Quantifying uncertainty for temperature maps derived from computer models

Lucia Paci[a,*], Alan E. Gelfand[b], Daniela Cocchi[a]

[a]*Department of Statistical Science at the University of Bologna, Via delle Belle Arti, 41 - 40126 Bologna, Italy*
[b]*Department of Statistical Science at Duke University, Box 90251, Durham NC 27708-0251, USA*

Abstract

Computer models are often deterministic simulators used to predict several environmental phenomena. Such computer models do not provide any measure of uncertainty associated to their output since they are derived from deterministic specifications. However, many sources of uncertainty exist in constructing and employing numerical models, so finding a way to quantify such uncertainty is important.

We are motivated by temperature maps arising from the Rapid Update Cycle (RUC) model, a regional short-term weather forecast model for the continental United States (US). These maps are provided by the National Climatic Data Center with no associated uncertainty in their forecasts.

Despite a rapidly growing literature on uncertainty quantification, there is little regarding statistical methods for attaching uncertainty to model output when we do not have information about how deterministic predictions are created. Although numerical models produce deterministic surfaces, the output will not be the 'true' value of the process and, given the true value and the model output, the associated error is not stochastic. However, under suitable stochastic modeling, this error can be interpreted as a random unknown about which we can infer using a Bayesian specification within a data fusion setting, fusing the computer model data with some external validation data collected independently over the same spatial domain as the deterministic

---

[*]Corresponding author. Tel.: +39 051 2098201; Fax: +39 051 232153.

Email addresses: `lucia.paci2@unibo.it` (Lucia Paci), `alan@stat.duke.edu` (Alan E. Gelfand), `daniela.cocchi@unibo.it` (Daniela Cocchi)

map. Our specifications reflect the intuitive fact that uncertainty associated with locations does not necessarily grow as we become more distant from the validation locations. We apply our modeling approach to obtain an uncertainty map associated with RUC forecasts over the northeastern US. **Abstract must be 200 words=Remove 53 words**

1. Introduction

Many computer models are deterministic simulation models developed, for example, to predict environmental phenomena such as temperatures or air pollution levels. In a spatial setting, numerical model outputs are displayed in the form of maps, provided as averages over grid cells, usually at high spatial and temporal resolution. Such computer models do not provide any measure of uncertainty associated to their output since they are derived from deterministic specifications. However, many sources of uncertainty exist in constructing and employing numerical models. In fact, with computer models providing spatially referenced output, uncertainty maps can be a useful tool to guide environmental agencies in refining and improving computer models. Furthermore, when we use model output as predictor for an environmental variable, we might seek to evaluate how these uncertainties propagate from the model output to the forecasting of the response. Altogether, it seems there is need for quantifying uncertainty in this setting.

The motivating context for us are temperature maps that arise from the Rapid Update Cycle (RUC) numerical weather model, a regional short term weather forecast model for the continental United States. This model yields maps that are publicly available, provided by the National Oceanic and Atmospheric Administration (NOAA)'s National Climatic Data Center (NCDC) but with no explicit detail regarding their development and no associated uncertainty in their forecasts. Our contribution, articulated in detail below, is to propose a hierarchical stochastic model along with the introduction of a validation data set consisting of temperature measurements collected at monitoring stations operating in the same study region. The model fuses the two data sources to enable assessment of uncertainty associated with RUC maps.

In applications, the sources of potential uncertainty associated with nu-

merical models include input uncertainty, function uncertainty, model discrepancy and observational error (Cumming and Goldstein, 2010). The Bayesian approach represents a natural way to account for all of these uncertainty sources and several methods have been developed to deal with the uncertainty analysis for complex computer models. Customarily, numerical models are implemented as computer codes, dependent upon a number of inputs which determine the nature of the output. These inputs represent unknown parameters and the uncertainty about them propagates through the numerical model, inducing uncertainty in the output.

A general statistical framework has been presented by Raftery et al. (1995) for mapping from a set of input parameters to a set of model outputs, the so-called Bayesian synthesis which was later revised leading to the Bayesian melding approach (Poole and Raftery, 2000) Also, statistical methods have been proposed to handle the sensitivity analysis which is concerned with understanding how the model output is influenced by changes in the model inputs (e.g. Oakley and O'Hagan, 2004). For deterministic numerical models, i.e., models with no random components, their predictions are subject to error because any model is a simplification of reality. So, model output will not equal the 'true' value of the process of interest and this discrepancy is well-known as model inadequacy (Kennedy and O'Hagan, 2001).

Structural uncertainty, which is introduced by scientific choices of model design and development, can be also quantified by analyzing multi-model ensembles. In this case, the output consists of different versions of a numerical model, i.e. a model is run several times with different initial conditions (scenarios). Statistical approaches for quantifying uncertainty with ensembles have recently received considerable attention (see e.g. Gneiting et al. 2005; Raftery et al. 2005; Berrocal et al. 2007; Smith et al. 2009; Di Narzo and Cocchi 2010; Kleiber et al. 2011; Sloughter et al. 2013).

There is little in the literature about statistical methods for attaching uncertainty to model output when we do not have information about how such deterministic predictions are created, about model inputs. Indeed, our contribution to uncertainty quantification builds upon the notion of uncertainty introduced by Ghosh et al. (2012) when numerical models are unavailable; rather, only deterministic outputs at some spatial resolution are provided. In other words, we do not know how the deterministic surfaces have been developed; for us, they come from a entirely unknown 'black box'. Ghosh et al. (2012) proposed a general Bayesian approach to associate uncertainties with deterministic interpolated surfaces which requires some external

3

validation data collected independently over the same spatial domain as the deterministic map.

Although numerical models produce deterministic surfaces, we have already highlighted that the output will not be the 'true' value of the process. In this framework, given the truth and the model output, the associated error is not stochastic. But, under suitable stochastic modeling, this error can be interpreted as a random unknown which we can infer about using a Bayesian specification within a data fusion setting, fusing the computer model data with the validation data.

Reiterating, the contribution here is to develop a Bayesian hierarchical model to provide spatially smoothed uncertainty associated with numerical model output, regardless of how it was created. We show how we can learn about such uncertainty through stochastic data fusion modeling using some external validating data. We also take into account the change of support problem (COSP; Gelfand et al. 2001), which arises from the spatial misalignment between the numerical model output and the validation data. It is important to note that our objective is not the calibration of numerical model output. Rather, we are interested in spatially smoothed uncertainties associated with the prediction maps. To attach such varying uncertainty across grid cells we offer a fully model-based approach that can be used to assign uncertainty to any deterministic surface. Two useful advances over the work in Ghosh et al. (2012) are introduced. First, our specifications here reflect the intuitive fact that uncertainty associated with grid cells progressively away from the validation locations does not necessarily increase with distance. Second, our specifications recognize the spatial structure in the validation data and deal with the spatial misalignment arising from the grid cell level model output and the point level station data.

The format of the paper is as follows. In Section 2 we describe the motivating deterministic data surfaces. In Section 3 we provide explicit model details while in Section 4 we formalize the uncertainty definition and its calculation. Section 5 provides the fitting details. Section 6 offers some simulation results, while Section 7 provides the uncertainty surface for the RUC data. We conclude with a brief summary and future research possibilities. Supplemental material including additional figures is available.

4

2. Data description

The RUC model (Benjamin et al., 2004) is a regional short-term weather forecast model for the Continental United States (CONUS) developed by the National Centers for Environmental Prediction (NCEP) to serve users needing frequently updated short-term weather forecasts. When it was first implemented in 1994, the model was run every three hours making forecasts out to 12 hours. By 2002, the RUC was run every hour, on the hour, producing 12-hour forecasts at 13 km spatial resolution. The output from the RUC model is available, for free, at the website: http://ruc.noaa.gov/.

As an illustration, we consider the daily forecast map on August 7th, 2011 obtained as average of 24 hourly temperature forecasts (°F) provided by RUC model from 00:00 to 23:00 on August 7th over the northeastern U.S., see Figure 1. There are 3,862 RUC grid cells spanning our study region. Moreover, land-based station data over U.S. is provided by the NOAA's National Climatic Data Center (NCDC). Here, we consider 24-hour averages of hourly temperature collected from 163 monitoring stations operating in the study region for the same period.
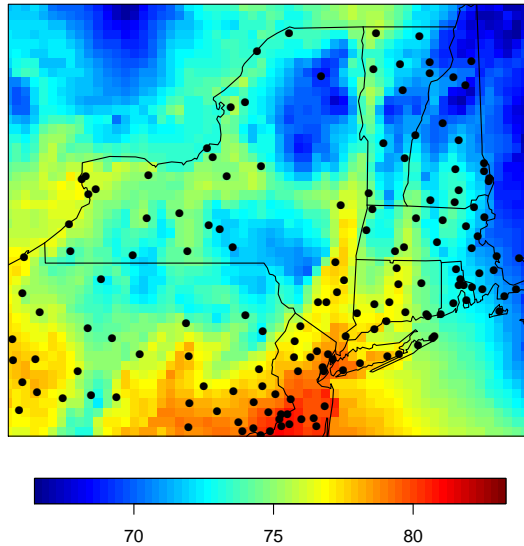

Figure 1: Temperature stations (black dots) and daily RUC output on August 7th, 2011 over Northeastern U.S..

3. Data fusion model

Let $R(A_i)$ denote the numerical model output (i.e. temperature predictions from the RUC model) over grid cell A_i, $(i = 1, \ldots, I)$. We interpret $R(A_i)$ as an average value over cell A_i, i.e. $R(A_i) = |A_i|^{-1} \int_{A_i} R(s)ds$.

First, we specify a measurement error model (MEM) for the numerical model output $R(A_i)$ relative to the truth. The measurement error model is also known as error-in-variables model; see for instance Fuller (1987) and references therein. In particular,

$$R(A_i) = \widetilde{R}(A_i) + \varepsilon_r(A_i) \tag{1}$$

where $\widetilde{R}(A_i)$ is the underlying process which represents the "true" average value for A_i and we assume $\varepsilon_r(A_i) \sim N\big(0, \sigma_r^2(A_i)\big)$ independently $\forall i, i = 1, \ldots, I$. The $\sigma_r^2(A_i)$ reflect the uncertainty we seek to attach to $R(A_i)$; specification of the model for these areal unit specific variances is a critical component and is elaborated below. The true average value $\widetilde{R}(A_i)$ arises from a Gaussian Markov Random Field (GMRF) equipped with a conditionally autoregressive structure (CAR) (Besag, 1974; Banerjee et al., 2004) that is:

$$\widetilde{R}(A_i) \,|\, \big\{ \widetilde{R}(A_{i'}) : i' \neq i \big\} \sim N\left(\sum_{i' \sim i} \frac{\widetilde{R}(A_{i'})}{w_i}, \frac{\tau^2}{w_i} \right) \tag{2}$$

where $i' \sim i$ denotes the cells $A_{i'}$ that are adjacent neighbors to cell A_i and w_i is the number of neighbors of cell A_i.

Let $V(\mathbf{s}_j)$ be the validation data (i.e. observed temperature) at location $\mathbf{s}_j$, $(j = 1, \ldots, n)$ gathered independently over the same region as the output, and $\widetilde{V}(\mathbf{s}_j)$ denotes the true value at $\mathbf{s}_j$. We allow for error in the validation data, assuming a spatial model given by:

$$V(\mathbf{s}_j) = \widetilde{V}(\mathbf{s}_j) + \varepsilon_v(\mathbf{s}_j). \tag{3}$$

Here $\boldsymbol{\varepsilon}_v' = (\varepsilon_v(\mathbf{s}_1), \ldots, \varepsilon_v(\mathbf{s}_n))$ is a zero-mean Gaussian process equipped with a spatial exponential correlation function, i.e. $\boldsymbol{\varepsilon}_v \sim N\big(\mathbf{0}, \sigma_v^2 \, H(\phi_v)\big)$ where $\big(H(\phi_v)\big)_{ij} = \exp\{-\phi_v \|\mathbf{s}_i - \mathbf{s}_j\|\}$. It seems appropriate to anticipate spatial dependence in the measurement errors.

Finally, we address the change of support problem between the observed data and the numerical model output by assuming a further measurement
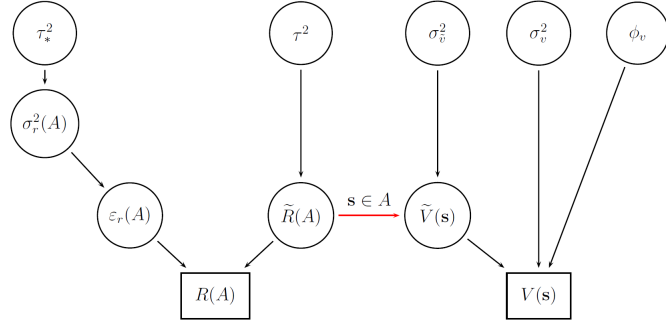
Figure 2: Graphical representation of model (1) - (4) under prior (7).

error model for $\widetilde{V}(\mathbf{s}_j)$. We avoid the integration problem associated with scaling from point to grid level by employing the downscaling approach (Berrocal et al., 2012; Paci et al., 2013) which associates to each site $\mathbf{s}_j$ the grid cell A_i that contains $\mathbf{s}_j$. Then, for each $j = 1, \ldots, n$ belonging to grid cell A_i we have:

$$\widetilde{V}(\mathbf{s}_j) = \widetilde{R}(A_i) + \varepsilon_{\tilde{v}}(\mathbf{s}_j) \tag{4}$$

where $\varepsilon_{\tilde{v}}(\mathbf{s}_j)$ are independent $N(0, \sigma_{\tilde{v}}^2)$. Figure 2 shows a graphical representation of the model described above. Now, we are ready to clarify what we mean by the uncertainty associated with the deterministic outputs, the $R(A_i)$.

4. Defining and modeling uncertainty associated with deterministic outputs

Recall that our primary goal is to provide a measure of uncertainty associated with numerical model output over grid cells. To clarify what we mean by uncertainty associated with a deterministic output, we say that the true error of $R(A_i)$ can be quantified by an appropriate residual defined below. When the expected squared value of this error is small for a grid cell, it implies small uncertainty associated to the numerical model prediction, when large we would imagine high uncertainty for such a cell.

To inform about the true error, we might rely on the *observed* residuals $R(A_i) - V(\mathbf{s}_j)$, i.e. comparing the numerical model output with the validation data for each grid cell that contains a site. Then, high 'disagreement' between $R(A_i)$ and $V(\mathbf{s}_j)$ for $s_j \in A_i$ suggests high uncertainty in A_i. Conversely, we

7

envision small uncertainty at grid cells where the disagreement between the numerical model output and the observed temperature is low.

However, two issues arise if we confine ourselves to the observed residuals: first, the comparison between the average $R(A_i)$ with the point-level measurement $V(\mathbf{s}_j)$ is inappropriate because of the different spatial support of the two data sources, i.e the change of support problem. Second, the observed residuals are available only for grid cells where monitoring sites lie, while our goal is to attach uncertainty to every grid cell.

To accomplish this, we consider instead the so-called *realized* residuals, that is $\varepsilon_r(A_i) = R(A_i) - \widetilde{R}(A_i)$ from equation (1). The term 'realized' is borrowed from Zellner (1975) where the problem of making inferences about the realized values of error terms in the linear normal regression model was analyzed; also see Chaloner and Brant (1988) for related work. Again, the true error for $R(A_i)$ is not known and, as usual within the Bayesian framework, we model unknowns as random and look at their posterior distributions for inference. In fact, under the specification above, we take the model for $\widetilde{R}(A_i)$ as the model for the truth and we look at the posterior distribution of the realized residuals, $\big[\varepsilon_r(A_i)|Data\big]$. We adopt the posterior variance, $var\left(\varepsilon_r(A_i)\,|\,Data\right)$, to provide the desired uncertainty, noting that it varies across grid cells.

We can obtain our local uncertainties by composition sampling, i.e. drawing posterior samples of $\varepsilon_r(A_i)$ and then computing their variance. Alternatively, we can obtain local uncertainties as the posterior means, $E\big(\sigma_r^2(A_i)\,|\,Data\big)$. Indeed, under model (1) - (4), we have

$$
\begin{aligned}
var\left(\varepsilon_r(A_i)\,|\,Data\right) =& E\left[var\left(\varepsilon_r(A_i)\,|\,\widetilde{R}(A_i),\sigma_r^2(A_i),Data\right)\right] \\
&+ var\left[E\left(\varepsilon_r(A_i)\,|\,\widetilde{R}(A_i),\sigma_r^2(A_i),Data\right)\right]
\end{aligned}
\tag{5}
$$

The second term in (5) is clearly 0 and the first reduces to $E\big(\sigma_r^2(A_i)\,|\,Data\big)$. So, Rao-Blackwellized estimates (Gelfand and Smith, 1990) can be directly obtained by computing the mean of the posterior sampled draws of $\sigma_r^2(A_i)$.

Since we are interested in the posterior distribution of $\sigma_r^2(A_i)$, the specification of its prior distribution becomes a crucial step. We consider two proposals for this prior.

First, we treat the collection of unknown variances as a collection of independent random effects. As usual, we adopt a specification which allows us to borrow strength across the $\sigma_r^2(A_i)$. If we work on the scale of the variances

we might adopt conditionally i.i.d. $\sigma_r^2(A_i)$, i.e., $\sigma_r^2(A_i)|a,b \sim IG(a,b)$. As a vague choice, we set $a = 2$ and let b be unknown. This provides random effects with mean b and infinite variance. An alternative is to take $\sigma_r^2(A_i)$ to R^1 by working with the $\log(\sigma_r^2(A_i))$. Now, we might assume i.i.d. $\log(\sigma_r^2(A_i))$ with

$$\log(\sigma_r^2(A_i))|\mu, \sigma^2 \sim N(\mu, \sigma^2). \tag{6}$$

Here, μ would correspond to $\log b$ and we would take σ^2 large, e.g. order 10^3.

As a second proposal, suppose we seek to attach high uncertainty to grid cells for which we expect large differences between the model output and the true value. In addition, given a large realized residual at grid cell A_i, we expect similar behavior in its neighborhood, equally for small realized residuals, i.e. we envision that changes in variance occur smoothly over space. This suggests the idea of offering a spatially structured specification for the $\log(\sigma_r^2(A_i))$. With areal unit level variances, this suggests a CAR model for the logarithm of $\sigma_r^2(A_i)$. That is,

$$\log(\sigma_r^2(A_i))| \left\{ \log(\sigma_r^2(A_{i'})) : i' \neq i \right\} \sim N \left(\sum_{i' \sim i} \frac{\log(\sigma_r^2(A_{i'}))}{w_i}, \frac{\tau_*^2}{w_i} \right) \tag{7}$$

where, following the notation in Section 3, $i' \sim i$ identifies the cell $A_{i'}$ adjacent to cell A_i and w_i is the number of neighbors of cell A_i. The logCAR prior model in (7) is analogous to the spatial stochastic volatility approach developed by Yan (2007) and extended by Reich and Hodges (2008) to capture spatial clustering in heteroscedasticity. The model in (7) enables us to explicitly impose spatially varying structure on the variances, allowing for both borrowing strength across grid cells and inducing local spatial smoothing of uncertainty estimates relative to their neighboring grid cells.

4.1. Comparing uncertainty assignments

The comparison of alternative models is traditionally performed with attention to uncertainty reduction. However, this is not an appropriate objective here. There is no intention to obtain arbitrarily small uncertainties. To clarify, for each A_i, there is a "true" $R_{true}(A_i)$. In the foregoing, $\widetilde{R}(A_i)$ is a model for $R_{true}(A_i)$ which will supply desired local uncertainties through the posterior variances of the realized residuals. Any model with local variances will provide such uncertainties; how can we say that one set of uncertainties is better than another?

An appropriate criterion needs to account for the trade-off between uncertainty and bias in $\widetilde{R}(A_i)$ that is

$$
\begin{aligned}
R(A_i) - R_{true}(A_i) = \Big(R(A_i) - \widetilde{R}(A_i) \Big) \\
+ \Big(\widetilde{R}(A_i) - R_{true}(A_i) \Big).
\end{aligned}
\tag{8}
$$

Therefore, model comparison should consider both the posterior variance arising from the first term in (8) and the squared bias associated with the second term. As pointed out by Ghosh et al. (2012), to inform about bias with available data, we can only compare $\widetilde{R}(A_i)$ with validation data $V(\mathbf{s}_j)$, for each $\mathbf{s}_j \in A_i$. Then, the balanced loss approach, as proposed in Gelfand and Ghosh (1998) yields the criterion

$$
\frac{1}{I} \sum_{i=1}^{I} var\big[\varepsilon_r(A_i) \,|\, data\big] + \frac{c}{n} \sum_{j=1}^{n} E\big[\big(\widetilde{R}(A_i) - V(\mathbf{s}_j)\big)^2 \,|\, data\big]
\tag{9}
$$

where c indicates the relative regret for the two losses. We choose the model yielding the smallest value of (9).

5. Fitting details

We recall that it is not possible to consistently estimate the decay and variance parameter in a spatial model with a covariance function belonging to the Matérn family (Zhang, 2004). In particular, with the exponential covariance function we adopt, we can only identify the product $\sigma_v^2 \phi_v$ and this product is what determines predictive behavior. So, we fix the decay parameter and we put a prior distribution on $\sigma_{\tilde{v}}^2$. We choose ϕ_v by connecting it to the scale of the spatial surface which we are working with; for instance, with our real data, we fix the decay parameter at roughly 60% of the maximum distance over the study region as suggested by explanatory analysis. Specifically, for the variance parameters σ_v^2 and $\sigma_{\tilde{v}}^2$, we place weak but partially data-driven conjugate inverse gamma priors $IG(a_\sigma, b_\sigma)$ where we choose a_σ and b_σ so that

$$
\frac{b_\sigma}{a_\sigma - 1} = \frac{MSE}{2} \quad \text{and} \quad \frac{b_\sigma^2}{(a_\sigma - 1)^2 (a_\sigma - 2)} = 10^2.
$$

Here, MSE is the mean square error arising from a simple linear regression of $V(s_j)$ on $R(A_i)$ for each $s_j \in A_i$. The prior distributions for τ^2 and τ_*^2 are specified as independent proper inverse gamma distributions $IG(a_\tau, b_\tau)$.

Recently, Sørbye and Rue (2013) proposed a general approach for choosing the prior distribution for the precision parameter of an intrinsic GMRF, according to the specific type of GMRF used. The authors suggested to select a scaled prior by mapping the precision parameter to the marginal standard deviation of the model, under linear constraints. In their applications, they showed that there were no significant differences in the estimated spatial effects using the default and the scaled priors for the precision parameter of a CAR process and their results were not sensitive to prior tuning. Due to the insensitivity to different choices of a_τ and b_τ, in our implementation we take $a_\tau = 2$ and $b_\tau = 1$, implying that these variance components have prior mean 1 and infinite variance.

Finally, a prior distribution for the parameter μ in (6) is needed. We assume that this parameter is sampled from a normal distribution with mean 0 and variance $g^2 = 10^3$.

Define $\mathbf{R} = \big(R(A_1), \ldots, R(A_I)\big)'$ and $\mathbf{V} = \big(V(\mathbf{s}_1), \ldots, V(\mathbf{s}_n)\big)'$; then the full distributional specification of model (1) - (4) using the logCAR prior model (7) is given by:

$$\left[\mathbf{R} \mid \widetilde{\mathbf{R}}, \boldsymbol{\sigma}_r^2\right] \left[\mathbf{V} \mid \widetilde{\mathbf{V}}, \sigma_v^2\right] \left[\widetilde{\mathbf{V}} \mid \widetilde{\mathbf{R}}, \sigma_{\tilde{v}}^2, \phi_v\right] \left[\widetilde{\mathbf{R}} \mid \tau^2\right] \left[\boldsymbol{\sigma}_r^2 \mid \tau_*^2\right] \tag{10}$$

where $\widetilde{\mathbf{R}} = \big(\widetilde{R}(A_1), \ldots, \widetilde{R}(A_I)\big)'$, $\widetilde{\mathbf{V}} = \big(\widetilde{V}(\mathbf{s}_1), \ldots, \widetilde{V}(\mathbf{s}_n)\big)'$ and $\boldsymbol{\sigma}_r^2 = \big(\sigma_r^2(A_1), \ldots, \sigma_r^2(A_I)\big)'$. Along with the prior distributions for all the unknown parameters, the Bayesian hierarchical model is completely specified. The model is fitted using Markov Chain Monte Carlo (MCMC) algorithm; details are deferred to the Appendix.

6. Simulation study

In this section, we consider simulation examples to illustrate the performance of the nonspatial and spatial variance specifications in Section 4. Since our attached uncertainty is model-based, it requires care to evaluate the modeling performance, as we discussed in Subsection 4.1. Under the simulation study we know the truth and so we know the true errors for assessing map uncertainty.

The simulation design is built from several sampling/fitting combinations allowing the investigation of different features. Simulation experiments are performed through the following steps:

1. We consider a unit square divided equally into 900 grid cells.

2. Using the centroids of the grid cells, we generate $\widetilde{R}(A_i)$ ($i = 1, \ldots, I = 900$) from the CAR model of expression (2) where $\tau^2 = 1$.

3. We generate $R(A_i)$ using relation (1). We consider different choices for variances of interest $\sigma_r^2(A_i)$:

   3.1. $\sigma_r^2(A_i) = 1$, $\forall A_i$;
   3.2. $\sigma_r^2(A_i) \sim \log N(\mu, \sigma^2)$ with $\mu = \log(0.5)$ and $\sigma^2 = 1$;
   3.3. $\sigma_r^2(A_i) \sim \mathrm{logCAR}(\tau_*^2)$, with $\tau_*^2 = 0.5$.

4. Then, two different sets of 200 locations are randomly generated within the unit square (hereafter, "Coords1" and "Coords 2").

5. For each location $\mathbf{s}_j$, ($j = 1, \ldots, n = 200$) belonging to a grid cell A_i, we generate $\widetilde{V}(\mathbf{s}_j)$ using relation (4) with $\sigma_{\tilde{v}}^2 = 1$ and fixed value of decay parameter ϕ_v. In particular, we set $\phi_v = 2.8$ or $\phi_v = 11.25$ corresponding, respectively, to spatial ranges of roughly 80% and 20% of the maximum distance over the region. We also consider the addition of some bias to (4) in a portion of the region (top right) when we do the sampling.

6. Finally, the validation data $V(\mathbf{s}_j)$ are generated from equation (3) where $\sigma_v^2 = 1$.

Given the sampling scheme described in the previous steps, we fit model (1) - (4) under both prior models (6) and (7). Moreover, we allow for the case when we fit the model setting the spatial decay parameter ϕ_v far away from its true value. Summarizing the sampling/fitting combinations, we consider the scenarios listed in Table 6.

We provide, as examples, the sampling design for scenarios (g) and (h). Figure 3 shows the absolute true errors, i.e. the absolute differences between the simulated $R(A_i)$ and the simulated $R_{true}(A_i)$ under both scenarios. Figure 4 shows the 900 estimated uncertainties (standard deviations) associated with the gridded data of scenario (g) and obtained under both priors (6) and

12

Table 1: Sampling/fitting simulation design.

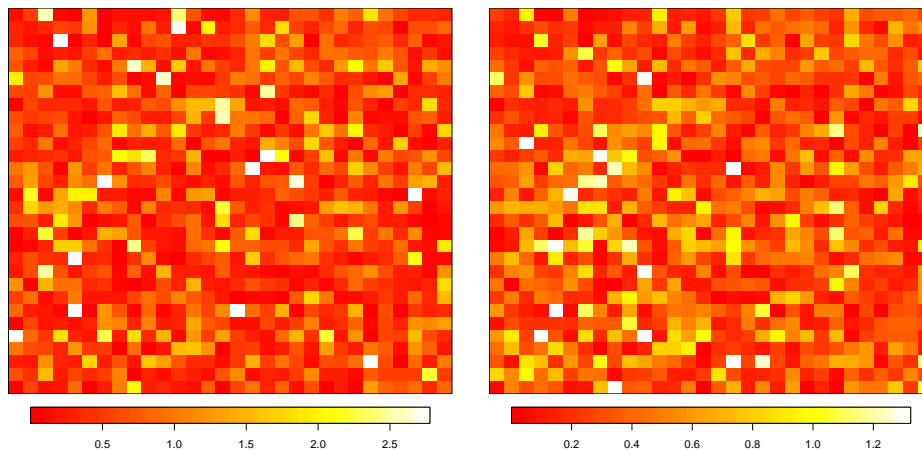| Scenario	Sampling				Fitting[1]
	$\sigma_r^2(A_i)$	Bias to (4)	ϕ_v	Validation data	$\hat{\phi}_v$
(a)	1, $\forall(A_i)$	NO	2.8	'Coords 1'	2.8
(b)	1, $\forall(A_i)$	YES	2.8	'Coords 1'	2.8
(c)	1, $\forall(A_i)$	NO	11.25	'Coords 1'	11.25
(d)	1, $\forall(A_i)$	YES	11.25	'Coords 1'	11.25
(e)	1, $\forall(A_i)$	NO	2.8	'Coords 1'	11.25
(f)	1, $\forall(A_i)$	NO	2.8	'Coords 2'	2.8
(g)	from $\log N(-0.69, 1)$	NO	2.8	'Coords 1'	2.8
(h)	from $\log\mathrm{CAR}(0.5)$	NO	2.8	'Coords 1'	2.8

Figure 3: True errors (absolute values) under scenarios (g) and (h) in the left and right panel, respectively.

(7). Equivalently, local posterior uncertainties for scenario (h) are shown in Figure 5. Such figures reveal that the estimated uncertainties tend to be larger at grid cells with higher true errors, in accord with the idea of uncertainty described in Section 4. Also, we might compare the local estimated uncertainties with the absolute true errors via the Spearman's correlation statistic ρ. For instance, with scenario (g) we have $\rho = 0.52$ and $\rho = 0.28$ under priors (6) and (7), respectively; given scenario (h), we obtain $\rho = 0.31$ and $\rho = 0.16$ for both priors. A full elaboration of the finding from this simulation study is presented in the supplementary materials. Here, we summarize with Table 2 which shows the comparison between the two alternative approaches via criterion (9) for all scenarios (here, $c = 1$). In the simulation
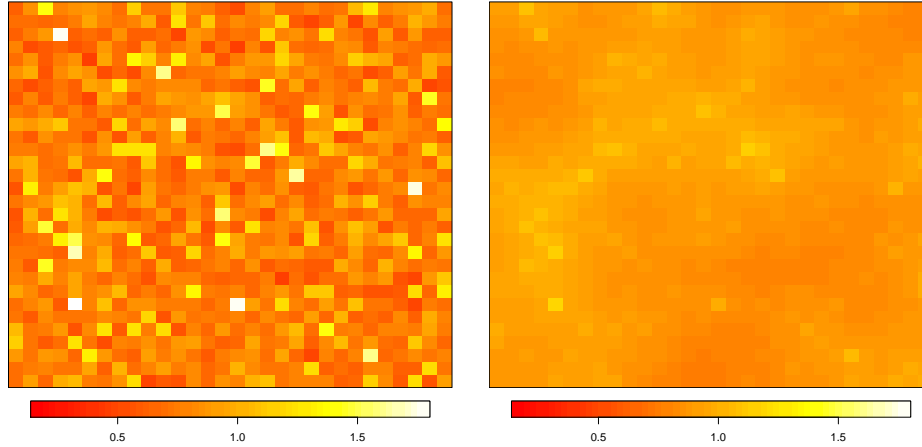
Figure 4: Scenario (g): estimated local uncertainties (standard deviations) under priors (6) and (7) in the left and right panel, respectively.

study, the true average $R_{true}(A_i)$ is available and it replaces $V(\mathbf{s}_j)$ in the criterion. The table reveals little differences between the two approaches for modeling, a priori, the variances of interest.

Table 2: Criterion (9) for different scenarios under the two alternative prior models.

Scenario	$\log N(\mu, \sigma^2)$	$\log\mathrm{CAR}(\tau_*^2)$
(a)	$1.066 + 0.187 = 1.253$	$1.072 + 0.158 = 1.230$
(b)	$1.205 + 0.174 = 1.380$	$1.101 + 0.159 = 1.260$
(c)	$1.018 + 0.196 = 1.214$	$1.079 + 0.159 = 1.238$
(d)	$1.076 + 0.190 = 1.266$	$1.084 + 0.159 = 1.244$
(e)	$1.092 + 0.182 = 1.274$	$1.085 + 0.157 = 1.242$
(f)	$1.162 + 0.160 = 1.322$	$1.010 + 0.152 = 1.252$
(g)	$0.825 + 0.140 = 0.965$	$0.832 + 0.142 = 0.974$
(h)	$0.228 + 0.101 = 0.329$	$0.240 + 0.098 = 0.338$

7. Attaching uncertainty to RUC output

Finally, we return to the RUC model output to illustrate our uncertainty quantification approach using the data fusion model of Section 3. From
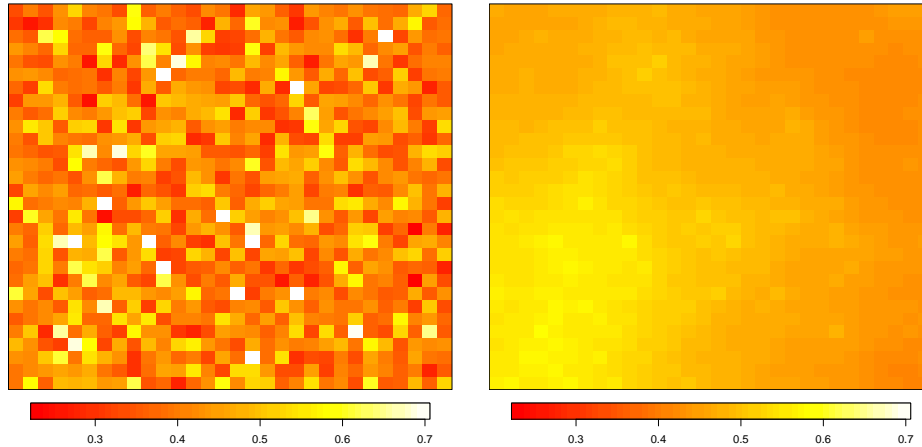
Figure 5: Scenario (h): estimated local uncertainties (standard deviations) under priors (6) and (7) in the left and right panel, respectively.

Section 2, we recall that RUC model produces weather short-term predictions for the conterminous U.S. over grid cells of size 13×13 kilometers.

We fit model (1) - (4) under both prior models (6) and (7). Regarding the spatial decay parameter ϕ_v, exploratory analysis suggested to set the parameter at roughly 60% of the maximum distance over the study region. The results are not sensitive to choice of ϕ_v; in fact, we also experimented with ϕ_v corresponding to spatial range of approximately 30% of the maximum distance, obtaining results indistinguishable from those presented in this section (see supplementary materials).

Posterior summaries of the unknown parameters are presented in Table 3 under both nonspatial and spatial variance specifications in Section 4. The posterior means of true temperature $\widetilde{R}$'s with specifications (6) and (7) are shown in Figure 6 in the left and right panel, respectively. Comparison of the two panels reveals little difference between posterior means of the $\widetilde{R}$'s under the two approaches except for somewhat smoother estimates under prior (6). This is consistent with similar posterior summaries of τ^2 in both models. Figure 7 shows the estimated uncertainty maps associated with RUC output under prior (6) and (7) in the left and right panel, respectively. The uncertainty map resulting from the logCAR prior model on σ_r^2's reveals smooth spatial variation, as we expected. It also worth noting from Figures 6 and 7 that the estimated $\widetilde{R}$'s and the attached uncertainties have differ-

15

Table 3: Posterior summaries of model parameters.

Parameters	$\log N(\mu, \sigma^2)$	$\mathrm{logCAR}(\tau_*^2)$
σ_v^2	0.933 (0.270, 2.221)	0.902 (0.283, 2.118)
$\sigma_{\tilde{v}}^2$	1.683 (1.203, 2.225)	1.701 (1.231, 2.242)
τ^2	0.787 (0.746, 0.828)	0.728 (0.694, 0.764)
μ	-6.248 (-7.327, -5.274)	
τ_*^2		10.868 (7.428, 14.910)

Table 4: Criteria (9) under the two alternative prior models.

$\log N(\mu, \sigma)$	$\mathrm{logCAR}(\tau_*^2)$
1.99 + 2.34 = 4.33	0.07 + 2.36 = 2.43

ent spatial patterns. In fact, high values in $\widetilde{R}(A_i)$ do not necessarily imply high uncertainty; rather, high uncertainty is linked to large realized residuals. Moreover, we have no reason to believe that the uncertainty should be proportional in some way to true temperature, that larger variances should be associated with larger responses. Finally, we note that our stochastic model does not attach higher uncertainty to RUC grid cells that contain sites. Intuitively, uncertainties associated with deterministic predictions need not be larger at grid cells farther from validation sites than at locations nearer.

The comparison between the two alternative approaches is shown in Table 4 revealing that model (7) is preferred according to the criterion (9). **The story here seems to be that, with criterion (9), the biases for (6) and (7) are essentially the same but the posterior variances are tiny under (7) compared with (6). Do we believe the uncertainties are that small? How can we explain the enormous difference in scale? Are you sure these numbers are correct? We do need to say something in the text.**
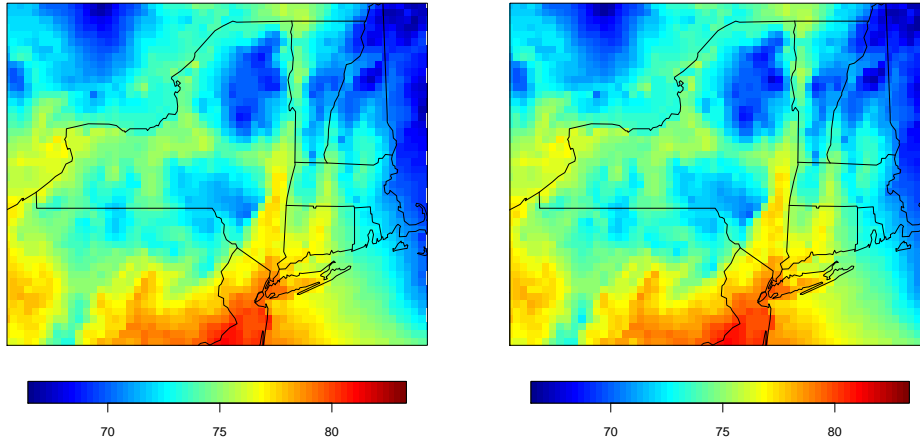
Figure 6: Posterior means of $\widetilde{R}(A_i)$ under priors (6) and (7) in left and right panel, respectively.
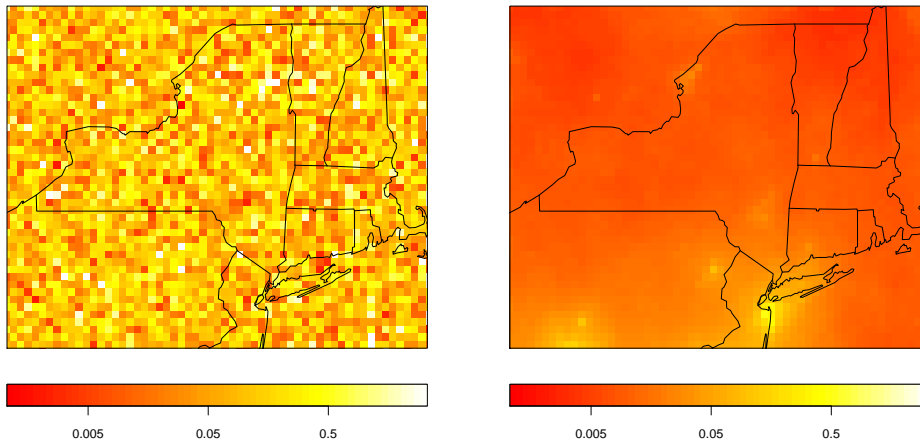

Figure 7: Estimated standard deviations under priors (6) and (7) in left and right panel, respectively.

8. Summary

With the goal of attaching uncertainty to deterministic surfaces, we have developed a hierarchical model to fuse the deterministic output with some validation data. We quantify the uncertainty using local posterior variances with priors on these variances that treat them as either independent or spatial structured random effects.

Future work will find us dealing with calibration of the RUC model, extending the simple measurement error specification in (4), jointly with the uncertainty assessment of the model output. Extension can also concern joint forecast maps, e.g. temperature and precipitation, attaching uncertainty through joint stochastic modeling. Regarding the attached uncertainty to RUC output, we are also interested in seasonal uncertainties, say winter or summer uncertainty maps. Finally, starting on May 1, 2012, the NCEP replaced the RUC model with the RAP numerical weather model, alleged to be a better deterministic model. Therefore, it will be of interest to compare uncertainty maps associated with RUC output against those attached to RAP predictions. **Should we mention the idea of dynamics, that with temp maps over time, we could potentially borrow strength across time to better attach uncertainty?**

Acknowledgements

Appendix A. Full conditional distributions

We partition $\widetilde{\mathbf{R}} = \left(\widetilde{\mathbf{R}}^{(1)}, \widetilde{\mathbf{R}}^{(2)}\right)$, where $\widetilde{\mathbf{R}}^{(1)}$ corresponds to the numerical model output for the n grid cells where monitoring stations are, while $\widetilde{\mathbf{R}}^{(2)}$ is the vector containing the numerical model output at $(I-n)$ grid cells where no observations are made.

The full conditional distributions for the inverse of the variance parame-

ters σ_v^2, $\sigma_{\tilde{v}}^2$, τ^2 and τ_*^2 are:

$$\frac{1}{\sigma_v^2}\,|rest \sim Ga\left(a_\sigma + \frac{n}{2},\ b_\sigma + \frac{1}{2}\left(\mathbf{V} - \widetilde{\mathbf{V}}\right)'H^{-1}(\phi_v)\left(\mathbf{V} - \widetilde{\mathbf{V}}\right)\right)$$

$$\frac{1}{\sigma_{\tilde{v}}^2}\,|rest \sim Ga\left(a_\sigma + \frac{n}{2},\ b_\sigma + \frac{1}{2}\left(\widetilde{\mathbf{V}} - \widetilde{\mathbf{R}}^{(1)}\right)'\left(\widetilde{\mathbf{V}} - \widetilde{\mathbf{R}}^{(1)}\right)\right)$$

$$\frac{1}{\tau^2}\,|rest \sim Ga\left(a_\tau + \frac{I}{2},\ b_\tau + \frac{1}{2}\widetilde{\mathbf{R}}'\left(D_w - W\right)\widetilde{\mathbf{R}}\right)$$

$$\frac{1}{\tau_*^2}\,|rest \sim Ga\left(a_\tau + \frac{I}{2},\ b_\tau + \frac{1}{2}\left(\log(\boldsymbol{\sigma}_r^2)\right)'\left(D_w - W\right)\left(\log(\boldsymbol{\sigma}_r^2)\right)\right)$$

where $\boldsymbol{\sigma}_r^2 = \left(\sigma_r^2(A_1), \ldots, \sigma_r^2(A_I)\right)'$, $D_w = diag(w_i)$ and W is the proximity matrix, according to the rook's neighborhood structure.

The posterior conditional distribution for $\widetilde{\mathbf{V}}$ is a multivariate normal distribution $N(D_{\tilde{v}}d_{\tilde{v}}, D_{\tilde{v}})$, where

$$D_{\tilde{v}}^{-1} = \frac{1}{\sigma_v^2}H^{-1}(\phi_v) + \frac{1}{\sigma_{\tilde{v}}^2}I_n$$

$$d_{\tilde{v}} = \frac{1}{\sigma_v^2}H^{-1}(\phi_v)\mathbf{V} + \frac{1}{\sigma_{\tilde{v}}^2}\widetilde{\mathbf{R}}^{(1)}$$

We sample the elements of $\widetilde{\mathbf{R}}$ using a univariate sampling scheme as following. If $\widetilde{R}(A_i) \in \widetilde{\mathbf{R}}^{(1)}$, the full conditional distribution for $\widetilde{R}(A_i)$ is the normal distribution $N(D_{r1}d_{r1}, D_{r1})$, where

$$D_{r1}^{-1} = \frac{1}{\sigma_r^2(A_i)} + \frac{1}{\sigma_{\tilde{v}}^2} + \frac{w_i}{\tau^2}$$

$$d_{r1} = \frac{R(A_i)}{\sigma_r^2(A_i)} + \frac{\widetilde{V}(A_i)}{\sigma_{\tilde{v}}^2} + \frac{1}{\tau^2}\sum_{i' \sim i}\widetilde{R}(A_{i'})$$

If $\widetilde{R}(A_i) \in \widetilde{\mathbf{R}}^{(2)}$, the full conditional distribution for $\widetilde{R}(A_i)$ is the normal distribution $N(D_{r2}d_{r2}, D_{r2})$, where

$$D_{r2}^{-1} = \frac{1}{\sigma_r^2(A_i)} + \frac{w_i}{\tau^2}$$

$$d_{r2} = \frac{R(A_i)}{\sigma_r^2(A_i)} + \frac{1}{\tau^2}\sum_{i' \sim i}\widetilde{R}(A_{i'})$$

19

The posterior conditional for μ is the normal distribution $N(D_\mu d_\mu, D_\mu)$, where

$$D_\mu^{-1} = \frac{1}{g^2} + \frac{I}{\sigma^2}$$

$$d_\mu = \frac{\sum_{i=1}^{I} \log\left(\sigma_r^2(A_i)\right)}{\sigma^2}$$

Finally, given both the log-normal and the logCAR prior models for $\sigma_r^2(A_i)$, the full conditionals cannot be obtained in closed form; so we use a random walk Metropolis proposal step to generate samples from their posteriors.

Banerjee, S., Carlin, B. P., Gelfand, A. E., 2004. Hierarchical Modeling and Analysis for Spatial Data. Chapman and Hall/CRC.

Benjamin, S. G., Dèvènyi, D., Weygandt, S. S., Brundage, K. J., Brown, J. M., Grell, G. A., Kim, D., Schwartz, B. E., Smirnova, T. G., Smith, T. L., Manikin, G. S., 2004. An hourly assimilation-forecast cycle: the RUC. Mon. Weather. Rev. 132, 495–518.

Berrocal, V. J., Gelfand, A. E., Holland, D. M., 2012. Space-time data fusion under error in computer model output: an application to modeling air quality. Biometrics 68, 837–848.

Berrocal, V. J., Raftery, A. E., Gneiting, T., 2007. Combining spatial statistical and ensemble information in probabilistic weather forecasts. Mon. Weather. Rev. 135, 1386–1402.

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. J. Roy. Stat. Soc. B Met. 36, 192–236.

Chaloner, K., Brant, R., 1988. A Bayesian approach to outlier detection and residual analysis. Biometrika 75, 651–659.

Cumming, J. A., Goldstein, M., 2010. Bayes linear uncertainty analysis for oil reservoirs based on multiscale computer experiments. In: Gelfand, A. E., Fuentes, M., Guttorp, P., Diggle, P. J. (Eds.), Handbook of Spatial Statistics. CRC Press, pp. 241–270.

Di Narzo, A. F., Cocchi, D., 2010. A Bayesian hierarchical approach to ensemble weather forecasting. J. Roy. Stat. Soc. C App. 59, 405–422.

Fuller, W. A., 1987. Measurement Error Models. Wiley.

Gelfand, A. E., Ghosh, S. K., 1998. Model choice: A minimum posterior predictive loss approach. Biometrika 85, 1–11.

Gelfand, A. E., Smith, A. F. M., 1990. Sampling-based approaches to calculating marginal densities. J. Am. Stat. Assoc. 85, 398–409.

Gelfand, A. E., Zhu, L., Carlin, B. P., 2001. On the change of support problem for spatio-temporal data. Biostatistics 2, 31–45.

Ghosh, S., Gelfand, A. E., Mølhave, T., 2012. Attaching uncertainty to deterministic spatial interpolations. Stat. Methodol. 9, 251–264.

Gneiting, T., Raftery, A. E., Westveld, A. H., Goldman, T., 2005. Calibrated probabilistic forecasting using ensemble model output statistics and minimum crps estimation. Mon. Weather. Rev. 133, 1098–1118.

Kennedy, M. C., O'Hagan, A., 2001. Bayesian calibration of computer models (with discussion). J. Roy. Stat. Soc. B Met. 63, 425–464.

Kleiber, W., Raftery, A. E., Baars, J., Gneiting, T., Mass, C. F., Grimit, E., 2011. Locally calibrated probabilistic temperature forecasting using geostatistical model averaging and local bayesian model averaging. Mon. Weather. Rev., 2630–2649.

Oakley, J. E., O'Hagan, A., 2004. Probabilistic sensitivity analysis of complex models: a Bayesian approach. J. Roy. Stat. Soc. B Met. 66, 751–769.

Paci, L., Gelfand, A. E., Holland, D. M., 2013. Spatio-temporal modeling for real-time ozone forecasting. Spat. Stat. 4, 79–93.

Poole, D., Raftery, A. E., 2000. Inference for deterministic simulation models: The Bayesian melding approach. J. Am. Stat. Assoc. 95, 1244–1255.

Raftery, A. E., Givens, G. H., Zeh, J. E., 1995. Inference from a deterministic population dynamics model for bowhead whales. J. Am. Stat. Assoc. 90, 402–416.

Raftery, A. E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian model averaging to calibrate forecast ensembles. Mon. Weather. Rev. 133, 1155–1174.

Reich, B. J., Hodges, J. S., 2008. Modeling longitudinal spatial periodontal data: a spatially adaptive model with tools for specifying priors and checking fit. Biometrics 64, 790–799.

Sloughter, J. M., Gneiting, T., Raftery, A. E., 2013. Probabilistic wind vector forecasting using ensembles and bayesian model averaging. Mon. Weather. Rev. 141, 2017–2119.

Smith, R. L., Tebaldi, C., Nychka, D. W., Mearns, L. O., 2009. Bayesian modeling of uncertainty in ensembles of climate models. J. Am. Stat. Assoc. 104, 97–116.

Sørbye, S. H., Rue, H., 2013. Scaling intrinsic Gaussian Markov random field priors in spatial modelling. Spat. Stat.

Yan, J., 2007. Spatial stochastic volatility for lattice data. J. Agric. Biol. Envir. S. 12, 25–40.

Zellner, A., 1975. Bayesian analysis of regression error terms. J. Am. Stat. Assoc. 70, 138–144.

Zhang, H., 2004. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. J. Am. Stat. Assoc. 99, 250–261.