NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

Spatial Modelling of Temperature and Humidity using Systems of Stochastic Partial Differential Equations

by

Xiangping Hu, Ingelin Steinsland, Daniel Simpson, Sara Martino and Håvard Rue

PREPRINT
STATISTICS NO. 5/2013

NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This preprint has URL http://www.math.ntnu.no/preprint/statistics/2013/S5-2013.pdf
Xiangping Hu has homepage: http://www.math.ntnu.no/∼xiangph
E-mail: xiangph@math.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and
Technology, N-7491 Trondheim, Norway.

Spatial Modelling of Temperature and Humidity using Systems of Stochastic Partial Differential Equations

Xiangping Hu*[1], Ingelin Steinsland[1], Daniel Simpson[1], Sara Martino[2] and Håvard Rue[1]

[1]Department of Mathematical Science, Norwegian University of Science and Technology, Trondheim, Norway
[2]SINTEF, Trondheim, Norway

July 4, 2013

Abstract

This work is motivated by constructing a weather simulator for precipitation. Temperature and humidity are two of the most important driving forces of precipitation, and the strategy is to have a stochastic model for temperature and humidity, and use a deterministic model to go from these variables to precipitation. Temperature and humidity are empirically positively correlated. Generally speaking, if variables are empirically dependent, then multivariate models should be considered. In this work we model humidity and temperature in southern Norway. We want to construct bivariate Gaussian random fields (GRFs) based on this dataset. The aim of our work is to use the bivariate GRFs to capture both the dependence structure between humidity and temperature as well as their spatial dependencies. One important feature for the dataset is that the humidity and temperature are not necessarily observed at the same locations. Both univariate and bivariate spatial models are fitted and compared. For modeling and inference the SPDE approach for univariate models and the systems of SPDEs approach for multivariate models have been used.

To evaluate the performance of the difference between the univariate and bivariate models, we compare predictive performance using some commonly used scoring rules: mean absolute error, mean-square error and continuous ranked probability score. The results illustrate that we can capture strong positive correlation between the temperature and the humidity. Furthermore, the results also agree with the physical or empirical knowledge. At the end, we conclude that using the bivariate GRFs to model this dataset is superior to the approach with independent univariate GRFs both when evaluating point predictions and for quantifying prediction uncertainty.

Keywords: Spatial statistics, SPDEs, bivariate random fields, covariance matrix, Gaussian random fields, Gaussian Markov random fields

*Corresponding author. Email: `Xiangping.Hu@math.ntnu.no`

1 Introduction

Using spatial statistical models for spatial datasets is of great importance in real-world applications. There are many different approaches for modelling spatial datasets. For instance, Cressie (1993) discussed many of the commonly used statistical methods in spatial statistics. Some theories for kriging were discussed by Stein (1999). Diggle and Ribeiro Jr (2006) gave a detailed discussion for geostatistical datasets from a model-based perspective. A handbook for spatial statistics (Gelfand et al., 2010) gave a comprehensive discussion on different methodologies in spatial statistics. A book written by Cressie and Wikle (2011) was emphasized on discussing on statistics methods for spatio-temporal data.

With the increasing requirement of prediction accuracy in spatial statistics, using *multivariate* models to capture the dependence between the components in the dataset is one of the common approaches to fulfill this requirement if the components are actually or empirically dependent. Multivariate models have been under research for a long time. For instance, Gneiting et al. (2010) and Hu et al. (2012b,a) proposed some methods to build stationary and isotropic models, and Gelfand et al. (2004) and Apanasovich et al. (2012) presented approaches to deal with nonstationarity in multivariate settings. There are many applications for multivariate random fields, such as in economics (Gelfand et al., 2004; Sain and Cressie, 2007), in the area of air quality (Brown et al., 1994; Schmidt and Gelfand, 2003), weather forecasting (Courtier et al., 1998; Reich and Fuentes, 2007) and quantitative genetics (Mcguigan, 2006; Konigsberg and Ousley, 2009).

Generally speaking, if the components in a dataset are empirically dependent, a multivariate model should be taken into consideration. In this paper we model the dataset with humidity and temperature from southern Norway. Since it is known that temperature and humidity are empirically positively correlated, and they are two of the most important driving forces of precipitation, we want to construct bivariate Gaussian random fields (GRFs) for the dataset. The strategy is to have a stochastic model for temperature and humidity, and use a deterministic model to go from these variables to precipitation. The aim of paper is to use the bivariate GRFs to model not only the marginal covariance functions for temperature and humidity but also the cross-covariance function between the temperature and humidity. The posterior mean (given the observations) surfaces of temperature and humidity are reconstructed which can then be used for simulating the precipitation. One important feature of the dataset is that the observations for humidity and temperature are not necessarily measured at the same locations. The dataset is fully discussed in Section 2.

We use the approach proposed by Hu et al. (2012b) to construct bivariate Gaussian random fields for humidity and temperature. With this approach systems of stochastic partial differential equations (SPDEs) are used to build Gaussian random fields for the dataset. There are two main advantages by using the proposed systems of SPDEs approach (Hu et al., 2012b). The first advantage is that the notorious nonnegative definiteness requirement for the covariance matrix is satisfied automatically since the constructed covariance matrix of the GRF from this approach is symmetric positive

Table 1: Number of observations for temperature and humidity

Year	2007	2008	2009	2010	2011
temperature	97	104	111	122	128
humidity	56	63	62	62	70

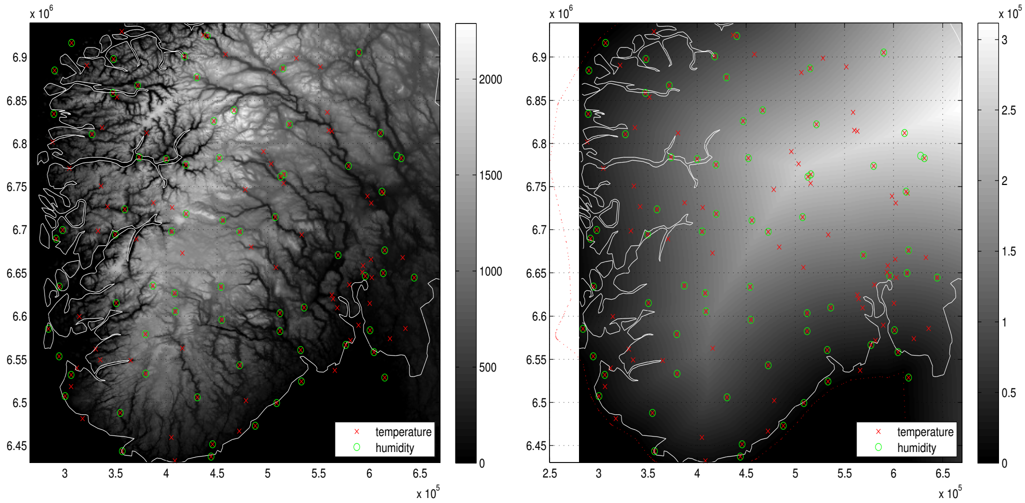
definite. The second advantage is that we can apply Gaussian Markov random field (GMRF) approximation to the constructed GRF. Since GMRF models are usually computationally efficient, this approach can be applied to large datasets. A brief introducton to GMRF is given in Section 3.1, and a detailed discussion about the models with SPDE approach is given in Section 3.2 - Section 3.4.

The rest of this paper is organized as followings. Section 2 describes the data. We review the knowledge about the SPDE approach for spatial statistics and introduce the spatial model for our dataset in Section 3. Section 4 discusses the evaluation procedure. Results are given in Section 5. Section 6 ends the paper with discussion and conclusion.

2 Data

The dataset contains daily mean temperature in Celsius degree and humidity recorded measured in mixing ratio for locations in southern Norway. The mixing ratio of humidity is defined as the mass of water vapor contained in a unit mass of dry air, and hence has a unit kg/kg. Two covariates are also included in the model: elevation at the measurement location and the distance to the ocean from the location. Both covariates are in meters. The dataset contains observations for temperature and humidity on 7th of December from year 2007 to year 2011, i.e. for 5 years. It is important to point it out that the observations are not necessarily at the same locations for all the 5 years. Most of humidity observations are measured at a subset of locations of temperature. Figure 1(a) and Figure 1(b) give an overview of locations for temperature and humidity. In addition, the elevation and distance to ocean are available at all location on a 1km by 1km grid. The elevation map of southern Norway and the distances to ocean are given in Figure 1(a) and Figure 1(b), respectively.

The dotted line is the base line for calculating the distance to ocean and the solid line is the coast line of southern Norway. We can clearly see that the distance to ocean is not the same as the distance to the coast. The cross marks ($\times$) and the circle marks ($\circ$) in Figure 1 are locations for temperature and humidity observations on 7th of December in 2011, respectively. The number of observations of temperature and humidity in different years is given in Table 1. Necessary pre-processing of the dataset has been done before modelling. More information about the pre-processing of the dataset together with some empirical data analysis can be found in Section 5.1.



(a) (b)

Figure 1: Locations of temperature and humidity observations on 7th of December in 2011 with elevation (a) and distances to ocean (b) on a 1km by 1km grid. The base line for calculating the distance to ocean (dotted-line) and the coast line (solid line) of southern Norway are also given. The cross marks ($\times$) and the circle marks ($\circ$) are locations for temperature and humidity observations, repectively.

3 Model using the SPDE approach

In this section we discuss the construction of spatial models for temperature and humidity using the SPDE approach and system of SPDEs approach. Three models are used and fitted to the data. The first model is a univariate GRF model. In this model we construct independent spatial random fields for temperature and humidity with the approach proposed by Lindgren et al. (2011). The second and the third models are bivariate models constructed with the approach proposed by Hu et al. (2012b), where we model temperature and humidity jointly. Since GMRFs are the main tool for achieving computational efficiency with models built by the SPDE approach, a brief introduction to GMRF is given in Section 3.1. The SPDE approach for spatial modelling univariate and multivariate GRFs are described in Section 3.2 and in Section 3.3, respectively. The spatial models used to model temperature and humidity in this paper are presented in Section 3.4.

3.1 Gaussian Markov random fields

A random vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ is a Gaussian random field with mean $\boldsymbol{\mu}$ and precision matrix $\boldsymbol{Q} > 0$ ($\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$) if and only if its density is

$$\pi(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}} |\boldsymbol{Q}|^{1/2} \exp\left(\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{Q}(\boldsymbol{x}-\boldsymbol{\mu})\right). \tag{1}$$

where $\boldsymbol{x}_{-ij}$ denotes for $\boldsymbol{x}_{-\{i,j\}}$. $\boldsymbol{Q} > 0$ denotes the precision matrix $\boldsymbol{Q}$ is positive definite. A Gaussian *Markov* random fields is a GRF with Markov property

$$Q_{ij} = 0 \Longleftrightarrow x_i \perp x_j | \boldsymbol{x}_{-ij}, \tag{2}$$

and hence the precision matrix $\boldsymbol{Q}$ for a GMRF is usually sparse. Therefore, numerical algorithms for sparse matrices can be applied when doing computations. Rue and Held (2005) gives a more detailed discussion on the theories for GMRFs. A condensed version about GMRFs can be found in Gelfand et al. (2010, Chapter 12).

3.2 SPDE approach for univariate GRFs

The main idea of the newly proposed approach by Lindgren et al. (2011) is to use SPDEs to construct GRFs for modelling spatial datasets. The SPDE used in this paper has the form

$$b(\kappa^2 - \Delta)^{\alpha/2} x(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathbb{R}^d, \quad \alpha = \nu + d/2, \quad \nu > 0, \tag{3}$$

where b is a parameter related to the variance of the random field $x(\boldsymbol{s})$, $\mathcal{W}(\boldsymbol{s})$ is a standard Gaussian white noise process, $(\kappa^2 - \Delta)^{\alpha/2}$ is a pseudo (fractional) differential operator and α must be a non-negative integer. Δ is the standard *Laplacian* with definition

$$\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}.$$

With this approach the most important relationship is that the stationary solution $x(\boldsymbol{s})$ to the SPDE (3) is a GRF with a Matérn covariance function. The Matérn covariance function has the form

$$M(\boldsymbol{h}|\nu, \kappa) = \frac{\sigma^2 2^{1-\nu}}{\Gamma(\nu)} (\kappa \|\boldsymbol{h}\|)^{\nu} K_{\nu}(\kappa \|\boldsymbol{h}\|), \tag{4}$$

where ν is the smoothness parameter, κ is the scaling parameter and K_ν is the modified Bessel function of second kind of order ν, $\|h\|$ denotes the Euclidean distance in $\mathbb{R}^d$ and σ^2 is the marginal variance. The closed form for σ^2 of random field $x(\boldsymbol{s})$ constructed from Equation (3) is

$$\sigma^2 = \frac{\Gamma(\nu)/\Gamma(\nu + d/2)}{(4\pi)^{d/2} b^2 \kappa^{2\nu}}.$$

The Matérn covariance function is isotropic and it is widely used in spatial statistics (Stein, 1999; Diggle and Ribeiro Jr, 2006; Simpson et al., 2010; Lindgren et al., 2011; Bolin and Lindgren, 2011; Ingebrigtsen et al., 2013; Hu et al., 2012b,a). We call a GRF with a Matérn covariance function a Matérn random field. Lindgren et al. (2011) pointed out that there was an explicit link between GRFs and GMRFs. They showed that GRFs can be represented by GMRFs. By using this technique, we can build the models theoretically with GRFs but doing computations with GMRFs. We use the finite element methods (FEMs) to solve the SPDE (3), and then apply the GMRF approximation to the solution in order to obtain computationally efficient GMRF models for fast inference. Bolin and Lindgren (2009) showed that the differences between the exact FEM representation and the GMRFs approximation are negligible. We refer to Zienkiewicz et al. (2005) and Bathe (2008) for more information on FEMs. Fuglstad (2011) and Ingebrigtsen et al. (2013) extended the SPDE approach to nonstationary GRFs. Nested SPDEs were proposed by Bolin and Lindgren (2011) for constructing a larger class of models for spatial datasets. Hu et al. (2012b) have extended the approaches from Lindgren et al. (2011) to construct multivariate GRFs. Hu et al. (2012a) proposed to use systems of SPDEs to construct multivariate GRFs with oscillating covariances functions.

Since the smoothness parameter ν is poorly identifiable (Diggle and Ribeiro Jr, 2006; Lindgren et al., 2011), we fix $\alpha_{11} = 2$ and $\alpha_{22} = 2$ when we do inference. With this univariate model for modelling humidity and temperature independently we have 4 parameters $\boldsymbol{\theta} = \{\kappa_{11}, \kappa_{22}, b_{11}, b_{22}\}$. The results for this model are given in Section 5.

3.3 Multivariate GRFs with systems of SPDEs

Hu et al. (2012b) extended the approach given by Lindgren et al. (2011) and proposed a new approach for constructing a multivariate GRF using a system of SPDEs. Hu et al. (2012b) claimed that this approach for constructing multivariate GRFs inherits both theoretical and computational advantages from the approach for univariate GRFs given by Lindgren et al. (2011). The system of SPDEs for constructing a multivariate GRF has the form

$$\begin{pmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \dots & \mathcal{L}_{1p} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \dots & \mathcal{L}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{L}_{p1} & \mathcal{L}_{p2} & \dots & \mathcal{L}_{pp} \end{pmatrix} \begin{pmatrix} x_1(\boldsymbol{s}) \\ x_2(\boldsymbol{s}) \\ \vdots \\ x_p(\boldsymbol{s}) \end{pmatrix} = \begin{pmatrix} \varepsilon_1(\boldsymbol{s}) \\ \varepsilon_2(\boldsymbol{s}) \\ \vdots \\ \varepsilon_p(\boldsymbol{s}) \end{pmatrix}, \tag{5}$$

where $\mathcal{L}_{ij} = b_{ij}(\kappa_{ij}^2 - \Delta)^{\alpha_{ij}/2}$ are similar differential operators as given in Equation (3) with $\{\alpha_{ij} = 0 \text{ or } 2; 1 \leq i, j \leq p\}$, $\{\varepsilon_i(\boldsymbol{s}); i, j = 1, \dots, p\}$ are Gaussian noise processes which are independent but not necessarily identically distributed. It was shown by Hu et al. (2012b) that the solution $\boldsymbol{x}(\boldsymbol{s}) = (x_1(\boldsymbol{s}), x_2(\boldsymbol{s}), \dots, x_p(\boldsymbol{s}))$ to the system of SPDE (5) is a multivariate GRF. The parameters $\{\kappa_{ij}; i, j = 1, \dots, p\}$ and $\{\nu_{ij}; i, j = 1, \dots, p\}$ are scaling parameters and smoothness parameters, respectively. $\{b_{ij}; i, j = 1, \dots, p\}$ are related to both the marginal covariance functions of the fields and the cross-covariance functions among the GRFs. On the theoretical side, similarly as discussed by Lindgren et al. (2011), the precision matrix $\boldsymbol{Q}$ for the multivariate GRF constructed from the system of SPDEs (5) satisfies the positive definite constraint automatically. Hu et al. (2012b) demonstrated that the link between the GMRFs and GRFs could be used, and hence we can construct models with GRFs but use GMRFs for computations. The precision matrix $\boldsymbol{Q}$ of the multivariate GMRF $\boldsymbol{x}(\boldsymbol{s})$ is sparse. Therefore, on the computational side, numerical algorithms for sparse matrices can be applied for sampling and fast inference. These are the main reasons why we have selected this approach for modelling the dataset with temperature and humidity in this paper.

The system of SPDEs we have actually chosen has the form

$$\begin{pmatrix} \mathcal{L}_{11} & \\ \mathcal{L}_{21} & \mathcal{L}_{22} \end{pmatrix} \begin{pmatrix} x_1(\boldsymbol{s}) \\ x_2(\boldsymbol{s}) \end{pmatrix} = \begin{pmatrix} \mathcal{W}_1(\boldsymbol{s}) \\ \mathcal{W}_2(\boldsymbol{s}) \end{pmatrix}, \tag{6}$$

where $\{\mathcal{W}_i(\boldsymbol{s}); i = 1, 2\}$ are standard Gaussian white noise processes. This is a special case of (5) with $\mathcal{L}_{12} = 0$ and $\{\varepsilon_i(\boldsymbol{s}) = \mathcal{W}_i(\boldsymbol{s}); i = 1, 2\}$ when $p = 2$. This system of SPDEs is called triangular system of SPDEs by Hu et al. (2012b,a). The advantage with a triangular systems of SPDEs is that this simplification makes both computations and interpretation easier. We refer to Hu et al. (2012b,a) for a detailed discussion on the triangular system of SPDEs. The smoothness parameters $\{\nu_{ij}; i, j = 1, 2\}$ are poorly identifiable, and we fix $\{\alpha_{ij} = 2; i, j = 1, 2, i > j\}$ (Diggle and Ribeiro Jr, 2006; Lindgren et al., 2011; Hu et al., 2012b,a).

With this setting we know that $x_1(\boldsymbol{s})$ is a Matérn random field and $x_2(\boldsymbol{s})$ is generally not a Matérn random field, but close to a Matérn random field (Hu et al., 2012b). This implies that the order of the random fields matters. Generally speaking, we need to choose the order of the random fields $x_1(\boldsymbol{s})$ and $x_2(\boldsymbol{s})$ for temperature and humidity, and this is usually done by a model selection test. Fit models with both orders and pick the one that minimizes some criterion, such as minimizing prediction error. In this paper we first set the first field $x_1(\boldsymbol{s})$ as temperature and the second field $x_2(\boldsymbol{s})$ as humidity. Then we switch the order of the fields, i.e., we set the first field $x_1(\boldsymbol{s})$ as humidity and the second field $x_2(\boldsymbol{s})$ as temperature.

Using the triangular system of SPDEs (6) for constructing a bivariate GRF, we have 6 parameters to estimate $\boldsymbol{\theta} = \{\kappa_{11}, \kappa_{21}, \kappa_{22}, b_{11}, b_{21}, b_{22}\}$ from the system of SPDEs when we model the temperature and humidity jointly.

Hu et al. (2012b) showed that the sign of cross-correlation between the humidity and temperature is only related to the product $b_{21}b_{22}$ with a triangular system of SPDE. In the extreme case, if b_{21} is zero, i.e., $x_1(\boldsymbol{s})$ and $x_2(\boldsymbol{s})$ are independent, then b_{22} can only be positive value. So restricting b_{22} to be only positive value is a natural choice. Therefore, the sign of the cross-correlation between the two random fields is only related to the sign of b_{21}. When $b_{21} < 0$, $x_1(\boldsymbol{s})$ and $x_2(\boldsymbol{s})$ are positively correlated, and when $b_{21} > 0$, $x_1(\boldsymbol{s})$ and $x_2(\boldsymbol{s})$ are negatively correlated. This setting is chosen in this paper. All results and corresponding discussion are given in Section 5.

3.4 Spatial model for temperature and humidity

As mentioned in Section 2, the dataset contains observations on 7th of December from year 2007 to year 2011 for both temperature and humidity in southern Norway. These observations are not necessary measured at the same locations in each year. We use a model of the following form

$$y_{ijk} = v_{ijk} + \xi_{ijk} + \varepsilon_{ijk}, \tag{7}$$

where i denotes the index of the observation, j denotes the index of the year, k denotes the index of field, v_{ijk} is the fix effect, ξ_{ijk} is the spatial effect and ε_{ijk} is the noise or the measurement error. The noise terms $\{\varepsilon_{ijk}; i = 1, 2, \ldots, n_k, j = 1, 2, \ldots, 5, k = 1, 2\}$ are independent and identically distributed (iid) with Gaussian distribution $\mathcal{N}(\mathbf{0}, \tau_{\epsilon k}^2)$, and are independent of the fix effect and spatial effect. $\{n_k; k = 1, 2\}$ denote the number of observations in all years for temperature and humidity, and $\{\tau_{\epsilon k}^2; k = 1, 2\}$ are the measurement error variances for temperature and humidity. Since we assume that the noise processes for temperature and humidity are independent, the precision matrix $\boldsymbol{Q}_\epsilon$ for the noise processes is a diagonal matrix. Model (7) can be written in vector form

$$\boldsymbol{y} = \boldsymbol{v} + \boldsymbol{A}\boldsymbol{\xi} + \boldsymbol{\varepsilon}, \tag{8}$$

where $\boldsymbol{v} = \boldsymbol{X}\boldsymbol{\beta}$ is the fixed effect with coefficients $\boldsymbol{\beta}$ and design matrix $\boldsymbol{X}$, and it consists of effect from covariates and from the year, i.e the year effect. The matrix $\boldsymbol{A}$ links the dense spatial fields to the observations. $\boldsymbol{\xi}$ is a spatial process with mean zero and precision matrix $\boldsymbol{Q}$. The precision matrix $\boldsymbol{Q}$ is constructed using the system of SPDEs (6) for bivariate model. For the univariate model the precision matrices for temperature and humidity are constructed by Equation (3) independently. $\boldsymbol{\varepsilon}$ is the unexplained random effects for humidity and temperature. This model can be formulated as a Bayesian hierarchical model, and it can be stated explicitly as

- Data model: $y_{ijk}|\eta_{ijk} \sim \mathcal{N}\left(\eta_{ijk}, \tau_{\epsilon k}^2\right)$. We assume that $\{\tau_{\epsilon k}^2; k = 1, 2\}$ are known;
- Process model: $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{A}\boldsymbol{\xi}$, where $\boldsymbol{\xi} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{Q}^{-1}\right)$. As discussed above, the precision matrix $\boldsymbol{Q}$ is constructed by the system of SPDEs (6) and the SPDE (3) for bivariate model and univariate model, respectively. Here we denote $\boldsymbol{\xi} \sim$ BSPDE$(\boldsymbol{b}, \boldsymbol{\kappa})$ which means that the spatial effects are construct by (6) for bivariate model, and correspondingly, we use $\boldsymbol{\xi} \sim$ USPDE$(\boldsymbol{b}, \boldsymbol{\kappa})$ for univariate model constructed from (3). We assume that the fixed effects from the covariates, i.e., elevation and distance to ocean, are the same for all 5 years with coefficients $\{\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}\}$. However, each year has different yearly effect for temperature and humidity $\left\{\beta_{10_j}, \beta_{20_j}; j = 1, 2, \ldots, 5\right\}$ in order to capture the multi-year effect.
- Parameter model: Specify the prior for parameters $\boldsymbol{\theta} = \{b_{11}, b_{21}, b_{22}, \kappa_{11}, \kappa_{21}, \kappa_{22}\}$ from the spatial effects for the bivariate model, and correspondingly for the univariate model $\boldsymbol{\theta} = \{b_{11}, b_{22}, \kappa_{11}, \kappa_{22}\}$. We also need to specify the prior distributions for the coefficients of the covariates $\{\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}\}$ and for the yearly effects $\left\{\beta_{10_j}, \beta_{20_j}; j = 1, 2, \ldots, 5\right\}$ for both the bivariate model and the univariate model.

The prior distributions of parameters are assumed to be independent and have the following distributions (if the parameter is included in the model),

- $\left\{\beta_{10_j}, \beta_{20_j}; j = 1, 2, \ldots, 5\right\}$: Gaussian distributions
- $\{\beta_{11}, \beta_{21}, \beta_{21}, \beta_{22}\}$: Gaussian distributions
- $\{b_{ii}, i = 1, 2\}$: Log-Gaussian distributions
- b_{21}: Gaussian distribution

- $\{\kappa_{11}, \kappa_{21}, \kappa_{22}\}$: Log-Gaussian distributions

This model formulation is similar to the ones in Hu et al. (2012b) and Hu et al. (2012a).

3.5 Statistical inference

We point out that since the coefficient parameters for the covariates can be modelled with Gaussian distributions, we can treat the coefficients $\boldsymbol{\beta}_j$ as the latent field together the spatial process $\boldsymbol{x}(\boldsymbol{s})$ and model them jointly instead of treating the coefficient parameters as hyper-parameters. Thus the hyper-parameters only contains the parameters from the systems of SPDEs, $\boldsymbol{\theta} = \{b_{11}, b_{21}, b_{22}, \kappa_{11}, \kappa_{21}, \kappa_{22}\}$ for bivariate model and $\boldsymbol{\theta} = \{b_{11}, b_{22}, \kappa_{11}, \kappa_{22}\}$ for univariate model, since we fix the values of $\{\alpha_{ij}; i, j = 1, 2\}$ for both the models. The latent field in this case is $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{\beta})^{\mathrm{T}}$, where T denotes the transpose of a vector or a matrix. This can speed up the optimization considerably since there are much few parameters in the numerical optimization. This is the commonly used setting in Rue et al. (2009).

Let $\boldsymbol{Q}(\boldsymbol{\theta})$ denote the precision matrix for the random fields constructed by the system of SPDEs (6) for the bivariate GRFs or the precision matrix for the univariate random fields with SPDE (3) with hyper-parameters $\boldsymbol{\theta}$. With the univariate model we construct the precision matrix $\boldsymbol{Q}(\boldsymbol{\theta})$ as a block diagonal precision matrix, then inference for this two univariate random fields can be done simultaneously. In this case we can use the same program for the bivariate model, and the univariate model has only one more constraint $b_{21} = 0$. Hu et al. (2012b) have shown that from the well known Bayesian formula

$$\pi(\boldsymbol{y}, \boldsymbol{\theta}) = \frac{\pi\left(\boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{y}\right)}{\pi\left(\boldsymbol{z} | \boldsymbol{y}, \boldsymbol{\theta}\right)}. \tag{9}$$

We can derive the posterior distribution for hyper-parameters

$$\begin{aligned} \log\left(\pi\left(\boldsymbol{\theta}|\boldsymbol{y}\right)\right) = \text{Const.} + \log\left(\pi\left(\boldsymbol{\theta}\right)\right) + \frac{1}{2}\log\left(|\boldsymbol{Q}(\boldsymbol{\theta})|\right) \\ - \frac{1}{2}\log\left(|\boldsymbol{Q}_c(\boldsymbol{\theta})|\right) + \frac{1}{2}\boldsymbol{\mu}_c^{\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{Q}_c(\boldsymbol{\theta})\boldsymbol{\mu}_c(\boldsymbol{\theta}), \end{aligned} \tag{10}$$

with $\boldsymbol{\mu}_c = \boldsymbol{Q}_c^{-1}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{Q}_\epsilon\boldsymbol{y}$, $\boldsymbol{Q}_c(\boldsymbol{\theta}) = \boldsymbol{Q}(\boldsymbol{\theta}) + \boldsymbol{C}^{\mathrm{T}}\boldsymbol{Q}_\epsilon\boldsymbol{C}$, and $\boldsymbol{C} = (\boldsymbol{A}, \boldsymbol{X})$. $\boldsymbol{A}$ is a sparse matrix which links the sparse observations of temperature and humidity to our bivariate GRF or univariate GRFs. $\boldsymbol{X}$ is the design matrix.

4 Evaluation

In this section we explain the evaluation schemes for comparing the results for three different settings given in Section 4.2 with three different models. We have two bivariate models. In the first one we set the first random field as temperature and the second random field as humidity. In the second one we switch the order of the random fields, i.e., we set the first random field as humidity and set the second random field as temperature. We also have one univariate model for comparison. Some commonly used scoring rules described in Section 4.1 are chosen in order to compare the predictive performances.

4.1 Scoring rules

In order to evaluate the results, scoring rules are used to compare the predictive performance between the univariate model and the bivariate model for temperature and humidity with different settings. In this paper the commonly used scoring rules mean absolute error (MAE), mean-square error (MSE) and the average of the continuous ranked probability score (CRPS) are used. Let $\hat{y}_{ijk}$ denote the prediction for the observations y_{ijk} for the observation i in year j for the kth field, and then the MAE and MSE for the kth field have the following definitions

$$\begin{aligned} \mathrm{MAE}_k &= \frac{1}{n_k}\sum_j\sum_i |y_{ijk} - \hat{y}_{ijk}|, \\ \mathrm{MSE}_k &= \frac{1}{n_k}\sum_j\sum_i (y_{ijk} - \hat{y}_{ijk})^2, \end{aligned}$$

The CRPS is also a commonly used scoring rule to evaluate the probabilistic forecasts, and it is the integral of the Brier scores for a continuous predictand at all possible threshold values p (Hersbach, 2000; Gneiting et al., 2005). We refer to Toth et al. (2003, Section 7.3.2) for detailed discussion about the Brier scores. Let F denote the predictive cumulative distribution function (CDF) and $H(p-y)$ be the Heaviside function with value 1 whenever $p-y > 0$ and value 0 otherwise. Then the continuous ranked probability score is defined as

$$\mathrm{crps}(F, y) = \int_{-\infty}^{\infty} (F(p) - H(p-y))^2 dp. \tag{11}$$

Gneiting et al. (2005) pointed out that if F is the CDF of a Gaussian distribution, then a closed form of the continuous ranked probability score can be obtained, and this form is usually used in applications. The average of continuous ranked probability score which is denoted as CRPS then has the form

$$\mathrm{CRPS}_k = \frac{1}{n_k}\sum_j\sum_i \mathrm{crps}(F_{ijk}, y_{ijk}). \tag{12}$$

We refer to Gneiting et al. (2005) for more information on scoring rules.

4.2 Cross validation scheme

In this section we explain the cross validation scheme for comparing the results from bivariate models with the results from univariate model. Three different settings have been chosen. We only consider the locations where both observations for temperature and humidity are presented in all settings.

Setting 1

In this setting we only predict the second field $x_2(\boldsymbol{s})$ at 20 locations removed from the dataset for each year from 2007 to 2011. The remaining observations from the second fields $x_2(\boldsymbol{s})$ together with all observations in the first field $x_1(\boldsymbol{s})$ are used to estimate parameters.

Setting 2

This setting is similar as the Setting 1. With this setting, however, 20 locations in each year in the first field $x_1(\boldsymbol{s})$ are left out for prediction and all observations in the second field $x_2(\boldsymbol{s})$ together with the rest of observations in the first field $x_1(\boldsymbol{s})$ are used to estimate the parameters.

Setting 3

In this setting we have left out observations at 20 locations from both fields $x_1(\boldsymbol{s})$ and $x_2(\boldsymbol{s})$ in each year for prediction. The rest of the observations are used to estimate the parameters.

The scoring rules defined in Section 4.1 are calculated for different models for each of these settings for both univariate and bivariate models. The results are presented in Section 5.

5 Results

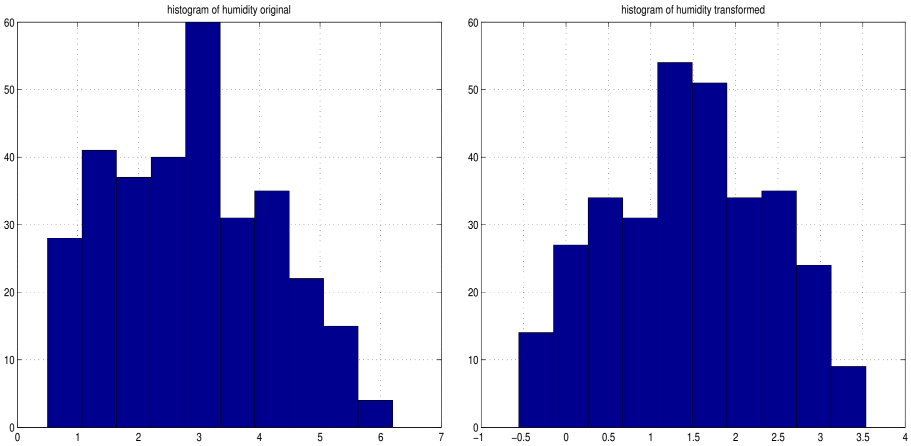
Results from different models with the three different settings are discussed in this section. Before modelling the dataset, some empirical data analysis have been conducted in Section 5.1. Inference results of the parameters are given in Section 5.2. Reconstruction of fields for temperature and humidity are shown in Section 5.3. The results for predictive performance are given in Section 5.4.

5.1 Empirical data analysis

Since the observations of humidity are positive, they are preprocessed with the widely used Box-Cox family of transformations given by Box and Cox (1964), in order to transform them to be approximately Gaussian distributed. We use the original observations of temperature. The Box-Cox family of transformations has the form

$$\hat{Y} = \begin{cases} \left(Y^{\lambda} - 1\right)/\lambda & \text{if } \lambda \neq 0 \\ \log(Y) & \text{if } \lambda = 0 \end{cases}. \tag{13}$$

The estimated values of λ for humidity for the Box-Cox transform is $\lambda = 0.66$, and the histograms for the humidity before the transformation and after the transformation are given in Figure 2(a) and Figure 2(b), respectively. From these two histograms, we can notice that the transformed humidity seems more reasonable to be modelled with Gaussian distribution. For more information about the Box-Cox transformation and other transformation methods, see, for example, Sakia (1992) and Diggle and Ribeiro Jr (2006).



(a) (b)

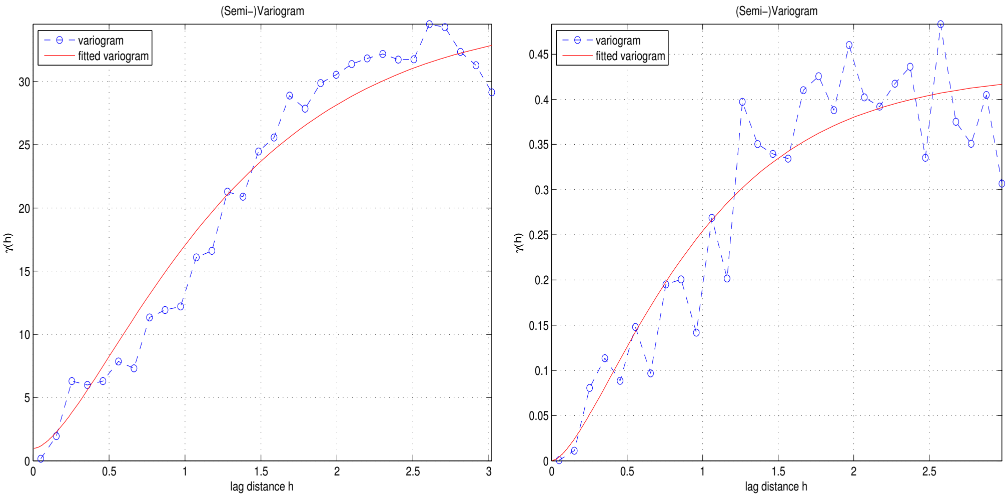
Figure 2: Histograms of original humidity observations(a) and transformed humidity observations (b).

The empirical variograms of both temperature and humidity have been calculated and fitted to theoretical variograms. In the theoretical variograms, we choose to fit the Matérn model. We only include the results for the dataset in 2011 and omit the others since they are similar. The empirical variograms and the corresponding fitted theoretical variograms for this year are shown in Figure 3. This analysis suggest the smoothness parameters for both the fields with $\nu = 1$ are reasonable, and hence fixing $\alpha = 2$ in our analysis is also reasonable. For a detailed discussion on variograms, we refer to Diggle and Ribeiro Jr (2006) and Banerjee et al. (2004). The nugget effects or the measurement error variances for temperature and humidity are assumed known and are fixed to $\tau_1 = 0.1$ and $\tau_2 = 0.01$,respectively, for both the bivariate model and univariate model.

5.2 Inference results of parameters

With Equation (10) together with the priors discussed in Section 3.4, the estimates, with the standard deviations given in the brackets, for the parameters for univariate model and for bivariate model are presented in Table 2 when we set $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity. When we switch the order of the fields, i.e., we set $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature, the corresponding results are given in Table 3. From Table 2 and Table 3 we can notice that temperature and humidity are positively correlated since $b_{21} < 0$ for both models.

With $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity, the correlations within temperature and humidity together with the cross-correlation between temperature and humidity can be calculated from the precision matrix $\boldsymbol{Q}$ for the bivariate model, and are given in Figure 4 (solid lines). The correlations within temperature and humidity for the univariate model can be obtained similarly, and are included in Figure 4 (dash-dot lines). From



(a) (b)

Figure 3: Empirical variograms for temperature and humidity (broken lines with circles) with fitted theoretical variograms (solid lines) for temperature (a) and humidity (b).

Figure 4 we can notice that the cross-correlation between temperature and humidity at the same location is $\gamma_{th} = 0.6351$. When we set $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature, the corresponding results for the correlations within humidity and temperature and cross-correlation between them are given in Figure 5 (solid lines). From Figure 5 we can get that the cross-correlation between the temperature and humidity at the same location is $\gamma_{ht} = 0.6613$, which is slightly higher than the previous model. From these results we notice that the cross-correlation between temperature and humidity are relatively high and indeed needed to be considered.

We define the correlation range as the correlation near 0.1 at distance ρ. The results of the correlation ranges for the above mentioned two models are given in Table 4 and Table 5. From these two tables we can notice that the correlation range for the temperature has been increased when we set $\boldsymbol{x}_1(\boldsymbol{s})$ as the humidity. The same conclusion can be drawn for the range of the cross-correlation between the humidity and the temperature with $\boldsymbol{x}_1(\boldsymbol{s})$ as humidity. The correlation range for the humidity, however, has been decreased when we set $\boldsymbol{x}_1(\boldsymbol{s})$ as the humidity. These differences indicates that we might need to investigate how to set the order of the fields in real-world applications when we used the triangular system of SPDEs (6). We return to this issue in Section 5.4.

As pointed out in Section 3.4, we treat the coefficients for the covariates and the yearly effects of temperature and humidity as parts of the latent field $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{\beta})^{\mathrm{T}}$. It

Table 2: Estimated hyper-parameters for bivariate model and for univariate model with $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity

	Bivariate	Univariate
b_{11}	$0.0104\ (8.106 \times 10^{-4})$	$0.0104\ (8.134 \times 10^{-4})$
b_{21}	$-0.0219\ (2.512 \times 10^{-3})$	
b_{22}	$0.3132\ (2.035 \times 10^{-2})$	$0.2149\ (1.460 \times 10^{-2})$
κ_{11}	7.6867 (0.6428)	7.6721 (0.6427)
κ_{21}	3.2291 (0.6037)	
κ_{22}	2.7981 (0.401)	3.1969 (0.2659)

Table 3: Estimated hyper-parameters for bivariate model and for univariate model with $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature

	Bivariate	Univariate
b_{11}	$0.1711\ (1.787 \times 10^{-2})$	$0.2149\ (1.460 \times 10^{-2})$
b_{21}	$-0.2230\ (2.875 \times 10^{-2})$	
b_{22}	$0.0198\ (2.843 \times 10^{-3})$	$0.0104\ (8.134 \times 10^{-4})$
κ_{11}	3.9537 (0.4118)	3.1969 (0.2659)
κ_{21}	2.5362 (0.5917)	
κ_{22}	5.6358 (0.7176)	7.6721 (0.6427)

Table 4: Correlation ranges for bivariate model and univariate model with $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity.

	ρ_t	ρ_h	ρ_{th}
Bivariate Model	39.4km	90.7km	35.2km
Univariate Model	39.4km	94.9km	

Table 5: Correlation ranges for bivariate model and univariate model with $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature.

	ρ_t	ρ_h	ρ_{ht}
Bivariate Model	43.7km	76.7km	43.8km
Univariate Model	39.4km	94.9km	

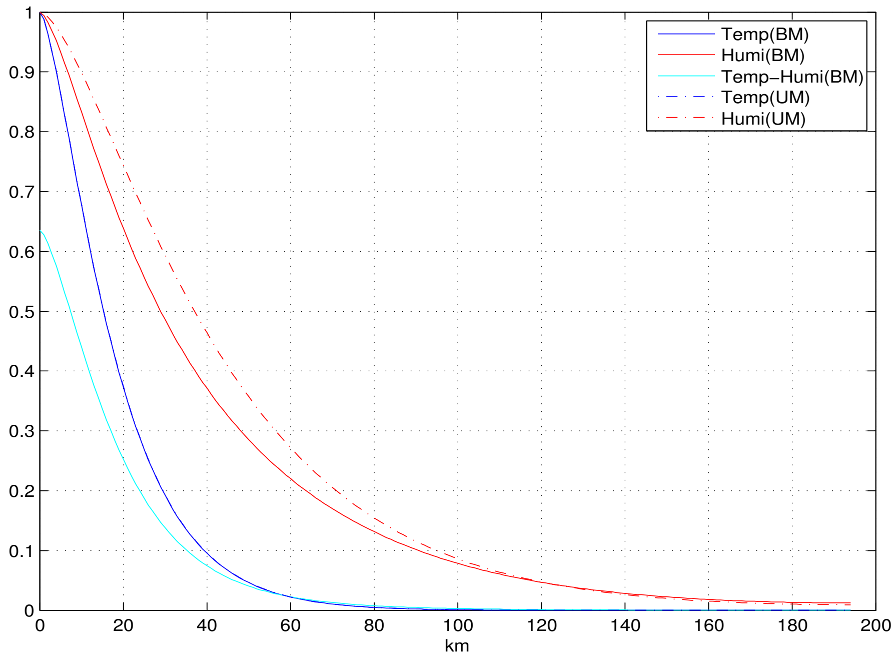


Figure 4: Correlations within temperature and humidity and cross-correlation between temperature and humidity for bivariate model together with the correlations within temperature and humidity for the univariate model when we set $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity. "BM" means Bivariate model. "UM" means univariate model. "Temp" means correlation within temperature. "Humi" means correlation within humidity. "Temp-Humi" means the cross-correlation between the temperature and humidity.

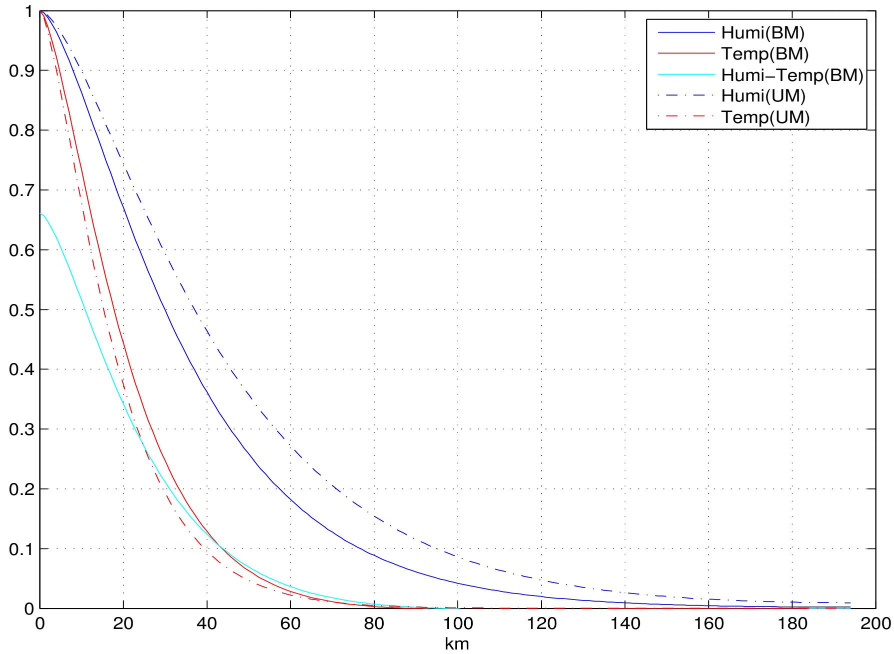


Figure 5: Correlations within temperature and humidity and cross-correlation between temperature and humidity for bivariate model together with the correlations within temperature and humidity for the univariate model when we set $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature.. "BM" means Bivariate model. "UM" means univariate model. "Temp" means correlation within temperature. "Humi" means correlation within humidity. "Humi-Temp" means the cross-correlation between the temperature and humidity.

Table 6: Estimated yearly effects for different years with bivariate model and univariate model with $\boldsymbol{x}_1$ as temperature and $\boldsymbol{x}_2$ as humidity

Model	Parameter	2007	2008	2009	2010	2011
Bivariate	β_{10}	8.808 (0.546)	-0.025 (0.541)	8.566 (0.532)	-5.981 (0.525)	0.894 (0.519)
	β_{20}	3.037 (0.112)	1.817 (0.112)	2.859 (0.111)	1.153 (0.111)	1.665 (0.110)
Univariate	β_{10}	8.794 (0.547)	-0.045 (0.542)	8.537 (0.532)	-6.004 (0.526)	0.874 (0.519)
	β_{20}	3.081 (0.118)	1.844 (0.118)	2.899 (0.117)	1.171 (0.117)	1.698 (0.116)

can be shown that

$$\begin{aligned}\pi(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{\theta}) &\propto \pi(\boldsymbol{z},\boldsymbol{y}|\boldsymbol{\theta}) \\ &= \pi(\boldsymbol{z}|\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{z},\boldsymbol{\theta}) \\ &\propto \exp\left(-\frac{1}{2}\left[\boldsymbol{z}^{\mathrm{T}}(\boldsymbol{Q}(\boldsymbol{\theta})+\boldsymbol{C}^{\mathrm{T}}\boldsymbol{Q}_n\boldsymbol{C})\boldsymbol{z}-2\boldsymbol{z}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{Q}_n\boldsymbol{y}\right]\right),\end{aligned} \tag{14}$$

and

$$\boldsymbol{z}|\boldsymbol{y},\boldsymbol{\theta} \sim \mathcal{N}\left(\boldsymbol{\mu}_c(\boldsymbol{\theta}),\boldsymbol{Q}_c(\boldsymbol{\theta})\right), \tag{15}$$

with $\boldsymbol{\mu}_c$, $\boldsymbol{Q}_c$ and $\boldsymbol{C}$ given in Section 3.4. From Equation (14) we can get the estimates for the yearly effects and for the coefficients of the covariates. When we set $\boldsymbol{x}_1$ as temperature and $\boldsymbol{x}_2$ as humidity, the estimates for the yearly effects are given in Table 6 with the standard deviations given in the brackets. Table 6 shows that the yearly effects are quite different. This explains the high temperature in 2007 but low temperature in 2010. The estimates of the coefficients of the covariates are given in Table 7. We can notice that the two covariates give negative contribution to both fields. When we set $\boldsymbol{x}_1$ as the humidity and $\boldsymbol{x}_2$ as the temperature, the corresponding results for the yearly effects and the coefficients for the covariates are given in Table 8 and Table 9. Similar conclusions can be drawn from Table 8 and Table 9 as from Table 6 and Table 7. These results agree with the empirical results, and we summarize as follows.

- The higher the elevation, the lower the temperature;
- The higher the elevation, the lower the humidity.
- The longer the distance to ocean, the lower the temperature;
- The longer the distance to ocean, the lower the humidity;

Table 7: Estimated coefficients for covariates and standard deviations with $\boldsymbol{x}_1$ as temperature and $\boldsymbol{x}_2$ as humidity

Model	Parameter	Estimate	Std. dev.
Bivariate	β_{11}	-6.825	0.683
	β_{12}	-9.859	0.721
	β_{21}	-0.199	0.092
	β_{22}	-1.460	0.128
Univariate	β_{11}	-6.832	0.683
	β_{12}	-9.815	0.721
	β_{21}	-0.493	0.107
	β_{22}	-1.438	0.138

Table 8: Estimated yearly effects for different years with bivariate model and univariate model with $\boldsymbol{x}_1$ as humidity and $\boldsymbol{x}_2$ as temperature

Model	Parameter	2007	2008	2009	2010	2011
Bivariate	β_{10}	3.031	1.819	2.852	1.158	1.663
		(0.101)	(0.100)	(0.099)	(0.099)	(0.098)
	β_{20}	8.843	0.048	8.634	-5.891	0.948
		(0.586)	(0.581)	(0.572)	(0.565)	(0.558)
Univariate	β_{10}	3.081	1.844	2.899	1.171	1.698
		(0.118)	(0.118)	(0.117)	(0.117)	(0.116)
	β_{20}	8.794	-0.045	8.537	-6.004	0.874
		(0.547)	(0.542)	(0.532)	(0.526)	(0.519)

Table 9: Estimated coefficients for the covariates and standard deviations with $\boldsymbol{x}_1$ as humidity and $\boldsymbol{x}_2$ as temperature

Model	Parameter	Estimate	Std. dev.
Bivariate	β_{11}	-0.212	0.093
	β_{12}	-1.438	0.138
	β_{21}	-6.467	0.656
	β_{22}	-10.108	0.764
Univariate	β_{11}	-0.493	0.107
	β_{12}	-1.438	0.138
	β_{21}	-6.832	0.683
	β_{22}	-9.815	0.721

5.3 Reconstruction of fields

With the estimates given in Section 5.2, we can reconstruct the fixed effects $\{\boldsymbol{v}_j; j = 1, 2, \ldots, 5\}$, the spatial effects $\{\boldsymbol{\xi}_j; j = 1, 2, \ldots, 5\}$ and the posterior mean of temperature and humidity $\boldsymbol{v}_j + \boldsymbol{\xi}_j$ for different years j. We show the fixed effect $\boldsymbol{v}_j$ for year 2008 in Figure 6(a) and Figure 6(b) for the bivariate model with 1km by 1km resolution when we set $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity. Figure 6(c) and Figure 6(d) give the corresponding results for the univariate model. The way for reconstructing the fixed effects with 1km by 1km resolution is that we first estimate the relevant parameters with the lower resolution model, and then plug in the estimates into the 1km by 1km resolution model. The fixed effects for other years are omitted since they are just shifted versions of each other because they just have different yearly effects but share the same coefficients from the two covariates. The corresponding results for the fixed effects with $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature are given in Figure 9(a) and Figure 9(b).

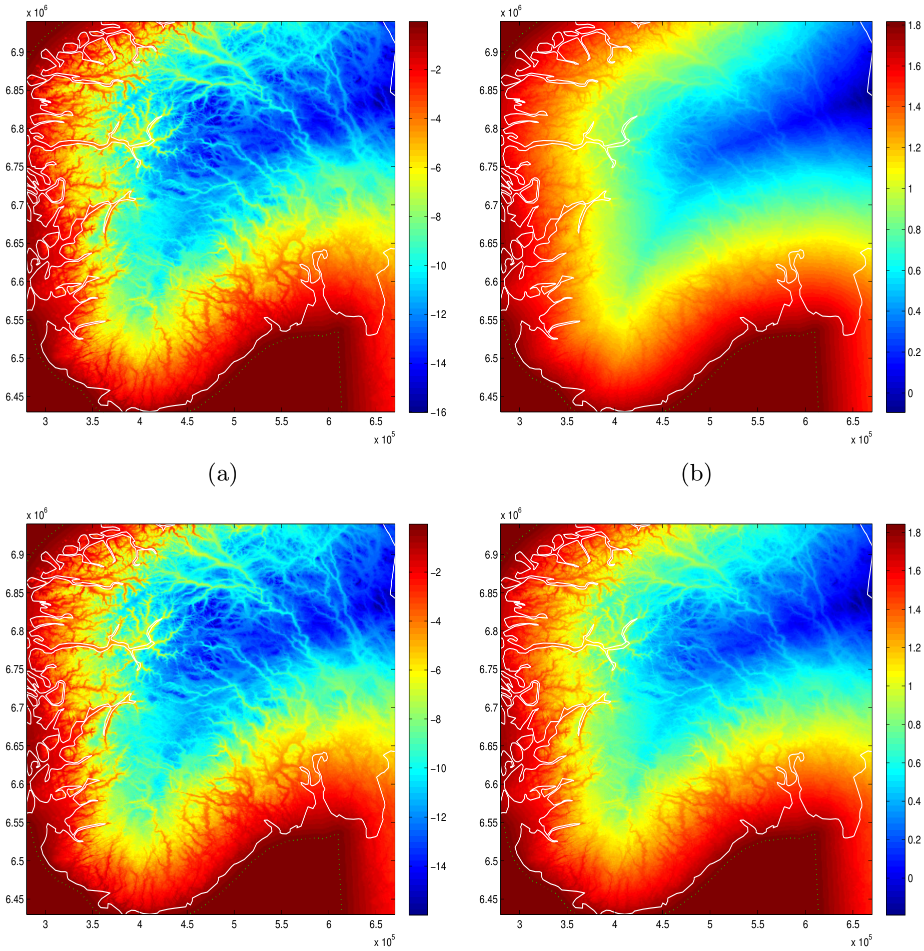
Using the same approach as the fixed effects, we can get the spatial effects $\{\boldsymbol{\xi}_j; j = 1, 2, \ldots, 5\}$ on high resolution with the estimates given in Table 2. Figure 7(a) and Figure 7(b) illustrate the spatial effects with 1km by 1km resolution for the bivariate model and Figure 7(c) and Figure 7(d) for the univariate model, respectively. The corresponding results for spatial effects with $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature are given in Figure 9(c) and Figure 9(d). As discussed in Section 3.4, the spatial effects are constructed by using the SPDEs or the systems of SPDEs with the same hyper-parameters. In other words, they are just different realizations of the same latent fields. We only show the results for year 2008 but emphasize that they are different from year to year.

With the fixed effects $\{\boldsymbol{v}_j; j = 1, 2, \ldots, 5\}$ and spatial effects $\{\boldsymbol{\xi}_j; j = 1, 2, \ldots, 5\}$ given in Figure 6 and Figure 7, we can get the posterior mean of temperature and humidity $\{\boldsymbol{\eta}_j = \boldsymbol{v}_j + \boldsymbol{\xi}_j; j = 1, 2 \ldots, 5\}$, and Figure 8(a) and Figure 8(b) illustrate the results for year 2008 for the bivariate model and Figure 8(c) and Figure 8(d) for the univariate model, respectively, with $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity. When we set $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature, we can get the posterior mean of temperature and humidity shown in Figure 10(a) and Figure 10(b) for bivariate model. We notice that the fixed effects, the spatial effects and the posterior mean for temperature and humidity in these two different bivariate models are not the same but quite similar.

5.4 Predictive performance

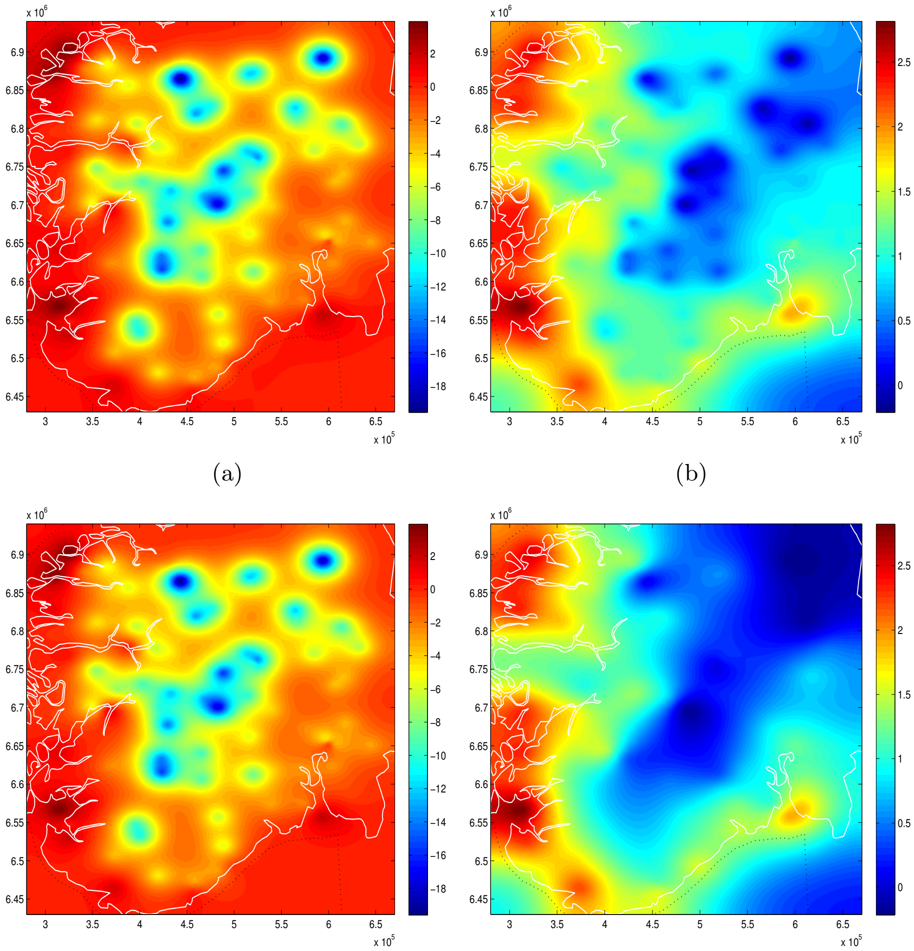
In this section the predictive performance for both the bivariate models and the univariate model with different settings, using the cross validation schemes given in Section 4.2, are discussed for both temperature and humidity. Using the left-out observations in different settings together with the scoring rules discussed in Section 4.1, we compare the values of scoring rules MAE, MSE and CRPS.

Table 10 illustrates the results for the scoring rules when we set $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity. In this table and the tables thereafter, "UM" means the results



(a) (b) (c) (d)

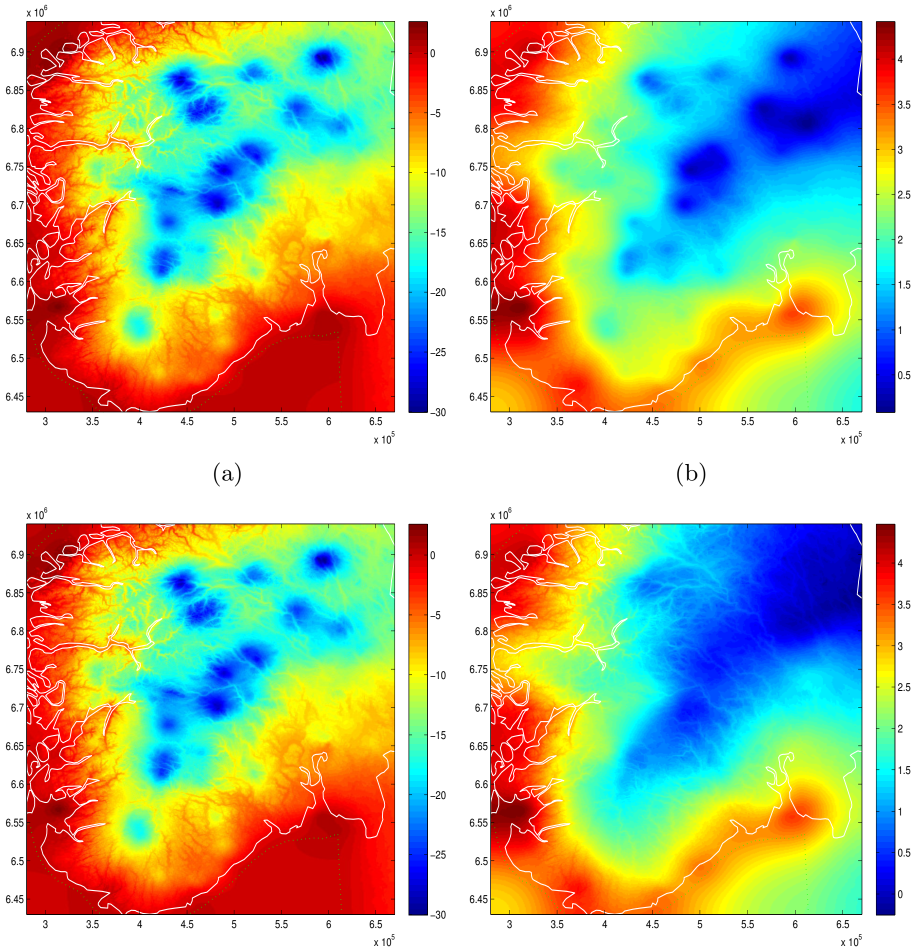
Figure 6: Fixed effects for bivariate model (a) - (b) and for the univariate model (c) - (d) of temperature and humidity in 2008 with 1km $\times$ 1km resolution. We set $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity.



(a) (b)

(c) (d)

Figure 7: Spatial effects for bivariate model (a) - (b) and for univariate model (c) - (d) of temperature and humidity in 2008 with 1km × 1km resolution. We set $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity.



(a) (b)

(c) (d)

Figure 8: Posterior mean of temperature and humidity in 2008 for bivariate model (a) - (b) ad for univariate model (c) - (d) with 1km × 1km resolution. We set $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity.

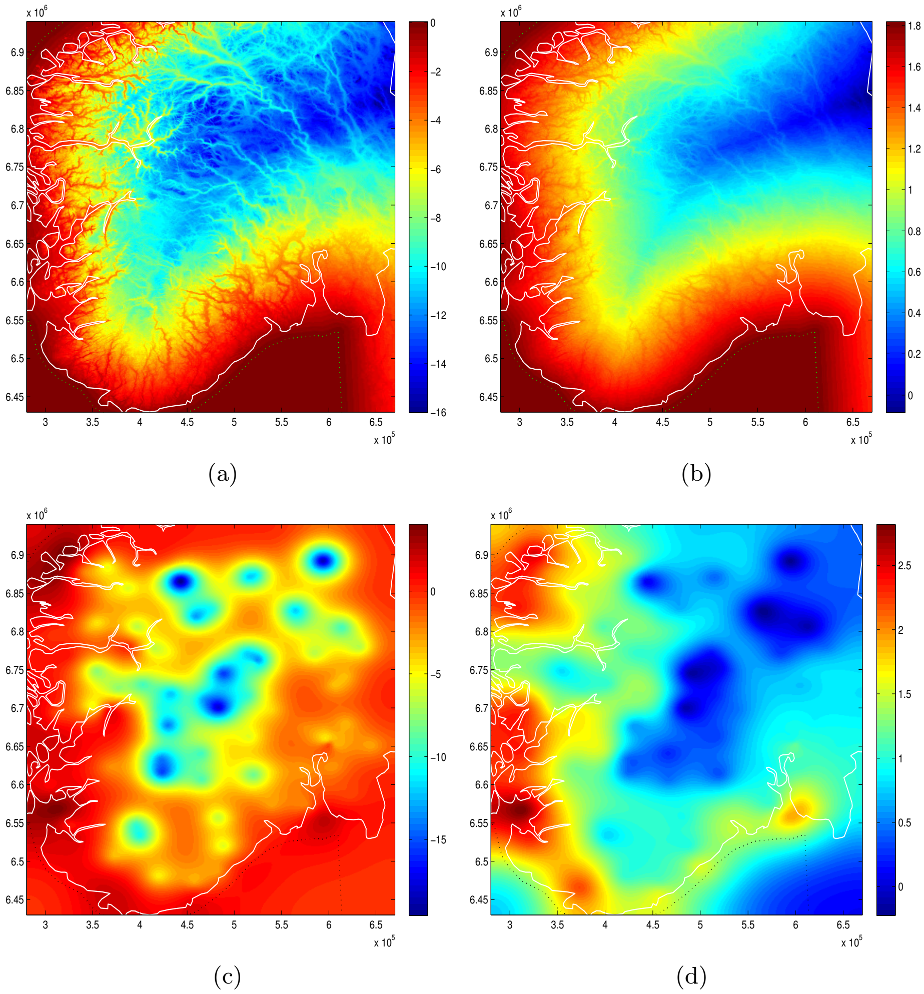
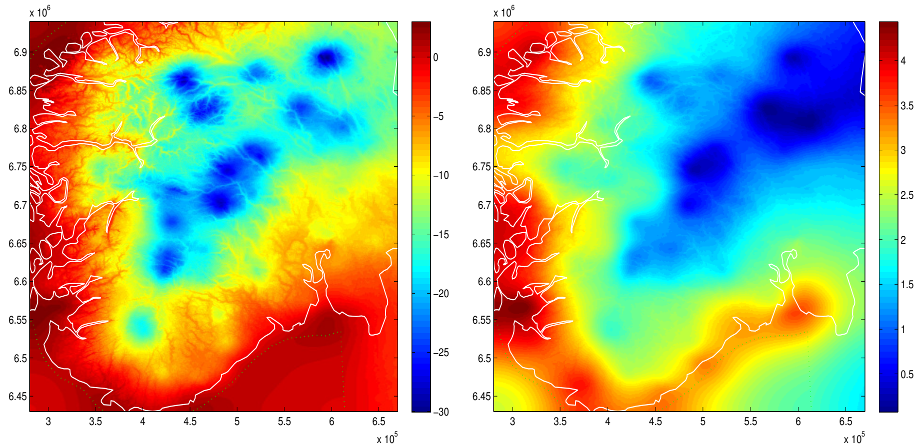


Figure 9: Fixed effects (a) - (b) and spatial effects (c) - (d) for bivariate model in 2008 with 1km $\times$ 1km resolution. We set $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature.



(a) (b)

Figure 10: Posterior mean of temperature (a) and humidity (b) in 2008 for bivariate model with 1km $\times$ 1km resolution. We set $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature.

are from the univariate model. "BM (Setting 3)" means the results are from the bivariate model with Setting 3, i.e., there are 20 locations left-out both for temperature and humidity in each year. " BM" means the results are from bivariate model with Setting 1 or Setting 2, i.e., there are 20 locations left-out from only temperature or humidity in each year. "T" and "H" denote the temperature and humidity, respectively. From Table 10 we can notice that the bivariate model performances uniformly better than the univariate model for humidity. We can also notice that "BM" outperforms the "BM(Setting 3)" uniformly, which means the observations from another field at the same locations improve the prediction accuracy. The scoring rules with "BM" is also uniformly better than the"UM" for temperature. However, the MSE and CRPS of temperature for bivariate model with Setting 3 is slightly higher than the corresponding results from univariate model.

Table 11 shows the predictive performance with $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature. We can notice that the bivariate model with "BM" performs uniformly better than the univariate model for temperature. But the bivariate model with Setting 3 perform a little worse than the univariate model for temperature. We can also notice that the bivariate model with Setting 3 performs uniformly better than the univariate model for the humidity. However, the bivariate model with "BM" performs the worst for the humidity.

Deeper analysis releases the reasons. We find that there is due to some "outliers" in the temperature observations in year 2009. Figure 11 has illustrated the left-out observations in year 2009. We can see that there are 5 locations with very high temperature but rather low humidity. The bivariate model has difficulties at these locations not only for predicting the temperature itself but also for humidity since the information from temperature leads the prediction of humidity in the wrong direction. The results of the

Table 10: Scoring rules for bivariate model and univariate model for temperature and humidity with $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity

Scoring rules for temperature (T) and humidity (H)						
	MAE		MSE		CRPS	
	T	H	T	H	T	H
UM	1.7485	0.2039	6.1102	0.0722	1.3524	0.1487
BM(Setting 3)	1.7463	0.1918	6.1164	0.0639	1.3539	0.1396
BM	1.5862	0.1501	4.3089	0.0352	1.1464	0.1060

scoring rules without year 2009 are given in Table 12 and 13. From these two tables, we can notice that the bivariate model performs uniformly better than the univariate model. Furthermore the bivariate model with "BM" performs uniformly better than the bivariate model with Setting 3. In addition, from Table 12 and Table 13, we can notice that we get better predictive performances when we setting the temperature as the second field when we need to predict temperature. This is also true with humidity.

The posterior standard deviations for temperature and humidity in 2011 with the bivariate model and the univariate model are presented in Figure 12(a) - Figure 12(b) and in Figure 12(c) - Figure 12(d), respectively, when we set $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity. With the bivariate model, we can notice that the posterior standard derivations for locations where we have the temperature observations but not the humidity observations are lower than the corresponding univariate models. Same conclusion can be drawn from Figure 13(a) and Figure 13(b), when we set $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature. The results for other years are similar and omitted here.

From all these results, we can notice that the bivariate models give better prediction accuracy than the univariate model. When the observations in one field is presented, it does not only improve the prediction accuracy but also have lower posterior standard deviations. We can also conclude that the order of the field has some influence for the prediction. The generally suggestion is that if we want to predict humidity or temperature, we should set it as the second field when we have enough time and computational resources. If time or computational resources is limited, then we do not need to consider about the order of fields, since the bivariate model can provide satisfiable results with both the orders.

6 Discussion and conclusion

We have modelled temperature and humidity in southern Norway based on the observations on 7th of December from 2007 to 2011. Three different models are used in this paper: two bivariate models for modelling them jointly and one univariate model for modelling them independently. The system of SPDEs approach proposed by Hu et al. (2012b) is chosen for constructing bivariate GRFs and the SPDE approach given by Lindgren et al. (2011) is chosen for constructing univariate GRFs. Three different set-

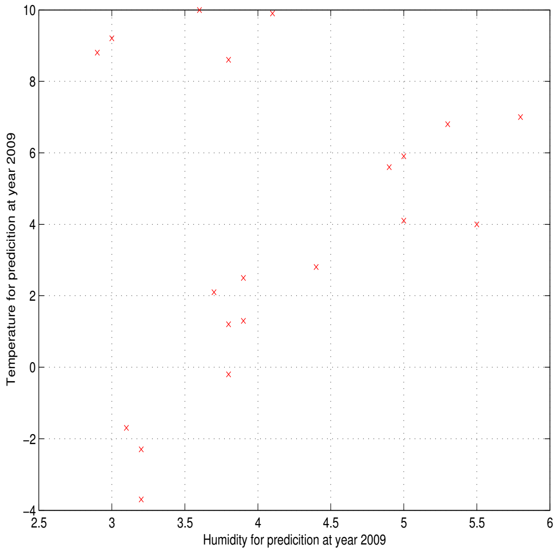


Figure 11: The observations of humidity and temperature for predictions at year 2009.

Table 11: Scoring rules for bivariate model and univariate model for temperature and humidity with $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature

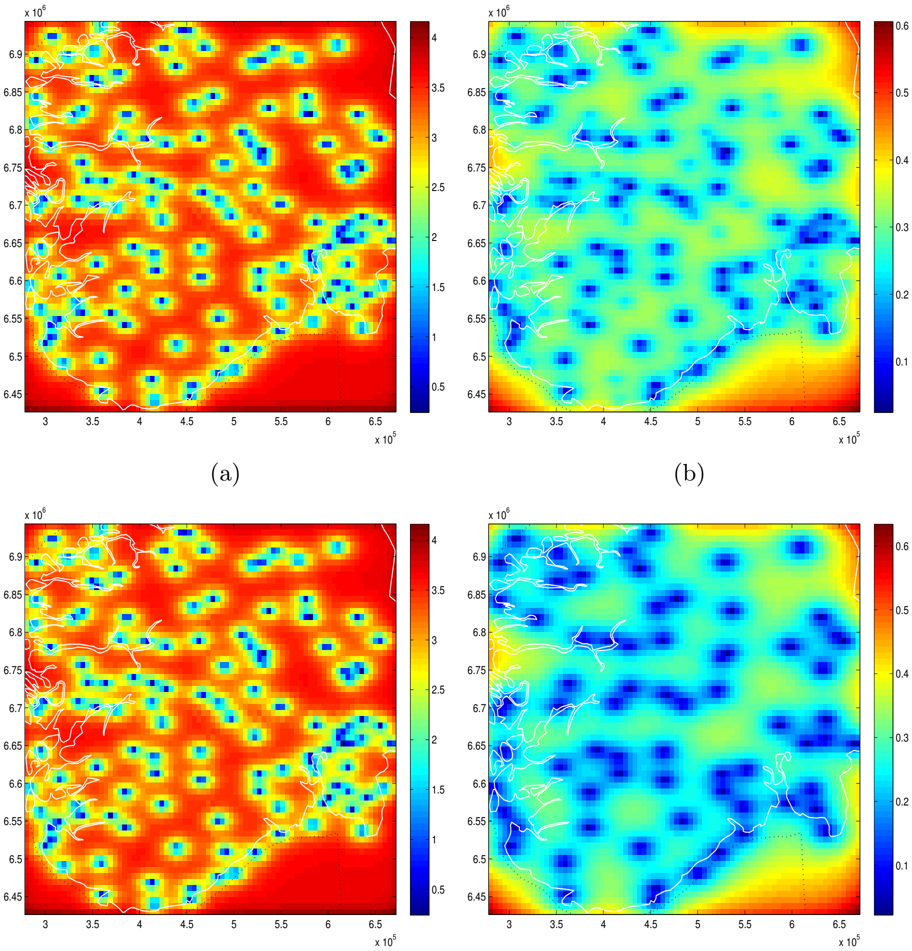
Scoring rules for temperature (T) and humidity (H)						
	MAE		MSE		CRPS	
	T	H	T	H	T	H
UM	1.7485	0.2039	6.1102	0.0722	1.3524	0.1487
BM(Setting 3)	1.7949	0.1929	6.2014	0.0638	1.3665	0.1402
BM	1.5564	0.2215	4.4149	0.0845	1.1579	0.1579

Table 12: Scoring rules for bivariate model and univariate model for temperature and humidity with $x_1(\boldsymbol{s})$ as temperature and $x_2(\boldsymbol{s})$ as humidity without year 2009

Scoring rules for temperature (T) and humidity (H)						
	MAE		MSE		CRPS	
	T	H	T	H	T	H
UM	1.7842	0.2095	6.7240	0.0791	1.4031	0.1553
BM(Setting 3)	1.7671	0.2055	6.6736	0.0737	1.3985	0.1481
BM	1.4693	0.1426	3.7366	0.0318	1.0677	0.1006

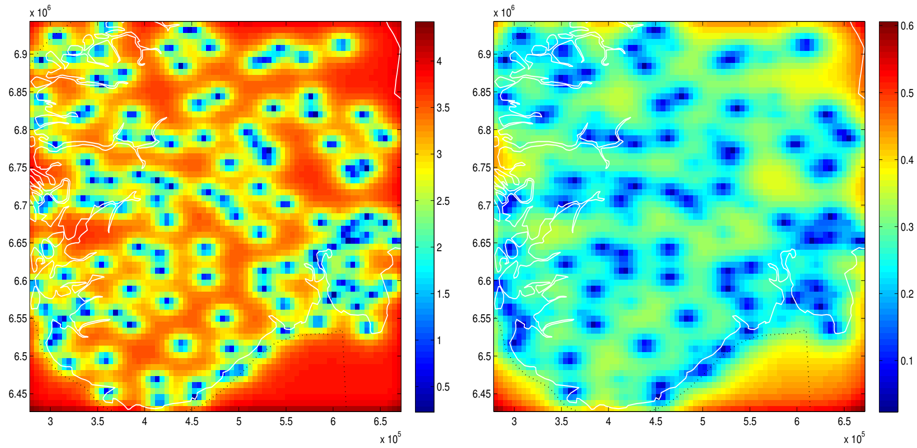
Table 13: Scoring rules for bivariate model and univariate model for temperature and humidity with $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature without year 2009

Scoring rules for temperature (T) and humidity (H)						
	MAE		MSE		CRPS	
	T	H	T	H	T	H
UM	1.7842	0.2095	6.7240	0.0791	1.4031	0.1553
BM(Setting 3)	1.7961	0.2010	6.5859	0.0713	1.3940	0.1467
BM	1.3837	0.1940	3.4967	0.0596	1.0513	0.1365



(a) (b) (c) (d)

Figure 12: Posterior standard deviation for temperature and humidity in 2011 by bivariate model (a) - (b) and by univariate model (c) - (d). We set the first field $x_1(\boldsymbol{s})$ as temperature and the second field $x_2(\boldsymbol{s})$ as humidity.



(a) (b)

Figure 13: The posterior standard deviation for temperature (a) and humidity (b) in 2011 by bivariate model. We set $x_1(\boldsymbol{s})$ as humidity and $x_2(\boldsymbol{s})$ as temperature.

tings are chosen in order to compare the predictive performance between the bivariate model and the univariate model with different settings. Computational efficiency is obtained by using the link between the GRFs and GMRFs. All our models are constructed with GRFs theoretically and all computations are conducted with GMRFs. The results illustrate that there is a strong positive correlation between temperature and humidity. The other results agree with the physical and empirical knowledge. We conclude that using a bivariate GRF to model temperature and humidity jointly is superior to model them independently using univariate GRFs, not only in term of prediction accuracy, but also in term of quantifying prediction uncertainty.

The results also illustrate that the order of fields seems relevant from the prediction point of view when we use a triangular system of SPDEs for constructing a bivariate field. However, since the results from both orders are satisfiable, we do not need to consider it if the computational resources or time is limited. There might be some other covariates, such as wind speed and solar radiation which needs to be included in our analysis. However, this is beyond the scope of this paper and we leave it for future research.

References

T.V. Apanasovich, M.G. Genton, and Y. Sun. A valid matérn class of cross-covariance functions for multivariate random fields with any number of components. *Journal of the American Statistical Association*, 107(497):180–193, 2012.

S. Banerjee, B.P. Carlin, and A.E. Gelfand. *Hierarchical modeling and analysis for spatial data.* Chapman & Hall, 2004. ISBN 158488410X.

K.J. Bathe. *Finite element method.* Wiley Online Library, 2008.

D. Bolin and F. Lindgren. Wavelet markov models as efficient alternatives to tapering and convolution fields. Technical report, Mathematical Statistics, Centre for Mathematical Sciences, Faculty of Engineering, Lund University, 2009.

D. Bolin and F. Lindgren. Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *The Annals of Applied Statistics*, 5 (1):523–550, 2011.

G.E.P. Box and D.R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 211–252, 1964.

P.J. Brown, N.D. Le, and J.V. Zidek. Multivariate spatial interpolation and exposure to air pollutants. *Canadian Journal of Statistics*, 22(4):489–509, 1994.

P. Courtier, E. Andersson, W. Heckley, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, M. Fisher, and J. Pailleux. The ecmwf implementation of three-dimensional variational assimilation (3d-var). i: Formulation. *Quarterly Journal of the Royal Meteorological Society*, 124 (550):1783–1807, 1998.

N. Cressie and C.K. Wikle. *Statistics for spatio-temporal data*, volume 465. Wiley, 2011.

N.A.C. Cressie. *Statistics for spatial data*, volume 298. Wiley-Interscience, 1993.

P.J. Diggle and P.J. Ribeiro Jr. *Model-based Geostatistics.* Springer, 2006.

G.A. Fuglstad. Spatial modelling and inference with spde-based gmrfs. Master's thesis, Department of Mathematical Sciences, Norwegian University of Science and Technology, 2011.

A.E. Gelfand, A.M. Schmidt, S. Banerjee, and CF Sirmans. Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, 13(2):263–312, 2004.

A.E. Gelfand, P.J. Diggle, M. Fuentes, and P. Guttorp. *Handbook of spatial statistics.* CRC Press, 2010.

T. Gneiting, A.E. Raftery, A.H. Westveld III, and T. Goldman. Calibrated probabilistic forecasting using ensemble model output statistics and minimum crps estimation. *Monthly Weather Review*, 133(5):1098–1118, 2005.

T. Gneiting, W. Kleiber, and M. Schlather. Matérn Cross-Covariance Functions for Multivariate Random Fields. *Journal of the American Statistical Association*, 105(491):1167–1177, 2010. ISSN 0162-1459.

H. Hersbach. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5):559–570, 2000.

X. Hu, F. Lindgren, D.P. Simpson, and H. Rue. Multivariate gaussian random fields with oscillating covariance functions using systems of stochastic partial differential equations. *statistical report, Norwegian University of Science and Technology*, 2012a.

X. Hu, D.P. Simpson, F. Lindgren, and H. Rue. Multivariate gaussian random fields using systems of stochastic partial differential equations. *statistical report, Norwegian University of Science and Technology*, 2012b.

R. Ingebrigtsen, F. Lindgren, and I. Steinsland. Using stochastic partial differential equation models for spatial reconstruction of annual precipitation. *submitted*, 2013.

L.W. Konigsberg and S.D. Ousley. Multivariate quantitative genetics of anthropometric traits from the boas data. *Human biology*, 81(5/6):579–594, 2009.

F. Lindgren, H. Rue, and J. Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.

K. Mcguigan. Studying phenotypic evolution using multivariate quantitative genetics. *Molecular ecology*, 15(4):883–896, 2006.

B.J. Reich and M. Fuentes. A multivariate semiparametric bayesian spatial modeling framework for hurricane surface wind fields. *The Annals of Applied Statistics*, 1(1):249–264, 2007.

H. Rue and L. Held. *Gaussian Markov random fields: theory and applications*. Chapman & Hall, 2005. ISBN 1584884320.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392, 2009. ISSN 1467-9868.

S.R. Sain and N. Cressie. A spatial model for multivariate lattice data. *Journal of Econometrics*, 140(1):226–259, 2007.

RM Sakia. The box-cox transformation technique: a review. *The statistician*, pages 169–178, 1992.

A.M. Schmidt and A.E. Gelfand. A bayesian coregionalization approach for multivariate pollutant data. *Journal of Geophysical Research*, 108(D24):8783, 2003.

D. Simpson, F. Lindgren, and H. Rue. In order to make spatial statistics computationally feasible, we need to forget about the covariance function. *Environmetrics*, 2010.

M.L. Stein. *Interpolation of Spatial Data: some theory for kriging*. Springer Verlag, 1999. ISBN 0387986294.

Z. Toth, O. Talagrand, G. Candille, and Y. Zhu. 7 probability and ensemble forecasts. 2003.

O.C. Zienkiewicz, R.L. Taylor, R.L. Taylor, and JZ Zhu. *The finite element method: its basis and fundamentals*, volume 1. Butterworth-heinemann, 2005.