

Bachelor project

Filip Schjerven

Spring 2017

Contents

0.0.1	Temporary layout	1
1	Preface	3
2	Spatial Gaussian Random Fields	4
2.1	Random Fields	4
2.2	Covariance functions	4
2.3	Gaussian processes	6
3	Spatial Design	7
3.1	Model design	7
3.2	Sampling design	8
4	Case	9
4.1	Case specification	9
4.1.1	Introduction and case goals	9
4.1.2	Input	9
4.1.3	Modelspecification and starting point of analysis	10
4.2	Analysis	10
4.2.1	Goals of analysis	10
4.2.2	Theory	11
4.2.3	Deriving the posterior GRF	12
4.2.4	Chossing prior-parameters	13
5	Conclusion	14
6	References	15

0.0.1 Temporary layout

- Introduction
- Spatial Gaussian Random Fields
- Spatial Design
- Case
- Conclusion

- References

Chapter 1

Preface

Chapter 2

Spatial Gaussian Random Fields

2.1 Random Fields

Dette stykket om random fields må vel forbedres

A random field is a set of random variables $Y(\vec{s}), \vec{s} \in \mathcal{R}^n$ that has a distribution function

$$F(Y(s_1) \leq y_1, \dots, Y(s_n) \leq y_n) \quad (2.1)$$

for any number n . A random field has typically some attribute that connects the random variables in $Y(\vec{s})$. Some examples of such attributes may be physical location, placement in a graph or just ordering in time.

Depending on the properties of the distribution function and the setting in which it is defined, we have many different types of random fields, e.g. Markov Random Fields, Gibbs Random Fields and Conditional Random Fields, all of which may again be used with different attributes connecting the variables of the field.

For this particular project, the main focus will revolve around a random field of type *Gaussian Random Field* (GRF). A GRF is set up so that the variables $Y(\vec{s})$ is governed by a *Gaussian process*. In addition, the Gaussian process **eller feltet?** will be assumed to be stationary, meaning that the associated probability distributions of the Gaussian process is invariant to changes in time. This is an important property when specifying the Gaussian process associated with the GRF.

2.2 Covariance functions

For a GRF with equal variance for all associated variables, two of them being X and Y , we have that the covariance is defined by:

$$\begin{aligned} \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} &= \frac{Cov(X, Y)}{\sigma^2} = Corr(X, Y) \\ \implies Cov(X, Y) &= \sigma^2 Corr(X, Y) \end{aligned}$$

with σ^2 being the variance parameter of the covariance function. So the covariance of two variables are linear their pairwise correlation.

Evidently, by defining a covariance function to use for a Gaussian process, one achieves an expression covariance with fewer parameters to estimate or set. This simplification is desirable and acceptable for many different models, as we now are down to estimate or set what is typically a couple of parameters compared to estimating $\frac{(n-1) \cdot (n-2)}{2}$ covariances for a Gaussian process with n variables.

In a spatial GRF setting as we will work with, the associated variables will be linked with a spatial location. The main argument for calculating the Euclidean distance between points, denoted as $d = |s_X - s_Y|$ where s_A is the spatial location for variable A. Combine this with the fact that we have the Gaussian process as a stationary one, the covariance function simplifies to a single-argument function of d :

$$\text{Corr}(X, Y) = C(|s_X - s_Y|) = C(d) \quad (2.2)$$

Whether this simplification is reasonable comes down to model specification. However, in a spatial setting where one is modelling geophysical attributes that only change significantly in time windows of thousands of years, there is a strong argument for.

For the choice of covariance functions there are many possibilities. A simple function is the exponential correlation function:

$$C(d) = \exp\left(-\frac{d}{\tau}\right)$$

where d is the Euclidean distance between two stochastic points X and Y in the GRF. τ is a *range* parameter, describing the intensity in which a certain should affect the correlation. For easy interpretation with this function, the distance $d = 3\tau$ is of interest as the resulting correlation is $\exp(-3) \approx 0.05$. The exponential covariance function is derived as a special case of the family of *Matern covariance functions*. The Matern covariance functions is defined in a stationary form, with distance d , range parameter τ and degrees of freedom ν , as:

$$C_\nu(d) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{d}{\tau}\right)^\nu K_\nu\left(\sqrt{2\nu} \frac{d}{\tau}\right) \quad (2.3)$$

where K_ν is the modified Bessel function of the second kind. The covariance functions obtained from the Matern family makes sample paths of an associated Gaussian process become $\nu - 1$ differentiable. The equation in 2.3 is seemingly a complicated affair to evaluate, but it is reducible to convenient forms for many choices of ν . Three of them are:

$$\begin{aligned} \nu = \frac{1}{2} : \quad C_{\frac{1}{2}}(d) &= \sigma^2 \exp\left(-\frac{d}{\tau}\right) \\ \nu = \frac{3}{2} : \quad C_{\frac{3}{2}}(d) &= \sigma^2 \left(1 + \frac{\sqrt{3}d}{\tau}\right) \exp\left(-\frac{\sqrt{3}d}{\tau}\right) \\ \nu = \frac{5}{2} : \quad C_{\frac{5}{2}}(d) &= \sigma^2 \left(1 + \frac{\sqrt{5}d}{\tau} + \frac{5d^2}{3\tau^2}\right) \exp\left(-\frac{\sqrt{5}d}{\tau}\right) \end{aligned} \quad (2.4)$$

We recognize the exponential covariance function as the Matern covariance function with d.o.f. $\nu = \frac{1}{2}$. This again implies that we will not have differentiable sample paths when taken from a Gaussian process using this covariance function, something that's illustrated in `eksempel/plot`.

2.3 Gaussian processes

Associated with every GRF is a underlying Gaussian process detailing the distributions of the variables contained in the field. The fundamental characterization of such a process is that all finite-dimensional joint distributions for the variables of the process is multivariate normally distributed (MVN), and especially that each variable is normally distributed. Being distributed as a MVN, a Gaussian process is completely determined by its mean and covariance functions **Her må det vel et par referanser?.** Thus, the mean and covariance functions of a process are often a focus point in dealing with a Gaussian process.

Amongst the many favourable properties of MVN distributions, we also have that the best predictor for an unobserved variable in a Gaussian process a linear function of the remaining observed variables. The predictor being a linear function of MVN distributed variables, the predictor as well is MVN distributed with an adjusted mean and covariance.

More precisely, if $\begin{bmatrix} A \\ B \end{bmatrix}$ is a vector of n stochastic variables associated with a Gaussian process, meaning $\begin{bmatrix} A \\ B \end{bmatrix}$ has a distribution on the form of $\mathcal{MVN}(\begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix}, \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{bmatrix})$, where $\Sigma_{AB} = \Sigma_{AB}^T$ denotes the covariance between the variables in A versus the variables in B . given B , we have exact results for the best linear predictor. $P(A|B = b)$ will then be *MVN* (or single Gaussian distributed if only one variable in A), with corrected mean $\mu_{A|B}$ and variance $\Sigma_{A|B}$ as:

$$\mu_{A|B} = \mu_A + \Sigma_{AB} \cdot \Sigma_{AA}^{-1}(b - \mu_B) \quad (2.5)$$

$$\Sigma_{A|B} = \Sigma_{AA} - \Sigma_{AB} \cdot \Sigma_{AA}^{-1} \cdot \Sigma_{BA} \quad (2.6)$$

In an uformal setting, one may interpret this as a linear smoothing of the parameters for the **reduced? glemt ordet** MVN distribution of A . Especially it's worth noting that since the covariance matrix Σ_{AA} is positive definite, its inverse is as well, meaning that $\Sigma_{AB} \cdot \Sigma_{AA}^{-1} \cdot \Sigma_{BA} \geq 0$. This implies that the adjusted variance of $P(A|B = b)$ is lower or equal to the old:

$$\Sigma_{AA} = \Sigma_{AA} - \Sigma_{AB} \cdot \Sigma_{AA}^{-1} \cdot \Sigma_{BA} \leq \Sigma_{AA} \quad (2.7)$$

The equality will only hold for $\Sigma_{AB} = \mathbf{0}$, which would mean that $Cov(A, B) = 0$. Intuitively, this makes sense as then the realizations of B shouldn't impact the values of A at all.

Evidently from the aforementioned property, when dealing with an Gaussian process it would be favourable to have well-defined expressions for its mean and covariance. One possible solution in the case of the process-covariance is to use a covariance function to define the covariance between variables given some criteria to evaluate.

Verdt å nevne noen flere ting om Gaussiske prosesser?

Chapter 3

Spatial Design

Proactive, retroactive

Minimaliserer forventet standardavvik

3.1 Model design

One common way of defining the model used to relate the process of dataacquirement to the properties of the GRF is by a hierarchical model. A hierarchical model divides the spatial design model into three parts:

- Data model - Describes the relation between process and datasampling
- Process model - Describes the model properties of the GRF
- Prior model - Priors used in the case of Bayesian modelling of parameters

An example of such a model may be given for the case example used to generate the sample paths in **Ref til sample paths eksempel ??**. In this model the parameters of the underlying GP is set, so 'Prior model' is omitted. Denoting Z as the sampled data, Y as the underlying GP and parameters $\mu, \tau, \sigma^2, \phi$:

- Data model: $Z(\vec{s}) \sim \mathcal{N}(F(\vec{s}) \cdot Y(\vec{s}), \phi^2 I(\vec{s}))$
- Process model: $Y(\vec{s}) \sim \mathcal{N}(\vec{\mu}, C_Y(D, \sigma^2, \tau))$

where $\vec{s} \in \mathcal{R}^n$ is the vector of points in question, D denotes a symmetric $n \times n$ matrix of distances between points in the field, F is a matrix indicating sampling of data and $C_Y(\dots)$ the chosen covariance function. For the sample path examples in ?? the three covariance functions in 2.4 have been used.

By using the linearity property of Gaussian variables, we may write the models as linear combinations of their means and covariances, i.e.

- Data model: $Z(\vec{s}) = F(\vec{s}) \cdot Y(\vec{s}) + \epsilon, \quad \epsilon \sim \mathcal{N}(\vec{0}, \phi^2 I(\vec{s}))$
- Process model: $Y(\vec{s}) = \vec{\mu} + \rho, \quad \rho \sim \mathcal{N}(\vec{0}, C_Y(D, \sigma^2, \tau))$

This is a unique property of models with a GRF and Gaussian distributed errors and is of no difference compared to the first model.

3.2 Sampling design

Dette er noe rotete In spatial predictioning, the choice of which spatial locations, or spatial design, to sample data from becomes critical. In order to achieve the best design, one would have to choose in cohesion with the model that is defined for the data situation and as to what criteria the model is trying to satisfy. If the overall goal is to minimize the expected variance over the whole field that is under investigation, the best option in many cases would be to sample a regular spaced grid to get measurements that cover the whole field. Some additional points placed with a random distance and placement to its neighbours in the grid would be effective for estimating parameters of the model, e.g. the nugget-effect, variance of the underlying process or range parameter. However, such measurements can be expensive and time consuming to perform, motivating the consideration of how the models spatial design is defined.

(Kanskje et avsnitt her for å sy de to sammen?)

Having performed the sampling of data and constructed the predictions, one could also evaluate ones spatial design to improve it. This can easily be done by using e.g. cross-validation techniques, so that one may reduce or alter ones design after some criteria. A typical criteria could be to decrease the number of sampling points in the design without significant losses in the information obtained. This could obviously be valuable if the data sampling is expensive or time consuming.

Chapter 4

Case

Motivasjon og kriterie, romlig forventet standardavvik, fjerne usikkerhet i parametervalg

4.1 Case specification

4.1.1 Introduction and case goals

In this case I've been presented with the task of performing a predictive analysis of an attribute that may be modelled by a random field. The scenario is that there is a desire in knowing the sealevel height gradient for a specific physical location, and one would like to perform predictive analysis with the following goals in mind:

- Perform satisfactory accurate predictions
- Minimize sailing needed to predict

4.1.2 Input

The only information that we're presented with beforehand is a variogram computed on the basis of some satellite data. This is however reckoned as slightly unreliable data and must be handled accordingly. The area in which the analysis will be based on is given as a list of xy-coordinates with each coordinate being given a z-variable denoting the numerical value to the attribute of interest. There is a total of 29116 coordinates, ordered into a rectangular grid with size 116 in x-direction and size 251 in y-direction. The directions will informally be called easting and northing when plotting. The variogram is presented in ?? along with the original data which we desire to predict.

Modelspesifikasjon og utgangspunkt

4.1.3 Modelspecification and starting point of analysis

As a necessary preliminary for the predictive analysis the height gradient attribute of all the coordinates is modelled as a stochastic variable. Attaining a Gaussian process to describe trend and variability of model one may view the grid as a GRF. Stationarity is reasonable to assume as the case is of predicting geophysical attributes that are unlikely to change within time-scales up to the thousands of years. The connection between variables in this model will revolve around spatial location within the grid, set in combination iwth stationarity we have that both the trendfunction and covariancefunction may be set to single-argument functions.

The model process may be fitted into the hierarchical model (HM) design described in 3.1. With $Y(\vec{s})$ denoting the gradient height variables at locations \vec{s} , that's set to be distributed according to the Gaussian process, the process model follows as:

$$Y(\vec{s}) \sim \mathcal{MVN}(\mu(\vec{s}), \Sigma(\vec{s})) \quad (4.1)$$

Here $\mu(\vec{s})$ denotes the trendfunction and $\Sigma(\vec{s})$ denotes the covariance function, both given easting-northing coordinates of the variables in question. **Noe mer her? Motivasjon for å velge Gaussisk fordeling**

With only a process model, one could have assumed noise-free sampling of data and started deriving results and predictions for the unobserved variables of the GRF. However, noise-free sampling may seem unreasonable in many settings. Gaussian distributed noise is a common way to model the unreliability of mearsurement equipment and/or random noise. Denoting the actual data being sampled from locations \vec{s} in the GRF as $Z(\vec{s})$, the data model is then defined as:

$$Z(\vec{s}) \sim \mathcal{MVN}(Y(\vec{s}), \epsilon^2 I) \quad (4.2)$$

Evidently, we assume that the noise is distributed independently on each sample with a single constant variance **Hvordan rimelig.**

Utlledning av framgangsmåte

4.2 Analysis

4.2.1 Goals of analysis

As a useful tool when performing the analysis, the goals in ?? will need to be framed to suit how the model has been specified. By doing so, one achieves concrete criterias in which to perform the analysis of data after. The goals in ?? can for instance be rephrased into "Minimizing the expected standard deviation of predictions on the basis of as few measurements as realistically possible", expressed in an equation as: **Her må noe skrives om. Å gjøre ny varians utregning vil være heavy**

$$I = \sum_{z_a} \sum_{i=1}^n \sqrt{\text{Var}_{y_a|z_a, z} P(z_a|z)} \quad (4.3)$$

where z is the original samples, z_a is the possible added samples, y_a denotes the underlying process of data in z_a and I is the total measure. There is trade-off between how many measurements one may be able to perform versus how much information that will give. One may measure this trade-off by the averaged standard deviation of predictions, meaning that we may evaluate said criteria by a cost-function if there were one associated to this analysis.

4.2.2 Theory

With the data model and process model somewhat defined, the issue of choosing trend- and covariancefunctions arises. The discussion of covariancefunctions was covered in ??, and results will be presented for all three covariance functions shown in 2.4. **Si noe om hvorfor disse velges?**

The trendfunction is another matter. As we are unaware of any data in the grid that may be related to the trend of the GRF model, the better option that is left is to estimate the trend based on the datacollection. The trendfunction is defined as a polynomial function of spatial location on the form $\mu(\vec{s}) = X(\vec{s})\vec{\beta}$, where X is a design matrix corresponding to the locations \vec{s} and $\vec{\beta}$ is vector of coefficients. By doing so we may utilize results from the theory of *Generalized Linear Models* (GLM) in the estimation of trend. One particular advantage of using GLM is that we are guaranteed by the Gauss-Markov Theorem that the GLM estimates is the *Best Linear Unbiased Estimator* (BLUE) for the vector of coefficients $\vec{\beta}$.

One of two caveats with using GLM estimates is that one still need to specify the polynomial trendfunction to fit the data, meaning there's still a selection process. As the trendfunction is to specified by spatial location for this case, one must face the second caveat. The GLM estimates are the BLUE only for the data and their associated spatial location used in the estimation, giving no guarantee for the prediction of data not part of the estimates. By this, the trendfunction will be selected to be a polynomial of lower order with simple interaction between easting- and northing-coordinates. This has been chosen as an prediction far away from the sample data will typically be over- or underpredicted by an unacceptable amount using a higher-order polynomial. This is due to the covariates of said polynomial will be fitted to the data that has been sampled, which in this case will be done mainly by specified design of low regularity. In sampling designs with high regularity over the field, this is effect is handled as then the estimation will try to adapt to grid points that are within a certain distance from each other. An example of this issue is shown in figures ?? along with the original data for comparison.

For the acutal selection of model, there has been used a ... R2-squared criteria etc etc

When the samples are drawn from a distribution with known correlation matrix $\Omega = \sigma^2 \cdot W$, with σ^2 is constant finite variance parameter and W the correlation matrix, the estimates of $\vec{\beta}$ converge in distribution to a MVN, or more precisely

$$\hat{\vec{\beta}} \xrightarrow{d} \mathcal{MVN}(\vec{\beta}, (X^T \Omega^{-1} X)^{-1}) \quad (4.4)$$

where X denotes the associated design matrix of the samples estimated.

4.2.3 Deriving the posterior GRF

The ultimate goal of the analysis is to provide a result for the posterior distribution of the underlying Gaussian process, given some sample data, that is $P(Y(\vec{s})|Z(\vec{s}_0))$ where \vec{s} denotes the whole coordinates of the whole field and \vec{s}_0 denotes the coordinates of samples obtained. This distribution is derived as

$$\begin{aligned} P(Y(\vec{s})|Z(\vec{s}_0)) &= \int \int P(Y(\vec{s}), \sigma^2, \tau | Z(\vec{s}_0)) d\sigma^2 d\tau \\ &= \int \int P(Y(\vec{s})|Z(\vec{s}_0), \sigma^2, \tau) \cdot P(\sigma^2, \tau|Z(\vec{s}_0)) d\sigma^2 d\tau \end{aligned}$$

Further, we use Bayes Theorem

$$\begin{aligned} P(\sigma^2, \tau|Z(\vec{s}_0)) &\propto P(Z(\vec{s}_0)|\sigma^2, \tau) \cdot P(\sigma^2, \tau) \\ &\propto P(Z(\vec{s}_0)|\sigma^2, \tau) \cdot P(\sigma^2) \cdot P(\tau) \\ \implies P(\sigma^2, \tau|Z(\vec{s}_0)) &= \frac{P(Z(\vec{s}_0)|\sigma^2, \tau) \cdot P(\sigma^2) \cdot P(\tau)}{\int \int P(Z(\vec{s}_0)|\sigma^2, \tau) \cdot P(\sigma^2) \cdot P(\tau) d\sigma^2 d\tau} \end{aligned} \quad (4.5)$$

This implies that

$$P(Y(\vec{s})|Z(\vec{s}_0)) = \int \int P(Y(\vec{s})|Z(\vec{s}_0), \sigma^2, \tau) \cdot \frac{P(Z(\vec{s}_0)|\sigma^2, \tau) \cdot P(\sigma^2) \cdot P(\tau)}{\int \int P(Z(\vec{s}_0)|\sigma^2, \tau) \cdot P(\sigma^2) \cdot P(\tau) d\sigma^2 d\tau} d\sigma^2 d\tau \quad (4.6)$$

An exact analytical result for the posterior found in both 4.6 and 4.5 may be found under the with the appropriate priors, but to accomodate any type of prior from a probability distribution, the integrals is approximated by a simple numerical procedure. This is done by discretizing the domain of σ^2 and τ and summing over the discretized values with some weight Δ for each discretization to ensure normalization, giving an approximation as

$$\begin{aligned} P(Y(\vec{s})|Z(\vec{s}_0)) &\approx \sum_{\tau_i} \sum_{\sigma_j^2} \Delta_{ij} \cdot P(Y(\vec{s})|Z(\vec{s}_0), \sigma_j^2, \tau_i) \cdot \frac{P(Z(\vec{s}_0)|\sigma_j^2, \tau_i) \cdot P(\sigma_j^2) \cdot P(\tau_i)}{\sum_{\tau_i} \sum_{\sigma_j^2} P(Z(\vec{s}_0)|\sigma_j^2, \tau_i) \cdot P(\sigma_j^2) \cdot P(\tau_i) \cdot \Delta_{ij}} \\ P(Y(\vec{s})|Z(\vec{s}_0)) &\approx \frac{1}{\Pi} \sum_{\tau_i} \sum_{\sigma_j^2} \Delta_{ij} \cdot P(Y(\vec{s})|Z(\vec{s}_0), \sigma_j^2, \tau_i) \cdot P(Z(\vec{s}_0)|\sigma_j^2, \tau_i) \cdot P(\sigma_j^2) \cdot P(\tau_i) \end{aligned}$$

With $\Pi = \sum_{\tau_i} \sum_{\sigma_j^2} P(Z(\vec{s}_0)|\sigma_j^2, \tau_i) \cdot P(\sigma_j^2) \cdot P(\tau_i) \cdot \Delta_{ij}$ as the normalizing constant.

For this integration to be performed, four distributions need to be defined: $P(\sigma_j^2)$ and $P(\tau_i)$ are prior that is set by some opinions of the parameters, $P(Z(\vec{s}_0)|\sigma_j^2, \tau_i)$ is known MVN by the model specification and lastly we have that $P(Y(\vec{s})|Z(\vec{s}_0), \sigma_j^2, \tau_i)$ is MVN by the results presented in 2.3 with $Y(\vec{s})$ as A and $Z(\vec{s}_0)$ as B .

When performing the numerical integration, it is noted that it's the mean and variance of $P(\sigma^2, \tau|Z(\vec{s}_0))$ that's of relevance for the predictions. By similar arguments as above,

one gets:

$$\begin{aligned}
\mathbf{E}(Y(\vec{s})|Z(\vec{s}_0)) &= \int \int \mathbf{E}(Y(\vec{s})|Z(\vec{s}_0), \sigma^2, \tau) \cdot \frac{P(Z(\vec{s}_0)|\sigma^2, \tau) \cdot P(\sigma^2) \cdot P(\tau)}{\int \int P(Z(\vec{s}_0)|\sigma^2, \tau) \cdot P(\sigma^2) \cdot P(\tau) d\sigma^2 d\tau} d\sigma^2 d\tau \\
&\approx \frac{1}{\Pi} \sum_{\tau_i} \sum_{\sigma_j^2} \Delta_{ij} \cdot \mathbf{E}(Y(\vec{s})|Z(\vec{s}_0), \sigma_j^2, \tau_i) \cdot P(Z(\vec{s}_0)|\sigma_j^2, \tau_i) \cdot P(\sigma_j^2) \cdot P(\tau_i) \\
\mathbf{Var}(Y(\vec{s})|Z(\vec{s}_0)) &= \int \int \mathbf{Var}(Y(\vec{s})|Z(\vec{s}_0), \sigma^2, \tau) \cdot \frac{P(Z(\vec{s}_0)|\sigma^2, \tau) \cdot P(\sigma^2) \cdot P(\tau)}{\int \int P(Z(\vec{s}_0)|\sigma^2, \tau) \cdot P(\sigma^2) \cdot P(\tau) d\sigma^2 d\tau} d\sigma^2 d\tau \\
&\approx \frac{1}{\Pi} \sum_{\tau_i} \sum_{\sigma_j^2} \Delta_{ij} \cdot \mathbf{Var}(Y(\vec{s})|Z(\vec{s}_0), \sigma_j^2, \tau_i) \cdot P(Z(\vec{s}_0)|\sigma_j^2, \tau_i) \cdot P(\sigma_j^2) \cdot P(\tau_i)
\end{aligned} \tag{4.7}$$

with Π defined as before. The expressions of $\mathbf{E}(Y(\vec{s})|Z(\vec{s}_0), \sigma^2, \tau)$ and $\mathbf{Var}(Y(\vec{s})|Z(\vec{s}_0), \sigma^2, \tau)$ is given in 2.5 and 2.6.

4.2.4 Chossing prior-parameters

As a consequence of

Resultater

Chapter 5

Conclusion

Summer opp

Alternativer - komputasjonsmessig kostbart å bruke normalfordeling

Chapter 6

References

Hvor nøye med refere til skriverier?