Master's thesis

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

Karina Lilleborge

Multivariate Spatial Modeling using SPDEs with Application to Ocean Sampling

Master's thesis in Applied Physics and Mathematics
Supervisor: Geir-Arne Fuglstad
June 2022

NTNU
Norwegian University of
Science and Technology

Karina Lilleborge

Multivariate Spatial Modeling using SPDEs with Application to Ocean Sampling

Master's thesis in Applied Physics and Mathematics
Supervisor: Geir-Arne Fuglstad
June 2022

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

NTNU
Norwegian University of
Science and Technology

Preface

This thesis is my master thesis symbolizing the end of a five-year study program in Applied Physics and Mathematics with a specialization in Industrial mathematics. My supervisor for the thesis was Geir-Arne Fuglstad, whom is the first person I would like to give my acknowledgement to. I have appreciated the constant support and helpful discussions throughout the thesis work. I am also thankful for the encouragement to apply for PhD.-positions, which resulted in a change in my plans for the future and four new years at NTNU in Trondheim.

Secondly, I would like to thank family and friends who have motivated me, cheering me on and been there for me when I needed them. A special thanks to my parents who have been great support when I was facing adversity due to stress or being “stuck”.

Sammendrag

I denne masteroppgaven vil vi foreslå multivariate romlige modeller for effektiv prediksjon av temperatur og salinitet i havet. Vi baserer modelleringen og analysen på modelløsninger fra en numerisk havmodell, SINMOD, utviklet av SINTEF siden 1987, og ekte data samlet inn i forbindelse med Maritime Autonomous Sampling and Control (MASCOT)-prosjektet ved bruk av en undervannsrobot.

Vi foreslår en Bayesiansk tilnærming for modelleringen, og bruk av systemer av lineære stokastiske partielle differensiallikninger (SPDE) blir brukt for bivariat beskrivelse av romlig avhengighet. Vi vil sammenlikne en univariat modell og to ulike bivariate modeller med utgangspunkt i å lære fra den numeriske havmodellen og på prediksjonsevne på ekte data.

De tre modellene presterer likt når vi sammenlikner deres evne til å lære fra den numeriske modellen. Sammenlikningen baserer seg på resultater fra "5-fold" kryss-validering og "hold-out region" kryss-validering, der mean square error (MSE) og continuous ranked probability score (CRPS) blir brukt som sammenlikningsgrunnlag. Basert på prediksjoner på data fra virkeligheten, finner vi at univariat modell presterer best. Dette blir målt ved å sammenlikne MSE og mean CRPS for alle tre modeller. Vi anbefaler videre arbeid med å formulere gyldige statistiske modeller for multivariat romlig modellering ved bruk av systemer av lineære SPDE-er.

Abstract

This thesis suggests multivariate spatial models for effective modelling of temperature and salinity in the ocean. We base our modeling and analysis on forecasts from a numerical ocean model, SINMOD, developed at SINTEF since 1987, and real data collected with an AUV through the Maritime Autonomous Sampling and Control (MASCOT) project.

We suggest a Bayesian approach for the modeling, and applying systems of linear stochastic partial differential equations (SPDEs) for the bivariate modeling of spatial dependency. We compare one univariate and two different bivariate models with regards to learning from the numerical ocean model and prediction on real data.

The three models perform equally when learning from the numerical model output when testing the models through 5-fold cross-validation and hold-out region cross-validation. The scoring rules are mean square error (MSE) and mean continuous ranked probability score (CRPS). Based off prediction on real data, we find that independent models perform the best when comparing MSE and mean CRPS for all three models. Further work in developing valid statistical models for multivariate spatial modeling using systems of linear SPDEs is encouraged.

Contents

Chapter 1

Introduction

Spatial modeling plays an important role in various scientific fields such as geology, meteorology and oceanography (Cressie and Wikle 2011; Gelfand et al. 2010). The objective of spatial modeling is to describe the spatial dependency within and between variables. By describing such spatial dependency, one gain a better understanding of the variable and obtain predictions at unobserved locations. Even if there is one variable of interest, we often collect data for more than one variable. By applying multivariate spatial modeling, we are able to use other variables as support. With multivariate spatial models, the cross-dependency between variables is included in addition to the dependency within each variable.

One application in multivariate spatial modeling is ocean modeling. Ocean data collection has previously been based mainly on static buoys, floats and ship-based methods, but in recent years autonomous underwater vehicles (AUVs) have become more prominent (Fossum et al. 2021). Several variables are of interest in ocean modeling, and among them are temperature and salinity. These two variables are important when creating models for how fresh water (low salinity) mixes with sea water (high salinity). When the variables are known to be correlated, this correlation should be included in the model to achieve a more complex understanding of the variables.

Defining models for various spatial processes have been of interest for hundreds of years (Gelfand et al. 2010). Also non-statistical methods have been central for describing processes such as waves in the ocean and weather forecasting. Normally such models provide a better understanding of the underlying process, or produce predictions with high accuracy. Building statistical models in a Bayesian hierarchical setting is popular when we deal with noisy observations. Describing a prior is central in Bayesian hierarchical modeling, and basing this knowledge of established models can be appropriate.

Adaptive sampling of the ocean with AUVs require a model which can be updated in real-time. The AUV will have to make decisions for the path of collecting new data, and such decisions require some foundation. This foundation should be as accurate as possible, and therefore be updated in real-time as new samples are collected. One established numerical ocean model is called SINMOD (SIN-

TEF 1987), and has been developed by SINTEF over several years. The numerical model obtains forecasts for several components of the ocean, where some examples are temperature, salinity and the velocity of the water. SINMOD is a complex numerical model and obtaining forecasts is time consuming. The SINMOD forecasts are also associated with some numerical error. In this thesis, we will suggest a statistical model which can learn a prior from numerical model output, and update the prior as observations are collected.

Gaussian Random Fields (GRFs) are often used in spatial modeling due to their predictive power in many applications (Cressie and Wikle 2011). Matérn cross-covariance (Gneiting, Kleiber et al. 2010) and co-regionalization (Schmidt and Gelfand 2003) are two approaches that have been popular for multivariate spatial modeling. Their weaknesses are parameter identifiability and specifying valid statistical models, which are two of the main challenges when building spatial statistical models.

Lindgren et al. (2011) proposed using Stochastic Partial Differential Equations (SPDEs) to model spatial fields. This approach is attractive because it creates an explicit link between GRFs and Gaussian Markov Random Fields (GMRFs). We can model using GRFs, but all computations are done with GMRFs, where the latter is associated with sparse precision matrices. This gives effective sampling and computations, which we explore further in Section 3.2. Obtaining physical interpretation of the model parameters include some spectral theory, which is presented in Section 3.4. Previous work with multivariate modeling using systems of linear SPDEs include Hu, Simpson et al. (2013) and Hu, Steinsland et al. (2013). We will propose three different models, two multivariate models with different cross-dependency structure and one univariate model. These models will be compared both with regards to learning the spatial dependency from a numerical model, and predictive ability on real data samples.

To evaluate the models, we need some quantitative measurements. We choose Mean Square Error (MSE) and Continuous Ranked Probability Score (CRPS) to evaluate the predictive power of the models. MSE provides a easily interpretable measure of the squared deviance between the predicted value and the true value. CRPS gives a measure for how well the target distribution fits the observed values. When evaluating the ability to learn the spatial dependency from numerical forecasts, we will perform both 5-fold cross validation (CV) and hold-out region CV. Evaluation of the predictive power on real data samples will be MSE and mean CRPS for predictions forward in time.

Chapter 2 will go into further detail of the data available for this thesis and describe the problem of interest. In Chapter 3 we will provide the necessary theoretic background to understand the application described in Chapter 4. The results for learning a statistical model from a numerical model are presented in Chapter 5. Chapter 6 applies a statistical model on real data. Lastly in Chapter 7 we discuss the results.

Chapter 2

Objectives and Data Sources

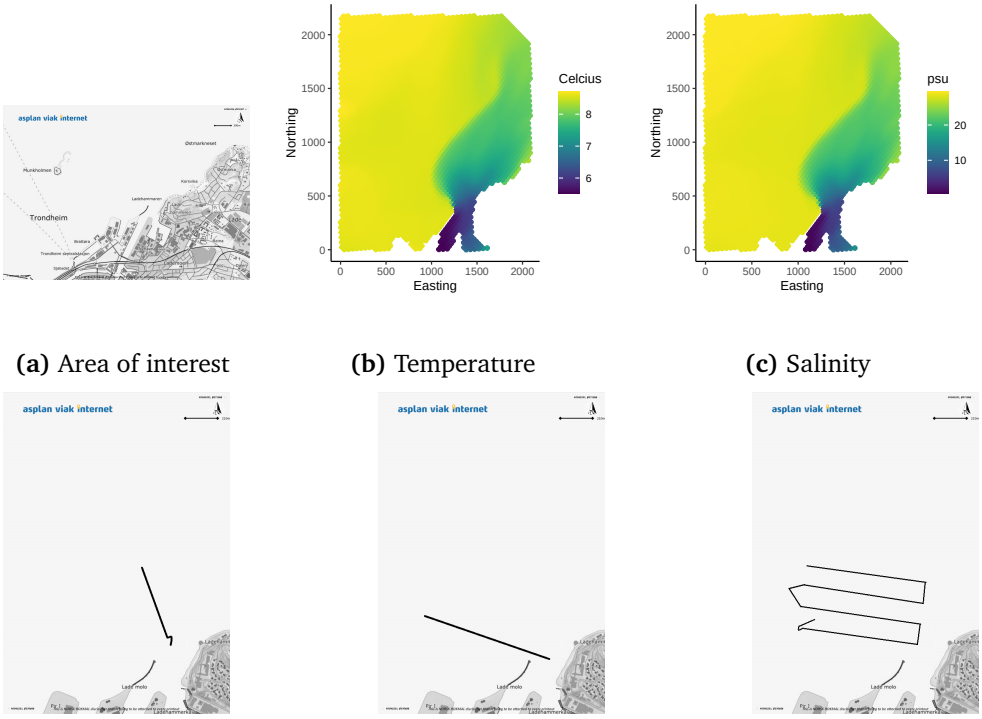
2.1 Application: Trondheim Plume

In Trondheimsfjorden, there is a plume where the river, Nidelven, meets the sea water in the fjord. This border between fresh water and sea water is only visible if weather conditions allow for it. However, it is possible to locate the plume using measurements of salinity and temperature in the water. This way of locating the plume is illustrated in Figure 2.1b and Figure 2.1c showing forecasts of the two fields from the complex numerical ocean model, SINMOD (SINTEF 1987). Dependency between the two outcomes is clearly present. Empirical correlation and cross-correlation, presented in Figure 2.2, amplify this suggestion, and we believe that multivariate spatial models are appropriate for modeling the two fields. We can divide the modeling into to objectives: (1) learning a statistical model for the spatial dependency from the numerical model outputs, and (2) updating the model as observations are collected to improve the model predictions.

2.2 Objective 1: Learn a Statistical Model from Numerical Model

The already established ocean model, SINMOD, comes from solving multiple differential equations numerically. The model is complex, and can give forecasts for both salinity and temperature at 25 layers and in a area of size 4736 m × 3776 m which is locally defined. The temperature unit is Kelvin, which is easily converted to Celcius, while salinity is forecasted in parts per thousand (ppt). Realizations are on a 149 × 119 grid which is rotated with an angle compared to the geographic coordinate system (GCS) (longitude, latitude). The time steps are of 10 minutes, giving 144 numerical model outputs each day. The computations are time-consuming, and getting updated forecasts in real-time is not feasible.

In this thesis, we aim to use the complex numerical ocean model's forecasts to construct a "prior" distribution for the two fields. To fit such prior, we extract the forecasts from 27^{th} of May 2021 at 11 a.m., as this corresponds to the date



(a) Area of interest **(b)** Temperature **(c)** Salinity

(d) Mission 1 **(e)** Mission 2 **(f)** Mission 3

Figure 2.1: (a) A map of Trondheimsfjorden outside of Trondheim. Map was downloaded from `https://marinegrunnkart.avinet.no`. Ocean model (SINMOD) forecasts for (b) temperature and (d) salinity for May 27th 2021 at 11 a.m. The x- and y-axis represents meters away in East and North direction from a reference point. (d)-(f) Illustrations of AUV mission paths collecting data outside of Nidelven in Trondheimsfjorden. Maps are downloaded from `https://marinegrunnkart.avinet.no`. The three missions were all performed on May 27th 2021.

and time where the AUV missions took place. Figure 2.1b and Figure 2.1c display these forecasts on the desired domain. We also compute a time average in each point from the time interval 9 a.m. to 1 p.m. By subtracting the time average from the forecasts for 11 a.m., we obtain two fields with zero mean. The two fields are used to explain the spatial variation and cross-dependency between the variables. Chapter 5 presents the results of learning such multivariate spatial model using systems of linear SPDEs.

Based of the SINMOD forecasts, we explore the correlation between temperature and salinity through an empirical cross-correlogram, presented in Figure 2.2. We observe how the fields are highly correlated for short distances and with range 300 m. Together with the cross-correlogram, empirical correlograms for each of the measurements is included. We observe that the spatial dependency for the two

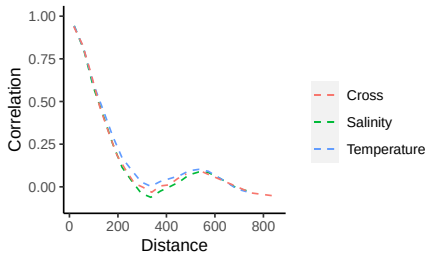


Figure 2.2: Estimated correlograms and cross-correlogram for temperature and salinity on 27th of May 2021 at 11 a.m. The data comes from forecasts by SINMOD for the given time, as visualized in Figure 2.1b and Figure 2.1c, where a time average in each grid point is subtracted.

fields are indistinguishable, where the individual ranges for both fields are both 300 m. We will suggest three different prior models, one with univariate fields and two prior models with cross-dependency between the two fields. This is primary objective 1 of the thesis.

2.3 Objective 2: Apply the Model on AUV Data

The real data used for this thesis was collected for the MASCOT-project, a collaboration project with NTNU and SINTEF (Eidsvik 2020). The AUV data samples is collected in three separate missions, all from May 27th, 2021. The first two missions are transect lines starting in fresh water moving into to sea water consisting of 7,461 and 10,491 measurements. The paths are illustrated in Figure 2.1d and Figure 2.1e. The third mission consists of 39,052 observations originally, but we filter out every 5th measurement such that we obtain 7,811 observations from this path. An illustration of this mission path is in Figure 2.1f. The AUV collects data and estimates its position based on a reference point defined before the mission starts. The estimated position is given in meters north and east relative to this reference point. The reference point is given in the GCS, more precisely WGS-84. By converting this reference point to the universal transverse mercator (UTM) coordinate system we get a neutral coordinate system for both data sources. The UTM-zone for the area of interest is 32V, and we determine the coordinates for each measurement from the AUV. The temperature is measured in degrees Celcius, and the salinity measurements are in *practical salinity unit* (psu). In practice, ppt and psu are used interchangeably.

When the prior from SINMOD data is established, we can start collecting samples from the ocean and update our prior knowledge. This gives us predictions for the current state of the ocean based on samples from the AUV and SINMOD data. The results from the adaptive sampling is found in Chapter 6, and concludes Objective 2 of the thesis.

Chapter 3

Background

3.1 Gaussian Random Fields

We will use GRFs to model temperature and salinity in the ocean. What follows is a short introduction to GRFs. We follow Rue and Held (2005), and define GRFs.

Definition 1 - GRF.

$\{u(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$ is called a GRF if for all finite sets $\{\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_n\} \subset \mathbb{R}^d$, for any $n \in \mathbb{N}$, the n-dimensional vector $[u(\mathbf{s}_1), u(\mathbf{s}_2), ..., u(\mathbf{s}_n)]^\mathsf{T}$ follows a multivariate normal distribution.

GRFs' popularity comes from its compact description. A GRF can be described through its mean function $\mu(\mathbf{s}) = \mathrm{E}[u(\mathbf{s})]$, $\mathbf{s} \in \mathbb{R}^d$ and covariance function $C(\mathbf{s}, \mathbf{t}) = \mathrm{Cov}[u(\mathbf{s}), u(\mathbf{t})]$, $\mathbf{s}, \mathbf{t} \in \mathbb{R}^d$, where $C(\mathbf{s}, \mathbf{t})$ is a symmetric, positive semi-definite function. One example of such valid covariance function is the Matérn covariance function,

$$\rho(d; \nu, \rho) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\sqrt{8\nu}\frac{d}{\rho}\right)^{\nu} K_\nu \left(\sqrt{8\nu}\frac{d}{\rho}\right), \quad d \geq 0 \tag{3.1}$$

where $d = \|\mathbf{s} - \mathbf{t}\|$ for two spatial locations $\mathbf{s}, \mathbf{t} \in \mathbb{R}^d$, $\|\cdot\|$ refers to the Euclidean distance, ν is a smoothness parameter and ρ is the range parameter defined such that the correlation is approximately 0.1 at distance ρ for all values of ν. $K_\nu(\cdot)$ is the modified Bessel function (Gneiting, Kleiber et al. 2010).

The likelihood for a n-variate normal random vector $\mathbf{x}$ with mean vector μ and covariance matrix $\mathbf{\Sigma}$ is given by

$$L(\mathbf{x}; \boldsymbol{\mu}, \mathbf{\Sigma}) = (2\pi)^{-n/2} \det(\mathbf{\Sigma})^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\mathsf{T} \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right\}.$$

The covariance matrix entries satisfy $\Sigma_{ij} = \mathrm{Cov}[x_i, x_j]$ for entries x_i and x_j in $\mathbf{x}$. The covariance matrix is often a dense matrix, and the computations of $\mathbf{\Sigma}^{-1}$ are costly when n is large. In some applications it is more useful to refer to the

precision matrix which is the inverse covariance matrix, $\mathbf{Q} = \Sigma^{-1}$. One of these useful applications are GMRFs, which we come back to in Section 3.2.

Often we have more than one variable of interest, and to construct models with cross-dependency, we use multivariate models. Including cross-dependency in the model can improve predictive power, due to the ability to borrow strength from the other fields. In the case of p variables of interest, we extend the definition from before for p-variate GRFs.

Definition 2 - Multivariate GRF.

Let $\{u_1(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$, ..., $\{u_p(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$ be p GRFs. If for any p finite sets of spatial locations $\{\mathbf{s}_1^{(1)}, ..., \mathbf{s}_{n_1}^{(1)}\}$, ..., $\{\mathbf{s}_1^{(p)}, ..., \mathbf{s}_{n_p}^{(p)}\} \subset \mathbb{R}^d$,

$$[u_1(\mathbf{s}_1^{(1)}), ..., u_1(\mathbf{s}_{n_1}^{(1)}), u_2(\mathbf{s}_1^{(2)}), u_2(\mathbf{s}_{n_2}^{(2)}), ..., u_p(\mathbf{s}_1^{(p)}), u_p(\mathbf{s}_{n_p}^{(p)})]^\mathsf{T}$$

follows a multivariate normal distribution, then $\{\mathbf{u}(\mathbf{s}) = [u_1(\mathbf{s}), ..., u_p(\mathbf{s})]^\mathsf{T} : \mathbf{s} \in \mathbb{R}^d\}$ is a p-variate GRF.

3.2 Gaussian Markov Random Fields

One approach to avoid working with large dense covariance matrices is GMRF. By relating a GRF to a connected graph, $\mathcal{G}$, we can describe a spatial connection between locations. The graph consists of two sets, the set of edges, $\mathcal{E}$ and the set of vertices, $\mathcal{V}$. Figure 3.1 provides a simple example of such simple graph. The

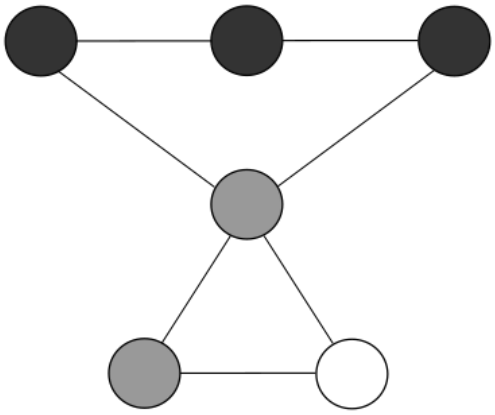


Figure 3.1: Illustration of a simple graph with six nodes and seven edges connecting the nodes. The nodes are categorized into three groups: black, gray and white. The three black nodes are connected such that the middle has two edges to each of the two blacks. The left and right black are directly connected to the middle black node. Gray nodes are directly connected to each other and to the white node. The white node is not directly connected to any black nodes, i.e. there is no single edge connecting the white node to any black node.

definition of a GMRF is as follows (Rue and Held 2005).

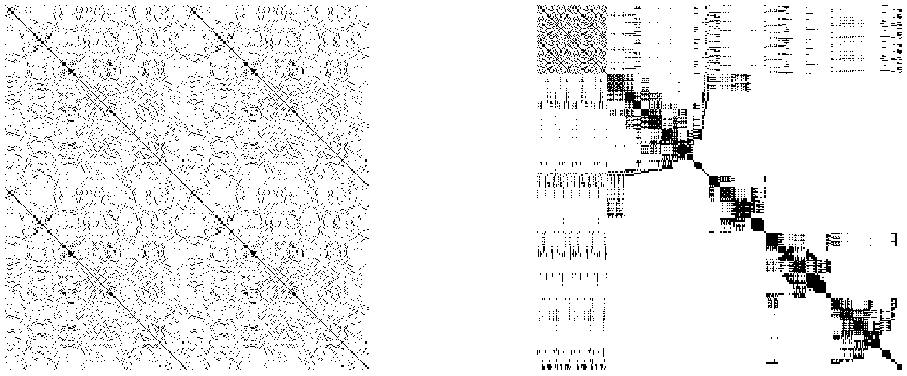
Definition 3 - GMRF.

A random vector $\mathbf{u} = [u_1, u_2, ..., u_n]^\mathsf{T} \in \mathbb{R}^n$ with mean $\boldsymbol{\mu}$ and precision matrix $\mathbf{Q}$ is called a GMRF with respect to a graph $\mathcal{G} = \{\mathcal{E}, \mathcal{V}\}$ iff its density is a multivariate normal distribution and the elements of the precision matrix satisfy

$$Q_{ij} \neq 0 \quad \Longleftrightarrow \quad (i,j) \in \mathcal{E}, \tag{3.2}$$

where Q_{ij} is the element of $\mathbf{Q}$ in row i and column j, and i, j denote indices for vertices in $\mathcal{V}$.

GMRFs are characterized by their sparse precision matrices, which are convenient both related to memory and computations. The computational benefits related to sparse precision matrices are easily available Cholesky factorization and likelihood computations.



(a) Non-permuted **(b)** Permuted

Figure 3.2: Non-zero elements of the precision matrices are represented by black squares, while white areas represent zero elements in the matrix. (a) is not permuted, while (b) is permuted.

The first advantage of working with sparse matrices is the easily obtained Cholesky decomposition. For dense matrices, this decomposition is of computational cost $\mathcal{O}(n^3)$ for a $n \times n$ matrix. For symmetric sparse precision matrices on two dimensional space this cost is reduced to $\mathcal{O}(n^{3/2})$ (Rue and Held 2005). When the matrix in addition is diagonally dominated, the complexity is further reduced to $n(p^2+3p)$, where p refers to the bandwidth of the diagonally dominated matrix (Rue and Held 2005, Section 2.4.1).

The precision matrix for a GMRF is sparse, but not generally diagonally dominated. With permutations of the matrix, we can obtain such diagonally dominated

sparse matrix. Then the Cholesky decomposition of a permuted $\mathbf{Q}$ becomes as follows,

$$\mathbf{PQP}^\mathsf{T} = \mathbf{LL}^\mathsf{T},$$

where $\mathbf{P}$ is referred to as a permutation matrix, which satisfies $\mathbf{PP}^\mathsf{T} = \mathbf{I}$, and $\mathbf{L}$ is the resulting lower triangular matrix from the Cholesky factorization. The permuted matrix, $\mathbf{PQP}^\mathsf{T}$, is constructed such that the bandwidth is optimally reduced. Figure 3.2a illustrates how such permutation changes the order of the elements in the precision matrix. The Cholesky decomposition is provided by the function `Cholesky()` from the `Matrix`-library (Bates et al. 2022).

In maximal likelihood computations, we often use the log-likelihood as an objective function. That is, for a Gaussian random vector $\mathbf{x}$ with mean vector $\boldsymbol{\mu}$ and precision matrix $\mathbf{Q}$, we maximize

$$l(\mathbf{x}; \boldsymbol{\mu}, \mathbf{Q}) = -\frac{n}{2}\log(2\pi) + \log(|\mathbf{Q}|^{1/2}) - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{Q}(\mathbf{x}-\boldsymbol{\mu}), \quad \mathbf{x} \in \mathbb{R}^n$$

instead. When the precision matrix is sparse, we can simplify the log-likelihood expression by using the Cholesky decomposition of $\mathbf{Q}$. That is,

$$l(\mathbf{x}; \boldsymbol{\mu}, \mathbf{Q}) = -\frac{n}{2}\log(2\pi) + \sum_{i=1}^{n}\log(L_{ii}) - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\mathsf{T}\mathbf{Q}(\mathbf{x}-\boldsymbol{\mu})$$

where L_{ii} refers to diagonal elements of the lower triangular matrix $\mathbf{L}$ such that $\mathbf{Q} = \mathbf{LL}^\mathsf{T}$.

To apply GMRFs for multivariate spatial modeling we need some way of describing these graphs and precision matrices. One such method is systems of linear SPDEs.

3.3 Systems of Linear SPDEs

The spatial dependency between temperature and salinity is modelled using systems of linear SPDEs. A system of linear SPDEs in two spatial dimensions will in general have the form

$$\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \dots & \mathcal{L}_{1p} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \dots & \mathcal{L}_{2p} \\ \vdots & & \ddots & \vdots \\ \mathcal{L}_{p1} & & \dots & \mathcal{L}_{pp} \end{bmatrix} \begin{bmatrix} u_1(\mathbf{s}) \\ u_2(\mathbf{s}) \\ \vdots \\ u_p(\mathbf{s}) \end{bmatrix} = \begin{bmatrix} \omega_1(\mathbf{s}) \\ \omega_2(\mathbf{s}) \\ \vdots \\ \omega_p(\mathbf{s}) \end{bmatrix}, \quad \mathbf{s} \in \mathbb{R}^2, \tag{3.3}$$

where $\omega_i(\mathbf{s}), i = 1, ..., p$ are independent Gaussian white noise processes, $\mathcal{L}_{ij} = \tau_{ij}(\kappa_{ij}^2 - \Delta)^{\alpha_{ij}/2}$, $i, j = 1, ..., p$ is the operator working on u_j and Δ refers to the Laplacian $\sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}$ (Hu and Steinsland 2016; Lindgren et al. 2011). We require $\tau_{ii}, \kappa_{ij} > 0$, $i, j = 1, ..., p$.

In this thesis, we will refer to the system of linear SPDEs as the bivariate case, $p = 2$, and we fix $\alpha_{ij} = 2$, $\forall\, i, j = 1, 2$. This simplification will have an effect

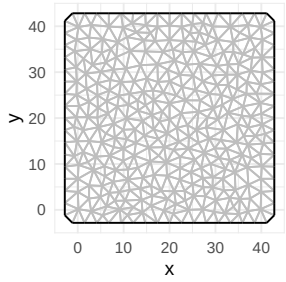


Figure 3.3: Visualization of a triangulation on a $[0, 40] \times [0, 40]$ domain. Here with 132 nodes.

on the smoothness of the fields. More specifically, the stationary solution to the $p = 1$ linear SPDE is a zero mean Matérn GRF when $\alpha > d/2$ (Whittle 1954; Whittle 1963). The smoothness parameter ν in Equation (3.1) is determined from $\nu = \alpha - d/2$, and with $d = 2$ and $\alpha = 2$ the spatial correlation will be restricted to smoothness $\nu = 1$ for the first field.

We simplify Equation (3.3) to a lower triangular operator matrix, as suggested by Hu and Steinsland (2016), and we get the system

$$\begin{bmatrix} \tau_{11}(\kappa_{11}^2 - \Delta) & 0 \\ \tau_{21}(\kappa_{21}^2 - \Delta) & \tau_{22}(\kappa_{22}^2 - \Delta) \end{bmatrix} \begin{bmatrix} u_1(\mathbf{s}) \\ u_2(\mathbf{s}) \end{bmatrix} = \begin{bmatrix} \omega_1(\mathbf{s}) \\ \omega_2(\mathbf{s}) \end{bmatrix}, \quad \mathbf{s} \in \mathbb{R}^2, \tag{3.4}$$

such that $u_1(\mathbf{s})$, $\mathbf{s} \in \mathbb{R}^2$ is determined from $\tau_{11}(\kappa_{11}^2 \Delta) u_1(\mathbf{s}) = \omega_1(\mathbf{s})$, $\mathbf{s} \in \mathbb{R}^2$. The first field is a zero mean Matérn GRF. The system in (3.4) can be represented in a compact manner,

$$\boldsymbol{\mathcal{L}}\mathbf{u}(\mathbf{s}) = \boldsymbol{\omega}(\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^2,$$

where $\boldsymbol{\mathcal{L}}$ is the operator matrix, $\boldsymbol{\omega}$ contains the Gaussian noise processes and $\mathbf{u}$ is the solution to the system of linear SPDEs. Next we provide a short description of the finite element method (FEM) to solve the system of SPDEs. A more thorough introduction is provided by Lilleborge (2021).

To solve the system of linear SPDEs we use FEM, which gives us the weak solution. First, we limit the solution to be on a closed domain, $\Omega \subset \mathbb{R}^2$. Simultaneously, we assume the domain to be homogeneous, which includes that the boundary conditions are homogeneous Neumann. We make a discretization of the domain Ω, called a triangulation, $\mathcal{T}$ with m vertices. An example of such triangulation is shown in Figure 3.3. Every triangle represents an element, and we look for a solution $\mathbf{u}$ of the form

$$\mathbf{u}(\mathbf{s}) = \sum_{i=1}^{2m} w_i \boldsymbol{\phi}_i(\mathbf{s}) \tag{3.5}$$

for weights w_i and some set of linear basis functions

$$\left\{ \begin{bmatrix} \phi_1(\mathbf{s}) \\ 0 \end{bmatrix}, \begin{bmatrix} \phi_2(\mathbf{s}) \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} \phi_m(\mathbf{s}) \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \phi_1(\mathbf{s}) \end{bmatrix}, \begin{bmatrix} 0 \\ \phi_2(\mathbf{s}) \end{bmatrix}, \ldots, \begin{bmatrix} 0 \\ \phi_m(\mathbf{s}) \end{bmatrix}, \right\}.$$

We require

$$\phi_k(\mathbf{s}_j) = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases}, \quad k = 1, ..., m, \tag{3.6}$$

for vertex locations $\mathbf{s}_j$, $j = 1, 2..., m$ in $\mathcal{T}$. This approximation of $\mathbf{u}$ gives us Gaussian distributed weights, $\mathbf{w} = [w_1, ..., w_{2m}]^\mathsf{T}$ with precision matrix given by

$$\mathbf{Q} = \mathbf{K}^\mathsf{T} \tilde{\mathbf{C}}^{-1} \mathbf{K}, \tag{3.7}$$

where $\mathbf{K}$ and $\tilde{\mathbf{C}}$ are sparse block matrices,

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{11} & \mathbf{0} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{bmatrix} \quad \tilde{\mathbf{C}} = \begin{bmatrix} \tilde{\mathbf{C}}_{11} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{C}}_{22} \end{bmatrix}.$$

The block elements $\mathbf{K}_{ij} = \tau_{ij}(\kappa_{ij}\mathbf{C} + \mathbf{G})$, and $\mathbf{C}$, $\tilde{\mathbf{C}}_{ii} = \mathbf{C}^*$ and $\mathbf{G}$ are the element matrices given by

$$C_{ij} = \langle \phi_i, \phi_j \rangle \quad C^*_{ii} = \langle \phi_i, 1 \rangle \quad G_{ij} = \langle \nabla\phi_i, \nabla\phi_j \rangle$$

where $\langle \cdot, \cdot \rangle$ refers to the inner product given by

$$\langle f, g \rangle = \int_\Omega f(\mathbf{s}) g(\mathbf{s}) \, d\mathbf{S}.$$

A thorough computation of the element contributions to the element matrices is provided by Lilleborge (2021), inspired by the work presented by Lindgren et al. (2011) and Hu, Simpson et al. (2013). We denote an element of the triangulation, $\mathcal{T}$, by T_{ijk}. This element consists of three vertices, and three edges connecting the vertices. We denote the vertices by $\mathbf{v}_i$, $\mathbf{v}_j$ and $\mathbf{v}_k$, and the edges $\mathbf{e}_i$, $\mathbf{e}_j$ and $\mathbf{e}_k$ such that $\mathbf{e}_i$ is the edge connecting $\mathbf{v}_j$ and $\mathbf{v}_k$. This is visualized in Figure 3.4. The

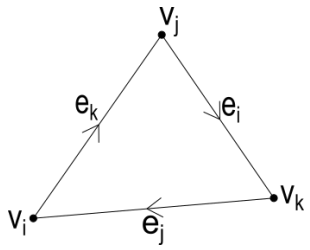


Figure 3.4: Illustration of a triangular element, T_{ijk} in the triangulation $\mathcal{T}$. The triangle consists of vertices $\mathbf{v}_i$, $\mathbf{v}_j$ and $\mathbf{v}_k$ with edges $\mathbf{e}_i$, $\mathbf{e}_j$ and $\mathbf{e}_k$ connecting the vertices.

element contributions from a triangle T_{ijk} to the element matrices, $\mathbf{C}$, $\mathbf{C}^*$ and $\mathbf{G}$

become

$$\{C_{\alpha\beta}\}_{\alpha,\beta=i,j,k} = \frac{1}{12}|T_{ijk}| \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

$$\{C^*{}_{\alpha\alpha}\}_{\alpha=i,j,k} = \frac{1}{3}|T_{ijk}|\mathbf{I}_3$$

$$\{G_{\alpha\beta}\}_{\alpha,\beta=i,j,k} = \frac{1}{4|T_{ijk}|}(\mathbf{e}_i\ \mathbf{e}_j\ \mathbf{e}_k)^\mathsf{T}(\mathbf{e}_i\ \mathbf{e}_j\ \mathbf{e}_k)$$

where $|T_{ijk}|$ denotes the area of the triangle element T_{ijk}.

Due to the construction of the triangulation, $\mathcal{T}$, and the choice of linear basis functions, the precision matrix $\mathbf{Q}$ will be sparse. For a fixed $i \in \{1,...,m\}$, the only non-zero elements of $\mathbf{K}$ are for $j,k \in \{1,...,m\}$ such that (i,j,k) makes a triangle T_{ijk} in $\mathcal{T}$. $\tilde{\mathbf{C}}$ is a diagonal matrix, and so is $\tilde{\mathbf{C}}^{-1}$. Then the construction of $\mathbf{Q} = \mathbf{K}^\mathsf{T}\tilde{\mathbf{C}}^{-1}\mathbf{K}$ will result in a sparse precision matrix.

3.4 Stationary Stochastic Processes

The SPDE parameters in (3.4) can not be interpreted physically, such as range and marginal variance. We would like to gain some understanding of these parameters, and before we can get there, we need some basic understanding of stationary stochastic processes. The processes of interest are real processes, but working with complex processes will simplify the computations of interest. We start by introducing a stationary stochastic process, which can be represented in the following form

$$u(\mathbf{t}) = x(\mathbf{t}) + \mathrm{i}y(\mathbf{t}),$$

where $\{x(\mathbf{t}) : \mathbf{t} \in \mathbb{R}^d\}$ and $\{y(\mathbf{t}) : \mathbf{t} \in \mathbb{R}^d\}$ are two real, stationary stochastic processes and $\mathrm{i} = \sqrt{-1}$ is the imaginary unit. Then

$$\mathrm{E}[u(\mathbf{t})] = \mathrm{E}[x(\mathbf{t})] + \mathrm{i}\,\mathrm{E}[y(\mathbf{t})] = m \quad \mathbf{t} \in \mathbb{R}^d$$

and

$$\mathrm{Cov}[u_1,u_2] = \mathrm{E}[(u_1 - m_1)\overline{(u_2 - m_2)}] = \mathrm{E}[u_1\overline{u_2} - m_1\overline{m_2}],$$

for two stationary stochastic processes, u_1 and u_2, with expectations m_1 and m_2, respectively. We define stationarity similarly to Lindgren (2013, see Definition 4.1).

Definition 4 - Stationarity.

Strictly stationary: For the stochastic process $\{u(\mathbf{t}) : \mathbf{t} \in \mathbb{R}^d\}$, where $u(\mathbf{t}) = x(\mathbf{t}) + \mathrm{i}y(\mathbf{t})$ for $x(\mathbf{t})$, $y(\mathbf{t})$ to be two real-valued processes. If all $2n$-dimensional distributions of $x(\mathbf{t}_1+\boldsymbol{\tau}), y(\mathbf{t}_1+\boldsymbol{\tau}), x(\mathbf{t}_2+\boldsymbol{\tau}), y(\mathbf{t}_2+\boldsymbol{\tau}), ..., x(\mathbf{t}_n+\boldsymbol{\tau}), y(\mathbf{t}_n+\boldsymbol{\tau})$

are independent of $\boldsymbol{\tau} \in \mathbb{R}^d$, we say that the process $\{x(\mathbf{t}) : t \in \mathbb{R}^d\}$ is strictly stationary.

Second order stationary: For some arbitrary process $\{u(\mathbf{t}) : \mathbf{t} \in \mathbb{R}\}$, where $\mathrm{E}[u(\mathbf{t})] = m$ and $\mathrm{E}[u(\mathbf{s})\overline{u(\mathbf{t})}] = C(u(\mathbf{s}), u(\mathbf{t})) + |m|^2$ depends only on the difference $\mathbf{s} - \mathbf{t}$.

For a single process, we define the covariance function as follows (Lindgren 2013).

Definition 5 - Covariance function.

The covariance function $r(\boldsymbol{\tau})$, $\mathbf{t} \in \mathbb{R}^d$ for a stationary stochastic process x with expectation m is given by

$$r(\boldsymbol{\tau}) = \mathrm{E}[(x(\mathbf{s}+\boldsymbol{\tau}) - m)\overline{(x(\mathbf{s}) - m)}], \quad \boldsymbol{\tau} \in \mathbb{R}^d,$$

for some location $\mathbf{s} \in \mathbb{R}^d$. $r(\boldsymbol{\tau})$ is Hermitian. That is, $r(\boldsymbol{\tau}) = \overline{r(\boldsymbol{\tau})}$.

The bivariate process of interest, $\mathbf{u}$ in Equation (3.4), is described through a system of linear SPDEs. To obtain a further understanding of this equation, we need to apply the spectral theorem. Provided is the theorem for a continuous stationary process defined on a one-dimensional domain (Lindgren 2013).

Theorem 1 - The spectral theorem.

If $\{x(t), t \in \mathbb{R}\}$ is a zero mean, continuous stationary process with spectral distribution $F(\omega)$, there exists a complex-valued spectral process $\{Z(\omega), \omega \in \mathbb{R}\}$, with orthogonal increments, such that

$$\mathrm{E}[|Z(\omega_2) - Z(\omega_1)|^2] = F(\omega_2) - F(\omega_1)$$

for all $\omega_1 < \omega_2$ and

$$x(t) = \int_{-\infty}^{\infty} e^{\mathrm{i}\omega t} dZ(\omega).$$

The spectral theorem naturally expands to d dimensional processes, where the integral is over all of $\mathbb{R}^d$. We provide an example of the use of the spectral theorem in $\mathbb{R}^2$. We define the univariate SPDE,

$$\tau(\kappa^2 - \Delta)u(\mathbf{s}) = \omega(\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^2$$

where parameters $\tau > 0$ and $\kappa > 0$, $\omega(\mathbf{s})$ is Gaussian white noise process and $u(\mathbf{s})$ is the process of interest. By inspection of the Fourier transform of the SPDE, we obtain

$$\tau(\kappa^2 + |\mathbf{w}|^2)\hat{u}(\mathbf{w}) = \hat{\omega}(\mathbf{w}), \quad \mathbf{w} \in \mathbb{R}^2,$$

where $\hat{u}$ and $\hat{\omega}$ are the Fourier transformed processes of u and ω. We are able to explicitly state the spectral process $\hat{u}$ by

$$\hat{u}(\mathbf{w}) = \frac{1}{\tau(\kappa^2 + |\mathbf{w}|^2)}\hat{\omega}(\mathbf{w}), \quad \mathbf{w} \in \mathbb{R}^2.$$

We are interested in the covariance function, $r(\mathbf{h})$ where $\mathbf{h} = \mathbf{s} - \mathbf{t}$ between two locations $\mathbf{s}, \mathbf{t} \in \mathbb{R}^2$, which can be found from

$$r(\mathbf{h}) = \iint_{\mathbb{R}^2} \exp\{\mathrm{i}(\mathbf{h} \cdot \mathbf{w}\}f(\mathbf{w})d\mathbf{w} \tag{3.8}$$

where $\cdot$ refers to the inner product, $f(\mathbf{w})$ is the spectral density function´(Lindgren 2013, see p. 16-17). The spectral density function can be found from evaluating $\mathrm{E}[\hat{u}\overline{\hat{u}}]$, where $\overline{a}$ is the conjugate of a complex number $a \in \mathbb{C}$. We find

$$\mathrm{E}[\hat{u}\overline{\hat{u}}] = \frac{1}{\tau^2(\kappa^2 + |\mathbf{w}|^2)^2}\,\mathrm{E}[\hat{\omega}\overline{\hat{\omega}}]$$

where $\mathrm{E}[\hat{\omega}\overline{\hat{\omega}}] = 1/(2\pi)^2$ since the process $\hat{\omega}$ is the Fourier transform of a Gaussian white noise process, $\omega(\mathbf{s})$, defined for $\mathbf{s} \in \mathbb{R}^2$. The covariance function is determined by

$$r(\mathbf{h}) = \frac{1}{\tau^2}\frac{1}{(2\pi)^2}\iint_{\mathbb{R}^2} \exp\{\mathrm{i}(\mathbf{h} \cdot \mathbf{w})\}\frac{1}{(\kappa_{22}^2 + |\mathbf{w}|^2)^2}d\mathbf{w}$$

which Whittle (1963, see Eq. (3.15)) evaluated to be

$$r(\mathbf{h}) = \frac{1}{\tau^2}\frac{|\mathbf{h}|}{4\pi\kappa}K_1(\kappa|\mathbf{h}|)$$

where $K_1(\cdot)$ refers to the modified Bessel function. Since $|\mathbf{h}|$ is the only appearance of $\mathbf{h}$, we can write $d = |\mathbf{h}|$,

$$r(d) = \frac{d}{4\pi\tau^2\kappa}K_1(\kappa d)$$

where we recognize the Matérn covariance function in Equation (3.1) with smoothness parameter, range and marginal variance given by

$$\nu = 1, \quad \rho = \sqrt{8}/\kappa, \quad \sigma^2 = 1/(4\pi\tau^2\kappa^2),$$

respectively.

3.5 Scoring Rules

To evaluate different models, we have to apply some kind of scoring rules. Here we will use mean square error (MSE) and continuous ranked probability score

(CRPS). MSE is commonly used and gives an interpretable measurement for the difference in the prediction and truth. The CRPS measures the prediction fit to the target distribution. The combination of these two measures gives a good understanding of the model's predictive power.

For a set of predictions $\hat{\mathbf{y}} = [\hat{y}_1, \hat{y}_2, ..., \hat{y}_n]^\mathsf{T}$ with true values $\mathbf{y} = [y_1, y_2, ..., y_n]^\mathsf{T}$, the MSE is found from

$$\text{MSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$

When comparing different models and their MSE values, we prefer the model with the lowest MSE score.

The CRPS is a measure of how well the observations y_i fits the target/predictive distribution, F. The general formula for the CRPS is

$$\text{CRPS}(F_i, y_i) = \int_{-\infty}^{\infty} [F(x) - I(x - y)]^2 dx,$$

where $I(x)$ is the Heaviside step function, which is 1 for $x \geq 0$ and 0 otherwise. We can compute the CRPS for an observation y_i when the target distribution is a multivariate normal distribution by the following formula,

$$\text{CRPS}(F_i = \Phi, y_i) = \sigma \left\{ \frac{y_i - \mu}{\sigma} \left[2\Phi\left(\frac{y_i - \mu}{\sigma} \right) - 1 \right] + 2\varphi\left(\frac{y_i - \mu}{\sigma} \right) - \frac{1}{\sqrt{\pi}} \right\},$$

where Φ and ϕ refers to the cumulative density function (CDF) and probability density function (PDF) of a standard normal distribution (Gneiting and Raftery 2007). The observation is denoted by y_i, μ is the mean value of the predictive distribution and σ is the standard deviation. For several predictions, we are interested in the mean CRPS, which is found from

$$\text{CRPS} = \frac{1}{n} \sum_{i=1}^{n} \text{CRPS}(F_i, y_i).$$

Similarly as for MSE, we will prefer the lowest CRPS value when comparing models.

Chapter 4

Bayesian Hierarchical Model

This chapter provides a further introduction to how we build the hierarchical model for real-time predictions during AUV missions. Though the application is for bivariate spatial modeling of temperature and salinity in the ocean, we will in this chapter describe a general Bayesian hierarchical model for $p = 2$ variables.

4.1 Structure

Hierarchical modeling is a popular approach to modeling when you are working with noisy observations of one or several processes. Normally, we collect all observations in a single $2n \times 1$ vector, where 2 refers to the number of variables and n to the number of observations from each variable. A hierarchical model with a Gaussian latent process will typically have the form

$$\begin{aligned} \mathbf{y} \mid \boldsymbol{\eta}, \boldsymbol{\theta} &\sim \pi(\mathbf{y} \mid \boldsymbol{\eta}, \boldsymbol{\theta}) \\ \boldsymbol{\eta} \mid \boldsymbol{\theta} &\sim \mathrm{N}(\boldsymbol{\eta} \mid \boldsymbol{\theta}) \\ \boldsymbol{\theta} &\sim \pi(\boldsymbol{\theta}) \end{aligned}$$

where $\mathbf{y}$ are referred to as the observations, $\boldsymbol{\eta}$ is a vector containing the two Gaussian latent processes and $\boldsymbol{\theta}$ contains the model parameters.

The hierarchical structure allows for different conditional distribution computations, such as integrated likelihood, $\boldsymbol{\theta} \mid \mathbf{y}$ and posterior distribution $\boldsymbol{\eta} \mid \mathbf{y}, \boldsymbol{\theta}$. These conditional distributions can be used for maximum likelihood and predictions. Bayes' rule becomes central in these computations, and we remind the reader

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)} \tag{4.1}$$

for events A and B.

The highest level includes the observations, conditioned on knowing the latent processes and the model parameters. This layer will explain how the observations are related to the latent processes. A common assumption is that the observations

are conditionally independent, that is we have the following relation

$$\pi(\mathbf{y} \mid \boldsymbol{\eta}, \boldsymbol{\theta}) = \prod_{i=1}^{n} \pi(y_i^{(1)} \mid \boldsymbol{\eta}, \boldsymbol{\theta}) \prod_{i=1}^{n} \pi(y_i^{(2)} \mid \boldsymbol{\eta}, \boldsymbol{\theta})$$

where $y_i^{(j)}$ refers to each observation of variable $j = 1, 2$ and n is the number of observation locations.

The observational model will be of the form

$$y_i^{(j)} = \eta^{(j)}(\mathbf{s}_i) + \varepsilon_i^{(j)}, \quad i = 1, ..., n,\ j = 1, 2$$

where $\varepsilon_1^{(1)}, ..., \varepsilon_n^{(1)} \overset{iid}{\sim} \mathrm{N}(0, \sigma_1^2)$ and $\varepsilon_1^{(2)}, ..., \varepsilon_n^{(2)} \overset{iid}{\sim} \mathrm{N}(0, \sigma_2^2)$ are noise related to the measurement, and $\eta^{(j)}(\mathbf{s})$ is the latent process for variable j evaluated at observation location $\mathbf{s}_i$. That is, if the latent process is known, we have $y_i^{(j)} \mid \eta^{(j)}(\mathbf{s}_i) \sim \mathrm{N}(\eta^{(j)}(\mathbf{s}_i), {\sigma_j}^2)$.

The middle layer in the hierarchical model is called the latent process. This process is not observed directly, but is the source for the observations. An example for such middle layer is the true distribution of salinity in the ocean.

The latent process for variable j is of the form

$$\eta^{(j)}(\mathbf{s}) = \mathbf{X}^{(j)}(\mathbf{s})\boldsymbol{\beta}^{(j)} + u^{(j)}(\mathbf{s}), \quad j = 1, 2,\ \mathbf{s} \in \mathcal{D}$$

where $\mathbf{X}^{(j)}(\mathbf{s})$ is a design matrix of spatial covariates, $\boldsymbol{\beta}^{(j)} = [\beta_0^{(j)}, \beta_1^{(j)}]^\mathsf{T}$ contains model parameters and $u^{(j)}(\mathbf{s})$ is a centered GRF.

The lowest level of the hierarchical model is $\boldsymbol{\theta}$. This is often referred to as the prior. The distribution will reflect the existing knowledge about the parameters included in the model. Often this includes some variance related to the observation error, $\boldsymbol{\theta} = [\sigma_1, \sigma_2]^\mathsf{T}$. The underlying GRF included in the latent processes will be modelled through a system of linear SPDEs. That is, the model parameters will additionally include the parameters from Equation (3.4).

4.2 Implementing Systems of Linear SPDEs

We assume that the true spatial fields of temperature and salinity in the ocean can be described through GRFs. The zero-mean GRFs, $u^{(j)}(\mathbf{s})$ for $j = 1, 2$, can be modelled through the system of linear SPDEs. In this thesis, we will refer to three different ways to model the two GRFs: independent model (IM), simple model (SM) and complex model (CM). A model summary is presented in Table 4.1. The models are nested, such that

$$\mathrm{IM} \subseteq \mathrm{SM} \subseteq \mathrm{CM},$$

which means that independency can be recognized for all three models, while IM and SM are not able to describe as complex dependency as CM allow for. The models are defined such that different model complexities can be assessed.

Table 4.1: The three model types together with the parameter limitations related to the system of linear SPDEs in Equation (3.4).

Model name	Description	Restrictions
IM	Independent GRFs	$\tau_{21} = 0$
SM	Simple cross dependence	$\kappa_{21} = \kappa_{11}$
CM	Complex cross dependence	-

We remind the reader of the system of SPDEs, represented by

$$\boldsymbol{\mathcal{L}}\mathbf{u}(\mathbf{s}) = \boldsymbol{\omega}(\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^2,$$

where $\boldsymbol{\mathcal{L}}$ is the operator matrix, $\mathbf{u}(\mathbf{s})$ contains the two fields and the two independent Gaussian white noise processes, $\boldsymbol{\omega}(\mathbf{s}) = [\omega_1(\mathbf{s}), \omega_2(\mathbf{s})]^\mathsf{T}$. We make further investigation of the SPDE parameter interpretation.

The Fourier transformation of the system of SPDEs is on the form

$$\boldsymbol{\mathcal{H}}\hat{\mathbf{u}}(\mathbf{w}) = \hat{\boldsymbol{\omega}}(\mathbf{w}), \quad \mathbf{w} \in \mathbb{R}^2 \tag{4.2}$$

where the elements of $\mathcal{H}_{ij} = \tau_{ij}(\kappa_{ij}^2 + |\mathbf{w}|^2)$, and $\hat{\mathbf{u}}(\mathbf{w})$ and $\hat{\boldsymbol{\omega}}$ are the Fourier transforms of $\mathbf{u}$ and $\boldsymbol{\omega}$, respectively. The spectral process is then found from $\hat{\mathbf{u}}(\mathbf{w}) = \boldsymbol{\mathcal{H}}^{-1}\hat{\boldsymbol{\omega}}(\mathbf{w})$. We are interested in the power spectrum, $\mathbf{S}_u(\mathbf{w}) = \mathrm{E}[\hat{\mathbf{u}}\hat{\mathbf{u}}^H]$, that is,

$$\mathbf{S}_u = \boldsymbol{\mathcal{H}}^{-1}\begin{bmatrix} 1/(2\pi)^2 & 0 \\ 0 & 1/(2\pi)^2 \end{bmatrix}\boldsymbol{\mathcal{H}}^{-H}$$

where $\mathbf{A}^{-H}$ refers to the inverse Hermitian transpose. Since $\boldsymbol{\mathcal{H}}$ is a real matrix, it is equivalent to the inverse of the transpose, $\boldsymbol{\mathcal{H}}^{-H} = \boldsymbol{\mathcal{H}}^{-\mathsf{T}}$. The Gaussian white noise processes are independent and $\mathrm{E}[\hat{\omega}(\mathbf{w})\hat{\omega}(\mathbf{w})^H] = \sigma_\omega^2/(2\pi)^2$ for the Fourier transformation of some Gaussian white noise, $\omega(\mathbf{s})$. Since τ_{ij} is directly scaling σ_ω^2, we choose $\sigma_\omega^2 = 1$. The resulting power spectrum becomes

$$\mathbf{S}_u = \frac{1}{(2\pi)^2}\begin{bmatrix} \frac{1}{\tau_{11}^2(\kappa_{11}^2+|\mathbf{w}|^2)^2} & \frac{-\tau_{21}(\kappa_{21}^2+|\mathbf{w}|^2)}{\tau_{11}^2(\kappa_{11}^2+|\mathbf{w}|^2)^2\tau_{22}(\kappa_{22}^2+|\mathbf{w}|^2)} \\ \frac{-\tau_{21}(\kappa_{21}^2+|\mathbf{w}|^2)^2}{\tau_{11}^2(\kappa_{11}^2+|\mathbf{w}|^2)^2\tau_{22}(\kappa_{22}^2+|\mathbf{w}|^2)} & \frac{1}{\tau_{22}^2(\kappa_{22}^2+|\mathbf{w}|^2)^2}\left(1+\frac{\tau_{21}^2(\kappa_{21}^2+|\mathbf{w}|^2)^2}{\tau_{11}^2(\kappa_{11}^2+|\mathbf{w}|^2)^2}\right) \end{bmatrix}, \tag{4.3}$$

and we can use these spectral density functions to find expressions for the marginal variances. These can be found from $r(\mathbf{h} = \mathbf{0})$ from Equation (3.8), that is,

$$\sigma_i^2 = \iint_{\mathbb{R}^2} (S_u)_{ii}(\mathbf{w})d\mathbf{w}$$

where $(S_u)_{ii}$ is the ith element of the diagonal of $\mathbf{S}_u$. This results in marginal variances

$$\sigma_1^2 = \frac{1}{4\pi\tau_{11}^2\kappa_{11}^2} \tag{4.4}$$

for the first field and

$$\sigma_2^2 = \frac{1}{4\pi}\Bigg\{\frac{1}{\tau_{22}^2\kappa_{22}^2} + \frac{\tau_{21}^2}{\tau_{11}^2\tau_{22}^2}\Bigg[2\frac{(\kappa_{22}^2-\kappa_{21}^2)(\kappa_{21}^2-\kappa_{11}^2)}{(\kappa_{11}^2-\kappa_{22}^2)^3}(\log\kappa_{11}^2-\log\kappa_{22}^2) + \frac{(\kappa_{21}^2-\kappa_{22}^2)^2}{\kappa_{22}^2(\kappa_{11}^2-\kappa_{22}^2)^2} + \frac{(\kappa_{21}^2-\kappa_{11}^2)^2}{\kappa_{11}^2(\kappa_{11}^2-\kappa_{22}^2)^2}\Bigg]\Bigg\}, \quad \kappa_{11}\neq\kappa_{22} \tag{4.5}$$

for the second field.

We observe that if we assume $\kappa_{11} = \kappa_{21}$ (SM) the expression for the second field simplifies to

$$\sigma_2^2 = \frac{1}{4\pi\tau_{22}^2\kappa_{22}^2}\left[1+\left(\frac{\tau_{21}}{\tau_{11}}\right)^2\right]$$

Similarly, we can find an expression for the cross-covariance from

$$\sigma_{12} = \iint (\mathbf{S}_u)_{12}(\mathbf{w})d\mathbf{w},$$

where $(\mathbf{S}_u)_{12}$ refers to the off-diagonal element of $\mathbf{S}_u$. An evaluation of the integral gives us the following expression

$$\sigma_{12} = -\frac{1}{4\pi}\frac{\tau_{21}}{\tau_{11}^2\tau_{22}}\left[\frac{\kappa_{22}^2-\kappa_{21}^2}{(\kappa_{22}^2-\kappa_{11}^2)^2}(\log\kappa_{22}^2-\log\kappa_{11}^2) + \frac{\kappa_{21}^2-\kappa_{11}^2}{\kappa_{11}^2(\kappa_{22}^2-\kappa_{11}^2)}\right] \tag{4.6}$$

for the cross-covariance. In Appendix A there is included a more thorough computation scheme for the evaluation of the integrals. In addition, the resulting expressions for marginal variance and cross-covariance if $\kappa_{11} = \kappa_{22}$.

4.3 Computational Model

In practice, we observe at n different locations, $\{\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_n\} \subset \mathbb{R}^2$, and we have observations of the two fields, $\mathbf{y}^{(1)} = [y_1^{(1)}, ..., y_n^{(1)}]^\mathsf{T}$ and $\mathbf{y}^{(2)} = [y_1^{(2)}, ..., y_n^{(2)}]^\mathsf{T}$, in all of the locations. For the case of missing data for one of the fields, one will simply have two sets of observation locations of different sizes. Without loss of generality, we will assume that we always observe both fields for all n spatial locations. The conditional model takes the form

$$\mathbf{y}^{(j)} \mid \mathbf{w}^{(j)}, \boldsymbol{\beta}^{(j)} \sim \mathrm{N}(\mathbf{A}^{(j)}\mathbf{w}^{(j)} + \mathbf{X}^{(j)}\boldsymbol{\beta}^{(j)}, \sigma_j^2\mathbf{I}) \quad j = 1, 2 \tag{4.7}$$

where $\mathbf{A}^{(j)}$ is an $n \times m$ matrix mapping between the triangulation, $\mathcal{T}$, and the observation locations, $\mathbf{w}^{(j)}$ are the weights for field j, $\mathbf{X}^{(j)}$ is the design matrix, $\boldsymbol{\beta}^{(j)}$ is the parameter vector, $\boldsymbol{\beta}^{(j)} = [\beta_0^{(j)}, \beta_1^{(j)}]^\mathsf{T}$ and $\mathbf{I}$ is the $n \times n$ identity matrix.

The weights and the model parameters for the two fields are assumed to be independent, and we can write $\mathbf{z} = [\mathbf{w}^\mathsf{T}, \boldsymbol{\beta}^\mathsf{T}]^\mathsf{T}$, where $\mathbf{w} = [\mathbf{w^{(1)}}^\mathsf{T}, \mathbf{w^{(2)}}^\mathsf{T}]^\mathsf{T}$ and

$\boldsymbol{\beta} = [\boldsymbol{\beta}^{(1)^\mathsf{T}}, \boldsymbol{\beta}^{(2)^\mathsf{T}}]^\mathsf{T}$. Then $\mathbf{z} \mid \boldsymbol{\theta} \sim \mathrm{N}(\boldsymbol{\mu}_{z|\theta}, \mathbf{Q}_{z|\theta})$ where

$$\mathbf{Q}_{z|\theta} = \begin{bmatrix} \mathbf{Q}_w & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_\beta \end{bmatrix}$$

is a $(2m+4) \times (2m+4)$ matrix.

By collecting all observations $\mathbf{y} = [\mathbf{y}^{(1)^\mathsf{T}}, \mathbf{y}^{(2)^\mathsf{T}}]^\mathsf{T}$, we can construct the hierarchical model as follows

$$\begin{aligned} \mathbf{y} \mid \mathbf{z}, \boldsymbol{\theta} &\sim \mathrm{N}(\mathbf{S}\mathbf{z}, \mathbf{Q}_y^{-1}) \\ \mathbf{z} \mid \boldsymbol{\theta} &\sim \mathrm{N}(\boldsymbol{\mu}_{z|\theta}, \mathbf{Q}_z^{-1}) \end{aligned}$$

where $\mathbf{S} = [\mathbf{A} \quad \mathbf{X}]$ where $\mathbf{A}$ is the block diagonal matrix of $\mathbf{A}^{(1)}$, and $\mathbf{A}^{(2)}$ and similarly for $\mathbf{X}$ with $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$. The precision matrix for $\mathbf{y} \mid \mathbf{z}, \boldsymbol{\theta}$ is written as $\mathbf{Q}_y$ and is a diagonal matrix where the first n entries are $1/\sigma_1^2$ and the last n are equal to $1/\sigma_2^2$. We can determine the distribution for $\mathbf{z} \mid \mathbf{y}, \boldsymbol{\theta}$ from Bayes' rule,

$$\begin{aligned} \pi(\mathbf{z} \mid \mathbf{y}, \boldsymbol{\theta}) &= \frac{\pi(\mathbf{y} \mid \mathbf{z}, \boldsymbol{\theta})\pi(\mathbf{z} \mid \boldsymbol{\theta})}{\pi(\mathbf{y} \mid \boldsymbol{\theta})} \\ &\propto \pi(\mathbf{y} \mid \mathbf{z}, \boldsymbol{\theta})\pi(\mathbf{z} \mid \boldsymbol{\theta}) \\ &\propto \exp\left\{-\frac{1}{2}\mathbf{z}^\mathsf{T}(\mathbf{Q}_{z|\theta} + \mathbf{S}^\mathsf{T}\mathbf{Q}_y\mathbf{S})\mathbf{z} + \mathbf{z}^\mathsf{T}(\mathbf{S}^\mathsf{T}\mathbf{Q}_y\mathbf{y} + \mathbf{Q}_{z|\theta}\boldsymbol{\mu}_{z|\theta})\right\} \end{aligned}$$

and we find that $\mathbf{z} \mid \mathbf{y}, \boldsymbol{\theta}$ is also Gaussian distributed with mean vector and precision matrix given by

$$\begin{aligned} \boldsymbol{\mu}_{z|y,\theta} &= \mathbf{Q}_{z|y,\theta}^{-1}(\mathbf{S}^\mathsf{T}\mathbf{Q}_y\mathbf{y} + \mathbf{Q}_{z|\theta}\boldsymbol{\mu}_{z|\theta}) \\ \mathbf{Q}_{z|y,\theta} &= \mathbf{Q}_{z|\theta} + \mathbf{S}^\mathsf{T}\mathbf{Q}_y\mathbf{S} \end{aligned} \tag{4.8}$$

where $\mathbf{Q}_{z|y,\theta}^{-1}$ is a dense matrix. This distribution is central for making predictions.

We predict the two fields for a set of n^* new locations, $\{\mathbf{s}_1^*, ..., \mathbf{s}_{n^*}^*\}$, using the density for $\mathbf{z} \mid \mathbf{y}, \boldsymbol{\theta}$. That is, we define some $\mathbf{S}_p = [\mathbf{A}_p \quad \mathbf{X}_p]$, where $\mathbf{A}_p$ is a block weight matrix mapping between triangulation and prediction locations, and $\mathbf{X}_p$ is a block design matrix for prediction locations. Then the prediction density for some set of prediction locations becomes

$$\mathbf{S}_p\mathbf{z} \mid \mathbf{y}, \boldsymbol{\theta} \sim \mathrm{N}(\mathbf{S}_p\boldsymbol{\mu}_{z|y,\theta}, (\mathbf{S}_p\mathbf{Q}_{z|y,\theta}\mathbf{S}_p^\mathsf{T})^{-1}), \tag{4.9}$$

where new predictions, $\hat{\mathbf{y}} = [\hat{y}_1^{(1)}, ..., \hat{y}_{\tilde{n}}^{(1)}, \hat{y}_1^{(2)}, ..., \hat{y}_{\tilde{n}}^{(2)}]^\mathsf{T}$, are given by $\mathbf{S}_p\boldsymbol{\mu}_{z|t,\theta}$ and the variance related to each prediction is found by the diagonal of $(\mathbf{S}_p\mathbf{Q}_{z|y,\theta}\mathbf{S}_p^\mathsf{T})^{-1}$, which is a dense matrix. We will avoid computing the inverse precision matrix in full, as we are only interested in the diagonal elements. We denote such diagonal element by, $\sigma_{\hat{y},k}^2$ for $k = 1, ..., 2\tilde{n}$. These diagonal elements can be obtained

without computing the inverse of the precision matrix, as described by Gelfand et al. (2010). The diagonal elements are computed as follows,

$$\sigma^2_{\hat{y},k} = \sum_i S_{p_i} \sum_j (\mathbf{Q}_{z|y,\theta})^{-1}_{ij} S_{p_j}, \quad k = 1, ..., 2\tilde{n}.$$

Due to the sparseness of $\mathbf{S}_p$, we only need the elements of $\mathbf{Q}^{-1}_{z|y,\theta}$ where the precision matrix elements are non-zero. Such partial-inverse matrix can be obtained from the `R-INLA`-library (Rue, Martino et al. 2009) (`www.r-inla.org`), using the `inla.qinv()` function.

Hierarchical models are also useful when we want to fit a model to observations through parameter estimation. We can express the integrated likelihood as follows,

$$\pi(\mathbf{y} \mid \boldsymbol{\theta}) = \frac{\pi(\mathbf{y}, \boldsymbol{\eta} \mid \boldsymbol{\theta})}{\pi(\boldsymbol{\eta} \mid \mathbf{y}, \boldsymbol{\theta})} = \frac{\pi(\mathbf{y} \mid \boldsymbol{\eta}, \boldsymbol{\theta})\pi(\boldsymbol{\eta} \mid \boldsymbol{\theta})}{\pi(\boldsymbol{\eta} \mid \mathbf{y}, \boldsymbol{\theta})},$$

which is independent of $\boldsymbol{\eta}$. This likelihood can be used to determine a set of parameters that satisfy

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \pi(\mathbf{y} \mid \boldsymbol{\theta}),$$

where $\mathbf{y}$ are the observations. This approach to parameter estimation is called maximum likelihood, and will be applied in Chapter 5 where the goal is to learn the spatial dependency of temperature and salinity through forecasts from the numerical ocean model, SINMOD.

Chapter 5

Learning a Statistical Model from Numerical Model

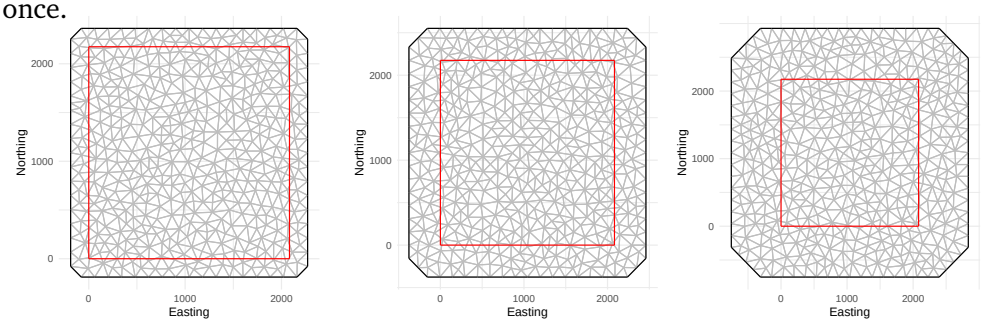
5.1 Goal

In this chapter, we aim to learn a statistical model from the numerical ocean model, SINMOD, using systems of SPDEs. That is, we use the forecasts from SINMOD which were presented in Figure 2.1b and Figure 2.1c for May 27$^{\text{th}}$ 2021 at 11 a.m., and we subtract the time average computed from the same day from 9 a.m and 1 p.m. We then obtain two fields with zero means. They consist of 3798 spatial data points. These two anomalies are used as observations, $\mathbf{y}$, when we estimate the parameters from MLE using $\mathbf{y} \mid \boldsymbol{\theta}$ from Section 4.3. Before we can start the optimization, we have to determine the triangularization $\mathcal{T}$. The mesh is created with R-INLA (Rue, Martino et al. 2009), `inla.mesh.2d()`. Specifying the border offset will have an effect on the parameter estimation, and this effect is investigated further in the following section.

5.2 Model Resolution

We want to determine the offset value that gives the more reliant parameter estimation. The larger boundary, the less boundary effects on the domain we are interested in (Lindgren et al. 2011). With a larger domain the computation time will increase. By repeated parameter estimation for different offset values, we find that a 6.25% boundary is too small, as the range estimates with resulting parameters are too large. For boundary offsets equal to 12.5% and 25% the parameters give better estimates for range. We decide to use boundary offset equal to 12.5% for our parameter estimates. The number of vertices, m, is chosen to be 7500. The `inla.mesh.2d()` function also requires a specification of the maximal edge length for the triangles, which is chosen to be 1.8% of the width of the area of interest. The mesh is created before making any observations, and can then be used even when more observations are added. This saves computations, as the

element matrices for building **Q** from Equation (3.7) only needs to be computed once.



(a) Small border **(b)** Medium border **(c)** Large border

Figure 5.1: Triangulations with 500 vertices with different border sizes. Left (a), middle (b) and right (c) has offset borders to be 6.25%, 12.5% and 25% of the diameter of the domain. Red box illustrates the area where observations are collected from. The x-axis is parallel with Easting and the y-axis is parallel with Northing in the UTM 32V zone. The reference point $(0, 0)$ is chosen such that all coordinates inside the read box have positive x and y values.

5.3 Parameter Estimation and Interpretation

The model presented in (4.7) takes the form

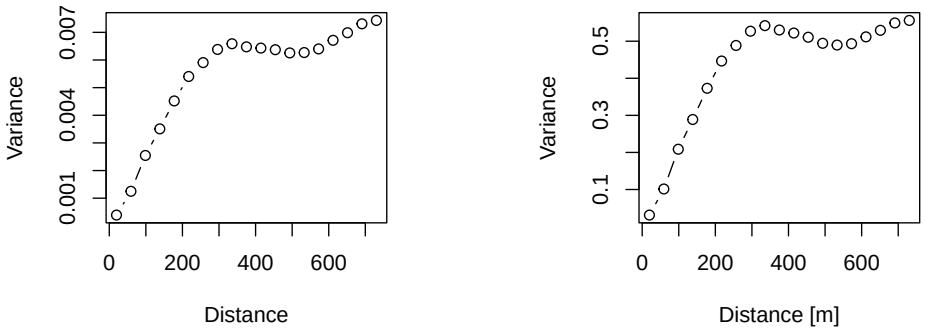
$$y_i^{(j)} \mid \mathbf{w}^{(j)}, \boldsymbol{\theta} = \mathbf{A}_i^{(j)} \mathbf{w}^{(j)} + \varepsilon_i^{(j)} \quad j = 1, 2, i = 1, ..., n \tag{5.1}$$

for the zero-mean SINMOD anomalies, for a location $\mathbf{s}_i$. That is, the design matrix becomes $\mathbf{X}_i^{(j)} = [0 \quad 0]^\mathsf{T}$, for $j = 1, 2$, and $i = 1, ..., n$. The observation errors are Gaussian distributed, $\varepsilon_i^{(j)} \overset{iid}{\sim} \mathrm{N}(0, {\sigma^{(j)}}^2)$. A description of the SINMOD forecasts can be found in Section 2. The parameters for the model becomes

$$\boldsymbol{\theta} = [\kappa_{11}, \kappa_{21}, \kappa_{22}, \tau_{11}, \tau_{21}, \tau_{22}, \sigma^{(1)}, \sigma^{(2)}]^\mathsf{T},$$

which are estimated from optimization using the log-likelihood as explained in Section 4.3.

We add noise to the SINMOD anomalies to smooth out the spatial field. Without this added noise, the optimization is not able to recognize the SPDE parameters accurately. The noise is Gaussian with standard deviation $\sigma^{(j)}$, $\varepsilon_i^{(j)} \sim \mathrm{N}(0, {\sigma^{(j)}}^2)$, where the variance is chosen to be 4% of the empirical marginal variance in the SINMOD data. From Figure 5.2a and Figure 5.2b we estimate this marginal variance to be 0.007 and 0.5 for temperature and salinity, respectively. The choice of 4% of the empirical marginal standard deviation in the data is from experimenting with different values. For lower values, for instance 1%, we found that the optimization had trouble identifying the observation variance correctly.



(a) Temperature **(b)** Salinity

Figure 5.2: Empirical variograms based on the SINMOD anomalies for temperature and salinity.

The initial parameter values are chosen such that field ranges and marginal variances are close to the estimated values. We use the relations described in Section 4.2. The initial parameter values are presented in Table 5.1. In addition we pass the true values for $\sigma^{(j)}$, $j = 1, 2$, that were added manually. The estimates

Table 5.1: Initial SPDE parameters passed to the optimization algorithm.

κ_{11}	κ_{21}	κ_{22}	τ_{11}	τ_{21}	τ_{22}
0.00943	0.00943	0.00943	358	-0.0001	42.3

for all six model parameters can be found in Table 5.2.

The estimated SPDE parameters can provide estimates for the marginal variance and range for the two GRFs, and the cross-covariance between them. For the first field, the marginal variance and range is simply estimated from the formula presented in Equation (4.4) and the empirical relation $\rho = \sqrt{8}/\kappa$ (Lindgren et al. 2011). For IM we can use these two formulas for both fields. For the coupled fields, we have more complex formulas for the marginal variance, as presented in Section 4.2. Next we present the estimated for all three models.

With IM we estimate the marginal variances to be 0.0137 for temperature and 1.01 for salinity. The estimated range for temperature is 580 m and for salinity it is 557 m. The estimated marginal variances and ranges through SPDE parameters

Table 5.2: Estimates for the SPDE parameters based on SINMOD anomalies at time of interest.

$\hat{\boldsymbol{\theta}}$ \ Model	IM	SM	CM
$\kappa_{11}(10^{-3})$	4.88	0.689	4.73
$\kappa_{21}(10^{-1})$	0	0.00689	2.71
$\kappa_{22}(10^{-2})$	0.508	0.468	4.57
τ_{11}	494	491	471
τ_{21}	0	-39.1	-11.1
τ_{22}	55.2	55.1	41.7

are higher than the estimated empirical variograms, which were 0.007 (marginal variance, temperature), 0.5 (marginal variance, salinity) and 300 m (range, temperature and salinity).

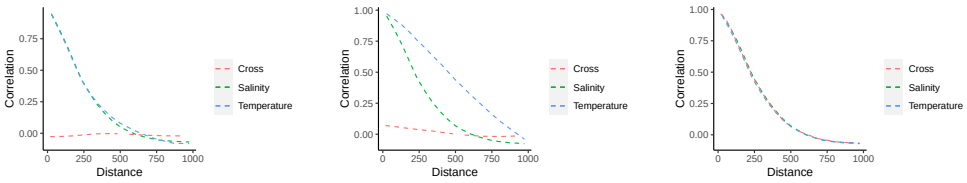
The SPDE parameters estimated with SM result in marginal variances equal to 0.696 and 1.20 for temperature and salinity, respectively. The cross-covariance is estimated to be 0.0182. This suggests that SM estimates two independent fields. The ranges are estimated to be 4100 m for first field, temperature. For salinity the estimated range is more complex to quantify exactly, but ranges from the dependency, κ_{21} is estimated to be 4100 m as well, while for κ_{22} we get 604 m. The first field is estimated to have a larger marginal variance and a much larger range than the ones obtained from empirical variograms. The second field is also estimated with a larger marginal variance than the empirical variogram suggests.

The marginal variances with CM are estimated to be 0.0161 and 1.25 for temperature and salinity, respectively, when using the formulas in Equation (4.4) and Equation (4.5). The range for temperature with CM is estimated to be 598 m. From the dependency term, the empirical range is 10.4 m, and from the second field we get 61.9 m. The actual ranges for each field is not known. The marginal variances are 0.0161 for temperature, 1.25 for salinity and 0.149 for the cross-covariance. That is, CM suggest correlation close to 1, and the fields are strongly dependent in the same location.

In total, all three models suggest a larger range than the empirical estimates of 300 m for temperature and salinity. The marginal variances are also estimated with SPDE parameters to be at least two times larger than the values from the empirical variograms. We note how SM is suggesting almost no dependency between the fields, and that the estimated marginal variance for temperature is close to 100 times larger than the empirical variogram suggests.

In addition to analytical computations using the SPDE parameters directly, we construct empirical correlograms and cross-correlograms for all three model types. These provide estimates for ranges and cross-ranges, as we are not able to directly estimate them from model parameters. Figure 5.3 display estimated correlograms for spatial fields of temperature and salinity and cross-correlogram between the two fields. These are obtained by repeatedly sampling from the Gaus-

sian distribution with precision matrix $\mathbf{Q}$ with the parameters estimated for each model. As expected, we observe that the cross-correlogram for both IM and SM are close to zero for all distances, suggesting completely independent fields. The range estimates for IM, in Figure 5.3a, suggests the range to be around 600 m for both fields. This corresponds well with the estimates using the empirical relation $\sqrt{8\nu}/\kappa$. Figure 5.3b we note that the range for salinity is shorter than the range for temperature. For CM, Figure 5.3c, we note that all empirical correlograms are indistinguishable, similar to the result for the SINMOD anomalies in Figure 2.2. The ranges are all estimated to be around 500 m, which is close to the empirical estimate for temperature when using SPDE parameters for CM.



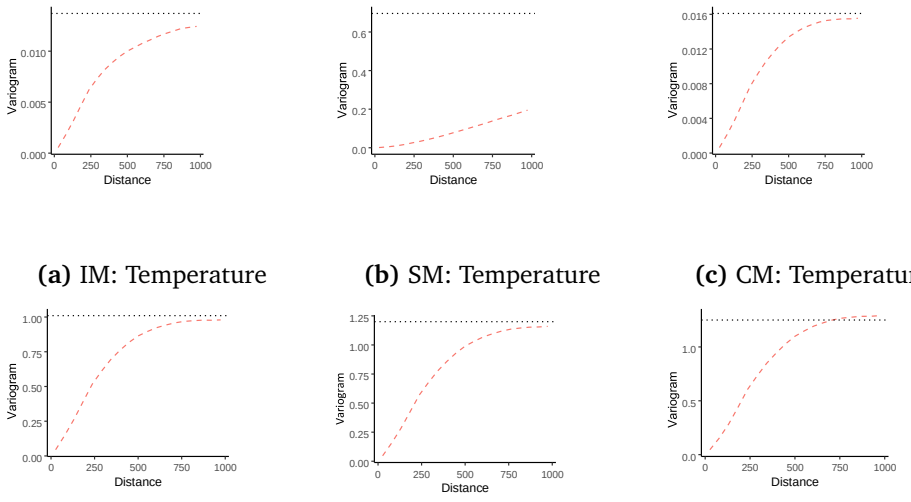
(a) IM **(b)** SM **(c)** CM

Figure 5.3: Estimated correlograms and cross-correlogram for temperature and salinity. The distance is in meters [m]. We sample with precision matrix as described in Equation (3.7), with the set of SPDE parameters in Table 5.2 and obtain 15 realizations, which are used to compute 15 correlograms for each field and cross-correlograms. The means of the 15 correlograms and cross-correlograms are presented.

We also validate with empirical variograms that the marginal variances in Equation (4.4) and Equation (4.5) provide accurate estimates. The result is visualized in Figure 5.4. For all models we find that the formulas provide accurate values, where the only exception is with SM for temperature. The estimated variogram is only for distances up to 1000 m, while the estimated range is 4100 m. The estimated marginal variance for temperature with SM will therefore not be reached within the first 1000 m of the variogram. Extending the variogram for distances over 1000 m will result in strange-looking variograms due to few data points to compute the empirical variogram with.

5.4 Evaluating Model Fit

Evaluating the model fit can be done in various ways. We will do two different approaches, k-fold cross-validation (CV) and hold-out regions CV. All evaluations



(a) IM: Temperature **(b)** SM: Temperature **(c)** CM: Temperature

(d) IM: Salinity **(e)** SM: Salinity **(f)** CM: Salinity

Figure 5.4: Empirical variograms (red, dotted line) for temperature and salinity and all three models. The black dotted line illustrates the estimated marginal variances computed using SPDE parameter estimates in Table 5.2 and formulas in Equation (4.4) and Equation (4.5). We sample with precision matrix as described in Equation (3.7), with the set of SPDE parameters in Table 5.2 and obtain 15 realizations, which are used to compute 15 empirical variograms for each field. The means of the 15 variograms are presented.

are performed with the same forecasts from SINMOD for 11 a.m. with the time average from 9 a.m. to 1 p.m. in each spatial location subtracted.

A regular approach for model evaluation is k-fold CV, which includes randomly dividing the data set into a train and test sets. The training set is used for fitting the model, and the test set is used to evaluate the model fit. A regular approach is to do 5-fold CV. The procedure is as follows: we randomly divide the data into five folds, and sequentially leave out one fold as the test set, and the rest is used as the training set.

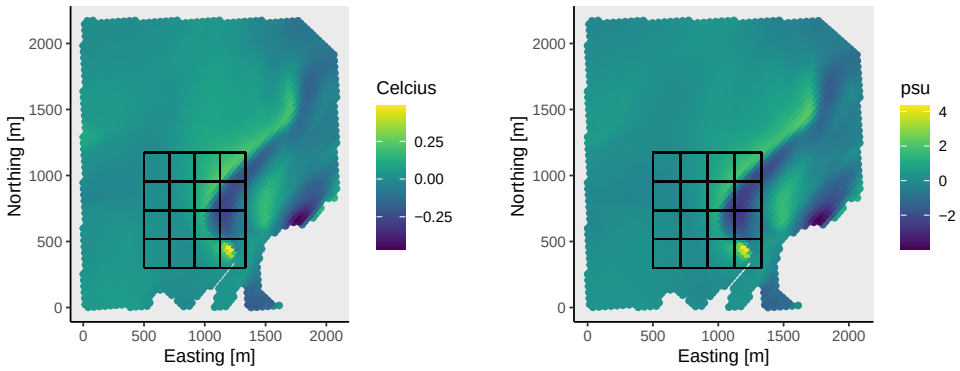
We repeat the 5-fold CV five times, giving us in total 25 parameter estimates, MSE scores and mean CRPS for each of the three model types, IM, SM and CM. The mean values for MSE and mean CRPS are presented in Table 5.3. We find that the models perform equally well, but SM is the model with the lowest mean MSE and CRPS. To make sure the models have the same foundation to perform well, we repeat the five sets of folds for all models.

A second approach to model evaluation is hold-out region CV. First, we determine the area where the plume is the most prominent. We divide this section into sixteen distinct areas, as shown in Figure 5.5. The data points from outside of the selected area are included in all parameter estimations. These forecasts are close to zero, as this area is mostly pure sea water and pure fresh water. For each

Table 5.3: Mean MSE and CRPS scores for first (temperature) and second (salinity) field from the repeated 5-fold CV evaluations on the SINMOD data.

	MSE (10^{-2})		CRPS (10^{-2})	
	1st Field	2nd Field	1st Field	2nd Field
IM	0.0468	3.21	1.18	9.75
SM	0.0450	3.09	1.16	9.56
CM	0.0456	3.13	1.17	9.61

of the 16 regions, we leave one of them out as a test set. The training set is the remaining 15 regions and the data from outside of the plume area. The training set is used to estimate SPDE parameters and based on these parameters we make predictions for the anomalies on the test set locations. The predictions are compared to the test set values, and MSE and mean CRPS are computed. The regions are the same for all models. Hold-out region CV will, in contrast to regular k-fold CV, give a better understanding of the predictive ability in regions where there are no observations. This is useful in our desired application, where the collection of data is limited to the path of the AUV.



(a) Temperature **(b)** Salinity

Figure 5.5: Illustrating the sixteen regions used for hold-out validation for both (a) temperature and (b) salinity. The data points are spatial forecasts from SINMOD where the time averages are subtracted for temperature and salinity.

The MSE and mean CRPS values for the hold-out CV are presented in Table 5.4. We observe that the MSE values decrease for the second field when making the model more complex, but the first field has the lowest MSE score for SM. For the average CRPS' we also see that SM is the model with the lowest score. This suggests that SM is the best at catching the trends in the SINMOD data, though the results are quite similar for all three models.

We observe that the optimization is numerically unstable, and the `nloptr` gives a status "Generic failure code" for some of the optimizations. This could be caused

Table 5.4: Mean MSE and CRPS scores for the first (temperature) and second (salinity) field from hold-out CV evaluations on the SINMOD data.

	MSE (10^{-1})		CRPS (10^{-1})	
	1st Field	2nd Field	1st Field	2nd Field
IM	0.0253	2.07	0.0220	2.00
SM	0.0245	2.05	0.0217	1.99
CM	0.0253	2.03	0.0219	2.00

by a flat likelihood, which results in a small gradient, making the optimization give up. One possible solution to this issue is to compute an analytic gradient for the objective function.

It seems from multiple parameter estimations that the κs are the most difficult parameters to estimate when we have dependent fields. The estimates for κ_{11}, κ_{21} and κ_{22} are varying in scale from 10^{-7} to 10^{-3}. The estimates for the τs are more consistent, where τ_{ii}, $i = 1, 2$ are all on the same scale, while τ_{21} is mostly scaled with 10^2, but in some cases 10^1 and 10^3.

Chapter 6

Applying Statistical Model on AUV Data

6.1 Goal

Now that the prior for the two fields utilizing the SINMOD forecasts is established, we want to make predictions on real measurements from the area of interest. That is, we will use data collected with an AUV in the ocean outside of Trondheim on May 27$^{\text{th}}$ 2021. By extracting observations for only parts of the missions, we observe how the statistical models learn from the collected data. We will provide a more precise description of the model for prediction of AUV data. We explore how well the model learns the fields when combining the prior established from SINMOD data and observations. The predictions are compared with the observations collected for the rest of the mission. This is done for all three missions presented in Chapter 2. Lastly, we also visually evaluate predictions on the area of interest. The true field values are not known for the whole area.

6.2 Prediction Model

We would now like to determine the model for the AUV collecting data sequentially. That is, we use the SINMOD model as a prior, and combine this with the observations in real-time. The model will be as described in Section 4.3. The design matrix from Equation (4.7) takes the form

$$\mathbf{X}^{(j)} = \begin{bmatrix} 1 & \overline{Z}_1^{(j)} \\ \vdots & \vdots \\ 1 & \overline{Z}_n^{(j)} \end{bmatrix}, \quad j = 1, 2,$$

where $\overline{Z}_i^{(j)}$, $i = 1, ..., n$ are the time average values for location $\mathbf{s}_i$ for field j computed from the SINMOD forecasts from 9 a.m. to 1 p.m. To get these time averages,

$\overline{Z}_i^{(j)}$ for all observation locations, $\mathbf{s}_i$, $i = 1, 2, ..., n$ and both fields, $j = 1, 2$ we apply linear interpolation of the SINMOD forecasts onto the observation locations.

The observation variances, σ_1^2 and σ_2^2, are estimated from the collected data. The procedure is as follows: we divide the observations into bins of a certain length, and compute the standard deviation of each bin. The average standard deviation across all bins is used as an estimate for the observation standard deviation. For this data set we find the observation variances to be $\sigma_1^2 = 0.05^2$ and $\sigma_2^2 = 0.15^2$ for temperature and salinity, respectively.

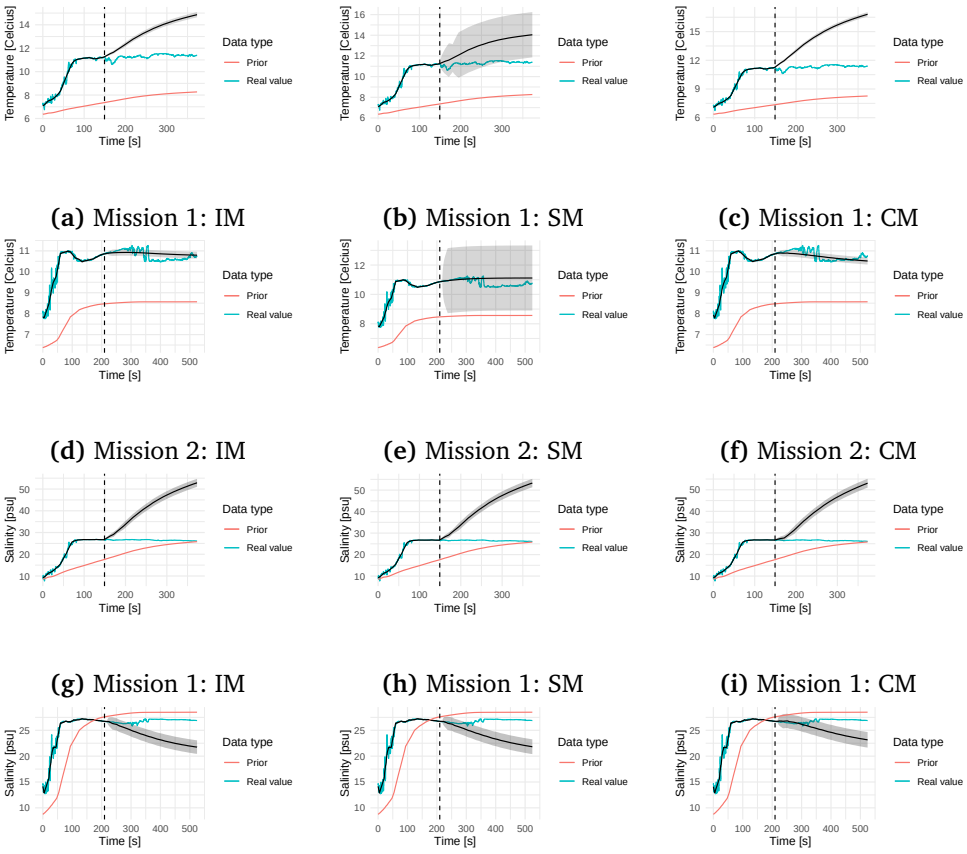
We make predictions, $\hat{\mathbf{y}}$, as described in Section 4.3. Computations of the density for $\mathbf{S}_p\mathbf{z} \mid \mathbf{y}, \boldsymbol{\theta}$ using Equation (4.9) gives us both predictions and the associated uncertainty for each spatial prediction.

6.3 Prediction on Transect Line Missions

We want to investigate how well the model predicts the temperature and salinity for future locations, $\{\mathbf{s}_1, ..., \mathbf{s}_n\}$, when the observations are from a different area, $\{\mathbf{s}_1^*, ..., \mathbf{s}_{n^*}^*\}$. That is, the observations are all collected along a single line, and we want to predict the measurements along a continuation of that same line. To evaluate the model prediction, we compare the prediction interval to the actual measurements collected by the AUV. This is visualized in Figure 6.1 for temperature and salinity, where 40% of the observations are used to learn the model. For the first mission, we note that the prior knowledge (red line) about the field is not very accurate for the level we actually observe. The location of the plume (quick shift in salinity and temperature) is not in the prior knowledge, and all three models use a scaled prior for predictions on unobserved locations. The prediction back in time suggest to follow the observations closely. The model recognizes that the prior does not fit the observations well, and will trust the observations more than the prior when updating the prior as new observations are sampled.

We note that SM gives a much wider prediction interval for temperature than the two other models. This is most likely caused by the estimated marginal variance for temperature, which is 0.696 from the estimated SPDE parameters with SM, compared to 0.0137 and 0.0161 for IM and CM respectively. SM will be more uncertain about future observations, as the GRF for temperature has a larger marginal variance than the two other models.

We check if the predictions improve with more observation time. We increase the observation time to 80% of the total number of observations, and the result is visualized in Figure 6.2 for temperature and salinity. As seen before the predictions back in time closely follow the observed values, while predictions forward in time are mainly based on the prior distribution.



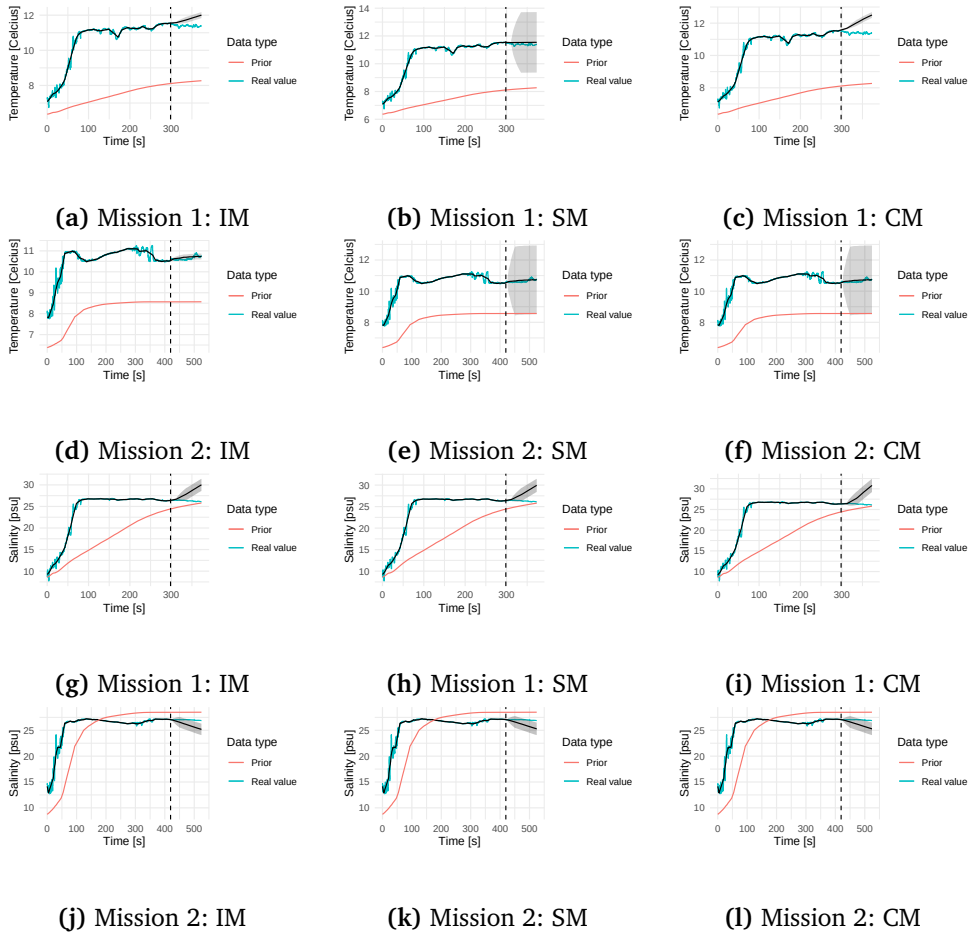
(a) Mission 1: IM **(b)** Mission 1: SM **(c)** Mission 1: CM

(d) Mission 2: IM **(e)** Mission 2: SM **(f)** Mission 2: CM

(g) Mission 1: IM **(h)** Mission 1: SM **(i)** Mission 1: CM

(j) Mission 2: IM **(k)** Mission 2: SM **(l)** Mission 2: CM

Figure 6.1: Predictions of (a)-(f) temperature and (g)-(l) salinity for the first and second mission. The observation time varies for each mission, and is determined such that the AUV has collected data from 40% of the total mission. The dashed line marks where the observed time stops. The prior (red line) represents the knowledge about the field before the mission started. The true value (blue line) shows the AUV measurements from the whole mission. The prediction (black line) is what the model predicts after observing the initial measurements. The gray area suggests the uncertainty associated with the prediction.



(a) Mission 1: IM **(b)** Mission 1: SM **(c)** Mission 1: CM

(d) Mission 2: IM **(e)** Mission 2: SM **(f)** Mission 2: CM

(g) Mission 1: IM **(h)** Mission 1: SM **(i)** Mission 1: CM

(j) Mission 2: IM **(k)** Mission 2: SM **(l)** Mission 2: CM

Figure 6.2: Predictions of (a)-(f) temperature and (g)-(l) salinity for the first and second mission. The observation time varies for each mission, and is determined such that the model learns from 80% of the total mission. The dashed line marks where the observed time stops. The prior (red line) represents the knowledge about the field before the mission started. The true values (blue line) shows the AUV measurements from the whole mission. The prediction (black line) is what the model predicts after observing the initial measurements. The gray area suggests the uncertainty associated with the prediction.

We find that the model will not necessarily learn from the observations, if only either sea water or fresh water is observed. One solution is to collect data from both levels before-hand, such that the model learns both levels, and can use the prior less when SINMOD is not providing an accurate forecasts. We test this using the third mission with a lawn mower path.

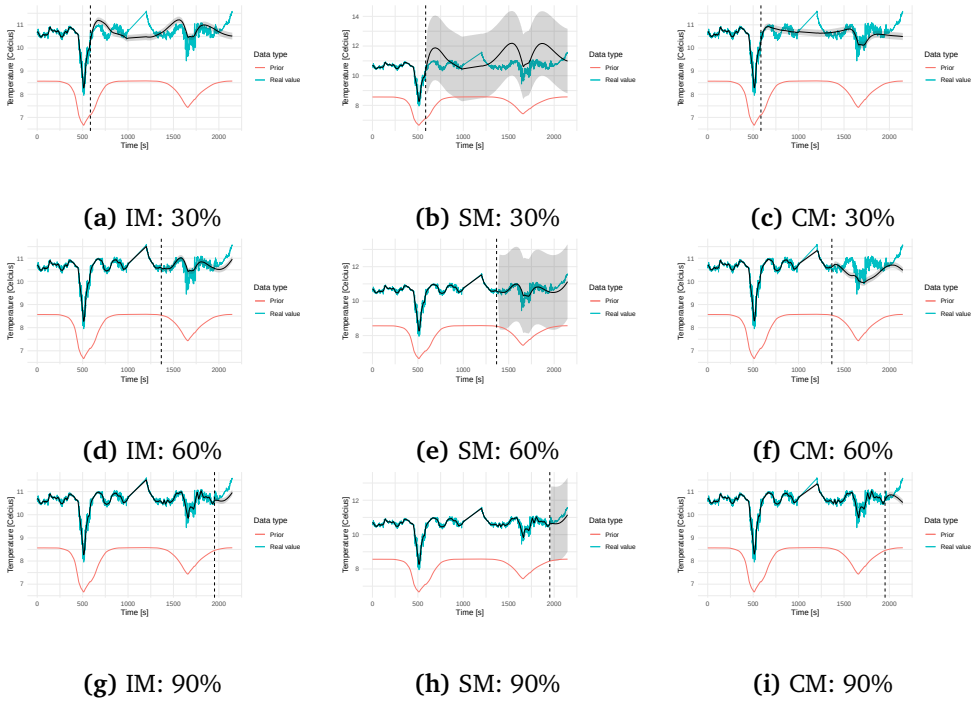
6.4 Prediction on Lawn Mower Mission

The data from the third mission is collected in a typical lawn mower pattern. This results in an early visitation of both sea water and fresh water from the river, such that the models are able to learn both levels in a shorter time. A visualization of how the three models perform with varying observation time is illustrated in Figure 6.3 and Figure 6.4 for temperature and salinity, respectively.

We compare the models' ability to make accurate predictions for temperature in three different scenarios. First, as seen before, SM has a wider prediction interval for temperature than the two other models. Figure 6.3b, Figure 6.3e and Figure 6.3h illustrate this. The prior knowledge is accurately locating the rapid change in temperature, though the levels are inaccurate. All models will quickly learn this level change, and give quite similar predictions for the future measurements. When comparing Figure 6.3a, Figure 6.3b and Figure 6.3c we see that IM is predicting the shifts in temperature the most accurately, while SM is predicting a more extreme increase in temperature than the true field.

When we increase the observation time, we note that all models are closely following the observed levels for previously visited locations. We note that the prediction means in Figure 6.3d and Figure 6.3e are very similar, while CM in Figure 6.3f is making predictions based off the prior knowledge. This could be caused by CM being a more complex model, which might need more observations to learn the field accurately. In Figure 6.3g-6.3i we find that IM and SM make similar predictions, while CM predicts that the temperature will not increase for the next future locations.

The first thing to note in Figure 6.4 is that the prior is predicting the distribution and levels of salinity very accurately for this mission. The predictions for a short observation time is mainly based on the prior, which means that the three models make quite similar predictions, as seen in Figure 6.4a, Figure 6.4b and Figure 6.4c. Figure 6.4d and 6.4e shows that the predictions from IM and SM on salinity are similar also for longer observation times. This is also the case in Figure 6.4g and Figure 6.4h. On the contrary, we see that CM predicts something closer to the prior when observing 60% of the mission data, as can be seen in Figure 6.4i. This could, as mentioned before, suggest that CM needs a longer observation time before learning the field.



(a) IM: 30% **(b)** SM: 30% **(c)** CM: 30%

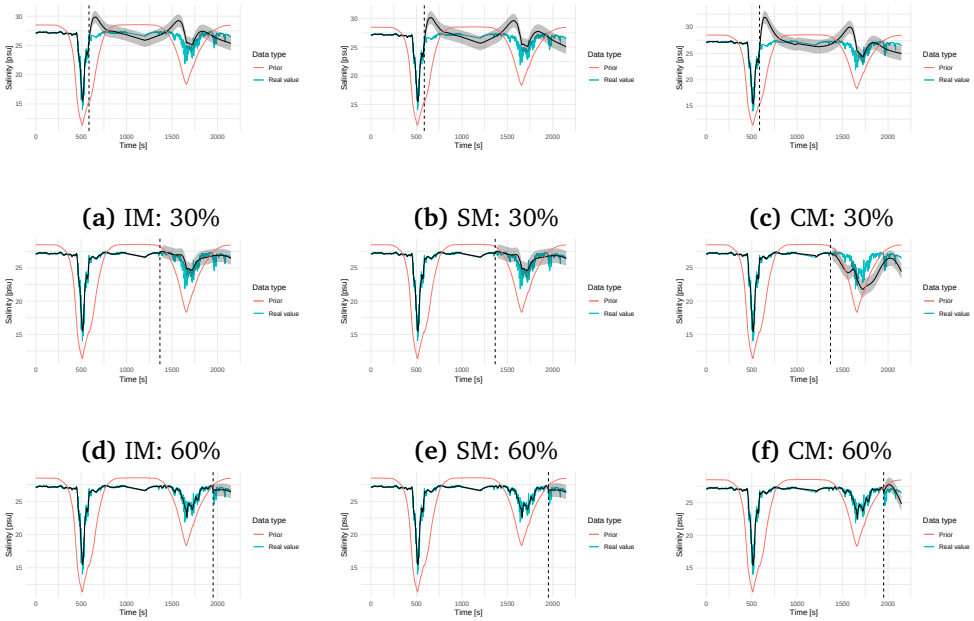
(d) IM: 60% **(e)** SM: 60% **(f)** CM: 60%

(g) IM: 90% **(h)** SM: 90% **(i)** CM: 90%

Figure 6.3: Predictions for temperature for the third mission. Each row corresponds to observing 30%, 60% and 90% of the mission before making predictions. Dashed vertical line marks where the observation time stops and predictions for unseen locations start. The columns represents the three models. Red line shows the prior knowledge, blue line is the true values measured with the AUV and black line is the model prediction. Gray area suggests the uncertainty related to the prediction, which is ±1s.d.

The associated MSE and CRPS scores for the predictions are presented in Table 6.1, Table 6.2 and Table 6.3. As we have seen in Figure 6.3b, Figure 6.3e and Figure 6.3h, the prediction interval with SM is much wider than the other model types. This will give a higher mean CRPS score, as this "safety" of a wide prediction interval is punished with a higher score. The CRPS scores for the first field is therefore higher for SM than the two other models in all three tables. The MSE scores for the first field with SM is lower in Table 6.2 and Table 6.3, which suggests that SM gives a more accurate prediction, though the model is more uncertain about the prediction.

For the second field, we find that the mean CRPS scores in all three tables, Table 6.1, Table 6.2 and Table 6.3, that IM is the preferred model, with SM in a close second. The least preferable model is CM based on this measure. Similarly the MSE scores for the second field suggest that CM is the model performing the worst on this data. IM is the model with the lowest MSE for the second field in all three observation situations, but SM is a close second. In total, we find that with this data there is not enough evidence that multivariate models improve the predictive ability.



(a) IM: 30% **(b)** SM: 30% **(c)** CM: 30%

(d) IM: 60% **(e)** SM: 60% **(f)** CM: 60%

(g) IM: 90% **(h)** SM: 90% **(i)** CM: 90%

Figure 6.4: Predictions for salinity on the third mission. Each row corresponds to observing 30%, 60% and 90% of the mission before making predictions. Dashed vertical line marks where the observation time stops and predictions for unseen locations start. The columns represents the three models. Red line shows the prior knowledge, blue line is the true values measured with the AUV and black line is the model prediction. Gray area suggests the uncertainty related to the prediction, which is ±1s.d.

Table 6.1: Mean MSE and CRPS for predictions of the third mission, when observing 30% of the data with three different model types and for temperature(1st field) and salinity(2nd field).

	MSE		CRPS	
	1st field	2nd field	1st field	2nd field
IM	0.121	1.940	0.220	0.781
SM	0.744	2.46	0.639	0.887
CM	0.0863	3.55	0.165	1.03

Table 6.2: Mean MSE and CRPS for predictions of the third mission, when observing 60% of the data with three different model types and for temperature(1st field) and salinity(2nd field).

	MSE		CRPS	
	1st field	2nd field	1st field	2nd field
IM	0.0656	0.492	0.147	0.392
SM	0.00539	0.497	0.502	0.406
CM	0.214	4.09	0.312	1.23

Table 6.3: Mean MSE and CRPS for predictions of the third mission, when observing 90% of the data with three different model types and for temperature(1st field) and salinity(2nd field).

	MSE		CRPS	
	1st field	2nd field	1st field	2nd field
IM	0.108	0.190	0.228	0.294
SM	0.0618	0.203	0.501	0.315
CM	0.160	0.824	0.238	0.524

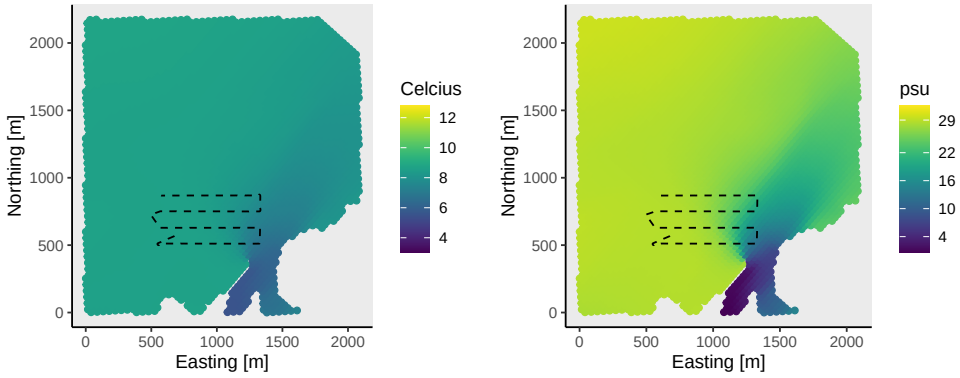
We would like to see how well the models predict on the whole spatial domain, and the results for IM is presented in the following section.

6.5 Spatial Prediction

Before making any observations, the predictions rely fully on the prior knowledge from the SINMOD predictions. These are visualized in Figure 6.5a and Figure 6.5b for temperature and salinity respectively. The uncertainty related to this prediction is mainly decided by the prior distribution of $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$, which are assigned to be

$$\boldsymbol{\beta}^{(j)} \sim \mathrm{N}\left(\begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}10^4 & 0\\0 & 1\end{bmatrix}\right), \quad j = 1,2$$

which is extremely large due to the lack of restriction on $\beta_0^{(j)}$, $j = 1,2$. A plot of this uncertainty is for that reason not included. We present spatial predictions when we have observed 30% (short) and 90% (long) of the total mission with IM.

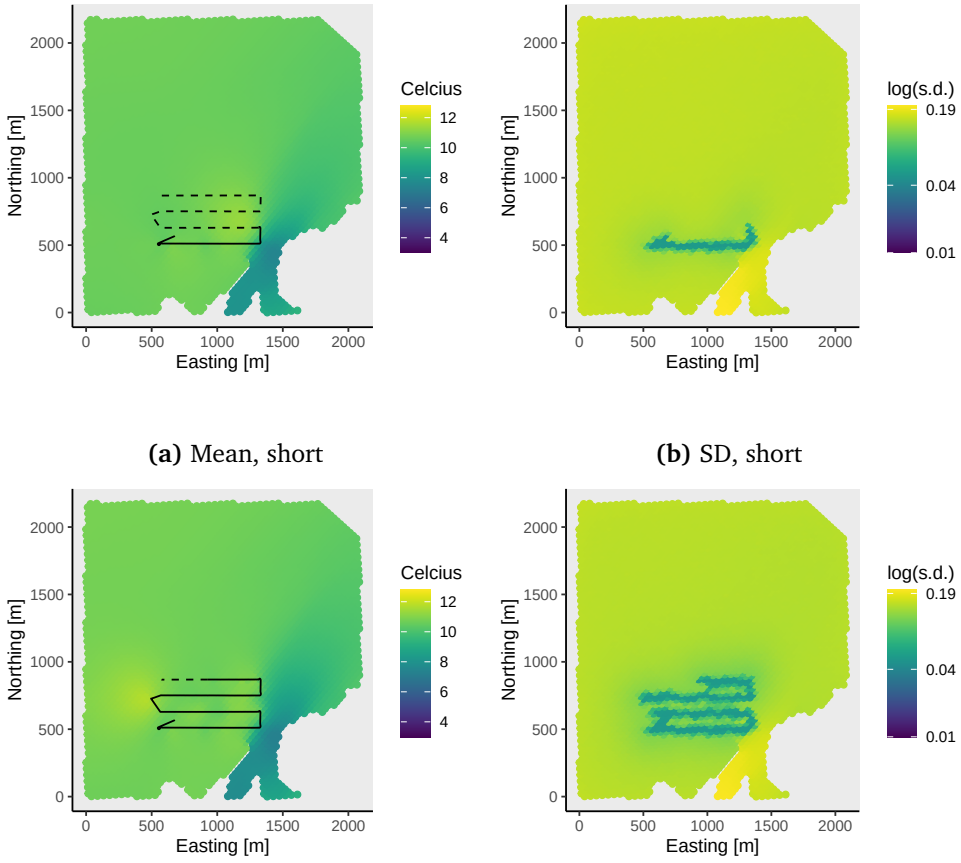


(a) Temperature **(b)** Salinity

Figure 6.5: Prediction for (a) temperature and (b) salinity before any observations are made. Dashed line visualizes the planned path of the mission.

Spatial predictions with IM are presented in Figure 6.6 and Figure 6.7 for temperature and salinity. In addition to predictions, the uncertainty related to each spatial prediction is included. As we saw in previous visualizations, the priors are not accurately predicting the temperature levels, but the locations of the plume is accurate for the mission of interest. We note that the uncertainty for each spatial prediction is tightly related to whether the AUV has visited this location already or not. The spatial uncertainty quickly increases when the distance to the visited locations increases.

Results for spatial predictions with SM and CM can be found in Appendix B. As expected, the predicted fields of temperature and salinity from SM are looking less dependent than the predictions with CM. We observe that SM has larger uncertainty related to all its spatial predictions of temperature further from the observation path. This increase in uncertainty is most likely caused by the marginal variance of the GRF for temperature with SM, which is greater than for IM and CM. This increase in marginal variance also result in more varied predictions, as can be seen for SMs predictions of temperature on the area of of interest. In general, we find that the spatial predictions of temperature with IM and CM are similar, and that SM is unlike the two others. For salinity all three models are making similar spatial predictions and with similar uncertainties related to each prediction.



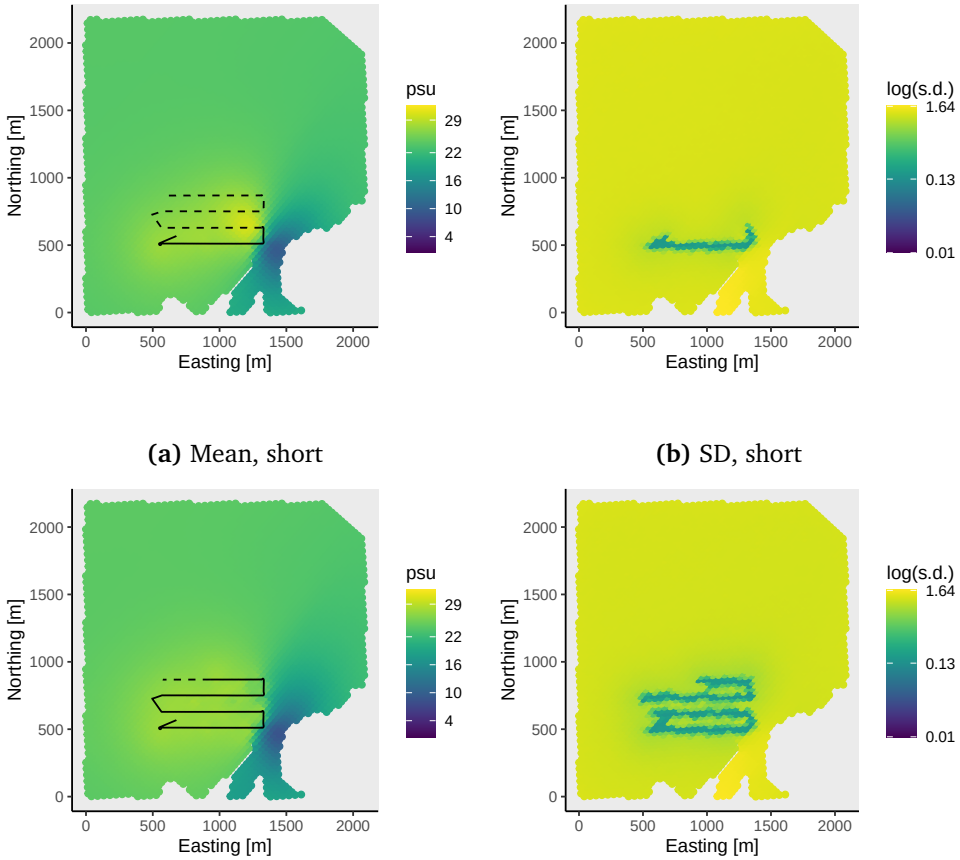
(a) Mean, short

(b) SD, short

(c) Mean, long

(d) SD, long

Figure 6.6: Spatial predictions with IM of temperature when observing (a) 30% and (c) 90% of the total mission data. The associated uncertainty for each prediction is visualized for (b) 30% and (d) 90% and is presented on log-scale. Solid line visualize where observations are made and dashed line is the planned path for the remaining of the mission.



(a) Mean, short

(b) SD, short

(c) Mean, long

(d) SD, long

Figure 6.7: Spatial predictions with IM of salinity when observing (a) 30% and (c) 90% of the total mission data. The associated uncertainty for each prediction is visualized for (b) 30% and (d) 90% and is presented on log-scale. Solid line visualize where observations are made and dashed line is the planned path for the remaining of the mission. The vertical path to the left in the figures (where the uncertainty is greater) is not associated with observations, as this was collected from a different depth.

Chapter 7

Discussion

We obtain a physical interpretation of the parameters related to the SPDE approach to bivariate spatial modeling. More precisely, we found expressions for how the SPDE parameters contribute to the marginal variances of the two fields, and to the marginal cross-covariance. These estimates are accurate, but the parameter estimation is found to be unstable. Three models with different complexity were considered, one univariate and two bivariate models with simple and complex cross-dependency. The models were compared with regards to learning from a numerical model and for how well it performed on real data samples. The numerical model in question is SINMOD, and the data samples are ocean samples of temperature and salinity collected with an AUV.

Learning the spatial dependency for the two fields from the complex numerical model is based on maximum likelihood. The optimization is unstable, which could be caused by a flat objective function. A proposed solution is to implement analytic gradient computations, as the current approach utilize a numerical gradient. For future implementations, we recommend looking into methods for stabilizing the optimization of the likelihood. The simple bivariate model was marginally better, by comparison of MSE and mean CRPS measurements, than the univariate and complex bivariate model, with regards to learning from the numerical model.

We compare the models' ability to learn from observations and make predictions for unobserved locations. We quickly discovered in Chapter 6 that the model has to learn from measurements made both of fresh water and salt water to make good predictions. By only observing one of the levels, we find that the model would have to strictly trust the prior for future observations. The necessity of a good prior becomes critical for accurate predictions, which is not always the case, as seen in the first two missions. When the prior is accurately predicting the location of the plume, we obtain accurate predictions even when observations are few and of only one of the levels. This suggest that the model is too dependent on a prior, which is not always providing a good representation. Then, the learning process becomes more important, and observing both levels of temperature and salinity from fresh water and sea water becomes crucial to obtain accurate predictions.

We did not find evidence for preferring multivariate models for temperature and salinity over univariate models. That is, modeling temperature and salinity separately gives better results with regards to predictions. Due to the increase in model complexity and computation time, and lack of better predictive power, we cannot argue that bivariate models are preferred in our application. Other studies, like Hu, Steinsland et al. (2013) and Hu and Steinsland (2016) found that multivariate models did better than univariate models when the variables of interest are known to be dependent. In the case of missing data from one of the fields, or if one is interested in the joint probability distribution, multivariate models would be favorable.

Simulation from the conditional distribution, $\mathbf{z} \mid (\mathbf{y},)\boldsymbol{\theta}$, gives unrealistic results with the independent model. From a physical point of view, we know that sea water and fresh water will have different temperature and different levels of salinity, and simulating these independently can produce unrealistic fields. In the case of predicting the relation between the two, for example $y_i^{(2)}/y_i^{(1)}$, $i = 1, ..., n^*$, using an independent model could be a bad idea. In the application on the AUV some decision rule for the path could be based on both fields, and a joint distribution would be useful. Currently, there is not such implementation on the AUV, but works in this area opens up new possibilities in the future for potential better solutions.

By increasing the model complexity, an increase in computation time naturally follows. Lilleborge (2021) reported run-times for parameter estimation for the three models in a simulation study. The computation time was more than doubled when expanding from univariate to bivariate models on observations that are known to be correlated. That is, we do suffer a great increase in computation time when the cross-dependency is estimated as well. The bottle neck for the described model is the first objective, where the model must learn the spatial dependency of the fields. With our approach, this can be done before collecting any data, and is only necessary to be executed once. When the SPDE parameters are established, we can update our prior knowledge as observations are collected. This update process is not of great computational cost, and can be done within few minutes on simple laptop.

When evaluating the models' ability to learn the numerical field, we chose two approaches: 5-fold CV and hold-out region CV. The scoring rules were MSE and mean CRPS. 5-fold CV gives us an indication of how the estimation will handle missing data at random. This is not a realistic scenario when working with deterministic predictions from numerical solutions to differential equations. In the desired application of adaptive sampling, hold-out region CV is a more interesting approach as this provides us with information of how sensitive the learning is to the choice of area. In our case, we found that the parameter estimates were highly influenced by the choice of area. Carefully choosing the area of interest when learning the fields become of great importance to obtain reliable parameter estimates. To evaluate predictions for real data, we compared MSE and mean CRPS for predictions forward in time. By choosing time points during the three

missions, we divided the data such that observations from before the chosen time points are used to update the prior, while all observations after the given time point are used as test data. Depending on the application of the model, different approached may be more suitable. If the model only needs accurate predictions for a restricted area surrounding the current position, we will not care as much for the accuracy further from the current area of observations.

To avoid being too dependent on the priors accuracy, we suggest that some measurements from both levels is collected before starting the mission. This will give the model information of the levels and how well the prior represents the truth, and improve predictions when the AUV is collecting data and maneuvering autonomously.

The model currently assumes stationary fields, which includes constant marginal variance. From the prior knowledge and awareness of where the fresh water comes from, we expect less variation in temperature and salinity further away from the plume. Near the plume we expect quick changes in temperature and salinity, and a spatially dependent marginal variance could be included in the model by letting the parameters be spatially dependent. Such extension to the model will acquire some careful consideration of the configuration, to ensure non-negative definite covariance functions, and is not covered in this thesis.

Secondly, the model could be expanded to $\mathbb{R}^3$, where multiple layers in the SINMOD data is considered. The spatial dependency is present between different depths as well, and it might improve the predictive power, at the cost of computational complexity. Also it seems that the fields are varying with time quite drastically. That is, the location of the plume is changing with time, and SINMOD cannot always give an accurate prediction of the plume. It would also be interesting to learn from multiple SINMOD predictions around the time of interest, and possibly locate the plume more accurately.

Lastly, we found that the range parameters are the least identifiable. The estimates for the range parameters are varying the most. This was also discussed by Lilleborge (2021), who found that parameters in the second equation in the system presented in Equation (3.4) are more difficult to identify. For the simple bivariate model we obtained parameter estimates suggesting a range ten times longer than prior knowledge suggested for temperature. In addition the marginal variance estimate for temperature with this model was much larger than anticipated from the SINMOD predictions.

Bibliography

Bates, Douglas, Martin Maechler, Mikael Jagan, Timothy A. Davis, Jens Oehlschlägel and Jason Riedy (2022). *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version >=3.5.0. URL: https://Matrix.R-forge.R-project.org/.

Cressie, N. and C.K. Wikle (2011). *Statistics for Spatio-Temporal Data*. CourseSmart Series. Wiley. ISBN: 9780471692744. URL: https://books.google.no/books?id=-kOC6D0DiNYC.

Eidsvik, Jo (2020). *MASCOT - Maritime Autonomous Sampling and Control*. URL: https://wiki.math.ntnu.no/mascot (visited on 17/04/2022).

Fossum, Trygve Olav, Cédric Travelletti, Jo Eidsvik, David Ginsbourger and Kanna Rajan (2021). 'Learning excursion sets of vector-valued Gaussian random fields for autonomous ocean sampling'. In: *The Annals of Applied Statistics* 15.2, pp. 597–618. DOI: 10.1214/21-AOAS1451. URL: https://doi.org/10.1214/21-AOAS1451.

Gelfand, Alan E., Peter J. Diggle, Montserrat Fuentes and Peter Guttorp, eds. (2010). *Handbook of spatial statistics*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, Boca Raton, FL, pp. xii+607. ISBN: 978-1-4200-7287-7. DOI: 10.1201/9781420072884. URL: https://doi.org/10.1201/9781420072884.

Gneiting, Tilmann, William Kleiber and Martin Schlather (2010). 'Matérn Cross-Covariance Functions for Multivariate Random Fields'. In: *J. Amer. Statist. Assoc.* 105.491, pp. 1167–1177. ISSN: 0162-1459. DOI: 10.1198/jasa.2010.tm09420. URL: https://doi.org/10.1198/jasa.2010.tm09420.

Gneiting, Tilmann and Adrian E. Raftery (2007). 'Strictly Proper Scoring Rules, Prediction, and Estimation'. In: *Journal of the American Statistical Association* 102.477, pp. 359–378. DOI: 10.1198/016214506000001437. eprint: https://doi.org/10.1198/016214506000001437. URL: https://doi.org/10.1198/016214506000001437.

Hu, Xiangping, Daniel Simpson, Finn Lindgren and Håvard Rue (2013). *Multivariate Gaussian Random Fields Using Systems of Stochastic Partial Differential Equations*. arXiv: 1307.1379 [stat.ME].

Hu, Xiangping and Ingelin Steinsland (2016). 'Spatial modeling with system of stochastic partial differential equations'. In: *Wiley Interdiscip. Rev. Comput.*

Stat. 8.2, pp. 112–125. ISSN: 1939-5108. DOI: 10.1002/wics.1378. URL: https://doi.org/10.1002/wics.1378.

Hu, Xiangping, Ingelin Steinsland, Daniel Simpson, Sara Martino and Håvard Rue (2013). *Spatial Modelling of Temperature and Humidity using Systems of Stochastic Partial Differential Equations*. DOI: 10.48550/ARXIV.1307.1402. URL: https://arxiv.org/abs/1307.1402.

Lilleborge, Karina (2021). *Bivariate Spatial Models using Stochastic Partial Differential Equations*. Department of Mathematical Sciences, NTNU. Project thesis.

Lindgren, Finn, Håvard Rue and Johan Lindström (2011). 'An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach'. In: *J. R. Stat. Soc. Ser. B Stat. Methodol.* 73.4. With discussion and a reply by the authors, pp. 423–498. ISSN: 1369-7412. DOI: 10.1111/j.1467-9868.2011.00777.x. URL: https://doi.org/10.1111/j.1467-9868.2011.00777.x.

Lindgren, Georg (2013). *Stationary stochastic processes*. Chapman & Hall/CRC Texts in Statistical Science Series. Theory and applications. CRC Press, Boca Raton, FL, pp. xxviii+347. ISBN: 978-1-4665-5779-6.

Rue, Håvard and Leonhard Held (2005). *Gaussian Markov random fields*. Vol. 104. Monographs on Statistics and Applied Probability. Theory and applications. Chapman & Hall/CRC, Boca Raton, FL, pp. xii+263. ISBN: 978-1-58488-432-3. DOI: 10.1201/9780203492024. URL: https://doi.org/10.1201/9780203492024.

Rue, Håvard, Sara Martino and Nicolas Chopin (Apr. 2009). 'Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations'. In: *Journal of the Royal Statistical Society Series B* 71, pp. 319–392. DOI: 10.1111/j.1467-9868.2008.00700.x.

Schmidt, Alexandra M. and Alan E. Gelfand (2003). 'A Bayesian coregionalization approach for multivariate pollutant data'. In: *Journal of Geophysical Research: Atmospheres* 108.D24. DOI: https://doi.org/10.1029/2002JD002905. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2002JD002905. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2002JD002905.

SINTEF (1987). *SINMOD*. URL: https://www.sintef.no/programvare/sinmod/ (visited on 18/05/2022).

Whittle, P. (Dec. 1954). 'On Stationary Processes in the Plane'. In: *Biometrika* 41.3-4, pp. 434–449. ISSN: 0006-3444. DOI: 10.1093/biomet/41.3-4.434. eprint: https://academic.oup.com/biomet/article-pdf/41/3-4/434/625593/41-3-4-434.pdf. URL: https://doi.org/10.1093/biomet/41.3-4.434.

Whittle, P. (1963). 'Stochastic processes in several dimensions'. In: *Bull. Inst. Internat. Statist.* 40, pp. 974–994.

Appendix A

Physical Interpretation of SPDE Parameters

Every continuous covariance function can be represented through a Fourier integral (Lindgren 2013),

$$r(\mathbf{h}) = \iint_{\mathbb{R}^2} \exp\{\mathrm{i}(\mathbf{h}\cdot\mathbf{w})\} f(\mathbf{w}) d\mathbf{w}$$

for a continuous spectral distribution function $F(\mathbf{w})$, and $f(\mathbf{w})$ is the spectral density function or the power spectrum, and $\cdot$ refers to the inner product between two vectors in $\mathbb{R}^2$.

We would like to find expressions for the marginal variances for the fields generated from the SPDE approach, using the SPDE parameters. To do so, we look at the Fourier transform of the linear set of equations presented in (3.4), and compute the power spectrum (procedure in Section 4.2). The power spectrum takes the form

$$\mathbf{S}_u = \frac{1}{(2\pi)^2} \begin{bmatrix} \frac{1}{\tau_{11}^2(\kappa_{11}^2+|\mathbf{w}|^2)^2} & \frac{-\tau_{21}(\kappa_{21}^2+|\mathbf{w}|^2)}{\tau_{11}^2(\kappa_{11}^2+|\mathbf{w}|^2)^2\tau_{22}(\kappa_{22}^2+|\mathbf{w}|^2)} \\ \frac{-\tau_{21}(\kappa_{21}^2+|\mathbf{w}|^2)^2}{\tau_{11}^2(\kappa_{11}^2+|\mathbf{w}|^2)^2\tau_{22}(\kappa_{22}^2+|\mathbf{w}|^2)} & \frac{1}{\tau_{22}^2(\kappa_{22}^2+|\mathbf{w}|^2)^2}\left(1+\frac{\tau_{21}^2(\kappa_{21}^2+|\mathbf{w}|^2)^2}{\tau_{11}^2(\kappa_{11}^2+|\mathbf{w}|^2)^2}\right) \end{bmatrix},$$

and we can use these expressions to find expressions for the marginal variances, that is $r(\mathbf{0})$. To get such expressions, we have to compute integrals for the following form,

$$\sigma_i^2 = \iint (S_u)_{ii}(\mathbf{w}) d\mathbf{w}$$

where $(S_u)_{ii}$ is the ith element of the diagonal of $\mathbf{S}_u$. The integrals simplifies to fractions with polynomials of order two where only the length of $\mathbf{w}$ appears. By changing to polar coordinates, the azimuth contribution to the integral is simply a

factor of 2π. The marginal variances are found from computation of the integrals

$$\sigma_1^2 = \int_0^\infty \frac{1}{2\pi} \frac{1}{\tau_{11}^2(\kappa_{11}^2 + r^2)^2} r\,dr, \tag{A.1}$$

$$\sigma_2^2 = \int_0^\infty \frac{1}{2\pi} \left(\frac{1}{\tau_{22}^2(\kappa_{22}^2 + r^2)^2} + \frac{\tau_{21}^2(\kappa_{21}^2 + r^2)^2}{\tau_{11}^2\tau_{22}^2(\kappa_{11}^2 + r^2)^2(\kappa_{22}^2 + r^2)^2} \right) r\,dr, \tag{A.2}$$

where the first integral is simply calculated by substituting $u = \kappa_{11}^2 + r^2$, and we find

$$\sigma_1^2 = \frac{1}{4\pi\tau_{11}^2\kappa_{11}^2}. \tag{A.3}$$

The second integral is calculated by finding the partial fractions for the second term. That is, we use the substitution $u = r^2$, and then determine constants A, B, C and D such that

$$\frac{(\kappa_{21}^2 + u)^2}{(\kappa_{11}^2 + u)^2(\kappa_{22}^2 + u)^2} = \frac{A}{(\kappa_{22}^2 + u)} + \frac{B}{(\kappa_{22}^2 + u)^2} + \frac{C}{(\kappa_{11}^2 + u)} + \frac{D}{(\kappa_{11}^2 + u)^2}. \tag{A.4}$$

After some calculus we arrive at

$$\begin{aligned} A &= 2\frac{(\kappa_{22}^2 - \kappa_{21}^2)(\kappa_{21}^2 - \kappa_{11}^2)}{(\kappa_{11}^2 - \kappa_{22}^2)^3}, \\ B &= \frac{(\kappa_{21}^2 - \kappa_{22}^2)^2}{(\kappa_{11}^2 - \kappa_{22}^2)^2}, \\ C &= -A, \\ D &= \frac{(\kappa_{21}^2 - \kappa_{11}^2)^2}{(\kappa_{11}^2 - \kappa_{22}^2)^2}, \end{aligned}$$

where we assume $\kappa_{11} \neq \kappa_{22}$. The marginal variance for the second field becomes

$$\sigma_2^2 = \frac{1}{4\pi}\left\{ \frac{1}{\tau_{22}^2\kappa_{22}^2} + \frac{\tau_{21}^2}{\tau_{11}^2\tau_{22}^2}\left[A(\log\kappa_{11}^2 - \log\kappa_{22}^2) + \frac{B}{\kappa_{22}^2} + \frac{D}{\kappa_{11}^2} \right] \right\}, \quad \kappa_{11} \neq \kappa_{22}. \tag{A.5}$$

We note that for $\kappa_{21} = \kappa_{11}$ (simple model), we get $A = C = D = 0$ and $B = 1$, which gives us the expected marginal variance,

$$\sigma_2^2 = \frac{1}{4\pi\tau_{22}^2\kappa_{22}^2}\left(1 + \frac{\tau_{21}^2}{\tau_{11}^2}\right).$$

If the two fields have the same scale parameter, $\kappa_{11} = \kappa_{22}$, we find

$$\frac{(\kappa_{21}^2 + u)^2}{(\kappa_{11}^2 + u)^4} = \frac{A'}{(\kappa_{11}^2 + u)} + \frac{B'}{(\kappa_{11}^2 + u)^2} + \frac{C'}{(\kappa_{11}^2 + u)^3} + \frac{D'}{(\kappa_{11}^2 + u)^4}, \tag{A.6}$$

which results in the following constants

$$\begin{aligned} A' &= 0 \\ B' &= 1 \\ C' &= -2(\kappa_{11}^2 - \kappa_{21}^2) \\ D' &= (\kappa_{11}^2 - \kappa_{21}^2)^2 \end{aligned}$$

and the marginal variance expressed through the SPDE parameters becomes

$$\sigma_2^2 = \frac{1}{4\pi}\left\{\frac{1}{\tau_{22}^2\kappa_{11}^2} + \frac{\tau_{21}^2}{\tau_{11}^2\tau_{22}^2}\left[\frac{1}{\kappa_{11}^2} - \frac{\kappa_{11}^2 - \kappa_{21}^2}{\kappa_{11}^4} + \frac{(\kappa_{11}^2 - \kappa_{21}^2)^2}{3\kappa_{11}^6}\right]\right\}. \tag{A.7}$$

The cross-covariance can be found by computing the same kind of integral for the off-diagonal entry of the power spectrum matrix. That is,

$$\sigma_{12} = \int_0^\infty \frac{1}{2\pi}\frac{-\tau_{21}}{\tau_{11}^2\tau_{22}}\frac{\kappa_{21}^2 + r^2}{(\kappa_{11}^2 + r^2)^2(\kappa_{22}^2 + r^2)} r\,dr,$$

and by following the same procedure as for the second field marginal variance, we find the partial fractions,

$$\frac{\kappa_{21}^2 + u}{(\kappa_{11}^2 + u)^2(\kappa_{22}^2 + u)} = \frac{\kappa_{22}^2 - \kappa_{21}^2}{(\kappa_{22}^2 - \kappa_{11}^2)^2}\left(\frac{1}{\kappa_{11}^2 + u} - \frac{1}{\kappa_{22}^2 + u}\right) + \frac{\kappa_{21}^2 - \kappa_{11}^2}{(\kappa_{22}^2 - \kappa_{11}^2)(\kappa_{11}^2 + u)^2},$$

where the substitution $u = r^2$ is performed. The computed integral then gives us the cross-covariance in the same spatial location to be

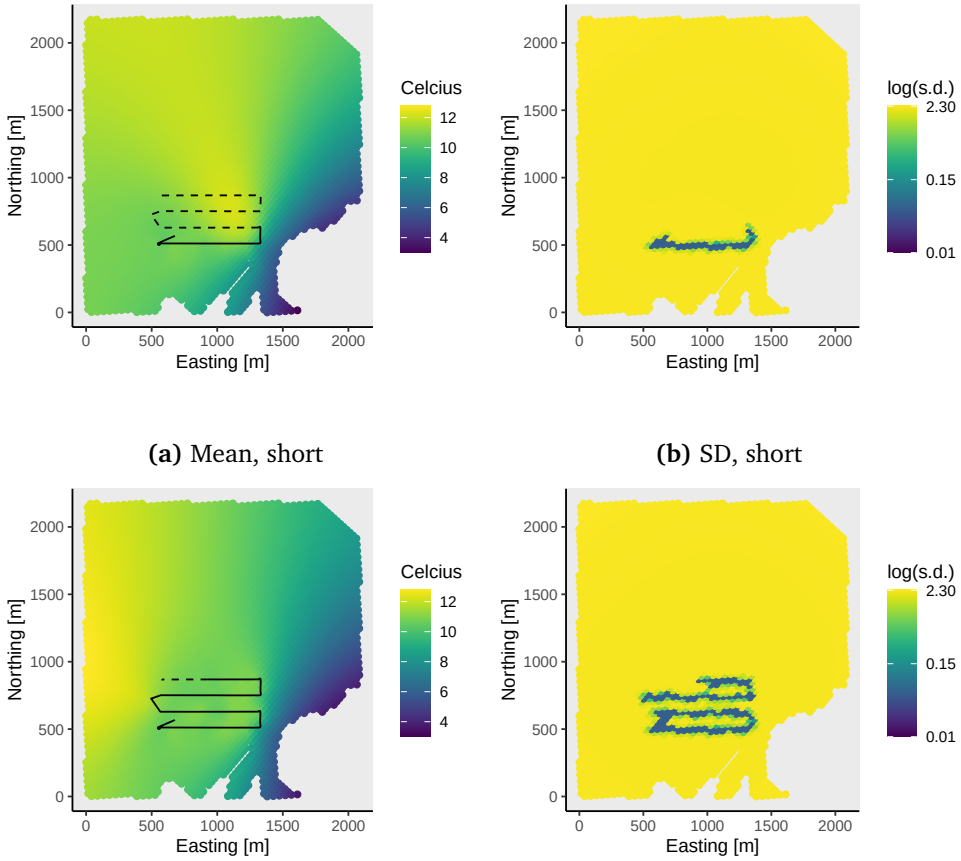
$$\sigma_{12} = \frac{-\tau_{21}}{4\pi\tau_{11}^2\tau_{22}}\left(\frac{\kappa_{22}^2 - \kappa_{21}^2}{(\kappa_{22}^2 - \kappa_{11}^2)^2}(\log\kappa_{22}^2 - \log\kappa_{11}^2) + \frac{\kappa_{21}^2 - \kappa_{11}^2}{\kappa_{11}^2(\kappa_{22}^2 - \kappa_{11}^2)}\right).$$

Appendix B

Spatial Predictions with Bivariate Models

The appendix provides additional spatial predictions with bivariate models, SM and CM. Figure B.1 and B.2 are spatial predictions for temperature and salinity with SM respectively. The predictions are the two fields to the left, and the uncertainty related to each prediction is displayed to the right.

Spatial predictions and uncertainty for CM is presented in Figure B.3 and Figure B.4 for temperature and salinity, respectively. We note that the scale for the uncertainties are different for the models, where Figure B.1b and Figure B.1d are on a larger scale than the uncertainties in Figure B.3b and Figure B.3d.



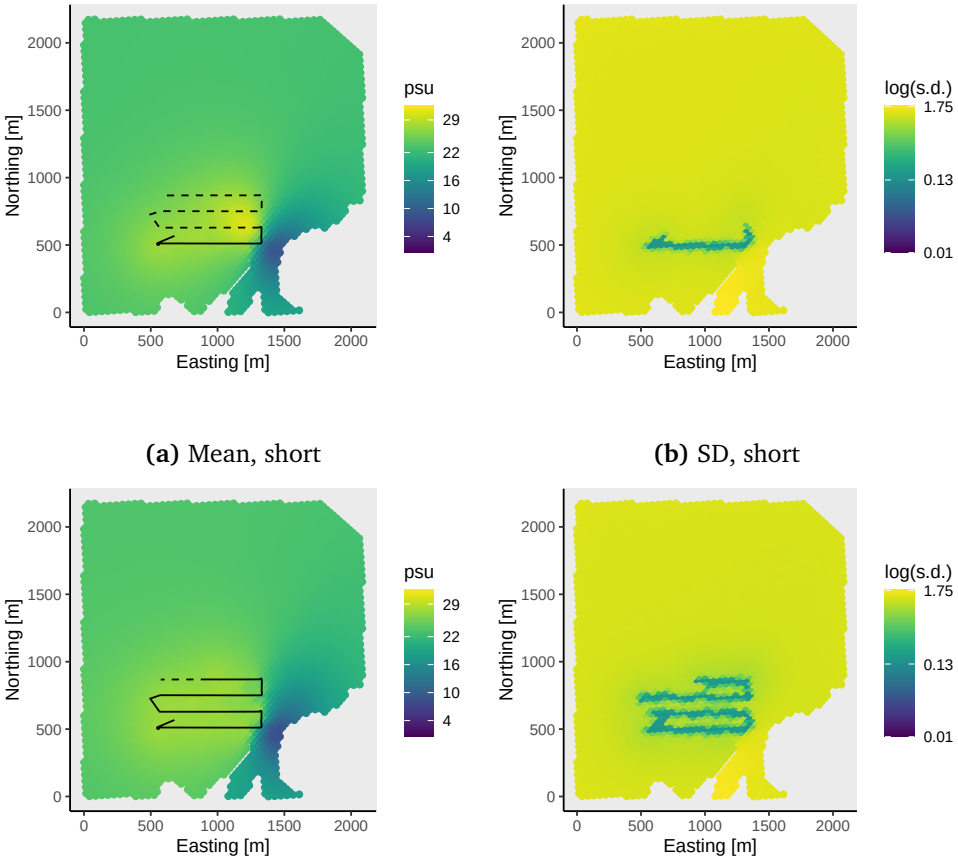
(a) Mean, short

(b) SD, short

(c) Mean, long

(d) SD, long

Figure B.1: Spatial prediction of temperature when observing (a) 30% and (c) 90% of the total mission data with SM. The associated uncertainty for each prediction is visualized for (b) 30% and (d) 90% and is presented on log-scale. Solid line visualize where observations are made and dashed line is the planned path for the remaining of the mission.



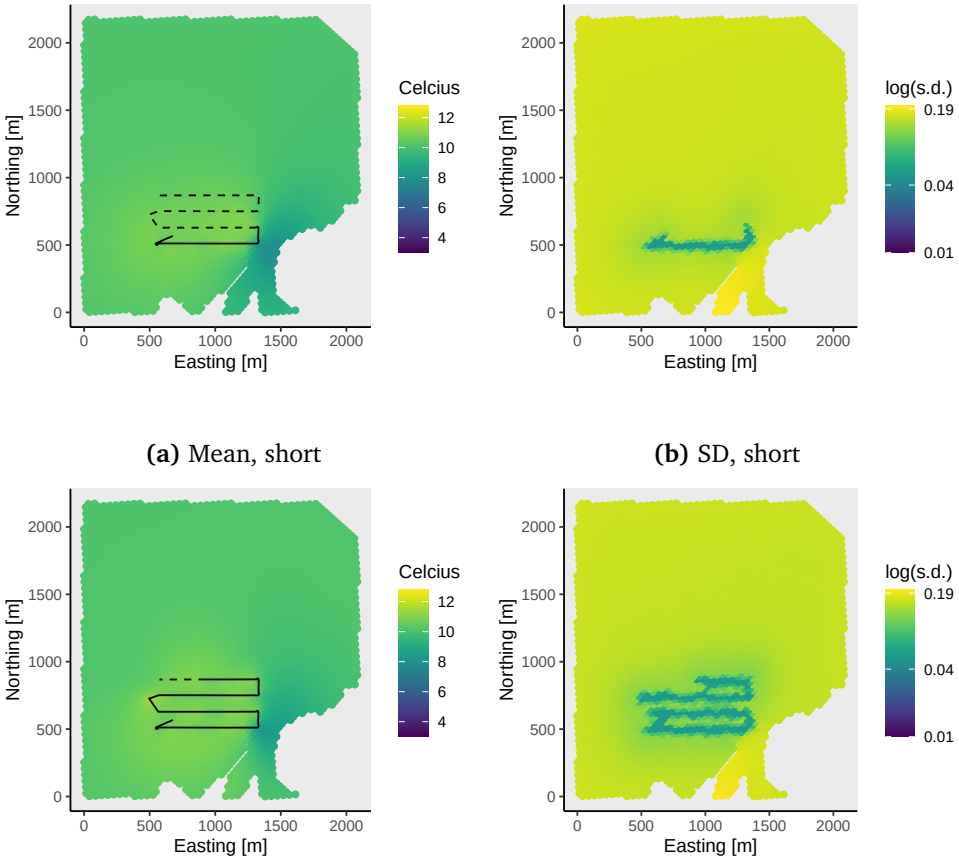
(a) Mean, short

(b) SD, short

(c) Mean, long

(d) SD, long

Figure B.2: Spatial prediction of salinity when observing (a) 30% and (c) 90% of the total mission data with SM. The associated uncertainty for each prediction is visualized for (b) 30% and (d) 90% and is presented on log-scale. Solid line visualize where observations are made and dashed line is the planned path for the remaining of the mission.



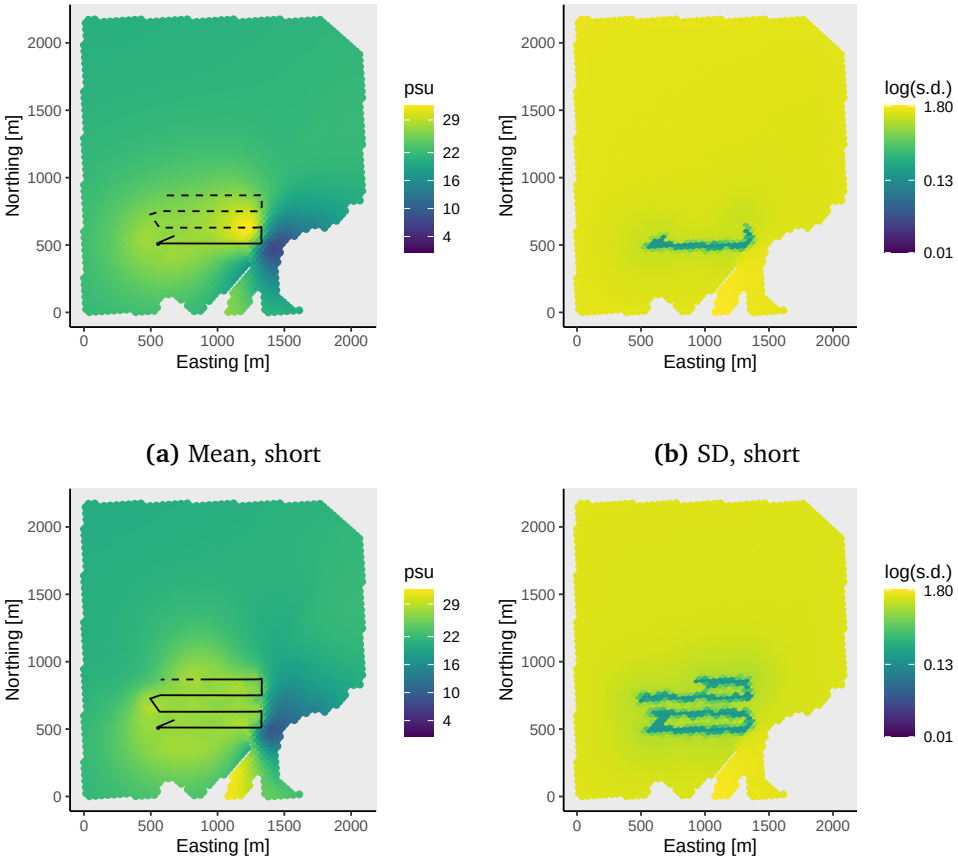
(a) Mean, short

(b) SD, short

(c) Mean, long

(d) SD, long

Figure B.3: Spatial prediction of temperature when observing (a) 30% and (c) 90% of the total mission data with CM. The associated uncertainty for each prediction is visualized for (b) 30% and (d) 90% and is presented on log-scale. Solid line visualize where observations are made and dashed line is the planned path for the remaining of the mission.



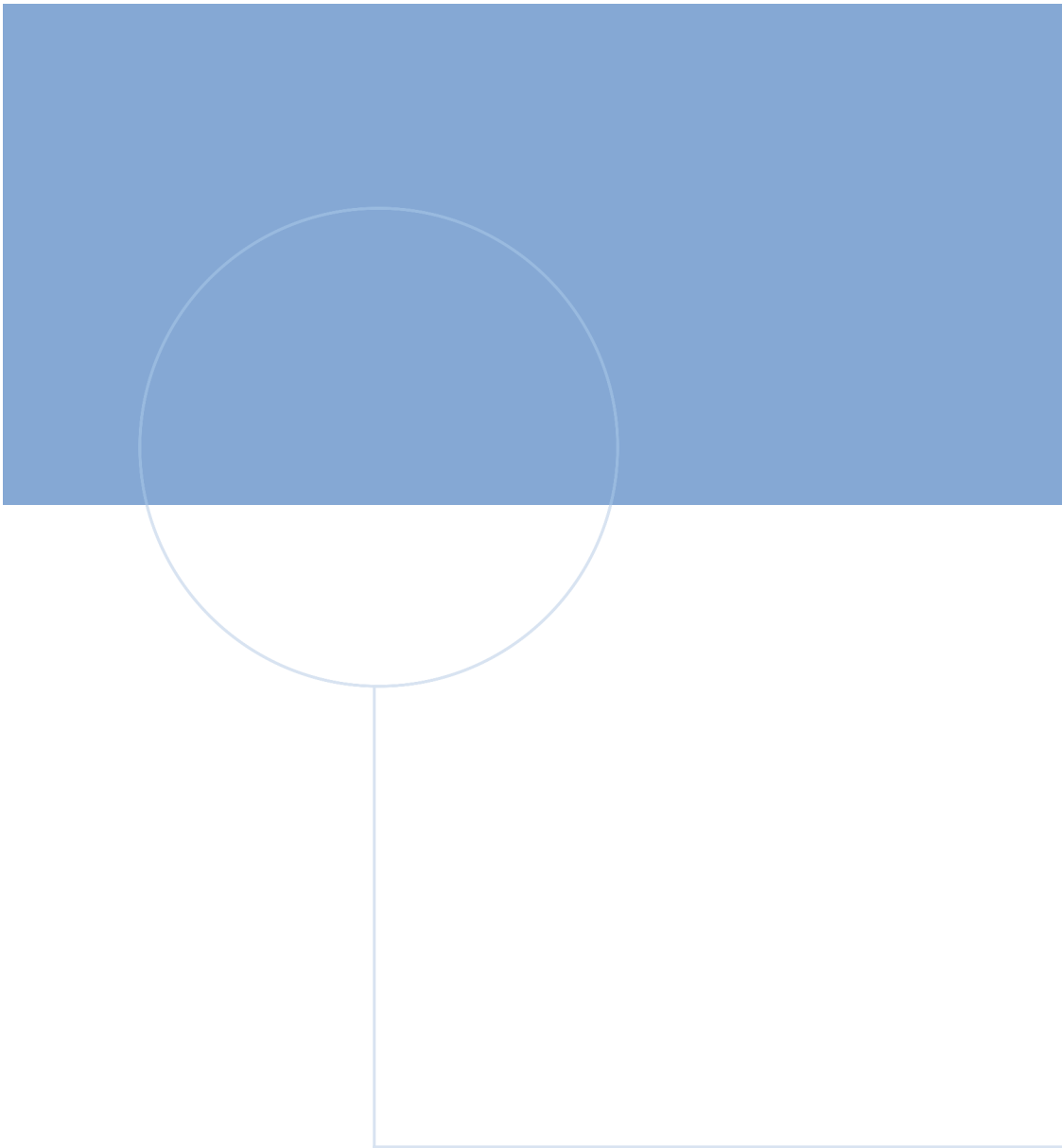
(a) Mean, short

(b) SD, short

(c) Mean, long

(d) SD, long

Figure B.4: Spatial prediction of salinity when observing (a) 30% and (c) 90% of the total mission data with CM. The associated uncertainty for each prediction is visualized for (b) 30% and (d) 90% and is presented on log-scale. Solid line visualize where observations are made and dashed line is the planned path for the remaining of the mission.



NTNU

Norwegian University of
Science and Technology