Master's thesis

2019

Karine Hagesæther Foss

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

Karine Hagesæther Foss

Spatio-temporal Gaussian processes and excursion sets for adaptive environmental sensing using underwater robotics

December 2019

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

NTNU

Norwegian University of
Science and Technology

Spatio-temporal Gaussian processes and excursion sets for adaptive environmental sensing using underwater robotics

Karine Hagesæther Foss

Applied Physics and Mathematics
Submission date: December 2019
Supervisor: Jo Eidsvik

Norwegian University of Science and Technology
Department of Mathematical Sciences

Abstract

Underwater robotics makes it possible to explore the ocean in new ways, and can now be used for environmental sensing. In this thesis, the ideas of underwater sensing are applied to concentrations of pollutants from a factory in Frænfjorden, Norway. Statistical models combined with data from the fjord may be used to monitor the contaminant distributions over space and time. The basis for the statistical modelling is spatio-temporal Gaussian processes developed from the advection-diffusion stochastic partial differential equation. The aim of the thesis is to use the statistical model when exploring different decision strategies for the movement of underwater vehicles. The model and strategies can be used on board the vehicles for efficient sampling of the fjord. The focus is on identifying excursion sets, and a myopic (greedy) decision criterion for minimising the probability of misclassification with respect to the excursion set is derived in closed form.

A simulation study of pollution concentration in Frænfjorden is performed, comparing different exploration strategies for autonomous underwater vehicles. When a vehicle is following a random surveying plan, the average misclassification rate with respect to the excursion set is $13.1\ \%$. This is reduced to $10.3\ \%$ when using the misclassification probability decision strategy. The vehicle then efficiently finds excursion set borders, but often does not explore enough of the domain to identify the whole set. Adding variance reduction to the decision criterion makes the vehicle explore more, but predetermining a whole path for the vehicle beforehand is seen to perform just as good for the Frænfjorden model. The misclassification rates are $9.9\ \%$ and $9.7\ \%$ using each of these two strategies, respectively.

Sammendrag

Utviklingen av undervannsrobotikk gjør det mulig å utforske havet på nye måter, og denne teknologien kan nå brukes i miljøovervåking. I denne oppgaven er undervannsovervåking anvendt på konsentrasjon av forurensing fra en fabrikk i Frænfjorden, Norge. Statistiske modeller kombinert med data fra fjorden kan brukes til å overvåke fordelingen av forurensingen i tid og rom. Gaussiske prosesser utviklet fra en stokastisk partiell differensialligning, adveksjon- diffusjonsligningen, brukes som en basis for modelleringen. Målet med arbeidet er å bruke den statistiske modellen i utforsking av forskjellige beslutningsstrategier for bevegelsen til undervannsfarkoster. Modellen og strategiene kan brukes om bord på farkostene for effektiv overvåking av fjorden. Fokuset er på overskridelsesmengder (excursion sets), og et myopisk (grådig) beslutningskriterium for å minimere sannsynligheten for feilklassifisering med hensyn til overskridelsesmengden er utledet med analytisk løsning.

Et simuleringsstudie for forurensingskonsentrasjon i Frænfjorden er gjennomført for å sammenligne forskjellige utforskningsstrategier for autonome undervannsfarkoster. Når en farkost velger posisjoner for målinger tilfeldig, blir i gjennomsnitt $13.1\ \%$ av området feilklassifisert. Dette reduseres til $10.3\ \%$ når strategien for å minimere sannsynligheten for feilklassifisering benyttes. Farkosten finner effektivt deler av overskridelsesmengden, men utforsker ofte for lite av domenet til å identifisere hele mengden. Ved å legge til variansreduksjon som en del av kriteriet, utforsker farkosten større områder, men å bestemme hele ruten til farkosten på forhånd viser seg å være like bra for modellen for Frænfjorden. Feilklassifiseringsratene er henholdsvis $9.9\ \%$ og $9.7\ \%$ for disse to strategiene.

Preface

This thesis is the final work of my Master of Science degree at NTNU, Norwegian University of Science and Technology. The degree is in Applied Physics and Mathematics with main profile Industrial Mathematics, specialising in Statistics.

The work has been interesting and exciting, and sometimes stressful and frustrating. I would like to thank my supervisor, Jo Eidsvik, for help and support throughout the process. There have been countless meetings and emails during the last two years, since we first talked about the thesis in January 2018. I am grateful for useful insights, ideas and conversations. I would also like to thank Gunhild Elisabeth Berget for being available throughout the thesis work, and for providing me with data from DREAM and SINMOD in the format that I needed.

Finally, I want to express my gratitude for getting the chance to study abroad during my time at NTNU. In addition to learning about science, mathematics and statistics, I had the opportunity to experience other cultures and views on science and society. I am grateful to graduate from NTNU more knowledgeable, and also (hopefully) wiser, than when I arrived in August 2014.

Table of contents

Chapter 1

Introduction

Recently, the environmental impact of human activities on oceans and coastal areas has been in focus. Oil spills are known to have big consequences for life at sea and along the coast, and macro and micro plastics have also proved to be extremely harmful. In Norway, disposal of mining waste into fjords is a never ending controversy. Environmental activists have protested against the planned seafills in Førdefjorden and Repparfjorden for some years, and the decisions have also been appealed. In late November 2019, however, the final appeal for Repparfjorden was dismissed, and the mining company Nussir ASA can start their work (Garvik, 2017; Verstad et al., 2019). When mine tailings are deposited into a fjord, it is important that the contamination is monitored, to make sure that the polluted area is not extended beyond what is permitted, and that the damage is minimal.

In this thesis, the focus is on ocean pollution in Frænfjorden. The factory of the mining company Omya Hustadmarmor AS is located in Elnesvågen in Frænfjorden, as shown in Figure 1.1, and waste is deposited into the fjord close to the factory. This is one out of five currently active seafills in Norway (Setså, 2019). Various tools can be used to monitor the movements of the pollutants, and the development of autonomous underwater vehicles (AUVs) opens for new techniques. In the past few years, AUVs have become less expensive and more robust, and can now be used in scientific field work (Fossum et al., 2018), as shown in Figure 1.2. AUVs may then also be used for environmental sensing.

This thesis considers spatio-temporal random processes for modelling the pollution dynamics in the inner part of Frænfjorden. These are processes explaining phenomena that evolve and have dependencies in both time and space. We focus on a *dynamical* spatio-temporal statistical model, where the process values are modelled based on the values in the past and by values in spatially near locations (Cressie and Wikle, 2011). According to Cressie and Wikle (2011), it may be convenient to think of the development in time for different spatial positions as multivariate time series, but spatio-temporal processes may also be seen as spatial fields moving in time.

Equations describing known physical phenomena are useful for developing models for spatio-temporal processes (Cressie and Wikle, 2011; Sigrist et al., 2015b). The process is often expressed through a partial differential equation (PDE) or a stochastic PDE (SPDE). The latter expresses partial derivatives in time and/or space in the same way as do PDEs, but uncertainty terms are also present. Then the resulting process is not deterministic, but a

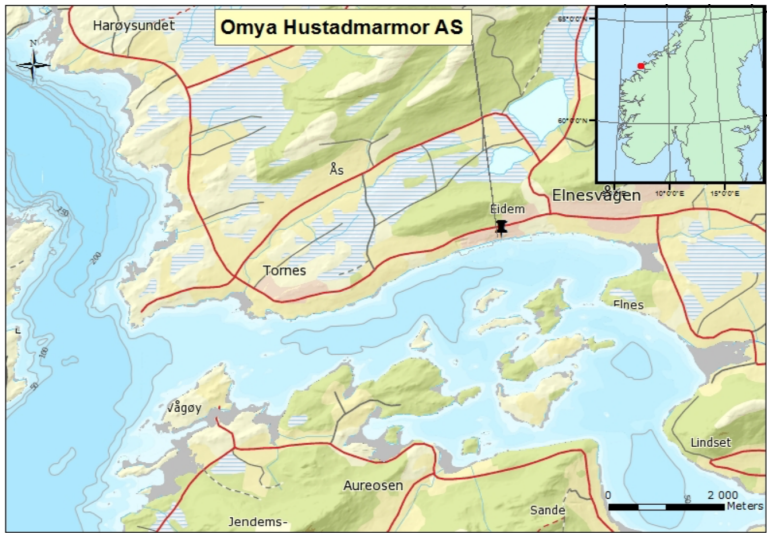


Figure 1.1: Position of Omya Hustadmarmor AS in Frænfjorden outside of Molde, on the west coast of Norway. Figure taken from Glette (2019).

Figure 1.2: An AUV in action. Courtesy of Gunhild Elisabeth Berget.

stochastic process. There is a one-to-one mapping between describing the process through a time-space covariance function and the use of SPDEs (Hartikainen et al., 2011). In this thesis, the advection-diffusion SPDE is the basis for the process model. The SPDE and a spatial covariance function are used for explaining the process.

The spatial field is modelled by probabilistic Gaussian random fields, also called Gaussian processes (GPs). The GPs are built up by jointly Gaussian distributed variables, and the class is closed under linear transformations; linear combinations of GPs are also GPs. The advection-diffusion SPDE is linear. Thus, if the stochastic noise term is Gaussian and the process is initialised by a GP, the process is determined by spatio-temporal GPs. Spatial and spatio-temporal GPs are both denoted GP. Spatio-temporal GPs are used in many applications, such as neuroimaging (Hyun et al., 2016), Machine Learning (Hartikainen et al., 2011; Rasmussen and Williams, 2006), geostatistics and environmental sciences such as precipitation analysis (Liu et al., 2019; Sigrist et al., 2015b).

An area where some quantity of interest is above a given threshold or critical limit is called an *excursion set* (ES). ESs may be of particular interest in environmental sensing. Bolin and Lindgren (2015) work on ESs in the application of air pollution exceeding the limits set by the European Union, which is one example where allowed levels of pollutants are law regulated. If levels are exceeded or if certain substances move into the wrong areas, the consequences may in many situations be large.

The aim of the thesis is to use spatio-temporal GPs to model the pollution in Frænfjorden, and explore different strategies for efficient ocean monitoring by AUVs. The AUV will then obtain data from positions chosen based on some decision criterion. Consider the ES of a system based on the GP model. A decision criterion for minimising the misclassification probability with respect to the ES is derived and used in a simulation study for pollution in Frænfjorden. In this way, the ideas of GPs, ESs and AUV technology are combined.

The thesis is organised as follows: In chapter 2, the underlying model for concentration is developed in a generic framework, while model parameters for the situation in Frænfjorden are found in chapter 3. The methods of ESs and decision strategies are presented and derived in chapter 4, and in chapter 5, the details of the simulation study for Frænfjorden are given. Finally, the results of the simulation study are presented and discussed in chapter 6 and chapter 7, respectively.

Chapter 2

Model specifications

In this chapter, theory of GPs in both a spatial and spatio-temporal setting is presented. First, a spatial model is obtained, and equations for data assimilation in the spatial setting are introduces. Further, a spatio-temporal process model is developed based on the advection-diffusion SPDE, and finally, handling new data in time is shown using the Kalman filter.

The random variable describing the quantity of interest in this thesis is X. It may depend on both time, t, and space, $\boldsymbol{s}$, and is then denoted $X(t, \boldsymbol{s})$. In the general case, both variables are continuous, but they are both discretised at some point in the thesis. The spatial component is two-dimensional, and given in the Cartesian coordinates *east* and *north*. That is, $\boldsymbol{s} = (s_e, s_n)^\intercal$. If the temporal component is fixed, it may be introduced as a subscript: $X_t(\boldsymbol{s})$, meaning that X_t is a spatial variable. This notation is often used in the discrete time case, to emphasise that the time variable may not vary arbitrarily, and is in that sense fixed. When the spatial component is fixed, $\boldsymbol{s} = \boldsymbol{s}_i$, regular notations, $X(t, \boldsymbol{s}_i)$ or $X_t(\boldsymbol{s}_i)$, are used. In a vector setting, it might sometimes also be convenient to refer to the vector itself rather than specific elements. Then brackets indicate the vector element, with $\boldsymbol{X}_t[i] = X_t(\boldsymbol{s}_i)$.

2.1 Spatial model

First consider the spatial setting for a fixed time t, so that $X(t, \boldsymbol{s}) = X_t(\boldsymbol{s})$.

2.1.1 Gaussian processes and initial state

Let $\mathcal{D} \subset \mathbb{R}^2$ be some finite 2-dimensional domain. The variable of interest is some continuous field, $\{X_t(\boldsymbol{s}); \boldsymbol{s} \in \mathcal{D} \subset \mathbb{R}^2\}$, where $X_t(\cdot)$ represents the concentration of pollutants. It could also represent other continuous spatial quantities, such as temperature or terrain height. The field is, as mentioned, modelled by a GP, meaning that $(X_t(\boldsymbol{s}_1), X_t(\boldsymbol{s}_2), \ldots, X_t(\boldsymbol{s}_k))$, for any configuration $(\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_k) \in \mathcal{D}$, $k = 1, 2, \ldots$ are jointly Gaussian,

$$(X_t(\boldsymbol{s}_1), X_t(\boldsymbol{s}_2), \ldots, X_t(\boldsymbol{s}_k)) \sim N_k(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t) \tag{2.1}$$

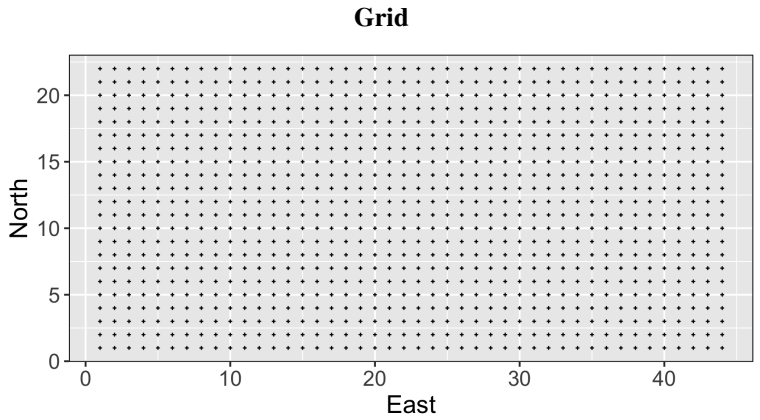


Figure 2.1: Regular spatial grid on the domain $\mathcal{D} = [1, 44] \times [1, 22] \subset \mathbb{R}^2$. The number of grid nodes is $N = 968$.

(Cressie and Wikle, 2011). Here, the notation $\sim$ means *follows the distribution of*, $N_k(\cdot)$ denotes a k-variate Gaussian distribution, $\boldsymbol{\mu}_t$ the k-dimensional mean vector and $\boldsymbol{\Sigma}_t$ the $k \times k$ covariance matrix of the distribution. The GP is discretised onto a regular grid covering $\mathcal{D}$, $\mathcal{D}_s$. The grid is given in Cartesian coordinates, with N_e grid points in the east direction and N_n in the east direction. Then $N = N_e \cdot N_n$ is the size of the grid, or, equivalently, the number of grid nodes. Let

$$\boldsymbol{X}_t = (X_t(\boldsymbol{s}_1), X_t(\boldsymbol{s}_2), \ldots, X_t(\boldsymbol{s}_N))^\intercal \tag{2.2}$$

for $\{\boldsymbol{s}_i \in \mathcal{D}_s \subset \mathbb{R}^2\}$ be the N-vector representing the discretised GP. Figure 2.1 illustrates the uniform grid of $N = 44 \cdot 22 = 968$ nodes on the domain $\mathcal{D} = [1, 44] \times [1, 22] \subset \mathbb{R}^2$.

In the spatio-temporal setting, the system model is updated as time passes and as new data gets available. To start the process, a model for the *initial state* is needed. The initial state is modelled by a spatial GP as described in this section, for time $t = 0$. We have

$$\boldsymbol{X}_0 = (X_0(\boldsymbol{s}_1), X_0(\boldsymbol{s}_2), \ldots, X_0(\boldsymbol{s}_N))^\intercal, \tag{2.3}$$

$$\boldsymbol{X}_0 \sim N_N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0). \tag{2.4}$$

The mean and covariance matrix for the initial state will be determined from available model data specific to the application, but some general features are discussed here. The initial state covariance matrix contains information about the level of noise in the system, i.e. variance for the value in each grid node. Further, the matrix takes into account the spatial correlation between the variables. It is common to assume isotropic correlation, where the correlation only depends on the Euclidean distance between variables on the grid. The distance between grid nodes $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ is denoted $||\boldsymbol{s}_i - \boldsymbol{s}_j||$. Let $\Delta\boldsymbol{s}$ denote the distance between two arbitrary points. The correlation is given by a positive definite correlation function, ensuring that the resulting matrix is positive definite. Denote the function $\rho(\Delta\boldsymbol{s})$ and the marginal variance (noise) σ_0^2. Then

$$\boldsymbol{\Sigma}_0[i, j] = \sigma_0^2 \rho(||\boldsymbol{s}_i - \boldsymbol{s}_j||), \tag{2.5}$$

where $[i, j]$ indicates row i and column j of the matrix.

A few common correlation functions are the *Exponential* and the *Matern* correlation functions (Cressie and Wikle, 2011). Further, the exponential correlation function can be combined with a cosine to get some periodic effect. The function can then be called *Damped Cosine*. See Table 2.1 for the details on the correlation functions considered here. The range parameter ϕ is the distance until values are approximately uncorrelated. When $\Delta \boldsymbol{s} = \phi$, we have correlations $\exp\{-3\} \approx 0.05$ and $(1+5) \cdot \exp\{-5\} \approx 0.04$.

Exponential	$\exp\{-3\Delta \boldsymbol{s}/\phi\}$
Matern	$(1 + 5\Delta \boldsymbol{s}/\phi)\exp\{-5\Delta \boldsymbol{s}/\phi\}$
Damped Cosine	$\exp\{-3\Delta \boldsymbol{s}/\phi\}\cos(3\Delta \boldsymbol{s}/\phi)$

Table 2.1: Some possible spatial correlation functions. The Matern correlation is shown for smoothness parameter 1.5 and ϕ is the range parameter controlling the distance until node values are uncorrelated.

To visualise how the spatial correlation changes as a function of the distance between points in the domain, variograms, $\gamma(\cdot)$, are often used. A variogram is defined as

$$\begin{aligned}\gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) &= \frac{1}{2}\mathrm{Var}(X_t(\boldsymbol{s}_i) - X_t(\boldsymbol{s}_j)) \\ &= \frac{1}{2}\left(\mathrm{Var}(X_t(\boldsymbol{s}_i)) + \mathrm{Var}(X_t(\boldsymbol{s}_j)) - 2\mathrm{Cov}(X_t(\boldsymbol{s}_i), X_t(\boldsymbol{s}_j))\right),\end{aligned} \tag{2.6}$$

(Cressie and Wikle, 2011). Inserting the variance and isotropic correlation function, we obtain

$$\gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) = \frac{1}{2}(2\sigma_0^2 - 2\sigma_0^2\rho(||\boldsymbol{s}_i - \boldsymbol{s}_j||) = \sigma_0^2(1 - \rho(||\boldsymbol{s}_i - \boldsymbol{s}_j||)). \tag{2.7}$$

Figure 2.2 shows variograms of the different correlation functions with parameters $\phi = 25$ and $\sigma_0^2 = 0.6$. All variograms stabilise at the horizontal line of σ_0^2 as the distance increases, meaning that values at positions far away from each other are uncorrelated. When the values are below this line, the correlation is positive. We see that the Damped Cosine function has a faster decrease in correlation, and that the correlation is negative (variogram above the horizontal line) for some distances before stabilising at 0 correlation. We further notice that the Matern correlation function has derivative 0 at $\Delta \boldsymbol{s} = 0$, so the correlation is larger for small distances than for the others.

For illustration, a test example is constructed. The grid is as in Figure 2.1, and the mean value of the initial state $\boldsymbol{X}_0$ is set as shown in Figure 2.3. Here, the concentration values decrease linearly with the Euclidean distance from grid point $(\boldsymbol{s}_e, \boldsymbol{s}_n)^\intercal = (11, 11)^\intercal$ until they reach 0. The maximum value is 12.

Each of the correlation functions described in Table 2.1 with parameters as in Figure 2.2 are used to construct covariance matrices, and we can sample from each corresponding distribution. Assume we can sample from a standard Gaussian distribution, $Z \sim N(0, 1)$, and let $\boldsymbol{z}$ be a vector containing N independent standard Gaussian samples. To sample from the multivariate distribution of $\boldsymbol{X}_0$, as given in equation (2.4), the Cholesky decomposition, $\boldsymbol{\Sigma}_0 = \boldsymbol{L}\boldsymbol{L}^\intercal$, is computed. The lower triangular matrix $\boldsymbol{L}$ is called the Cholesky

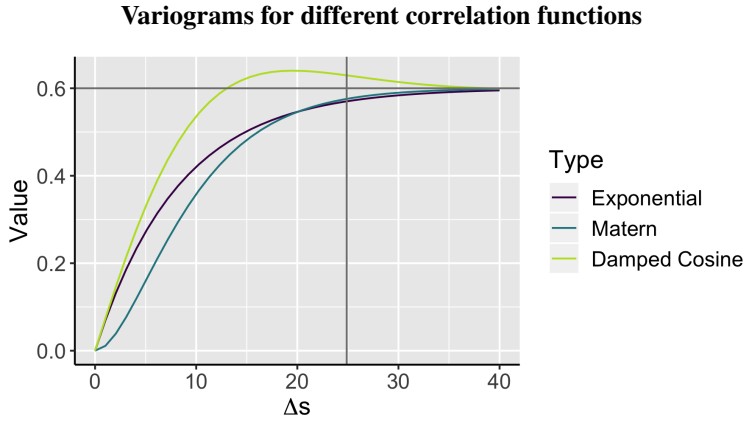


Figure 2.2: Variograms for different correlation functions with variance $\sigma_0^2 = 0.6$ and range $\phi = 25$. The variance is added as a horizontal line and the range as a vertical line in the display.

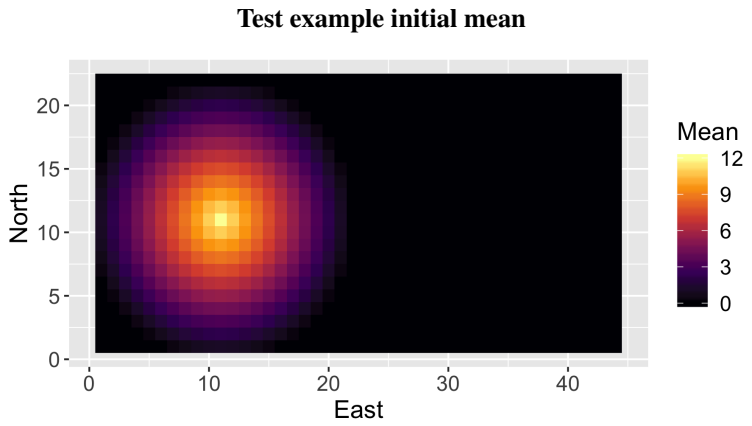
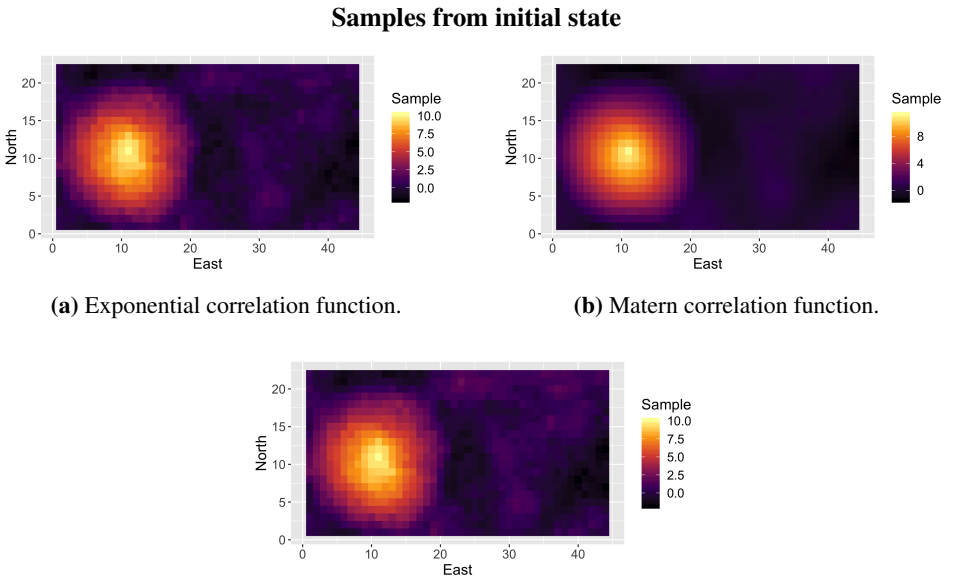


Figure 2.3: Initial mean in test example. The values decrease linearly from position $(\boldsymbol{s}_e, \boldsymbol{s}_n)^\intercal = (11, 11)^\intercal$.



(a) Exponential correlation function.

(b) Matern correlation function.

(c) Damped Cosine correlation function.

Figure 2.4: Samples from initial state of the test example. The samples are identical except use of correlation function in the covariance matrix.

factorisation of $\boldsymbol{\Sigma}_0$ and a sample of $\boldsymbol{X}_0$ is

$$\boldsymbol{x}_0 = \boldsymbol{\mu}_0 + \boldsymbol{L}\boldsymbol{z} \tag{2.8}$$

(Eidsvik et al., 2015). One sample from each distribution is shown in Figure 2.4. The only difference between these samples is the correlation function. The sample with the Matern correlation function in Figure 2.4b stands out by being smoother than the others, which is a result of the 0 derivative at distance 0. The effect of negative correlation in the Damped Cosine compared to the Exponential function is not easy to spot in these samples. In the following, the Matern correlation function is used.

2.1.2 Data assimilation

The GP is used in a Bayesian setting. Assume that a measurement from the system is available at time t. In a Bayesian and spatial setting, a model reflecting the knowledge of the system prior to obtaining this new data is called the *prior* distribution or model. The model is updated to include the new information, and this model is called the *posterior* model. In addition, the probability function of the new data (given the underlying system) is called the *likelihood*. We will see, in section 2.2.2, that in a spatio-temporal setting, the prior and posterior distributions for each time step are often called *predictive* and *filtering* distributions. In this setting, the prior and posterior are mostly referred to before and after the whole process.

Now we consider the case where the state is fixed at time t and the data is obtained at this time. We have $\boldsymbol{X}_t \sim N_N(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$. The new data is obtained from measuring one or several points on the grid, corresponding to vector elements $\boldsymbol{d} = (d_1, \ldots, d_m)^\intercal$, $1 \leq m \leq N$. That is, the value in $\boldsymbol{s}_{d_j} \in \mathcal{D}_s$ is measured for all $j \in 1, \ldots m$. In this thesis, $m = 1$ for the most part, meaning that only one position is measured at a time. The expected value of the measurement is the true system value in grid points $\boldsymbol{s}_{d_1}, \ldots, \boldsymbol{s}_{d_m}$. In addition, there is some measurement error, $\boldsymbol{\varepsilon}$. Let the random vector of measurements be denoted $\boldsymbol{Y}$, such that the measurement at time t is

$$\boldsymbol{Y}_t = \boldsymbol{H}_t \boldsymbol{X}_t + \boldsymbol{\varepsilon}_t, \tag{2.9}$$

where $\boldsymbol{H}_t$ is a $m \times N$ matrix specifying which grid point is measured at time step t. Each row has a 1 on the vector position of one of the measured values, and the rest are zeros. We call this the *design matrix*. Each measurement error ε_j is Gaussian distributed with mean zero and standard deviation τ, and they are independent. That is, $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \tau^2 \boldsymbol{I}_m)$, where $\boldsymbol{0}$ is a vector of only zeros, and $\boldsymbol{I}_m$ is the $m \times m$ identity matrix. Assume that $\boldsymbol{\varepsilon}$ and $\boldsymbol{X}$ are independent. Further, let $\boldsymbol{Y}_t \,|\, \boldsymbol{X}_t$ denote a random vector that follows the distribution of $\boldsymbol{Y}_t$ when $\boldsymbol{X}_t$ is given. Since linear combinations of Gaussian distributed random variables are Gaussian distributed themselves,

$$\boldsymbol{Y}_t \,|\, \boldsymbol{X}_t \sim N(\boldsymbol{H}_t \boldsymbol{X}_t, \tau^2 \boldsymbol{I}_m). \tag{2.10}$$

This is the likelihood in the Bayesian framework, and may also be called the *data model*. Of course, the real values in the system are unknown. By the formulas for double expectation and variance (Casella and Berger, 2002), $\boldsymbol{Y}_t$ is marginally Gaussian distributed with mean and covariance matrix

$$\begin{aligned} \mathrm{E}[\boldsymbol{Y}_t] &= \mathrm{E}[\mathrm{E}[\boldsymbol{Y}_t \,|\, \boldsymbol{X}_t]] \\ &= \mathrm{E}[\boldsymbol{H}_t \boldsymbol{X}_t] = \boldsymbol{H}_t \boldsymbol{\mu}_t, \end{aligned} \tag{2.11}$$

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{Y}_t) &= \mathrm{E}[\mathrm{Cov}(\boldsymbol{Y}_t \,|\, \boldsymbol{X}_t)] + \mathrm{Cov}(\mathrm{E}[\boldsymbol{Y}_t \,|\, \boldsymbol{X}_t]) \\ &= \mathrm{E}[\tau^2 \boldsymbol{I}_m] + \mathrm{Cov}(\boldsymbol{H}_t \boldsymbol{X}_t) \\ &= \tau^2 \boldsymbol{I}_m + \boldsymbol{H}_t \boldsymbol{\Sigma}_t \boldsymbol{H}_t^\intercal. \end{aligned} \tag{2.12}$$

We want to take the measured values into account and update the model for the whole system. That is, we are interested in finding the distribution of $\boldsymbol{X}_t \,|\, \boldsymbol{Y}_t$, i.e. that of $\boldsymbol{X}_t$ given the observed $\boldsymbol{Y}_t$, or the so-called posterior distribution in the Bayesian setting. If two random vectors have joint distribution

$$\begin{bmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{bmatrix} \sim N_{N_u + N_v} \left(\begin{bmatrix} \boldsymbol{\mu_u} \\ \boldsymbol{\mu_v} \end{bmatrix}, \begin{bmatrix} \boldsymbol{A} & \boldsymbol{C} \\ \boldsymbol{C}^\intercal & \boldsymbol{B} \end{bmatrix} \right), \tag{2.13}$$

with $\boldsymbol{A} = \mathrm{Cov}(\boldsymbol{U})$, $\boldsymbol{B} = \mathrm{Cov}(\boldsymbol{V})$ and $\boldsymbol{C} = \mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y})$, the distribution of $\boldsymbol{U} \,|\, \boldsymbol{V}$ is

$$\boldsymbol{U} \,|\, \boldsymbol{V} \sim N_{N_u}(\boldsymbol{\mu_u} + \boldsymbol{C}\boldsymbol{B}^{-1}(\boldsymbol{v} - \boldsymbol{\mu_v}), \boldsymbol{A} - \boldsymbol{C}\boldsymbol{B}^{-1}\boldsymbol{C}^\intercal), \tag{2.14}$$

with $\boldsymbol{v}$ the realisation of $\boldsymbol{V}$; see equation A.3 in appendix A of Eidsvik et al. (2015). By matrix computation of

$$\mathrm{Cov}(\boldsymbol{X}_t, \boldsymbol{Y}_t) = \mathrm{E}\left[(\boldsymbol{X}_t - \boldsymbol{\mu}_t)(\boldsymbol{Y}_t - \boldsymbol{H}_t\boldsymbol{\mu}_t)^\intercal\right] \tag{2.15}$$

with $Y_t = \boldsymbol{H}_t\boldsymbol{X}_t + \boldsymbol{\varepsilon}$,

$$\mathrm{Cov}(\boldsymbol{X}_t, \boldsymbol{Y}_t) = \boldsymbol{\Sigma}_t\boldsymbol{H}_t^\intercal. \tag{2.16}$$

Let the realisation of $\boldsymbol{Y}_t$ be $\boldsymbol{y}_t$. Then the distribution of $\boldsymbol{X}_t \,|\, \boldsymbol{Y}_t$ is

$$\boldsymbol{X}_t \,|\, \boldsymbol{Y}_t \sim N_N(\boldsymbol{\mu}_{t\,|\,\boldsymbol{y}_t}, \boldsymbol{\Sigma}_{t\,|\,\boldsymbol{y}_t}), \tag{2.17}$$

$$\boldsymbol{\mu}_{t\,|\,\boldsymbol{y}_t} = \boldsymbol{\mu}_t + \boldsymbol{\Sigma}_t\boldsymbol{H}_t^\intercal(\tau^2\boldsymbol{I}_m + \boldsymbol{H}_t\boldsymbol{\Sigma}_t\boldsymbol{H}_t^\intercal)^{-1}(\boldsymbol{y}_t - \boldsymbol{H}_t\boldsymbol{\mu}_t), \tag{2.18}$$

$$\boldsymbol{\Sigma}_{t\,|\,\boldsymbol{y}_t} = \boldsymbol{\Sigma}_t - \boldsymbol{\Sigma}_t\boldsymbol{H}_t^\intercal(\tau^2\boldsymbol{I}_m + \boldsymbol{H}_t\boldsymbol{\Sigma}_t\boldsymbol{H}_t^\intercal)^{-1}\boldsymbol{H}_t\boldsymbol{\Sigma}_t^\intercal, \tag{2.19}$$

which is the posterior distribution at time t. Observe that when the data model and the prior distribution are Gaussian, then so is the posterior distribution.

2.2 Spatio-temporal model

2.2.1 Process model

Now time is introduced as a variable and spatio-temporal GPs considered. A spatio-temporal stochastic process may be in continuous or discrete time. In this thesis we mostly consider the discrete case, though some derivations start out in the continuous case. For discrete or fixed time t, the system is denoted as in equations (2.1) and (2.2), while for an arbitrary time and space location, denote the state variable $X(t, \boldsymbol{s})$.

A common model for explaining development over time, it the Autoregressive model of order 1, AR(1). This is one of the simplest models used in time series theory and is a Markov process where the state at the next time step only depends on the previous one. The process is stationary in time, with all state vectors having the same distribution as long as no data is introduced (Cressie and Wikle, 2011). Let time be discrete, with $t = 0, 1, 2, \ldots$. If the mean at position $\boldsymbol{s}$ is $\mu_{\boldsymbol{s}}$, the AR(1) model is on the form

$$X_{t+1}(\boldsymbol{s}) - \mu_{\boldsymbol{s}} = \phi \cdot (X_t(\boldsymbol{s}) - \mu_{\boldsymbol{s}}) + \epsilon(\boldsymbol{s}). \tag{2.20}$$

The constant ϕ is the correlation between $X_{t+1}(\boldsymbol{s})$ and $X_t(\boldsymbol{s})$ and is thus restricted by $|\phi| \leq 1$. The ϵs are zero-mean random variables. They are independent and identically distributed in time, but correlated in space. We let ϕ be the same for all time steps and spatial positions. The state variables can be collected in a vector like for the initial state, see equation (2.2). We have the initial state $\boldsymbol{X}_0 \sim N_N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$. The spatial AR(1) model is then

$$\boldsymbol{X}_{t+1} - \boldsymbol{\mu}_0 = \phi(\boldsymbol{X}_t - \boldsymbol{\mu}_0) + \boldsymbol{\epsilon}. \tag{2.21}$$

In multivariate time series theory, it is called a *vector autoregressive* model (Cressie and Wikle, 2011), and $\phi\boldsymbol{I}_N$ is the *propagator matrix*. We say that $\boldsymbol{\epsilon}$ is "white" (uncorrelated) in time and "coloured" (correlated) in space. Its covariance matrix is determined by the

fact that all state vectors, $\boldsymbol{X}_0, \boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ are equally distributed, $\boldsymbol{X}_t \sim N_N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$. We have

$$\begin{aligned} \mathrm{E}[\boldsymbol{X}_{t+1}] &= \boldsymbol{\mu}_0 + \phi(\mathrm{E}[\boldsymbol{X}_t] - \boldsymbol{\mu}_0) = \boldsymbol{\mu}_0, && (2.22) \\ \mathrm{Cov}(\boldsymbol{X}_{t+1}) = \boldsymbol{\Sigma} &= \phi^2 \mathrm{Cov}(\boldsymbol{X}_t) + \mathrm{Cov}(\boldsymbol{\varepsilon}) \\ &= \phi^2 \boldsymbol{\Sigma}_0 + \mathrm{Cov}(\boldsymbol{\varepsilon}) && (2.23) \\ \implies \mathrm{Cov}(\boldsymbol{\epsilon}) &= (1-\phi^2)\boldsymbol{\Sigma}_0. \end{aligned}$$

In this thesis, we consider a more complex relationship between consecutive state vectors, but the process is still linear, and has the Markov property in time. The spatio-temporal process of spreading of a medium in a fluid can be expressed by the advection-diffusion SPDE (Sigrist et al., 2015b), and the process model is based on this. The equation is

$$\frac{\partial}{\partial t} X(t, \boldsymbol{s}) = -\boldsymbol{v}^\intercal \nabla X(t, \boldsymbol{s}) + \nabla \cdot \boldsymbol{D} \nabla X(t, \boldsymbol{s}) + \zeta X(t, \boldsymbol{s}) + \tilde{\eta}(t, \boldsymbol{s}), \tag{2.24}$$

with $\boldsymbol{v}$ the drift vector for advection, $\boldsymbol{D}$ the diffusion matrix, $\zeta \in [-1, 0]$ a damping constant controlling the auto regressive relationship between state vectors in time (Richardson, 2017) and $\tilde{\eta}(t, \boldsymbol{s})$ is the noise term of the SPDE, and is also called the *innovation term*. It is assumed a temporarily white and spatially coloured GP (Sigrist et al., 2015b). All terms collected to a N-vector gives $\tilde{\boldsymbol{\eta}}(t, \cdot) \sim N_N(\boldsymbol{0}, \tilde{\boldsymbol{Q}}_t)$. The covariance matrix is constructed in the same way as the initial state covariance matrix (equation (2.5)), but the parameters and correlation function may differ. These depend on the application, and the specifics of the model used in this thesis are described in chapter 3. One can have additional information about the system. This may be a point source for the pollutants or information along the boundary. This is added as a term $\tilde{R}(t, \boldsymbol{s})$ to the right hand of the SPDE,

$$\frac{\partial}{\partial t} X(t, \boldsymbol{s}) = -\boldsymbol{v}^\intercal \nabla X(t, \boldsymbol{s}) + \nabla \cdot \boldsymbol{D} \nabla X(t, \boldsymbol{s}) + \zeta X(t, \boldsymbol{s}) + \tilde{R}(t, \boldsymbol{s}) + \tilde{\eta}(t, \boldsymbol{s}), \tag{2.25}$$

where $\tilde{R}(t, \boldsymbol{s}) = 0$ if we have no additional information in $\boldsymbol{s}$ at time t.

The SPDE in equation (2.25) has been solved in different ways. Sigrist et al. (2015b) use spectral representation with Fourier terms, and this solution is also implemented as a package for the statistical software R (Sigrist et al., 2015a). Solutions are fast and accurate, but stationarity in space is required, which is not a plausible assumption in the application for Frænfjorden. Another simple approach is to use a method of finite differences, and this is done in this thesis. Let the spatial component be discretised as before as $\boldsymbol{s} = (s_e, s_n)^\intercal$ and likewise $\boldsymbol{v} = (v_e, v_n)^\intercal$. Then

$$\boldsymbol{v}^\intercal \nabla X(t, \boldsymbol{s}) = v_e \frac{\partial}{\partial s_e} X(t, \boldsymbol{s}) + v_n \frac{\partial}{\partial s_n} X(t, \boldsymbol{s}). \tag{2.26}$$

Further, in this thesis we let the diffusion be isotropic. Then the diffusion matrix $\boldsymbol{D}$ can be replaced by a constant D, and

$$\nabla \cdot \boldsymbol{D} \nabla X(t, \boldsymbol{s}) = D \Delta X(t, \boldsymbol{s}) = D \left(\frac{\partial^2}{\partial s_e^2} X(t, \boldsymbol{s}) + \frac{\partial^2}{\partial s_n^2} X(t, \boldsymbol{s}) \right). \tag{2.27}$$

Forward differences are used in time and central differences in space. Let the spatial grid $\mathcal{D}_s$ have resolution ds_e in east direction and ds_n in the north direction. Further, let the time be discretised with steps of dt. Thus

$$\begin{aligned}
\frac{\partial}{\partial t}X(t,\boldsymbol{s}) &\approx \frac{X(t+dt,\boldsymbol{s})-X(t,\boldsymbol{s})}{dt},\\
\frac{\partial}{\partial s_e}X(t,\boldsymbol{s}) &\approx \frac{X(t,\boldsymbol{s}+(ds_e,0)^\intercal)-X(t,\boldsymbol{s}-(ds_e,0)^\intercal)}{2ds_e},\\
\frac{\partial}{\partial s_n}X(t,\boldsymbol{s}) &\approx \frac{X(t,\boldsymbol{s}+(0,ds_n)^\intercal)-X(t,\boldsymbol{s}-(0,ds_n)^\intercal)}{2ds_n},\\
\frac{\partial^2}{\partial s_e^2}X(t,\boldsymbol{s}) &\approx \frac{X(t,\boldsymbol{s}+(ds_e,0)^\intercal)-2X(t,\boldsymbol{s})+X(t,\boldsymbol{s}-(ds_e,0)^\intercal)}{ds_e^2},\\
\frac{\partial^2}{\partial s_n^2}X(t,\boldsymbol{s}) &\approx \frac{X(t,\boldsymbol{s}+(0,ds_n)^\intercal)-2X(t,\boldsymbol{s})+X(t,\boldsymbol{s}-(0,ds_n)^\intercal)}{ds_n^2}.
\end{aligned} \tag{2.28}$$

All inserted in (2.24), we obtain

$$\begin{aligned}
X(t+dt,\boldsymbol{s}) = X(t,\boldsymbol{s}) &+ dt\tilde{R}(t,\boldsymbol{s}) + dt\tilde{\eta}(t,\boldsymbol{s})\\
&+ dt\left(\zeta - \frac{2D}{ds_e^2} - \frac{2D}{ds_n^2}\right)X(t,\boldsymbol{s})\\
&+ dt\left(\frac{v_e}{2ds_e} + \frac{D}{ds_e^2}\right)X(t,\boldsymbol{s}-(ds_e,0)^\intercal)\\
&+ dt\left(-\frac{v_e}{2ds_e} + \frac{D}{ds_e^2}\right)X(t,\boldsymbol{s}+(ds_e,0)^\intercal)\\
&+ dt\left(\frac{v_n}{2ds_n} + \frac{D}{ds_n^2}\right)X(t,\boldsymbol{s}-(0,ds_n)^\intercal)\\
&+ dt\left(-\frac{v_n}{2ds_n} + \frac{D}{ds_n^2}\right)X(t,\boldsymbol{s}+(0,ds_n)^\intercal).
\end{aligned} \tag{2.29}$$

It is a usual approach to apply forward differences in time, but Richardson (2017) proposes to use backward differences, to ensure not having to invert dense matrices in filtering and posterior sampling. It has also been suggested to let the discretisation in space depend on the direction of the advection field in a so-called "upwind scheme". The central differences are here used so that it can be used in all advection directions.

Now let us introduce some scaling of the time discretisation, so that time changes can be denoted $t+1$ instead of $t+dt$. If the time is really $\{t = t_0 + j \cdot dt; j = 0,1,2\dots\}$, with t_0 the time for the initial state, let us refer to the time steps $j = 0,1,2,\dots$. This variable is henceforth denoted t. In addition, we only consider a finite number of time steps, $t = 0,1,\dots,T$, $T \in \mathbb{N}$. Ignoring the boundary for now, the terms of equation (2.29) can be collected for all spatial positions in the domain to obtain a process model on the form

$$\boldsymbol{X}_{t+1} = \boldsymbol{A}_{t+1}\boldsymbol{X}_t + \boldsymbol{R}_{t+1} + \boldsymbol{\eta}_{t+1}. \tag{2.30}$$

Here, $\boldsymbol{X}_t$ is the state N-vector at time step t for all spatial positions and $\boldsymbol{A}_{t+1}$ is the $N \times N$ propagator matrix for time step $t+1$ including the information about advection,

diffusion and damping from the finite differences. We see that the state at one time step only depends on the state one time step earlier, so the process is a Markov process in time. It is possible to let the advection and diffusion vary with time and spatial position: $\boldsymbol{v} = \boldsymbol{v}(t, \boldsymbol{s}) = (v_e(t, \boldsymbol{s}), v_n(t, \boldsymbol{s}))^\intercal$, $D = D(t, \boldsymbol{s})$. In the general model, we therefore write $\boldsymbol{A}_{t+1}$, since the propagator matrix may vary with time. Let $\boldsymbol{\eta}_{t+1}$ be the innovation N-vector for the same time step. We have $\boldsymbol{\eta}_{t+1}[i] = dt \cdot \tilde{\eta}(t, \boldsymbol{s}_i)$. Thus $\boldsymbol{\eta}_{t+1} \sim N_N(\boldsymbol{0}, \boldsymbol{Q}_{t+1})$, $\boldsymbol{Q}_{t+1} = dt^2 \tilde{\boldsymbol{Q}}_{t+1}$. Further, $\boldsymbol{\eta}_{t+1}$ is independent of $\boldsymbol{X}_t$. The N-vector $\boldsymbol{R}_{t+1}$ is simply $dt \cdot \tilde{\boldsymbol{R}}_{t+1}$, where $\tilde{\boldsymbol{R}}_{t+1}$ is the collection of all terms $\tilde{R}(t, \boldsymbol{s})$. The parameters $\boldsymbol{v}$, D, ζ, $\boldsymbol{R}$ and $\boldsymbol{Q}$ are fixed, but unknown, and must be specified in the context of an application. In this thesis, we assume that they are all independent of time, as long as the time step length dt is fixed. Then the process model can be simplified as

$$\boldsymbol{X}_{t+1} = \boldsymbol{A}\boldsymbol{X}_t + \boldsymbol{R} + \boldsymbol{\eta}. \tag{2.31}$$

This process model describes the time dynamics of the state vector, and the model is (in general) not stationary in time. We let the advection or *drift field* vary in space: $\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{s}) = (v_e(\boldsymbol{s}), v_n(\boldsymbol{s}))^\intercal$, so the elements of $\boldsymbol{A}$ vary for each row, but $\boldsymbol{A}$ is the same for all time steps.

In the test example, we let the drift of pollution be in the east direction, with a constant speed of $v_e(\boldsymbol{s}) = 0.05$ m/s, diffusion constant $D = 0.1$ m^2/s and $\zeta = -0.0001$. The grid and time scales are set to be similar to the situation in Frænfjorden, which will be clear in chapter 3. The distance between each grid node is 20.1 m and one time unit is 10 minutes. We use 1 minute time steps, i.e. $dt = 0.1$, and let the covariance matrix of $\tilde{\boldsymbol{\eta}}$ be the same as the initial state covariance matrix.

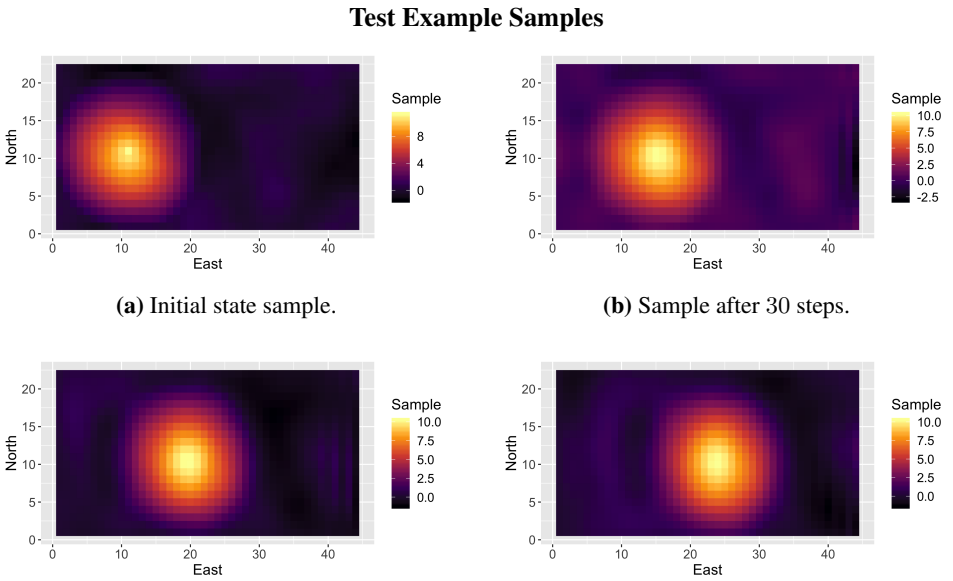
When parameters are found so all terms are known in the propagator matrix $\boldsymbol{A}$, the covariance matrix $\boldsymbol{Q}$ and the constant vector $\boldsymbol{R}$ for all time steps $1, \dots T$, the process model can be used to update the state model as time passes. Assume that $\boldsymbol{X}_t \sim N_N(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$. Then since $\boldsymbol{X}_t$ is a GP and the model is linear, $\boldsymbol{X}_{t+1}$ is also a GP. The mean and covariance matrix are

$$\mathrm{E}[\boldsymbol{X}_{t+1}] = \mathrm{E}[\boldsymbol{A}\boldsymbol{X}_t] + \mathrm{E}[\boldsymbol{R}] + \mathrm{E}[\boldsymbol{\eta}] = \boldsymbol{A}\boldsymbol{\mu}_t + \boldsymbol{R}, \tag{2.32}$$

$$\mathrm{Cov}(\boldsymbol{X}_{t+1}) = \mathrm{Cov}(\boldsymbol{A}\boldsymbol{X}_t) + \mathrm{Cov}(\boldsymbol{\eta}) = \boldsymbol{A}\boldsymbol{\Sigma}_t\boldsymbol{A}^\intercal + \boldsymbol{Q}. \tag{2.33}$$

The latter is correct because of the independence between $\boldsymbol{\eta}$ and $\boldsymbol{X}_t$. By simple induction, when the initial state is Gaussian distributed, all states $\boldsymbol{X}_t$ are Gaussian distributed.

The boundary must of course also be considered, in form of a set of boundary conditions (BCs) for the SPDE. There are several approaches for setting the BCs, depending on what information is available or what kind of physical properties the boundary has. For the test example, 0 Dirichlet BCs are used. Dirichlet BCs require knowledge of the state values, and can be seen as a constant source along the boundaries. In the discretisation for the boundary nodes, the state random variable for the position outside the grid is replaced by the known value at the boundary. For the left boundary, $X(t, \boldsymbol{s} - (ds_e, 0)^\intercal)$ is outside the grid, and replaced by the known boundary value at this position, say $\beta_{\boldsymbol{s}}$. Then the $X(t, \boldsymbol{s} - (ds_e, 0)^\intercal)$-term in equation (2.29) becomes $dt(\frac{v_e}{2ds_e} + \frac{D}{ds_e^2})\beta_{\boldsymbol{s}}$. This term is not part of the propagator matrix; all BC terms are collected in the $\boldsymbol{R}$ vector on the right



(a) Initial state sample.

(b) Sample after 30 steps.

(c) Sample after 60 steps.

(d) Sample after 90 steps.

Figure 2.5: Samples from the test example using the advection-diffusion SPDE process model. The time steps are of 1 minute, the distance between each grid node is approximately 20 m and the advection is 0.05 m/s in the east direction.

hand side of equation (2.31). With 0 along the boundary, no information is contained in this vector for the test example. A few snapshots of the process over time is shown in Figure 2.5, starting from the initial state. Some boundary effects can be seen close to the right boundary as pollutants move in that direction, but these are small and far from the pollutants. The scale in Figure 2.5b is slightly changed because of these numerical effects.

Using Dirichlet BCs is convenient, but if the values along the boundary are in fact not known, the approach will add a sense of false certainty in the values near the boundary. For the initial state of the test example, the concentration of pollution is 0 or very close to 0 along the boundary. We use Dirichlet BCs with 0 concentration along the boundary for the whole process. How appropriate this is depends on the movement of the pollutants. Consider a case where the advection is in the north direction. After a while, the concentration is far from 0 near the boundary, and we get some numerical effect close to the boundary, as shown in Figure 2.6. When the movement is in the east direction, however, the pollutants can move quite far without getting close to boundary, and the BCs are more realistic.

Figure 2.7 shows the mean and variances for concentrations in the test example after some time steps. We see that the mean pollution is influenced by the advection, and moves east over time. The variance is affected by the BCs. Along the north and south boundary, the variance very close to the boundary is small, since the BCs state that we know the values on the boundary. Further, the drift of the pollutants is in the east direction. This means that setting BCs on the west boundary to 0, equals having incoming 0 concentration at all times here. After some time, we therefore have close to 0 concentration further

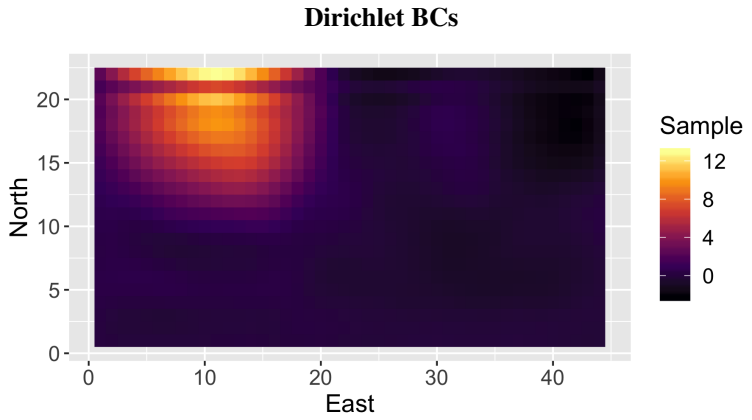
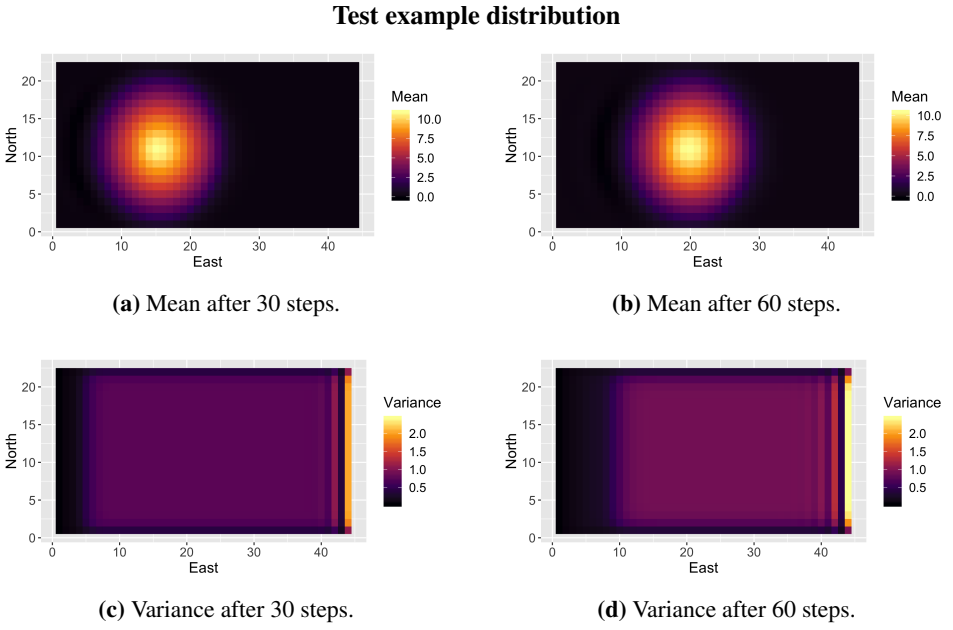


Figure 2.6: An example where the advection is in the north direction with 0 Dirichlet BCs. The 0 concentration along the north boundary then becomes unrealistic.



(a) Mean after 30 steps.

(b) Mean after 60 steps.

(c) Variance after 30 steps.

(d) Variance after 60 steps.

Figure 2.7: Test example mean and variance after 30 and 60 time steps using the SPDE model. The mean follows the advection, and the Dirichlet BCs affect the variance along the boundaries, in particular on the western side of the domain.

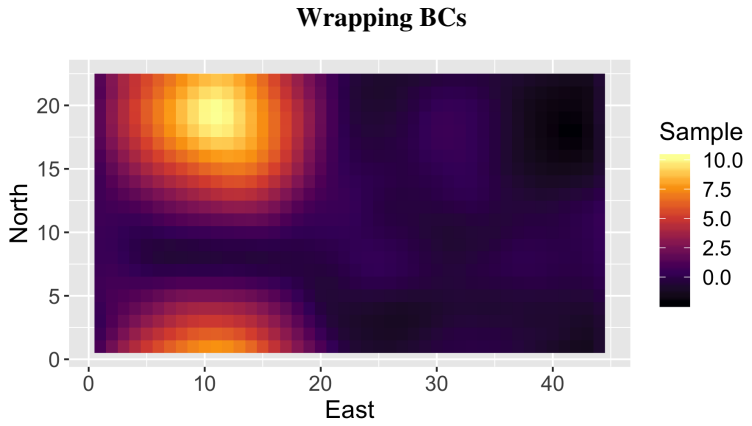


Figure 2.8: An example where the advection is in the north direction with wrapping BCs. The pollution that moves out of the domain through the north boundary enters through the south boundary.

into the area, and the variance here is small. Along the right boundary, we get higher variance due to the numerical effects seen when pollution is about to leave the area, and 0 concentration along the boundary is unrealistic at times. In the rest of the area, the variance increases from 0.6 initially to 0.76 after 30 steps and 0.91 after 60 steps.

Other BCs may also be considered. One solution that is often used in spatial statistics is so-called torus or wrapping BCs. Then the domain is assumed wrapped around itself like a torus, so that the upper and lower boundaries are connected, and likewise for the left and right boundaries. In this case, the neighbours needed in the finite differences discretisation are always present, though sometimes on the other side of the domain. The advantage with this approach is that the model does not require additional information about the boundary. Rectangular grid wrapped around a torus is also indirectly implied when solving the SPDE using the fast Fourier transform, as done by Sigrist et al. (2015b). The disadvantage is that the different parts of the domain are affected by each other. Fluids moving out of the domain through one boundary, reenters through the opposite, as demonstrated in Figure 2.8 for advection in the north direction. To avoid this problem, the domain should be made large enough for the real area of interested not to be affected too much. This means that the SPDE must be solved on a system that is e.g. 4 times bigger than the one really of interest, which is computer demanding.

If the domain boundary is modelling a solid boundary, like land around the fjord, setting the normal derivative (derivative orthogonal to the boundary) to 0, will be a good BC. This is called a Neumann BC. Then no fluid/pollutants can flow out of the domain. The BC for the eastern boundary in a square domain is

$$\frac{\partial}{\partial s_e} X(t, \boldsymbol{s}) = 0, \tag{2.34}$$

for $\boldsymbol{s} \in \mathcal{B}_e$, with $\mathcal{B}_e$ the set of grid nodes directly near the eastern boundary. Discretised by central differences this becomes

$$\begin{aligned} &\frac{X(t, \boldsymbol{s} + (ds_e, 0)^\intercal) - X(t, \boldsymbol{s} - (ds_e, 0)^\intercal)}{2ds_e} = 0 \\ \implies &X(t, \boldsymbol{s} + (ds_e, 0)^\intercal) = X(t, \boldsymbol{s} - (ds_e, 0)^\intercal). \end{aligned} \tag{2.35}$$

Equation (2.35) can be inserted into equation (2.29) for the eastern boundary nodes, and similar expressions derived for the other boundaries. If we have a solid boundary, the advection field should also reflect this by not pointing out of the domain. If not, similar effects as for the Dirichlet BCs will be present along the boundary. It is also possible to let the derivative be non-zero, but this again requires additional information. We then get constants that can be included in the vector $\boldsymbol{R}$ of equation (2.31), just as for Dirichlet BCs. It is also possible to have different BCs for different time steps, but this is not considered in this thesis.

2.2.2 Data assimilation – Kalman filter

Now we have the tools needed for updating the model both as time goes by and as new information gets available. The set of equations evolving from combining these is called a Kalman filter, and is derived in this section on the form of those of Cressie and Wikle (2011) and Särkkä (2013), though some names are changed. First, some new notation is introduced.

Let $\boldsymbol{Y}_{1:t}$ be all available data from time step 1 to t and define

$$\boldsymbol{\mu}_{t\,|\,t-1} = \mathrm{E}[\boldsymbol{X}_t \,|\, \boldsymbol{Y}_{1:t-1}], \tag{2.36}$$

$$\boldsymbol{P}_{t\,|\,t-1} = \mathrm{Cov}(\boldsymbol{X}_t \,|\, \boldsymbol{Y}_{1:t-1}) = \mathrm{E}[(\boldsymbol{X}_t - \boldsymbol{\mu}_{t\,|\,t-1})(\boldsymbol{X}_t - \boldsymbol{\mu}_{t\,|\,t-1})^\intercal \,|\, \boldsymbol{Y}_{1:t-1}], \tag{2.37}$$

and likewise

$$\boldsymbol{\mu}_{t\,|\,t} = \mathrm{E}[\boldsymbol{X}_t \,|\, \boldsymbol{Y}_{1:t}], \tag{2.38}$$

$$\boldsymbol{P}_{t\,|\,t} = \mathrm{Cov}(\boldsymbol{X}_t \,|\, \boldsymbol{Y}_{1:t}) = \mathrm{E}[(\boldsymbol{X}_t - \boldsymbol{\mu}_{t\,|\,t})(\boldsymbol{X}_t - \boldsymbol{\mu}_{t\,|\,t})^\intercal \,|\, \boldsymbol{Y}_{1:t}]. \tag{2.39}$$

We have seen that the state vector is always Gaussian distributed, since the processes of updating the state are always linear in Gaussian distributed random vectors. In the Kalman filter, updates with respect to time and new data are computed sequentially, and the state vector is still always Gaussian distributed.

Consider the state at time t, and assume that the distribution of $\boldsymbol{X}_{t-1}$ given $\boldsymbol{Y}_{1:t-1}$ is known. We have

$$\boldsymbol{X}_t \,|\, \boldsymbol{Y}_{1:t-1} \sim N_N(\boldsymbol{\mu}_{t\,|\,t-1}, \boldsymbol{P}_{t\,|\,t-1}), \tag{2.40}$$

$$\boldsymbol{X}_t \,|\, \boldsymbol{Y}_{1:t} \sim N_N(\boldsymbol{\mu}_{t\,|\,t}, \boldsymbol{P}_{t\,|\,t}). \tag{2.41}$$

The Kalman filter consists of updating the distribution of $\boldsymbol{X}_t$ in two steps. First, in the prediction step, the *forecasting* or *predictive distribution* of equation (2.40) is obtained. In this step, the process is updated in time, without adding any new data from time $t-1$

to time t. Thus, the expected value and the covariance matrix are updated according to equations (2.32) and (2.33). With $\boldsymbol{X}_{t-1} \,|\, \boldsymbol{Y}_{1:t-1} \sim N_N(\boldsymbol{\mu}_{t-1\,|\,t-1}, \boldsymbol{P}_{t-1\,|\,t-1})$, we get

$$\boldsymbol{\mu}_{t\,|\,t-1} = \boldsymbol{A}\boldsymbol{\mu}_{t-1\,|\,t-1} + \boldsymbol{R}, \tag{2.42}$$

$$\boldsymbol{P}_{t\,|\,t-1} = \boldsymbol{A}\boldsymbol{P}_{t-1\,|\,t-1}\boldsymbol{A}^\intercal + \boldsymbol{Q}. \tag{2.43}$$

Figure 2.9 shows a drawing of one step prediction and dependencies between the state vectors and the data. The data needed to obtain the forecasting distribution of $\boldsymbol{X}_t$ is also indicated in the display.

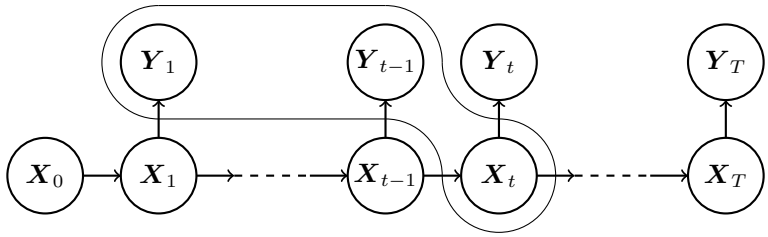


Figure 2.9: One step prediction in the Kalman filter. Data up to time $t-1$ is needed to calculate the forecasting distribution of $\boldsymbol{X}_t$, as marked in the display. Arrows indicate dependencies between the state vectors and obtained data.

The second step, called the *filtering* step, accounts for new data at time t, and we call the distribution in equation (2.41) the *filtering distribution*. The formulas for the mean and covariance matrix correspond to the ones in equations (2.18) and (2.19) with the predictive distribution of $\boldsymbol{X}_t \,|\, \boldsymbol{Y}_{1:t-1}$ as the prior, but are usually written in a slightly different way. We found the marginal distribution of $\boldsymbol{Y}_t$ in equations (2.11) and (2.12), and similarly we can find that

$$\boldsymbol{Y}_t \,|\, \boldsymbol{Y}_{1:t-1} \sim N(\boldsymbol{H}_t\boldsymbol{\mu}_{t\,|\,t-1}, \boldsymbol{S}_t), \tag{2.44}$$

$$\boldsymbol{S}_t = \boldsymbol{H}_t\boldsymbol{P}_{t\,|\,t-1}\boldsymbol{H}_t^\intercal + \tau^2\boldsymbol{I}_m. \tag{2.45}$$

The covariance matrix, $\boldsymbol{S}_t$, has dimensions $m \times m$, where m is the number of measured values, i.e. the dimension of $\boldsymbol{Y}_t$. Using the new notation and predictive distribution, equations (2.18) and (2.19) become

$$\boldsymbol{\mu}_{t\,|\,t} = \boldsymbol{\mu}_{t\,|\,t-1} + \boldsymbol{P}_{t\,|\,t-1}\boldsymbol{H}_t^\intercal\boldsymbol{S}_t^{-1}(\boldsymbol{y}_t - \boldsymbol{H}_t\boldsymbol{\mu}_{t\,|\,t-1}), \tag{2.46}$$

$$\boldsymbol{P}_{t\,|\,t} = \boldsymbol{P}_{t\,|\,t-1} - \boldsymbol{P}_{t\,|\,t-1}\boldsymbol{H}_t^\intercal\boldsymbol{S}_t^{-1}\boldsymbol{H}_t\boldsymbol{P}_{t\,|\,t-1}^\intercal. \tag{2.47}$$

Further, we define the $N \times m$ matrix $\boldsymbol{K}_t = \boldsymbol{P}_{t\,|\,t-1}\boldsymbol{H}_t^\intercal\boldsymbol{S}_t^{-1}$ the *Kalman gain* and note that $\boldsymbol{P}_{t\,|\,t-1}^\intercal = \boldsymbol{P}_{t\,|\,t-1}$ because of covariance symmetry. Then finally the filtering step equations are

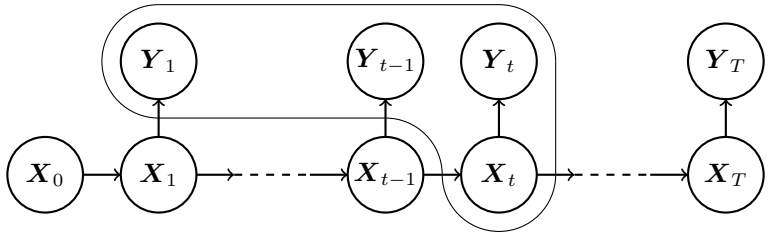


Figure 2.10: Filtering step in Kalman filter. Data up to time t is needed to calculate the filtering distribution of $\boldsymbol{X}_t$, as marked in the display. Arrows indicate dependencies between the state vectors and obtained data.

$$\boldsymbol{\mu}_{t\,|\,t} = \boldsymbol{\mu}_{t\,|\,t-1} + \boldsymbol{K}_t(\boldsymbol{y}_t - \boldsymbol{H}_t\boldsymbol{\mu}_{t\,|\,t-1}), \tag{2.48}$$

$$\boldsymbol{P}_{t\,|\,t} = (\boldsymbol{I}_N - \boldsymbol{K}_t\boldsymbol{H}_t)\boldsymbol{P}_{t\,|\,t-1}. \tag{2.49}$$

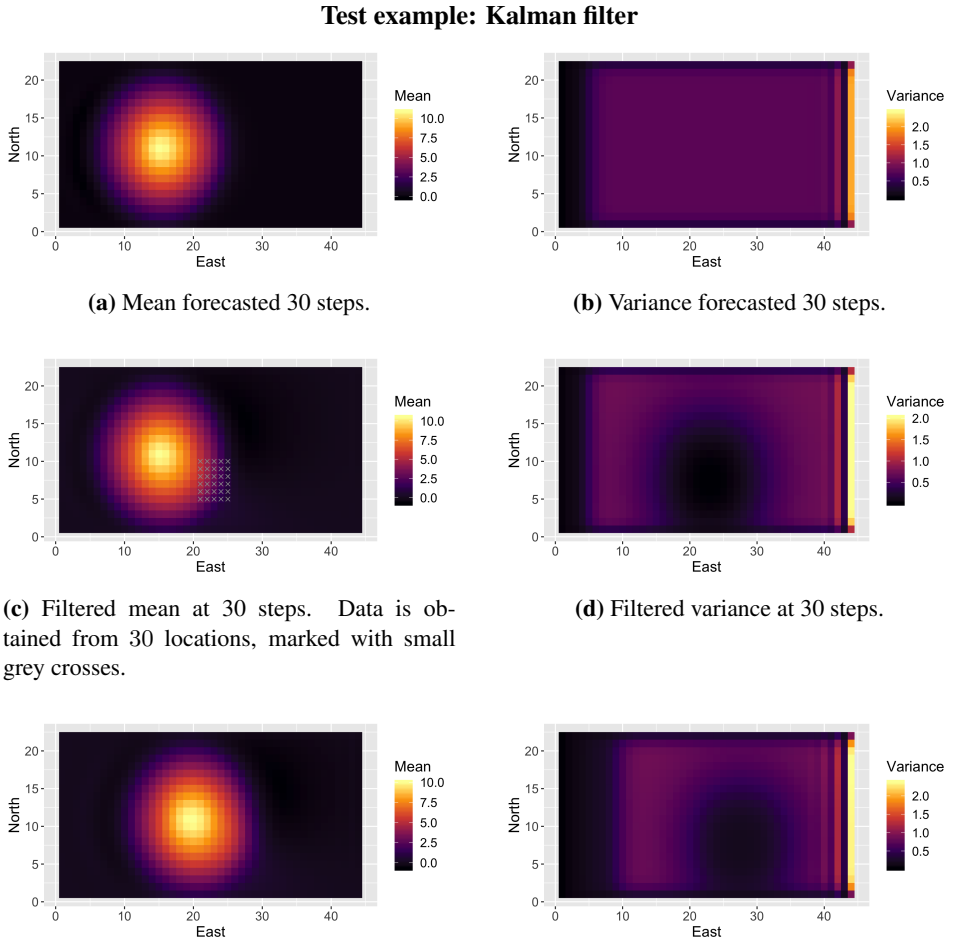
The dependencies are demonstrated in Figure 2.10. It is also possible to obtain these equations by first defining the filtering mean as the predictive mean weighted by the new measurement, as in equation (2.48). The weight, the Kalman gain, can then be found using least squares method. The result is the same either way. In this thesis, the dimension of $\boldsymbol{Y}_t$ is usually 1 for all t, and the covariance matrix $\boldsymbol{S}_t$ is then also of dimensions 1×1.

Figure 2.11 shows the effect on the distribution of obtaining data. It is mainly visible in the variance displays. There is some noise in the measurement, but the variance is very small in the area of measurements just after the data is obtained; around 0.01 in Figure 2.11d. The variance is gradually increased away from the measured area. As time passes, the distribution is forecasted and the uncertainty is increases. The variances in the measured locations after forecasting 30 new steps are around 0.2-0.4. The area of low variance is also moved due to the advection. The measured locations are marked in Figure 2.11c.

If we were to use the AR(1) model for developing the state model over time, the Kalman filter is slightly different. As long as no data is ever introduced, the state distribution is the same for every time step. When introducing data, the distribution is changed according to the Kalman filtering. The filtering equations are still valid as given in equations (2.48) and (2.49), but the forecasting equations are different. The forecasting distribution is characterised by

$$\mathrm{E}[\boldsymbol{X}_t \,|\, \boldsymbol{Y}_{1:t-1}] = \boldsymbol{\mu}_{t\,|\,t-1} = (1-\phi)\boldsymbol{\mu}_0 + \phi\boldsymbol{\mu}_{t-1\,|\,t-1}, \tag{2.50}$$

$$\mathrm{Cov}(\boldsymbol{X}_t \,|\, \boldsymbol{Y}_{1:t-1}) = \boldsymbol{P}_{t\,|\,t-1} = \phi^2\boldsymbol{P}_{t-1\,|\,t-1} + (1-\phi^2)\boldsymbol{\Sigma}_0. \tag{2.51}$$



(a) Mean forecasted 30 steps.

(b) Variance forecasted 30 steps.

(c) Filtered mean at 30 steps. Data is obtained from 30 locations, marked with small grey crosses.

(d) Filtered variance at 30 steps.

(e) Mean after forecasting 30 more steps.

(f) Variance after forecasting 30 more steps.

Figure 2.11: Demonstration of the distribution mean and variances when the model is updated in time and due to new data. The equations specifying the mean and covariance matrices are called a Kalman filter.

Chapter 3

Frænfjorden model

The GP model described in chapter 2 is now applied to a case of concentration of pollutants from mining waste in Frænfjorden. The mining company Omya Hustadmarmor AS produces calcium carbonate (lime) and releases mining deposits into the fjord. The permitted deposition area is shown by the orange dashed line in Figure 3.1. The transport lines for the tailings from the production site, as used in 2018, are marked in black in the same display.

Outside the deposition area, there is an upper limit to the permitted turbidity, or particle concentration. Monitoring stations are placed around the area to control the deposits and the environmental impacts (Glette, 2019). The research organisation SINTEF and NTNU have an ongoing project where adaptive AUV sampling is tested in Frænfjorden, and more field work in the area is expected. Berget et al. (2018) build a GP model for Frænfjorden and simulate sampling from an AUV, and identifying ESs would be particularly useful near the limit of the permitted area. A picture of the AUV used in the work by NTNU in Frænfjorden is shown in Figure 3.2. It is possible to use many different sensors with the AUV, but in this thesis, only the concentration of contaminants is of interest.

In this chapter, parameters for the initial state model and SPDE process model for Frænfjorden are estimated based on model data from the fjord.

3.1 Parameter estimation

Data from two SINTEF ocean models is used to build a model for Frænfjorden. Data from these models is also used by by Berget et al. (2018). Drift data is from the numerical ocean model SINMOD (Slagstad and McClimans, 2005) at 01.04.2013 and 02.04.2013. Pollution concentration data is provided by the model DREAM (Dose-related Risk and Effect Assessment Model) from 26.03.2013 to 04.04.2013. Among others, DREAM calculates the transport, exposure, dose and effects of chemicals in marine environments. The model is 3-dimensional and time-dependent (Rye et al., 2008), and the ocean dynamics in this model is delivered by SINMOD (Berget et al., 2018). The drift data is used as the advection field in the SPDE, while the pollution concentration is used to build an initial model and estimate the covariance matrix for the innovation term in the process model.

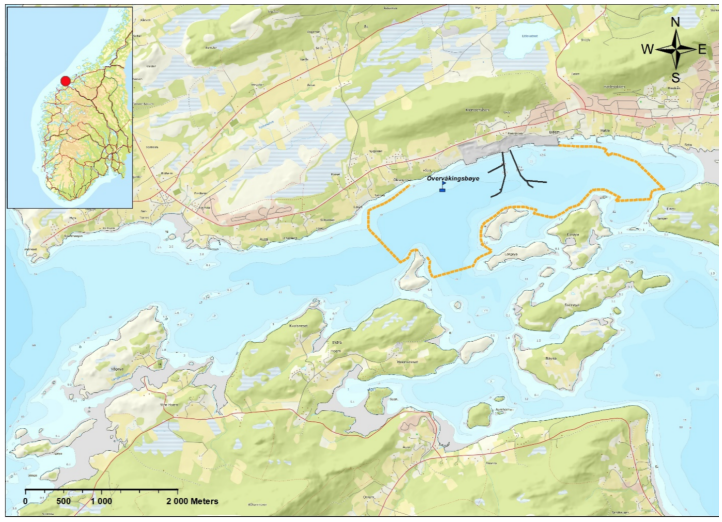
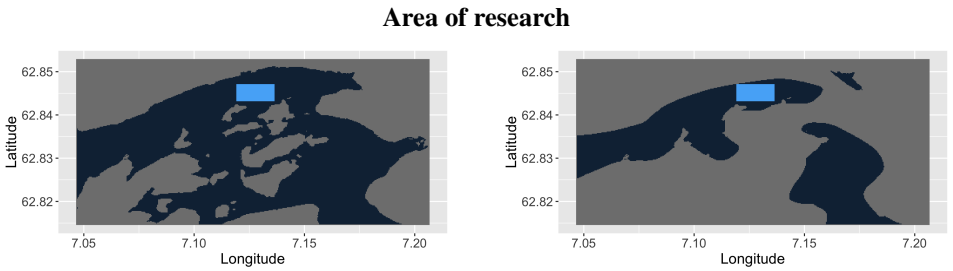


Figure 3.1: Area for tailing placements from Omya Hustadmarmor AS in Frænfjorden marked by the orange dashed line. Black lines show transport of tailings. Figure taken from Glette (2019).

Figure 3.2: The AUV used in Frænfjorden work. Picture taken from AUR lab (nd).



(a) 2 m below fjord surface.

(b) 22 m below fjord surface.

Figure 3.3: Frænfjorden at 2 and 22 m below the surface. Grey area indicates land, and the fjord is noticeably slimmer at depth 22 m than at the surface. The blue area indicates the area studied in this thesis. This area is in the inner part of the fjord, close to the location of Omya Hustadmarmor AS and outlet positions.

The outlets of pollutants happen at depth approximately 20 m (Glette, 2019), and a 2-dimensional grid is made by considering only depth 22 m of the DREAM data. For the convenience of having a regular, rectangular spatial grid, we only consider a small area in the inner part of the fjord, close to outlet positions and Omya Hustadmarmor AS. The area is shown in Figure 3.3 and displays the fjord just below the surface and at depth 22 m. For the latter, the fjord is quite slim, and the considered area is about the same width as the fjord in south-north direction. By choosing this small area, there is water and potentially pollution in every grid node. The grid is the same as shown for the test example in Figure 2.1, with $N_e = 44$ grid nodes in the east-west direction and $N_n = 22$ nodes in the north-south direction. The nodes are in the DREAM data given as coordinates of longitude and latitude. By the calculator from National Geospatial-Intelligence agency (Length of degree, nd), at latitude 62.845 degrees, one degree of latitude equals 111459 m and one degree of longitude equals 50941 m. This means there are approximately 20.1 m between each grid node in both directions.

The drift varies in both space and time. We therefore consider some shorter time periods, and assume that the drift is constant in time. Assuming that tidal streams are the main source of difference in the advection field and concentration flow, the data is split into periods starting at high tides, so that we have 18 periods for concentration and 4 for advection. It is found that the first period of concentration differs a lot from the other periods, especially in correlation, so this period is disregarded, leaving 17 relevant periods of concentration data. We assume the AUV is in the water sampling for 30 minutes, and for estimating parameters it is decided to consider data for 1-hour periods starting from a fixed time in the tide cycle. Qualitative study of some of the high to low tide periods of DREAM concentration leads to the choice of considering the hour starting from 5 hours after high tide, that is, approximately the last hour before low tide. In this period, inspection of concentration indicates that the drift is approximately constant. There are also some larger changes in the concentration field which might make the analysis interesting. The DREAM data has temporal resolution of 10 minutes, so we have 7 realisations of DREAM concentration data for each of the 17 1-hour periods. The drift data has one realisation every 20 minutes, so for this we have 4 realisations for each of 4 time periods. We

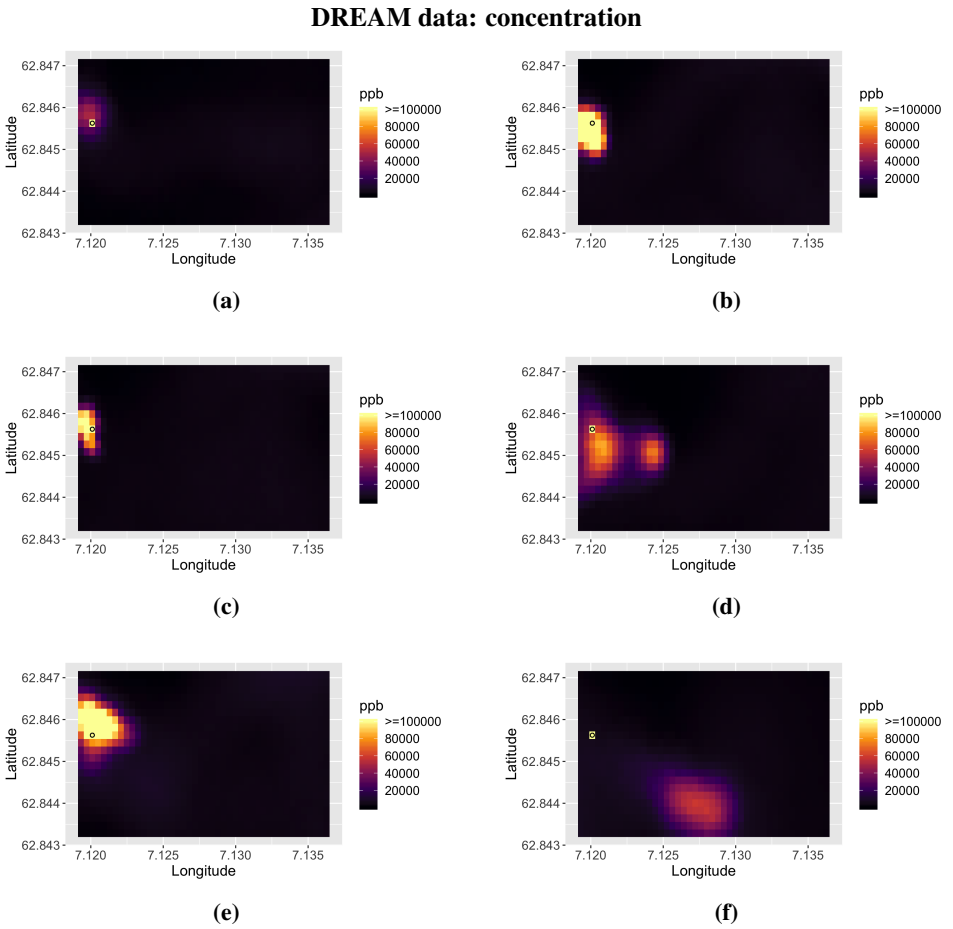


Figure 3.4: Concentrations from different time series at 5 hours after high tides. A source position with particularly high concentration of pollution is marked by a black circle.

assume realisations from different time periods are independent, since the time between the periods are as much as approximately 12 hours.

Figure 3.4 shows some realisations of concentration from DREAM at 5 hours after high tide, and Figure 3.5 shows the same realisations, but with concentrations on log scale ($\log(\text{concentration} + 1)$). We use concentration $+ 1$ here to avoid getting $-\infty$ values where the concentration is 0. There is a source position, or a deposition position, where the concentration is especially high. This marked by a black circle in the displays. The concentrations are given in *parts per billion* (ppb).

We see that the concentrations in Figure 3.4 decrease very fast from the source, and consider the concentration on log scale for the rest of the thesis. DREAM data is first used to build an initial state model for Frænfjorden, and then further to estimate parameters needed in the SPDE process model, as given in equation (2.25). We let the DREAM data

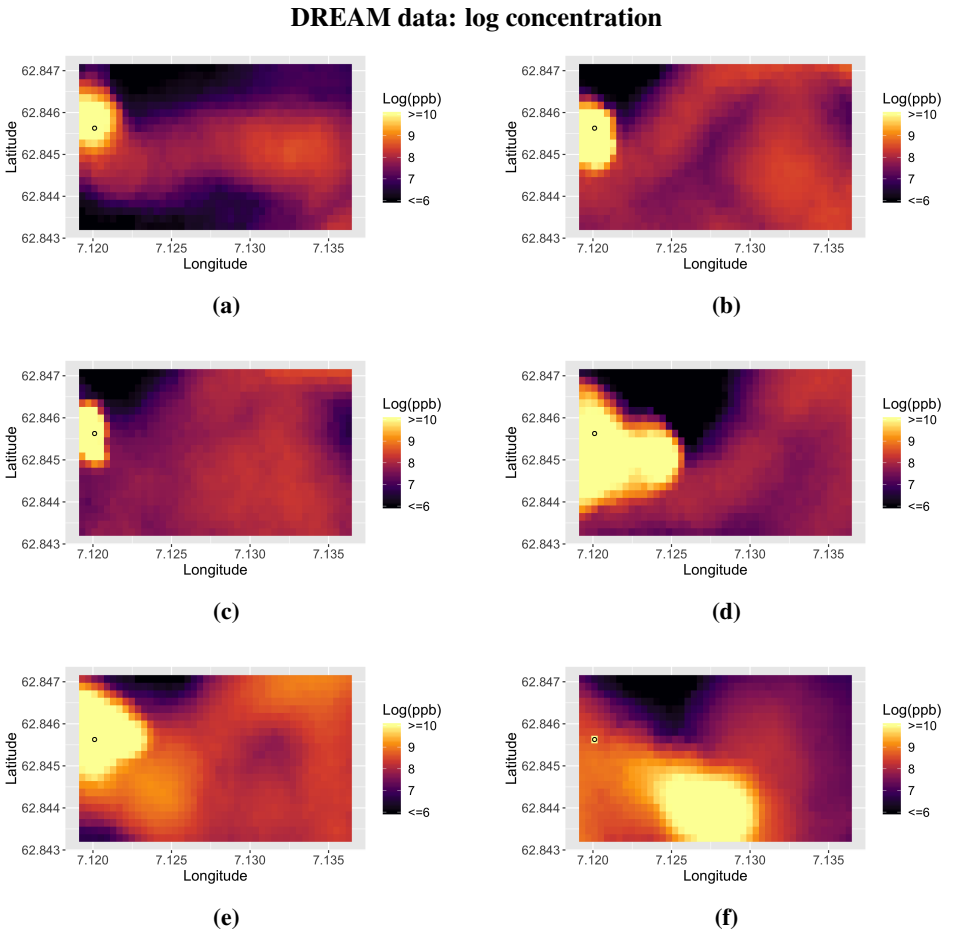


Figure 3.5: Dream data at 5 hours past high tides. A source position with particularly high concentration of pollution is marked by a black circle. The values are on the form $(\log(\text{concentration} + 1)$, and these are used to build the initial state model for Frænfjorden.

represent exact observations of the concentration in the domain. To use the notation of chapter 2, we assume we have knowledge of a set of $\boldsymbol{X}$s, not $\boldsymbol{Y}$s. This type of data is easier to use for parameter estimation.

3.1.1 Initial state model

To build the initial state model, we consider the DREAM data on log scale at five hours after each high tide, some of which as shown in Figure 3.5. We already know that the initial state should be a GP, and need to determine the mean vector and covariance matrix. The initial mean is taken to be the mean for each position over the 17 realisations, and is shown in Figure 3.6. The mean varies from position to position, so the model is not stationary in space.

Figure 3.7 shows the empirical standard deviation of concentration for each position. We see that even though the concentrations vary somewhat through the realisations, the standard deviation is typically below 1. For the initial state model, we want a covariance matrix on the form of equation (2.5). Thus we need to decide on one marginal variance σ_0^2 and a correlation function with range parameter. We do this by fitting theoretical variograms to empirical variograms. Figure 3.8 shows empirical variograms for the 17 initial state realisations subtracted the mean. They mostly have approximately the same size and shape. The mean variogram is used to fit the parameters of the covariance matrix.

At large distances, the empirical variograms differ a bit, due to little data at these distances, but the mean variogram does not have a dip in the end indicating negative correlation before stabilising at 0 correlation. Thus, the regular Matern or Exponential correlation functions seem better choices than the Damped Cosine. Figure 3.9 displays the same mean empirical variogram together with variograms for Matern and Exponential correlation functions from Table 2.1. The range parameter, ϕ, and noise σ_0^2 are marked as a vertical and a horizontal black line, respectively. These are set to fit the empirical variogram as well as possible. Both theoretical variograms actually match the empirical one very well. We make a covariance matrix for each correlation function and draw a sample as in equation (2.8) from the corresponding initial state models. These are shown in Figure 3.10. We see that the sample with Matern correlation is a lot smoother than the sample with Exponential correlation and also more similar to the DREAM data from Figure 3.5. Though the variograms do not match completely, the essence of the DREAM data seems to be captured. Thus, the initial state covariance matrix is given by equation (2.5), inserted $\sigma_0^2 = 0.6\ \log(\text{ppb})^2$ and the Matern correlation function from Table 2.1 with $\phi = 500\ \text{m}$.

3.1.2 SPDE parameters

Now the initial state distribution is determined, and the parameters for the SPDE process model are to be estimated. The parameters $\boldsymbol{v}$ and D are provided by SINMOD, and do not need estimation. The unknown parameters from equation (2.25) are the damping constant ζ, any information in the constant vector $\tilde{\boldsymbol{R}}$ and the covariance matrix of $\tilde{\boldsymbol{\eta}}$, $\tilde{\boldsymbol{Q}}$.

First we consider the advection parameter $\boldsymbol{v}$ and diffusion constant D, as given by SINMOD. We let the advection vary in space but be constant in time. The mean and standard deviation of drift data across all 4 periods between 5 and 6 hours after high tide for each grid position are shown in Figure 3.11 for both directions. The displays show

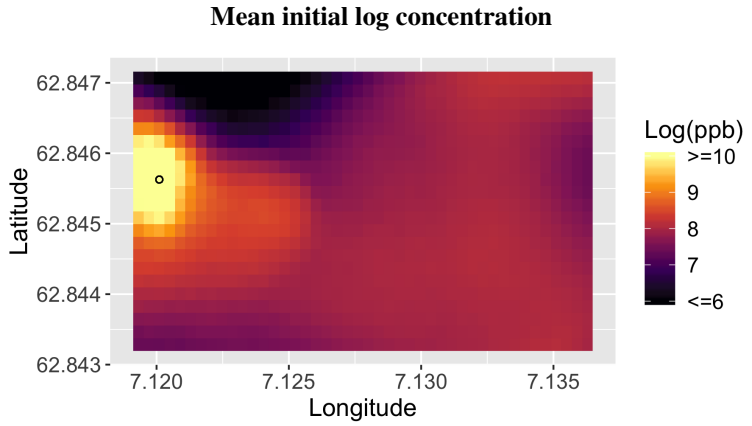


Figure 3.6: Mean log concentration of pollution in Frænfjorden at 5 hours past high tide. The mean is found from 17 realisations of DREAM data. The position marked by a black circle has particularly high initial concentration.

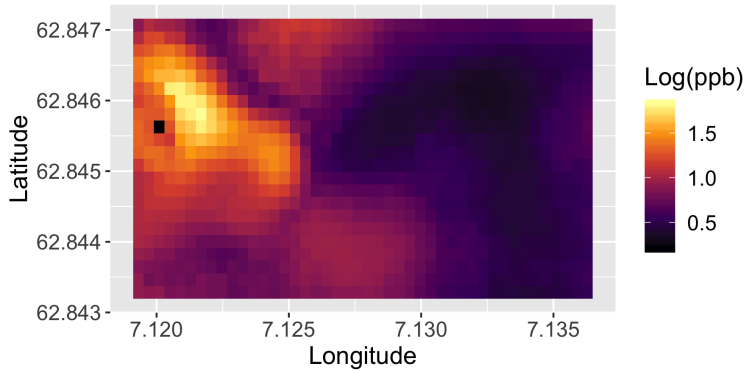


Figure 3.7: Empirical standard deviation of log concentration for each grid position over the 17 realisations of DREAM data for initial state.

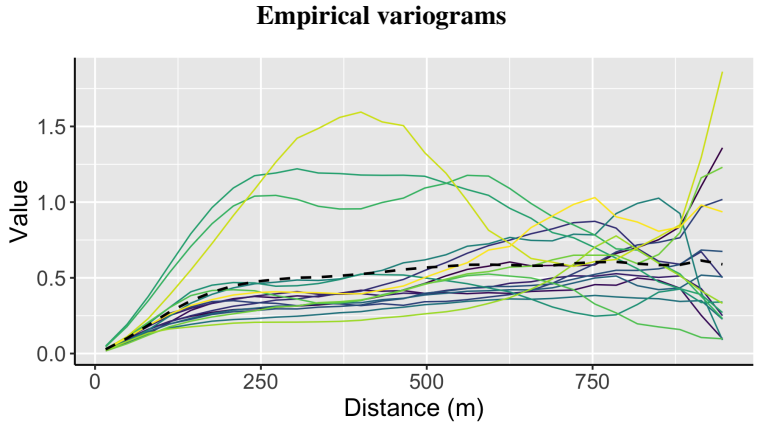


Figure 3.8: Empirical variograms for the 17 realisations of DREAM concentrations on log scale for initial state model. The mean variogram is marked by a dashed black line.

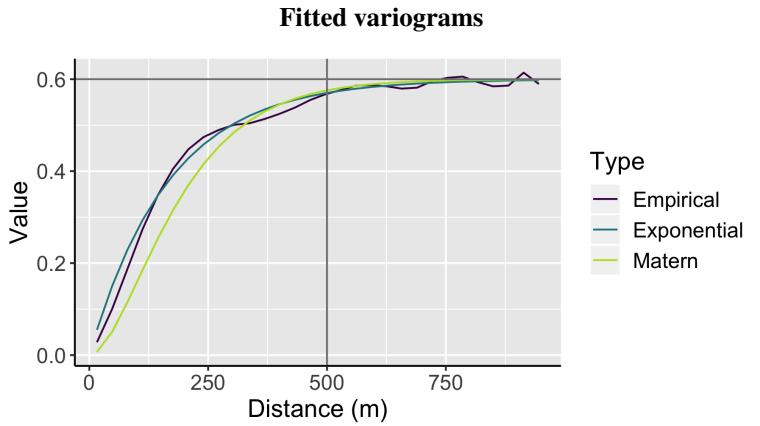
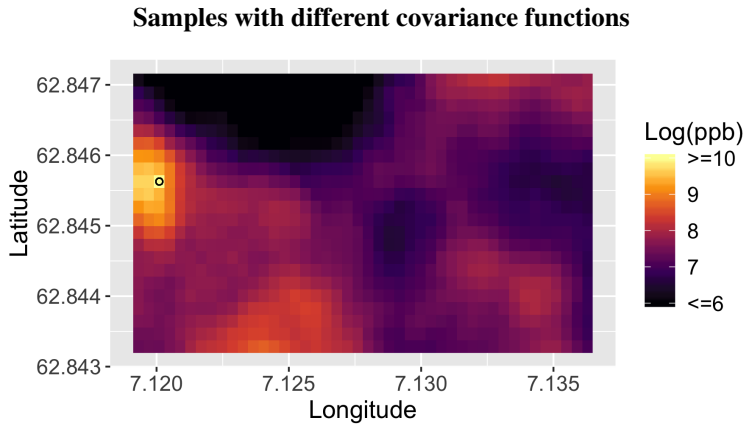
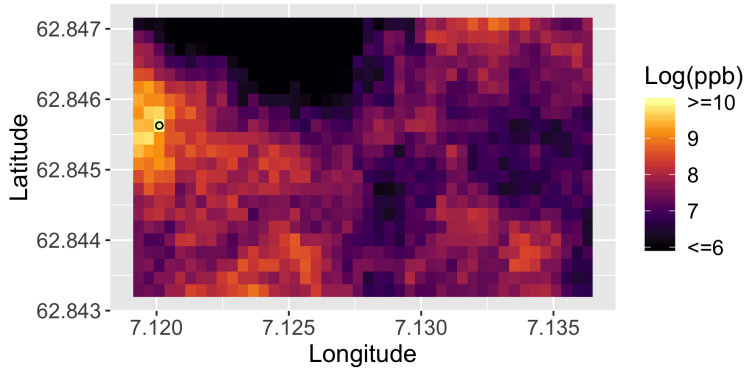


Figure 3.9: Mean empirical variogram together with variograms of the Matern and Exponential correlation functions. The range parameter, $\phi = 500$ m, is marked by a vertical line, and variance $\sigma_0^2 = 0.6\ \log(\text{ppb})^2$ by a horizontal line.

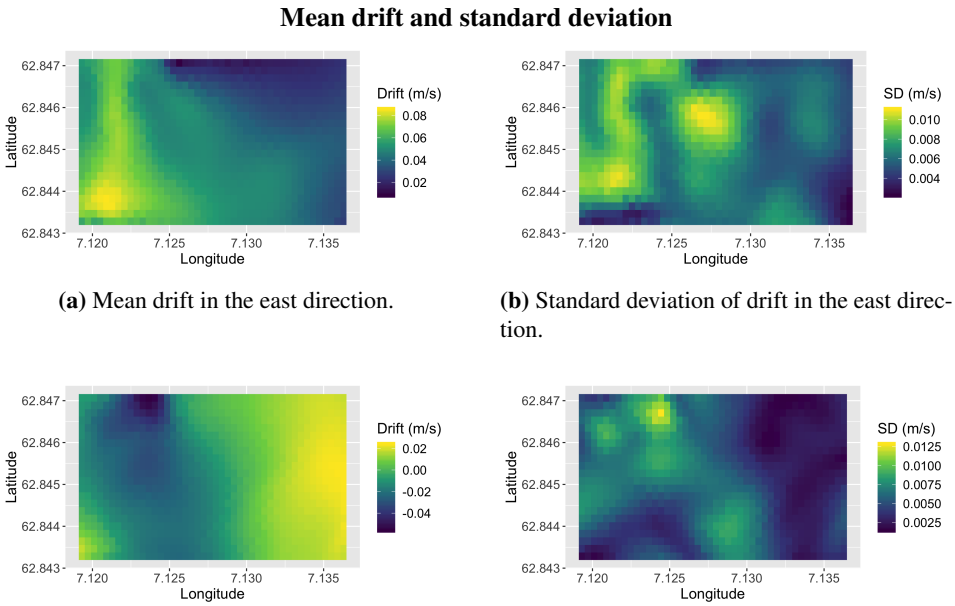


(a) Sample with Matern correlation.



(b) Sample with Exponential correlation.

Figure 3.10: Samples from the initial state distribution for Frænfjorden using two different correlation functions. The sample with Matern correlation is a lot smoother than the one with Exponential correlation.



(a) Mean drift in the east direction.

(b) Standard deviation of drift in the east direction.

(c) Mean drift in the north direction.

(d) Standard deviation of drift in the north direction.

Figure 3.11: Mean and standard deviation for advection over 4 1-hour periods.

that the standard deviations are quite large in some areas, meaning that the drift is actually changing quite a bit also within this 1-hour period. It seems that assuming constant drift field is not a very realistic assumption, as the field varies also for smaller periods. The variation is seen both within one period and between different periods. However, if we let the field vary with time, the propagator matrix $\boldsymbol{A}$ from equation (2.31) must be recalculated for each change in the advection, which means a lot more computations. We stick with the assumption of time independent parameters in this thesis. The means in Figure 3.11 are used as the fixed advection field in the model.

Figure 3.12 shows the directions of the average advection field. The lengths of the arrows indicate the relative speed of the drift. The drift in the east direction is positive for all positions, so the pollutants flow towards the east, though in different speeds for different positions in the domain. In the north direction, the pollution flows south for some positions and north for others. In general, pollutants located to the west in the domain drift towards the south. In the middle of the domain, the advection speed is low and only in the east direction, while the drift is towards the north-east in the eastern part of the domain.

The diffusion constant, D, is also given as $0.1\ \mathrm{m}^2/\mathrm{s}$. Before using the drift field and diffusion constant, we must scale them to fit the time and space units used in the DREAM data. As earlier stated, we have DREAM realisations every 10 minute ($600\ \mathrm{s}$), and the distance between each grid node is $20.1\ \mathrm{m}$. In simulation, we use 1 minute time steps, i.e. $dt = 0.1$.

Now consider the parameters that are not already given. From Figure 3.5, we see that

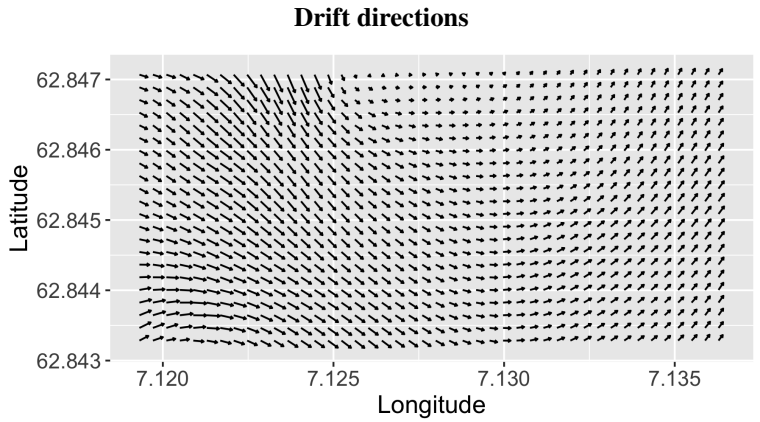


Figure 3.12: Arrows indicating the drift direction in each grid position and the relative drift speed.

there is incoming pollution from the left/west, and it varies a bit how far it has come at the initial time 5 hours after high tide. The source position in the area, though with very high concentration itself, does not seem to supply the area with more pollutants. The contaminants seem to be coming from an outlet to the west of the domain. Therefore, we do not add a point source to $\boldsymbol{R}$, but rather account for pollution drifting into the area through BCs. Only information from the BCs are then contained in $\boldsymbol{R}$.

As we have seen, we use only a small area of the fjord where there is no land, so that concentration data is found in every grid point. The advantage of this is that we get a regular grid, and the propagator matrix from section 2.2.1 is easier to make. However, choosing BCs is not simple. Assuming 0 concentration along the boundary does not look realistic from the data. Using other Dirichlet BCs or non-zero Neumann BCs require information about the area that is not available. Dirichlet BCs are chosen for the western boundary where pollutants are coming into the area. Dirichlet BCs should reflect absolute knowledge about the concentration in this area. This is of course not true in practice. In the model, we assume that the concentrations along the western boundary are constant in time and equal the initial mean, as this is our best guess. On the other boundaries, 0 Neumann BCs are applied, meaning that there can be no flow out of these boundaries. We can see in the drift direction plot, Figure 3.12, that there is actually flow across the boundaries, but the largest concentration values are located far from the boundary, and we assume the effect will be small. The southern and northern boundaries of the domain are quite close to fjord boundaries, as seen in Figure 3.3b, so here the assumptions should be quite good. Finding the ideal BCs is not the focus in this thesis, but some other BCs are also tried. Wrapping BCs are disregarded because of the large domain needed, as explained in section 2.2.1. Various Dirichlet BCs on all boundaries are tested, as well as non-zero Neumann BCs where the normal derivatives along the boundaries are set equal the drift in that direction. In the end, the combination of Dirichlet and 0 Neumann BCs are considered the most convenient choice for the work in this thesis.

Figure 3.13 shows the development of one DREAM data series within the 1-hour period. The corresponding initial state is found in Figure 3.5d. The high concentrated pol-

lution seems to move towards south-east throughout the hour, just as seen in Figure 3.12. There does not seem to be any damping in the DREAM data, so ideally we would set $\zeta = 0$. However, for stability reasons (Richardson, 2017), there should be a slight damping. Therefore, we set $\zeta = -0.0001$ in simulations. This small damping is not included when estimating the innovation term covariance.

For simplicity, we assume that the covariance matrix $\tilde{\boldsymbol{Q}}$ has the same correlation structure as the covariance matrix for the initial state, i.e. that it follows a Matern correlation function. To find the range and variance for $\tilde{\boldsymbol{Q}}$, we take a similar approach as with the initial state, and use empirical variograms to fit the Matern correlation function. We still disregard the first series, and thus have 17 1-hour series consisting in 7 realisations in each.

Let $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t+1}$ be consecutive realisations of DREAM data. The innovation term should explain why $\boldsymbol{x}_{t+1}$ differs from the prediction $\hat{\boldsymbol{x}}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{R}$ from equation (2.31). Since we have set the source constant to 0, $\boldsymbol{R}$ only consists of constants from the Dirichlet BCs. There are no unknown parameters in the predictions. From the DREAM data, we can find the residuals $\boldsymbol{x}_{t+1} - (\boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{R})$. Figure 3.14 shows the previous data $\boldsymbol{x}_t$ (Figure 3.14a), the prediction $\hat{\boldsymbol{x}}_{t+1}$ (Figure 3.14c), true $\boldsymbol{x}_{t+1}$ (Figure 3.14b) and the corresponding residuals (Figure 3.14d) for one step in the series of DREAM data from Figure 3.13. We see that the error in the prediction is mainly around the front of the highly concentrated pollution, and for higher concentrations, the size of the error tends also to be bigger. The Dirichlet BCs on the left boundary will never be completely correct using the initial state expected values, and the residuals are therefore also larger along this boundary. The residuals should be explained by the innovation term if the process follows the SPDE model. It seems that the innovation term variance should be larger for positions of high concentration than for positions of low concentration. In this thesis, however, the variance is considered the same across the whole grid.

We use the residuals to make empirical variograms. The variograms for the realisations within one time series tend to be similar, as Figure 3.15 demonstrates. Further, the mean variograms for each series are displayed in Figure 3.16 together with the overall mean in dotted line. We see that for large distances the variograms blow up. However, there is not much data this far apart, so the variograms are not reliable for large distances. Vertical lines have been added at distance $500\ \mathrm{m}$, and we consider only distances smaller than this. Like for the initial state, the variograms are following approximately the same shape, and we therefore assume the mean variogram represents the system, and use this to fit the parameters of the Matern correlation function. The result is found in Figure 3.17. Considering only the smaller distances before the empirical variogram blows up, the range and variance are both quite small. The parameters are set to $\phi = 120\ \mathrm{m}$ and $\sigma^2 = 0.1\ \log(\mathrm{ppb})^2$. It looks like the Matern-based variogram does not fit the data perfectly, and even that the empirical variogram has a dip that might suggest negative correlation for some distances. However, the scale on the y-axis is very small, so the difference is minimal. We have seen that the Matern correlation function results in smoother samples than the Exponential and the Damped Cosine, fitting the general appearance in the DREAM data. It could be considered if the same feature does not have to hold for the innovation term. However, in this thesis, the Matern correlation function is kept and used in simulations.

Since the movement of the DREAM data is based on SINMOD, and we use the drift field from SINMOD, it might not be very surprising that the noise in the DREAM data

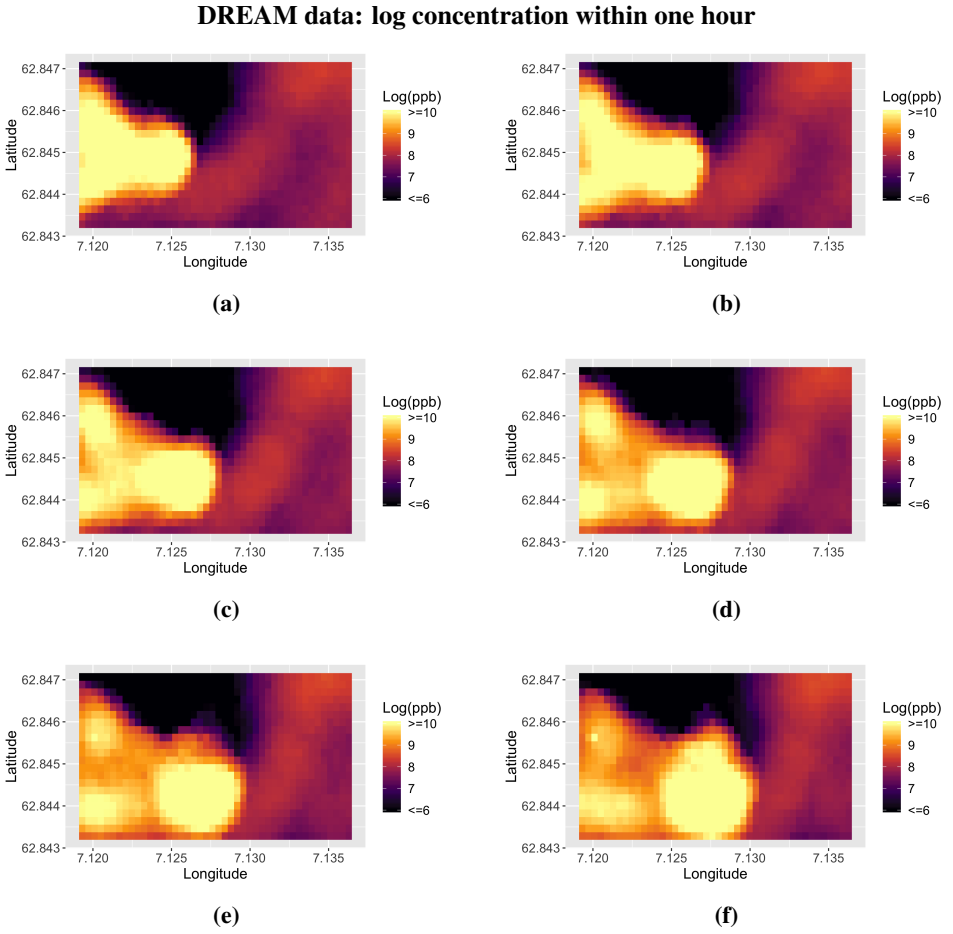
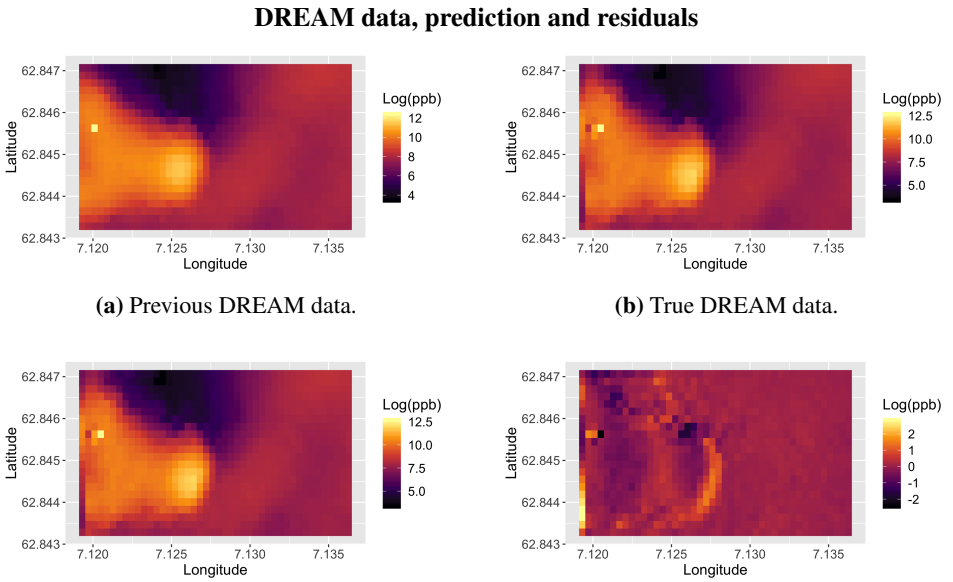


Figure 3.13: The concentration development of one series of DREAM data, following the displays in alphabetical order. The series starts at 10 minutes after the time of initial state. Incoming pollution from the western side moves in south-east direction.



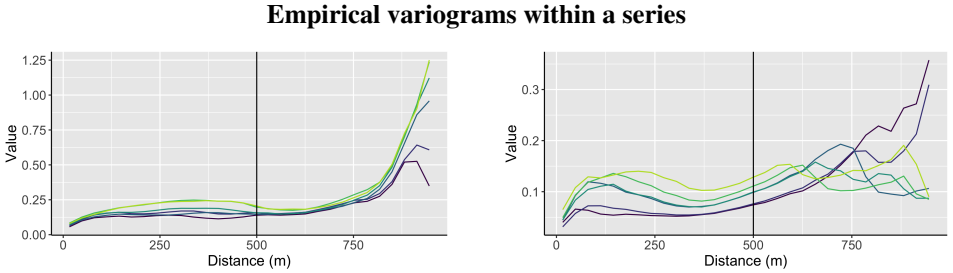
(a) Previous DREAM data.

(b) True DREAM data.

(c) Prediction.

(d) Residuals.

Figure 3.14: One demonstration of prediction and residuals when using the SPDE model to predict concentration for the next step. The DREAM data for time t is used for the prediction, and the prediction is compared to the true DREAM data for time $t + 1$. The residuals should be due to the innovation term if the data follows the SPDE model.



(a)

(b)

Figure 3.15: Empirical variograms for two series of DREAM data. The shapes are very similar within each series. We only consider the distances to the left of the vertical line, as the variograms tend to blow up for large distances, and there is nearly no data for such distances.

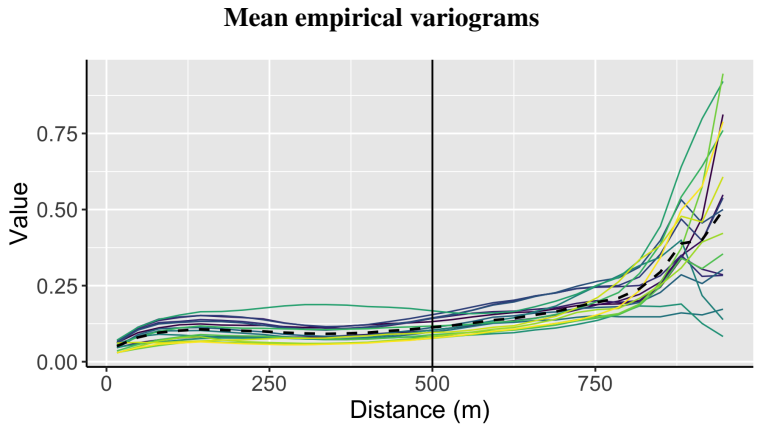


Figure 3.16: Mean empirical variograms for the 17 series of DREAM log concentration. We only consider the distances to the left of the vertical line.

is small. In this setting when DREAM is considered the "truth", we therefore get small variance for the innovation term. In practice, this might be a bit larger. To help ensure numerical stability and positive definiteness, we add a small nugget effect of 0.01^2 to the innovation term variances in simulations.

Finally, all necessary parameters are set to make a model for the pollution development in Frænfjorden. In chapter 5, this model will be used in a simulation study concerning data collected by AUVs.

3.2 Model variations

Some variations on the Frænfjorden model will also be used in simulations. Parameter values that could be possible for Frænfjorden, but are on the outer extremes of what one can expect, are considered. The reason is to check the model sensitivity in how best to obtain measurements in the area. It is seen that the advection data from SINMOD is not constant, so the advection speed may be both higher and lower than what is considered in the preceding section. Therefore, this seems the most pressing model parameter to test. We focus on varying the amplitude of the speed, and keep the direction field the same, for best being able to compare the different scenarios.

Let the drift velocity be v_e in the east direction and v_n in the north direction, as before. The advection for the Frænfjorden base case, as already found, is determined as the mean advection in each direction. For variation, the absolute speed ($\sqrt{v_e^2 + v_n^2}$) is computed for each position and all realisations of SINMOD data. We compute the mean speed for each position over the relevant realisations, as well as the empirical variance. Then we can choose "extreme" values that are still in the range of what is possible based on the SINMOD data. We use the speed of the base case ± 3 standard deviations. The standard deviations are different in each position, i.e. we scale the speed in each position differently, but the direction in each position stays the same. The direction in each position is

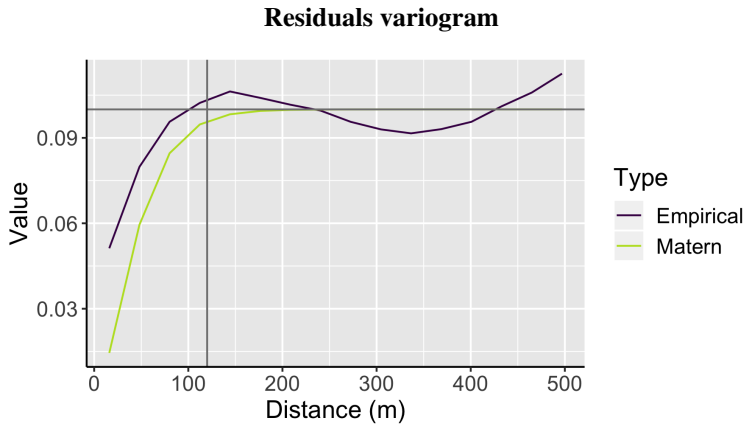


Figure 3.17: Mean empirical variogram over the 17 time series together with the variogram for the fitted Matern correlation function. The variance, $\sigma^2 = 0.1\ \log(\text{ppb})^2$ is marked by a horizontal line, and the range, $\phi = 120$ m, by a vertical line.

ensured by calculating direction (angle) based on the drift from the base case. Subtracting 3 standard deviations gives two slight negative numbers (-0.0475 and -0.0039), and since the speed must be non-negative, we simply truncate these to 0. Because of the small values (in amplitude), this is not noticeable in practice. The mean speed is $0.051556\ \text{m/s}$ for the Frænfjorden base case, $0.07307\ \text{m/s}$ for the high advection case and $0.03004\ \text{m/s}$ for the low advection case.

For further comparison, an AR(1) model for Frænfjorden will also be used. The model is given in equation (2.21), and given the initial distribution, only the parameter ϕ is unknown. The parameter is estimated by finding the sample correlation between consecutive realisations of DREAM data for each grid position. We have $K = 17$ assumed independent time series, and $T = 7$ realisations in each time series. As mentioned, we let the DREAM data be considered the complete accurate concentrations for the respective times, i.e. we have realisations $\boldsymbol{x}_t^{(k)}$ for times $t = 1, \dots T$ and time series $k = 1, \dots K$. Further, let $\overline{x}_{ti}$ be the average concentration in grid node $\boldsymbol{s}_i$ at time t over the K time series. We can then calculate the sample correlation between the concentration for some times t and $t+1$ for position $\boldsymbol{s}_i$ as

$$\text{Corr}\left(X_t(\boldsymbol{s}_i), X_{t+1}(\boldsymbol{s}_i)\right) \approx \frac{\sum_{k=1}^{K}\left(x_t^{(k)}(\boldsymbol{s}_i) - \overline{x}_{ti}\right)\left(x_{t+1}^{(k)}(\boldsymbol{s}_i) - \overline{x}_{t+1i}\right)}{\sqrt{\sum_{k=1}^{K}\left(x_t^{(k)}(\boldsymbol{s}_i) - \overline{x}_{ti}\right)^2 \sum_{k=1}^{K}\left(x_{t+1}^{(k)}(\boldsymbol{s}_i) - \overline{x}_{t+1i}\right)^2}}.$$

We estimate ϕ as the mean correlation over all positions and times, which gives $\phi_1 = 0.9522$. We call this ϕ_1 because it corresponds to the correlation between concentration 10 minutes apart. As we will use 1-minute intervals in simulations, we need some scaling. Assume the relation $\phi = \exp\{-dt/\theta\}$ for some constant scale θ. Since $dt = 1$ corresponds to 10 minutes, we have $\phi_1 = \exp\{-1/\theta\}$ and can find $\theta = -1/\phi_1$. Finally, $\phi = \exp\{-0.1/\theta\} = 0.9951$ is obtained for use in simulations. We see that the correla-

tion is very large with this data, but it is hard to determine whether it is *too* large. It might simply be that the AR(1) model, which assumes stationarity, is not a good model for the data.

Chapter 4

Method

The work in this thesis is on sensing of the concentration of pollution in Frænfjorden using AUVs. The aim is to choose locations from where the AUV obtains concentration measurements to get the best possible picture of where the concentration is above some critical limit. To do this, we need some tools for mapping where the concentration is below and above the limit. This is done in the first section of this chapter.

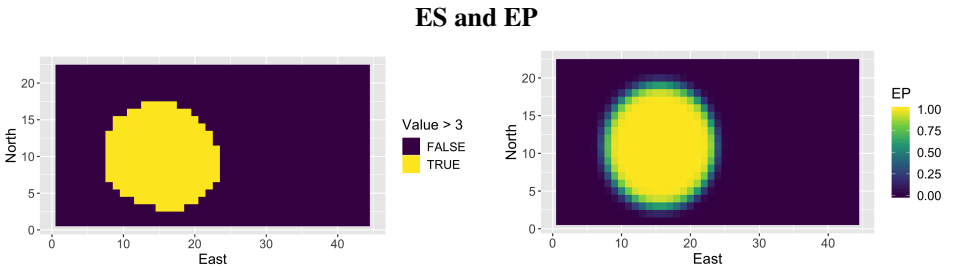
Further, we need some decision strategies for the movements of the AUV. The focus is mainly on adaptive decision strategies, and one main criterion is derived in this chapter. The ideas of this chapter are to be used together with the model for Frænfjorden from chapter 3 in a simulation study. Therefore, finally, the ideas of Monte Carlo simulation are presented.

4.1 Excursion sets and probabilities

ESs tell where the values in a system are above some critical limit. The aim in this thesis is to be able to make a good prediction of the ES of the system based on measurements. Identifying ESs is important in several environmental applications (Angulo and Madrid, 2010; French and Sain, 2013). Sommerfeld et al. (2018) have temperature change as application, and French and Sain (2013) explore both temperature and extreme precipitation. In the spatial setting, there exists a lot of work on ESs, on providing conservative estimates (Azzimonti et al., 2015) and credible regions (French and Hoeting, 2016). Bolin and Lindgren (2015) work in a Bayesian Gaussian setting, using properties of the posterior distribution, and they have also developed an R package for probabilistic ESs (Bolin and Lindgren, 2018). Let l be the critical limit of the system. A random ES over $\mathcal{D}_s$ can be defined as

$$\text{ES}_{\text{random}} = \{\boldsymbol{s}_i \in \mathcal{D}_s : X(\boldsymbol{s}_i) > l\}, \tag{4.1}$$

with $X(\boldsymbol{s}_i)$ the random variable for the state in position $\boldsymbol{s}_i$. In this thesis, we are interested in predicting such ESs. Considering the ESs fixed, with the true value at position $\boldsymbol{s}_i$ given as $x(\boldsymbol{s}_i)$, we have:



(a) True ES_T.

(b) EP for time $T = 30$ when no data is obtained, i.e. $\text{EP}_T^{(0)}$.

Figure 4.1: ES and EP for time $T = 30$ and concentration limit 3.

$$\text{ES}_{\text{fixed}} = \{\boldsymbol{s}_i \in \mathcal{D}_s : x(\boldsymbol{s}_i) > l\}. \tag{4.2}$$

In a system changing over time, different definitions of the ES are used. There exists work where the ES is defined very general, e.g. by Angulo and Madrid (2010) and Adler (2000), so that the ES is higher dimensional when adding temporal dependencies. If the interest is the ES at some fixed time, say after the final time step T, we can use a fixed-time ES:

$$\text{ES}_T = \{\boldsymbol{s}_i \in \mathcal{D}_s : x_T(\boldsymbol{s}_i) > l\}. \tag{4.3}$$

This can be useful if some decision is to be made at time step T, so that the situation at that point in time is particularly important. A similar definition is used by French and Sain (2013), and this ES is also the focus in this thesis.

Another possibility is to map where the system value is ever above the critical limit. Discretised in time, this means that a position is part of the ES if the system value is above l at any time step $0, 1, \ldots, T$. Call this the maximal ES,

$$\text{ES}_{\text{max}} = \{\boldsymbol{s}_i \in \mathcal{D}_s : \max_{t=0,\ldots,T} x_t(\boldsymbol{s}_i) > l\}. \tag{4.4}$$

This approach is useful if we want to map the total area where values may be high, or if there are some positions that are especially sensitive to high values, and is somewhat similar to the motivation of Adler (2000). However, extreme value statistics is a whole field in itself, and computations with this ES is outside the scope of this thesis.
An ES is usually represented by a N-vector with ones for all $\boldsymbol{s}_i$ with $x(\boldsymbol{s}_i) > l$ and zeros otherwise. We may also write

$$\text{ES}(\boldsymbol{s}_i) = \begin{cases} 1 & x(\boldsymbol{s}_i) > l, \\ 0 & \text{otherwise.} \end{cases} \tag{4.5}$$

For the sample from the test example at time $T = 30$ in Figure 2.5b, the ES_T is shown in Figure 4.1a.

Since the true systems are not available in practice, it is useful to define the *excursion probability* (EP), which is the probability of the value in a position to be above the critical

limit. Let $\text{EP}_{t'}^{(t'')}(\boldsymbol{s}_i)$ be this probability for the state value in grid location $\boldsymbol{s}_i$ at time t' given measurements up to time step t''. That is,

$$\begin{aligned}\text{EP}_{t'}^{(t'')}(\boldsymbol{s}_i) &= P(X_{t'}(\boldsymbol{s}_i) > l \,|\, \boldsymbol{Y}_{1:t''}) \\ &= 1 - \Phi\left(\frac{l - \text{E}[X_{t'}(\boldsymbol{s}_i) \,|\, \boldsymbol{Y}_{1:t''}]}{\sqrt{\text{Var}(X_{t'}(\boldsymbol{s}_i) \,|\, \boldsymbol{Y}_{1:t''})}}\right),\end{aligned} \tag{4.6}$$

where $\Phi(\cdot)$ is the univariate standard Gaussian cumulative distribution function (CDF), as shown in Figure 4.2 and $P(\cdot)$ denotes probability. We always have $t' \geq t''$, and the

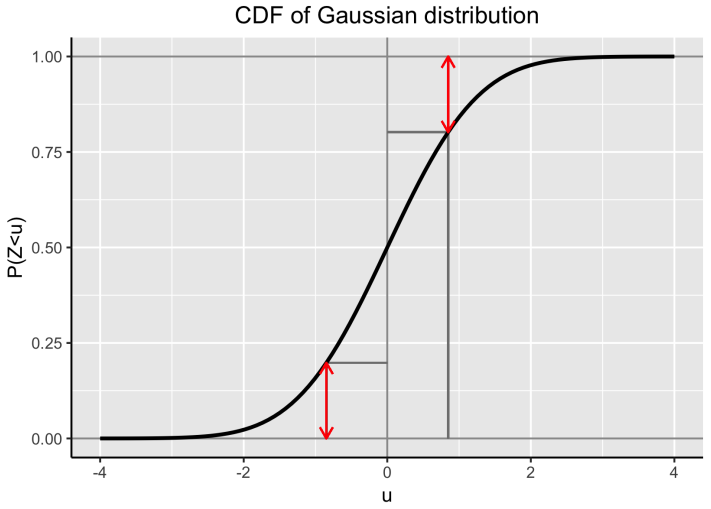


Figure 4.2: Univariate standard Gaussian CDF. The symmetry $1 - \Phi(u) = \Phi(-u)$ is demonstrated by red arrows.

dependencies are as shown in Figure 2.9 if we insert $t' = t, t'' = t - 1$. The set of these probabilities for all $\boldsymbol{s}_i \in \mathcal{D}_s$ is denoted $\text{EP}_{t'}^{(t'')}$. EPs will be used to predict ESs, by using a cutoff for classification at $\text{EP}_T^{(T)} = 0.5$. That is, if $\text{EP}_T^{(T)}(\boldsymbol{s}_i) \geq 0.5$, we predict $\hat{\text{ES}}_T(\boldsymbol{s}_i) = 1$. In Figure 4.1b, $\text{EP}_T^{(0)}$ is shown for the test example sample at time $T = 30$ in Figure 2.5b.

4.2 Decision-making strategies

Let a path denote the set of grid points that measurements are obtained from, sorted in the order they are visited in. In this thesis, the distances between consecutive grid nodes in the path are approximately equal, assuming the AUV travels at approximately constant speed and obtains measurements at uniform time intervals. From one position, the AUV can move in any direction, though not out of the domain. Consider the domain of Frænfjorden and assume the AUV moves 2.8-3.2 grid node distances, 56-64 m, in the time interval

between each measurement. Then from a position away from the domain boundaries, 16 positions can be reached for the next measurement, all located on an approximate circle around the current position. The grid and positions are shown in Figure 4.3. The number of possible positions is smaller close to the boundary.

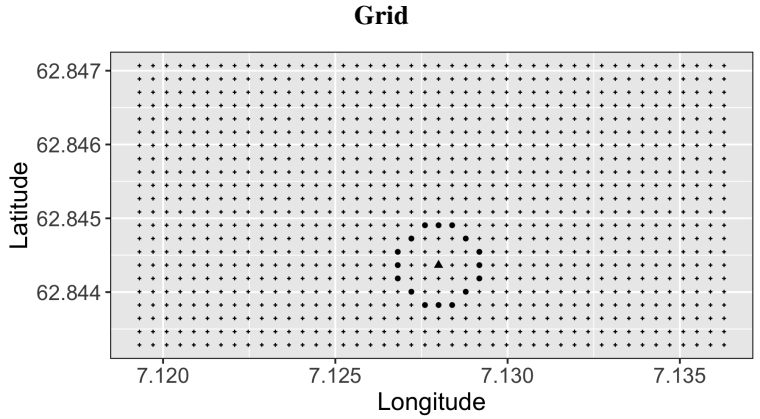


Figure 4.3: Grid showing the possible positions for an AUV to move. The current position is marked by a triangle, and the positions to consider are marked with circles.

Deciding where the AUV should move, depends on the final aim of the process, and decisions in one step affect the possible decisions and outcome in future steps. In general, all possible steps of the future must be taken into account to find the overall optimal solution in sequential decision making. This is not feasible in practice, as all possibilities grow exponentially for each step. The diagram of Figure 4.4 demonstrates this for three steps forward in time when the number of possible positions in each step is 2. In the situation of Figure 4.3, where there are 16 possible positions, the number of combinations grows extremely fast. It is therefore common to use *greedy*, or *myopic*, strategies, where the next position is chosen as though it is the last step. Only one step forward in Figure 4.4 is then seen, as though the graph is cut after time $t + 1$.

In the decision strategy, some sort of criterion or objective function is needed to evaluate the possible decisions. Different types of objective functions have been studied for assessing the quality of obtaining different information, the *value of information*, as called by Bhattacharjya et al. (2013) and Eidsvik et al. (2015). In this thesis, the value of letting the AUV visit one position compared to another needs assessment. Fossum et al. (2018) use a combination of variance reduction and temperature gradient intensity as objective function for deciding where to send AUVs. The aim is often to reduce uncertainty of some sort (Azzimonti et al., 2015; Bhattacharjya et al., 2013; Chevalier et al., 2013), and Bect et al. (2012) suggest criteria for minimising both the variance and the probability of misclassification with respect to the ES. Chevalier et al. (2013) later use the expected variance reduction of EP as objective function. The focus in this thesis is on reducing the probability of misclassification.

In the following section, a new myopic decision criterion based on the expected mean misclassification probability (EMMP) is derived. Using the EMMP as decision criterion,

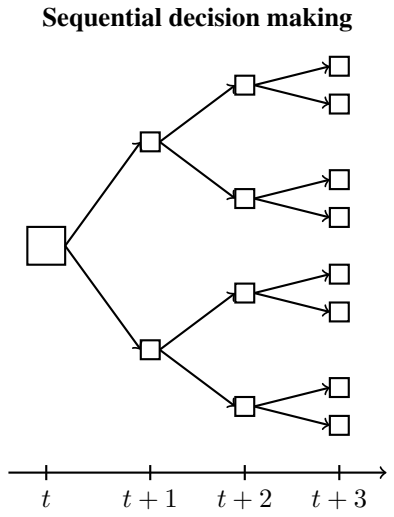


Figure 4.4: Number of possible decisions increases exponentially for each increased number of future steps to consider. This diagram shows the development when the number of positions to consider in each step is 2 and at 3 steps forward in time the number of possible combinations is 8, i.e. the number of nodes at position $t + 3$.

the path can be made adaptively as the measurements are obtained. When data at one position is obtained, the EMMP for all possible next positions should be computed, and the one with the lowest EMMP chosen. The chosen positions become part of the path. This procedure is repeated until some stopping criterion is reached. In this thesis, the focus is on time based operations, meaning that the operation is stopped after some fixed time. The EMMP criterion for adaptive sampling is the main strategy in this thesis, but a few others are also considered.

One simple adaptive approach is to choose the position with the current EP_t^{t-1} (equation (4.6)) closest to 0.5, so that data is obtained from the position with the highest misclassification probability at time t given all data up to time $t-1$, $\boldsymbol{Y}_{1:t-1}$. The most naive adaptive strategy is to choose positions at random. This makes out a good reference case for comparison, as all strategy should at least be better than randomly drawing a position among the possibilities for each step. It is also possible to choose all positions before any measurements are obtained. Some such predetermined paths are made to compare with the adaptive ones. In a later section in this chapter, a hybrid criterion partly based on the EMMP criterion is also discussed.

4.2.1 Expected mean misclassification probability

It is desired to obtain a state model that predicts the ES at time step T as well as possible. Optimally, we would want to choose measurement positions so that the $\text{EP}_T^{(T)}$ is the best possible approximation of the of the ES_T. A criterion for where to draw measurements based on the probability of misclassification is now derived. In chapter 2, the number of

new data points at time t is m, i.e. the data model is given for the random vector $\boldsymbol{Y}_t$. This notation would also be valid in the following derivation. However, the decision strategy is to be used in adaptive decision making, and in this thesis only one position will be chosen for each time step. Therefore, $m = 1$ in the following, and the data is given by the random variable Y. Further, the decision is made time step by time step, and the time is fixed for one set of calculations. To simplify notations somewhat, the time subscripts are not included during derivation.

Given some state vector $\boldsymbol{X} = (X(\boldsymbol{s}_1), \dots X(\boldsymbol{s}_N))^\intercal$, one could predict an ES based on the EP for $\boldsymbol{X}$. Assume classification cutoff at 0.5,we want to push the EPs towards 0 or 1 to reduce the probability of misclassification. We call $\boldsymbol{X}$ the *target state vector* if it is used to predict ESs in this way, and its distribution the *target distribution*. Let the grid node in which to obtain the next measurement be identified by d, i.e. the measurement is obtained in position $\boldsymbol{s}_d$. The random variable for data in $\boldsymbol{s}_d$ is denoted Y_d and one realisation of Y_d is y_d. We are interested in the distribution of $\boldsymbol{X}$ given Y_d. More specific, we want to choose d such that $P(X(\boldsymbol{s}_i) > l \,|\, Y_d)$ is as close as possible to 0 or 1 for all $\boldsymbol{s}_i \in \mathcal{D}_s$. Consider $\min\{1 - P(X(\boldsymbol{s}_i) > l \,|\, Y_d), P(X(\boldsymbol{s}_i) > l \,|\, Y_d)\}$. This is the probability of misclassification in grid node $\boldsymbol{s}_i$ at cutoff 0.5. From this we can define the mean misclassification probability (MMP) for the ES with the target distribution of $\boldsymbol{X}$,

$$\text{MMP}_{y_d} = \frac{1}{N} \sum_{i=1}^{N} \min\{1 - P(X(\boldsymbol{s}_i) > l \,|\, Y_d), P(X(\boldsymbol{s}_i) > l \,|\, Y_d)\}. \tag{4.7}$$

As before, let $\boldsymbol{X} \,|\, Y_d$ denote the random vector following the distribution of $\boldsymbol{X}$ given Y_d. Assume $\boldsymbol{X} \,|\, Y_d \sim N_N(\boldsymbol{\mu}_{x \,|\, y_d}, \boldsymbol{\Sigma}_{x \,|\, y_d})$ with $\boldsymbol{\mu}_{x \,|\, y_d} = (\mu_{x \,|\, y_d 1}, \dots \mu_{x \,|\, y_d N})^\intercal$ and $\text{diag}(\boldsymbol{\Sigma}_{x \,|\, y_d}) = (\sigma^2_{x \,|\, y_d 1}, \dots \sigma^2_{x \,|\, y_d N})^\intercal$. Then

$$\begin{aligned} \text{MMP}_{y_d} &= \frac{1}{N} \sum_{i=1}^{N} \min \left\{ \Phi \left(\frac{l - \mu_{x \,|\, y_d i}}{\sigma_{x \,|\, y_d i}} \right), 1 - \Phi \left(\frac{l - \mu_{x \,|\, y_d i}}{\sigma_{x \,|\, y_d i}} \right) \right\} \\ &= \frac{1}{N} \sum_{i=1}^{N} \text{MP}_{y_d i}, \end{aligned} \tag{4.8}$$

where $\text{MP}_{y_d i}$ is the misclassification probability in grid node $\boldsymbol{s}_i$ if one measures the value y_d in position $\boldsymbol{s}_d$.

We want to choose d such that the MMP is as small as possible. However, we cannot evaluate the MMP, since the data at position $\boldsymbol{s}_d$, y_d, is not available at the decision time. Instead, we calculate the expected MMP, EMMP, over all possible values Y_d. Then different positions can be compared based on the expected value. For a fixed d, we need

$$\text{EMMP}_d = \text{E}_{Y_d}[\text{MMP}_{y_d}] = \frac{1}{N} \sum_{i=1}^{N} \text{E}_{Y_d}[\text{MP}_{y_d i}] = \frac{1}{N} \sum_{i=1}^{N} \int_{y_d} \text{MP}_{y_d i} p(y_d) dy_d, \tag{4.9}$$

with $p(y_d)$ the probability density function (PDF) of Y_d. Since we only consider obtaining new data in one position per time step, the integral is one-dimensional. Nevertheless, some observations and simplifications are convenient.

First, define a new univariate variable for grid point $\boldsymbol{s}_i$ as follows:

$$U_{di} = \frac{l - \mu_{x \mid y_d i}}{\sigma_{x \mid y_d i}}. \tag{4.10}$$

To calculate the expected value and variance of U_{di}, the distribution of $\boldsymbol{X}$ is needed. We will see, however, when considering specific target states, that $\sigma_{x \mid y_d i}$ really only depends on position $\boldsymbol{s}_d$, not the value y_d. Further, $\mu_{x \mid y_d i}$ is linear in Y_d. With d fixed, we then have

$$\mathrm{E}[U_{di}] = \mu_{u_{di}} = \frac{l - \mathrm{E}[\mu_{x \mid y_d i}]}{\sigma_{x \mid y_d i}}, \tag{4.11}$$

$$\mathrm{Var}(U_{di}) = \sigma^2_{u_{di}} = \frac{1}{\sigma^2_{x \mid y_d i}} \mathrm{Var}(\mu_{x \mid y_d i}). \tag{4.12}$$

Thus, the distribution of U_{di} does not depend on the actual value of Y_d. Since U_{di} is linear in Y_d, which is a Gaussian distributed random variable, U_{di} is also Gaussian distributed:

$$U_{di} \sim N(\mu_{u_{di}}, \sigma^2_{u_{di}}). \tag{4.13}$$

Now observe that $\mathrm{MP}_{y_d i}$ (equation (4.8)) depends only on the random variable U_{di}. Therefore, the expected value in equation (4.9) is really only the mean of expected values with respect to the distribution of variables $U_{di}, i = 1, ...N$.

$$\frac{1}{N} \sum_{i=1}^{N} \mathrm{E}_{Y_d}[\mathrm{MP}_{y_d i}] = \frac{1}{N} \sum_{i=1}^{N} \mathrm{E}_{U_{di}}[\mathrm{MP}_{y_d i}]. \tag{4.14}$$

That is, each element i of EMMP_d is the integral over all possible values of U_{di}. With this knowledge, inserting $\mathrm{MP}_{y_d i}$ from equation (4.8) into equation (4.9) leaves the problem

$$\begin{aligned} \mathrm{EMMP}_d &= \frac{1}{N} \sum_{i=1}^{N} \int_{y_d} \min\left\{\Phi(u_{di}), 1 - \Phi(u_{di})\right\} p(y_d) dy_d \\ &= \frac{1}{N} \sum_{i=1}^{N} \int_{u_{di}} \min\left\{\Phi(u_{di}), 1 - \Phi(u_{di})\right\} p(u_{di}) du_{di} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{EMP}_{di}. \end{aligned} \tag{4.15}$$

With the same notation as before, EMP_{di} is the expected misclassification probability in position $\boldsymbol{s}_i$ if one chooses to measure in position $\boldsymbol{s}_d$. It does not depend on the actual measured value in this position, and can therefore be evaluated prior to obtaining the measurement.

For convenience, EMP_{di}, U_{di} and u_{di} are in the following denoted by EMP_i, U_i and u_i. The problem to be solved is

$$\mathrm{EMP}_i = \int_{-\infty}^{\infty} \min\left\{\Phi(u_i), 1 - \Phi(u_i)\right\} p(u_i) du_i. \tag{4.16}$$

Because of the symmetry of Φ, as shown in Figure 4.2, $1 - \Phi(u_i) = \Phi(-u_i)$. Also, Φ is monotonically increasing, so $\Phi(U_i) \leq \Phi(-U_i)$ whenever $U_i \leq -U_i$. This means,

$\min\{\Phi(U_i), \Phi(-U_i)\}$ equals $\Phi(U_i)$ when $U_i < 0$ and $\Phi(-U_i)$ when $U_i > 0$. Thus, equation (4.16) can be reformulated into

$$\text{EMP}_i = \int_{-\infty}^{0} \Phi(u_i)p(u_i)du_i + \int_{0}^{\infty} \Phi(-u_i)p(u_i)du_i. \tag{4.17}$$

Let Z be a random variable with standard Gaussian distribution. Then

$$\Phi(u_i) = P(Z < u_i) = \int_{-\infty}^{u_i} p(z)dz, \tag{4.18}$$

and

$$\int_{-\infty}^{0} \Phi(u_i)p(u_i)du_i = \int_{-\infty}^{0}\int_{-\infty}^{u_i} p(z)p(u_i)dzdu_i = P(Z < U_i, U_i < 0). \tag{4.19}$$

Observe that Z and U_i are chosen to be independent, so that $p(z)p(u_i)$ is their joint PDF. By the same argument for the second integral in equation (4.17), it is obtained that

$$\begin{aligned}\text{EMP}_i &= P(Z < U_i, U_i < 0) + P(Z < -U_i, U_i > 0)\\ &= P(Z - U_i < 0, U_i < 0) + P(Z + U_i < 0, -U_i < 0).\end{aligned} \tag{4.20}$$

Define

$$\begin{aligned} V_i &= Z - U_i, & A_i &= Z + U_i,\\ W_i &= U_i, & B_i &= -U_i.\end{aligned}$$

Then

$$\text{EMP}_i = P(V_i < 0, W_i < 0) + P(A_i < 0, B_i < 0). \tag{4.21}$$

Here, A_i is unrelated to the propagator matrix in the process model of section 2.2.1. The joint distribution of V_i and W_i can be obtained by finding the expected values, variances and the covariance between the variables. As they are both linear combinations of univariate Gaussian distributed variables, they jointly follow a bivariate Gaussian distribution. We have $\mathrm{E}[V_i] = \mathrm{E}[Z] - \mathrm{E}[U_i] = 0 - \mu_{u_i} = -\mu_{u_i}$, $\mathrm{Var}(V_i) = \mathrm{Var}(Z) + \mathrm{Var}(U_i) = 1 + \sigma^2_{u_i}$, $\mathrm{E}[W_i] = \mu_{u_i}$, $\mathrm{Var}(W_i) = \sigma^2_{u_i}$ and $\mathrm{Cov}(V_i, W_i) = \mathrm{Cov}(Z - U_i, U_i) = \mathrm{Cov}(Z, U_i) - \mathrm{Var}(U_i) = 0 - \sigma^2_{u_i} = -\sigma^2_{u_i}$. Thus,

$$\begin{bmatrix} V_i \\ W_i \end{bmatrix} \sim N_2\left(\begin{bmatrix} -\mu_{u_i} \\ \mu_{u_i} \end{bmatrix}, \begin{bmatrix} \sigma^2_{u_i} + 1 & -\sigma^2_{u_i} \\ -\sigma^2_{u_i} & \sigma^2_{u_i} \end{bmatrix}\right) = N_2(\boldsymbol{\mu_i^*}, \boldsymbol{\Sigma_i^*}), \tag{4.22}$$

and similarly,

$$\begin{bmatrix} A_i \\ B_i \end{bmatrix} \sim N_2(-\boldsymbol{\mu_i^*}, \boldsymbol{\Sigma_i^*}). \tag{4.23}$$

Further,

$$P(V_i < v_i, W_i < w_i) = \Phi_2\left(\begin{bmatrix} v_i \\ w_i \end{bmatrix}; \boldsymbol{\mu_i^*}, \boldsymbol{\Sigma_i^*}\right), \tag{4.24}$$

with $\Phi_2(\cdot; \boldsymbol{\mu}_{\boldsymbol{i}}^*, \boldsymbol{\Sigma}_{\boldsymbol{i}}^*)$ the bivariate Gaussian CDF with expected value $\boldsymbol{\mu}_{\boldsymbol{i}}^*$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{i}}^*$. Then

$$\text{EMP}_i = \Phi_2\left(\mathbf{0}; \boldsymbol{\mu}_{\boldsymbol{i}}^*, \boldsymbol{\Sigma}_{\boldsymbol{i}}^*\right) + \Phi_2\left(\mathbf{0}; -\boldsymbol{\mu}_{\boldsymbol{i}}^*, \boldsymbol{\Sigma}_{\boldsymbol{i}}^*\right), \tag{4.25}$$

where $\mathbf{0} = (0, 0)^\intercal$.

For computational reasons using statistical software, it is convenient to have the bivariate distribution on standard form, with means 0 and marginal variances 1. We transform V_i, W_i, A_i and B_i so they each marginally are univariate standard Gaussian. For V_i and W_i we obtain

$$\begin{aligned}
\tilde{V}_i &= \frac{V_i - (-\mu_{u_i})}{\sqrt{\sigma_{u_i}^2 + 1}} \sim N(0,1), \\
\tilde{W}_i &= \frac{W_i - \mu_{u_i}}{\sigma_{u_i}} \sim N(0,1), \\
\text{Cov}(\tilde{V}_i, \tilde{W}_i) &= -\frac{\sigma_{u_i}}{\sqrt{\sigma_{u_i}^2 + 1}}, \\
P(V_i < 0, W_i < 0) &= P\left(\tilde{V}_i < \frac{0 - (-\mu_{u_i})}{\sqrt{\sigma_{u_i}^2 + 1}}, \tilde{W}_i < \frac{0 - \mu_{u_i}}{\sigma_{u_i}}\right),
\end{aligned} \tag{4.26}$$

and likewise for A_i and B_i. Define

$$\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{i}} = \begin{bmatrix} 1 & \tilde{\rho} \\ \tilde{\rho} & 1 \end{bmatrix} \quad \text{and} \quad \tilde{\boldsymbol{z}}_{\boldsymbol{i}} = \begin{bmatrix} \frac{\mu_{u_i}}{\sqrt{\sigma_{u_i}^2 + 1}} \\ \frac{-\mu_{u_i}}{\sigma_{u_i}} \end{bmatrix} \tag{4.27}$$

with $\tilde{\rho} = -\frac{\sigma_{u_i}}{\sqrt{\sigma_{u_i}^2 + 1}}$. Then finally

$$\text{EMP}_i = \Phi_2\left(\tilde{\boldsymbol{z}}_{\boldsymbol{i}}; \mathbf{0}, \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{i}}\right) + \Phi_2\left(-\tilde{\boldsymbol{z}}_{\boldsymbol{i}}; \mathbf{0}, \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{i}}\right). \tag{4.28}$$

Now reintroduce the notation with d to emphasise that $\tilde{\boldsymbol{z}}_{\boldsymbol{i}} = \tilde{\boldsymbol{z}}_{\boldsymbol{di}}$ and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{i}} = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{di}}$ depend on the choice of position $\boldsymbol{s}_d$ through $U_i = U_{di}$. Insert equation (4.28) into equation (4.15) to obtain

$$\text{EMMP}_d = \frac{1}{N} \sum_{i=1}^{N} \left(\Phi_2\left(\tilde{\boldsymbol{z}}_{\boldsymbol{i}}; \mathbf{0}, \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{i}}\right) + \Phi_2\left(-\tilde{\boldsymbol{z}}_{\boldsymbol{i}}; \mathbf{0}, \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{i}}\right)\right). \tag{4.29}$$

We have now derived a general criterion which can be used to evaluate different possible measurement positions. The lower EMMP_d, the better is the position considered to be. The aim is to use the criterion to decide where to draw the *next* measurement. That is, the EMMP criterion should be used dynamically to choose position for one time step at a time. To compute EMMP_d, the target state vector $\boldsymbol{X}$ has to be defined and from there the distribution of U_{di} must be derived. In the next sections, some possible target models are defined, and the corresponding distributions derived.

4.2.2 EMMP with forecasted distributions

As mentioned in section 4.1, the aim in this thesis is to predict an ES as accurately as possible, and the focus is on fixed ES for time T (equation (4.3)). Because of this, we use the state vector at time T, $\boldsymbol{X}_T$, as our target state. Further, we would optimally use the distribution of this given the measured values for all time steps, $\boldsymbol{Y}_{1:T}$. This means that we would want to choose the position for time step t such that $\text{EP}_T^{(T)}(\boldsymbol{s}_i) = P(X_T(\boldsymbol{s}_i) > l \,|\, \boldsymbol{Y}_{1:T}), i = 1, \dots, N$ (equation (4.6)) is as close as possible to 0 or 1. For choosing grid node for step t, we should then consider all possible positions and measurement values for steps $t+1, \dots, T$ as well. This is computationally infeasible, as the possible positions and values to consider will be immense, as seen in Figure 4.4.

Instead we focus on a heuristic, myopic method, meaning that we choose a local optimum at time step t considering only the data for the next measurement when choosing position. Let the target state at time step t be $\boldsymbol{X}_T$, but now condition only on measurements for time steps $1, \dots, t$, $\boldsymbol{Y}_{1:t}$. At time t, the measurements for $t = 1, \dots t-1$, $\boldsymbol{Y}_{1:t-1}$, are already obtained. We know the distribution of $\boldsymbol{X}_T \,|\, \boldsymbol{Y}_{1:t}$ is Gaussian, and the only unknown part is Y_t. The analytic derivation of the base criterion from section 4.2.1 can be used to evaluate different grid node candidates for the measurement at time step t. To calculate EMMP_d for grid node $\boldsymbol{s}_d$ we need the mean and variance of variable U_{di} in equation (4.10) and for that we need the mean and variance of all elements of the target vector, up to the unknown value in position $\boldsymbol{s}_d$ at time step t, y_{td}.

At time step t with data from all time steps up to $t-1$ available, the distribution of $\boldsymbol{X}_t \,|\, \boldsymbol{Y}_{1:t-1}$ is known, and reflects current knowledge of the state. From there we can take $T-t$ prediction steps to obtain the distribution of $\boldsymbol{X}_T \,|\, \boldsymbol{Y}_{1:t-1}$. To find the distribution of $\boldsymbol{X}_T \,|\, \boldsymbol{Y}_{1:t}$, we must take one filtering step from $\boldsymbol{X}_t \,|\, \boldsymbol{Y}_{1:t-1}$ to find $\boldsymbol{X}_t \,|\, \boldsymbol{Y}_{1:t}$ and then $T-t$ prediction steps to get to the distribution of $\boldsymbol{X}_T \,|\, \boldsymbol{Y}_{1:t}$. The procedure is shown schematically in Figure 4.5.

Let the notation of mean and variance be as in section 2.2.2. Then $\boldsymbol{\mu}_{t\,|\,t-1}$ and $\boldsymbol{P}_{t\,|\,t-1}$ are known, and $\boldsymbol{\mu}_{t\,|\,t}$ and $\boldsymbol{P}_{t\,|\,t}$ are given by the filtering equations (2.48) and (2.49). Note that $\boldsymbol{\mu}_{t\,|\,t}$ depends on the measured value at time t, while $\boldsymbol{P}_{t\,|\,t}$ does not. Both depend on d, but the dependency is dropped in the notation for readability. We let the propagator matrix $\boldsymbol{A}$, the source vector $\boldsymbol{R}$ and the covariance matrix of the innovation term $\boldsymbol{\eta}$ of the process model (equation (2.31)) be constant for all time steps, but add the time step subscript to keep track of the predictive steps. Performing $T-t$ prediction steps, like in equations (2.42) and (2.43), results in

$$\begin{aligned}\boldsymbol{\mu}_{T\,|\,t} &= \boldsymbol{A}_T \dots (\boldsymbol{A}_{t+2}(\boldsymbol{A}_{t+1}\boldsymbol{\mu}_{t\,|\,t} + \boldsymbol{R}_{t+1}) + \boldsymbol{R}_{t+2}) \cdots + \boldsymbol{R}_T \\ &= \boldsymbol{A}^{T-t}\boldsymbol{\mu}_{t\,|\,t} + (\boldsymbol{A}^{T-t-1} + \boldsymbol{A}^{T-t-2} + \cdots + \boldsymbol{A} + 1)\boldsymbol{R},\end{aligned} \tag{4.30}$$

$$\begin{aligned}\boldsymbol{P}_{T\,|\,t} &= \boldsymbol{A}_T(\dots(\boldsymbol{A}_{t+2}(\boldsymbol{A}_{t+1}\boldsymbol{P}_{t\,|\,t}\boldsymbol{A}_{t+1}^\intercal + \boldsymbol{Q}_{t+1})\boldsymbol{A}_{t+2}^\intercal + \boldsymbol{Q}_{t+2})\dots)\boldsymbol{A}_T^\intercal + \boldsymbol{Q}_T \\ &= \boldsymbol{A}(\dots(\boldsymbol{A}(\boldsymbol{A}\boldsymbol{P}_{t\,|\,t}\boldsymbol{A}^\intercal + \boldsymbol{Q})\boldsymbol{A}^\intercal + \boldsymbol{Q})\dots)\boldsymbol{A}^\intercal + \boldsymbol{Q}.\end{aligned} \tag{4.31}$$

Also note that

$$\boldsymbol{\mu}_{T\,|\,t-1} = \boldsymbol{A}^{T-t}\boldsymbol{\mu}_{t\,|\,t-1} + (\boldsymbol{A}^{T-t-1} + \boldsymbol{A}^{T-t-2} + \cdots + \boldsymbol{A} + 1)\boldsymbol{R}, \tag{4.32}$$

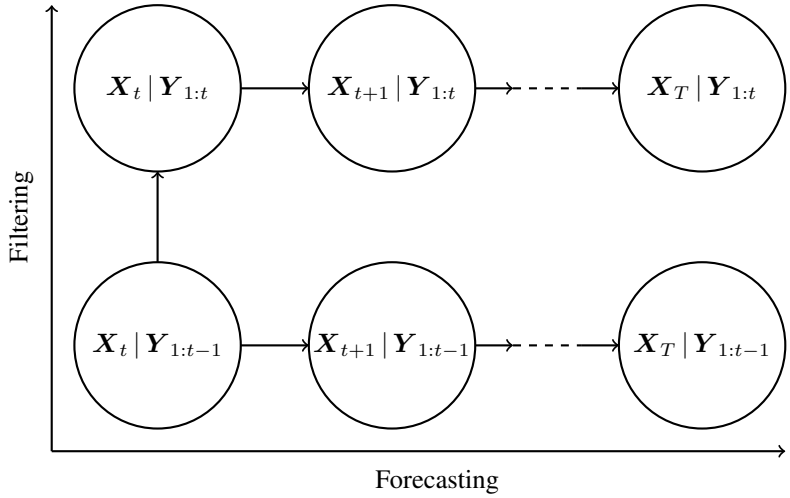


Figure 4.5: Schematic drawing of Kalman filter forecasting and filtering to reach the distribution of $\boldsymbol{X}$ given data up to time $t-1$ or time t.

by same calculations.

The random variable U_{di} from equation (4.10) for target state vector $\boldsymbol{X}_T$, conditioned on $\boldsymbol{Y}_{1:t}$ is

$$U_{di} = \frac{l - \boldsymbol{\mu}_{T\,|\,t}[i]}{\sqrt{\boldsymbol{P}_{T\,|\,t}[i,i]}}. \tag{4.33}$$

We need to calculate the mean and variance of U_{di}. At time step t, all measurements up to time $t-1$ are known, and all calculations are therefore really given $\boldsymbol{Y}_{1:t-1}$. As seen in equations (4.11) and (4.12), finding the distribution of U_{di} given $\boldsymbol{Y}_{1:t-1}$ involves computing the mean and variance for $\boldsymbol{\mu}_{T\,|\,t}$, and since $\boldsymbol{\mu}_{T\,|\,t}$ is a function of $\boldsymbol{\mu}_{t\,|\,t}$ (equation (4.30)), the same must be calculated for the latter. Note that for these computations, $\boldsymbol{\mu}_{t\,|\,t}$ and $\boldsymbol{\mu}_{T\,|\,t}$ must be considered random variables, and in the expression for $\boldsymbol{\mu}_{t\,|\,t}$ in equation (2.48), the observed $\boldsymbol{y}_t$ should be interchanged with the random variable $\boldsymbol{Y}_t$. The notation $\mathrm{E}[\boldsymbol{\mu} \mid \boldsymbol{Y}]$ can be seen as the expected value of a random vector $\boldsymbol{\mu} \mid \boldsymbol{Y}$, or as $\mathrm{E}[\boldsymbol{\mu}]$ when $\boldsymbol{Y}$ is given. In this setting, the latter is more convenient.

From equations (2.44) and (2.45), $Y_{td} \mid \boldsymbol{Y}_{1:t-1} \sim N(\boldsymbol{H}_{td}\boldsymbol{\mu}_{t\,|\,t-1}, S_{td})$, and in the one-dimensional data case, $S_{td} = \boldsymbol{H}_{td}\boldsymbol{P}_{t\,|\,t-1}\boldsymbol{H}_{td}^{\intercal} + \tau^2$. The d in the notation is added to emphasise dependence on the position $\boldsymbol{s}_d$. Likewise, the Kalman gain in equations (2.48) and (2.49) is in the following denoted $\boldsymbol{K}_{td}$. Using these equations and the distribution of $Y_{td} \mid \boldsymbol{Y}_{1:t-1}$, we find

$$\begin{aligned}
\mathrm{E}[\boldsymbol{\mu}_{t\,|\,t} \,|\, \boldsymbol{Y}_{1:t-1}] &= \mathrm{E}[\boldsymbol{\mu}_{t\,|\,t-1} + \boldsymbol{K}_{td}(Y_{td} \,|\, \boldsymbol{Y}_{1:t-1} - \boldsymbol{H}_{td}\boldsymbol{\mu}_{t\,|\,t-1})] \\
&= \boldsymbol{\mu}_{t\,|\,t-1} + \boldsymbol{K}_{td}(\mathrm{E}[Y_{td} \,|\, \boldsymbol{Y}_{1:t-1}] - \boldsymbol{H}_{td}\boldsymbol{\mu}_{t\,|\,t-1}) \\
&= \boldsymbol{\mu}_{t\,|\,t-1} + \boldsymbol{K}_{td}(\boldsymbol{H}_{td}\boldsymbol{\mu}_{t\,|\,t-1} - \boldsymbol{H}_{td}\boldsymbol{\mu}_{t\,|\,t-1}) \\
&= \boldsymbol{\mu}_{t\,|\,t-1},
\end{aligned} \tag{4.34}$$

$$\begin{aligned}
\mathrm{Cov}(\boldsymbol{\mu}_{t\,|\,t} \,|\, \boldsymbol{Y}_{1:t-1}) &= \mathrm{Cov}(\boldsymbol{\mu}_{t\,|\,t-1} + \boldsymbol{K}_{td}(Y_{td} \,|\, \boldsymbol{Y}_{1:t-1} - \boldsymbol{H}_{td}\boldsymbol{\mu}_{t\,|\,t-1})) \\
&= \boldsymbol{K}_{td}\mathrm{Var}(Y_{td} \,|\, \boldsymbol{Y}_{1:t-1})\boldsymbol{K}_{td}^{\intercal} = \boldsymbol{K}_{td}S_{td}\boldsymbol{K}_{td}^{\intercal}.
\end{aligned} \tag{4.35}$$

Then we also obtain

$$\begin{aligned}
\mathrm{E}[\boldsymbol{\mu}_{T\,|\,t} \,|\, \boldsymbol{Y}_{1:t-1}] &= \boldsymbol{A}^{T-t}\mathrm{E}[\boldsymbol{\mu}_{t\,|\,t} \,|\, \boldsymbol{Y}_{1:t-1}] \\
&\quad + (\boldsymbol{A}^{T-t-1} + \boldsymbol{A}^{T-t-2} + \cdots + \boldsymbol{A} + 1)\boldsymbol{R} \\
&= \boldsymbol{A}^{T-t}\boldsymbol{\mu}_{t\,|\,t-1} + (\boldsymbol{A}^{T-t-1} + \boldsymbol{A}^{T-t-2} + \cdots + \boldsymbol{A} + 1)\boldsymbol{R} \\
&= \boldsymbol{\mu}_{T\,|\,t-1},
\end{aligned} \tag{4.36}$$

$$\begin{aligned}
\mathrm{Cov}(\boldsymbol{\mu}_{T\,|\,t} \,|\, \boldsymbol{Y}_{1:t-1}) &= \boldsymbol{A}^{T-t}\mathrm{Cov}(\boldsymbol{\mu}_{t\,|\,t} \,|\, \boldsymbol{Y}_{1:t-1})(\boldsymbol{A}^{T-t})^{\intercal} \\
&= \boldsymbol{A}^{T-t}\boldsymbol{K}_{td}S_{td}\boldsymbol{K}_{td}^{\intercal}(\boldsymbol{A}^{T-t})^{\intercal}.
\end{aligned} \tag{4.37}$$

Finally, we can find the mean and variance of $U_{di} \,|\, \boldsymbol{Y}_{1:t-1}$:

$$\mathrm{E}[U_{di} \,|\, \boldsymbol{Y}_{1:t-1}] = \frac{l - \mathrm{E}[\boldsymbol{\mu}_{T\,|\,t}[i] \,|\, \boldsymbol{Y}_{1:t-1}]}{\sqrt{\boldsymbol{P}_{T\,|\,t}[i,i]}} = \frac{l - \boldsymbol{\mu}_{T\,|\,t-1}[i]}{\sqrt{\boldsymbol{P}_{T\,|\,t}[i,i]}}, \tag{4.38}$$

$$\begin{aligned}
\mathrm{Var}(U_{di} \,|\, \boldsymbol{Y}_{1:t-1}) &= \frac{1}{\boldsymbol{P}_{T\,|\,t}[i,i]}\mathrm{Var}(\boldsymbol{\mu}_{T\,|\,t}[i] \,|\, \boldsymbol{Y}_{1:t-1}) \\
&= \frac{1}{\boldsymbol{P}_{T\,|\,t}[i,i]}\boldsymbol{A}^{T-t}[i,]\boldsymbol{K}_{td}S_{td}\boldsymbol{K}_{td}^{\intercal}(\boldsymbol{A}^{T-t}[i,])^{\intercal},
\end{aligned} \tag{4.39}$$

where $[i,]$ indicates row i of the matrix. From here, the calculation of the criterion is given in section 4.2.1.

If we want to consider the same target, but with the spatial AR(1) model instead of the SPDE model, the formulas are somewhat simplified. Equation (4.30) becomes

$$\boldsymbol{\mu}_{T\,|\,t} = \phi^{T-t}\boldsymbol{\mu}_{t\,|\,t} + \sum_{k=0}^{T-t-1} \phi^k(1-\phi)\boldsymbol{\mu}_0, \tag{4.40}$$

with $\boldsymbol{\mu}_{t\,|\,t}$ as in equation (2.48), and equation (4.31) becomes

$$\boldsymbol{P}_{T\,|\,t} = \phi^{2(T-t)}\boldsymbol{P}_{t\,|\,t} + \sum_{k=0}^{T-t-1} \phi^{2k}(1-\phi^2)\boldsymbol{\Sigma}_0. \tag{4.41}$$

Likewise we get

$$\mathrm{Cov}(\boldsymbol{\mu}_{T\,|\,t} \,|\, \boldsymbol{Y}_{1:t-1}) = \phi^{2(T-t)}\mathrm{Cov}(\boldsymbol{\mu}_{t\,|\,t} \,|\, \boldsymbol{Y}_{1:t-1}), \tag{4.42}$$

with $\mathrm{Cov}(\boldsymbol{\mu}_{t\,|\,t} \,|\, \boldsymbol{Y}_{1:t-1})$ and the rest of the criterion like before.

4.2.3 EMMP with other target states

We again consider the SPDE model, but look at some simpler target distributions. It might be interesting to investigate how important it is to focus on the predictive distribution of $\boldsymbol{X}_T$. What if we only try to optimise the current state, that is use $\boldsymbol{X}_t$ as target state? How much does this affect the final approximation of ES_T? By doing this, we skip all the forecasting steps of equations (4.30) and (4.31). A lot of computation time and possible storage capacity are saved in this approach. We still condition on $\boldsymbol{Y}_{1:t}$ in the criterion, and $\boldsymbol{Y}_{1:t-1}$ is given. In this case,

$$U_{di} = \frac{l - \boldsymbol{\mu}_{t\,|\,t}[i]}{\sqrt{\boldsymbol{P}_{t\,|\,t}[i,i]}}, \tag{4.43}$$

where $\boldsymbol{P}_{t\,|\,t}$ is given in equation (2.49). We already computed the mean and variance of $\boldsymbol{\mu}_{t\,|\,t}$ given $\boldsymbol{Y}_{1:t-1}$ in equations (4.34) and (4.35), and we get

$$\text{E}[U_{di}\,|\,\boldsymbol{Y}_{1:t-1}] = \frac{l - \boldsymbol{\mu}_{t\,|\,t-1}[i]}{\sqrt{\boldsymbol{P}_{t\,|\,t}[i,i]}}, \tag{4.44}$$

$$\text{Var}(U_{di}\,|\,\boldsymbol{Y}_{1:t-1}) = \frac{1}{\boldsymbol{P}_{t\,|\,t}[i,i]}(\boldsymbol{K}_{td}S_{td}\boldsymbol{K}_{td}^\intercal)[i,i]. \tag{4.45}$$

We can imagine an even simpler case where we assume a pure spatial state in our criterion calculations. Let the initial state model $\boldsymbol{X}_0$ be the prior distribution, and assume no time dynamics. Then the model is only updated due to new data. Let us denote the model given the $t-1$ first measurements $\boldsymbol{X}_0\,|\,\boldsymbol{Y}_{1:t-1}$, though the subscript 0 does no longer indicates time step 0, but rather that this is the updated initial state model. Then let the target state be $\boldsymbol{X}_0$, and condition on $\boldsymbol{Y}_{1:t}$. The mean and covariance matrix of $\boldsymbol{X}_0\,|\,\boldsymbol{Y}_{1:t-1} \sim N_n(\boldsymbol{\mu}_{0\,|\,t-1}, \boldsymbol{P}_{0\,|\,t-1})$ are known. To find the mean and variance of the variable

$$U_{di} = \frac{l - \boldsymbol{\mu}_{0\,|\,t}[i]}{\sqrt{\boldsymbol{P}_{0\,|\,t}[i,i]}} \tag{4.46}$$

given $\boldsymbol{Y}_{1:t-1}$, we can use the filtering equations from section 2.2.2, or equivalently the equations for the posterior distribution of section 2.1.2. Let $S_{td} = \boldsymbol{H}_{td}\boldsymbol{P}_{0\,|\,t-1}\boldsymbol{H}_{td}^\intercal + \tau^2$. Then $Y_t\,|\,\boldsymbol{Y}_{1:t-1} \sim N(\boldsymbol{H}_{td}\boldsymbol{\mu}_{0\,|\,t-1}, S_{td})$ and

$$\boldsymbol{\mu}_{0\,|\,t} = \boldsymbol{\mu}_{0\,|\,t-1} + \boldsymbol{P}_{0\,|\,t-1}\boldsymbol{H}_{td}^\intercal S_{td}^{-1}(Y_{td} - \boldsymbol{H}_{td}\boldsymbol{\mu}_{0\,|\,t-1}). \tag{4.47}$$

We, analogously to the previous derivations, find

$$\text{E}[\boldsymbol{\mu}_{0\,|\,t}\,|\,\boldsymbol{Y}_{1:t-1}] = \boldsymbol{\mu}_{0\,|\,t-1}, \tag{4.48}$$

$$\begin{aligned}\text{Cov}(\boldsymbol{\mu}_{0\,|\,t}\,|\,\boldsymbol{Y}_{1:t-1}) &= (\boldsymbol{P}_{0\,|\,t-1}\boldsymbol{H}_{td}^\intercal S_{td}^{-1})\text{Var}(Y_{td}\,|\,\boldsymbol{Y}_{1:t-1})(\boldsymbol{P}_{0\,|\,t-1}\boldsymbol{H}_{td}^\intercal S_{td}^{-1})^\intercal \\ &= \boldsymbol{P}_{0\,|\,t-1}\boldsymbol{H}_{td}^\intercal S_{td}^{-1})S_{td}(\boldsymbol{P}_{0\,|\,t-1}\boldsymbol{H}_{td}^\intercal S_{td}^{-1})^\intercal \\ &= \boldsymbol{P}_{0\,|\,t-1}\boldsymbol{H}_{td}^\intercal S_{td}^{-1}\boldsymbol{H}_{td}\boldsymbol{P}_{0\,|\,t-1}^\intercal,\end{aligned} \tag{4.49}$$

which gives

$$\mathrm{E}[U_{di} \mid \boldsymbol{Y}_{1:t-1}] = \frac{l - \boldsymbol{\mu}_{0 \mid t-1}[i]}{\sqrt{\boldsymbol{P}_{0 \mid t}[i,i]}}, \tag{4.50}$$

$$\mathrm{Var}(U_{di} \mid \boldsymbol{Y}_{1:t-1}) = \frac{1}{\boldsymbol{P}_{0 \mid t}[i,i]} \boldsymbol{P}_{0 \mid t-1}[i,] \boldsymbol{H}_{td}^{\intercal} S_{td}^{-1} \boldsymbol{H}_{td} \boldsymbol{P}_{0 \mid t-1}[i,]^{\intercal}. \tag{4.51}$$

4.2.4 Hybrid criterion

A feature of the myopic approaches is that they look for local optima, that is, solutions that give immediate reward, and they fail to see solutions that might bring a better reward in the future. In this context, the AUV may get stuck in one area of the domain and fail explore other parts. Then important features might go undiscovered. Increasing the number of steps results in exponential growth of computational complexity, and it is therefore infeasible planning far ahead.

Instead of extending the model to look several steps ahead, one can set up a hybrid strategy, trying to balance the trade-off between exploration of the domain and exploitation of known information (Zhang and Yu, 2013). One approach is to use a ε-greedy approach, where one chooses the next step e.g. at random for exploration with probability ε, and otherwise optimises the next step based on short-term reward, as called by Zhang and Yu (2013). It is also possible to combine different decision criteria by adding weights to each, and thus balancing the need of exploration and the EMMP criterion. Berget et al. (2018) use this weighted approach. One of the criteria they use is to choose positions with high variance. Then the overall variance in the system is reduced, and it also ensures that the AUV travels to places in the domain so far unexplored. As noted in section 4.2, variance is often used as a decision criterion, and seems to be a good exploration criterion.

In this thesis, a variant of the ε-greedy approach is tested. The idea is to start with the predictive EMMP criterion from section 4.2.2 for k steps, and then use the same criterion for the next k steps with probability ε. Otherwise, variance is used in a decision criterion for the next k steps. More specific, the total variance reduction for the next step is compared:

$$\vartheta(t,d) = \mathrm{trace}(\boldsymbol{P}_{t \mid t-1} - \boldsymbol{P}_{t \mid td}), \tag{4.52}$$

where d is the identifier for a possible grid node for next step, and trace is the sum of the diagonal elements of a matrix. The matrices $\boldsymbol{P}_{t \mid t-1}$ and $\boldsymbol{P}_{t \mid td}$ are the forecasting and filtering covariance matrices for time step t, respectively, and the subscript d is added to $\boldsymbol{P}_{t \mid td}$ to emphasise dependence of position $\boldsymbol{s}_d$. The formulas are found in equations (2.43) and (2.49). The AUV will move to the position $\boldsymbol{s}_{d^*}$ with the largest $\vartheta(t, d^*)$.

Optimally, we would want to use the EMMP criterion as much as possible, since this is the criterion developed in this thesis which aims to find the ES_T and reduce the misclassification rate. Thus, we want to keep using this criterion with large probability until the degree of exploration is too low, or the AUV is stuck in some small area. If this happens, we want to reduce ε to increase the probability of switching to the variance criterion. There are many different approaches for tuning ε. In this thesis, an approach considering the earlier visited positions is considered.

From the beginning, the probability ε is quite high, as we want the default to be the EMMP criterion, call this ε_0. For each k time steps, the type of decision criterion is reevaluated. The number of times the AUV has been inside a circle of radius r from the current position is then found. This number, n, is at least 1, since the AUV is currently inside the circle. The probability ε is then reduced by a factor $\frac{1}{n}$, i.e.

$$\varepsilon = \varepsilon \cdot \frac{1}{n}. \tag{4.53}$$

The probability is decreased until the decision criterion is changed. Then it is set back to $\varepsilon = \varepsilon_0$. The procedure is shown schematically in Algorithm 1.

Algorithm 1: Algorithm for choosing position using the hybrid decision strategy at time t. The default probability of using the EMMP criterion is ε_0, and the probability is updated every kth step.

if $t > 1$ *and* $t \mod k == 1$ **then**
  if *Strategy* $==$ *Variance* **then**
    $\varepsilon = \varepsilon_0$;
  end
  neighbourhood $=$ grid nodes inside circle with radius r from current AUV position;
  $n =$ number of times AUV has visited neighbourhood;
  $\varepsilon = \varepsilon \cdot \frac{1}{n}$;
  draw u from $Uniform(0, 1)$ distribution;
end
if $u < \varepsilon$ **then**
  Strategy $=$ EMMP;
  Choose $position$ with smallest EMMP
else
  Strategy $=$ Variance;
  Choose $position$ with largest $\vartheta(t, d)$;
end
return $position$

The idea with this hybrid decision strategy is to let the AUV explore more of the domain and not be stuck in one small area. If the AUV has been very close to the current position earlier, this means that the AUV has turned and come back to the same area, as long as we set r smaller than the distance between two consecutive visited positions. We assume this is an indication of the AUV having to explore more. However, there is some randomness in the strategy, and one risks moving away from interesting areas too early, or in less desirable directions.

4.3 Monte Carlo simulation

We aim to compare various decision strategies for AUV sampling in Frænfjorden. There is no way to analytically assess the different strategies, since most strategies are adaptive and

each decision depends on the updated model with all measured values up to that step. For the predetermined paths, we could compute the EMMP for the whole path if all positions were visited at the same time, but when the path is visited over time, this is no longer possible. We therefore assess the strategies by performing a simulation study where paths are simulated using several time series samples from the Frænfjorden SPDE model. This is called Monte Carlo simulation, or "playing the game of Monte Carlo".

For the simulation study, time series replicates from the model must be made. Using the model parameters specific for Frænfjorden, i.e. those found in chapter 3, the propagator matrix $\boldsymbol{A}$, the vector $\boldsymbol{R}$ and the innovation term covariance matrix, $\boldsymbol{Q}$, are found. Algorithm 2 shows how to simulate one spatial time series from the model. The final time step is T. Sampling from the initial state distribution and the distribution of the innovation term, $\boldsymbol{\eta}$, is performed as in equation (2.8), with $\boldsymbol{L}$ the Cholesky factorisation of the relevant covariance matrix.

Algorithm 2: Simulating one time series replicate from SPDE process model.

Find $\boldsymbol{A}$, $\boldsymbol{R}$ and $\boldsymbol{Q}$;
Sample $\boldsymbol{x}_0$ from initial state distribution $N_N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$;
for $t = 1, \dots T$ **do**
    Sample $\boldsymbol{\eta}^*$ from innovation distribution $N_N(\boldsymbol{0}, \boldsymbol{Q})$;
    $\boldsymbol{x}_t = \boldsymbol{A}\boldsymbol{x}_{t-1} + \boldsymbol{R} + \boldsymbol{\eta}^*$;
end
$\boldsymbol{x} = (\boldsymbol{x}_0, \boldsymbol{x}_1, \dots \boldsymbol{x}_T)$;
return $\boldsymbol{x}$

For each such time series sample, a decision strategy can be used to simulate an AUV path. The time series replicate is then considered the true spatio-temporal system. Since the true systems are available, it is easy to evaluate the decision strategy based on the posterior model after data is obtained in all T positions of the path. The procedure should be repeated many times when evaluating, so that the individual Monte Carlo samples are not important for the result. We are interested in the performance of a strategy on average for the model, not the specific simulated time series. Algorithm 3 shows how to simulate one path in general. The model on board the AUV is updated according to the Kalman filter of section 2.2.2. For each new time step, the model is *forecasted* according to equations (2.42) and (2.43), and after each measurement is obtained, the model is *filtered* according to equations (2.48) and (2.49). Here, the distribution of the on-board model is simply called the *AUV distribution*. The measurement variance is τ^2.

Algorithm 3: Algorithm for simulating measurements and a path for the AUV when the true system is known for all time steps. The distribution of the model on board the AUV is called the *AUV distribution* for all time steps.

Let $\boldsymbol{x} = (\boldsymbol{x}_0, \boldsymbol{x}_1, \dots \boldsymbol{x}_T)$ be a the simulated spatial time series;
AUV distribution = initial state distribution;
for $t = 1, \dots T$ **do**
    Forecast *AUV distribution* one time step;
    Choose position $\boldsymbol{s}_{d_t}$ for step t using given *decision strategy*;
    Let $\boldsymbol{H}_{td}$ be the design matrix indicating the position $\boldsymbol{s}_{d_t}$;
    Sample measurement y_t from data model distribution $N(\boldsymbol{H}_{td}\boldsymbol{x}_t, \tau^2)$;
    Filter *AUV distribution* to include y_t at $\boldsymbol{s}_{d_t}$;
end
$\boldsymbol{Y}_{1:T} = (y_1, \dots y_T)^\intercal$;
return *AUV distribution* for time step T given $\boldsymbol{Y}_{1:T}$.

Chapter 5

Simulation study for Frænfjorden

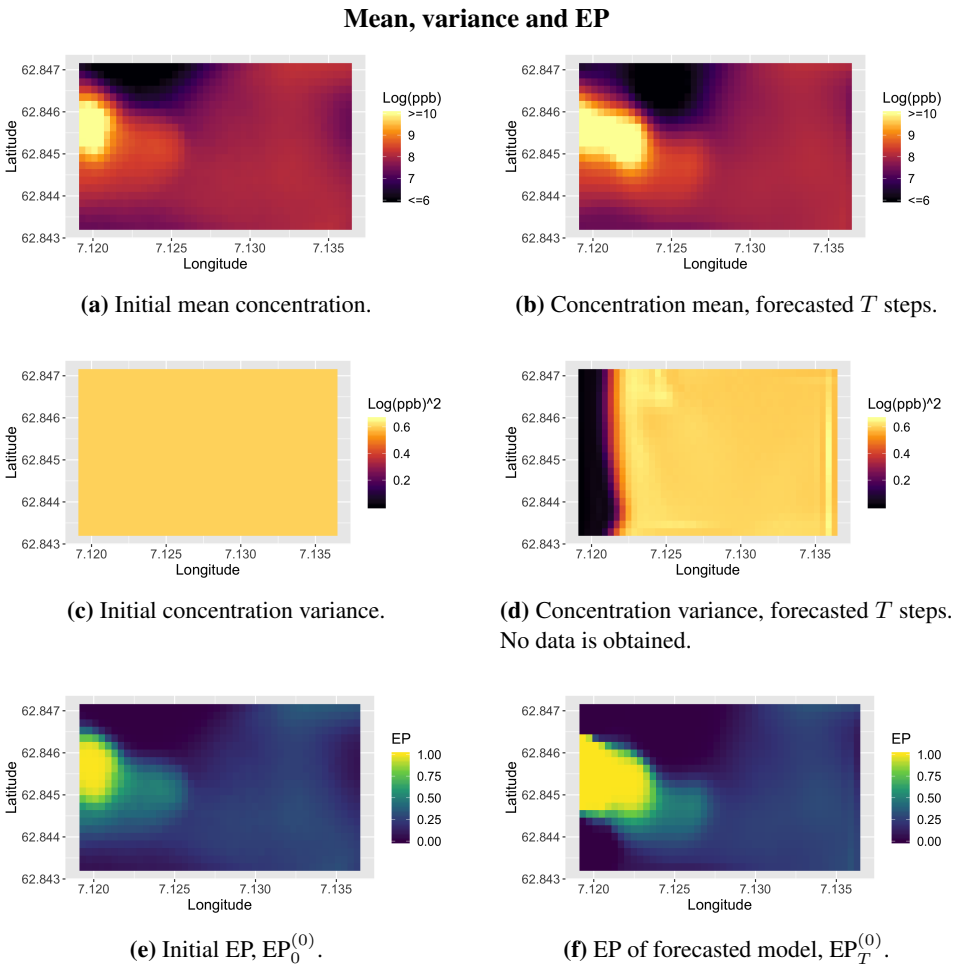
A simulation study is performed to study how to conduct efficient sampling of concentration in Frænfjorden to get an accurate prediction of the ES for a fixed time T. In this chapter, the different tests and setup for the simulations, as well as some numerical aspects of the study, are explained.

5.1 Simulation setup

For the simulation study, we assume that the model of Frænfjorden from chapter 3 describes the concentration state in Frænfjorden perfectly, and simulate spatial time series from this model. In total 1000 realisations from the initial state distribution are drawn, and then the process model (equation (2.31)) is used to create time series, as described in Algorithm 2. Because of the small variance of $\boldsymbol{\eta}$ estimated in chapter 3, the variations in the series are mainly due to the different realisations of the initial state vector. We set Dirichlet BCs along the western boundary and Neumann BCs along the others, as described in section 3.1.2.

The ES at time step T, ES_T from equation (4.3), is of interest. To define the ES, the critical limit, l, of the area is needed. For simplicity, we set $l = 8.5$ $\log(\text{ppb})$, as the values in the area are usually between 5 and 11 $\log(\text{ppb})$. In real applications, the limit should be clear from the operational objective. Figure 5.1 shows a comparison of the model at initial state and at time step T when no data is present. We see that the high concentrations in the model mean have expanded from the initial state to time step T, due to the movement included in the SPDE and the constant incoming pollutants from the Dirichlet BCs at the west boundary. Over time, the incoming known concentrations result in a very low variance on the west side of the domain, as seen in Figure 5.1d. There is also some numerical effect along the eastern boundary, but these are small, as assumed in section 3.1.2.

We let the time period for AUV operations be 30 minutes, with one sample every 1 minute, starting 1 minute after initial state. Thus, each of the 1000 time series consists



(a) Initial mean concentration.

(b) Concentration mean, forecasted T steps.

(c) Initial concentration variance.

(d) Concentration variance, forecasted T steps. No data is obtained.

(e) Initial EP, $\text{EP}_0^{(0)}$.

(f) EP of forecasted model, $\text{EP}_T^{(0)}$.

Figure 5.1: Model concentration, variance and EP initially and after $T = 30$ time steps. Critical limit for the EP is $l = 8.5$ log(ppb).

of 31 samples, for time steps $t = 0, 1, \dots T$ ($T = 30$), and measurements are drawn for $t = 1, 2, \dots T$. The AUV moves with speed between 0.5 and 2 $\mathrm{m/s}$ (AUR lab, nd). Assuming it can have an average speed of a bit over 1 $\mathrm{m/s}$, it should be able to spend some seconds deciding where to go next, move about 60 m, and do one measurement every 1 minute. We let the possible positions for each step be as explained in section 4.2 and shown in Figure 4.3. In practice, the AUV would probably do measurements continuously along the way, but only decide the next direction at some points. For simplicity, we assume the measurements are only obtained in the selected positions. In all simulations, we let the AUV do its first measurement at the middle position on the southern border of the sampling area.

The measurement variance τ^2 is not easy to set, as there are several factors that are uncertain. The measurement errors in the sensor are probably small compared to variability in the AUV position and in the concentration within the discretised grid node. Calibration of sensors attached to the AUV has shown hard (Cetinić et al., 2009; Rogowski et al., 2013). In addition, what is measured is optical backscattering in form of attenuation coefficient, as a measure of turbidity, which has to be converted to a measure of (log) concentration, and there is some uncertainty in this transformation as well. The Wet Labs ECO 3 sensor user manual (Eco 3, 2007) suggests a attenuation error of about $4\ \%$ at attenuation coefficient 1 m^{-1}. Since the log concentrations in the area vary between approximately 5 and 11 $\log(\text{ppb})$, assuming measurement standard deviation of $4\ \%$ of the log concentration values gives τ between 0.2 and 0.44 $\log(\text{ppb})$. Since this is a simulation study, it is important that some information is indeed gained from doing observations, and at the same time that the observations are not too close to being exact. The initial variance in a position is 0.6 $\log(\text{ppb})^2$. The random noise is very small, and the variance over time is not changed much for most of the domain, see Figure 5.1c and 5.1d. For measurements to be informative, their variance should be considerably smaller than the original concentration variance. We choose $\tau^2 = 0.1\ \log(\text{ppb})^2$, i.e. $\tau \approx 0.316\ \log(\text{ppb})$. Then the variance is of the same size order as the original concentration variance, but considerably smaller, and the standard deviation is within the $4\ \%$ range.

5.2 Test cases

Several models and strategies for deciding the AUV paths are introduced. From these, we construct test cases combining model and decision strategy. These test cases are used for choosing AUV paths in the simulation study, and are explained in this section. Most strategies are myopic. For each test case, paths are simulated according to Algorithm 3 for each time series replicate.

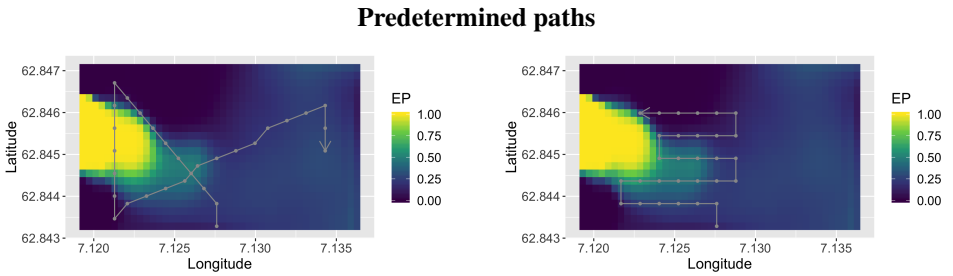
First consider two strategies where the AUV is using the SPDE model from section 2.2.1 and the EMMP criterion from section 4.2.1 to decide where to obtain data. That is, the SPDE process model is assumed to describe the concentration dynamics in the system, and since this is a simulation study, this model is the same as the underlying model for the simulated spatial time series. For the decision criterion, we consider one case where the forecasted distribution of $\boldsymbol{X}_T$ as target distribution and condition on all data $\boldsymbol{Y}_{1:t}$. Details can be found in section 4.2.2. This test case is the main focus in this thesis, as it uses both the main model and criterion described and developed in the thesis. The test case

is denoted $\{$SPDE: EMMP $\boldsymbol{X}_T\}$, indicating the model and the strategy used to create the results. We also consider a similar case where the SPDE model is used, but the forecasted distribution is ignored. Instead, the distribution of $\boldsymbol{X}_t$ is used as target distribution for step t, still conditioned on data $\boldsymbol{Y}_{1:t}$. This method is less accurate, but also more computer efficient. The test case is denoted $\{$SPDE: EMMP $\boldsymbol{X}_t\}$, and details are found in section 4.2.3.

Consider the EMMP criterion using the distribution of the state at time T, but let this distribution be determined by the simpler spatial AR(1) time series model introduced in section 2.2.1. In this case, the model differs from the underlying model in the simulated concentration states, but it is simpler, so computations are faster than for the $\{$SPDE: EMMP $\boldsymbol{X}_T\}$ case. When the AR(1) model is used, the calculations needed for the EMMP criterion are found in equations (4.40) to (4.42), and we denote the test case $\{$AR(1): EMMP$\boldsymbol{X}_T\}$. Further, we consider a spatial model, where the true state is assumed constant over time, and the initial state model for Frænfjorden from section 3.1.1 is used for the AUV sensing. One then assumes that the changes forward in time the next $T = 30$ minutes are negligible. There is not much sense in forecasting to time step T, as the model stays the same as long as no new data is obtained. We therefore denote this test case $\{$Spatial $\boldsymbol{X}_0$: EMMP$\}$, where $\boldsymbol{X}_0$ represents the initial state model, and is the target state for the EMMP criterion, conditioned on $\boldsymbol{Y}_{1:t}$.

Now we have described four test cases using the EMMP decision criterion but different models. For the remaining 5 test cases, the SPDE model is the foundation for the AUV paths, but the decision strategies vary. We consider two cases with the simple decision criteria mentioned in section 4.2. The test cases are denoted $\{$SPDE: $\text{EP}_t^{(t-1)}0.5\}$ and $\{$SPDE: random$\}$. We also make two cases where the paths are predetermined, and the AUV thus visits the same positions in each spatial time series replicate. We let each predetermined path be as comparable as possible to the adaptive paths. That is, they consist of 30 measurement points with a distance of about $60\ \mathrm{m}$ between each point, and the time between each obtained measurement is 1 minute, just as before. The first position is also at the middle position of the southern border. Two different approaches for predetermining measurement points are considered. In the first one aims to explore an as big part of the domain as possible. It is seen in earlier work that the AUV is set to measure along the diagonals of the research areas (Fossum et al., 2018), so this strategy is used for this path. In addition, the path is intended to move across areas with $\text{EP}_T^{(0)}$ close to 0.5. The path is shown together with $\text{EP}_T^{(0)}$ in Figure 5.2a, and the test case is called $\{$SPDE: Predetermined diagonals$\}$. The second predetermined path intends to explore the area where $\text{EP}_T^{(0)}$ is close to 0.5 uniformly, and the result is shown in Figure 5.2b. The test case is called $\{$SPDE: Predetermined uniform$\}$.

Finally, one test case using the hybrid decision criterion described in section 4.2.4 is made. The distance between each position visited by the AUV is still approximately $60\ \mathrm{m}$. To set the probability ε, we need to define the radius of the neighbourhood around the current position where we count the number of earlier measurements. This should be small enough to be certain that the AUV has returned to the same area several times. Thus, we need the radius, r, smaller than 2.8 grid node distances. We set $r = 40.2\ \mathrm{m}$, or 2 grid node distances. Further, we set the initial probability $\varepsilon_0 = 0.9$ and the number of steps before ε is updated and the strategy might be changed, to be $k = 5$. The idea is that the



(a) Path along diagonals.

(b) Uniform path near positions of $\text{EP}_T^{(0)}$ close to 0.5.

Figure 5.2: Decision of positions for predetermined paths. Both use $\text{EP}_T^{(0)}$ as a basis.

AUV then gets some time to explore locally or move to another part of the domain before the strategy is again (potentially) changed. We denote this test case {SPDE: Hybrid}.

5.3 Time evaluation

All code for calculations in this thesis is written using the statistical software R. The various strategies demands different amount of calculations, and thus different amount of time is needed to use the strategies. Table 5.1 shows approximate timings from simulating one path with each strategy using a MacBook Pro 2015 laptop with i5 core. We assume these are comparable to computer time needed by the AUV. What is most relevant in practice, is the *Choose next step* column, as this shows the approximate time needed from obtaining one measurement until the next position is decided and the AUV can start moving in that direction. One would like to spend the time the AUV is in the water as efficiently as possible, meaning that this pure calculation time should be small. If it exceeds about 15 s, the AUV will not be able to move the distance assumed in this simulation study and still measure every 1 minute. In addition, the AUV is affected by currents, so that it is unfeasible to stand still.

For all paths using the EMMP strategies, time to compute the filtering and forecasting distributions (equations (2.40) and (2.41)) for each step is part of the decision time, as this is needed after a measurement is obtained before the AUV can decide where to go next. For the test case {Spatial $\boldsymbol{X}_0$: EMMP}, there is no forecasting since there is no time dynamic, and time is therefore saved. For cases {AR(1): EMMP $\boldsymbol{X}_T$} and {SPDE: EMMP $\boldsymbol{X}_T$}, evaluating each position is much slower, as for each position considered for the next step, covariance matrices forecasted to time step T are needed. The times vary for the {AR(1): EMMP $\boldsymbol{X}_T$} test case because the forecasting to time step T is coded with a for loop, and the number of times to loop decreases as the current time step t increases. Thus, the time needed to decide is shorter as time passes.

The most computer demanding test case is {SPDE: EMMP $\boldsymbol{X}_T$}. Forecasting $T - t$ steps ahead in time for each time step includes a lot of matrix multiplications, which require computer resources. Even after several attempts on making computations faster

	Consider 1 position	*Choose next step*
SPDE: EMMP $\boldsymbol{X}_T$	2.5 s	40 s
SPDE: EMMP $\boldsymbol{X}_t$	0.05 s	3 s
Spatial $\boldsymbol{X}_0$: EMMP	0.05 s	1.5 s
AR(1): EMMP $\boldsymbol{X}_T$	0.4-0.6 s	8-11 s
SPDE: $\text{EP}_t^{(t-1)}0.5$	0 s	2 s
SPDE: Random	0 s	0 s
SPDE: Predetermined diagonals	-	-
SPDE: Predetermined uniform	-	-
SPDE: Hybrid	0.5/2.5 s	10/40 s

Table 5.1: Approximate running times using a MacBook Pro 2015 laptop with i5 core. The times are assumed comparable to computing times on board the AUV. For thy hybrid test case, the numbers represent times for the variance and EMMP criterion, respectively.

and computing most of the matrix multiplications beforehand, decision time is about 40 s for each step for case {SPDE: EMMP $\boldsymbol{X}_T$}. This is of course a lot more than is feasible for real-time computations on board an AUV. It should be noted that the less accurate {SPDE: EMMP $\boldsymbol{X}_t$} case only needs 3 s to choose next step, which is acceptable for real-time calculations.

The test cases {SPDE: $\text{EP}_t^{(t-1)}0.5$} and {SPDE: Random} spends no time evaluating each possible position, but for {SPDE: $\text{EP}_t^{(t-1)}0.5$}, the distributions need to be updated before the $\text{EP}_t^{(t-1)}$ can be computed, and thus the AUV spends some time deciding the next step. For the predetermined paths and test case {SPDE: Random}, the AUV does not have to stop and wait at all. For the {SPDE: Hybrid} case, there is quite a difference in time between using the EMMP criterion and the variance reduction criterion. The numbers before the slash in Table 5.1 represent the variance computations, which are a lot faster. The strategy is designed so use the EMMP criterion most of the time, so one average one does not save much time, however.

When running on a more powerful server, and running code in parallel, the time could be reduced by up to 90 %. Because of this, more time is not spent on improving the code for this simulation study. However, calculations on board an AUV is the goal. If to be used for field work, the code must be translated into the language `Python` to be compatible with currently running code, and further improvements can then be tested. Some possibilities for further improvements are exploiting that the propagator matrix, $\boldsymbol{A}$, from the process model is a sparse matrix, considering the misclassification probability on only parts of the domain when choosing positions, to deal with smaller matrices in the decision computations, or writing the algorithms in another language, such as `C`. One can also let the number of positions to consider in each step be fewer.

5.4 Numerical issues

During the simulations of test case {SPDE: EMMP X_T}, some numerical issues are encountered. The issues arise when forecasting to time step T to evaluate the EMMP of

potential next positions. In this calculation, the covariance matrix of $\boldsymbol{\mu}_{T\,|\,t}$ is needed, as given in equation (4.37). Numerically, this matrix is not always positive definite, as recognised by negative eigenvalues. The error is clearly numerical, and a small nugget effect in the innovation term covariance matrix and a slight damping in the SPDE are added trying to push values away from zero. This does not always solve the problem. Since the matrix is not part of the main model, but only used in the decision strategy, it is used without amendment, but during simulations, it happens that diagonal elements, i.e. variances, also become numerically negative. In these cases, the computation is aborted and the whole replicate skipped. Fortunately, this issue arise for only 1 out of 1000 time series replicates. As a result, we have 999 paths for the $\{$SPDE: EMMP $X_T\}$ test case.

Chapter 6

Results and analysis

AUV sensing is simulated as in Algorithm 3 on the 1000 time series samples using the parameters estimated for Frænfjorden. The results from this main simulation study are presented in this chapter. Each section considers a few of the test cases introduced in section 5.2. The goal is to reduce prediction error with respect to ES_T. At time step T, data from T positions is obtained, $\boldsymbol{Y}_{1:T}$, one measurement for each time step, and the distribution of the state vector $\boldsymbol{X}_T \,|\, \boldsymbol{Y}_{1:T}$ should be as informative as possible regarding ES_T.

The main decision strategy discussed in this thesis is the EMMP decision criterion, and after each simulated path, we can compute the posterior MMP. Analogous to equation (4.7), we obtain

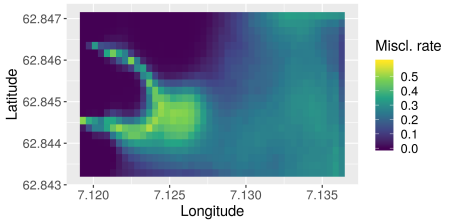
$$\text{Posterior MMP} = \frac{1}{N}\sum_{i=1}^{N} \min\{1 - P(X_T(\boldsymbol{s}_i) > l \,|\, \boldsymbol{Y}_{1:T}), P(X_T(\boldsymbol{s}_i) > l \,|\, \boldsymbol{Y}_{1:T})\}. \tag{6.1}$$

In this chapter, the Posterior MMP is referred to simply as MMP. Further, actual misclassification rate for the paths can be calculated since the true systems are available. For the classification, a probability cutoff of 0.5 on $\text{EP}_T^{(T)}$ is used, matching the EMMP criterion. The misclassification rate after one simulated path is

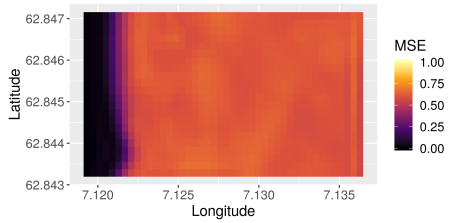
$$\text{Misclassification rate} = \frac{1}{N}\sum_{i=1}^{N} \left((\text{EP}_T^{(T)}(\boldsymbol{s}_i) \geq 0.5) \neq \text{ES}_T(\boldsymbol{s}_i) \right), \tag{6.2}$$

where $\text{ES}_T(\boldsymbol{s}_i)$ is as in equation (4.5) at time T. We assume the logical *True* and *False* are also given as 1 and 0, respectively, and can thus be summed over. We want to evaluate the strategies on average, and thus focus on the mean of all calculated MMPs or misclassification rates. We also study the spatial misclassification rate. Then, the rate is calculated over all $B = 1000$ samples for each grid node separately, to see where in the domain misclassification is more common, and

$$\text{Spatial misclassification rate}(\boldsymbol{s}_i) = \frac{1}{B}\sum_{b=1}^{B} \left((\text{EP}_T^{(T)b}(\boldsymbol{s}_i) \geq 0.5) \neq \text{ES}_T^b(\boldsymbol{s}_i) \right) \tag{6.3}$$



(a) Misclassification rate in each grid node over all 1000 samples using $\text{EP}_T^{(0)}$ for classification.



(b) MSE in each grid node over all 1000 samples when prediction is the mean of the SPDE forecasted distribution at time T, with no data available.

Figure 6.1: Spatial misclassification rate and MSE using the distribution from the SPDE model at time T without obtaining any data.

is the misclassification rate for grid node $\boldsymbol{s}_i$. Superscript b indicates time series replicate number b. Considering the SPDE model distribution at time T without obtaining any data, results in an average misclassification rate of $18.2\ \%$ with standard deviation $11.9\ \%$. The spatial misclassification rate is displayed in Figure 6.1a. We will see, in the proceeding sections, how much these misclassification rates are improved when obtaining data according to the various test cases.

Though not the main focus, we can also compare how well the posterior distribution at time step T predicts the actual concentration at that time. We consider the concentration variance at the end of each run, and also the mean squared error (MSE). After one simulated path,

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(\text{E}[X_T(\boldsymbol{s}_i)\,|\,\boldsymbol{Y}_{1:T}] - x_T(\boldsymbol{s}_i)\right)^2, \tag{6.4}$$

and mean MSE is the average of this over the 1000 samples. Like for misclassification rate, MSE for each spatial position is also considered, showing the average squared error of concentration in each grid node over all samples. This is shown for the case where no data is obtained in Figure 6.1b. The misclassification rate and MSE can be computed since this is a simulation study, and the true full systems are available. In other situations, these are not available, and the MMP and model variances are left for evaluating the strategies. Expected squared error can be factorised into a bias term and a variance term, explaining the bias-variance trade-off in statistical prediction and machine learning (James et al., 2013). The bias is how much the expected value of the predictor differs from the expected value of the concentration. In this study, the predictor is the posterior mean $\boldsymbol{\mu}_{T\,|\,T}$. If the bias is small, the mean posterior variance and the spatial MSE are quite similar. The unit for the MSE and the variances is $\log(\text{ppb})^2$, but this is not included in plots and discussion in this chapter, since it is not important for interpretation. The mean posterior variance is, for grid node $\boldsymbol{s}_i$,

$$\text{Mean variance} = \frac{1}{B}\sum_{b=1}^{B}\boldsymbol{P}_{T\,|\,T}[i,i]^b. \tag{6.5}$$

6.1 SPDE model and EMMP strategy

First consider the two test cases $\{$SPDE: EMMP $\boldsymbol{X}_T\}$ and $\{$SPDE: EMMP $\boldsymbol{X}_t\}$ using the SPDE model and the EMMP criterion.

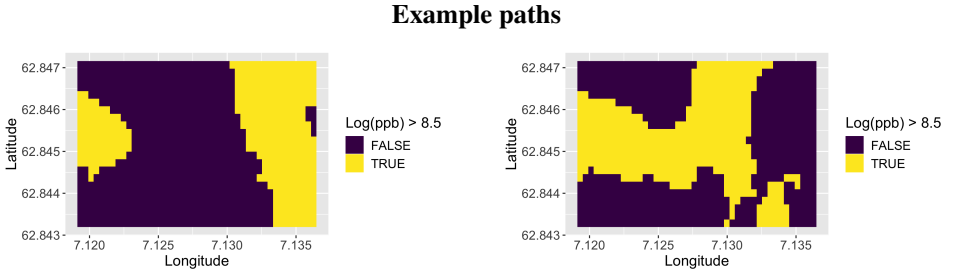
Test case	Mean MMP	Mean misclass. rate	SD misclass. rate
SPDE: EMMP $\boldsymbol{X}_T$	0.104	0.103	0.073
SPDE: EMMP $\boldsymbol{X}_t$	0.103	0.105	0.075

Table 6.1: Mean MMP and misclassification rate for the two test cases using the SPDE model and the EMMP strategy. The misclassification rate is for cutoff at probability 0.5.

The mean final MMP and misclassification rate and the misclassification standard deviation are displayed in Table 6.1. The two test cases are not significantly different. The mean misclassification rate is about $10\,\%$ for both test cases. The misclassification rate is smaller for the predictive $\boldsymbol{X}_T$ case, but not by much. The variation in misclassification rate is also quite similar for the two cases, and so is the mean MMP. The sample standard deviation for the MMP is approximately 0.026 for both cases. The MMP is actually slightly smaller for the $\boldsymbol{X}_t$ case, showing that minimising the (expected) probability of misclassification does not necessarily lead to actual lower misclassification rate. The MMP is dependent on the underlying model, and since these two cases use the same model, the MMPs are comparable. Misclassification rate is a better means of comparing the test cases, but when this is not possible, we see that the MMP is quite similar to the misclassification rate.

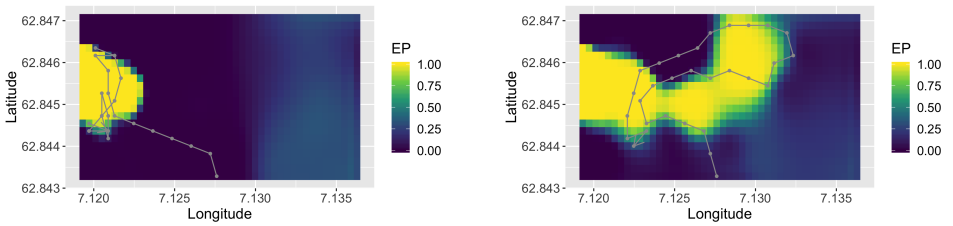
Figure 6.2 displays some paths made by the two different adaptive strategies. The paths are drawn upon the final $\text{EP}_T^{(T)}$, and the true ES_T is also shown for each sample.
In a lot of the simulated replicates, the ES is divided into two separate parts of the domain, one in the western part of the domain and one in the eastern. In these cases, the AUV rarely discovers both parts, but gets stuck in one area, mostly the western part. Because the strategies are myopic, the AUV only looks one step ahead, and therefore prioritises to move to places with immediate reward, in form of reduction in EMMP. To discover other areas of high concentration, the AUV must move across areas where the probability of high concentration is very low, so there is little immediate gain. Figure 6.2c shows an example of a case where the western side of the domain is researched first, and the AUV never moves away from this area again. The $\text{EP}_T^{(T)}$ is below 0.5 in the eastern area, and the positions are classified as 0, or *not part of the* ES_T. The true ES_T is shown in Figure 6.2a, and most of the eastern half of the system is misclassified in this case. In the system shown by the ES_T in Figure 6.2b, the AUV moves around a lot more for both test cases.

Though paths are different for the two strategies, on average they explore the area in a very similar way. We can see this by studying the mean variance after T time steps for both test cases, see equation (6.5). These are shown in Figure 6.3a and 6.3b for the $\{$SPDE: EMMP $\boldsymbol{X}_T\}$ and the $\{$SPDE: EMMP $\boldsymbol{X}_t\}$ case, respectively. The variance is small near the path of the AUV, so a mean path is visible in the variance plot. It should be noted that the positions with small variance move as time passes by, due to the forecasting process model (equation (2.31)). Thus the exact path is not visible, but some general features are nevertheless interesting. The starting position near the central south node is recognised as the purple area close to the middle of the southern border. Further, the



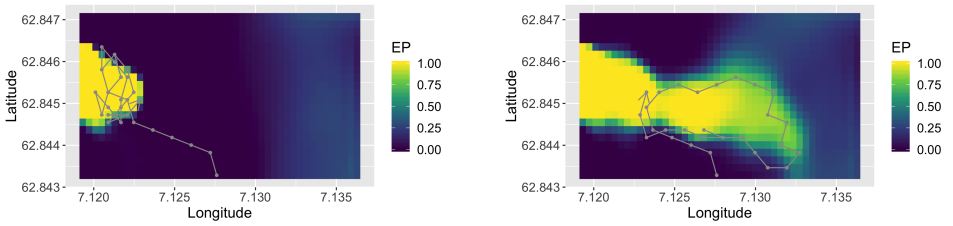
(a) ES_T for time series 1.

(b) ES_T for time series 2.



(c) Final path using the {SPDE: EMMP $\boldsymbol{X}_T$} test case on time series 1 together with $\mathrm{EP}_T^{(T)}$.

(d) Final path using the {SPDE: EMMP $\boldsymbol{X}_T$} test case on time series 2 together with $\mathrm{EP}_T^{(T)}$.

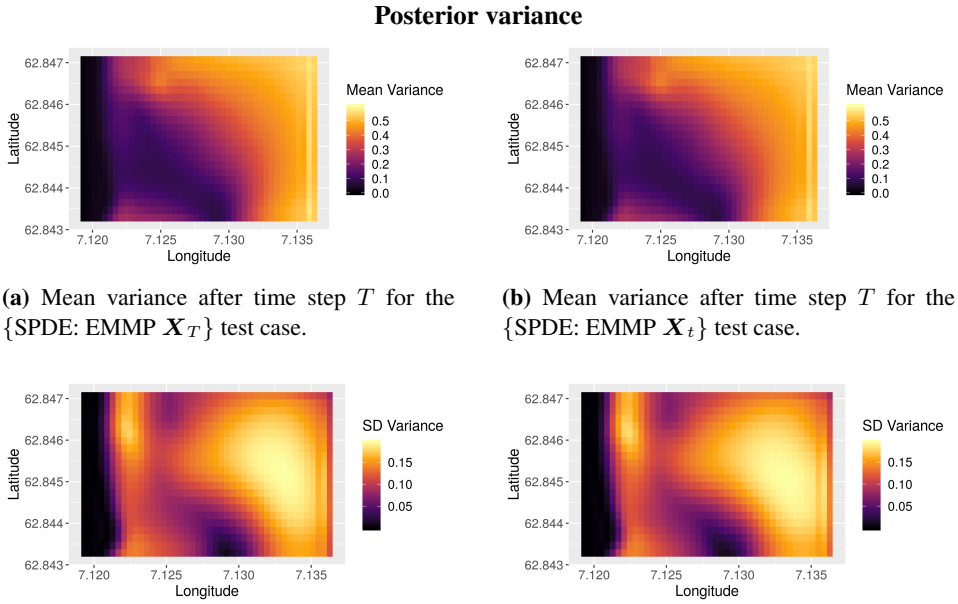


(e) Final path using the {SPDE: EMMP $\boldsymbol{X}_t$} test case on time series 1 together with $\mathrm{EP}_T^{(T)}$.

(f) Final path using the {SPDE: EMMP $\boldsymbol{X}_t$} test case on time series 2 together with $\mathrm{EP}_T^{(T)}$.

Figure 6.2: Final simulated paths for two sampled time series and the two test cases {SPDE: EMMP $\boldsymbol{X}_T$} and {SPDE: EMMP $\boldsymbol{X}_t$}. The paths are displayed together with the corresponding $\mathrm{EP}_T^{(T)}$s, and the ESs at time step T for the two samples are also shown. The $\mathrm{EP}_T^{(T)}$ should look as much like the ES as possible.

AUV mostly moves west in both cases, so the variance is small in the north-west direction from the starting position. The mean path is shifted slightly to the east for the predictive $\{\text{SPDE: EMMP } \boldsymbol{X}_T\}$ test case compared with the $\{\text{SPDE: EMMP } \boldsymbol{X}_t\}$ case, though this is barely visible. The shift is due to the prediction forward in time. The movement of pollutants in the area is, roughly speaking, to the east, so the area with $\text{EP}_T^{(t)}$ close to 0.5 is farther to the east than for $\text{EP}_t^{(t)}$.



(a) Mean variance after time step T for the $\{\text{SPDE: EMMP } \boldsymbol{X}_T\}$ test case.

(b) Mean variance after time step T for the $\{\text{SPDE: EMMP } \boldsymbol{X}_t\}$ test case.

(c) Sample standard deviation for the posterior variance after time step T for the $\{\text{SPDE: EMMP } \boldsymbol{X}_T\}$ test case.

(d) Sample standard deviation for the posterior variance after time step T for the $\{\text{SPDE: EMMP } \boldsymbol{X}_t\}$ test case.

Figure 6.3: Mean and standard deviation for the posterior variance in each grid node after T time steps for the two test cases using the SPDE model and the EMMP decision strategy.

Figure 6.3c and 6.3d show the corresponding sample standard deviation of the posterior variance. These displays show the variation in variance for each position over all simulated spatial time series, and tell more about where the AUV moves in each case. The standard deviation is very small close to the starting position, since the AUV always starts at the same place. Further, we can see that the AUV usually chooses the same positions in the beginning, due to the small standard deviation in the north-west direction from the starting position. Since very little data is obtained at this point, the models are very similar in each run, and thus the decisions are similar as well. The standard deviation is relatively high in the north-eastern part of the domain. This tells us that, even though the AUV on average does not obtain much information from this area due to the large mean variance, this varies much from sample to sample. Mainly, it seems the AUV moves into this area when there is a connected area of high concentration across a bigger part of the domain, as in 6.2d and 6.2f. Another interesting observation from Figure 6.3 is the movement of the AUV in

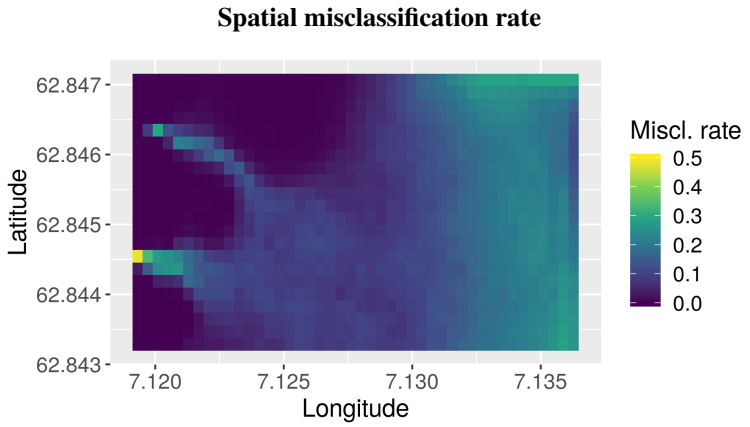


Figure 6.4: Misclassification rate in each grid node over all 999 samples for test case {SPDE: EMMP $\boldsymbol{X}_T$}.

the northern part of the domain. There is a small area of small standard deviation (purple) in the middle/west part, close to the northern boundary. The mean variance in this area is relatively high. This means that the AUV very rarely moves into this area. Just to the west of this area, however, the mean variance is relatively small, but the variation quite big. In all displays of Figure 6.3, the incoming small variance in the western part is due to the Dirichlet BCs, and movement of the AUV in this area does not make much of a difference on the variance. Again, the plots for the two test cases are without substantial differences. Because the AUV in both test cases tends to get stuck in the western area, this area is explored thoroughly in both cases, and the advantage of predicting forward in time seems to be lost.

We may also study the misclassification rate in space, computed as in equation (6.3). We have seen that the movements of the AUV in the two test cases are very similar, so the misclassification rates are also very similar in space. Figure 6.4 shows the misclassification rate at each grid node for test case {SPDE: EMMP $\boldsymbol{X}_T$}. The misclassification rate is largest at a point on the left boundary. This is because the concentration here, according to the BCs, is 8.485 $\log(\text{ppb})$, and due to the noise term in the SPDE (equation (2.31)), there is some randomness and the concentration may be higher or lower than the limit of 8.5 $\log(\text{ppb})$ around this point. This makes the concentration in this area hard to predict. The misclassification rate is quite large just around the western concentration front. Because of this, there will also be immediate reward in exploring this area in more detail, and the AUV thus fails to explore other parts of the domain. The whole east side of the domain also has quite large misclassification rates. As we have already seen, the AUV often does not explore this area.

Table 6.2 shows the average MSE at the final time step and the corresponding sample standard deviation. The MSEs are close to identical for the two approaches. Since this is a simulation study, and the model on board the AUV is the same as the model used to simulation the concentration data, the concentration predictions should not be biased. The MSEs for each spatial grid point are very similar to the variances in Figure 6.3a and

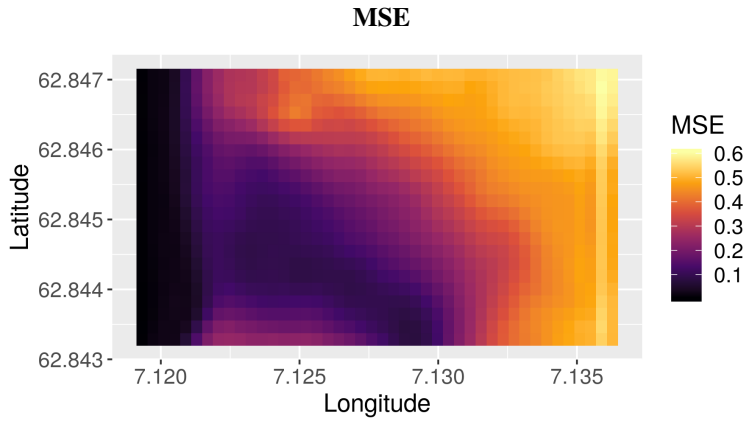


Figure 6.5: MSE in each grid node over all 999 samples for the {SPDE: EMMP $\boldsymbol{X}_T$} test case.

6.3b, as then should be expected. The MSE for case {SPDE: EMMP $\boldsymbol{X}_T$} is shown in Figure 6.5. The MSE is largest to the north-east of the domain, while it is very small in the western part of the domain. Comparing Figure 6.5 to the spatial misclassification rate in Figure 6.4, we see that the misclassification rate is large in some areas where the MSE is small. This shows that the high misclassification rate is then due to the concentration being very close to the limit for the ES, and not due to bad concentration prediction. We also note a line of large MSE close to the eastern boundary due to the BCs. The same type of effect can be found in the variance plots in Figure 6.3.

Test case	Mean MSE	SD MSE
SPDE: EMMP $\boldsymbol{X}_T$	0.273	0.158
SPDE: EMMP $\boldsymbol{X}_t$	0.276	0.153

Table 6.2: Mean and standard deviation for MSE for the two test cases using the SPDE model and the EMMP decision strategy.

The analysis of these two first test cases does not indicate any advantage in using the forecasting strategy, as the results are close to identical. Further, the forecasting is quite computer demanding, and as explained in section 5.4, this sometimes also leads to numerical issues regarding positive definiteness. For efficient computations on board AUVs, using the distribution of $\boldsymbol{X}_t$ instead of $\boldsymbol{X}_T$ is a lot more convenient. However, the similarity seems to be due to the fact that the AUV only explores one small area when the ES is disconnected, and there might be more to gain on the forecasting strategy in different situations.

6.2 EMMP strategy with simpler models

Now we will look at the effect of using simpler distributions when modelling the concentrations on board the AUV, and consider the test cases {Spatial $\boldsymbol{X}_0$: EMMP} and {AR(1):

EMMP $\boldsymbol{X}_T\}$. The mean MMP, mean misclassification rate and the standard deviation of the misclassification rate are displayed in Table 6.3. Since the models used in these test cases are different from the underlying SPDE model, and also different from each other, the mean MMPs, which are calculated from the on-board model (see equation (6.1)), are not comparable, and cannot be used for evaluating the test cases. The misclassification rates, however, can still be used, and it can be seen that these are larger for both these test cases than for the test cases from section 6.1, as found in Table 6.1. The misclassification rate is slightly smaller using the AR(1) model than the spatial model, but the difference is not very big. Immediately, the impression is that it is important to include the temporal advection-diffusion dynamics in the model. The simple use of an AR(1) model, which assumes a stationary time series, does not capture the dynamics nearly as well, though this is slightly better than ignoring temporal effects completely.

Test case	Mean MMP	Mean misclass. rate	SD misclass. rate
Spatial $\boldsymbol{X}_0$: EMMP	0.112	0.140	0.079
AR(1): EMMP $\boldsymbol{X}_T$	0.132	0.135	0.079

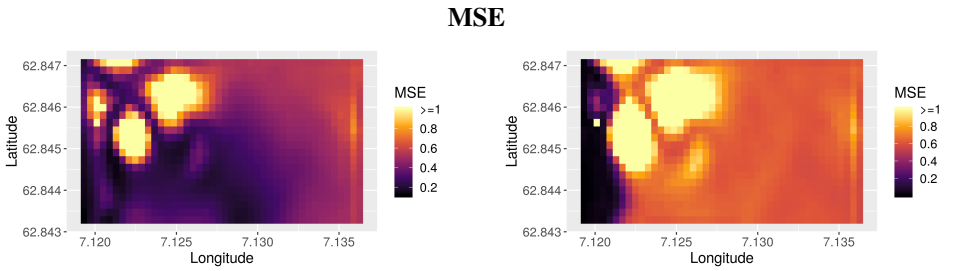
Table 6.3: Mean MMP and mean misclassification rate for two test cases using the EMMP decision strategy, but different models. The misclassification rate is for cutoff at probability 0.5.

Analysing other features substantiate this hypothesis. The MSE is significantly larger for test cases $\{$Spatial $\boldsymbol{X}_0$: EMMP$\}$ and $\{$AR(1): EMMP $\boldsymbol{X}_T\}$ than $\{$SPDE: EMMP $\boldsymbol{X}_T\}$ and $\{$SPDE: EMMP $\boldsymbol{X}_t\}$, and the difference between the spatial case and the AR(1) case is very small, as can be seen in Table 6.4.

Test case	Mean MSE	SD MSE
Spatial $\boldsymbol{X}_0$: EMMP	0.486	0.191
AR(1): EMMP $\boldsymbol{X}_T$	0.482	0.190

Table 6.4: Mean and standard deviation for MSE for two test cases using the EMMP decision strategy, but different models.

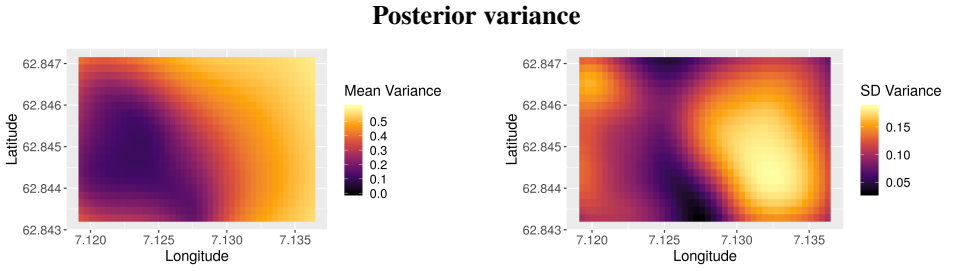
The large average MSE is due to very large MSE in some parts of the domain, as the spatial MSE for test case $\{$AR(1): EMMP $\boldsymbol{X}_T\}$ in Figure 6.6a shows. The prediction in the areas with large MSE is particularly bad as a result of large concentration variation and incorrect dynamics in the model. If no data is obtained, the concentration prediction is the distribution mean at time T. With the correct process model, the MSEs for the various spatial positions are as in Figure 6.1b, while if one uses simply a spatial model with no movement from the initial state, the MSE is as shown in Figure 6.6b. Comparing Figure 6.6a to the latter two plots, it is clear that the $\{$AR(1): EMMP $\boldsymbol{X}_T\}$ test case does not have the correct model, but that prediction is improved by obtaining new information, as the areas with large MSE are smaller in Figure 6.6a than in Figure 6.6b. The mean variance for test case $\{$AR(1): EMMP $\boldsymbol{X}_T\}$ is shown in Figure 6.7a. The variances and spatial MSE are not similar, due to the difference in models. Formulated with the bias-variance factorisation, the biases of the prediction models are large. The results are similar for the two test cases $\{$AR(1): EMMP $\boldsymbol{X}_T\}$ and $\{$Spatial $\boldsymbol{X}_0$: EMMP$\}$.



(a) MSE in each grid node over all 1000 samples for the {AR(1): EMMP $\boldsymbol{X}_T$} test case.

(b) MSE in each grid node over all 1000 samples when prediction is the initial state mean with no data available.

Figure 6.6: MSE in each grid node for {AR(1): EMMP $\boldsymbol{X}_T$} test case, compared to prediction only based on the initial state distribution. Because the initial state and the AR(1) model do not include the advection-diffusion time dynamics, the MSE is very large in some areas.



(a) Mean variance after time step T for the {AR(1): EMMP $\boldsymbol{X}_T$} test case.

(b) Sample standard deviation for the posterior variance after time step T for the {AR(1): EMMP $\boldsymbol{X}_T$} test case.

Figure 6.7: Mean and standard deviation for the posterior variance after T time steps for two test cases using the EMMP decision strategy, but different models.

These simple models are not based on solving (S)PDEs, and thus have less boundary effects in their distributions. BCs are usually chosen based on assumptions, and boundary effects are often unwelcome. Having the opportunity of making a model without the use of BCs may therefore be of interest. In a simulation study like this, however, where the incoming pollution at the west border is known, the simpler models miss out on this information. The lack of information is visible in the mean variance in Figure 6.7a. The belt with small variance close to the western border is missing, compared to the variances in Figure 6.3. The standard deviation of the variance is also quite large in this area, as Figure 6.7b shows, since the variance depends on whether the area is visited or not. The misclassification rates are bigger in this area as well, compared to the test cases based of the SPDE model; see Figure 6.8. The biggest difference between Figure 6.8 and 6.4, however, is that the misclassification rates in the middle area are larger for the AR(1) model case. In other words, the AR(1) model does not manage to predict the extent of the pollutants in

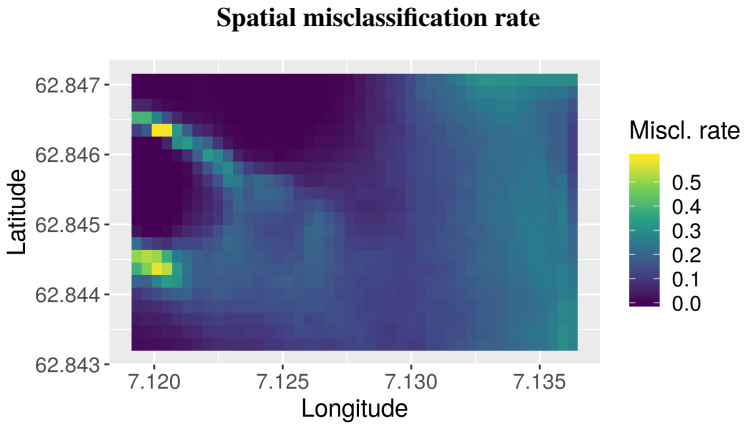


Figure 6.8: Misclassification rate in each grid node over all 1000 samples for the {AR(1): EMMP $\boldsymbol{X}_T$} test case.

the area at time T, since the model does not capture the movement over time.

Some example paths for the two test cases are shown in Figure 6.9. The AUV decision strategy is myopic EMMP, and the general features are the same as for the previous test cases. The paths are shown with the $\text{EP}_T^{(T)}$s calculated from the spatial and the AR(1) models, respectively.

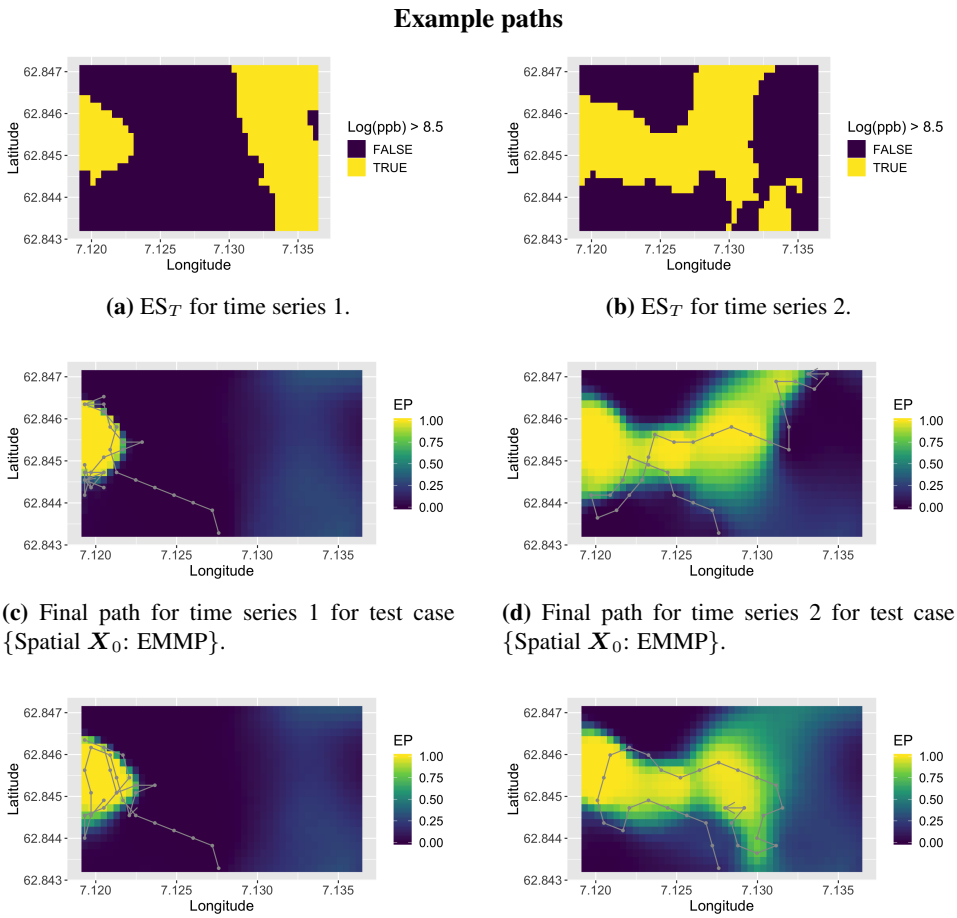
6.3 SPDE model with simpler decision strategies

Two test cases with simple, dynamic decision strategies are considered, as well as two test cases where the path is predetermined, so that no decisions are made during simulation. All these test cases use the SPDE model for the AUV; the same as the underlying model of the time series replicates. We first have a quick look at the simple dynamic strategies, test cases {SPDE: $\text{EP}_t^{(t-1)}0.5$} and {SPDE: Random}, before moving on to the predetermined paths, test cases {SPDE: Predetermined diagonals} and {SPDE: Predetermined uniform}.

Test case	Mean MMP	Mean misclass. rate	SD misclass. rate
SPDE: $\text{EP}_t^{(t-1)}0.5$	0.115	0.119	0.080
SPDE: Random	0.131	0.131	0.080

Table 6.5: Mean MMP and mean misclassification rate for two test cases using the SPDE model, but simpler decision strategies. The misclassification rate is for cutoff at probability 0.5.

Table 6.5 shows comparable MMPs and misclassification rates for the two test cases {SPDE: $\text{EP}_t^{(t-1)}0.5$} and {SPDE: Random}. Again, we see that the MMP gives a fair estimate of the misclassification rate, but since the misclassification rate is available in the simulation study, we use this for evaluation. Both test cases have larger classification error than the test cases from section 6.1, but smaller than those in section 6.2. In particular,



(a) ES_T for time series 1.

(b) ES_T for time series 2.

(c) Final path for time series 1 for test case $\{$Spatial $\boldsymbol{X}_0$: EMMP$\}$.

(d) Final path for time series 2 for test case $\{$Spatial $\boldsymbol{X}_0$: EMMP$\}$.

(e) Final path for time series 1 for test case $\{$AR(1): EMMP$\boldsymbol{X}_T\}$.

(f) Final path for time series 2 for test case $\{$AR(1): EMMP$\boldsymbol{X}_T\}$.

Figure 6.9: Final simulated paths for two sampled time series and the two test cases $\{$AR(1): EMMP $\boldsymbol{X}_T\}$ and $\{$Spatial $\boldsymbol{X}_0$: EMMP$\}$. The paths are displayed together with the corresponding $\text{EP}_T^{(T)}$s, and the ES at time step T for the two samples are also shown. The $\text{EP}_T^{(T)}$ should look as much like the ES as possible.

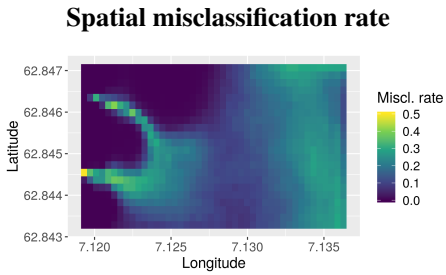


Figure 6.10: Misclassification rate in each grid node over all 1000 samples for the {SPDE: Random} test case.

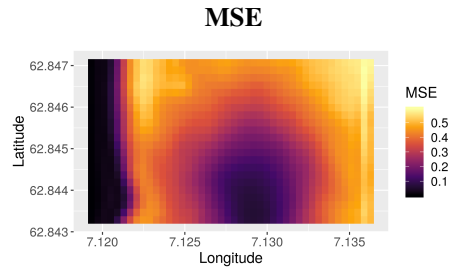


Figure 6.11: MSE for each grid node over all 1000 samples for the {SPDE: Random} test case.

we note that having the right model and choosing positions at random is better on average than the EMMP decision strategy when the model is incorrect, see Table 6.3. Spatial misclassification rate for case {SPDE: Random} is shown in Figure 6.10. Comparing to that of the {AR(1): EMMP $\boldsymbol{X}_T$} test case in Figure 6.8, we see that the misclassification rate is smaller to the far west for the {SPDE: Random}, since the model on board the AUV is correct in this case. And further, compared to not obtaining any data, as seen in Figure 6.1a, we can clearly see that misclassification rates are decreased when adding data, even if the positions are chosen at random. All so far discussed test cases have considerably lower misclassification rate than if no data is obtained, resulting in the average misclassification rate of $18.2\,\%$.

Now consider only the test cases using the SPDE model. Choosing the next position at random gives good results for comparing. All strategies trying to lower the misclassification rate should on average give lower rate than the random strategy; if not it can be disregarded at once. The test case {SPDE: Random} has $13.1\,\%$, and {SPDE: $\text{EP}_t^{(t-1)}0.5$} $11.9\,\%$, so the latter clearly has a more efficient strategy. The EMMP strategies are visibly better than both, with $10.3\,\%$ misclassification rate for test case {SPDE: EMMP $\boldsymbol{X}_T$}.

Test case	Mean MSE	SD MSE
SPDE: $\text{EP}_t^{(t-1)}0.5$	0.331	0.171
SPDE: Random	0.316	0.159

Table 6.6: Mean and standard deviation for MSE for the two test cases using the SPDE model, but simpler decision strategies.

Average MSE for test cases {SPDE: $\text{EP}_t^{(t-1)}0.5$} and {SPDE: Random} are shown in Table 6.6. These are quite similar for the two test cases, and slightly larger than for the {SPDE: EMMP $\boldsymbol{X}_T$} and {SPDE: EMMP $\boldsymbol{X}_t$} cases. Neither of the four strategies aims to minimise the MSE, to the small difference is not very surprising. The spatial MSE for test case {SPDE: Random} is displayed in Figure 6.11, and it is here clearly visible that on average, the random path moves around in a small area close to the starting position, as one expects from a spatial random walk. The MSE and the variances are again similar, since the assumptions of the model on board the AUV match the underlying true model.

Paths for the same simulated time series as in Figure 6.2 and 6.9 are shown in Figure 6.12 for test cases {SPDE: $\text{EP}_t^{(t-1)}0.5$} and {SPDE: Random}. The movements in these test cases are different from those we have seen so far, but it is also clear that the {SPDE: $\text{EP}_t^{(t-1)}0.5$} test case tries to explore the domain in a similar fashion to the EMMP strategy, while the other of course chooses steps at random. The difference between them can in particular be seen in Figure 6.12c and 6.12e.

Now consider the two test cases using predetermined paths. The paths are shown in Figure 5.2, and the key MMP and misclassification numbers are found in Table 6.7. The {SPDE: Predetermined uniform} test case sampling gives larger misclassification rate

Test case	Mean MMP	Mean misclass. rate	SD misclass. rate
SPDE: Predet. diagonals	0.095	0.097	0.059
SPDE: Predet. uniform	0.118	0.119	0.080

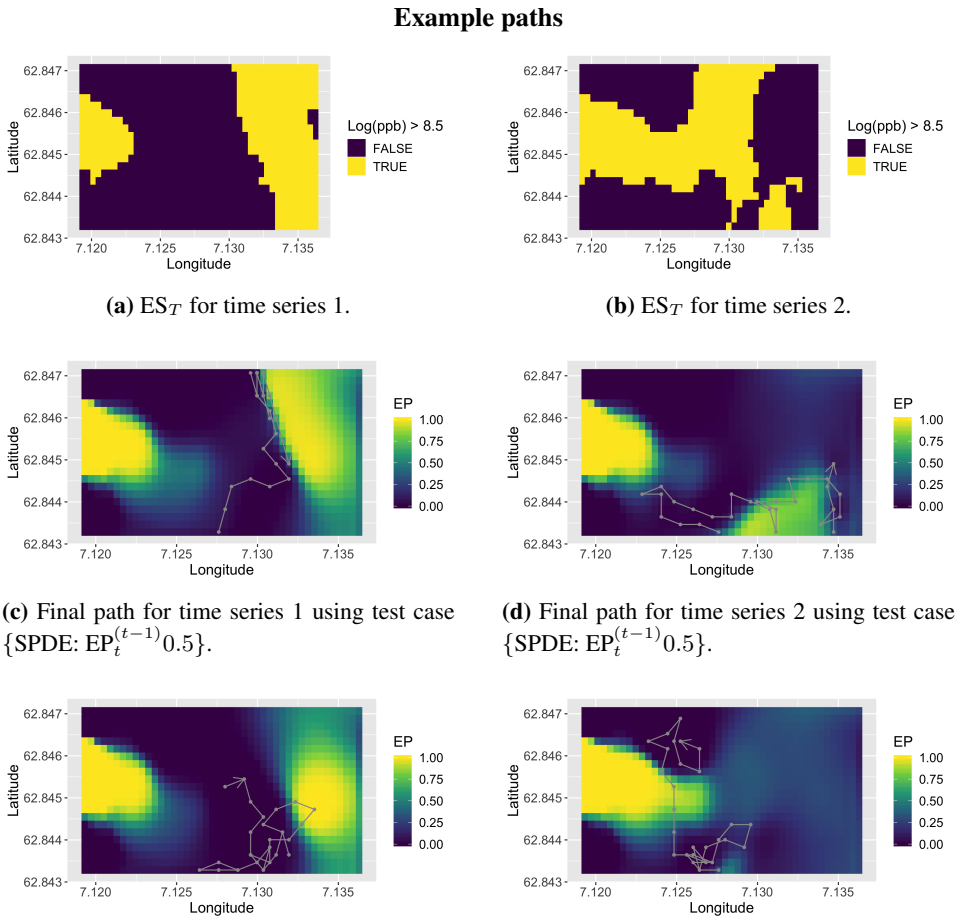
Table 6.7: Mean MMP and misclassification rate for the two test cases using the SPDE model and where the paths are predetermined. The misclassification rate is for cutoff at probability 0.5.

than the test cases{SPDE: EMMP $\boldsymbol{X}_T$} and {SPDE: EMMP $\boldsymbol{X}_t$}, while using test case {SPDE: Predetermined diagonals} results in the lowest misclassification rate on average. In this test case, the path is designed to explore as much of the domain as possible with the available restrictions of step length and time. As already discussed, the ES is often divided into separate parts, so that there are areas of high concentration both at the west and the east side of the domain. Both are rarely discovered using the myopic EMMP decision strategy. In the {SPDE: Predetermined diagonals} case, however, the big areas are usually discovered, though some details might be missing.

Spatial misclassification rate is displayed for test case {SPDE: Predetermined diagonals} in Figure 6.13. Comparing this display to the one for test case {SPDE: EMMP $\boldsymbol{X}_T$} in Figure 6.4, the main difference is that the rate is lower in the eastern area for the {SPDE: Predetermined diagonals} case. We have already seen that the details around the western high concentration front are hard to predict, and now it can be noted that even though the predetermined path explores the western area in less detail than {SPDE: EMMP $\boldsymbol{X}_T$}, the misclassification rates are very similar in this area. It seems there is not much to gain by staying in the same area for a long time, like the AUV often does using the EMMP strategy. The {SPDE: Predetermined diagonals} test case is of course helped by the western BCs, so that concentrations are known with large certainty in positions that are never visited. Figure 6.14 shows some examples of what the AUV discovers in the two test cases with predetermined paths.

The MSE for the same two test cases are listed in Table 6.8. It is smallest for test case {SPDE: Predetermined diagonals}, but also the test case {SPDE: Predetermined uniform} predicts the concentration better than the EMMP strategy test cases, on average. The EMMP criterion is developed to reduce the probability of misclassification with respect to ESs, and prediction of concentration at positions where it is certain to be above or below the critical limit of the ES does not help classification. Thus, it is as must be expected that strategies with paths covering more of the domain result in smaller MSE.

Spatial MSEs for the two test cases using predetermined paths are shown in Figure 6.15. The paths are clearly visible. Figure 6.15b reveals very few areas with high MSE for



(a) ES_T for time series 1.

(b) ES_T for time series 2.

(c) Final path for time series 1 using test case $\{\text{SPDE: EP}_t^{(t-1)}0.5\}$.

(d) Final path for time series 2 using test case $\{\text{SPDE: EP}_t^{(t-1)}0.5\}$.

(e) Final path for time series 1 using test case {SPDE: Random}.

(f) Final path for time series 2 using test case {SPDE: Random}.

Figure 6.12: Final simulated paths for two sampled time series and the two test cases $\{\text{SPDE: EP}_t^{(t-1)}0.5\}$ and {SPDE: Random}. The paths are displayed together with the corresponding $\text{EP}_T^{(T)}$s, and the ES at time step T for the two samples are also shown. The $\text{EP}_T^{(T)}$ should look as much like the ES as possible.

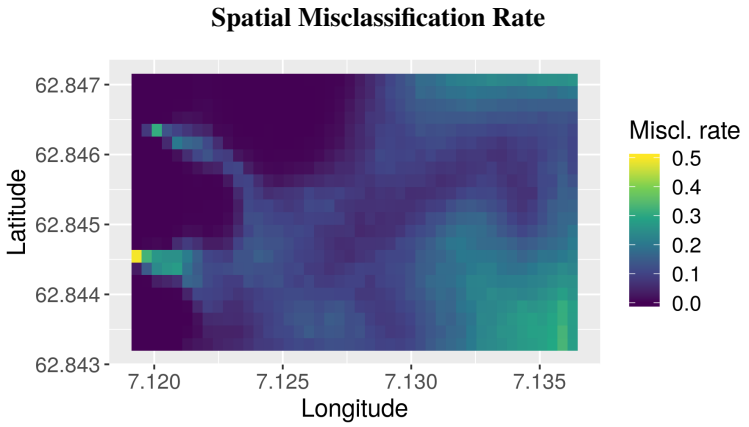


Figure 6.13: Misclassification rate for each grid node over all 1000 samples for the {SPDE: Predetermined diagonals} test case.

Test case	Mean MSE	SD MSE
SPDE: Predetermined diagonals	0.165	0.060
SPDE: Predetermined uniform	0.232	0.122

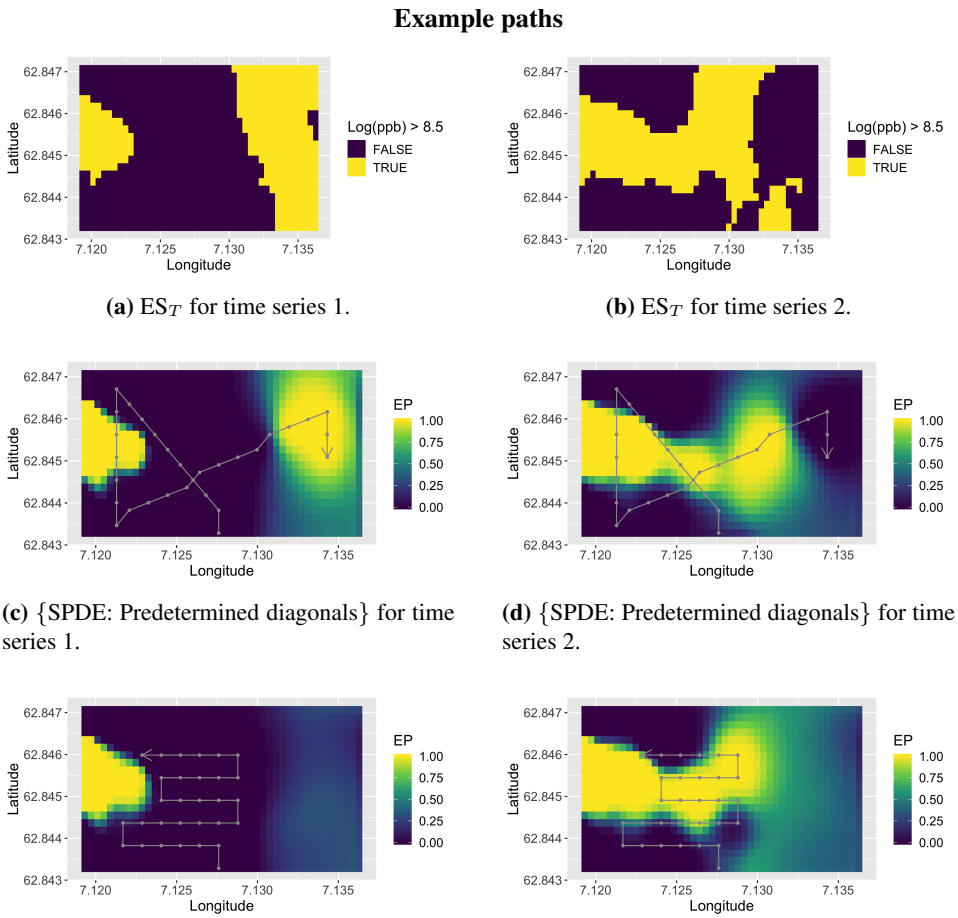
Table 6.8: Mean and standard deviation for MSE for the two test cases using the SPDE model and where the paths are predetermined.

the {SPDE: Predetermined diagonals} test case. The variance is similar. Since the path is the same in all runs and the variance is only dependent on the positions for measurements, not the measured values themselves (equation (2.49)), the variance is the same at the end of every run, and variance standard deviation is 0 in every grid node.

6.4 SPDE model with hybrid decision strategy

We have seen that for minimising misclassification rate in the Frænfjorden situation, exploration of the whole domain is important, as predetermining a path exploring big parts of the domain has so far resulted in the lowest misclassification rate, as well as the best concentration predictions (lowest MSE). Second best is using the myopic EMMP strategy. The hybrid decision strategy introduced in section 4.2.4 is developed to remedy these results, in an attempt to combine the EMMP strategy with a strategy for exploring new areas. The strategy {SPDE: Hybrid} is tested on the same 1000 time series replicates as all other test cases. The key numbers for this case are listed in Table 6.9. The mean misclassification rate is not quite small enough to beat the predetermined diagonals path, but the results are very close, and there is a clear improvement from the pure EMMP strategy. The spatial misclassification rate is shown in Figure 6.16. In the eastern part of the domain, the error is smaller than that of case {SPDE: EMMP X_T} in Figure 6.4.

The mean posterior variance in Figure 6.17a shows that the hybrid strategy often results



(a) ES_T for time series 1.

(b) ES_T for time series 2.

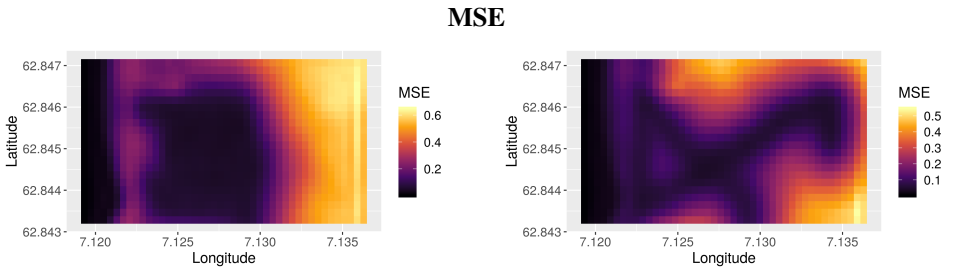
(c) {SPDE: Predetermined diagonals} for time series 1.

(d) {SPDE: Predetermined diagonals} for time series 2.

(e) {SPDE: Predetermined uniform} for time series 1.

(f) {SPDE: Predetermined uniform} for time series 2.

Figure 6.14: Final simulated paths for two sampled time series and the two test cases {SPDE: Predetermined diagonals} and {SPDE: Predetermined uniform}. The paths are displayed together with the corresponding $\mathrm{EP}_T^{(T)}$s, and the ES at time step T for the two samples are also shown. The $\mathrm{EP}_T^{(T)}$ should look as much like the ES as possible.



(a) {SPDE: Uniform}. **(b)** {SPDE: Predetermined diagonals}.

Figure 6.15: MSE for each grid node over all 1000 samples for the two test cases using the SPDE model where the path is predetermined. The shapes of the paths are clearly visible as the MSE is very small in the measured positions.

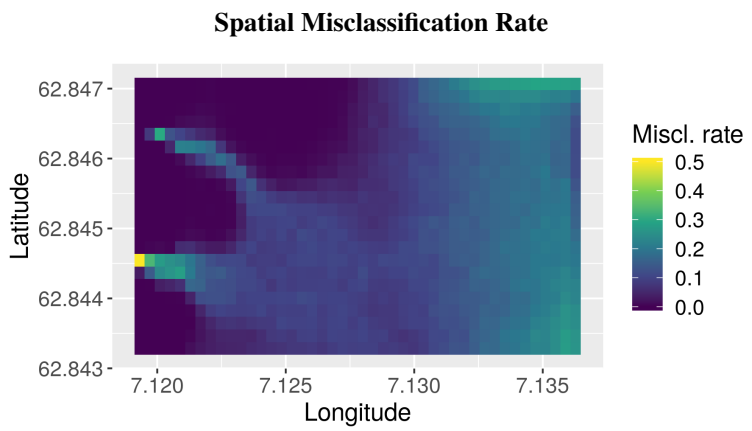
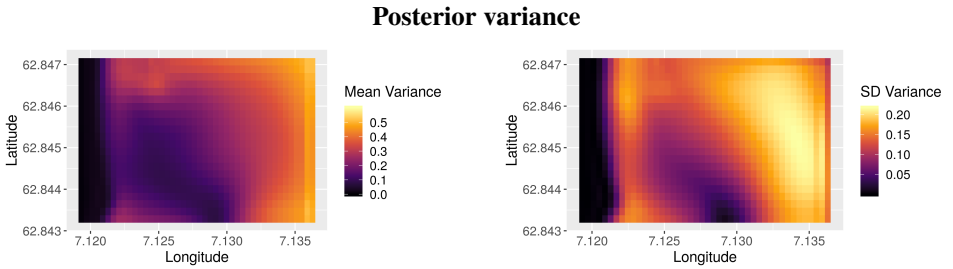


Figure 6.16: Misclassification rates for each grid node over all 1000 samples for the test case using the SPDE model and the hybrid decision strategy.

Test case	Mean MMP	Mean misclass. rate	SD misclass. rate
SPDE: Hybrid	0.097	0.099	0.069

Table 6.9: Mean MMP and misclassification rate for the test case using the SPDE model and the hybrid decision strategy. The misclassification rate is for cutoff at probability 0.5.



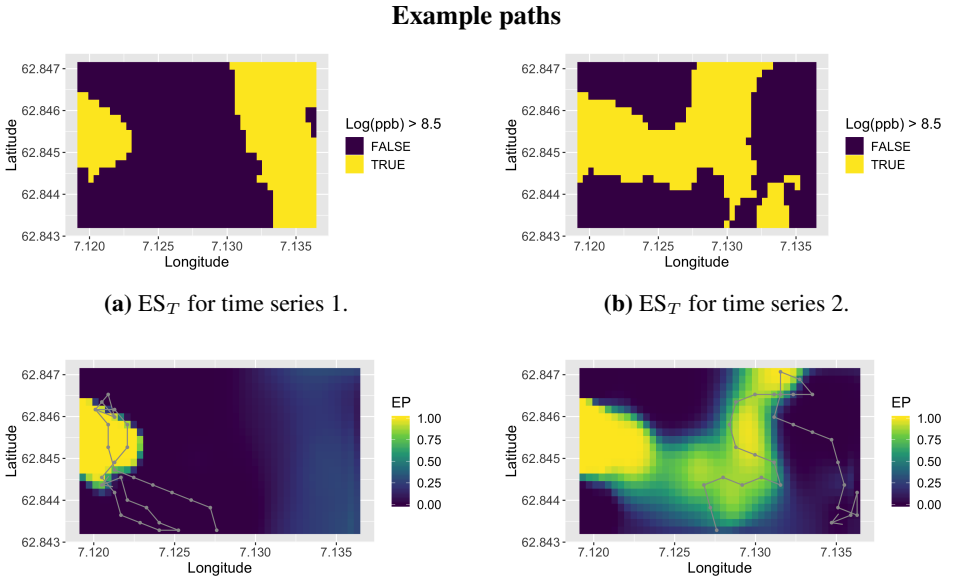
(a) Mean posterior variance.

(b) Sample standard deviation for the posterior variance.

Figure 6.17: Mean and standard deviation for the posterior variance after T time steps for the {SPDE: Hybrid} test case.

in paths going towards the west, like in the {SPDE: EMMP X_T} case, but on average the variance is lower in the north-east for the hybrid case, meaning that the AUV is more often moving into this area. The same can be observed by studying the MSE. The mean MSE is 0.224 with sample standard deviation 0.124, which is an improvement from the pure EMMP strategy test cases. The prediction error is still not quite as small as for the {SPDE: Predetermined diagonals} test case, but this hybrid strategy makes the AUV explore the domain more than does the pure EMMP strategy. In Figure 6.17b the standard deviation of the variance is displayed. The standard deviation is larger for the hybrid case than the pure EMMP cases in some areas, indicating that the hybrid strategy paths are more varying, as is expected because of the randomness and the variance criterion, as described in section 4.2.4. The example paths made from the {SPDE: Hybrid} test case in Figure 6.18 show that the pattern of moving along the ES boundary is sometimes broken, so that the AUV explores new areas. Figure 6.18c shows an example where the strategy is maybe not the most successful, while the AUV is actually quite close to discovering a new area of high concentration in Figure 6.18d.

There are also cases where the strategy works extremely well, as Figure 6.19 shows. Here, the paths for the {SPDE: EMMP X_T} and {SPDE: Hybrid} test cases are both shown, as well as the actual ES at time T. While the AUV stays in the western area using the pure EMMP strategy, it moves both east and south and discovers several areas of high pollution concentration using the hybrid strategy. Because of the randomness in the hybrid strategy, it will sometimes work really well and sometimes not. On average, we see that for this model for Frænfjorden, exploration is important, and using the hybrid criterion improves the results of identifying the ES_T, and also lowers the MSE, compared to a pure EMMP strategy.



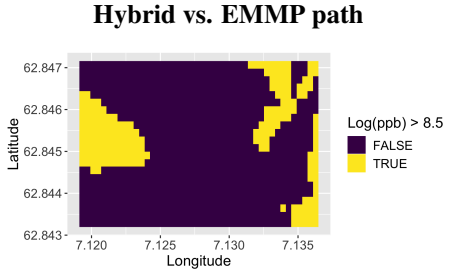
(a) ES_T for time series 1.

(b) ES_T for time series 2.

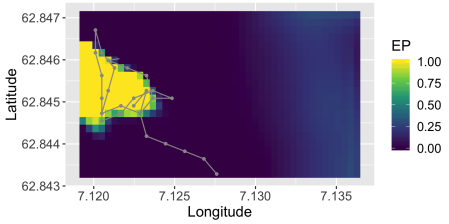
(c) Final path for time series 1 using test case {SPDE: Hybrid}.

(d) Final path for time series 2 using test case {SPDE: Hybrid}.

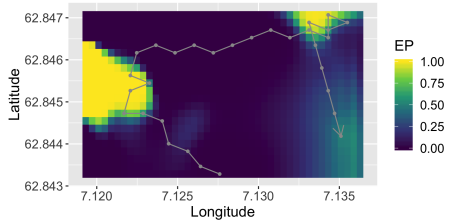
Figure 6.18: Final simulated paths for two sampled time series and the test case {SPDE: Hybrid}. The paths are displayed together with the corresponding $\mathrm{EP}_T^{(T)}$s, and the ES at time step T for the two samples are also shown. The $\mathrm{EP}_T^{(T)}$ should look as much like the ES as possible.



(a) ES_T.



(b) {SPDE: EMMP X_T}.



(c) {SPDE: Hybrid}.

Figure 6.19: Final paths for a time series replicate where the hybrid decision strategy leads the AUV to discover several high concentration areas that it misses using the pure EMMP strategy. The paths are shown together with the respective $\mathrm{EP}_T^{(T)}$s, which should look as much as the displayed ES_T as possible.

Chapter 7

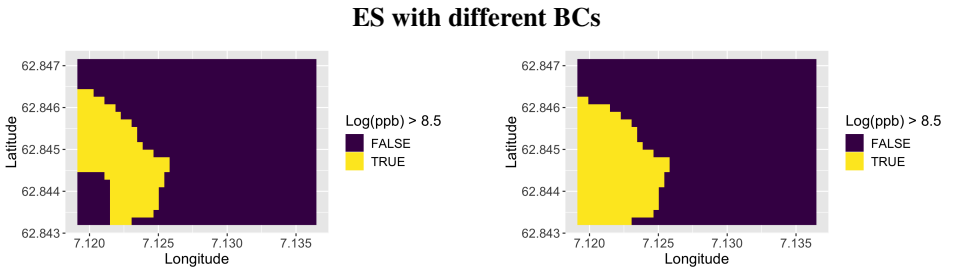
Further discussion

In chapter 6, we found that having the correct SPDE model is important for the classification results. The AR(1) and the spatial model with EMMP decision criterion result in larger misclassification rates than choosing positions at random using the SPDE model. The AR(1) is performing slightly better than the pure spatial model, since the AR(1) model includes some temporal variation. The test cases using the SPDE model with EMMP decision criterion perform fairly well, and better than more naive strategies, but the AUV often fails to explore the domain enough to find all areas of high concentration. Because of this, the predetermined path exploring a big fraction of the domain has the lowest misclassification rate, though a hybrid strategy using the EMMP and a variance criterion has almost equal performance.

For further discussion, some other simulation studies are performed and presented in this chapter. The results of these lead to discussions of the Frænfjorden model and the performance of the different strategies. BCs and model assumptions are discussed first, before moving on to the sensitivity of the input model parameters, and at the end there are some remarks about the pure spatial model and starting position for the AUV paths.

7.1 Boundary conditions and model assumptions

A clear challenge in the building of the Frænfjorden model is to choose realistic BCs, as discussed in sections 2.2.1 and 3.1.2. The largest effects we have seen are from the western boundary, where Dirichlet BCs are used. We have perfect data along the boundary, and this propagates into the area through the process model and advection field. Over time, the concentrations in the area to the far west of the domain are almost certain, as we have seen several times, e.g. in Figure 5.1d. Such perfect information is not realistic in practice, but allows us to have a perfect model in the simulation study, so that the results are focused on different strategies for the AUV. The initial state model expected values for the grid nodes farthest to the west are used for the Dirichlet BCs. A consequence in the simulation of time series replicates is that the transit from the initial state sample concentrations to the concentrations on the boundary may be quite sharp, and there is sometimes a sort of border between the concentration of pollutants already in the domain at initial time and



(a) Initial state *mean* concentrations along the western boundary.

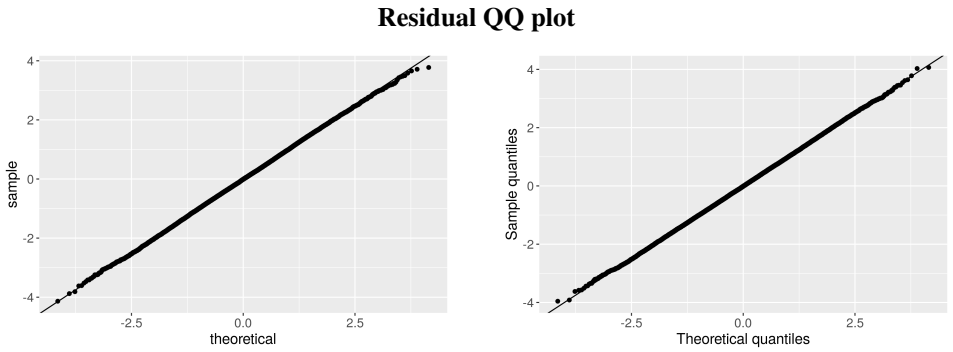
(b) Initial state *sample* concentration along the western boundary.

Figure 7.1: ES_T from one simulated time series with two different BCs.

the incoming pollution as time passes. This means that some samples lack the smoothness seen in the Frænfjorden data. An example is displayed in Figure 7.1a. In this ES it is very easy to see the limit between initial pollution and where the concentration comes from the boundary, and it does not look very realistic. Such a hard transit is fortunately not very common.

A new study is performed on time series replicates simulated in a slightly different way. To make the samples somewhat more realistic, the BCs are set to the initial concentration for the respective initial state sample. The ES in Figure 7.1b shows how the new BCs follow the initial state sample, so that the transit is smooth. The model on board the AUV still uses initial mean along the boundary. This leads to unfortunate results in the western area, as the SPDE model on board the AUV now does not match the underlying model of the simulated time series in this area. This simulation study shows some of the things that may happen if one uses Dirichlet BCs without really having this information available, and the situation may thus be closer to what one would experience when dealing with real data.

The AUV paths are not the same in the two studies, but the differences are not substantial. What makes the AUV choose different positions are the different observed values in the western area. Otherwise everything is the same from the AUV test case perspective. However, the model on board the AUV reflects a false certainty about the concentration close to the western boundary, since the incoming concentration according to the model is not the actual incoming concentration. Therefore, the variance is too small in this area, and concentration predictions and classifications are not corrected by new data. The misclassification rate is therefore overall higher, and the MMP is not a good estimate for the classification error. The MMP is calculated from the model on board the AUV, and is therefore very similar in the two simulation studies, while the misclassification rate of e.g. the {SPDE: EMMP X_T} test case is on average $11.7\ \%$ in the new study, from $10.3\ \%$ when the BCs are correct. Likewise, test case {SPDE: Random} and {SPDE: Predetermined diagonals} get misclassification rates $14.6\ \%$ and $11.1\ \%$, respectively. The spatial MSE is large in the western area, and does not have the same shape as the model variance. The increase in mean MSE for test case {SPDE: EMMP X_T} is from 0.273 to 0.338, test case {SPDE: Random} has MSE 0.390 and {SPDE: Predetermined diagonals} 0.231. The



(a) {SPDE: EMMP X_T}. **(b)** {SPDE: Hybrid}.

Figure 7.2: QQ-plots for the standardised measurement residuals for test cases {SPDE: EMMP X_T} and {SPDE: Hybrid} of the original simulation study. The residuals are Gaussian distributed, and the theoretical and sample quantiles match almost perfectly. Each display shows 30000 residuals.

benefit of using the SPDE model compared to the AR(1) or spatial model is smaller when the BCs are incorrect. The AR(1) and spatial model do not have any BCs, and the misclassification rates are not changed much for the test cases using these models. For the spatial test case, the mean misclassification rate is now 13.6%, which is actually a small decrease. The constant, smooth, incoming pollution might be easier to predict for the cases without BCs.

One way of checking model assumptions is to study the residuals between measured values and their predicted values, which should be Gaussian distributed with mean zero if assumptions hold. The predicted value for the measurement is the mean of $Y_t \mid \boldsymbol{Y}_{1:t-1}$, as given in equation (2.44), and divided by the variance in the same equation, the residuals should be standard Gaussian distributed. Since this is a simulation study, we already know that everything should be Gaussian. One can check the Gaussian assumption by making a QQ-plot (quantile-quantile-plot). Figure 7.2 shows QQ-plots with the standardised residuals from test cases {SPDE: EMMP X_T} and {SPDE: Hybrid} from the original simulation study discussed in chapters 5 and 6. The standardised residuals from a standard Gaussian distribution should, when sorted and plotted against theoretical residuals, approximately follow the straight line, and we see that the result is almost perfect here. With 30 residuals for each time series replicate, and 1000 (or 999) replicates, this is not surprising. For one single time series or with fewer samples, there will of course be some deviation from the line even when all model assumptions are correct.

When simulating with incorrect BCs, however, the QQ-plots become as shown in Figure 7.3. There is clear displacement from the straight line at the ends. The displacement tells that the sample quantiles are more extreme than the theoretical quantiles, suggesting that the sample residuals come from a distribution with heavier tails than the standard Gaussian, so that the residuals become more extreme. What happens is that due to the incorrectly small variance from the western boundary, where the BCs give a false certainty, the residuals from measurements near this boundary are too big. In other words, we can clearly see the result of having different models for the simulated time series replicates

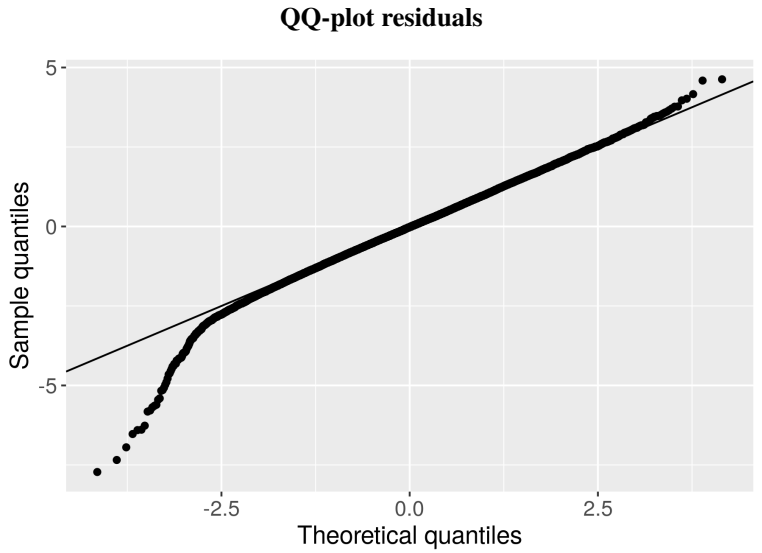
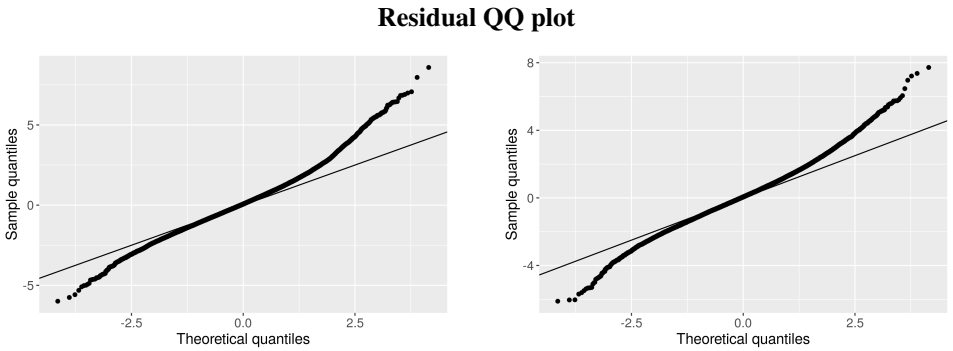


Figure 7.3: QQ-plot for the standardised measurement residuals of the study with different BCs for test case {SPDE: EMMP $\boldsymbol{X}_T$}. The sorted residuals deviate from the straight line at the ends, indicating heavier tails on the sample distribution.

and the AUV sampling, even though the difference is only in the BCs. For the test cases using the AR(1) model and the spatial model, the measurements are not drawn from the distribution used by the on-board model, and the residuals after subtracting mean and dividing by the variance in the model, are therefore not standard Gaussian in the original simulation study either. We can see this in Figure 7.4. The deviation from the line is large, with more extreme residuals than one expects in the standard Gaussian distribution. It is easy to see that the model assumptions in these cases do not match the (simulated) reality. Doing some diagnostic plotting of the residuals can therefore be a good way of evaluating model assumptions.

We have seen that the misclassification rate in the simulation study for Frænfjorden is affected by the BCs. In this model, there is incoming pollution of high concentration, so the BCs greatly affect the dynamics in the model and the resulting ES. The domain was chosen because there seemed to be interesting pollutant dynamics. In hindsight, for a simulation study it would be a better idea to choose a domain where most of the pollution stays inside the domain during the relevant time period, as the model would then not be so dependent on the boundary. It is safer to choose a domain for the model where one does not expect anything interesting to happen along the boundary, and which is large enough for boundary effect not to affect the area of interest significantly. However, the model in this thesis is based on a real situation, and such are rarely ideal.



(a) {Spatial X_0: EMMP}. **(b)** {AR(1): EMMP X_T}.

Figure 7.4: QQ-plots for the standardised measurement residuals for test cases {Spatial X_0: EMMP} and {AR(1): EMMP X_T} of the original simulation study. The sorted residuals do not follow the straight line, indicating that the model assumptions are not fulfilled and that the tails of the sample distributions are too heavy.

7.2 Effect of parameter change

To test model sensitivity, all test cases are run on 100 time series replicates for higher and lower advection speed, as described in section 3.2. We use the same seeds for the randomness, so that the initial state and innovation (noise) samples are the same as for the 100 first sample in the original study, and the only difference in the time series is essentially related to the changed advection parameter. The aim is to see how the results of the test cases depend on the parameters in the model. The same advection is used both when simulating the time series and in the model on board the AUV for the test cases using the SPDE model. The AR(1) parameter ϕ is not straightforward to update corresponding to the increased or decreased drift field, so for this test case, the same parameter is used for the on-board model. Since only 100 replicates are used, there is more uncertainty in the results for these simulation studies, but some general tendencies are discussed in this section.

In general, the change in misclassification rate when varying the advection slightly is, at least partly, due to the effect of the BCs in the western area. When the speed of the drift field is larger, the known concentration on the boundary reaches further into the domain, and the misclassification rate is lower. Approximately two columns more from the west in the grid are affected by the BCs for high advection than in the original Frænfjorden simulation study, and likewise two columns less when the advection speed is low. The numerical boundary effects along the eastern boundary are also larger for high advection, resulting in larger variance in this area. This is as expected, since this means the speed of the advection towards this boundary is larger, and the condition of no flow across the boundary less realistic and harder to fulfill numerically. In the rest of the domain, the model variance does not vary significantly for the three models.

Because of the boundary effects, some interesting features might be lost, as it is hard to know exactly how much of the change is due to the BCs. There are still some tendencies

of change between the different test cases. In particular, we may study what happens to the test cases {Spatial $\boldsymbol{X}_0$: EMMP} and {AR(1): EMMP $\boldsymbol{X}_T$}. While the test cases using the SPDE model have higher misclassification rates with low advection and lower for high advection, the results are opposite for these two test cases. As one would expect, when the drift speed is small, the error of not including time dynamics is smaller than when the drift speed is large. The test cases with these simpler models for low advection still do not perform as well as the EMMP strategy with SPDE model, but the difference is significantly lower. The {Spatial $\boldsymbol{X}_0$: EMMP} test case now has about the same performance as the {SPDE: $\text{EP}_t^{(t-1)}0.5$ } test case, with misclassification rates of $12.8\,\%$ and $12.7\,\%$, respectively. The {SPDE: Random} case now has a misclassification rate of $13.6\,\%$ on average. The {SPDE: EMMP $\boldsymbol{X}_t$} and {SPDE: EMMP $\boldsymbol{X}_T$} test cases both have misclassification rates of about $11\,\%$ when advection speed is low.

There is a slight indication that the test case {SPDE: Predetermined diagonals}, which has given the best results so far, is not outperforming the {SPDE: Hybrid} test case, at least when parameters are changed. The predetermined path is designed to move across the areas with the most prior uncertainty with respect to the ES, i.e. where the original $\text{EP}_T^{(0)}$ is close to 0.5. When the advection speed is changed, the positions no longer capture the same features. When drift speed is low, the two cases both have a misclassification rate of $10.7\,\%$. For high advection, the test case {SPDE: Hybrid} has average misclassification rate of $9.4\,\%$, which is lower than the $9.6\,\%$ of the {SPDE: Predetermined diagonals} case. The difference is not very big, and with a sample size of only 100, we can only conclude that the performance of {SPDE: Hybrid} and {SPDE: Predetermined diagonals} are quite similar. For high advection, the {SPDE: EMMP $\boldsymbol{X}_t$} and {SPDE: EMMP $\boldsymbol{X}_T$} test cases have misclassification rates of $9.9\,\%$ and $10.1\,\%$, respectively. In general, the internal variations between the different test cases are small. It seems, therefore, that the performance of the different decision strategies are not dependent on the exact parameter configuration used in this thesis.

The test cases are also run with smaller measurement variance τ^2 on 100 original time series replicates. In this case, we use $\tau = 0.05$. The results are, unsurprising, improved for all test cases. The internal order between the test cases are still mostly the same, but again the hybrid test case has lower mean misclassification rate than the predetermined diagonals test case. The mean misclassification rates for these 100 time series are $8.4\,\%$ for test case {SPDE: Predetermined diagonals}, 8.1% for test case {SPDE: Hybrid}, 8.8% for {SPDE: EMMP X_T} and $8.9\,\%$ for {SPDE: EMMP X_t}.

7.3 Other remarks

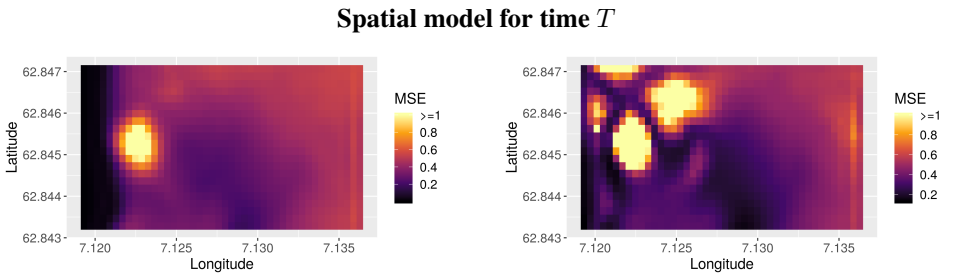
The pure spatial model considered in this thesis is the initial state model for Frænfjorden. We have seen that the test case using this model does not perform as well as the cases using the SPDE model that includes temporal changes. It does not capture the movement of the pollution, and does not have information about the concentration along the boundary provided by the BCs. What if we, instead of using the model for the initial time step, consider a spatial model based on the prior knowledge of the state at time T? Would the spatial model then predict ES_T more accurately?

As a test, the initial state model is forecasted $T = 30$ steps using the SPDE model, and

this forecasted distribution of $\boldsymbol{X}_T$ is used in a new spatial test case, denoted {Spatial X_T: EMMP}. Paths are simulated for this test case on the original 1000 time series replicates. The key numbers we then get are mean MMP 0.107, mean misclassification rate 0.126 and mean MSE 0.362. For this new spatial model, there are still no time dynamics included in the model on board the AUV, but the boundary effects are correct at time step T and the spatial model is closer to the reality at time T than when considering the initial state model. Therefore, both the misclassification rate and the MSE are improved quite a bit, as can be seen by comparing to the results in Table 6.3 and 6.4.

The results are still quite far from the results of test cases {SPDE: EMMP X_T} and {SPDE: EMMP X_t}, and even further from {SPDE: Hybrid} and {SPDE: Predetermined diagonals}. This shows that the SPDE model is not better only because of the much discussed BCs. Including the time dynamics in the on-board model really is the best approach. Figure 7.5a shows the spatial MSE from test case {Spatial X_T: EMMP}. We see that the concentration prediction is now very accurate close to the western boundary, and compared to the MSE from test case {Spatial X_0: EMMP} in Figure 7.5b, most of the areas with really bad predictions are gone. Nevertheless, it is clear from the QQ-plot of residuals in Figure 7.5c that model assumptions are still not correct. A pure spatial model can perhaps be considered acceptable if the advection speed is very small or the time period in question very little. Especially if good BCs are in addition proving hard to find, spatial models are convenient, since choosing bad BCs may lead to unfortunate results, as discussed in section 7.1.

Starting positions are not studied thoroughly in this thesis, but the AUV paths seem to be sensitive to starting position, as it tends to get stuck along the boundary of the part of the ES that it discovers first. If there is only one connected ES, the AUV moves around the whole boundary a lot more, and there is reason to believe that the starting position is not very important in these cases. Also, the gain of thoroughly exploring the western ES front is quite small in this simulation study, as a big part of that area is well-known without any additional data because of the Dirichlet BCs. In other scenarios, one may expect the EMMP decision strategy to be beneficial compared to paths such as the predetermined diagonals, which explores big parts of the domain, but may then fail to discover details along the ES borders.



(a) MSE in each grid node for the {Spatial X_T: EMMP} test case.

(b) MSE in each grid node for the {Spatial X_0: EMMP} test case.

(c) QQ-plot for the standardised measurement residuals for test case {Spatial X_T: EMMP}. The tails of the sample distribution are too heavy.

Figure 7.5: MSE and residuals QQ-plot for the test case {Spatial X_T: EMMP}. MSE for the test case {Spatial X_0: EMMP} is displayed for comparison.

Chapter 8

Conclusion

In this thesis, spatio-temporal GPs are used to model the flow of contaminant concentration in Frænfjorden, Norway. The advection-diffusion SPDE is solved by finite differences to create the model, and parameters are found from the numerical data models DREAM and SINMOD. Further, simulations of different strategies for AUV environmental sensing are explored for the established model, with the aim of gaining information on the concentration ES after T time steps, ES_T. A myopic decision strategy based on the probability of misclassification, EMMP, is derived, and compared with more naive strategies. Using the EMMP strategy with the forecasted distribution of $\boldsymbol{X}_T$ is quite computer demanding.

Letting the AUV use the SPDE model, EMMP strategies are performing better than other naive dynamic strategies in predicting the ES. Choosing positions at random results in an average misclassification rate of $13.1\ \%$, and using the EMMP criterion reduces it to 10.3-$10.5\ \%$. If a spatial model or a spatial AR(1) time series model is used on board the AUV, however, the predictions are not as good, and the mean misclassification rate is approximately $14\ \%$. It seems that even over a small time period like 30 minutes, it is important to consider time dynamics. The simpler models perform better when the advection speed is lower, as the change in time is then smaller. Simulating from the Frænfjorden model, there are often several areas of high concentration in the area, so that the ES_T is split in several disconnected parts. The test cases using the EMMP strategy efficiently finds one part, but often fails to explore the rest of the domain. Because of this, predetermining a path exploring a large part of the domain often predicts the ES_T better, with an average misclassification rate of $9.7\ \%$. Some initial work is done on finding an adaptive strategy that explores greater parts of the domain. A hybrid strategy is introduced, where the position for the next step is sometimes chosen based on reduction in variance instead of the EMMP, and though the results are not better than the predetermined diagonals test case, it seems to have potential. Using the hybrid strategy, the mean misclassification rate is $9.9\ \%$ in the simulation study. The test cases exploring the most of the domain are also seen to have the lowest MSEs. In practice, the misclassification rate and the MSE are not available, since the true concentrations are unknown. However, we have seen that if the model assumptions hold, these are on average very similar to the MMP and the posterior model variances, respectively. In the developed model, the BCs on the western side of the domain have large effects on the concentrations near the boundary and also on the results,

as they affect the ES in this part of the domain. When perfect information is not available, as it rarely is in practice, one should, if possible, choose the domain so that the features of interest are not greatly affected by the boundary.

As we have seen that correct time dynamics are important for the results, future work should include improving the model for Frænfjorden. The advection field is considered constant in this thesis, but letting the field vary in time, or considering anisotropy in the diffusion, are some suggestions for a more nuanced model. The finite differences scheme could be improved, and more accurate model parameters could perhaps also be found by e.g. maximum likelihood estimation. Further work on improving the EMMP strategy is needed, and testing it in different situations could be interesting. Though the AUV tends to get stuck in one part of the domain in the situation where several areas have high concentrations, it seems to be a good strategy if there is only one ES border to be found. So far, forecasting to time T to use the distribution of $\boldsymbol{X}_T$ is not seen to have any advantage over using the distribution $\boldsymbol{X}_t$ when making decisions for time t. This should, however, be tested in situations where there is only one area of high concentration. For the EMMP strategy to be feasible to use on board an AUV, some improvements on the computations of the EMMP criterion are needed. This is in particular important if the distribution of $\boldsymbol{X}_T$ is to be used. Future work for Frænfjorden should also include more research on the hybrid decision strategy. Some tuning of parameters is needed to balance the two decision strategies for better performance. Adaptive criteria are dynamic and may discover features unforeseen by the model or explore areas in greater detail. These are advantages over predetermined paths, where one only has prior knowledge when making the decisions. Nevertheless, as it is seen that the predetermined paths are actually performing quite well, further work could also exploit this knowledge, since live computations from the AUV are then not required. Simple AUV strategies may in the future be combined with other environmental sensing techniques to optimise ocean monitoring.

Bibliography

Adler, R. J. (2000). On excursion sets, tube formulas and maxima of random fields. *The Annals of Applied Probability*, 10(1):1–74.

Angulo, J. M. and Madrid, A. E. (2010). Structural analysis of spatio-temporal threshold exceedances. *Environmetrics*, 21:415–438.

AUR lab (n.d.). *LAUV Harald*. NTNU. `https://www.ntnu.edu/aur-lab/lauv-harald`.

Azzimonti, D., Ginsbourger, D., Chevalier, C., Bect, J., and Y., R. (2015). Excursion and contour uncertainty regions for latent Gaussian models. *Journal of the Royal Statistical Society*, 77(1):85–106.

Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vazquez, E. (2012). Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22(3):773–793.

Berget, G. E., Fossum, T. O., Johansen, T. A., Eidsvik, J., and Rajan, K. (2018). Adaptive sampling of ocean processes using an AUV with a Gaussian proxy model. *IFAC-PapersOnLine*, 51(29):238–243.

Bhattacharjya, D., Eidsvik, J., and Mukerji, T. (2013). The value of information in portfolio problems with dependent projects. *Decision Analysis*, 10(4):341–351.

Bolin, D. and Lindgren, F. (2015). Excursion and contour uncertainty regions for latent Gaussian models. *Journal of the Royal Statistical Society: Series B*, 77(1):85–106.

Bolin, D. and Lindgren, F. (2018). Calculating probabilistic excursion sets and related quantities using excursions. *Journal of Statistical Software*, 86(5).

Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Brooks/Cole, Cenage Learning, 2 edition.

Cetinić, I., Toro-Farmer, G., Ragan, M., Oberg, C., and Jones, B. H. (2009). Calibration procedure for Slocum Glider deployed optical instruments. *Optics express*, 17(18):15420–15430.

Chevalier, C., Bect, J., Ginsbourger, D., Vazquez, E., Picheny, V., and Richet, Y. (2013). Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. *Technometrics*, 56(4):455–465.

Cressie, N. A. C. and Wikle, C. K. (2011). *Statistics for Spatio-temporal data*. Wiley.

Eco 3 (2007). *Eco 3-Measurement Sensor; User's Guide*. WET Labs, Inc.

Eidsvik, J., Mukerji, T., and Bhattacharjya, D. (2015). *Value of Information in the Earth Sciences*. Cambridge University Press.

Fossum, T. O., Eidsvik, J., Ellingsen, I., Alver, M. O., Fragoso, G. M., Johnsen, G., Mendes, R., Ludvigsen, M., and Rajan, K. (2018). Information-driven robotic sampling in the coastal ocean. *Journal of Field Robotics*, 35(7):1101–1121.

French, J. P. and Hoeting, J. A. (2016). Credible regions for exceedance sets of geostatistical data. *Environmetrics*, 27(1):4–14.

French, J. P. and Sain, S. R. (2013). Spatio-temporal exceedance locations and confidence regions. *The Annals of Applied Probability*, 7(3):1421–1449.

Garvik, O. (2017). Gruvekonflikten i Førdefjorden. *Snl.no*. `https://snl.no/Gruvekonflikten_i_F%C3%B8rdefjorden`.

Glette, T. (2019). Miljøovervåking i Frænfjorden 2018. Environmental Risk Management 2019-0219, DNV-GL.

Hartikainen, J., Riihimäki, J., and Särkkä, S. (2011). Sparse spatio-temporal Gaussian processes with general likelihoods. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 6791, page 193–200.

Hyun, J. W., Li, Y., Huang, C., Styner, M., Lin, W., and Zhu, H. (2016). STGP: Spatio-temporal Gaussian process models for longitudinal neuroimaging data. *NeuroImage*, 134:550–562.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer.

Length of degree (n.d.). *Length of a Degree of Latitude and Longitude*. National Geospatial-Intelligence Agency. `https://msi.nga.mil/MSISiteContent/StaticFiles/Calculators/degree.html`.

Liu, H., Hitchcock, D. B., and Samadi, S. Z. (2019). Spatial and spatio-temporal analysis of precipitation data from South Carolina. In *Modern Statistical Methods for Spatial and Multivariate Data*. Springer International Publisher.

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. MIT Press.

Richardson, R. A. (2017). Sparsity in nonlinear dynamic spatiotemporal models using implied advection. *Environmetrics*, 28(6).

Rogowski, P., Terrill, E., Otero, M., Hazard, L., and Middleton, W. (2013). Ocean outfall plume characterization using an autonomous underwater vehicle. *Water Science and Technology*, 67(4):925–933.

Rye, H., Reed, M., Frost, T. K., Smit, M. G. D., Durgut, I., Johansen, O., and Ditlevsen, M. K. (2008). Development of a numerical model for calculating exposure to toxic and nontoxic stressors in the water column and sediment from drilling discharges. *Integrated Environmental Assessment and Management*, 4(2):194–203.

Setså, R. (2019). Et kvantesprang for sjødeponi. *geoforskning.no*. https://www.geoforskning.no/nyheter/bergverk/2058-et-kvantesprang-for-sjodeponi.

Sigrist, F., Künsch, H. R., and Stahel, W. A. (2015a). spate: An R package for spatio-temporal modeling with a stochastic advection-diffusion process. *Journal of statistical Software*, 63(14):1–23.

Sigrist, F., Künsch, H. R., and Stahel, W. A. (2015b). Stochastic partial differential equation based modelling of large space-time data sets. *Journal of the Royal Statistical Society Series B*, 77(1):3–33.

Slagstad, D. and McClimans, T. A. (2005). Modeling the ecosystem dynamics of the barents sea including the marginal ice zone: I. physical and chemical oceanography. *Journal of Marine Systems*, 58(1):1–18.

Sommerfeld, M., Sain, S., and Schwartzman, A. (2018). Confidence regions for spatial excursion sets from repeated random field observations, with an application to climate. *Journal of the American Statistical Association*, 113(523).

Särkkä, S. (2013). *Bayesian Filtering and Smoothing*. Cambridge University Press.

Verstad, A. B., Rostad, I. L., and Tronsen, J. H. (2019). Avviste klage fra naturvernerne – nå kan Nussir starte opp driften. *Nrk.no*. https://www.nrk.no/sapmi/nussir-med-driftskonsesjon-_-na-kan-de-starte-opp-driften-1.14801021.

Zhang, S. and Yu, A. (2013). Forgetful Bayes and myopic planning: Human learning and decision-making in a bandit setting. In *Advances in Neural Information Processing Systems 26*. Neural information processing systems foundation.