAD	Mean absolute difference
MAP	Maximum a posteriori
MC	Monte Carlo
ML	Maximum likelihood
MLE	Maximum likelihood estimation
MMSE	Minimum mean square error
MPF	Mass point filter
MSE	Mean square error
PDA	Probabilistic data association
PDAF	Probabilistic data association filter
PDE	Partial differential equation
PDF	Probability density function
PF	Particle filter
PVDF	Polyvinylidene fluoride
RLG	Ring laser gyro
RMS	Root mean square
SITAN	Sandia inertial terrain aided navigation
SMC	Sequential Monte Carlo filter
TERNAV	Terrain navigation
TERCOM	Terrain contour matching
TERPROM	Terrain profile matching
UKF	Unscented Kalman filter
UT	Unscented transform
VAR	Variance

Bibliography

- [AMGC02] S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, *A Tutorial on Particle Filters for On-line Non-linear/Non-Gaussian Bayesian Tracking*, IEEE Transactions on Signal Processing, Vol. 50, No. 2, February 2002.
- [AN00] C.-J. Andersson, “*Navigation System for an Autonomous Underwater Vehicle*”, Master thesis in Automatic Control, Royal Institute of Technology, Stockholm, Sweden 2000.
- [AN79] B. D. O. Anderson, J. B. Moore, *Optimal Filtering*, Prentice-Hall, Inc., USA, 1979.
- [ANIN03] C.-J. Andersson, I. Nygren, *A Method for Terrain Positioning of Underwater Vehicles*, Proc. Undersea Defence Technology Europe (UDT), 2003.
- [ANTR02] K. Andersson, F. Traugott, *Implementation of a terrain correlator for an underwater vehicle in an FPGA*, Master thesis, Department of Electronics. Mälardalens University, Västerås, Sweden, 2002.

- [AOA02] A. Stuart, J. K. Ord, S Arnold, *Classical Inference & the Linear Model*, Kendall's Advanced Theory of Statistics, Arnold, London, UK, 2002.
- [AR88] A. J. Robins, *Recent developments in the "TERPROM" Integrated Navigation System*, Proceedings of the 44th Annual Meeting, Annapolis, Maryland. June 1988.
- [AS01] B. Azimi-Sadjadi, *Approximate Nonlinear Filtering with Applications to Navigation*, Ph.D. Thesis, Institute for System Research, University of Maryland, USA, 2001.
- [AS72] D.L. Alspach and H.W. Sorenson. *Nonlinear Bayesian estimation using Gaussian sum approximation*, *IEEE Transactions on Automatic Control*, 17:439-448, 1972.
- [AZ96] A. Azzalini, *Statistical Inference, Based on the likelihood*, Chapman & Hall, Monographs on Statistics and Applied Probability, 1996.
- [BA85] C. A. Baird, *Designs Techniques for Improved Map-Aided Navigation*, Proceedings of the IEEE National Aerospace and Electronic Conference, Naecon, 1985.
- [BAAB84] C. A. Baird, M. R. Abramson, *A Comparison of Several Digital Map-Aided Navigation Techniques*, Proceedings of the 1984 IEEE Position, Location And Navigation Symposium (PLANS'84), San Diego, CA, 1984.
- [BE01] N. Bergman, *Posterior Cramér-Rao bounds for sequential estimation*, in *Sequential Monte Carlo Methods in Practice*, A. Doucet. H. deFreitas, N. Gordon, Eds. New York Springer-Verlag, 2001.
- [BE98] N. Bergman, *Deterministic and Stochastic Bayesian Methods in Terrain Navigation*, In Proc. 37th IEEE Conference on Decision and Control 1998.
- [BE99] N. Bergman, *Recursive Bayesian Estimation: Navigation*

and Tracking Applications, Dissertation 579, Linköpings Univ., Linköping, Sweden, 1999.

- [BER93] O. Bergem, *Bathymetric Navigation of Autonomous Underwater Vehicles Using a Multibeam Sonar and a Kalman Filter with Relative Measurement Covariance Matrices*, Dr. Scient thesis, University of Trondheim, Norway, 1993.
- [BI99] D. J. Biezad, *Integrated Navigation and Guidance Systems*, AIAA Education Series, USA, 1999.
- [BJ74] Å. Björck, *Numerical Methods*, Prentice-Hall Series in Automatic Computation, Prentice-Hall Inc. 1974.
- [BMB00] R. Beckman, A. Martinez, B. Bourgeois, *AUV positioning using bathymetry matching*, Conference proceedings, MTS/IEEE Oceans 2000, 11-14 Sept. 2000.
- [BP71] J. S. Bendat, A. G. Piersol, *Random Data: Analysis and Measurement Procedures*, Wiley-Interscience, John Wiley & Sons, New York, 1971.
- [BR86] R. N. Bracewell, *The Fourier Transform and Its Applications*. Second edition, revised, MacGraw-Hill, 1986, Ltd., 2002.
- [BR96] D. Brigo, *Filtering by Projection on the Manifold of Exponential Densities*, Dr thesis, Vrije University, Amsterdam, Holland, 1996.
- [BRDA91] P. J. Brockwell, R. A. Davis, *Time Series: Theory and Methods*, 2nd Ed. Springer-Verlag, New York, 1991.
- [BRHW97] R. G. Brown, P. Y. C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, John Wiley & Sons 1997.

- [BRK01] Y. Bar-Shalom, X. Rong Li, T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*, Wiley Interscience Publication, USA, 2001.
- [BS02] J. M. Bernardo, A. F. M. Smith, *Bayesian Theory*, Wiley Series in Probability and Statistics, Wiley Series in Probability and Statistics, John Wiley & Sons, New York, 2000.
- [BSFO88] Y. Bar-Shalom, T. E. Fortman, *Tracking and Data Association*, Mathematics in Science and Engineering, VOL. 179, Academic Press, Inc., 1988.
- [BU91] W. S. Burdic, *Underwater Acoustic System Analysis*, 2nd Ed. Prentice-Hall, 1991.
- [BUSE71] R.S. Bucy and K.D. Senne, *Digital synthesis of nonlinear filters. Automatica*, 7:287-298, 1971.
- [CA88] P. E. Caines, *Linear Stochastic Systems*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York, 1988.
- [CANY05] J. Carlström, I. Nygren, *Terrain Navigation of the Swedish AUV62F*, Proceedings, 14th International Symposium UUST05, Durham, New Hampshire, USA, 2005.
- [CKN92] I. Cato, B. Kjellin, K. Nordberg, *Havets bottnar och sediment, (The bottoms and the sediments of the sea)*, Sveriges Nationalatlas, Hav och kust. Temareaktör Björn Sjöberg, SNA Förlag, Kulturgeografiska institutionen, Stockholms universitet, Stockholm, 1992.
- [CO04] Comsol AB, *Partial Differential Equation Toolbox for Use with Matlab*, The Math Works Inc. USA, 2004.
- [CR46] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, USA, 1946.

-
- [CT92] C. H. Therrien, *Discrete Random Signals and Statistical Signal Processing*, Prentice Hall, Signal Processing Series, Englewood Cliffs, New Jersey, USA.
- [CYY92] Z. Chen, P. J. Yu, H. Yang, *BUAA inertial terrain-aided navigation (BITAN) algorithm*, Proceedings of the 18th ICAS Congress, Beijing, China, Sept. 20-25, 1992.
- [DABJ74] G. Dahlquist, Å. Björck, *Numerical Methods*, Prentice Hall, Inc., Englewood Cliffs, N.J. USA, 1974.
- [DFG01] A. Doucet, H. deFreitas, N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, New York Springer-Verlag, 2001.
- [DOBS90] D. Lerro, Y. Bar-Shalom, *Automated Tracking with Target Amplitude Information*, Proc. of the 1990 American Control Conference, San Diego, CA. USA, 1990.
- [EN01] P. Engman, *Felfortplantning i ekolodsystem, (Error Propagation in Bathymetric Systems)*, Thesis FHS HNU 2001 (in Swedish), Försvarshögskolan, Stockholm, Sweden, 2001.
- [FD05] T. S. Ferguson, *A Course in Large Sample Theory*, Chapman & Hall, 1996.
- [FE96] F. Daum, *Nonlinear Filters: Beyond the Kalman Filter*, IEEE Aerospace and Electronic Systems Magazine, August 2005, Volume 20, No. 8. Part 2.
- [FG00] F. Gustafsson, *Adaptive Filtering and Change Detection*, John Wiley & Sons, Ltd. UK, 2000.
- [FJ03] H. Fjaereide, *Navigation Aided by Terrain Correlation for Submarines*, Proc. Undersea Defence Technology Europe (UDT), 2003.

- [FK88] F. Knabe, *Untersuchung zum TERCOM Navigationsverfahren anhand von Flugversuchen über den Testgebieten Deister und Harz*, Deutschen Forschungs- und Versuchsanstalt für Luft- und Raumfahrt. 1988.
- [FO97] J. R. Fountain, *Digital Terrain System*, IEE Airborne Navigation System Workshop, UK, 1997.
- [GA04] C. W. Gardiner, *Handbook of Stochastic Methods*, 3rd Edition, Springer-Verlag Berlin, Heidelberg, New York, 2004.
- [GDW00] S. Godsill, A. Doucet, M. West, *Maximum a posteriori sequence estimation using Monte Carlo particle filters*, Ann. Inst. Stat. Math., Cambridge, UK, Vol. 52, No.1, 2001.
- [GGBFJ
KN02] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, P.-J. Nordlund, *Particle Filters for Positioning, Navigation, and Tracking*, IEEE Transactions on Signal Processing, Vol. 50, No. 2, February 2002.
- [GO80] J. Golden, *Terrain contour matching (TERCOM): A cruise missile guidance aid*, SPIE Vol. 238 Image Processing for Missile Guidance, 1980.
- [GOOR92] G. H. Golub, J. M. Ortega, *Scientific Computing and Differential Equations*, Academic Press Inc., 1992.
- [GRAN01] M. S. Grewal, A. P. Andrews, *Kalman Filtering: Theory and Practice*, Wiley-Interscience Publication, USA, 2001.
- [GWA01] M. S. Grewal, L. R. Weill, A. P. Andrews, *Global Positioning Systems, Inertial Navigation and Integration*, John Wiley & Sons, Inc., 2001.

-
- [HA01] S. Haykin, *Kalman Filtering and Neural Networks*, John Wiley & Sons, Inc. New York, USA, 2001.
- [HAFO04] A. O'Hagan, J. Forster, *Bayesian Inference*, Kendall's Advanced Theory of Statistics, Arnold, London, UK, 2004.
- [HL01] Edited by D. L. Hall, J. Llinas, *Handbook of Multisensor Data Fusion*, CRC Press, 2001.
- [HN28] H. Nyquist, *Certain Topics in Telegraph Transmission Theory*, AIEE Transactions, 1928, p.617, USA.
- [HO90] J. A. Hollowell, *Heli/SITAN: A Terrain Referenced Navigation Algorithm for Helicopters*, Proc. of the IEEE 1990 PLANS - Position, Location, And Navigation Symposium, 1990, Las Vegas, NV, USA.
- [HO91] J. Hollowell, *Terrain Referenced Navigation for Cruise Missiles*, The 1991 Cruise Missiles Technology Symposium (CMTS-91), Laurel, MD, USA, 1991.
- [HOAN85] L. D. Hostetler, R. D. Andreas, *Nonlinear Kalman Filtering Techniques for Terrain-Aided Navigation*, in H.W. Sorensen, editor, *Kalman filtering: Theory and Application*, IEEE Press 1985.
- [JA70] A. Jazwinski. *Stochastic Process and Filtering Theory*, volume 64 of *Mathematics in Science and Engineering*, Academic Press, New York, 1970.
- [JA89] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall International Editions, 1989.
- [JE01] C. Jekeli, *Inertial Navigation Systems with Geodetic Applications*, Walter de Gruyter Gmb H & Co. Berlin, Germany, 2001.

- [JMHP04] B. Jalving, M. Mandt, O. K. Hagen, F. Pöhner, *Terrain Referenced Navigation of AUVs and Submarines Using Multibeam Echo Sounders*, Proc. Undersea Defence Technology Europe (UDT), 2004.
- [JO87] C. Johnson, *Numerical solution of partial differential equations by the finite element method*, Studentlitteratur, Lund, Sweden, 1987.
- [JTH90] N. Johnson, W. Tang, G. Howell, *Terrain Aided Navigation Using Maximum A Posteriori Estimation*, Proceedings of the 1990 IEEE Position, Location And Navigation Symposium (PLANS'90) Las Vegas, Nevada, USA, March 1990.
- [JU97] S. J. Julier, *Process Models for the Navigation of High-Speed Land Vehicles*, Ph.D. Thesis, Dept. of Engineering Science, University of Oxford, UK 1977.
- [JUUH04] S. J. Julier, J. K. Uhlmann, *Unscented Filtering and Nonlinear Estimation*, Proceedings of the IEEE, Vol. 92, NO. 3, March 2004.
- [KA05] R. Karlsson, *Particle Filtering for Positioning and Tracking Applications*, Dissertation 924, Linköpings Univ., Linköping, Sweden, 2005.
- [KA93] S. M. Kay, *Fundamentals of Statistical Signal Processing, Estimation Theory*, Prentice Hall, 1993.
- [KH49] A. I. Khinchin, *Mathematical Foundations of Statistical Mechanics*, Dover Publications, Inc., New York, 1949.
- [KICH91] D. Kincaid, W. Cheney, *Numerical Analysis*, Brooks/Cole Publishing Company, Pacific Grove, Cal., USA, 1991.
- [KL01] F. Klebaner, *Introduction to Stochastic Calculus with*

Applications, Imperial College Press, London, 2001.

- [KSH00] T. Kailath, A. H. Sayed, D. Hassibi, *Linear Estimation*, Prentice Hall Inc., New Jersey, USA, 2000.
- [KSL93] T. P. Krauss, L. Shure, J. N. Little, *Signal Processing Toolbox for use with Matlab*, The Math Works Inc. USA, 1993.
- [KU67] H. J. Kushner, *Nonlinear Filtering: The Exact Dynamical Equations Satisfied by the Conditional Mode*, IEEE Transactions on Automatic Control, VOL. AC-12, No. 3. June 1967.
- [LA88] H.-D. Larche, *LATAN, a Modern Terrain Following and Flight Control System*, Messerschmitt-Boelkow-Blohm (MBB) GMBH, Ottobrunn, Germany 1988.
- [LA92] Anthony Lawrence, *Modern Inertial Technology, Navigation, Guidance, and Control*, Second edition, Springer-Verlag, New York 1992.
- [LJ91] L. Ljung, *System identification Toolbox for Use with Matlab*, The Math Works Inc. USA, 1991.
- [LJ99] L. Ljung, *System Identification. Theory for the User*, Second Ed., Prentice Hall Inc., New Jersey, USA, 1999.
- [LÜ96] H. Lütkepohl, *Handbook of Matrices*, John Wiley & Sons Ltd., UK, 1996.
- [MA82a] P. S. Maybeck, *Stochastic models, estimation and control*, Vol. 1, Academic Press Inc., New York, USA, 1982.
- [MA82b] P. S. Maybeck, *Stochastic models, estimation and control*, Vol. 2, Academic Press Inc., New York, USA, 1982.

- [MA97] D.E. di Massa, *Terrain-relative navigation for autonomous underwater vehicles*, Ph.D. Thesis MIT/Woods Hole Oceanographic Institution, 1997, USA, 1997.
- [MAST97] D.E. di Massa, W. K. Stewart, Jr., *Terrain-relative navigation for autonomous underwater vehicles*, Conference proceedings, MTS/IEEE Oceans 1997, 6-9 Oct. 1997.
- [META83] G. L. Mealy, W. Tang, *Application of Multiple Model Estimation to a Recursive Terrain Height Correlation System*, IEEE Transactions on Automatic Control, VOL. AC-28, March 1983.
- [METR02] J. Metzger, G. F. Trommer, *Studies on Four Terrain Referenced Navigation Techniques*, Symposium Gyro Technology 2002, Stuttgart, Germany, 2002.
- [MDFW00] R. van der Merwe, A. Doucet, N. de Freitas, E. Wan, *The Unscented Particle Filter*, Technical Report CUED/F-INFENG/TR 380, Cambridge University Eng. Dept. 2000.
- [MI00] T. Mikosch, *Elementary Stochastic Calculus with Finance in View*, World Scientific Publishing Co. Pty. Ltd., Singapore, 2000.
- [MIMI93] G. Minkler and J. Minkler, *Theory and Application of Kalman Filtering*, Magellan Book Company, 1993.
- [MLG99] N. Bergman, L. Ljung, F. Gustafsson, *Terrain navigation using Bayesian statistics*, in Proc. IEEE Int. Control Magazine, Vol. 19, no. 3, pp. 33-40, 1999
- [MM01] M. Mandt, *Terrengreferert posisjonering for undervannsfarkoster*, FFI/Rapport-2001/05900, NO-2027 Kjeller, Norway, 2001.
- [MU99] M. Unser, *Splines, A Perfect Fit for Signal and Image Processing*, IEEE Signal Processing Magazine, November, 1999.

-
- [MUOT82] J. F. Muller, W. F. Othling, *The Cruise-Missile Technical Challenge*, Astronautics and Aeronautics, v20 n1 p24-30, 53 Jan 1982.
- [MWT02] J. Metzger, J. Wendel, G. F. Trommer, *Comparison of Different Terrain Referenced Navigation Techniques*, ION 58th Annual Meeting, Albuquerque, 2002.
- [NA69] N. E. Nahi, *Estimation Theory and Applications*, John Wiley & Sons, Inc. New York, 1969.
- [NE75] D. E. Newland, *Introduction to Random Vibration and Spectral Analysis*, Longman 1975.
- [NKB94] L. Norman, S. Kotz, N. Balakrishnan, *Continuous Univariate Distributions*, John Wiley & Sons, 1994.
- [NY02] I. Nygren, *Recursive Terrain Navigation, Application of the Correlator Method*, FOI-R—0764—SE, Sweden, August 2002.
- [NY99] I. Nygren, *Terrängnavigering för undervattensfarkoster. Resultat av fältförsök 1998* (Terrain Navigation of Underwater Vehicles. Results from Sea-Trial 1998), FOA—R—99-01298-313—SE, December 1999.
- [NY01a] I. Nygren, *A Method for Terrain Navigation of an AUV*, Conference proceedings, MTS/IEEE Oceans 2001.
- [NY01b] I. Nygren, *A Method for Terrain Positioning of an AUV*, Proceedings, 12th International Symposium UUST01, Durham, New Hampshire, USA, 2001.
- [NYJA03a] I. Nygren, M. Jansson, *Recursive Terrain Navigation with the Correlator Method for High Position Accuracy*, Proceedings, 13th International Symposium UUST03, Durham, New Hampshire, USA, 2003.
- [NYJA03b] I. Nygren, M. Jansson, *Robust Terrain Navigation with*

the Correlator Method for High Position Accuracy,
Conference proceedings, MTS/IEEE Oceans 2003.

- [NYJA04a] I. Nygren, M. Jansson, *Terrain Navigation Using the Correlator Method*, Conference proceedings, 2004 IEEE Position Location And Navigation Symposium, April 27-29 Monterey, Cal., USA.
- [NYJA04b] I. Nygren, M. Jansson, *Terrain Navigation for Underwater Vehicles Using the Correlation Method*, IEEE Journal of Oceanic Engineering, Vol. 29, No. 3. July 2004.
- [NYJA04c] I. Nygren, M. Jansson, *A Terrain Navigation Method for UAVs and AUVs Based on Correlation*, IEEE Transactions on Aerospace and Electronic Systems, to be published.
- [OJ72] A. V. Oppenheim, D. H. Johnson, *Discrete Representation of Signals*, Proc. IEEE, Vol. 60, No. 6. June 1972.
- [OW97] A. V. Oppenheim, A. S. Willsky, *Signals & Systems*, Prentice Hall, Signal Processing Series, 1997, USA.
- [PE04] M. Pikander, P. Eskelinen, *Differential GPS Dynamic Location Experiment at Sea*, IEEE Aerospace & Electronic Systems Magazine, March 2004, USA.
- [PE95] V. V. Petrov, *Limit Theorems of Probability Theory, Sequences of Independent Random Variables*, Oxford Studies in Probability 4, Claredon Press, Oxford, Uk, 1995.
- [PK92] P. Kraniuskas, *Transforms in Signals and Systems*, Addison-Wesley Publishers Ltd. Wokingham, England, 1992.
- [PP02] A. Papoulis, S. Unnikrishna Pillai, *Probability, Random Variables and Stochastic Processes*, MacGraw-Hill, New York, NY, USA, 2002.
- [PUJP00] P. Pucar, J. Palmqvist, *Saab NINS/NILS an automatic*

-
- landing system for Gripen*, Conf. Proceedings of the 2000 IEEE Position, Location And Navigation Symposium (PLANS'00), 1200.
- [RA45] C. R. Rao, *Information and the Accuracy Attainable in the Estimation of Statistical Parameters*, Bulletin of the Calcutta Mathematical Society, Vol 37, 1945.
- [RI84] H. Risken, *The Fokker Planck Equation*, Springer-Verlag Berlin, Heidelberg, New York, 1984.
- [RO01] C. P. Robert, *The Bayesian Choice*, 2nd edition, Springer Texts in Statistics, Springer Verlag, 2001.
- [RU85] A. R. Runnals, *A Bayesian Approach to Terrain Contour Navigation*, AGARD Unclassified paper No. 43, Guidance and Control Panel, 40th Symposium, May 1985.
- [RÅWE95] L. Råde, B. Westergren, *Mathematical Handbook for Science and Engineering*, Studentlitteratur, Sweden, 1995.
- [SC91] L. L. Scharf, *Statistical Signal Processing, Detection, Estimation, and Time Series Analysis*, Addison-Wesley Publishing Co. New York, USA, 1991
- [SCA99] O. Strauss, F. Comby, M. J. Aldon, "Multibeam sonar image matching for terrain-based underwater navigation", Conference proceedings, MTS/IEEE Oceans 1999, Riding the Crest into the 21st Century, 13-16 Sept. 1999.
- [SESI93] P. K. Sen, J. M. Singer, *Large Sample Methods in Statistics, An Introduction with Applications*, Chapman & Hall, UK, 1993.

- [SEV00] T. A. Severini, *Likelihood Methods in Statistics*, Oxford Statistical science series, 22. Oxford University press, UK, 2000.
- [SH90] S. D. Stearns, D. R. Hush, *Digital Signal Analysis*, Prentice Hall, 1990.
- [SIL96] D.S. Silvia, *Data Analyses, A Bayesian Tutorial*, Oxford Science Publications, Claredon Press, Oxford, UK, 1996.
- [SK85] E. Skarman, *Flight test with a terrain aided navigation system*, Proceedings of the 5th International Conference, Remotely piloted vehicles, Bristol, England, 1985,
- [SM03] G. D. Smith, *Numerical Solution of Partial Differential Equations, Finite Difference Methods*, Third Edition, Clarendon Press, Oxford, UK, 2003.
- [SO96] S. P. Stoica, B. Ottersten, *The evil of super-efficiency*, Signal Processing, Vol. 55, Issue 1 November 1996, Elsevier North-Holland Inc., Amsterdam, The Netherlands.
- [SOA99] M. Sistiaga, J. Opderbecke, M. J. Aldon, *Depth image matching for underwater vehicle navigation*, Conference proceedings, In international Conference on Image Analysis and Processing, 1999, 27-29 Sept. 1999.
- [SÖST89] T. Söderström, P. Stoica, *System Identification*, Prentice-Hall, 1989.
- [ST97] V. Stohny, *Symmetry Properties and Exact Solutions of the Fokker-Planck Equation*, Nonlinear Mathematical Physics, 1997, V.4. N 1-2, pp.132-136, Kiev Polytechnical Institute, Kiev, Ukraina.
- [STBO97] G. Strang, K. Borre, *Linear Algebra, Geodesy, and GPS*, Wellesley-Cambridge Press, USA, 1997.

-
- [SY98] H. Stark, Y. Yang, *Vector Space Projections*, John Wiley & Sons, Inc. USA, 1998.
- [TALL03] F. Traugott, K. Andersson, A. Löfgren, L. Lindh, *Successful Prototyping of a Real-Time Hardware Based Terrain Navigation Correlator Algorithm*, Conference Proceedings, DSD2003 Euromicro Symposium On Digital System Design, Architectures, Methods and Tools, Turkey, Antalya, september 2003.
- [TH97] V. Thomée, *Galerkin Finite Element Methods for Parabolic Problems*, Springer-Verlag, Berlin, Heidelberg, New York, 1997.
- [TIWE97] D. H. Titterton, J. L. Weston, *Strapdown inertial navigation technology*, IEE Radar, Sonar, Navigation and Avionic series 5, Peter Peregrinus Ltd. London UK, 1997
- [TS02] T. Söderström, *Discrete-time Stochastic Systems*, Springer-Verlag, London Ltd., UK, 2002.
- [UR83] R. J. Uric, *Principles of underwater sound*, 3rd Ed, McGraw-Hill Book Company, 1983.
- [VA96] S. V. Vaseghi, *Advanced Signal Processing and Digital Noise Reduction*, Wiley-Teubner, 1996.
- [VA98] A. W. van der Vart, *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 1998.
- [VT02] H. L. Van Trees, *Optimum Array Processing, Estimation and Modulation Theory*, Wiley-Interscience Publication, New York, USA, 2002.
- [VT68] H. L. Van Trees, *Detection, Estimation and Modulation Theory*. Wiley, New York, 1968.

- [WL90] E. Waltz and J. Llinas, *Multisensor data fusion*, Artech House, Boston, 1990.
- [YP01] Yudi Patwitan, *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, Oxford Science Publications, Clarendon Press. Oxford, UK, 2001.
- [YCH91] P. Yu, Z. Chen, J. Hung, *Performance evaluation of six terrain stochastic linearization techniques for TAN*. IEEE Aerospace and Electronics Conference 1991.