Using Stochastic Partial Differential Equation Models for Spatial Reconstruction of Annual Precipitation

Rikke Ingebrigtsen, Finn Lindgren and Ingelin Steinsland

Abstract This work is motivated by the needs of spatial reconstruction of climate, especially precipitation, in hydropower planning. Traditionally, statistical models for reconstruction of precipitation only include topographical attributes, such as altitude, in the expectation term and not in the dependency structure. These models will therefore often miss the topography's impact on the precipitation level. In this work, we present a model that incorporate topography, not only as an altitude covariate, but also in the covariance structure. Annual precipitation is a non-stationary process which depends locally on the topography and dominating wind direction. To allow for flexible enough modelling of the covariance structure, we will use a stochastic partial differential equation (SPDE) approach to represent the Gaussian random field. The benefit of such an approach is twofold. First, the SPDE allows for a non-stationary precipitation field using covariates that capture the local topographical dependence in a physical meaningful way. Second, we obtain a Markov representation of the Gaussian random field which makes computations feasible. The objective of this work is to reconstruct an expectation map (with uncertainty) of annual precipitation over Southern Norway using data from 2008–2009.

Rikke Ingebrigtsen
Norwegian University of Science and Technology (NTNU), Trondheim, Norway,
e-mail: `Rikke.Ingebrigtsen@math.ntnu.no`

Finn Lindgren
Norwegian University of Science and Technology (NTNU), Trondheim, Norway,
e-mail: `Finn.Lindgren@math.ntnu.no`

Ingelin Steinsland
Norwegian University of Science and Technology (NTNU), Trondheim, Norway,
e-mail: `Ingelin.Steinsland@math.ntnu.no`

Ninth International Geostatistics Congress, Oslo, Norway, June 11. – 15., 2012

1 Introduction

In Norway, most of the power production comes from hydropower. Hydropower is a renewable and clean energy source, and it is desirable to exploit the existing hydropower plants to their full extent. The power production is influenced by changes in the amount of precipitation from season to season, and from year to year. For optimal planning of the power production, spatial reconstruction of precipitation is important, because it is used as input data in hydrological models. However, of even more importance is realistic modelling, and good understanding, of the uncertainty in the precipitation process.

The Norwegian climate can be described by extremely large variations in precipitation. The wettest part of the country is the western part, where the annual normal for one of the weather stations is as high as 3573 mm. The amount of precipitation in this region is among the highest in Europe. The weather station with the lowest annual normal lies in the eastern part of the country. The amount of precipitation there is only 278 mm.

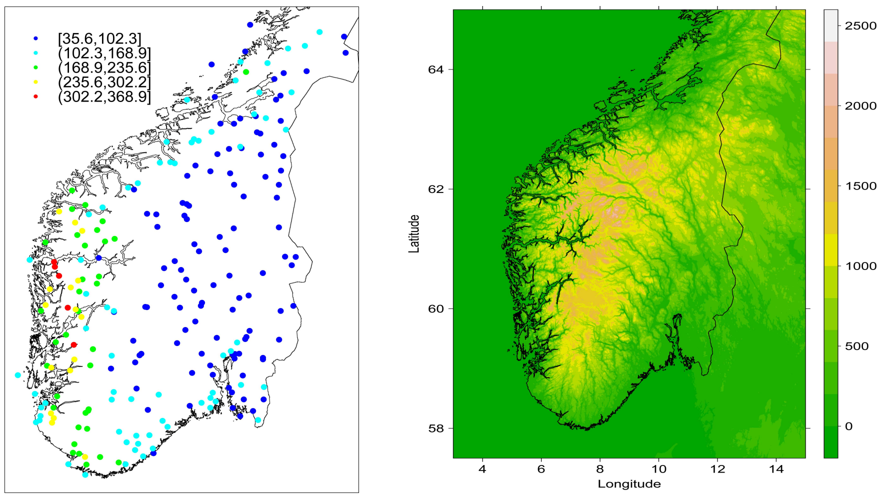
This large difference can be explained by the Norwegian topography and prevailing westerly winds. Norway is a mountainous country, where the western and eastern parts of Southern Norway are separated by the mountain range Langfjella. Due to the prevailing westerly winds, most of the large weather systems that hit Norway come from the west. Hence, the weather systems first reach the steep west coast of Norway. The humid air coming in from over the ocean is forced to ascend because of the topography, and the air will cool down and release precipitation in form of rain or snow. This phenomenon is known as orographic precipitation. The eastern part of Norway is a leeward region in relation to the weather systems coming from the west, which explains the low annual normals in the “rain shadow”.

In this work, we present a Bayesian hierarchical model for the annual precipitation in Southern Norway. We demonstrate how a non-stationary and anisotropic spatial model with dependency structure governed by the local topography can be obtained by the use of a stochastic partial differential equation [2], allowing fast computations based on Gaussian Markov random fields [4]. The Bayesian inference is performed using integrated nested Laplace approximations [5]. The approach is a computationally efficient Bayesian version of geostatistics.

2 Data

The precipitation data used in this study were obtained from a web portal provided by the the Norwegian Meteorological Institute (`eklima.met.no`). Daily precipitation observations, in the one year period 2008-09-01 – 2009-08-31, from stations in Southern Norway (up to and included Nord-Trøndelag) were summarized to obtain the annual values. Stations with incomplete records were removed. The remaining data set consists of observations from 233 weather stations. The annual precipitation data are presented in Fig. 1a.

The data used to model the topography of Southern Norway is based on a 30-arc-second (1-km) gridded, global digital elevation Model [1]. A topographical map can be seen in Fig. 1b.



(a) (b)

Fig. 1: (a) : Annual precipitation observations (in cm) from Southern Norway. The data are from the Norwegian Meteorological Institute and consists of observations from 233 weather stations. (b) : A topographical map over Southern Norway. The elevation is in metres.

3 Theory

The main ingredient in our geostatistical model is a Gaussian random field (GRF). Consider a spatial domain $\mathscr{D} \in \mathbb{R}^d$, where d is typically 2 or 3, then the random field $\{x(\boldsymbol{s}) : \boldsymbol{s} \in \mathscr{D}\}$ is Gaussian if all finite dimensional distributions of the field are Gaussian. More precisely, for all $n \in \mathbb{N}$ and all choices of locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \in \mathscr{D}$ the vector $[x(\boldsymbol{s}_1), \ldots, x(\boldsymbol{s}_n)]$ follows a multivariate Gaussian distribution. The Gaussian distribution is characterized by its mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. A GRF is, in most cases, specified using a mean function $\mu(\cdot)$ and a covariance function $C(\cdot,\cdot)$, which yields $\boldsymbol{\mu} = [\mu(\boldsymbol{s}_i)]_{i=1,\ldots,n}$ and $\boldsymbol{\Sigma} = [C(\boldsymbol{s}_i, \boldsymbol{s}_j)]_{i,j=1,\ldots,n}$ for the finite dimensional distribution. If the covariance function is only a function of the relative position between two locations, i.e. $C(\boldsymbol{s}_i, \boldsymbol{s}_j) = C(\boldsymbol{s}_i - \boldsymbol{s}_j)$, C is said to be stationary. The covariance function is isotropic if it only depends on the Euclidean distance between the locations, i.e. $C(\boldsymbol{s}_i, \boldsymbol{s}_j) = C(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|)$. The covariance function gives the strength of the dependency between two locations, and for a stationary and isotropic field this relationship is the same throughout the domain and in all directions.

Many environmental phenomena are non-stationary and anisotropic by nature. Thus, in these cases, stationary and isotropic GRFs are inappropriate models. There have been numerous suggestions on how to model non-stationarity and anisotropy, and most of them involve some sort of altering of the covariance function. In this work, we present a non-stationary and anisotropic model for the annual precipitation data making use of recent developments in the field of statistics.

In 2011, Lindgren et al. [2] introduced a novel approach to geostatistics: the SPDE approach. The origin of this idea can be dated back to 1954 (and 1963), when Whittle [6, 7] proved that the solution to the following stochastic partial differential equation (SPDE)

$$(\kappa^2-\Delta)^{\alpha/2}x(\boldsymbol{s})=\mathscr{W}(\boldsymbol{s}),\quad \boldsymbol{s}\in\mathbb{R}^d,\quad \alpha=\nu+d/2,\quad \kappa>0,\quad \nu>0,\quad (1)$$

is a Gaussian random field with Matérn covariance function.[1] The innovation process $\mathscr{W}$ on the right hand side of Eq. (1) is spatial Gaussian white noise, and Δ is the Laplace operator.

In [2] this theoretical result is made applicable by the use of a basis function representation of the GRF on a triangulation of the domain $\mathscr{D}$ (see Fig. 2 for an example of a triangulation). The approximate stochastic weak solution to the SPDE provides an explicit link between some Gaussian fields in the Matérn class and the computational more favorable Gaussian Markov random fields (GMRFs). For details about the SPDE approach we refer to the original paper [2], and for an introduction to GMRFs we refer to [4].

One of the main advantages with the SPDE approach is that it allows us to alter the SPDE instead of the covariance function, yielding e.g. GRFs on manifolds. In this work, we modify the SPDE in Eq. 1 to obtain a non-stationary and anisotropic field for the precipitation process. We fix $\alpha=2$, when $d=2$ this is the same as setting $\nu=1$ in the Matérn covariance function. The integer value of ν determines the mean-square differentiability of the field. This parameter influences the predictions made by the model, but it is usually fixed since it is difficult to identify.

The parameter κ is a scaling parameter. However, it is linked to the range ρ by the empirically derived relationship $\rho=\sqrt{8\nu}/\kappa$. Here, the spatial correlation is 0.1 at the distance ρ for a GRF with Matérn covariance with parameters κ and ν. Thus, we can think of κ as a range parameter governing the spatial dependency structure. Let τ be a variance parameter and rescale the field x to obtain the following SPDE

$$(\kappa^2-\Delta)(\tau x(\boldsymbol{s}))=\mathscr{W}(\boldsymbol{s}),\quad \boldsymbol{s}\in\mathbb{R}^2.\quad (2)$$

[1] The Matérn covariance function between locations $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ in $\mathbb{R}^d$ is

$$C(\boldsymbol{s}_1,\boldsymbol{s}_2)=\frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa\|\boldsymbol{s}_2-\boldsymbol{s}_1\|)^\nu K_\nu(\kappa\|\boldsymbol{s}_2-\boldsymbol{s}_1\|),$$

where K_ν is the modified Bessel function of the second kind and order $\nu>0$. $\kappa>0$ is a scaling parameter and σ^2 is the marginal variance.

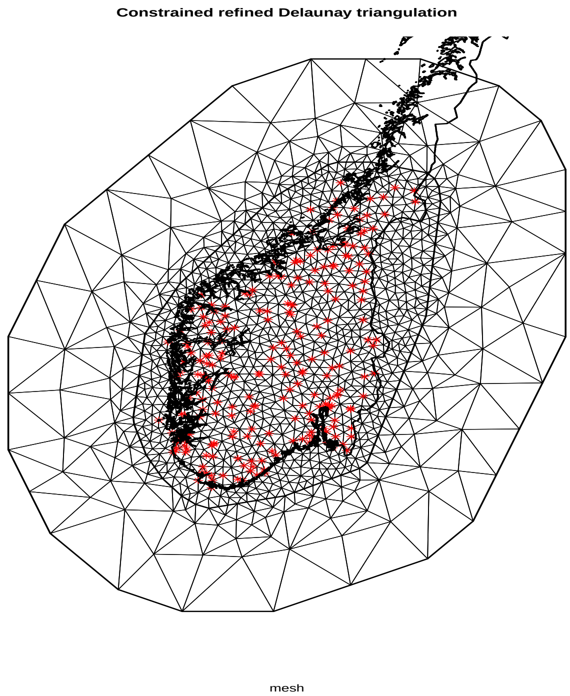


Fig. 2: Triangulation of the spatial domain $\mathscr{D}$, in this case Southern Norway. The red points are the observation locations.

We get a non-stationary and isotropic field by letting the variance and range vary with location. Define $\log\tau(\boldsymbol{s})$ and $\log\kappa(\boldsymbol{s})$ as a sum of basis functions,

$$\log\tau(\boldsymbol{s}) = \sum_{i=1}^{p} B_i^{\tau}(\boldsymbol{s})\theta_i, \quad \log\kappa(\boldsymbol{s}) = \sum_{i=1}^{p} B_i^{\kappa}(\boldsymbol{s})\theta_{i+p}, \tag{3}$$

where the basis functions $B_i^{\tau}(\cdot)$ and $B_i^{\kappa}(\cdot)$ are defined on the triangulated domain, and $\theta_1,\ldots,\theta_{2p}$ are weight parameters.

4 Model

In this section, we present a Bayesian hierarchical model for the annual precipitation data. Let the spatial process $\{\xi(\boldsymbol{s})\colon \boldsymbol{s}\in\mathscr{D}\}$, represent the true level of annual precipitation in Southern Norway. We assume that this process is observed with additive measurement error at the $n = 233$ weather stations. This yields the following data model for the observations $y_1,\ldots,y_{233}$

$$y_i = \xi(\boldsymbol{s}_i) + \varepsilon_i, \quad i = 1, \ldots, 233, \tag{4}$$

where the noise terms $\varepsilon_1, \ldots, \varepsilon_{233}$ are iid $N(0, \sigma_\varepsilon^2)$, and independent of $\xi(\cdot)$.

Furthermore, we assume that the precipitation process can be modelled by three parts: an intercept (β), a smooth effect of altitude (z) and a non-stationary and anisotropic spatial field (x) capturing the spatial dependency structure. The process model can be written as

$$\xi(\boldsymbol{s}) = \beta + z(h_{\boldsymbol{s}}) + x(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathscr{D}, \tag{5}$$

where we assume a RW2 model for z, where $h_{\boldsymbol{s}}$ is the elevation at location $\boldsymbol{s}$, and a spatial field as introduced in Eq. 2 and Eq. 3 for x.

The RW2 model is an intrinsic GMRF [4] with one parameter, the precision, which we denote by τ_z. This model allows estimation of a smooth non-linear effect of station elevation. For identifiability with β, z is constrained to integrate to zero.

The SPDE based spatial model x depends on two parameters: κ and τ. To obtain non-stationarity and anisotropy we let these parameters vary in space as defined in Eq. (3). We introduce topography in the dependency structure by the use of topographical explanatory variables as basis functions on the triangulated domain. We have chosen to use the gradient and elevation field, both based on the digital elevation model [1]. Hence, $p = 3$, with basis functions $B_1^{\tau,\kappa}(\cdot) = 1$, $B_2^{\tau,\kappa}(\cdot) = \texttt{gradient}$ and $B_3^{\tau,\kappa}(\cdot) = \texttt{elevation}$, all defined on the mesh in Fig. 2.

To complete the specification of the Bayesian hierarchical model we need a model for the parameters, $\boldsymbol{\theta} = [\beta, \sigma_\varepsilon^2, \tau_z, \theta_1, \ldots, \theta_6]$. We assume that the parameters are a priori independent with the following distributions

σ_ε^2	Inverse Gamma
β	Gaussian
τ_z	Gamma
$\theta_1, \ldots, \theta_6$	Gaussian

The joint posterior distribution for the parameters $\boldsymbol{\theta}$ and the precipitation process $\boldsymbol{\xi} = [\xi(\boldsymbol{s}_1), \ldots, \xi(\boldsymbol{s}_{233})]$ is given by

$$\pi(\boldsymbol{\theta}, \boldsymbol{\xi} | \boldsymbol{y}) \propto \pi(\boldsymbol{y} | \boldsymbol{\xi}, \boldsymbol{\theta}) \pi(\boldsymbol{\xi} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}). \tag{6}$$

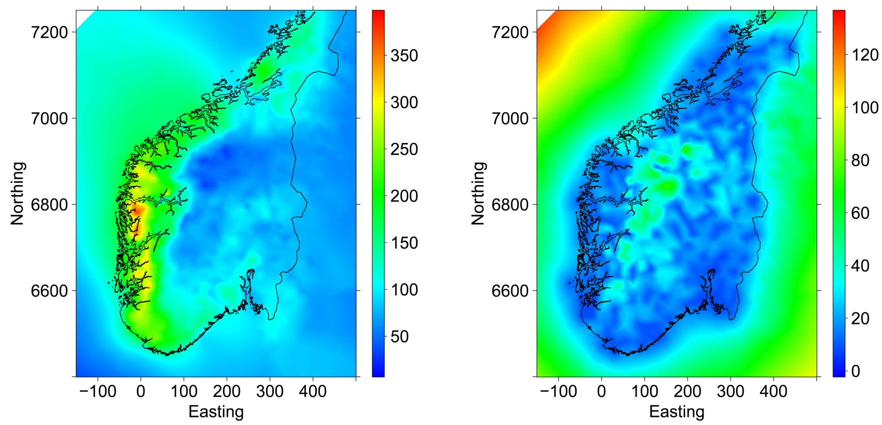
The model defined in this section belongs to a class of latent Gaussian models for which we can estimate the posterior marginals using integrated nested Laplace approximations [5]. These efficient approximations are available in the R [3] package INLA (see `www.r-inla.org`).

5 Results

We have used the R-INLA package to define the model from Sec. 4 and to perform the full Bayesian inference.

In Fig. 3 are the posterior mean and standard deviation of the annual precipitation process ξ. There are clearly different climates in the western and eastern part of Norway. The inland mountain regions are the driest regions, while the wettest region is on the west coast, between the two longest fjords: Hardangerfjorden and Sognefjorden.

It is interesting to observe the standard deviation in Fig. 3b. Clearly, the uncertainty is lowest at the observation locations. However, the interesting part is that the uncertainty is quite high in the mountain regions. This quantifies what meteorologists and hydrologists already know; in the mountain regions there are few observations, and often large differences in altitude between the prediction location and the nearest observation locations, this causes high uncertainty in the prediction.



(a) (b)

Fig. 3: Map of the posterior mean (left) and standard deviation (right) of the annual precipitation process ξ. The annual precipitation is in cm. The coordinate reference system is UTM33 in kilometres.

Fig. 4 illustrates the non-stationarity and anisotropy of the field. The spatial correlation was computed for a reference point in west, east, north and south of Southern Norway. The correlation range is shorter on the west coast, where the terrain is steep, than in the flat region east in Norway. The correlation range is not the same in all directions, and it is shortest in the direction where it hits mountains.

The non-linear effect of altitude, z, can be seen in Fig. 5. There is high uncertainty above 1200 m, because there are no observations at these altitudes.

The posterior estimates of the parameters are in Tab. 1.

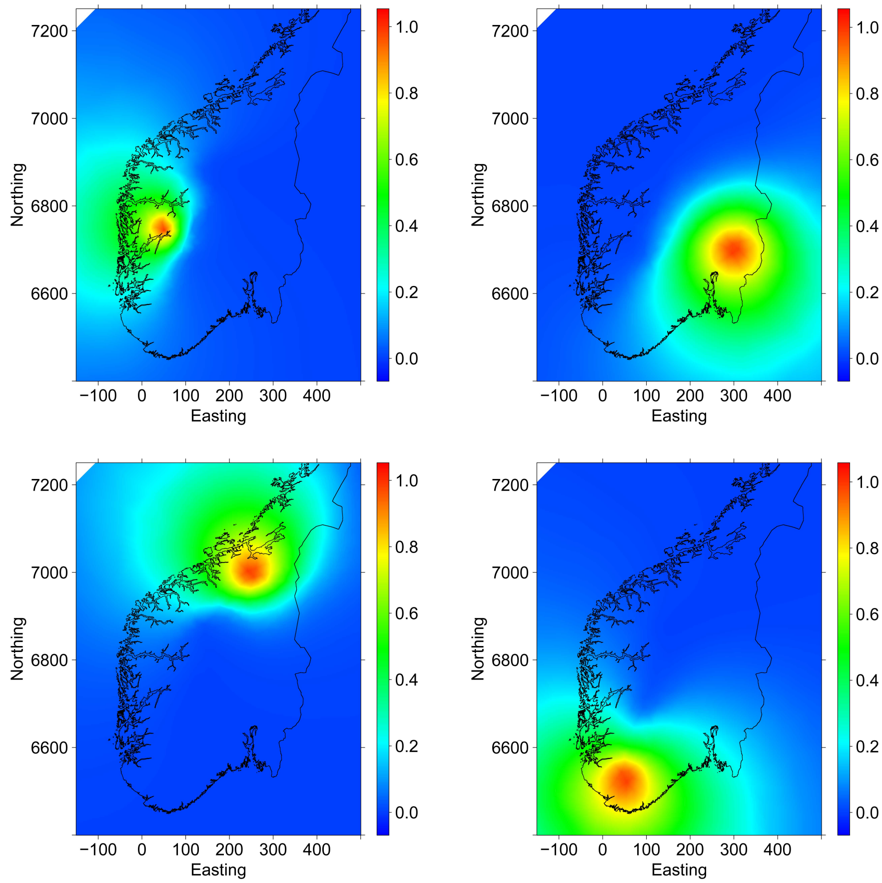


Fig. 4: Spatial correlation between four reference points and all other locations in the domain.

Table 1: Posterior estimates of the parameters

Parameter	Mean	Standard deviation
β	46.70	19.28
σ_{ε}^2	0.008748	0.002274
τ_z	1.188	1.576
θ_1	-0.6577	0.1127
θ_2	-0.4475	0.1215
θ_3	-0.9677	0.2065
θ_4	-4.992	0.4286
θ_5	-1.843	0.8328
θ_6	3.647	0.5889

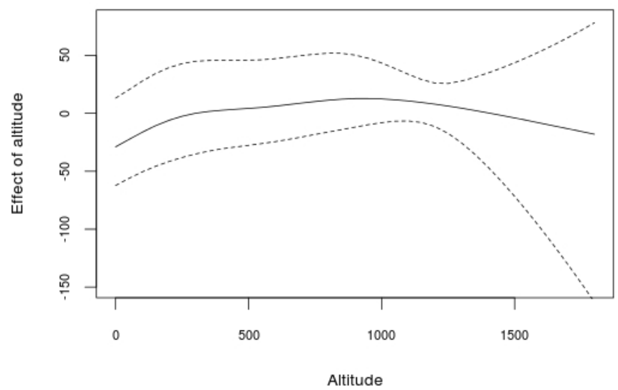


Fig. 5: The estimated effect of altitude, z, posterior mean (solid line) and 95% credible interval (dashed line).

6 Discussion

We have presented a spatial model for annual precipitation in Southern Norway with dependency structure governed by the local topography. The topography is included in the dependency structure, in form of gradient and elevation, by use of the SPDE approach. We obtain a non-stationary and anisotropic model for the annual precipitation. In Fig. 4 it can be seen that the correlation range is shorter on the steep west coast, than on the more flat eastern part of the country. It can also be seen that the mountains in the middle of Norway (Langfjella) prevents an isotropic correlation structure. This is in agreement with the physical knowledge about the precipitation pattern over Norway: the mountains block for dependency between east and west.

We could have included more covariates, e.g. wind direction, in our model. However, the objective has been to study the effect of topography and we have chosen to ignore other covariates in the current study. The model we present is based on data from only one year, it would be interesting to use data from more years, as well as modelling monthly and daily precipitation.

Acknowledgements The research is funded by the Research Council of Norway.

References

1. GLOBE Task Team and others (Hastings, David A., Paula K. Dunbar, Gerald M. Elphingstone, Mark Bootz, Hiroshi Murakami, Hiroshi Maruyama, Hiroshi Masaharu, Peter Holland, John Payne, Nevin A. Bryant, Thomas L. Logan, J.-P. Muller, Gunter Schreier, and John S. MacDonald), eds., 1999. The Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Version 1.0. National Oceanic and Atmospheric Administration, National Geophysical Data Center, 325 Broadway, Boulder, Colorado 80305-3328, U.S.A. Digital data base on the World Wide Web: `http://www.ngdc.noaa.gov/mgg/topo/globe.html` and CD-ROMs.
2. Lindgren F., Rue H., Lindström J.: An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion). J. R. Statist. Soc. B. **73**, Part 4, pp. 423–498 (2011)
3. R Development Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, `http://www.R-project.org/` (2012)
4. Rue H., Held L.: Gaussian Markov Random Fields. Theory and Applications. Chapman & Hall. (2005)
5. Rue H., Martino S., Chopin N.: Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). J. R. Statist. Soc. B. **71**, Part 2, pp. 319–392 (2009)
6. Whittle, P.: On stationary processes in the plane. Biometrika. **41**, pp. 434–449 (1954)
7. Whittle, P.: Stochastic processes in several dimensions. Bull. Inst. Int. Statist. **40**, pp. 974–994 (1963)