

**The University of British Columbia**  
**Department of Statistics**  
**Technical Report #233**

**An Appraisal of Bayesian Melding for Physical-Statistical Modeling**

Zhong Liu, Nhu D. Le, James V. Zidek

July 12, 2007

Department of Statistics

University of British Columbia

333-6356 Agricultural Road

Vancouver, BC

V6T 1Z2

Canada

# An Appraisal of Bayesian Melding for Physical-Statistical Modeling

Zhong Liu<sup>1</sup>, Nhu D. Le<sup>2</sup>, James V. Zidek<sup>3</sup>

**Abstract.** In this technical report, we review the Bayesian melding model, proposed by Fuentes and Raftery (2005). We give technical details about using Gibbs sampling algorithm to fit the model and critically assess the model through various simulation studies and applications of combining ozone measurements with simulated ozone levels from deterministic model AQM.

keywords: Bayesian melding, Bayesian hierarchical model, deterministic model, MCMC, non-stationary spatial process.

## 1 Introduction

This report explores through simulation studies and applications, the use of a method called “melding” for combining the output from deterministic models (“simulated data” or “modeling output”) with measurements from sites that monitor spatial processes. We also refer to measurements as observations or observed values in the rest of this report. The melding method as proposed by Fuentes and Raftery (2005) and used by them for sulfur dioxide, one of the five criteria pollutants regulated in the US by the National Ambient Air Quality Standards or NAAQS. However, our application concerns ozone, another of the criteria pollutants whose standards are currently under review by the US Environmental Protection Agency. More precisely, we explore melding as a tool for modeling hourly, daily and weekly ozone concentrations over the eastern and central United States. That challenging application provides us with a critical assessment of the method.

This report presents that assessment along along with a detailed description of its im-

---

<sup>1</sup>Department of Statistics, U. of British Columbia. Email: zliu@stat.ubc.ca

<sup>2</sup>British Columbia Cancer Research Centre. Email: nle@bccrc.ca

<sup>3</sup>Department of Statistics, U. of British Columbia. Email: jim@stat.ubc.ca

plementation as incorporated in our purpose-built online software that could be used to replicate our findings or model other environmental processes that resemble the ozone field we study. Since the problem of combining statistical and deterministic models is the subject of considerable current interest, this report may well have implications in a much broader context than the one of concern here.

Deterministic models have long been used in climate and environmental studies, one example being the Regional Climate Model described in Caya et al. (1995). Another example, the MAQSIP (Multiscale Air Quality Simulation Platform) model is described in Odman and Ingram (1996). The chemical transport model (CTM) called GEOS-CHEM, which models hourly ozone fields like the one of interest in this report, is suggested in the current review of NAAQS for ozone as a method for estimating ozone's policy relevant background (PRB) level (Garner et al. (2005)). The PRB level, a baseline for the NAAQS, is the ozone concentration that would obtain if there were no anthropogenic sources of ozone in North America. Since the PRB cannot be measured, it has to be modeled. Moreover the model generates useful byproducts such as estimates of the fraction of the ozone field from the vertical transport of stratospheric ozone to the troposphere.

The models above are called deterministic because repeated runs with fixed inputs yield the same output and lacking randomness, they differ from statistical models. However, unlike statistical models, they do attempt to capture the fundamental dynamic processes that govern the phenomenon of interest. The complexity of the equations that describe such dynamics can render analytic solution impossible and force the use of numerical models, typically with lots of complicated computer code and slow running times.

Deterministic models have a number of advantages. First and foremost they, unlike statistical models, do incorporate prior knowledge about the underlying processes. Since their parameters and inputs are adjustable, they enable scenario analysis under hypothetical changes to an existing regime, say as a result of abatement strategies. Run on a computer, they enable experiments to be run when real experiments would not be feasible due to ethical

or other constraints. Such experiments can enable the study of input - output relationships and possibly suggest optimal control settings for use in the real phenomenon they represent.

However, concerns inevitably arise in the application of deterministic models. First the much quoted remark George Box made about statistical models, that “all are wrong, but some are useful” also applies to these models. Thus in applications, two fundamental issues arise. How wrong and how useful? They in turn point to the need for some kind of model validation or assessment. Kasibhatla and Chameides (2000); Hogrefe et al. (2001b,a) address these issues in meteorological and environmental science contexts by using traditional scatter plots and least squares analyses of observed values and modeling output. Guttorp and Walden (1987) exam these issues within a statistical framework. Of particular relevance to this report is the work of Fuentes et al. (2003); Fuentes and Raftery (2002, 2005), in which the melding approach is used to combine observed and modeling output while respecting their intrinsic differences. The combination highlights the discrepancies and similarities between the values to be assessed. Moreover this method can be used in other ways such as predicting of the phenomenon of interest using both observed and model output values. Another Bayesian approach is taken by Sanso and Guenni (2002) who like Fuentes et al. (2003); Fuentes and Raftery (2002, 2005), postulates a “true underlying process” but unlike them is specifically concerned with rainfall and uses a truncated model for the relationship between the true underlying process and the data. In our application, melding is used to combine real ozone measurements and modeling output from a deterministic model. Section 5 presents more details about this deterministic model.

This report is organized as follows. Section 2 gives some background about stationary and non-stationary spatial covariance. This section also includes details about a class of non-stationary spatial model proposed by Fuentes and Smith (2001). Section 3 describes melding in detail along with its extension to ensembles of multiple deterministic models and measuring devices. Section 4 describes the results of an extensive simulation study that analyzes melding’s performance under a variety of conditions, investigates various

issues and discovers possible improvements in the estimation of the covariance parameters. Section 5 applies melding to the ozone air pollution data in the eastern and central USA considered by Kasibhatla and Chameides (2000).

## 2 Stationary and Non-Stationary Spatial Covariance

### 2.1 Definition and Estimation of The Variogram

Suppose we have a real-valued random process  $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ , which is observed at locations  $\{\mathbf{s}_i : i = 1, \dots, n\}$  over a geographic region  $D \subset \mathbb{R}^d$ ,  $d$  being a positive integer. The random process  $\{Z\}$  is defined as second order stationarity if it satisfies for all  $\mathbf{s}$

$$\begin{aligned} \mathbb{E}(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})) &= 0, \\ \text{Var}(Z(\mathbf{s} + \mathbf{h})) &= \text{Var}(Z(\mathbf{s})), \\ \text{Cov}(Z(\mathbf{s} + \mathbf{h}), Z(\mathbf{s})) &= C(\mathbf{h}), \end{aligned}$$

where  $C(\cdot)$  is a covariance function. In other words, the correlation between responses at two locations depends only on their degree of separation. In addition, the first and second order moments of the random process  $Z(\mathbf{s})$  are the same for all  $\mathbf{s}$ . Furthermore, if  $C(\mathbf{h}) = C(\|\mathbf{h}\|)$ , where  $\|\mathbf{h}\|$  is the length of  $\mathbf{h}$ , the covariance function  $C(\cdot)$  is called isotropic. Another very important quantity used in spatial statistics, the variogram, is related to the spatial covariance between two locations. We have

$$\text{Var}(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})) = 2\gamma(\mathbf{h}),$$

$2\gamma(\mathbf{h})$  being known as a variogram. The so-called semi-variogram refers to  $\gamma(\mathbf{h})$ .

The classical estimator of the variogram proposed by Matheron (1962) is obtained by using the method of moments. The estimator is

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} (Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2, \quad (1)$$

where the sum is over  $N(\mathbf{h}) = \{(i, j) : \mathbf{h} - \boldsymbol{\delta} \leq |\mathbf{s}_i - \mathbf{s}_j| \leq \mathbf{h}\}$  and  $|N(\mathbf{h})|$  is the number of distinct elements in  $N(\mathbf{h})$ . Although unbiased, this classical estimator is highly affected by

outliers due to the squared term in the summation. Hawkins and Cressie (1984) presented a more robust estimator

$$2\bar{\gamma}(\mathbf{h}) = \frac{\left\{ \frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} |Z(\mathbf{s}_i) - Z(\mathbf{s}_j)|^{1/2} \right\}^4}{0.457 + \frac{0.494}{|N(\mathbf{h})|}}. \quad (2)$$

From our definition of the semi-variogram, it is clear that  $\gamma(\mathbf{0}) = 0$ . However, more generally  $\gamma(\mathbf{h}) \rightarrow \tau > 0$  maybe allowed as  $\mathbf{h} \rightarrow \mathbf{0}$ , in which case  $\tau$  is the called nugget effect by Matheron (1962). One of the sources of the nugget effect is measurement error and commonly measurements are assumed to be the ground truth  $Z(\mathbf{s})$  plus some measurement error. The book by Cressie (1993) gives more details about the nugget effect.

The covariance function only needs to ensure the covariance matrix it generates is positive definite and symmetric. In the book Stein (1999), Bochner's theorem specifies that the positive definite covariance functions are those which are Fourier transforms of non-negative Borel measures. One commonly used covariance function proposed by Matern (1960) has the following form:

$$C_{\boldsymbol{\theta}}(d) = \frac{\sigma}{2^{\nu-1}\Gamma(\nu)} (2\nu^{1/2}|d|/\rho)^{\nu} K_{\nu}(2\nu^{1/2}|d|/\rho),$$

$d$  being the distance between any two locations. For convenience, we let  $\boldsymbol{\theta} = (\sigma, \rho, \nu)$ . Here  $\sigma$ , the *sill* parameter, represents the variance of random process  $Z(\mathbf{s})$  while  $\rho$ , the *range* parameter, determines how fast the correlation decreases when distance  $d$  increases and  $\nu$ , the *smoothing* parameter controls the smoothness of the covariance function.  $K_{\nu}$  is the modified Bessel function of type III as described by Abramowitz and Stegun (1972). Matern covariance functions have the advantage of flexibility. Certain choices of the smoothing parameter  $\nu$  reduce it to simple well known covariance functions. For example,  $\nu_s = 1/2$  gives the so-called exponential covariance function,

$$C_{\boldsymbol{\theta}}(d) = \begin{cases} \sigma \exp(-|d|/\rho) & \text{if } |d| > 0; \\ \sigma + \tau & \text{if } |d| = 0. \end{cases} \quad (3)$$

The *nugget* parameter is  $\tau$ . It is obvious that the covariance functions defined above are only suitable for stationary random fields because we assume the covariance  $C_{\theta}(d)$  is only a function of the distance  $d$ . However, in many cases, it is not reasonable to assume second order stationarity. In recent years, the non-stationarity problem has received a substantial amount of attention. One of the earliest and most important papers is due to Sampson and Guttorp (1992), who obtain a semi-parametric estimate of the non-stationary spatial covariance function by transforming the original geographical map into another, the deformed plane. However that deformation method relies on repeated measurements to give an empirical estimate of the variogram, a serious limitation in the context of geostatistics, for example, where typically just one realization of a space - time process is available. Haas (1995, 1998) circumvents that problem with a moving windows method as does Higdon et al. (1999) with a convolution approach although these methods may require data from a large number of monitoring sites to be effective. The paper of Gelfand et al. (2004) considers the non-stationarity problem for a multivariate case. Fuentes and Smith (2001) propose a class of non-stationary spatial models used in Fuentes et al. (2003), Fuentes and Raftery (2002), Fuentes and Raftery (2005). This approach's appeal derives not only from its its circumvention of the need for replicate measurements but as well, from its simple intuitive, easy to understand formulation as well as the ease with which it can be implemented. The next subsection presents more detail and discusses the method.

## 2.2 A Class of Non-stationary Spatial Models

This subsection presents more detail of the non-stationary covariance model proposed by Fuentes and Smith (2001). It assumes the existence at a number of locations of latent, independent, stationary random processes that need not have any physical meaning. The observed non-stationary process is represented as a weighted average or convolution of these latent processes. Each latent random process has its own covariance parameters which vary from one latent process (location) to another. The weight attached to each of the

stationary processes in the representation of the observed non-stationary process depends on its location. So the covariance of the process between any two locations depends not only on their distance  $d$  but also on their locations. The rest of this section gives a detailed account of the non-stationary model of Fuentes and Smith (2001). We include it to enable us to introduce the detailed MCMC algorithm used to fit this model in Section 3.2.

Suppose the latent stationary processes are  $Z_i, i = 1, \dots, K$ , with  $\text{Cov}(Z_i(\mathbf{s}), Z_j(\mathbf{s})) = 0$  for  $i \neq j$ . Then the observed process  $Z(\mathbf{s})$  can be expressed as

$$Z(\mathbf{s}) = \sum_{i=1}^K Z_i(\mathbf{s})w_i(\mathbf{s}), \quad \sum_{i=1}^K w_i = 1$$

in which  $w_i$  is the weight attached to stationary process  $Z_i$ . The above equation can be extended to an integral by replacing the weight with a kernel function as follows:

$$Z(\mathbf{x}) = \int_D K(\mathbf{x} - \mathbf{s})Z_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x})d\mathbf{s},$$

where  $K(\mathbf{X} - \mathbf{s})$  is a kernel function. So, the weight of latent process  $Z_{\boldsymbol{\theta}(\mathbf{s})}$  depends on the location difference vector between  $\mathbf{x}$  and  $\mathbf{s}$ . The spatial covariance parameter vector of the latent stationary process  $\{Z_{\boldsymbol{\theta}(\mathbf{s})}\}$  is  $\boldsymbol{\theta}(\mathbf{s})$  depending on its center location  $\mathbf{s}$ . That is why these latent stationary processes are called ‘‘locally stationary’’.

Since the covariance between  $Z_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x}_1)$  and  $Z_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x}_2)$  is

$$\text{Cov}(Z_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x}_1), Z_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x}_2)) = C_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x}_1 - \mathbf{x}_2),$$

the covariance between  $Z(\mathbf{x}_1)$  and  $Z(\mathbf{x}_2)$  in the non-stationary process  $Z$  can be expressed as a convolution of the local covariance  $C_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x}_1 - \mathbf{x}_2)$ :

$$C(Z(\mathbf{x}_1), Z(\mathbf{x}_2)) = \int_D K(\mathbf{x}_1 - \mathbf{s})K(\mathbf{x}_2 - \mathbf{s})C_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x}_1 - \mathbf{x}_2)d\mathbf{s}. \quad (4)$$

The kernel function can be any of the form  $K(u) = \frac{1}{h^2}K_0(\frac{u}{h})$ ,  $K_0$  being any non-negative

function with integral 1. A particularly convenient choice and the one we select in our application, is the quadratic weight function

$$K_0(\mathbf{u}) = \frac{3}{4}(1 - u_1^2)_+ \frac{3}{4}(1 - u_2^2)_+,$$

with  $\mathbf{u} = (u_1, u_2)$  and in general,  $a_+ = \max\{0, a\}$  for any scalar valued quantity  $a$ . The bandwidth parameter  $h$  can be any positive scalar valued quantity subject to certain restrictions explained later in this subsection. Also in Subsection 4.4, a simulation is carried out to examine the effect of different choices of the bandwidth  $h$  on prediction and parameter estimation.

Our choice of the kernel function implies that for a given pair of locations  $\mathbf{x}_1$  and  $\mathbf{x}_2$ , only the local stationary process whose center location  $\mathbf{s}$  is within circles with origins  $\mathbf{x}_1$ ,  $\mathbf{x}_2$  and radius  $h$  will have an effect on their covariance.

In the paper by Fuentes and Smith (2001), the integral (4) is approximated by an average. First, the center locations  $\mathbf{s}_m$ ,  $m = 1, \dots, M$  of those latent stationary processes are chosen as points on a regular grid over the map. The kernel integral (4) is then replaced by

$$C_M(\mathbf{x}_1, \mathbf{x}_2; \boldsymbol{\theta}) = M^{-1} \sum_{m=1}^M K(\mathbf{x}_1 - \mathbf{s}_m)K(\mathbf{x}_2 - \mathbf{s}_m)C_{\boldsymbol{\theta}(\mathbf{s}_m)}(\mathbf{x}_1 - \mathbf{x}_2). \quad (5)$$

In the non-stationary model, the parameter  $\boldsymbol{\theta}$  of the latent stationary processes is a function of its center location  $\mathbf{s}$ . This function can be smooth but instead Fuentes and Smith (2001) assume an additive ‘‘ANOVA’’ type model for  $\boldsymbol{\theta}$  in terms of center locations  $\mathbf{s}$ . The stationary points are points of a regular grid over the map. For example, for the exponential covariance function, with just two parameters, the *sill*  $\sigma$  and the *range*  $\rho$ , the ‘‘ANOVA’’

forms would be:

$$\begin{aligned}
\sigma_{i,j} &= \alpha + r_i + c_j + \epsilon_{i,j} \quad i = 1, \dots, n_1 \\
\rho_{i,j} &= \alpha' + r'_i + c'_j + \epsilon'_{i,j} \quad j = 1, \dots, n_2 \\
\boldsymbol{\mu}_\sigma &= (\alpha, r_1, \dots, r_{n_1}, c_1, \dots, c_{n_2}) \\
\boldsymbol{\mu}_\rho &= (\alpha', r'_1, \dots, r'_{n_1}, c'_1, \dots, c'_{n_2}) \\
\epsilon_{i,j} &\sim N(0, \Sigma(\tau_\sigma, \eta_\sigma)) \\
\epsilon'_{i,j} &\sim N(0, \Sigma(\tau_\rho, \eta_\rho)).
\end{aligned} \tag{6}$$

In the above model,  $n_1$  and  $n_2$  are the numbers of points in the horizontal and vertical directions respectively. As well,  $r_i$  and  $c_j$  are main effects of center location's longitude and latitude on  $\sigma$ . The main effects of the center location's longitude and latitude on  $\rho$  are  $r'_i$  and  $c'_j$ . In this hierarchical model,  $\sigma$  and  $\rho$  might well have normal prior distributions. Then the hyper-parameters would be  $r_i, r'_i, c_j, c'_j$  and  $\tau_\sigma, \eta_\sigma, \tau_\rho, \eta_\rho$ . Subsection 3.2 explores an algorithm for fitting this non-stationary model.

### 3 Melding

This section presents an alternative to Kriging, the classical approach to spatial interpolation in the point-referenced data setting, a name given to it by Matheron (1963) in honor of D.G. Krige, a South African mining engineer. Kriging, the best linear unbiased predictor (BLUP), weights the available observations in accordance with the distance between the locations where they are made and that of the response to be interpolated. These BLUP weights are obtained by minimizing the variance of the interpolation error assuming the spatial covariance is stationary and known. In practice, the spatial covariance function is estimated empirically by (1) or (2). Its flexibility, simplicity and ease of computation have made Kriging very popular and hence a valuable tool in situations where it is applicable. The Kriging method is available in the R package "geoR" developed by Ribeiro and Diggle (2001), that we use to implement the method in our simulation studies and data analyses later in this report.

However, Kriging relies on a known variogram to compute its weights, treating estimated parameters as fixed. So the true uncertainty in interpolation is underestimated. The simulation study in Section 4.1 demonstrates that disadvantage.

More importantly, grid cell data from a deterministic model will be on coarser scales of resolution than the micro scale on which measurements are made. For example, the deterministic MAQSIP model studied in this report output its hourly ozone concentration data at a resolution  $6 \times 6$  km<sup>2</sup>. The mismatch of scales leads to a need to re-calibrate the simulated data when combining it with the data for interpolation. While this could be done in an ad hoc fashion with Kriging, it is not designed to deal with that issue in a fundamental way.

In contrast to Kriging, melding, the Bayesian method developed and studied by Fuentes and Raftery (2005), Fuentes et al. (2003), as well as Fuentes and Raftery (2002), is designed to do that. It combines measurements and deterministic modeling output in a Bayesian

framework that enables deterministic model assessment and spatial prediction including interpolation.

Melding links processes with responses on mismatched scales through an underlying true process  $\{Z(\mathbf{s}) : \mathbf{s} \in \mathfrak{R}^D\}$  called the “truth”,  $D$  being the dimension of the domain. If the location  $\mathbf{s}$  only has longitude and latitude then  $D = 2$ , while if the location also has altitude then  $D = 3$ . The underlying true (latent) process, being unobservable, must itself be estimated. Denote the measurement process by  $\{\hat{Z}(\mathbf{s}) : \mathbf{s} \in \mathfrak{R}^D\}$ , and the deterministic modeling output process by  $\{\tilde{Z}(B)\}$ ,  $B$  being the grid cell. To match  $Z(\mathbf{s})$ , we also hypothesize the existence of deterministic modeling output process  $\{\tilde{Z}(\mathbf{s}) : \mathbf{s} \in \mathfrak{R}^D\}$  based on locations  $\mathbf{s}$ . Of course, the purely conceptual process  $\{\tilde{Z}(\mathbf{s})\}$  does not actually exist. Its purpose: to link the modeling output with the truth at the micro- scale, thereby enabling its representation as an integral of  $\{\tilde{Z}(\mathbf{s})\}$ . The truth thus serves as the common basis for both processes  $\{\hat{Z}(\mathbf{s})\}$  and  $\{\tilde{Z}(\mathbf{s})\}$ . Moreover, having overcome the mismatched scales problem, we can “meld”  $\{\hat{Z}(\mathbf{s})\}$  with  $\{\tilde{Z}(\mathbf{s})\}$  via the true underlying process, hence the name “melding”.

The Bayesian model has the following mathematical form:

$$\begin{aligned}
\hat{Z}(\mathbf{s}) &= Z(\mathbf{s}) + e(\mathbf{s}) \\
Z(\mathbf{s}) &= \mu(\mathbf{s}) + \epsilon(\mathbf{s}) \\
Z(B) &= \frac{1}{|B|} \int_B Z(\mathbf{s}) d\mathbf{s} \\
\tilde{Z}(\mathbf{s}) &= a(\mathbf{s}) + b(\mathbf{s})Z(\mathbf{s}) + \delta(\mathbf{s}) \\
\tilde{Z}(B) &= \frac{1}{|B|} \int_B a(\mathbf{s}) d\mathbf{s} + \frac{1}{|B|} \int_B b(\mathbf{s})Z(\mathbf{s}) d\mathbf{s} + \frac{1}{|B|} \int_B \delta(\mathbf{s}) d\mathbf{s} \\
\mu(\mathbf{s}) &= X(\mathbf{s})\boldsymbol{\beta}.
\end{aligned} \tag{7}$$

In the above model, the measurements error and modeling output error are independent of each other. The measurement errors,  $e(\mathbf{s})$ , are independent and identically distributed, having a normal distribution  $N(0, \sigma_e^2)$ . The modeling output errors,  $\delta(\mathbf{s})$ , are independent

and identically distributed with a normal distribution  $N(0, \sigma_\epsilon^2)$ . The spatially correlated residuals,  $\epsilon(\mathbf{s})$ , have zero mean and covariance matrix  $\Sigma(\boldsymbol{\theta})$ , where  $\boldsymbol{\theta}$  is the covariance parameter vector. The number of locations is  $n$ .  $Z(B)$  and  $\tilde{Z}(B)$  are integrals of  $Z(\mathbf{s})$  and  $\tilde{Z}(\mathbf{s})$  over grid cell  $B$ . We only observe realizations of process  $\hat{Z}(\mathbf{s})$  and  $\tilde{Z}(B)$  at measured stations and modeling output for grid cells.

The mean of the true underlying process is  $\mu(\mathbf{s}) = X(\mathbf{s})\boldsymbol{\beta}$ , where  $X(\mathbf{s})$  is a polynomial function of the coordinates at location  $\mathbf{s}$  and  $\boldsymbol{\beta}$  is the corresponding coefficient vector. The covariance matrix  $\Sigma(\boldsymbol{\theta})$  is constructed with some covariance functions having a parameters vector  $\boldsymbol{\theta}$ . (See the previous section for a detailed discussion of covariance functions in stationary and non-stationary cases.) Throughout this report, we use the exponential covariance function (3). However, the software we have developed and provided can use either the Matern, Gaussian or exponential covariance function. In Bayesian melding Model (7), measurement  $\hat{Z}(\mathbf{s})$  is modeled as the truth  $Z(\mathbf{s})$  plus measurement error  $e(\mathbf{s})$  and modeling output  $\tilde{Z}(\mathbf{s})$ , as  $Z(\mathbf{s})$  times a multiplicative calibration parameter  $b(\mathbf{s})$  plus additive calibration parameter  $a(\mathbf{s})$  and random error  $\delta(\mathbf{s})$ . Here we use the term “calibration”, not the more prejudicial term “bias” that would seriously misrepresent what the deterministic models aim to do. For example, in the equations for climate models that are embraced by our theory, simplifications are made that filter out micro - scale effects. It would be inappropriate (in fact, pejorative) to call their output biased since by symmetry, the observations would also “biased”. Instead, being on different scales, neither is correctly calibrated with respect to the other.

In general, we can assume that the calibration parameters are functions of the location of  $\mathbf{s}$ , namely  $b(\mathbf{s})$  and  $a(\mathbf{s})$  in the model specification, to take into account their variability with respect to locations. For simplicity, we assume  $a$  and  $b$  are constants throughout the rest of this chapter for derivation and in the simulation study. But later in the data analysis, we assume  $a$  is a function of the coordinates of the location  $\mathbf{s}$  and  $b$  still remains constant. The reason to keep  $b$  constant is that  $a$  varies over space much more than  $b$ .

Fuentes and Raftery (2005) suggest and we use a Monte Carlo method to approximate the integrals in model (7). For example, we could sample  $L$  points  $\mathbf{s}_{1,B}, \dots, \mathbf{s}_{L,B}$  within grid cell  $B$ . From now on, we call these points “sampling points” to distinguish them from stations. Then,  $Z(B)$  can be approximated by the average of the values of process evaluated at the sampling points within  $B$ :

$$Z(B) \approx \frac{1}{L} \sum_{j=1}^L Z(\mathbf{s}_{j,B}). \quad (8)$$

### 3.1 Bayesian Melding in Stationary Case

The Bayesian paradigm primarily seeks the posterior distribution of all the unknowns given the data as a description of their uncertainty. These unknowns include the random error variances  $\sigma_e^2, \sigma_\delta^2$ , coefficient vector  $\boldsymbol{\beta}$ , the true underlying process  $\mathbf{Z}$ , calibration parameters  $a, b$  and covariance parameter vector  $\boldsymbol{\theta}$ . We use vector  $\mathbf{Z}$  to stand for realizations of the true underlying process at the stations and sampling points within grid cells. If we have  $n$  stations and  $m$  grid cells, the dimension of  $\mathbf{Z}$  is  $n + m \times L$ . The dimensions of the measurements  $\hat{\mathbf{Z}}$  and modeling output  $\tilde{\mathbf{Z}}$  are  $n$  and  $m$  respectively. Let  $H = \{\mathbf{X}, \hat{\mathbf{Z}}, \tilde{\mathbf{Z}}\}$  represent all the data, where  $\mathbf{X}$  is the covariate matrix in model (7).

By using the above notation, the joint distribution of all the unknowns and available data can be decomposed as follows:

$$\begin{aligned} & p(\hat{\mathbf{Z}}, \tilde{\mathbf{Z}}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\theta}, a, b, \sigma_e^2, \sigma_\delta^2) \\ &= p(\hat{\mathbf{Z}}|\mathbf{Z}, \sigma_e^2) p(\tilde{\mathbf{Z}}|\mathbf{Z}, a, b, \sigma_\delta^2) p(\mathbf{Z}|\boldsymbol{\beta}, \boldsymbol{\theta}) p(\sigma_e^2, \sigma_\delta^2, \boldsymbol{\beta}, \boldsymbol{\theta}) \\ &= \Phi_{\boldsymbol{\Sigma}_1}(\hat{\mathbf{Z}} - A_0\mathbf{Z}) \Phi_{\boldsymbol{\Sigma}_2}(\tilde{\mathbf{Z}} - a - bA_1\mathbf{Z}) \Phi_{\boldsymbol{\Sigma}_3}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}) p(\sigma_e^2) p(\sigma_\delta^2) p(\boldsymbol{\beta}) p(\boldsymbol{\theta}). \end{aligned} \quad (9)$$

Note that given  $\mathbf{Z}$ ,  $\hat{\mathbf{Z}}$  and  $\tilde{\mathbf{Z}}$  are independent. In (9),  $\Phi_{\boldsymbol{\Sigma}}(\boldsymbol{\mu})$  stands for the multivariate normal density with mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ . We take the components of  $(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_e^2, \sigma_\delta^2)$  to have independent prior distributions, that is,  $p(\sigma_e^2, \sigma_\delta^2, \boldsymbol{\beta}, \boldsymbol{\theta}) =$

$p(\sigma_e^2)p(\sigma_\delta^2)p(\boldsymbol{\beta})p(\boldsymbol{\theta})$ . The matrices  $A_0$  and  $A_1$  are

$$A_0 = \begin{pmatrix} 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & 0 & \dots & 0 \end{pmatrix}_{n \times (mL+n)} ;$$

and

$$A_1 = \begin{pmatrix} 0 & \dots & 0 & \frac{1}{L} & \dots & \frac{1}{L} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & \frac{1}{L} & \dots & \frac{1}{L} \end{pmatrix}_{m \times (mL+n)} .$$

In matrix  $A_0$  the first  $n$  columns form an identity matrix and all other elements are all zero. In matrix  $A_1$ , the elements in row  $i$  are  $\frac{1}{L}$  from column  $n + 1 + (i - 1) \times L$  to  $n + 1 + i \times L$  and all other elements are all zero. We get (9) by using the approximation (8).  $\boldsymbol{\Sigma}_1 = \sigma_e^2 \mathbf{I}$  is the covariance matrix of measurement error vector  $\mathbf{e} = [e(\mathbf{s}_1), \dots, e(\mathbf{s}_n)]^\top$ ,  $\boldsymbol{\Sigma}_2 = \sigma_\delta^2 \mathbf{I}$ , the covariance matrix of  $\boldsymbol{\delta} = [\delta(B_1), \dots, \delta(B_n)]^\top$  and  $\boldsymbol{\Sigma}_3 = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ , the covariance matrix of  $\mathbf{Z}$  while  $\mathbf{I}$  is the identity matrix.

The density of the joint conditional distribution for  $(\hat{\mathbf{Z}}, \tilde{\mathbf{Z}}, \mathbf{Z} | \boldsymbol{\beta}, \boldsymbol{\theta}, a, b, \sigma_e^2, \sigma_\delta^2)$  is

$$\begin{aligned} & p(\hat{\mathbf{Z}}, \tilde{\mathbf{Z}}, \mathbf{Z} | \boldsymbol{\beta}, \boldsymbol{\theta}, a, b, \sigma_e^2, \sigma_\delta^2) \\ & \propto \exp \left\{ -\frac{1}{2} \left[ \left( \hat{\mathbf{Z}} - A_0 \mathbf{Z} \right)^\top \boldsymbol{\Sigma}_1^{-1} \left( \hat{\mathbf{Z}} - A_0 \mathbf{Z} \right) \right. \right. \\ & + \left( \tilde{\mathbf{Z}} - a - b A_1 \mathbf{Z} \right)^\top \boldsymbol{\Sigma}_2^{-1} \left( \tilde{\mathbf{Z}} - a - b A_1 \mathbf{Z} \right) \\ & \left. \left. + \left( \mathbf{Z} - \boldsymbol{\mu} \right)^\top \boldsymbol{\Sigma}_3^{-1} \left( \mathbf{Z} - \boldsymbol{\mu} \right) \right] \right\} \\ & = \exp \left\{ -\frac{1}{2} \left[ -\mathbf{Z}^\top \left( A_0^\top \boldsymbol{\Sigma}_1^{-1} \hat{\mathbf{Z}} + b A_1^\top \boldsymbol{\Sigma}_2^{-1} (\tilde{\mathbf{Z}} - a) + \boldsymbol{\Sigma}_3^{-1} \boldsymbol{\mu} \right) \right. \right. \\ & - \left( \hat{\mathbf{Z}}^\top \boldsymbol{\Sigma}_1^{-1} A_0 + b (\tilde{\mathbf{Z}} - a) \boldsymbol{\Sigma}_2^{-1} A_1 + \boldsymbol{\mu}^\top \boldsymbol{\Sigma}_3^{-1} \right) \mathbf{Z} \\ & \left. \left. + \mathbf{Z}^\top \left( A_0^\top \boldsymbol{\Sigma}_1^{-1} A_0 + b^2 A_1^\top \boldsymbol{\Sigma}_2^{-1} A_1 + \boldsymbol{\Sigma}_3^{-1} \right) \mathbf{Z} \right] \right\} + C, \end{aligned} \quad (10)$$

where  $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$  is the mean vector of  $\mathbf{Z}$  and  $C$  has the other terms without  $\mathbf{Z}$ .

As is well known, the normal prior distribution is conjugate to the normal sampling distribution. The full conditional distribution of  $\mathbf{Z}$  conditional on all the other unknown parameters and available data is also a normal with mean vector  $\tilde{\boldsymbol{\mu}}$  and variance matrix  $\tilde{\boldsymbol{\Sigma}}$ . So, the full conditional distribution of  $\mathbf{Z}$  must be

$$p(\mathbf{Z}|\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_e^2, \sigma_\delta^2, a, b, H) \propto \exp\left\{-\frac{1}{2}\left[\mathbf{Z}^T \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{Z} - \tilde{\boldsymbol{\mu}}^T \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{Z}\right]\right\}. \quad (11)$$

But what are  $\tilde{\boldsymbol{\mu}}$  and  $\tilde{\boldsymbol{\Sigma}}$ ? By matching the terms containing  $\mathbf{Z}$  and  $\mathbf{Z}^T$  in (11) and (10), we identify the required  $\tilde{\boldsymbol{\mu}}$  and  $\tilde{\boldsymbol{\Sigma}}$  for the full conditional distribution of  $\mathbf{Z}$  as the following:

$$\begin{aligned} \mathbf{Z} | (\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_e^2, \sigma_\delta^2, a, b, H) &\sim \text{MVN}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}); \\ \tilde{\boldsymbol{\Sigma}}^{-1} &= (A_0^T \boldsymbol{\Sigma}_1^{-1} A_0 + b^2 A_1^T \boldsymbol{\Sigma}_2^{-1} A_1 + \boldsymbol{\Sigma}_3^{-1}); \\ \tilde{\boldsymbol{\mu}} &= \tilde{\boldsymbol{\Sigma}} \left( A_0^T \boldsymbol{\Sigma}_1^{-1} \hat{\mathbf{Z}} + b A_1^T \boldsymbol{\Sigma}_2^{-1} (\tilde{\mathbf{Z}} - a) + \boldsymbol{\Sigma}_3^{-1} \boldsymbol{\mu} \right). \end{aligned} \quad (12)$$

If the prior for  $\boldsymbol{\beta}$  were  $p(\boldsymbol{\beta}) \sim \text{MVN}(\boldsymbol{\beta}_0, \mathbf{F})$ , then we could claim the full conditional distribution of  $\boldsymbol{\beta}$  would be

$$\boldsymbol{\beta} | (\boldsymbol{\theta}, \mathbf{Z}, \text{other parameters}) \sim \text{NVN}(\mathbf{B}\mathbf{b}, \mathbf{B}), \quad (13)$$

where  $\mathbf{B}^{-1} = \mathbf{X}^T \boldsymbol{\Sigma}_3^{-1} \mathbf{X} + \mathbf{F}^{-1}$  and  $\mathbf{b} = \mathbf{X}^T \boldsymbol{\Sigma}_3^{-1} \mathbf{Z} + \mathbf{F}^{-1} \boldsymbol{\beta}_0$ . The proof of that claim now follows.

Given  $\boldsymbol{\Sigma}_3^{-1}$  and  $\mathbf{Z}$ ,  $\boldsymbol{\beta}$  is independent of other parameters. So the full conditional distribution of  $\boldsymbol{\beta}$  is

$$\begin{aligned} p(\boldsymbol{\beta}|\mathbf{Z}, \boldsymbol{\Sigma}_3) &\propto p(\mathbf{Z}|\boldsymbol{\beta}, \boldsymbol{\Sigma}_3^{-1})p(\boldsymbol{\beta}) \\ &\propto \exp\left\{-\frac{1}{2}\left[(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}_3^{-1} (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}) + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{F}^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right]\right\}. \end{aligned}$$

We can find the mean and variance of  $p(\boldsymbol{\beta}|\mathbf{Z}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$  by an approach of Lindley and Smith

(1972); Smith (1973).

For the exponential covariance function, the two components  $\sigma$  and  $\rho$  of  $\boldsymbol{\theta}$  are independent with inverse gamma and gamma prior distributions respectively. The inverse gamma distribution has the density function,

$$p(x; \alpha, \gamma) = \frac{1}{\Gamma(\alpha)\gamma^\alpha} \frac{1}{x^{\alpha+1}} e^{-1/(\gamma x)},$$

$\alpha > 0$  and  $\gamma > 0$  being the shape and scale parameters, respectively.

The full conditional posterior density function of  $\boldsymbol{\theta}$ , given the other variables and the data is

$$p(\boldsymbol{\theta}|\boldsymbol{\beta}, \mathbf{Z}) \propto p(\boldsymbol{\theta})|\boldsymbol{\Sigma}_3|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_3^{-1}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})\right], \quad (14)$$

where  $p(\boldsymbol{\theta})$  is the prior density for  $\boldsymbol{\theta}$ . In general, we have  $\boldsymbol{\Sigma}_3 = \sigma g(\rho, \mathbf{C})$ , where  $\mathbf{C}$  is the Euclidean distance matrix between the stations and sampling points and  $g(\cdot)$  is the spatial correlation function. Then the full conditional distribution of  $\boldsymbol{\theta}$  can be written as

$$p(\boldsymbol{\theta}|\boldsymbol{\beta}, \mathbf{Z}) \propto p(\sigma)p(\rho)|\sigma g(\rho, \mathbf{C})|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})^\top (\sigma g(\rho, \mathbf{C}))^{-1}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})\right]. \quad (15)$$

From (15) the conjugacy of the inverse gamma as  $\sigma$ 's prior becomes apparent. With an exponential correlation structure, the spatial correlation matrix is  $g(\rho, \mathbf{C}) = \exp(-\mathbf{C}/\rho)$  and so (15) can be written as

$$p(\boldsymbol{\theta}|\boldsymbol{\beta}, \mathbf{Z}) \propto p(\rho)\sigma^{-\alpha-1-n/2}|e^{-\frac{\mathbf{C}}{\rho}}|^{-\frac{1}{2}} \exp\left[-\frac{1}{\sigma}\left(\frac{1}{\gamma} + \frac{1}{2}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})^\top (e^{-\frac{\mathbf{C}}{\rho}})^{-1}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})\right)\right].$$

Thus the full conditional distribution of  $\sigma$  also has an inverse gamma with parameter

$$\begin{aligned} \tilde{\alpha} &= \alpha + n/2 \\ \tilde{\gamma} &= \left(\frac{1}{\gamma} + \frac{1}{2}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})^\top (e^{-\frac{\mathbf{C}}{\rho}})^{-1}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})\right)^{-1}. \end{aligned}$$

However  $\rho$  does not have a standard full conditional distribution, forcing us to use Metropolis-Hasting algorithm proposed by Hasting (1970) to sample  $\rho$  from (15) with  $\sigma$  fixed.

If  $\sigma_e^2$  and  $\sigma_\delta^2$  have inverse gamma priors with parameters (shape= $\alpha_1$ ,scale= $1/\lambda_1$ ) and (shape= $\alpha_2$ ,scale= $1/\lambda_2$ ) respectively, the full conditional distributions of  $\sigma_e^2$  and  $\sigma_\delta^2$  become

$$\begin{aligned} \sigma_\delta^2 | (a, b, \tilde{\mathbf{Z}}, \mathbf{Z}) &\sim \text{IG} \left( \text{shape} = \alpha_1 + m/2, \text{scale} = (\lambda_1 + \frac{1}{2}\tilde{\lambda})^{-1} \right) \\ \text{with } \tilde{\lambda} &= (\tilde{\mathbf{Z}} - a - bA_1\mathbf{Z})^\top (\tilde{\mathbf{Z}} - a - bA_1\mathbf{Z}) \text{ and} \\ \sigma_e^2 | (\hat{\mathbf{Z}}, \mathbf{Z}_2) &\sim \text{IG} \left( \text{shape} = \alpha_2 + n/2, \text{scale} = (\lambda_2 + \frac{1}{2}\gamma)^{-1} \right) \\ \text{with } \gamma &= (\hat{\mathbf{Z}} - A_0\mathbf{Z})^\top (\hat{\mathbf{Z}} - A_0\mathbf{Z}). \end{aligned} \tag{16}$$

Letting the prior for  $a, b$  be

$$\begin{pmatrix} a \\ b \end{pmatrix} \sim \text{MVN}(\bar{\boldsymbol{\beta}}, \bar{\mathbf{F}})$$

yields the full conditional distribution of the calibration parameters  $a, b$  as

$$\begin{pmatrix} a \\ b \end{pmatrix} | \tilde{\mathbf{Z}}, A_1\mathbf{Z}, \boldsymbol{\Sigma}_2 \sim \text{MVN}(\bar{\mathbf{B}}\bar{\mathbf{C}}, \bar{\mathbf{B}}),$$

where

$$\bar{\mathbf{B}} = \begin{pmatrix} \mathbf{1}^\top \\ (A_1\mathbf{Z})^\top \end{pmatrix} (\boldsymbol{\Sigma}_2)^{-1} \begin{pmatrix} \mathbf{1}, & A_1\mathbf{Z} \end{pmatrix} + \bar{\mathbf{F}}^{-1}$$

and

$$\bar{\mathbf{C}} = \begin{pmatrix} \mathbf{1}^\top \\ (A_1\mathbf{Z})^\top \end{pmatrix} (\boldsymbol{\Sigma}_2)^{-1} \begin{pmatrix} \mathbf{1}, & \tilde{\mathbf{Z}} \end{pmatrix} + \bar{\mathbf{F}}^{-1}\bar{\boldsymbol{\beta}}.$$

We define  $\mathbf{1}$  to be an column vector having the same dimension as  $\tilde{\mathbf{Z}}$ .

### 3.2 Bayesian Melding in the Non-stationary Case

The previous subsection presents the full conditional distributions of all the parameters for a stationary true underlying process distribution. The full conditional distributions are the same for the non-stationary case except for the spatial covariance parameters ( $\sigma$  and  $\rho$ ). In Model (6), the priors for the “main-effects” ( $\alpha$ ,  $r_i$ ,  $c_j$  and  $\alpha'$ ,  $r'_i$ ,  $c'_j$ ) are independent multivariate normal distributions with means  $\boldsymbol{\mu}_\sigma$  and  $\boldsymbol{\mu}_\rho$ .

Fuentes and Smith (2001) propose the non-stationary Model (6) without giving much detail about fitting the model. In this subsection, we derive the full conditional distribution of  $\boldsymbol{\sigma}$ ,  $\boldsymbol{\mu}_\sigma$ ,  $\tau_\sigma$  and  $\eta_\sigma$ . For that purpose we let  $\mathbf{Y}$  be the design matrix in Model (6). The parameters associated with the *sill* include  $\boldsymbol{\sigma}$ ,  $\boldsymbol{\mu}_\sigma$ ,  $\tau_\sigma$  and  $\eta_\sigma$ , while those associated with the *range* include  $\boldsymbol{\rho}$ ,  $\boldsymbol{\mu}_\rho$ ,  $\tau_\rho$  and  $\eta_\rho$ . Because the full conditional distribution of the parameters associated with the *sill* are analogous to those associated with the *ranges*, we only give the derivation of the full conditional distributions of parameters associated with the former. In that derivation,  $H$  represents all the other parameters as well as the available data. The full conditional distribution of  $\boldsymbol{\sigma}$  is the following:

$$p(\boldsymbol{\sigma}|H) \propto |\boldsymbol{\Sigma}(\boldsymbol{\sigma}, \boldsymbol{\rho})|^{-\frac{1}{2}} |\boldsymbol{\Sigma}(\tau_\sigma, \eta_\sigma)|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} [(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}(\boldsymbol{\sigma}, \boldsymbol{\rho})^{-1} (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})]\right\} \\ \exp\left\{-\frac{1}{2} [(\boldsymbol{\sigma} - \mathbf{Y}\boldsymbol{\mu}_\sigma)^\top \boldsymbol{\Sigma}(\tau_\sigma, \eta_\sigma)^{-1} (\boldsymbol{\sigma} - \mathbf{Y}\boldsymbol{\mu}_\sigma)]\right\},$$

where  $\boldsymbol{\Sigma}(\boldsymbol{\sigma}, \boldsymbol{\rho})$ , the spatial covariance matrix of  $\mathbf{Z}$  and  $\boldsymbol{\Sigma}(\tau_\sigma, \eta_\sigma)$  is the spatial covariance of the vector  $\boldsymbol{\sigma}$  for the latent stationary processes. In the non-stationary case, the full conditional distribution of  $\boldsymbol{\sigma}$  is no longer a standard distribution, necessitating use of the Metropolis-Hasting algorithm to update  $\boldsymbol{\sigma}$  as a block.

Since  $\boldsymbol{\mu}_\sigma$  has a normal prior with mean  $\boldsymbol{\beta}_\sigma$  and variance matrix  $\mathbf{F}_\sigma$ , its full conditional

distribution is

$$\begin{aligned}\boldsymbol{\mu}_\sigma | \boldsymbol{\sigma} &\sim \text{MVN}(Bb, B) \\ B^{-1} &= \mathbf{Y}^\top \boldsymbol{\Sigma}(\tau_\sigma, \eta_\sigma)^{-1} \mathbf{Y} + \mathbf{F}_\sigma^{-1} \\ b &= \mathbf{Y}^\top \boldsymbol{\Sigma}(\tau_\sigma, \eta_\sigma)^{-1} \boldsymbol{\sigma} + \mathbf{F}_\sigma^{-1} \boldsymbol{\beta}_\sigma.\end{aligned}$$

The full conditional distribution of  $\tau_\sigma$  is

$$p(\tau_\sigma | \mathbf{Z}, \eta_\sigma) \propto p(\tau_\sigma, \eta_\sigma) |\boldsymbol{\Sigma}(\tau_\sigma, \eta_\sigma)|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\boldsymbol{\sigma} - \mathbf{Y} \boldsymbol{\mu}_\sigma)^\top \boldsymbol{\Sigma}(\tau_\sigma, \eta_\sigma)^{-1} (\boldsymbol{\sigma} - \mathbf{Y} \boldsymbol{\mu}_\sigma) \right].$$

Our Gibbs sampling algorithm must include Metropolis-Hasting steps to sample from the full conditional distributions whenever they cannot be specified in a closed form.

### 3.3 MCMC Algorithm

To sample from the posterior distribution of the parameters in the melding Model (7), we use the Gibbs sampling algorithm proposed by Gelfand and Smith (1990) as implemented by Fuentes and Raftery (2005). First, we choose some arbitrary initial values for  $\boldsymbol{\theta}$ ,  $\boldsymbol{\beta}$  as  $\boldsymbol{\theta}^{(1)}$ ,  $\boldsymbol{\beta}^{(1)}$ . Then, the Gibbs sampling is implemented in the following three stages.

- **Stage 1:** Given all the other parameters and the available data, realizations of the true underlying process  $\{\mathbf{Z}(\mathbf{s})\}$  are updated at  $n$  stations and  $L$  sampling points within each of the  $m$  grid cells. In this stage, a random sample of  $\mathbf{Z}$  is generated by using (12).
- **Stage 2:** First  $\boldsymbol{\theta} = (\sigma, \rho)$  is updated given  $\boldsymbol{\beta}$  and  $\mathbf{Z}$  obtained in Stage 1. Second,  $\boldsymbol{\beta}$  given  $\boldsymbol{\theta}$  and  $\mathbf{Z}$  is updated by (13). Updating  $\boldsymbol{\beta}$  and  $\sigma$  is easy because their full conditional distributions are multivariate normal and inverse gamma respectively. However, the full conditional distribution of  $\rho$  does not have a closed form, so we have to use Metropolis-Hasting algorithm to update it.

- **Stage 3:** In this stage,  $\sigma_e^2, \sigma_\delta^2$  and  $a, b$  are easily updated given all the other parameters since their full conditional distributions are either normal or inverse gamma. So, it is easy to update them.

The above Gibbs sampling algorithm is nearly identical for the stationary and non-stationary cases, the only difference being in the updating of  $\theta$  in Stage 2.

### 3.4 Spatial Prediction

This subsection indicates how melding can be used to predict realizations of the underlying process at unmonitored sites using the available data.

Denote by  $\mathbf{Z}_u$ , realizations of the true underlying process at the unmonitored sites of interest. Finding the interpolation procedure entails finding the posterior distribution of  $\mathbf{Z}_u|\mathbf{Z}_g$ ,  $\mathbf{Z}_g$  being realizations of the true underlying process at monitoring stations and sampling points within grid cells. Observe that  $E(\mathbf{Z}_g) = \boldsymbol{\mu}_g$  and  $E(\mathbf{Z}_u) = \boldsymbol{\mu}_u$ . Thus the conditional mean and variance of  $\mathbf{Z}_u|\mathbf{Z}_g, \boldsymbol{\beta}, \boldsymbol{\theta}$  are

$$\begin{aligned} E(\mathbf{Z}_u|\mathbf{Z}_g, \boldsymbol{\beta}, \boldsymbol{\theta}) &= \boldsymbol{\mu}_u + \boldsymbol{\Sigma}_{ug}\boldsymbol{\Sigma}_g^{-1}(\mathbf{Z}_g - \boldsymbol{\mu}_g), \text{ and} \\ \text{Var}(\mathbf{Z}_u|\mathbf{Z}_g, \boldsymbol{\beta}, \boldsymbol{\theta}) &= \boldsymbol{\Sigma}_u - \boldsymbol{\Sigma}_{ug}\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Sigma}_{gu}. \end{aligned}$$

The covariance matrix of  $\mathbf{Z}_u$  is  $\boldsymbol{\Sigma}_u$  and the covariance matrix between  $\mathbf{Z}_u$  and  $\mathbf{Z}_g$  is  $\boldsymbol{\Sigma}_{ug}$ . Let  $\boldsymbol{\Sigma}_{gu}$  denote the transpose of  $\boldsymbol{\Sigma}_{ug}$ . In the melding model (7), we have  $\hat{Z}(\mathbf{s}) = Z(\mathbf{s}) + e(\mathbf{s})$ . So, the conditional mean and variance of  $\hat{\mathbf{Z}}_u|\mathbf{Z}_g, \boldsymbol{\beta}, \boldsymbol{\theta}$  are

$$\begin{aligned} E(\hat{\mathbf{Z}}_u|\mathbf{Z}_g, \boldsymbol{\beta}, \boldsymbol{\theta}) &= \boldsymbol{\mu}_u + \boldsymbol{\Sigma}_{ug}\boldsymbol{\Sigma}_g^{-1}(\mathbf{Z}_g - \boldsymbol{\mu}_g), \text{ and} \\ \text{Var}(\hat{\mathbf{Z}}_u|\mathbf{Z}_g, \boldsymbol{\beta}, \boldsymbol{\theta}) &= \boldsymbol{\Sigma}_u - \boldsymbol{\Sigma}_{ug}\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Sigma}_{gu} + \sigma_e^2\mathbf{I}, \end{aligned}$$

in which the dimension of the identity matrix  $\mathbf{I}$  is the number of stations to be predicted.

The marginal distribution of  $\hat{\mathbf{Z}}_u | \hat{\mathbf{Z}}, \tilde{\mathbf{Z}}$  is

$$p(\hat{\mathbf{Z}}_u | \hat{\mathbf{Z}}, \tilde{\mathbf{Z}}) = \int p(\mathbf{Z}_u | \phi) p(\phi | \hat{\mathbf{Z}}, \tilde{\mathbf{Z}}) d\phi, \quad (17)$$

in which  $\phi$  stands for all the other parameters such as  $\mathbf{Z}_g$ ,  $\beta$  and  $\theta$  that can be approximated by

$$p(\hat{\mathbf{Z}}_u | \mathbf{Z}_g) \approx \frac{1}{n} \sum_{i=1}^n p(\hat{\mathbf{Z}}_u | \mathbf{Z}_g, \phi^i),$$

where  $\phi^i (i = 1, \dots, n)$  is the  $i$ -th MCMC sample of all the parameters.

### 3.5 Extension to Ensembles

Previous sections consider the problem of combining measurements from monitoring sites made with a single instrument with output from just one deterministic model. In fact, melding can be extended to combine data from ensembles of measuring instruments and of deterministic models. For simplicity we consider the case of just a single measuring instrument and multiple deterministic models although the extension to incorporate multiple measuring instruments will be obvious.

To extend Model (7), suppose  $\tilde{\mathbf{Z}}_1, \dots, \tilde{\mathbf{Z}}_p$  are output from an ensemble of  $p$  deterministic models. To ensure a non-singular spatial covariance matrix, suppose no overlap between the monitoring sites and sampling points within each grid cell for all the different deterministic models. For simplicity, we assume the number of sampling points within each grid cell is the same,  $L$ , for all the deterministic models although extension to differing numbers is straightforward. We also assume each deterministic model has modeling output on  $m$  grid cells. Then the dimension of  $\mathbf{Z}$  is  $n + p \times m \times L$ . As in (9), the joint posterior distribution

can be decomposed as follows:

$$\begin{aligned}
& p(\hat{\mathbf{Z}}, \tilde{\mathbf{Z}}_1, \dots, \tilde{\mathbf{Z}}_p, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\theta}, a_1, \dots, a_p, b_1, \dots, b_p, \sigma_e^2, \sigma_{\delta,1}^2, \dots, \sigma_{\delta,p}^2) \\
&= p(\hat{\mathbf{Z}}|\mathbf{Z}, \sigma_e^2) \prod_{i=1}^p p(\tilde{\mathbf{Z}}_i|\mathbf{Z}, a_i, b_i, \sigma_{\delta,i}^2) p(\mathbf{Z}|\boldsymbol{\beta}, \boldsymbol{\theta}) p(\sigma_e^2, \sigma_{\delta,1}^2, \dots, \sigma_{\delta,p}^2, \boldsymbol{\beta}, \boldsymbol{\theta}) \\
&= \Phi_{\boldsymbol{\Sigma}_0}(\hat{\mathbf{Z}} - \mathbf{Z}_0) \prod_{i=1}^p \Phi_{\boldsymbol{\Sigma}_i}(\tilde{\mathbf{Z}}_i - a_i - b_i A_i \mathbf{Z}) p(\sigma_{\delta,i}^2) \Phi_{\boldsymbol{\Sigma}(\boldsymbol{\theta})}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}) p(\sigma_e^2) p(\boldsymbol{\theta}) p(\boldsymbol{\beta}).
\end{aligned}$$

Deterministic model  $i$  has calibration parameters  $a_i$  and  $b_i$ , modeling output error variance parameter  $\sigma_{\delta,i}^2$ . The covariance matrix of measurement error  $\mathbf{e}$  is  $\boldsymbol{\Sigma}_0 = \sigma_e^2 \mathbf{I}$  and the covariance matrix of the modeling output error vector  $\boldsymbol{\delta}_i$  of the  $i$ -th deterministic model is  $\boldsymbol{\Sigma}_i = \sigma_{\delta,i}^2 \mathbf{I}$ . The dimension of matrix  $A_0$  is  $n \times (n + p \times m \times L)$  and that of matrix  $A_i$  ( $i = 1, \dots, p$ ) is  $m \times (n + p \times m \times L)$ . The first  $n$  columns of matrix  $A_0$  form an identity matrix and the remaining elements are all 0. Row  $j$  of matrix  $A_i$  ( $i = 1, \dots, p$ ) has elements  $\frac{1}{L}$  from column  $n + 1 + (i - 1) \times m \times L + (j - 1) \times L$  to  $n + 1 + (i - 1) \times m \times L + j \times L$  and all other elements are zero.

The density of the conditional distribution of  $(\hat{\mathbf{Z}}, \tilde{\mathbf{Z}}_1, \dots, \tilde{\mathbf{Z}}_p, \mathbf{Z})$  given  $(\boldsymbol{\beta}, \boldsymbol{\theta}, a_1, \dots, a_p, b_1, \dots, b_p, \sigma_e^2, \sigma_{\delta,1}^2, \dots, \sigma_{\delta,p}^2)$  is

$$\begin{aligned}
& p(\hat{\mathbf{Z}}, \tilde{\mathbf{Z}}_1, \dots, \tilde{\mathbf{Z}}_p, \mathbf{Z}) | \boldsymbol{\beta}, \boldsymbol{\theta}, a_1, \dots, a_p, b_1, \dots, b_p, \sigma_e^2, \sigma_{\delta,1}^2, \dots, \sigma_{\delta,p}^2 \\
& \propto \exp \left\{ -\frac{1}{2} \left[ (\hat{\mathbf{Z}} - A_0 \mathbf{Z})^\top \boldsymbol{\Sigma}_0^{-1} (\hat{\mathbf{Z}} - A_0 \mathbf{Z}) \right. \right. \\
& + \sum_{i=1}^p \left. \left. (\tilde{\mathbf{Z}}_i - a_i - b_i A_i \mathbf{Z})^\top \boldsymbol{\Sigma}_i^{-1} (\tilde{\mathbf{Z}}_i - a_i - b_i A_i \mathbf{Z}) \right. \right. \\
& + \left. \left. (\mathbf{Z} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \boldsymbol{\mu}) \right] \right\} \\
&= \exp \left\{ -\frac{1}{2} \left[ -\mathbf{Z}^\top \left( A_0^\top \boldsymbol{\Sigma}_0^{-1} \hat{\mathbf{Z}} + \sum_{i=1}^p b_i A_i^\top \boldsymbol{\Sigma}_i^{-1} (\tilde{\mathbf{Z}}_i - a_i) + \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right) \right. \right. \\
& - \left. \left. \left( \hat{\mathbf{Z}}^\top \boldsymbol{\Sigma}_0^{-1} A_0 + \sum_{i=1}^p b_i (\tilde{\mathbf{Z}}_i - a_i) \boldsymbol{\Sigma}_i^{-1} A_i + \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \right) \mathbf{Z} \right. \right. \\
& + \left. \left. \mathbf{Z}^\top \left( A_0^\top \boldsymbol{\Sigma}_0^{-1} A_0 + \sum_{i=1}^p b_i^2 A_i^\top \boldsymbol{\Sigma}_i^{-1} A_i + \boldsymbol{\Sigma}^{-1} \right) \mathbf{Z} \right] \right\} + C, \tag{18}
\end{aligned}$$

$\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  being the mean and variance of  $\mathbf{Z}$ ,  $C$ , the other terms without  $\mathbf{Z}$ .

Like (12), the full conditional distribution of  $\mathbf{Z}$  is the following.

$$\begin{aligned} \mathbf{Z} \mid \text{other parameters and all data} &\sim \text{MVN}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}) \\ \tilde{\boldsymbol{\Sigma}}^{-1} &= (A_0^\top \boldsymbol{\Sigma}_0^{-1} A_0 + \sum_{i=1}^p b_i^2 A_i^\top \boldsymbol{\Sigma}_i^{-1} A_i + \boldsymbol{\Sigma}^{-1}) \\ \tilde{\boldsymbol{\mu}} &= \tilde{\boldsymbol{\Sigma}} \left( A_0^\top \boldsymbol{\Sigma}_0^{-1} \hat{\mathbf{Z}} + \sum_{i=1}^p b_i A_i^\top \boldsymbol{\Sigma}_i^{-1} (\tilde{\mathbf{Z}}_i - a_i) + \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right). \end{aligned}$$

The prior distribution of  $(a_i, b_i)$ ,  $i = 1, \dots, p$ , is

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} \sim \text{MVN}(\bar{\boldsymbol{\beta}}_i, \bar{\mathbf{F}}).$$

Their full conditional distribution is

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} \mid \tilde{\mathbf{Z}}_i, \mathbf{Z}, \boldsymbol{\Sigma}_i \sim \text{MVN}(\bar{\mathbf{B}}\bar{\mathbf{C}}, \bar{\mathbf{B}}),$$

where

$$\bar{\mathbf{B}} = \begin{pmatrix} \mathbf{1}^\top \\ (A_i \mathbf{Z})^\top \end{pmatrix} (\boldsymbol{\Sigma}_i)^{-1} \begin{pmatrix} \mathbf{1}, & A_i \mathbf{Z} \end{pmatrix} + \bar{\mathbf{F}}_i^{-1}$$

and

$$\bar{\mathbf{C}} = \begin{pmatrix} \mathbf{1}^\top \\ (A_i \mathbf{Z})^\top \end{pmatrix} (\boldsymbol{\Sigma}_i)^{-1} \begin{pmatrix} \mathbf{1}, & \tilde{\mathbf{Z}}_i \end{pmatrix} + \bar{\mathbf{F}}_i^{-1} \bar{\boldsymbol{\beta}}_i.$$

### 3.6 Reversible Jump MCMC

Previous sections assume a known degree in the polynomial representing the mean function  $\mu(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta}$ . In other words, the coefficient vector  $\boldsymbol{\beta}$  has known dimension. Green (1995) proposes the reversible jump MCMC to allow the dimension of  $\boldsymbol{\beta}$  to be an unknown

parameter. For the melding model, we let  $k$  be  $\boldsymbol{\beta}$ 's dimension and  $\boldsymbol{\eta}_k = (\boldsymbol{\beta}_k, \boldsymbol{\theta})$ . The reversible jump MCMC allows  $k$  to vary, while the dimension of all other parameters are fixed. The objective: to sample from the joint posterior distribution of  $k$  and  $\boldsymbol{\eta}_k$ , that is,  $p(k, \boldsymbol{\eta}_k | \text{data})$ . To achieve that, we first find the full joint conditional distribution of  $k$  and  $\boldsymbol{\eta}_k$ . Given  $\mathbf{Z}$ , that full conditional distribution is

$$p(k, \boldsymbol{\eta}_k | \mathbf{Z}) \propto p(\mathbf{Z}, \boldsymbol{\eta}_k, k) = p(\mathbf{Z} | k, \boldsymbol{\eta}_k) p(\boldsymbol{\eta}_k | k) p(k),$$

$p(k)$  being the prior distribution of the dimension  $k$  and  $p(\boldsymbol{\eta}_k | k)$ , the prior distribution of  $\boldsymbol{\eta}_k$  given  $k$ . Suppose initially the dimension of  $\boldsymbol{\beta}$  is  $p$  ( $p > 0$ ). Then the next iteration has two possible outcomes for  $k$ : increase the dimension by one (a “birth”) or decrease it by one (a “death”). The probability of a “birth” or “death” is  $1/2$ . The jump scheme can be quite arbitrarily but we make the above choice because we favor “birth” or “death” equally. After choosing either “birth” or “death”, we accept/reject the jump with probability  $\alpha/(1 - \alpha)$ , the specification of  $\alpha$ , being explained below. If we reject the jump, then we stay with the current dimension  $p$ ; otherwise, the dimension  $k$  will be either  $p + 1$  or  $p - 1$ . In the same iteration, the next step updates the parameters  $\boldsymbol{\eta}$  given  $k$  as in the fixed dimensional case. By using the formulas provided in Green (1995), the acceptance probabilities of “birth” and “death” are the following.

- “Birth”. To jump from the previous dimension  $k = p$  to a new dimension  $k^* = p + 1$ , we need to propose one extra coefficient  $\beta_{new}$  for one extra covariate. That extra coefficient  $\beta_{new}$  is proposed by a proposal distribution with density  $q(\cdot)$ . Because  $\beta_{new}$  is proposed independently of other coefficients the Jacobian of transforming from  $\boldsymbol{\beta}$  to  $\boldsymbol{\beta}^*$  is 1. So, the new coefficient vector is  $\boldsymbol{\beta}^* = (\boldsymbol{\beta}, \beta_{new})$  and the acceptance probability of  $k^* = p + 1$  is

$$\alpha = \min\left\{1, \frac{p(k^*)p(\boldsymbol{\beta}^* | k^*)p(\mathbf{Z} | k^*, \boldsymbol{\beta}^*)}{p(k)p(\boldsymbol{\beta} | k)p(\mathbf{Z} | k, \boldsymbol{\beta})q(\beta_{new})}\right\}.$$

- “Death”. To jump from the previous dimension  $k = p$  to a new dimension  $k^* = p - 1$ , we need to delete one coefficient from the coefficient vector. We choose to delete the last coefficient of the vector  $\boldsymbol{\beta}$ . The reason is that we arrange the covariates from lower order of the coordinates to higher order and usually the higher order of the coordinates are more likely to be insignificant than lower ones. Again, the Jacobin is 1 and the acceptance probability of  $k^* = p - 1$  is

$$\alpha = \min\left\{1, \frac{p(k^*)p(\boldsymbol{\beta}^*|k^*)p(\mathbf{Z}|k^*, \boldsymbol{\beta}^*)q(\beta_p)}{p(k)p(\boldsymbol{\beta}|k)p(\mathbf{Z}|k, \boldsymbol{\beta})}\right\}.$$

After choosing dimension  $k$ , we update all other parameters by using Gibbs sampling as in the fixed dimension case. In Subsection 4.3, we conduct a simulation to see how well the reversible jump MCMC works when incorporated into the Bayesian melding model.

The discussion above shows how to use the reversible jump MCMC algorithm to choose the dimension of  $\boldsymbol{\beta}$ . That algorithm can also be used to choose the dimension of coefficients of the additive calibration  $a(\mathbf{s})$ . If we let  $a(\mathbf{s}) = \mathbf{Y}(\mathbf{s})\boldsymbol{\beta}_a$ ,  $\mathbf{Y}(\mathbf{s})$  being the polynomial function of the coordinate at location  $\mathbf{s}$ , then we can use reversible jump MCMC to choose the dimension of coefficients  $\boldsymbol{\beta}_a$ . The detailed algorithm is very similar to the above when reversible jump MCMC is applied to choose the dimension of  $\boldsymbol{\beta}$ . The current software does not include the reversible jump MCMC to choose the dimension of coefficients  $\boldsymbol{\beta}_a$ , but we are planning to incorporate that into the software in the near future.

### 3.7 Melding Software

We write the melding program in R which is developed by R Development Core Team (2006). The melding program is online at <http://enviro.stat.ubc.ca/melding/meldingcode.zip>. The code for stationary melding model is in the directory “mcmc” and the code for stationary melding model incorporating the reversible jump MCMC is in the directory “rjmc”. We will soon upload the code for the non-stationary melding model.

The subdirectory “debug” in directory “mcmc” is used to debug the R programs. That directory includes the file “simudata.s” simulating the measurements and modeling outputs, “mcmc.s”, the main function to implement the MCMC algorithm, “update.s”, various functions to sample parameters from their full conditional distributions. The description of the main function “melding” in file “mcmc.s” is the following.

Dependence:  $R \geq 2.3.0$  and packages “MASS”.

Usage: `melding(m1,m2,nm,sam.sloc,sam.sloc1,ton,burnin,zhat,Zbtilde,degree,cov.model)`

Arguments:

- nm: number of sampling points in each grid cell of the modeling output.
- sam.sloc: coordinates of the sampling points and the monitoring stations. It should be a  $(m2 \times nm + m1) \times 2$  matrix. The first  $m2 \times nm$  rows are the coordinates of the sampling points in grid cells. The last  $m1$  rows are the coordinates of the monitoring stations.
- sam.sloc1: coordinates of the unmonitored stations where the measurements are to be predicted.
- ton: number of MCMC iterations in the Gibbs sampling.
- burnin: the “burn-in” period of the Gibbs sampling.
- zhat: the measurements vector.
- Zbtilde: the modeling outputs vector.
- degree: degree of the polynomial function  $f(\cdot)$  for the mean  $\mu(\mathbf{s}) = f(\mathbf{s})\boldsymbol{\beta}$  and the following options are allowed:

0 the mean is assumed constant across space.

1 the mean is assumed to be a first order polynomial on the coordinates:  $\mu(\mathbf{s}) = \beta_0 + \beta_1 s_1 + \beta_2 s_2$ .

2 the mean is assumed to be a second order polynomial on the coordinates:  $\mu(\mathbf{s}) = \beta_0 + \beta_1 s_1 + \beta_2 s_2 + \beta_3 s_1^2 + \beta_4 s_2^2 + \beta_5 s_1 s_2$ .

- `cov.model`: a string with the name of the correlation function. The options are one of the following three functions:

“exponential”:  $\exp(-d/rho)$ ,  $d$  is the distance and  $rho$  is the range parameter.

“Gaussian”:  $\exp(-d^2/rho)$ ,  $d$  is the distance and  $rho$  is the range parameter.

“Matern”:  $(2^{(\nu-1)}\Gamma(\nu))^{-1}(d/\rho)^\nu K_\nu(d/\rho)$ .  $\nu$  is the smoothing parameter and  $K_\nu(\cdot)$  denotes the modified Bessel function of the third kind of order  $\nu$ .

The function “`melding`” will return a list of the following objects:

- “`beta.est`”: posterior mean of the coefficient vector  $\beta$ .
- “`beta.est.sd`”: posterior standard deviation of the coefficient vector  $\beta$ .
- “`theta.est`”: posterior mean of the spatial correlation vector  $\theta$ .
- “`theta.est.sd`”: posterior standard deviation of the spatial correlation vector  $\theta$ .
- “`prediction`”: posterior mean of the spatial prediction.
- “`pred.q1`”: 5% quantile of the posterior distribution of the spatial prediction.
- “`pred.q2`”: 95% quantile of the posterior distribution of the spatial prediction.
- “`ab.est`”: posterior mean of the additive and multiplicative parameters  $a$  and  $b$ .
- “`ab.est.sd`”: posterior standard deviation of the additive and multiplicative parameters  $a$  and  $b$ .
- “`sigmae.est`”: posterior mean of the measurement error variance parameter  $\sigma_e^2$ .
- “`sigmae.est.sd`”: posterior standard deviation of the measurement error variance parameter  $\sigma_e^2$ .

- “sigmad.est”: posterior mean of the modeling output error variance parameter  $\sigma_{\delta}^2$ .
- “sigmad.est.sd”: posterior standard deviation of the modeling output error variance parameter  $\sigma_{\delta}^2$ .

The description of various functions in file “update.s” is the following.

- “updatebeta(...)”: generate MCMC sample from the full conditional distribution of  $\beta$ . The arguments of this function include
  - y: the realizations of the true underlying process  $\{\mathbf{Z}(\mathbf{s})\}$  at monitored stations and sampling points within grid cells.
  - x: the covariate matrix at monitored stations and sampling points within grid cells.
  - prior.mean: prior mean of the coefficient vector  $\beta$ .
  - prior.var.solve: inverse of the prior variance matrix of the coefficient vector  $\beta$ .
  - sigma.solve: inverse of the spatial covariance matrix of true underlying process  $\{\mathbf{Z}\}$ .
- “updatetheta(...)”: generate MCMC sample from the full conditional distribution of  $\theta$ . The arguments of this function include
  - diff: the residuals of the true underlying process.
  - theta: the values of the covariance parameters *theta* from the previous MCMC iteration.
  - Distance: the Euclidean distance matrix between all the locations including monitored stations and sampling points within grid cells.
  - n: number of monitored stations and sampling points within grid cells.
  - cov.model: one of the three possible choices for the covariance function: “exponential”, “Gaussian” and “matern”.

- “updatesigma(...)”: generate MCMC sample from the full conditional distribution of  $\sigma_{\delta}^2$  and  $\sigma_e^2$ . The arguments of this function include
  - zhat: the measurements vector.
  - Zbtilde: the modeling outputs vector.
  - nm: number of sampling points in each grid cell of the modeling output; n: number of monitored stations and sampling points within grid cells; m2: number of grid cells.
  - a: additive bias parameter  $a$ ; b: multiplicative bias parameter  $b$ .
  - A2: matrix  $A_1$  as described in Subsection 3.1.
- “updateab(...)”: generate MCMC sample from the full conditional distribution of  $a$  and  $b$ . The arguments of this function include
  - Zbtilde: same as in the function “updatesigma(...)”. ,y,ab0,fb.solve,nm,m2,A2.
  - sigmad: the output error variance parameter  $\sigma_{\delta}^2$ .
  - y: same as in the function “updatesigma(...)”.
  - n,m2,A2: same as in the function updatesigma(...).
- “updatezs(...)”: generate MCMC sample from the full conditional distribution of  $\mathbf{Z}$ . The arguments of this function include
  - X: the covariate matrix at monitored stations and sampling points within grid cells.
  - zhat, Zbtilde, nm,m2: same as in the function “updatesigma(...)”.
  - sigmad, sigmae: the output error variance parameter  $\sigma_{\delta}^2$  and measurement error variance parameter  $\sigma_e^2$ .
  - sigma3.solve: inverse of the spatial covariance matrix of true underlying process  $\{\mathbf{Z}\}$ .

Reversible Jump MCMC is used to do the variable selection in a Bayesian framework. It is functioning similar to the step-wise regression in classical framework. All the function in "update.s" are the same as in melding model described above. In the function "rjmelding(...)", we do not specify the "degree", which is estimated by the reversible jump MCMC. The function "rjmeding(...)" only returns the dimension of the coefficient  $\beta$  in the mean of underlying true process.

## 4 Simulation Studies

This section presents an extensive simulation study of melding under a variety of scenarios that assess the method while checking the R code developed. We begin with the stationary melding model, one deterministic model and one measuring instrument. Then we move on to the case with an ensemble of deterministic models. We include a simulation study that incorporates a reversible jump MCMC into the Bayesian melding model. In addition, studies are devoted to several other issues such as how the estimation of covariance parameters can be improved by better monitored site layouts and the effect of choosing bandwidth  $h$  in non-stationary melding model.

### 4.1 A single deterministic model

This simulation, which assumes a stationary true underlying process, has two purposes. First, it validates the MCMC algorithm for implementing the melding model. Second, it compares the results of prediction using melding and Kriging.

#### **Simulation settings**

In the simulation, we have 20 monitored sites and 100 un-monitored sites whose responses are to be predicted. The number of available grid cells increase from 0 to 50 through the sequence of 0, 2, 10, 20, 30, 50. To reduce our computational burden, we only have one sampling point within each grid cell. The coordinates of the available/unavailable stations and sampling points within grid cells are generated uniformly on  $[-5, 5] \times [-5, 5]$ . Figure 1 depicts the locations of sites (monitored and un-monitored) and sampling points. For each combination of sites and sampling points, we generate 50 independent datasets.

The data generated for the 100 unmeasured stations will not be used in the MCMC prediction but left for validation. As the number of grid cells increases, we investigate changes in prediction accuracy and coverage probability for different prediction intervals at

various credibility levels (95%, 90%, 80%, 70%, 60%, 40%).

In generating the data, we let the mean of the true underlying process be

$$E[Z(\mathbf{s})] = \mu(\mathbf{s}) = \beta_0 + \beta_1 s_1 + \beta_2 s_2, \quad (19)$$

$s_1, s_2$  being the coordinates of location  $\mathbf{s}$ . We assume  $\{Z(\mathbf{s})\}$  has an exponential covariance structure. At the same time, we give  $\beta$  a prior normal distribution with mean  $\beta_0 = (1, 1, 1)$  and diagonal covariance matrix with diagonal elements (100, 100, 100). The covariance parameters  $\sigma$  and  $\rho$  have independent prior distributions;  $\sigma$  has an inverse gamma prior distribution with shape = 2, scale = 0.2, while  $\rho$  has a gamma prior distribution with shape = 3.5, scale = 2. The calibration parameters  $a$  and  $b$  have independent normal prior distributions with means  $\mu_a = 0$ ,  $\mu_b = 1$  and variances  $\sigma_a^2 = 100$ ,  $\sigma_b^2 = 100$ . The random variables  $\sigma_\delta^2$  and  $\sigma_e^2$  have identical and independent inverse gamma distributions with shape = 3 and scale = 0.1.

Using the MCMC algorithm to sample from the joint posterior distribution of all parameters, we take 1000 Gibbs sampling iterations (the first 100 being for “burn-in”), which is enough to ensure convergence as shown in Figure 3 and Figure 4. That is fortunate since computational times are long and necessity for a longer series of iterations would have greatly limited the scope of our simulation study. The software used in our study is available online (<http://enviro.stat.ubc.ca/melding/meldingcode.zip>) to enable independent verification of our findings.

## Simulation Results

Tables 1, 2 and 3 present estimation results for the parameters. Table 4 presents the SSPE (sum of squared prediction errors) for the 100 unmeasured stations. Our results lead to the following over observations.

- The estimation of  $a$ ,  $b$  and  $\beta$  are very close to their true values.

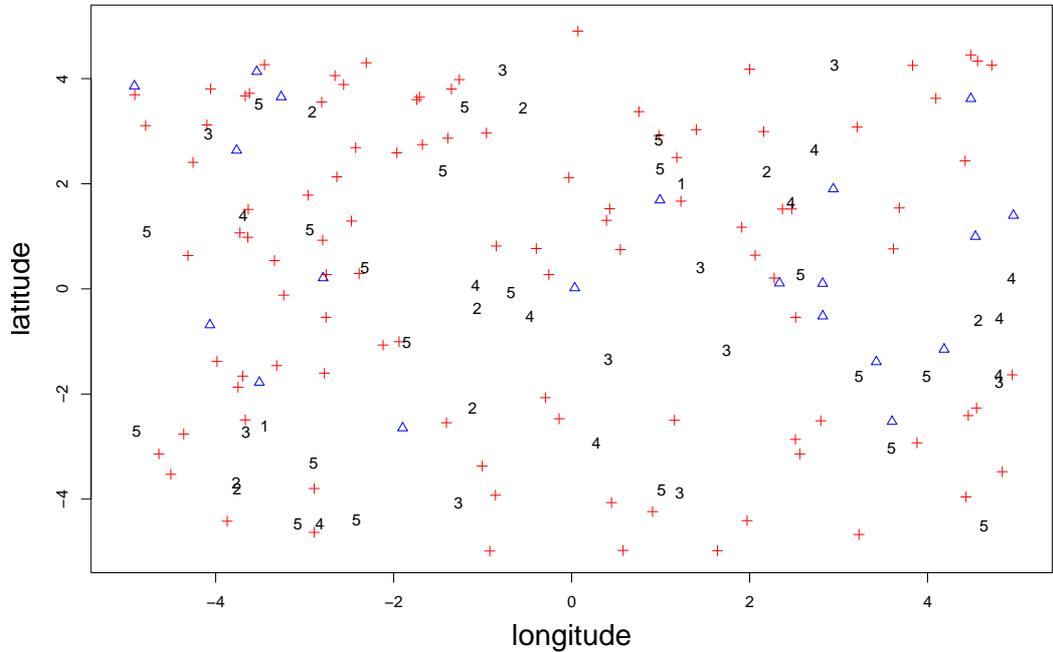


Figure 1: Locations of 120 (20 monitored, 100 to be predicted) sites and points sampled from up to 50 grid cells. Monitored sites ( $\Delta$ ), sites with responses to be predicted: +, 1: the first 2 points from grid cells, 2: 3-10 points, 3: 11-20 points, 4: 21-30 points, 5: 31-50 points.

- Based on Figure 2, which shows the true ozone values versus the melding prediction in the case of 50 grid cells, we can see the melding predictions are quite close to the true values.
- The estimates of the covariance parameters  $\theta = (\sigma, \rho)$  are reasonably good. However, the *range* parameter  $\rho$  proves much more difficult to estimate than the *sill* parameter  $\sigma$ . The posterior distribution in Figure 5 shows  $\rho$  to be widely dispersed.
- Table 4 reveals a significant variation in the SSPE from one simulated dataset to another.
- Table 4 shows that adding more grid cells decreases the average SSPE for the melding model. The average is computed over 50 independent datasets. In each, the SSPE

does not necessarily decrease in a monotonic fashion as the number of cells increases. However, variation from the non-decreasing pattern in the SSPE may just be due to the sampling variation between the generated datasets. Table 5 shows the SSPE obtained by using the true values of the parameters in the predictor. In spite of that advantage, the SSPE for even this predictor does not always decrease monotonically as the number of grid cells increases. (See hours 4, 6, and 6, for example.)

- Table 6 shows that the empirical coverage probability for the melding predictor comes close to the nominal level when we have a reasonably large number of cells (at least 10). However that is not the case when we have just 0 or 2 cells, not surprising given the paucity of data in those situations. Kriging’s coverage probability turns out to be much smaller than that for the melding predictor even when no simulated data are available from the deterministic model (so that the two methods compete “head-on”). That result would be anticipated since, as is well known, Kriging underestimates the uncertainty in its spatial predictions.
- Figures 3, 4 and 5 show the MCMC samples from posterior distributions of  $a$ ,  $b$ ,  $\beta$ ,  $\sigma$  and  $\rho$  in the case of 50 grid cells. From these three figures, we can see the Markov chain converges after a few iterations.

Table 1: Estimates of  $a$  and  $b$  with different numbers of modeling outputs available. The true values are  $a = 5.0$  and  $b = 2.5$ . Columns 2-5 are the averages of the posterior means and standard deviations. The averages are computed over the 50 dataset generated in the simulation study.

Number of cells	$\bar{a}$	$\bar{sd}$	$\bar{b}$	$\bar{sd}$
2	5.61	4.14	2.49	0.22
10	5.11	1.40	2.50	0.09
20	5.24	1.02	2.50	0.08
30	5.12	0.82	2.50	0.07
50	5.15	0.63	2.50	0.06

Table 2: Estimates of the coefficients  $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$  with different numbers of modeling outputs available. The true values are  $\beta_0 = 2.50$ ,  $\beta_1 = 2.90$  and  $\beta_2 = 3.20$ . Columns 2-7 are the averages of posterior means and standard deviations. The averages are computed over the 50 dataset generated in the simulation study.

Number of cells	$\hat{\beta}_0$	$\bar{sd}$	$\hat{\beta}_1$	$\bar{sd}$	$\hat{\beta}_2$	$\bar{sd}$
0	2.35	0.46	2.89	0.12	3.12	0.12
2	2.13	0.75	2.88	0.13	3.11	0.14
10	2.11	0.77	2.88	0.13	3.11	0.14
20	2.09	0.80	2.89	0.13	3.12	0.14
30	2.03	0.82	2.88	0.13	3.12	0.14
50	2.03	0.83	2.88	0.13	3.12	0.14

### Improving covariance parameter estimates

Table 3 shows that although the posterior mean of  $\rho$  gets closer to the true value as the number of grid cells increases from 0 to 50, the posterior standard deviation remains relatively large even with 50 grid cells. That may well be due to insufficiently many sampling points in close proximity to one another. In other words, we do not have enough information about the small scale process variability needed to accurately estimate the variogram. To explore that conjecture, we carry out a small simulation study with 20 monitoring sites and 50 sampling points in grid cells as in the previous simulation. However, in contrast to the previous case, we concentrate 25 sampling points in a very small region given by  $[-0.05, 0.05] \times [-0.05, 0.05]$ . Except for that variation, we generate these data in precisely the same way as in the previous simulation. Table 7 shows the results for the estimators of the covariance parameters. From that table, we can see that the standard error of estimator for  $\rho$  is reduced significantly compared with the result in Table 3, thereby adding support to our conjecture.

Table 3: Estimates of the covariance parameters in  $\theta = (\sigma, \rho)$  with different numbers of modeling outputs available. The true values are  $\sigma = 1.50$ ,  $\rho = 5.00$ . Columns 2-5 are the averages of posterior means and standard deviations. The averages are computed over the 50 dataset generated in the simulation study.

Number of cells	$\bar{\sigma}$	$\bar{sd}$	$\bar{\rho}$	$\bar{sd}$
0	1.42	0.26	1.36	1.46
2	1.54	0.31	2.41	2.22
10	1.52	0.35	2.98	2.01
20	1.53	0.33	3.61	2.03
30	1.54	0.33	4.41	2.17
50	1.52	0.27	5.34	2.32

## Conclusions

Our simulation results point to strengths and weaknesses in the melding approach.

### Strengths:

- The Bayesian melding model can estimate the calibration parameters of the modeling output very well, given a reasonable number of monitoring sites and grid cells.
- In general, increasing the number of grid cells improves spatial prediction accuracy at un-monitored sites.
- Estimates of the coefficients of the process mean are very good.
- The melding spatial predictor gives much more realistic estimates of the prediction uncertainty than does the classical Kriging approach.

### Weaknesses:

- Melding does not estimate the covariance parameters ( $\sigma$  and  $\rho$ ) well in terms of their standard errors unless an appreciable number of sampling points are very close to one

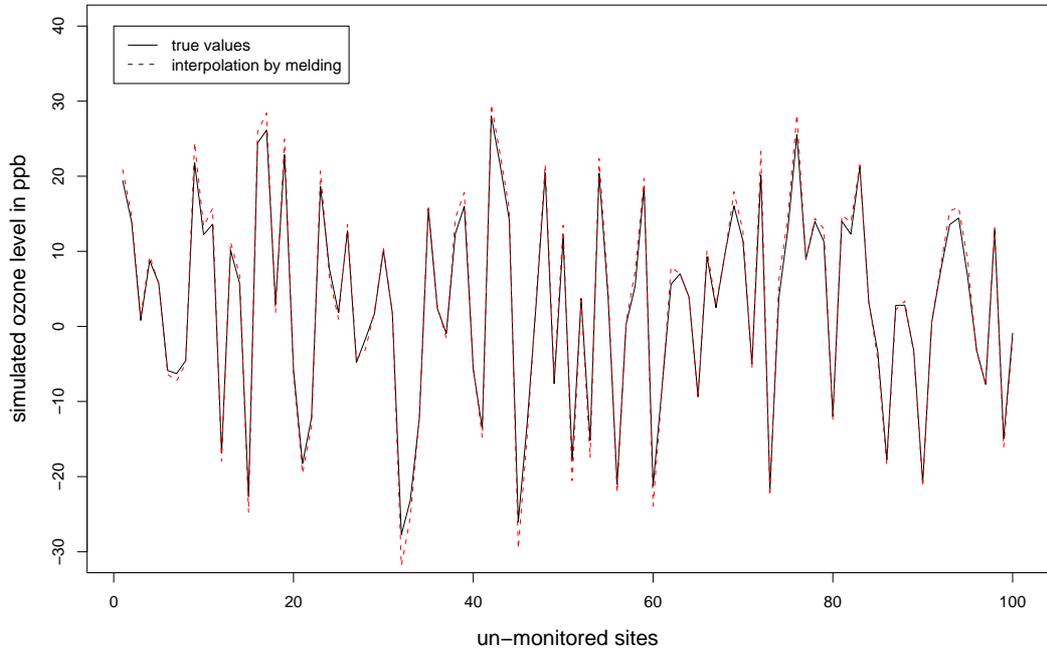


Figure 2: Spatial predictions for 100 un-monitored sites Bayesian melding in the case of 50 grid cells.

another. (As seen from the previous section, forcing a number of such points into close proximity does decrease their standard errors substantially).

- The current melding model (7) does not include temporal information. For the ozone pollution data studied by Kasibhatla and Chameides (2000), both the real measurements and modeling output are hourly data, which does have an obvious daily cycle and substantial autocorrelations. Extending the melding approach to embrace random space - time fields would be desirable, as it would enable potentially great strength to be borrowed across time as well as space.
- The computational burden imposed by melding approach limits its practicality. That burden stems from the need to invert a large dimensional spatial covariance matrices three times in each Gibbs sampling iteration. For example, 20 monitoring sites and 100 grid cells with 4 sampling points in each leads to a spatial covariance matrix with

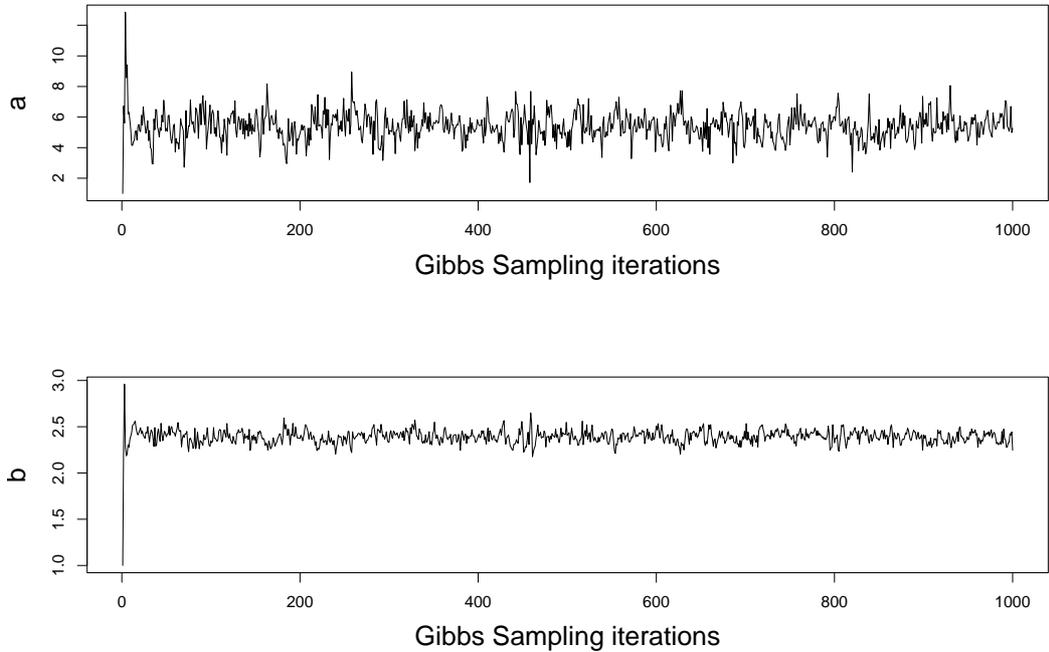


Figure 3: MCMC Gibbs samplers of the additive and multiplicative calibration parameters,  $a$  and  $b$ , respectively, in the case of 50 grid cells.

$20 + 100 \times 4 = 420$  rows and columns. MCMC does not provide a “free Bayesian lunch” as might naively be suggested by its very elegant theory.

## 4.2 Ensemble studies

### Simulation settings

This subsection presents a simulation study illustrating the use of melding to combine measurements with modeling outputs from more than one deterministic models. To make this simulation study closer to the real life case, we use locations from a dataset studied by Kasibhatla and Chameides (2000); Hogrefe et al. (2001b,a). There, measurements come from monitoring sites (“stations”) and modeling output from deterministic models. In fact, we use only a subset of these stations and grid cells in our simulation, with 50 monitored sites treated as stations and 100 un-monitored sites. For simplicity we assume only two

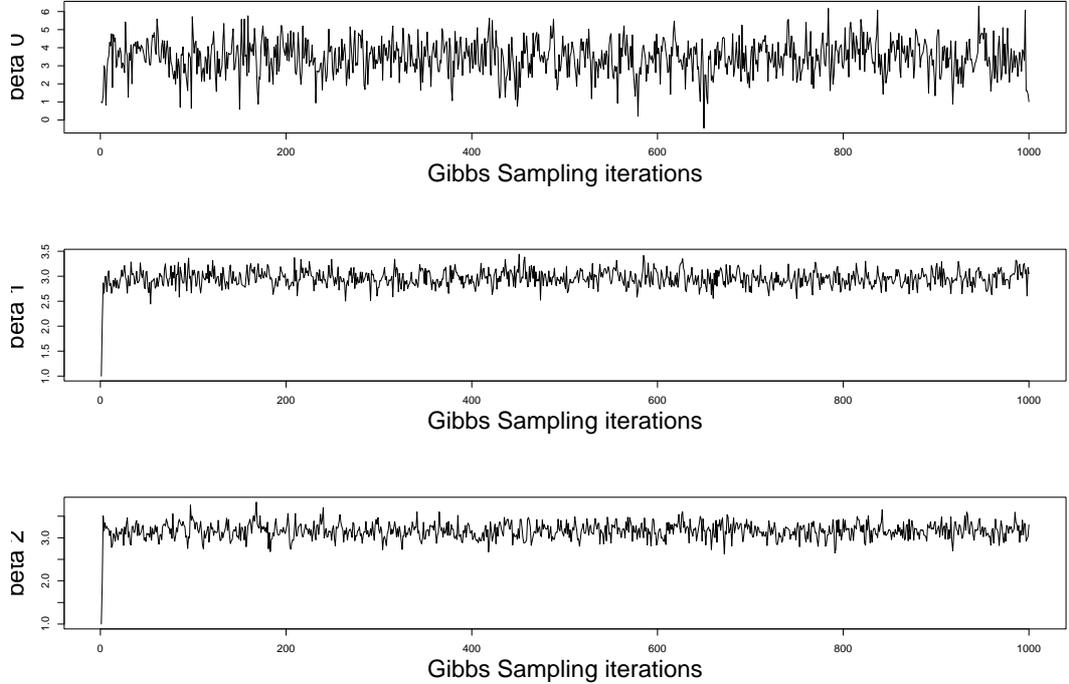


Figure 4: MCMC Gibbs samplers of coefficients  $\beta$  in the case of 50 grid cells.

deterministic models in our simulation study yielding output in 20 grid cells and each of them has two sampling points. Figure 7 shows the locations of the sites (monitored/unmonitored) and the sampling points within grid cells.

The mean of the true underlying random process  $Z$  is a polynomial function of the coordinates:

$$E[Z(\mathbf{s})] = \mu(\mathbf{s}) = \beta_0 + \beta_1 s_1 + \beta_2 s_2 + \beta_3 s_1^2 + \beta_4 s_2^2 + \beta_5 s_1 s_2,$$

$s_1$  and  $s_2$  being the longitude and latitude in degrees of the location  $\mathbf{s}$ . This second degree polynomial mean function of the coordinates is one degree higher than the mean function (19) used in the previous simulation. This more complicated choice permits us to see how well melding works no matter how complicated the mean function.

In this simulation, we generate 15 independent datasets. Initially we tried 1000 iterations

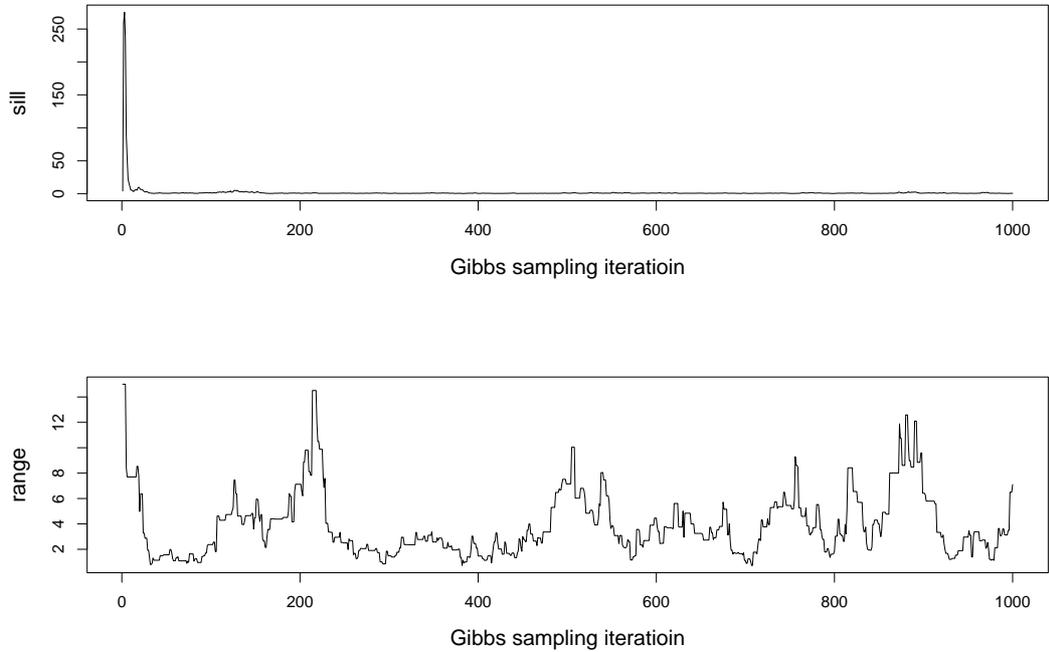


Figure 5: MCMC Gibbs samplers of covariance parameters  $\theta = (\sigma, \rho)$  in the case of 50 grid cells.

of the MCMC algorithm but found it failed to converge. Much better results obtained after we extended the number of iterations to 10,000 with a “burn-in” period of 1000. In the case of ensemble deterministic models, there are more parameters in the Melding model. The complexity of ensemble Melding model requires a much longer Markov chain to reach convergence.

## Results

Table 8 gives the estimated values for the parameters and Table 9, the SSPE (sum of squared prediction errors) for the 100 unmonitored stations. These results, suggest the following conclusions.

- Estimates of the additive and multiplicative biases for both deterministic models are quite accurate (Table 8).

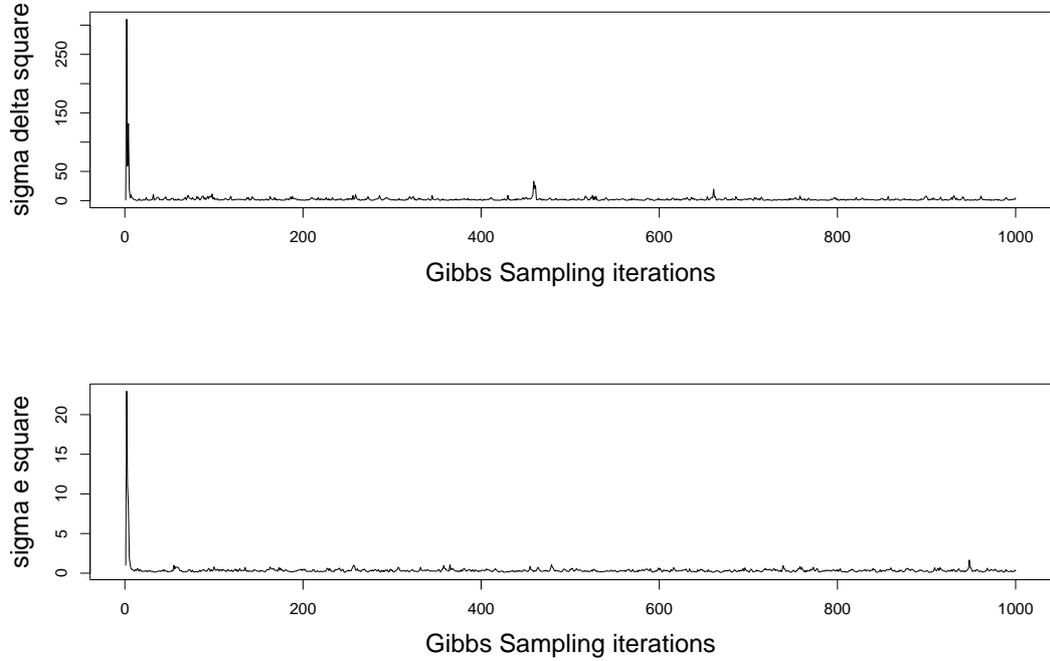


Figure 6: MCMC Gibbs samplers of error variance  $\sigma_\delta^2$  and  $\sigma_e^2$  in the case of 50 grid cells.

- The estimate of the *sill* parameter  $\sigma$  is very close to the true value while that of the *range* parameter  $\rho$  is not. However, the true value does lie within its 90% credibility interval.
- Estimates of  $\sigma_{\delta,2}^2$  is reasonably accurate. However, the estimate of  $\sigma_{\delta,1}^2$  exceeds the true value by quite a margin.
- Melding gives better predictions than Kriging, measured by SSPE, because the calibration parameters are estimated very well, meaning in effect, that the modeling output helps to achieve better prediction.

In summary, melding seems to work quite well when we have more than one deterministic model although the estimate of  $\sigma_{\delta,1}^2$  is not very close to the true value.

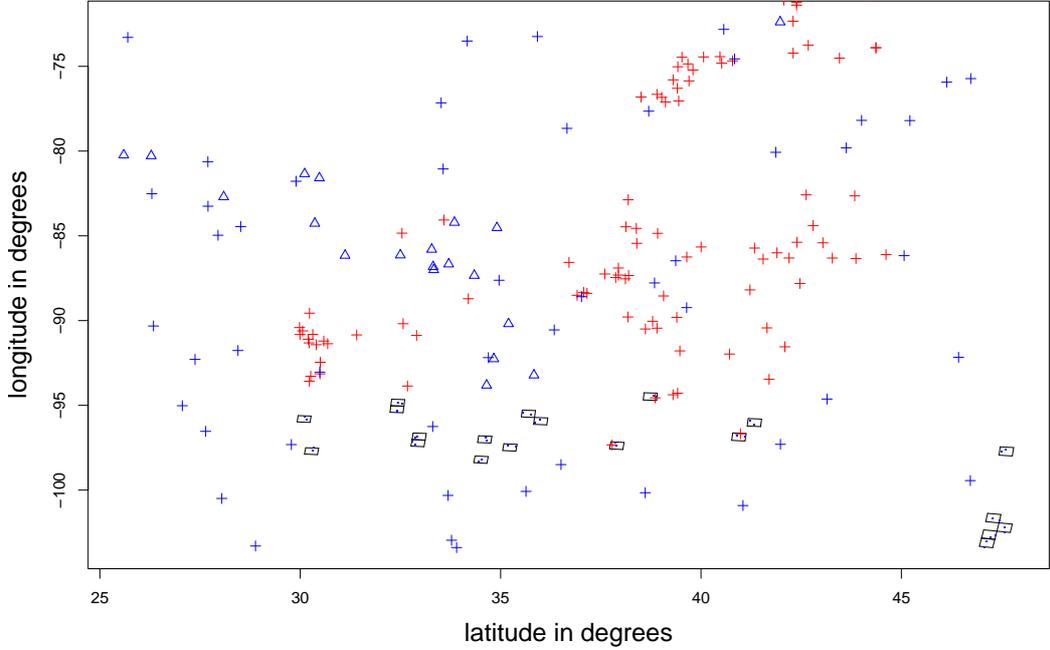


Figure 7: Locations of 150 sites (50 monitored, 100 to be predicted) and 20 grid cells. Stations:  $\Delta$ ; sites with sites to be predicted:  $+$ ; grid cells: rectangles. Each grid cell has two sampling points in it.

### 4.3 Reversible jump MCMC: the stationary case

This simulation involves 20 stations and 50 grid cells. Each grid cell has only one sampling point to reduce the computation burden. All stations (monitoring sites) as well as sampling points are generated uniformly on  $[-5, 5] \times [-5, 5]$ . The mean function of the underlying true process is

$$\mathbb{E}[Z(\mathbf{s})] = \mu(\mathbf{s}) = \beta_0 + \beta_1 s_1 + \beta_2 s_2 + \beta_3 s_1 * s_2$$

with true parameters  $\boldsymbol{\beta} = (1.3, 1.2, 0.5, 0.3)$ . The true parameters of the exponential covariance function are  $\sigma = 1.5$  and  $\rho = 5.0$ . The calibration parameters are  $a = 5.0$  and  $b = 2.5$ . The variances of the measurement error and modeling output error are  $\sigma_e^2 = 0.25$  and  $\sigma_\delta^2 = 0.25$  respectively. We simulate 15 independent datasets in total.

Figure 8 shows the MCMC plot and histogram of the dimension  $k$ . Table 10 shows the

estimation results. They lead us to make the following observations.

- The Markov chain for  $k$  converges to the true dimension  $k = 4$  (Figure 8).
- The estimates of  $\beta$  are averaged over all the MCMC iterations in which  $\beta$  has dimension of 4 and that the estimates are close to the true value of  $\beta$ .
- The point estimates of all the parameters are accurate except for those of  $\sigma_e^2$  and  $\sigma_\delta^2$ .

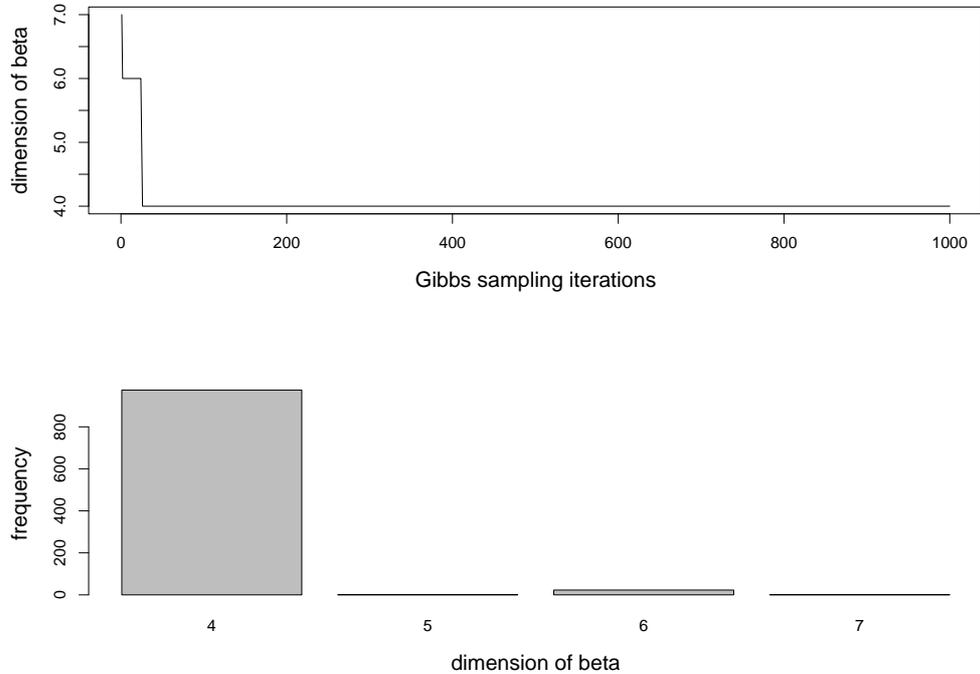


Figure 8: The upper plot: the dimension  $k$  of the coefficient  $\beta$ , as a function of the number of MCMC iterations. The lower plot: histogram of the posterior MCMC samples of  $k$ .

#### 4.4 Non-stationarity Melding Model

In Equation (4) the bandwidth  $h$  has to be chosen to represent the non-stationary process by convolution of latent stationary processes. A minimal requirement for  $h$ : ensure the positive definiteness of the covariance matrix. Yet if  $h$  is so small that two locations will

have no stationary processes between them, the covariance between these two locations will be zero. So, if  $h$  is too small then too many zeros will appear in the covariance matrix of the non-stationary process, resulting in numerical problems when this covariance matrix is inverted to evaluate the likelihood of the non-stationary process.

On the other hand, if  $h$  is too big, the covariance between two locations could also be too small because the kernel function is too flat and  $M$  in (5) is not big enough. In practice  $M$  cannot be chosen too big because of the computational burden. This simulation studies the effect of varying  $h$ . Primary interests focuses on spatial prediction error as well as on the accuracy of the  $a$  and  $b$  estimates.

Table 11 and 12 present the prediction and estimation results. They show that as long as  $h$  is not too small, both the SSPE and estimates of  $a$  and  $b$  are not much different. However, when the chosen  $h$  is too big, say, more than twice the true bandwidth used to simulate the data, the SSPE is much bigger. Also, the standard error of estimators of  $a$  and  $b$  are much bigger with small  $h$  (40% of the true  $h$ ) than the estimators with a bigger  $h$ . If  $h$  is chosen too small, say, less than half of the true bandwidth used to simulated the data, even the MCMC algorithm will "crash" because the covariance matrix becomes singular. So as a conservative strategy, we recommend choosing a large  $h$ . That choice of  $h$  is less likely to produce too many zeros in the spatial covariance matrix, leaving it more numerically stable.

Table 4: SSPE (Sum of squared prediction errors) by Kriging and Melding with different number of grid cells.

Dataset	Kriging	0cells	2 cells	10 cells	20 cells	30 cells	50 cells
1	116.03	74.11	74.21	75.83	61.75	56.67	71.86
2	104.50	42.16	42.73	40.57	40.60	41.17	39.46
3	223.32	104.80	115.60	107.57	94.54	88.72	78.28
4	119.76	78.73	76.37	73.50	73.48	69.00	63.57
5	182.26	80.53	90.61	72.47	73.20	69.82	62.21
6	135.59	59.63	64.79	52.65	52.60	53.14	51.06
7	150.89	68.86	68.71	70.23	68.27	67.88	67.92
8	127.38	92.14	90.36	94.48	86.12	89.74	79.00
9	115.09	65.81	59.11	62.19	57.87	60.31	56.87
10	121.31	120.22	117.87	106.56	103.36	91.41	82.43
11	107.19	56.38	67.78	57.15	54.68	55.63	51.33
12	172.63	100.56	96.47	96.12	93.62	89.37	86.77
13	140.90	71.37	70.93	69.01	63.82	62.55	56.94
14	194.51	126.80	134.38	134.74	120.24	123.25	104.89
15	181.34	92.89	89.98	83.47	84.37	83.54	78.83
16	257.84	151.52	141.73	129.46	118.79	105.07	95.62
17	141.69	81.64	80.08	77.18	73.68	72.90	57.69
18	170.47	90.68	89.63	91.66	84.61	79.97	73.82
19	113.81	73.24	74.97	74.34	69.21	67.39	66.16
20	99.03	83.48	85.73	82.64	80.42	79.52	73.01
21	249.37	163.09	147.83	156.51	141.12	131.03	105.89
22	138.46	85.70	95.09	81.80	79.31	72.60	64.44
23	138.66	82.25	73.79	81.62	69.68	68.94	61.82
24	110.81	76.74	75.58	79.02	81.83	81.40	69.30
25	110.84	92.90	98.19	88.23	84.67	82.00	76.09
26	162.32	112.48	114.98	110.75	102.47	91.22	85.60
27	203.55	112.85	113.07	116.73	123.22	112.19	105.84
28	138.53	72.54	74.38	73.53	69.50	68.60	66.67
29	117.12	82.46	82.15	101.03	75.22	74.61	69.97
30	152.81	135.98	133.59	122.34	117.14	93.48	72.49
31	157.05	117.43	117.13	107.43	106.66	105.21	92.00
32	140.36	101.53	100.57	93.45	88.29	93.38	88.65
33	142.60	103.28	101.85	104.96	106.54	111.43	107.54
34	102.68	93.99	86.41	88.18	74.78	77.84	67.23
35	153.52	80.53	78.76	75.80	74.19	76.07	70.82
36	171.43	75.52	73.94	70.59	65.64	63.30	52.16
37	105.98	85.41	85.17	79.46	72.99	78.09	66.47
38	144.45	62.95	71.70	63.52	68.56	66.12	60.56
39	257.85	173.02	170.12	163.96	179.29	154.24	132.83
40	158.86	123.88	119.54	109.30	95.47	91.86	104.50
41	145.30	93.46	100.74	101.24	94.68	84.76	65.31
42	128.23	72.57	73.41	73.23	69.72	67.10	65.88
43	193.23	80.03	83.14	82.81	82.26	80.17	75.48
44	130.01	72.40	72.02	74.59	74.97	70.93	67.97
45	112.00	79.34	80.79	72.00	76.52	67.87	81.12
46	168.65	121.84	95.57	91.42	86.39	91.86	76.49
47	213.19	137.96	139.81	122.81	122.18	112.25	93.99
48	100.05	84.13	74.76	76.80	68.59	65.79	63.60
49	136.49	70.64	74.48	66.42	62.05	66.89	54.86
50	113.47	110.61	92.53	89.05	85.96	79.55	73.53
mean	149.47	93.46	92.66	89.41	85.10	81.76	74.74

Table 5: The SSPE when the true values of the parameters ( $a$ ,  $b$ ,  $\beta$ ,  $\theta$ ) are used. That is, we use the conditional mean of the unavailable measurements based on the available measurements to predict at the 100 unavailable stations.

Dataset	2 cells	10 cells	20 cells	30 cells	50 cells
1	71.45	67.06	58.71	43.00	43.47
2	40.87	44.22	40.07	41.19	38.79
3	91.50	89.48	83.86	78.88	59.20
4	66.72	62.83	62.69	55.20	56.90
5	82.61	61.58	61.24	61.48	45.55
6	53.37	52.38	49.85	49.16	51.02
7	80.49	78.99	72.17	65.16	62.48
8	75.61	78.63	66.16	67.84	64.87
9	52.14	51.78	50.65	54.46	56.31
10	74.02	71.02	58.54	50.27	44.31
11	59.88	55.76	48.13	46.95	48.71
12	95.59	94.42	75.94	67.41	59.28
13	63.23	61.81	57.52	52.86	55.63
14	91.40	86.96	85.65	69.78	54.15
15	87.79	90.01	85.94	81.63	62.36
16	57.62	56.83	54.90	50.70	44.47
17	66.08	67.89	58.25	60.47	45.96
18	85.65	81.95	61.91	61.83	50.13
19	70.43	72.75	65.97	57.36	48.66
20	77.18	66.47	61.21	56.51	51.70
21	105.28	99.17	103.48	71.71	54.18
22	72.46	75.67	59.15	62.64	53.83
23	53.52	52.24	46.91	48.51	47.59
24	72.67	68.45	64.94	67.81	45.60
25	77.12	68.43	68.92	59.51	60.16
26	89.49	84.16	70.41	68.56	50.87
27	112.37	103.01	72.94	72.33	54.01
28	69.11	58.72	60.26	60.95	56.24
29	63.89	65.79	54.27	52.92	55.52
30	69.43	62.33	59.45	45.74	43.89
31	94.46	89.88	83.92	80.52	74.34
32	80.29	56.97	57.68	47.16	47.38
33	94.49	84.18	86.67	53.36	48.66
34	73.97	62.18	54.61	53.44	45.64
35	77.03	68.71	62.90	65.48	55.35
36	49.02	48.34	45.57	40.47	39.14
37	57.07	56.01	52.06	51.31	54.31
38	54.79	53.17	55.68	58.89	43.27
39	107.22	84.85	67.01	71.63	72.91
40	79.55	75.74	65.22	66.18	57.71
41	93.32	93.24	52.51	54.55	44.87
42	60.07	64.15	59.39	50.81	45.13
43	71.83	74.13	73.27	68.45	51.47
44	71.14	71.26	71.38	64.28	61.59
45	69.84	57.97	54.07	52.08	48.51
46	69.84	61.69	67.31	75.64	70.61
47	78.45	71.75	70.10	56.13	56.87
48	65.29	56.04	45.04	43.28	38.92
49	60.39	55.84	51.19	43.97	42.54
50	80.41	75.18	64.94	58.50	47.76
mean	74.35	69.84	63.21	58.78	52.26

Table 6: Coverage probability of the credible interval for the simulation study in Section 4.1, with 20 monitored sites and up to 50 grid cells to predict 100 un-monitored sites. The first column is the nominal coverage probability of the credible interval.

	kriging	0 cell	2 cells	10 cells	20 cells	30 cells	50 cells
0.95	0.37	0.68	0.91	0.94	0.94	0.97	0.95
0.90	0.32	0.59	0.84	0.89	0.88	0.86	0.91
0.80	0.26	0.49	0.73	0.79	0.78	0.76	0.80
0.70	0.21	0.40	0.62	0.70	0.68	0.66	0.71
0.60	0.18	0.33	0.52	0.57	0.56	0.55	0.59
0.40	0.12	0.22	0.34	0.39	0.38	0.36	0.40

Table 7: Estimates are based on 20 monitored sites and 50 sampling points. There are 25 sampling points in close proximity to one another. The averages are computed over 50 datasets. True values are  $\sigma = 1.50$  and  $\rho = 5.00$ .

	averaged posterior mean	averaged posterior sd
$\sigma$	0.97	0.25
$\rho$	4.69	0.76

Table 8: Parameter estimates in the simulation study of ensembles of deterministic models in Subsection 4.2. Columns 3-4 give the averages of posterior means and standard deviations for  $a_i, b_i, i = 1, 2$ , the calibration parameters of the modeling output for two deterministic models.  $\sigma_{\delta,1}$  and  $\sigma_{\delta,2}$  are the variances of two modeling output error processes. The averages are computed over 15 datasets.

Parameters	True value	mean	sd
$a_1$	5.00	4.94	0.48
$b_1$	2.50	2.50	0.02
$a_2$	4.00	4.00	0.68
$b_2$	3.40	3.39	0.03
$\sigma$	1.50	1.51	0.28
$\rho$	5.00	6.73	0.91
$\beta_0$	2.50	1.98	0.59
$\beta_1$	2.90	2.92	0.17
$\beta_2$	3.20	3.20	0.16
$\beta_3$	0.80	0.80	0.03
$\beta_4$	1.10	1.11	0.03
$\beta_5$	1.30	0.31	0.03
$\sigma_e$	0.25	0.33	0.13
$\sigma_{\delta,1}$	0.25	0.24	0.46
$\sigma_{\delta,2}$	0.25	4.51	0.91

Table 9: SSPE of Melding and Kriging in the simulation of multi-deterministic models. We have 50 monitored sites and 100 unmonitored ones. There are 20 grid cells each of them having two sampling points inside.

Dataset	Kriging	Bayesian Melding
1	42.57	33.13
2	74.07	70.99
3	85.39	59.90
4	56.63	48.59
5	81.75	45.36
6	81.91	73.19
7	56.64	55.89
8	55.31	48.74
9	85.39	56.79
10	72.93	57.10
11	60.41	49.30
12	65.04	59.70
13	97.69	50.43
14	62.23	60.55
15	60.79	41.45
mean	69.25	54.07

Table 10: Estimates in the stationary case using reversible jump Melding.

Parameters	True value	Estimate	Standard Error
$k$	4	4.079	0.41
$a_1$	5.00	5.32	0.52
$b_1$	2.50	2.45	0.07
$\sigma$	1.50	1.10	0.27
$\rho$	5.00	4.94	0.67
$\beta_0$	1.30	0.96	0.74
$\beta_1$	1.20	1.06	0.49
$\beta_2$	0.50	0.51	0.10
$\beta_3$	0.30	0.32	0.04
$\sigma_e^2$	0.25	0.39	0.06
$\sigma_\delta^2$	0.25	4.52	0.83

Table 11: SSPEs with different choices of  $h$  in non-stationary Melding. Columns 2-6 give the SSPEs for different ratios of the true bandwidth generating the data over the chosen bandwidth in the non-stationary melding model.

dataset	1.67	1.33	1.00	0.50	0.40
1	7.99	7.74	8.09	8.89	10.73
2	3.39	3.01	2.92	3.26	5.71
3	3.50	3.78	7.40	2.59	4.63
4	4.30	3.60	3.26	3.28	4.24
5	4.11	4.79	4.44	5.18	17.33
6	2.04	4.01	2.18	3.61	2.71
7	3.87	3.87	5.12	4.19	5.58
8	6.15	6.13	7.70	6.40	8.36
9	2.90	3.04	2.49	5.09	18.33
10	4.63	5.94	4.96	7.03	6.68
11	2.49	3.15	3.27	5.11	3.08
12	3.29	2.40	2.56	5.09	7.56
13	4.13	4.40	5.46	5.07	3.50
14	3.27	3.49	3.81	2.90	7.70
15	7.45	7.26	7.87	7.04	6.26
mean	4.23	4.44	4.77	4.98	7.49

Table 12: Estimates of  $a$  and  $b$  for varying values of  $h$  in non-stationary Melding. Columns 2-5 give the averages of posterior means and standard deviations. The averages are computed over the 15 datasets generated in the simulation study. The true values are  $a = 5.00$  and  $b = 2.50$ .

(true h)/(chosen h)	$\bar{a}$	$\bar{sd}$	$\hat{b}$	$\bar{sd}$
1.67	5.04	0.33	2.49	0.05
1.33	5.03	0.32	2.49	0.05
1.00	5.01	0.31	2.49	0.05
0.50	5.00	0.30	2.50	0.05
0.40	5.51	1.23	2.72	0.58

## 5 Application

This section presents an analysis of the results obtained by using the melding model to combine simulated ozone levels from a deterministic model with measurements from ozone monitoring sites located in the eastern and central USA.

### 5.1 Data description

The data in this section comes from two sources: regional surface ozone concentration measurements and modeling output from a deterministic model AQM (air quality model), a non-hydrostatic version of the MAQSIP (Multiscale Air Quality Simulation Platform) model. This AQM system has been described in detail by Wheeler and Houyoux. (1998). The AQM modeling output is based on grid cells with resolution  $6 \times 6$  km<sup>2</sup>. The measurements are from the Air Quality System (AQS) monitoring network. Both the measurements and modeling output are hourly data starting from May 15 to September 11, 1995 over a 120-day period. The dataset represents 375 monitoring stations in the AIRS network and 307 grid cells in the AQM output. Besides the fact that measurements and modeling output are based on different supports, the data represent different time standards, modeling output being based on the GMT (Greenwich Mean Time) time standard, the measurements on local time. Ignoring this time difference would result in poor correlation between measurements and modeling output. The next subsection shows the necessity to adjust this discrepancy.

This dataset has been analyzed by Kasibhatla and Chameides (2000), Hogrefe et al. (2001b) as well as Hogrefe et al. (2001a). The first paper compares the Pearson correlation coefficients between quantiles (10th, 25th, 50th, 75th and 90th) of the measurements and AQM modeling output for daily average data. The latter two papers also analyze the correlation between measurements and modeling output, but after decomposing the hourly time series into sub-series on different time scales, using the Kolmogorov-Zurbenko filter of

Zurbenko (1986). They reach similar conclusions, that the AQM modeling output represents the measurement better at longer time scales.

However in estimating the correlation between measurements and model output, the above analysis ignores the temporal and spatial correlation in the data, thus leaving some uncertainty about their validity. Moreover correlation analysis does not help in the assessment of the model calibration (additive and multiplicative) since the correlation is by definition invariant under the transformation of measurement scales. Finally our interest lies not merely in the degree of linear association between measurements and modeling output measured by the correlation, but rather in predicting the ozone level at unmonitored locations. These factors motivate a new data analysis using the Bayesian melding model.

## 5.2 Preliminary data analysis

The previous subsection points out that the measurements are based on local time while the modeling output are based on GMT (Greenwich Mean Time). So we need to adjust the modeling output to the local time. Being between May and September, those local times must be in daylight savings for the zone of the relevant monitoring site, that is one of: Eastern Daylight Time (EDT), Central Daylight Time (CDT) and Mountain Daylight Time (MDT).

Prior to adjustment, the correlation between measurements and modeling output can be quite small. Consider for example, the measurements from monitoring site (or station for brevity) #290470003 that lies inside the model grid cell #1847. That station, being in Mississippi, is on CDT, that is GMT -5 hours before adjustment. The scatter plots and their Pearson correlation coefficient in Figure 9 show little linear association between the two series prior to realigning the times. Correcting the misalignment increases considerably, the correlation between that measurements and simulated data for grid cell #1847. Figure 9 demonstrates the necessity to adjust the modeling output to the local time standard. The dataset has 119 days after the adjustment.

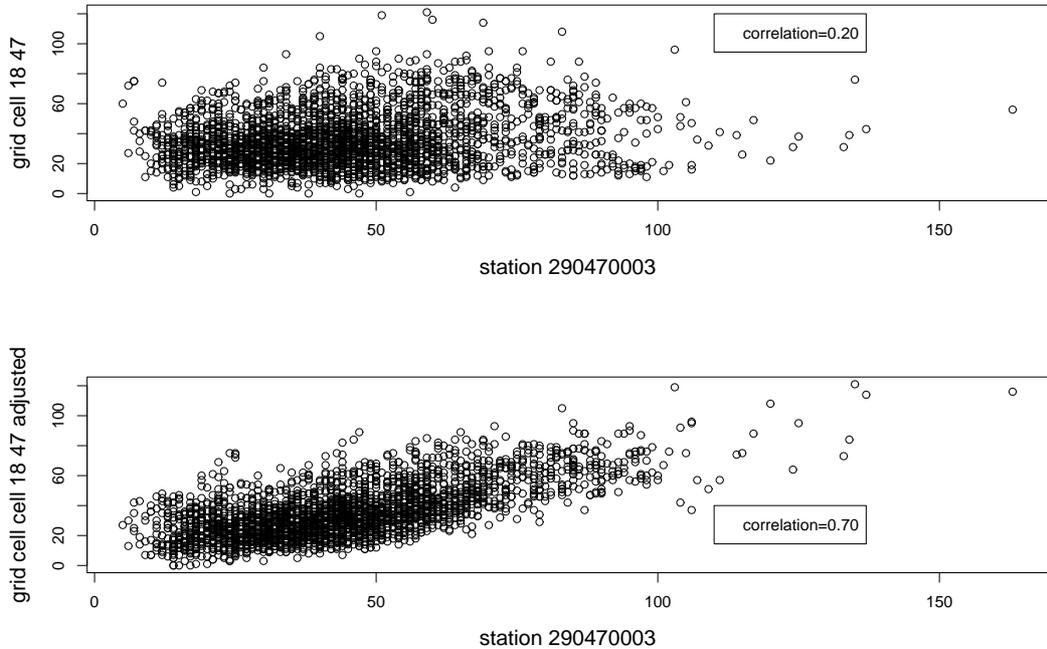


Figure 9: Scatter plots and Pearson correlations between measurements of station #290470003 and modeling output for cell #1847. The x-axis stands for the modeling output of grid cell #1847 and the y-axis stands for the measurements at station #290470003. The station #290470003 is inside grid cell #1847. The upper panel is before adjusting the modeling output to the local time and the lower is after. Observe that the adjustment results in marked increase in the correlation from 0.2 to 0.7.

The measurement series, unlike that for the modeling output, have missing values. For example, all the measurements from station 550730005 (in Wisconsin state) are missing. To deal with the missing values, we first choose those stations that have no more than 100 hours of missing measurements. Second, we use the 24 hour mean to fill in the missing values. For example, if the missing value occurs at 10 AM, then we use the average of the available values at 10 AM every day to fill in this missing value. After adjusting for different time standards and ignoring the stations with more than 100 missing measurements, we have measurements at 81 stations and modeling output on 375 grid cells of 2856 hours (119 days). In these 375 grid cells, there are 78 grid cells which contain one and only one station. To enable us understand better the role of model-to-measurement correlation in spatial

prediction, from now on the data always will focus on the 78 grid cells with 78 stations inside the grid cells. Although the measurements and modeling output are available during the 119 day period, we only focus on the 30 days of July, when the ozone concentration is at a high level due to high temperatures.

The ozone concentration level at night and early in the morning is much lower than during the day. Figure 10 shows the 24 side-by-side hourly box plots of measured ozone levels at station #10731005 and simulated ozone levels at grid cell #3529. The station #10731005 is inside grid cell #3529. Both the observed and simulated ozone level are at a peak during the 8 hours from 10 AM to 17 PM. Figure 11 shows the histograms of the model-to-measurement correlation at all 24 hours and at the 8 hours. Based on Figures 10 and 11, we focus on the analysis of 8-hour measurements and modeling output, since otherwise the two data sources are quite dissimilar. Moreover, little interest obtains in the hours outside that period. From now on, the 8-hour average measurements or modeling output is also referred to as the "daily average". We apply the melding model to analyze the hourly, daily and weekly average ozone concentration levels for the selected 8-hour time period. Kasibhatla and Chameides (2000); Fiore et al. (2003, 2004) focus their analysis on similar daily average data. In particular, Fiore et al. (2003) use the empirical orthogonal function (EOF) method, a principal component analysis (PCA), to compare the first two principal components of the measurement and modeling output.

The simulated ozone levels at the 78 grid cells and measurements at 48 stations are used to fit the Bayesian melding Model (7). The measurements at the remaining 30 stations are used as validation data. These 30 stations are called "un-monitoring" sites from now on. Figure 13 maps the grid cells and monitoring/un-monitoring sites.

We compare the results of spatial prediction between two competing approaches: Bayesian melding and Kriging. In fact, we have two different versions of Kriging: Kriging using measurements and Kriging using modeling output. The RMSPE (root mean square prediction error) measures the predictive performance. At time  $t$ , which could be hour, day or week,

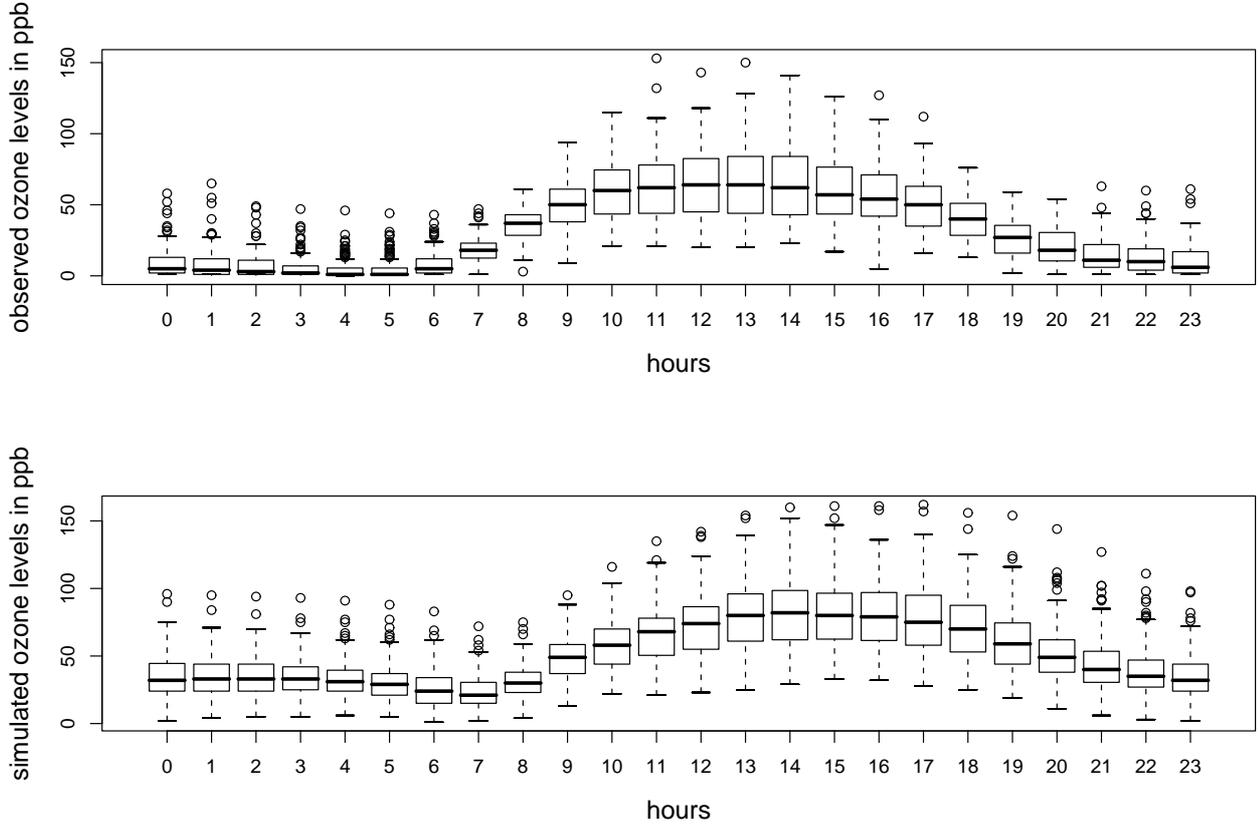


Figure 10: The upper panel is the 24 side-by-side hourly box plots of measurements at station #10731005. The lower panel is the 24 side-by-side hourly box plots of modeling output (simulated ozone levels) in grid cell #3529. The station #10731005 lies inside grid cell #3529.

we define the RMSPE by

$$\text{RMSPE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - \hat{O}_i)^2},$$

$n$  being the number of un-monitoring sites to be predicted,  $O_i$ , the measurement at station  $i$  and  $\hat{O}_i$ , the prediction.

### 5.3 Analysis of Hourly Data

For the 30 days from July 1 to July 30, 1995, we have 240 hourly measurements and modeling outputs at the daily 8-hour time period (10 AM-17 PM) selected for analysis.

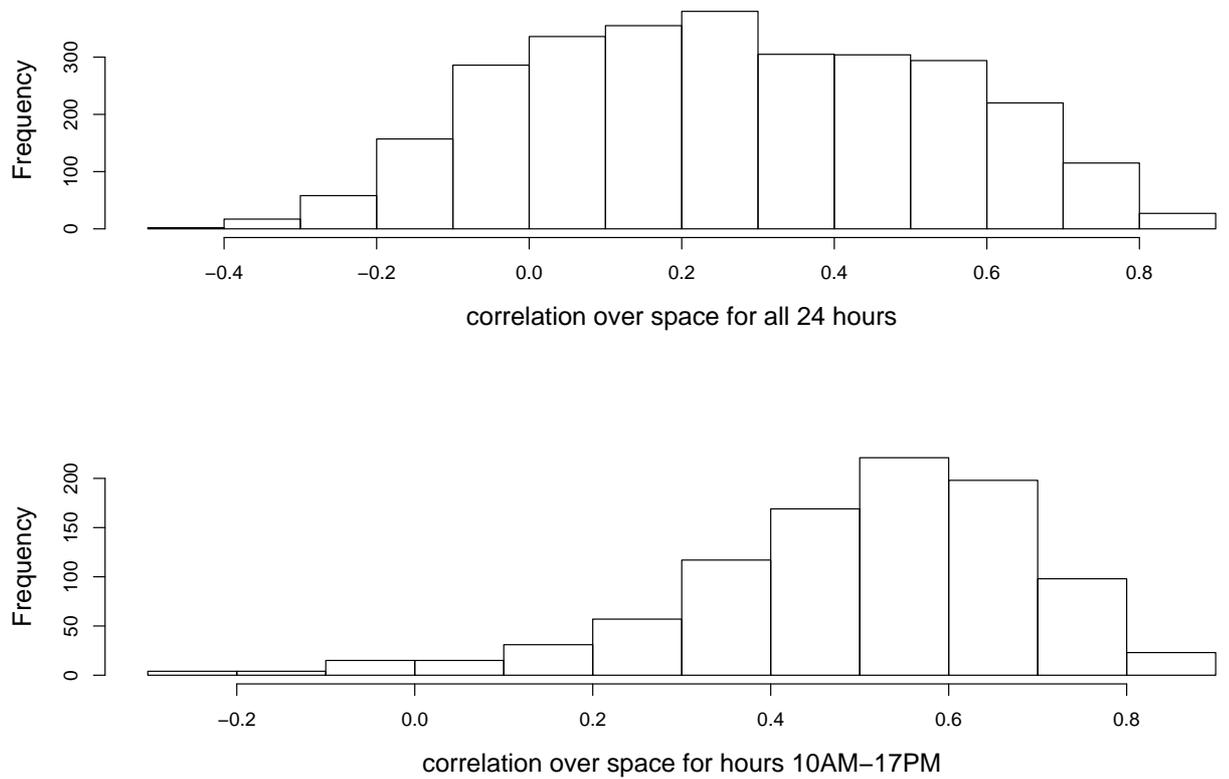


Figure 11: Histograms of correlation over space at each hour. The upper histogram is for all the 24 hours. The lower histogram is for the 8-hour day time. Notice the markedly larger model-to-measurement correlations during the hours 10AM-17PM.

We apply melding and Kriging to each hour separately. For brevity, we list the average RMSPE for melding and Kriging in Table 13. The averages are computed over the 240 hours. Figure 12 shows the plot of the RMSPE difference between melding and Kriging using only measurements versus the correlation between measurements and modeling output over space. We offer the following observations about the analysis of hourly data.

- Table 13 shows that on average, the melding predictor has the smallest RMSPE among all the competitive prediction approaches.
- The melding predictor seems marginally better than Kriging using only the measurements. The RMSPE for melding is 15.82 and for Kriging using measurements, 16.36.

As shown in the next subsection, the melding predictor does much better than that version of Kriging in the analysis of daily average data.

- Kriging using modeling output has the biggest RMSPE, pointing to the desirability of calibrating the modeling output.
- Figure 12 shows that in general, for spatial prediction melding outperforms Kriging using measurements when the correlation between measurements and modeling output is 0.6 or bigger.
- The coverage probability of melding’s 90% predictive interval is 86.72%, which is fairly good.

Table 13: Average RMSPE of hourly ozone predictions at 30 stations. Column 1: RMSPE for Bayesian melding. Column 2: RMSPE for Kriging using measurements. Column 3: RMSPE for Kriging using modeling output.

melding	Kriging 1	Kriging 2
15.82	16.36	16.91

## 5.4 Analysis of Daily Average Data

This section presents our analysis of daily 8-hour (10AM-17PM) measurements and modeling output averages. We would make the following observations about our analysis of daily average data.

- Table 14 shows that for daily averages, melding has the smallest RMSPE among all competitive prediction approaches. The RMSPE of melding is smaller than Kriging using measurements on 18 out of the 30 days. The average RMSPE for Kriging using modeling output is the biggest, showing the need to calibrate the modeling output.

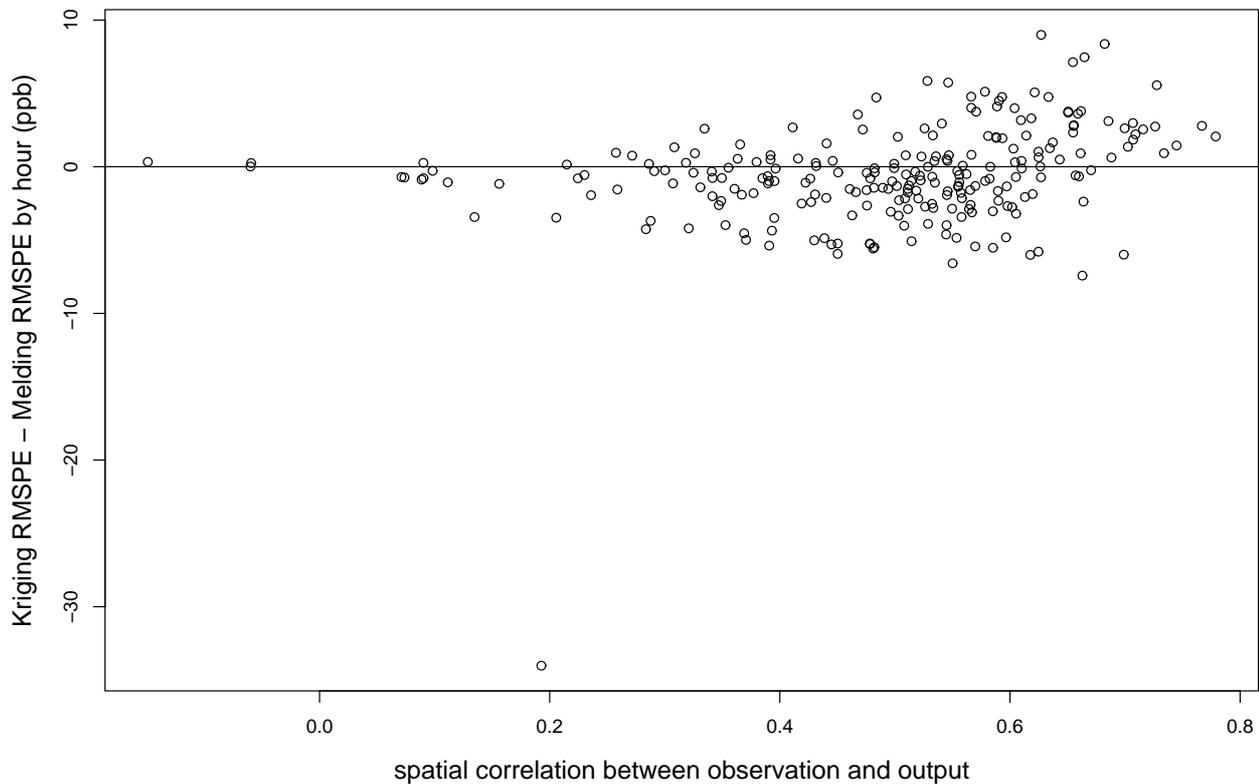


Figure 12: Hourly RMSPE (root mean square prediction error) difference between Kriging using measurements and melding versus the correlation over space between measurements and modeling output. Points above the horizontal line in the plot represent victories for Melding over Kriging.

- Figure 14 confirms the intuitively plausible result that in general the prediction performance of melding improves as the correlation between measurements and modeling output increases. When the correlation exceeds 0.6, melding performs substantially better than Kriging on most days.
- In implementing the melding model, we assume the additive calibration parameter  $a(\mathbf{s})$  is a polynomial function of the coordinates at location  $\mathbf{s}$ , that is  $a(\mathbf{s}) = f(\mathbf{s})\boldsymbol{\beta}_a$ . We use the reversible jump MCMC to choose the degree of the polynomial function. The reversible jump MCMC can sample the dimension of  $\boldsymbol{\beta}_a$  from its posterior distribution. Section 3.6 describes this reversible jump MCMC in detail. For all 30 days,

the posterior distribution of  $\beta_a$  ranges between 1 to 3. So, we can assume  $a(\mathbf{s})$  is a linear function of the coordinates at location  $\mathbf{s}$ , that is,  $a(\mathbf{s}) = a_0 + a_1s_1 + a_2s_2$ ,  $s_1$  and  $s_2$  being the longitude and latitude in degrees at location  $\mathbf{s}$ . By assuming  $a$  is a linear function coordinates at location  $\mathbf{s}$ , we get a smaller RMSPE than by assuming  $a$  is a constant across the space.

- In the melding Model (7), we assume that the multiplicative calibration parameter  $b$  is constant across stations. The more realistic this assumption, the better will be the melding predictor's performance against Kriging. In the dataset, each station is located in a grid cell, so the modeling output for that grid cell can be treated as the modeling output at the station inside the grid cell. Thus, at each grid cell  $\mathbf{B}$ , we can plausibly estimate  $b$  by

$$\hat{b}(\mathbf{B}) = \frac{\tilde{Z}(\mathbf{B}) - \hat{a}(\mathbf{B})}{\hat{Z}(\mathbf{B})}, \quad (20)$$

where  $\hat{Z}(B) = \int_B \hat{Z}(\mathbf{s})d\mathbf{s}$  is the integral of  $\{\hat{Z}(\mathbf{s})\}$  over the grid cell  $B$ . Because we only have one sampling point in each grid cell, this integral is just  $\hat{Z}(\mathbf{s})$ , that is, the measurement at the station inside grid cell  $B$ . After obtaining  $\hat{b}(B)$  from the above formula, we can compute the sampling variance of  $\{\hat{b}(B)\}$  and the absolute mean difference of  $\{\hat{b}(B)\}$  between grid cells having measured or unmeasured stations. Figure 15 and Figure 16 show plots of (Kriging RMSPE - melding RMSPE) versus  $\hat{\text{Var}}(\hat{b})$  and the absolute mean difference for  $\hat{b}$ , respectively, for the 30 days in July. From these plots we can see that melding prediction is better than Kriging when  $\hat{\text{Var}}(\hat{b})$  and the absolute mean difference of  $\hat{b}$  are small. This finding is expected because the melding prediction is better when the model assumptions are more justified.

- The coverage probability of the 90% credible for the melding prediction is 87.33%, which is fairly good.

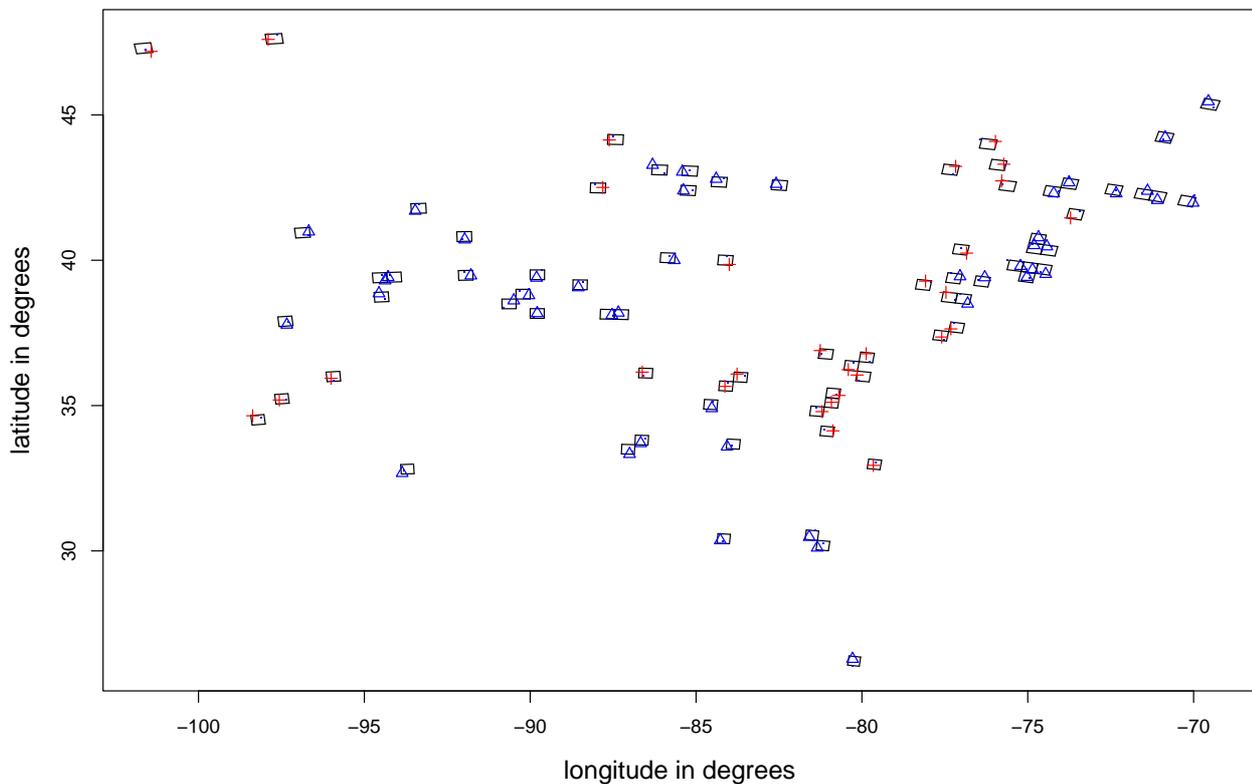


Figure 13: Locations of 48 available stations, 78 grid cells and 30 unavailable stations.  $\Delta$ : available stations. Rectangle: grid cells. +: unavailable stations to be predicted.

### 5.5 Analysis of weekly Average Data

Fuentes and Raftery (2005) analyze weekly average  $\text{SO}_2$  measurements and modeling output in USA. The available/unavailable stations and grid cells used in the analysis of weekly average data is the same as in the analysis of daily average data. However, with averages over longer time scales, the predictions of both melding and Kriging are expected to improve because of better normality of the data distribution and smaller variation of the average data over the longer time scale. The improvement of the melding prediction also lies in the better performance of deterministic model for the longer time scale as noted by Kasibhatla and Chameides (2000); Hogrefe et al. (2001b,a). Table 15 presents that the RMSPE of melding, Kriging with only measurements, Kriging with measurement plus modeling output without

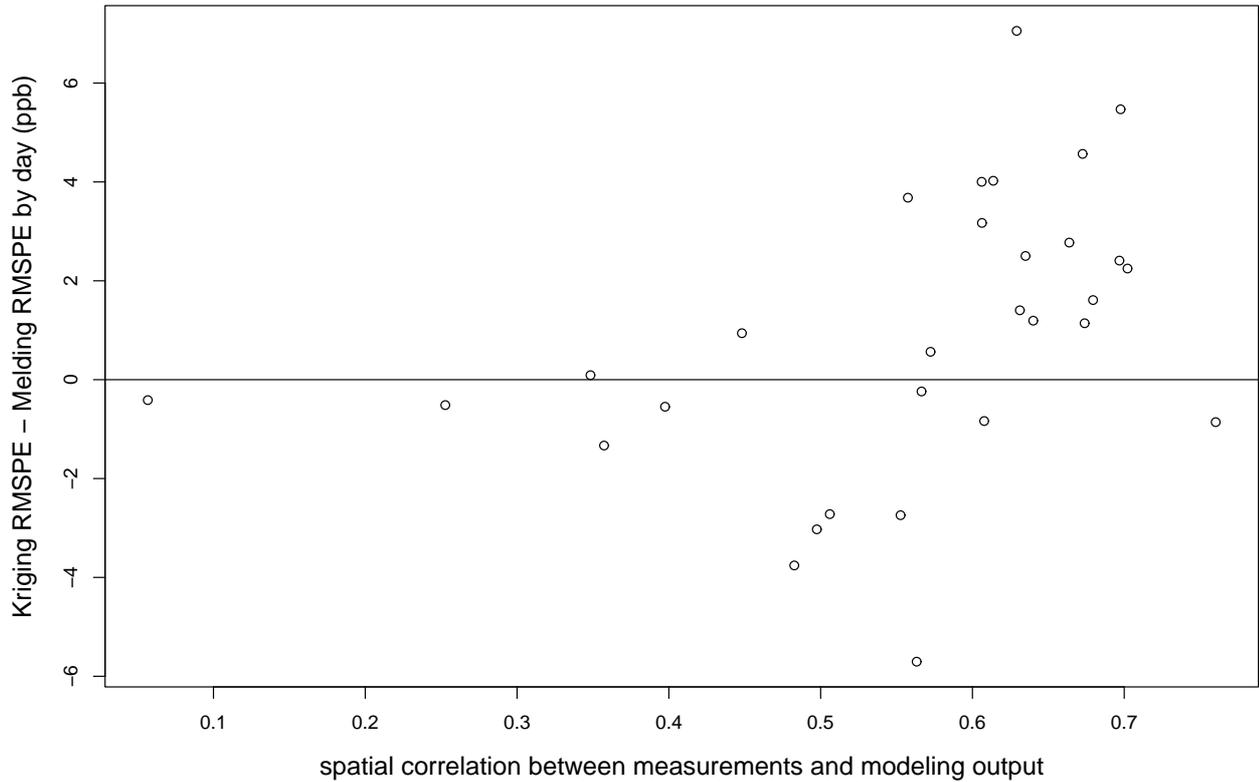


Figure 14: RMSPE difference between Kriging prediction with measurements and melding prediction versus the correlation over space between measurements and modeling output. Points above the plotted horizontal line represented victory for Melding. Notice the supremacy of Melding when correlation exceeds 0.6.

calibration, Kriging with only modeling output without calibration. We have the following findings from the analysis of weekly average data.

- Table 15 shows that melding achieves the smallest RMSPE in week 2 and week 4. On average over all 4 weeks, Melding has smallest RMSPE.
- As expected, the RMSPE for all the predictors are smaller than RMSPE in the analysis of daily average data.
- Melding’s 90% predictive interval has a reasonably good coverage probability of 91.67%.
- The RMSPE for the melding model in Tables 13, 14 and 15 tells us that the melding

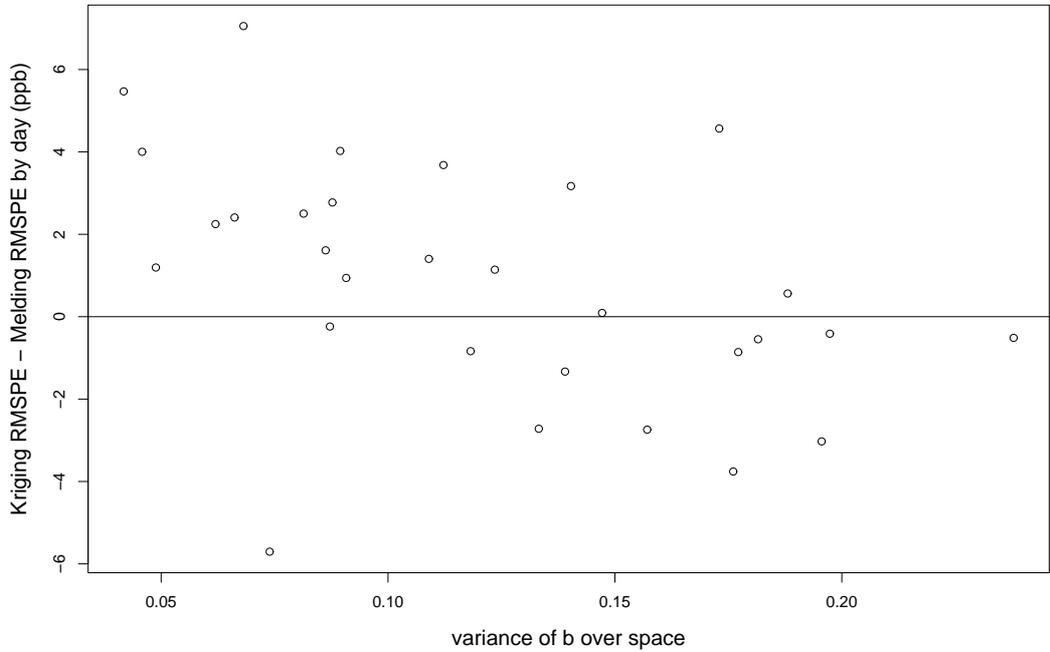


Figure 15: Daily RMSPE difference between Kriging with measurements and melding versus the variance of  $\hat{b}$  (defined by formula (20) ) across space. Points above the plotted horizontal line represented victory for Melding. Notice the supremacy of Melding when the variance of  $\hat{b}$  over space is small.

model’s predictor improves as the averaging time scale increases. This agrees with the findings in Kasibhatla and Chameides (2000); Hogrefe et al. (2001a), that is, the modeling output represents the measurements better on larger time scales.

## 6 Summary and Conclusions

Previous sections present the details of the Bayesian melding model and the MCMC algorithm used to fit the model. We have also conducted a comprehensive simulation and applied the melding model to the ozone data in different time scales.

We see that the Bayesian melding model has its strengths and weakness. Its strengths are the following.

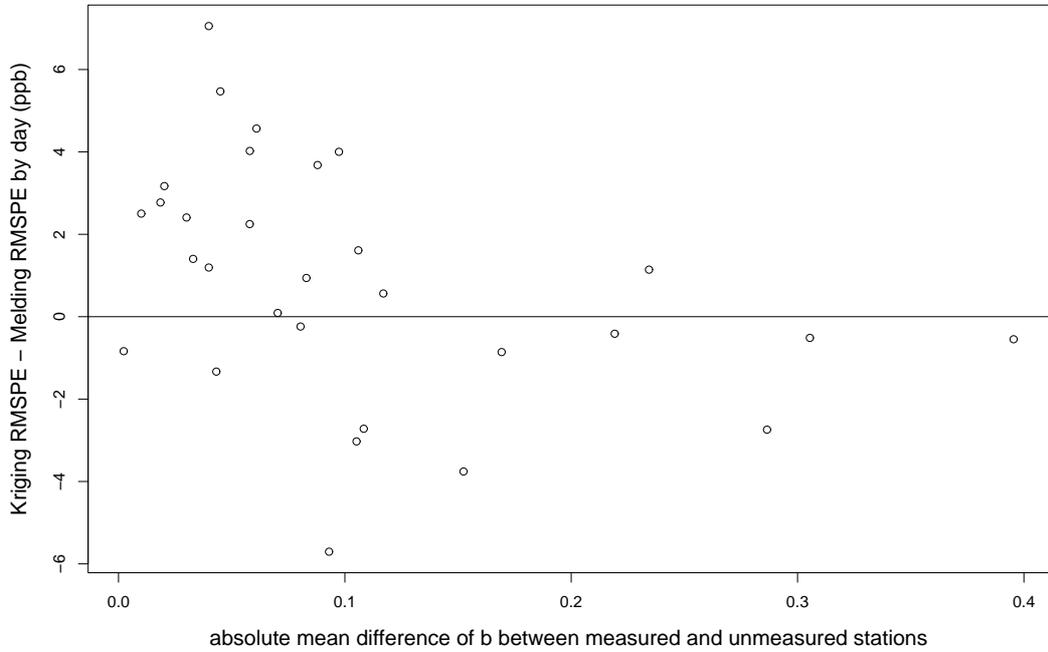


Figure 16: Daily RMSPE difference between Kriging with only measurement and melding versus the absolute mean difference of  $\hat{b}$  (defined by formula (20)) between the measured and unmeasured stations. Points above the plotted horizontal line represented victory for Melding. Notice the supremacy of Melding when the absolute mean difference of  $\hat{b}$  over space is small.

- The novel method connects the measurement process with the modeling output process by assuming the existence of an underlying process. This approach addresses the difference-in-support problem in a fundamental way.
- Melding can do a variety of things such as predict unmeasured responses, assessing deterministic modeling output, detecting spatial trends and estimating spatial correlation.
- The model better estimates prediction uncertainty than the classical approach, Kriging. In our data analysis, the coverage probability of Melding’s predictive interval is reasonably close to the nominal level.
- Its Bayesian framework makes melding relatively easy to extend to incorporate other

things like ensembles, the reversible jump MCMC and non-stationary spatial correlation.

However, melding model also has some weakness as following.

- The computational price is high. By sampling points within grid cells, the dimension of the spatial correlation matrix can be very big even with a modest number of grid cells. Inverting the spatial covariance matrix three times in each MCMC iteration takes a lot of computation time.
- Melding does not yet cover space-time processes and hence it can not "borrow strength" over time. Ozone data are recorded hourly, which obviously has strong periodicity and strong auto-correlation. The main challenge to create a "space-time" Bayesian melding model lies in the computation burden. With temporal correlation, the space-time correlation matrix will be huge and hence more likely to be ill-conditioned. Inverting such a matrix will be both difficult and computationally expensive.
- Melding's normality assumption poses a problem. For the measurements, a transformation maybe used to validate that assumption. However, with the modeling output data, non-linear transformations cannot be used since the modeling output is represented by such an integral of the true process.
- The locations of sampling points within grid cells change because of the random sampling scheme. Because the mean of the underlying process depends on its geographical coordinates in general (like universal Kriging), the prediction result have some variation every time the Bayesian melding is used for the same data.

Finally, we offer the following observations.

- Where some of the stations are within grid cells, it would seem better to use them instead of sampling points within the grid cells to reduce the dimension of the spatial correlation matrix.

- The meaning of the concept of a “true underlying process” seems unclear since it is merely a conceptual construct rather than a physically meaningful process. Thus its may will be criticized.
- We assume the underlying process mean to be a polynomial function of the geographical coordinates, which may not be enough if other variable such as temperature also affects the ozone level. We can side-step the assumption of a true underlying process by regressing the measurements on the modeling output directly.

## **7 Acknowledgments**

We thank Prasad Kasibhatla for providing the ozone air pollution data and MAQSIP modeling output for the applications in Section 5, and for stimulating conversations about the topic of this report with the third author while the latter was visiting the Statistical and Applied Mathematical Sciences Institute.

Table 14: RMSPE (root mean square prediction error) for predicting daily average of ozone levels at 30 stations. Column 1: days. Column 2: RMSPE for Bayesian melding. Column 3: RMSPE for Kriging with measurements. Column 4: RMSPE for Kriging with modeling output. The number with \* is the smallest number in each row.

day	melding	Kriging 1	Kriging 2
1	14.50	13.98	13.91*
2	8.99	8.58*	12.86
3	11.43	8.71*	11.84
4	13.53	12.19*	15.51
5	13.65	10.63*	15.59
6	14.43	11.69*	16.54
7	13.59	13.04*	13.54
8	11.73*	11.82	12.89
9	17.86	14.11	15.99
10	15.14	9.43	12.76
11	11.59	15.60	11.32
12	15.30*	16.24	15.30
13	14.14*	19.61	15.16
14	15.61*	22.67	17.99
15	18.55	21.06	18.45*
16	17.95	19.36	17.60*
17	11.37*	14.14	11.70
18	13.67	16.84	11.65*
19	6.98 *	11.54	10.51
20	12.84*	14.03	15.45
21	17.44	16.61*	17.33
22	9.48*	11.88	12.54
23	10.04*	11.65	8.65
24	9.60*	9.37	11.99
25	11.76*	12.90	21.11
26	17.55*	18.11	19.03
27	12.26*	15.94	14.95
28	14.31*	18.33	18.07
29	13.49	12.63*	19.10
30	12.41*	14.66	14.26
mean	13.37*	14.24	14.79

Table 15: RMSPE (root mean square prediction error) for predicting weekly average of ozone levels at 30 stations. Column 1: days. Column 2: RMSPE for Bayesian melding. Column 3: RMSPE for Kriging with measurements. Column 4: RMSPE for Kriging with modeling output. The number with \* is the smallest number in each row.

week	melding	Kriging 1	Kriging 2
1	9.43	7.98*	13.77
2	9.72*	11.22	10.51
3	10.32	10.21	9.83*
4	8.08*	11.76	12.71
mean	9.38*	10.29	11.71

## References

- Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions*, Dover.
- Caya, D., Laprise, R., Giguere, M., Bergeron, G., Blanchet, J., Stocks, B., Boer, G. and McFarlane, N. (1995). Description of the canadian regional climate model, *Water, Air and Soil Pollution* **82**: 477–482.
- Cressie, N. A. (1993). *Statistics for Spatial Data*, Wiley-Interscience.
- Fiore, A. M., Jacob, D., Liu, H., Yantosca, R. M., Fairlie, T. D. and Li, Q. (2004). Variability in surface ozone background over the united states: Implications for air quality policy, *Journal of Geophysical Research*.
- Fiore, A. M., Jacob, D., Mathur, R. and Martin, R. V. (2003). Application of empirical orthogonal functions to evaluate ozone simulation with regional and global models, *Journal of Geophysical Research*.
- Fuentes, M. and Raftery, A. (2002). Model validation and spatial interpolation by combining observations with outputs from numerical models via bayesian melding, *Technical Report 403*, Department of Statistics, University of Washington.
- Fuentes, M. and Raftery, A. (2005). Model evaluation and spatial interpolation by bayesian combination of observations with outputs from numerical models, *Biometrics* **61**: 36–45.
- Fuentes, M. and Smith, R. (2001). A new class of nonstationary spatial models, *Journal of the American Statistical Association*, *being reviewed*.
- Fuentes, M., Guttorp, P. and Challenor, P. (2003). Statistical assesment of numerical models, *Technical Report 076*, NRCSE.
- Garner, J., Lewis, T., Hogsett, W. and Andersen, C. (2005). Air quality criteria for ozone and related photochemical oxidants (second external review draft) volume iii, *Technical report*, U.S. Environmental Protection Agency.

- Gelfand, A. and Smith, A. (1990). Sampling-based approaches to calculating marginal densities, *Journal of American Statistical Association* **85**: 398–409.
- Gelfand, A., Schmidt, A., Banerjee, S. and Sirmans, C. (2004). Nonstationary multivariate process modelling through spatially varying coregionalization (with discussion), *Test* **13**: 1–50.
- Green, P. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination, *Biometrika* **82**: 711–732.
- Guttorp, P. and Walden, A. (1987). On the evaluation of geophysical models, *Geophysical Journal of the Royal Astronomical Society* **91**: 201–210.
- Haas, T. (1995). Local prediction of a spatio-temporal process with an application to wet sulfate deposition, *Journal of the American Statistical Association* **90**: 1189–1199.
- Haas, T. (1998). Statistical assessment of spatio-temporal pollutant trends and meteorological transport models, *Atmospheric Environment* **32**: 1865–1879.
- Hasting, W. (1970). Monte carlo sampling methods using markov chains and their application., *Biometrika* **57**: 97–109.
- Hawkins, D. and Cressie, N. (1984). Robust kriging-a proposal, *Journal of the International Association for Mathematical Geology* **16**: 3–18.
- Higdon, D., Swall, J. and Kerv, J. (1999). Non-stationary spatial modelling, *Bayesian Statistics* pp. 761–768.
- Hogrefe, C., Rao, S., Kasibhatla, P., Hao, W., Sistla, G., R., M. and McHenry, J. (2001a). Evaluating the performance of regional-scale photochemical modeling systems: Part ii-ozone predictions, *Atmos. Environ.* **35**: 4175–4188.
- Hogrefe, C., Rao, S., Kasibhatla, P., Kallos, G., Tremback, C., Hao, W., Olerud, D., Xiu, A., McHenry, J. and Alapaty, K. (2001b). Evaluating the performance of regional-

- scale photochemical modeling systems: Part i-meteorological predictions, *Atmos. Environ.* **35**: 4159–4174.
- Kasibhatla, P. and Chameides, W. L. (2000). Seasonal modeling of regional ozone pollution in the eastern united states, *Geophys. Res. Letter* **27**: 1415–1418.
- Lindley, D. and Smith, A. (1972). Bayes estimate for the linear model, *Journal of the Royal Statistical Society. Series B (Methodological)*.
- Matern, B. (1960). Spatial variation, *Technical Report 49*, Meddelanden fran Statens Skogsforskningsinstitut.
- Matheron, G. (1962). Trait de geostatistique appliquee, tome i, *Memoires du Bureau de Recherches Geologiques et Minieres*.
- Matheron, G. (1963). Traite de geostatistique appliquee, *Technical report*, Memoires du Bureau de Recherches Geologiques et Minieres.
- Odman, M. and Ingram, C. (1996). Multiscale air quality simulation platform (maqsip) source code documentation and validation, *Technical report*, Environmental Programs, MCNC-North Carolina Supercomputing Center.
- R Development Core Team (2006). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Ribeiro, P. J. and Diggle, P. (2001). geor: a package for geostatistical analysis, *R-NEWS* **1**(2): 14–18. ISSN 1609-3631.
- Sampson, P. and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure, *Journal of the American Statistical Association* **87**: 108–119.
- Sanso, B. and Guenni, L. (2002). Combining observed rainfall and deterministic prediction using a bayesian approach, *Technical report*, CESMA.

- Smith, A. (1973). A general bayesian linear model, *Journal of the Royal Statistical Society. Series B (Methodological)*.
- Stein, M. (1999). *Interpolation of spatial data : some theory for kriging*, New York : Springer.
- Wheeler, N. and Houyoux., M. (1998). Development and implementation of a seasonal model for regional air quality, *Proc. Air and Waste Manage Association*.
- Zurbenko, I. (1986). *The Spectral Analysis of Time Series*, North-Holland, Amsterdam.