1

Bayesian Forecasting Using Spatio-temporal Models with Applications to Ozone Concentration Levels in the Eastern United States

Sujit Kumar Sahu
Southampton Statistical Sciences Research Institute,
University of Southampton,
Southampton, SO17 1BJ, UK.
Khandoker Shuvo Bakar
CSIRO Mathematics, Informatics and Statistics,
GPO Box 664, Canberra, ACT 2601, Australia.
Norhashidah Awang
School of Mathematical Sciences, Universiti Sains Malaysia,
11800 USM Penang, Malaysia.

1.1 Summary

Bayesian forecasting in time and interpolation in space is a challenging task due to the complex nature of spatio-temporal dependencies that need to be modeled for better understanding and description of the underlying processes. The problem exacerbates further when the geographical study region, such as the one in the Eastern United States considered in this chapter, is vast and the training data set for forecasting, and modelling, is rich in both space and time. This chapter develops forecasting methods for three recently

proposed hierarchical Bayesian models for spatio-temporal data sets. The chapter also develops Markov chain Monte Carlo based computation methods for estimating a number of relevant forecast calibration measures that facilitates rigorous comparisons of the Bayesian forecasting methods. The methods are illustrated with a test data set on daily maximum eight hour average ozone concentration levels observed over a study region in the Eastern United States. Forecast validations, using several moving windows, find a model developed using an approximate Gaussian predictive process to be the best and it is the only viable method for large data sets when computing speed is also taken into account. The methods are implemented in a recently developed software package, spTimer, which is a publicly available contributed R package that has wider applicability.

1.2 Introduction

Bayesian forecasting methods are very much in demand in many application areas in environmental monitoring and surveillance. Consequently, model based forecasting has attracted much attention in the literature, see e.g., Bauer et al. (2001); Damon and Guillas (2002); Feister and Balzer (1991); Huerta et al. (2004); Kumar and Ridder (2010); McMillan et al. (2005); Sahu and Bakar (2012a); Sahu and Mardia (2005a,b); Sahu et al. (2009, 2011); Sousa et al. (2009); Stroud et al. (2001); West and Harrison (1997) and Zidek et al. (2012). Some of these papers also consider space-time modelling for the purposes of forecasting. However, the methods proposed in these articles are not able to handle the computational burden associated with large space-time data sets that we model in this chapter for forecasting purposes.

For point referenced spatial data from a large number of locations, exact likelihood based inference becomes unstable and infeasible since that involves computing quadratic forms and determinants associated with a high dimensional variance-covariance matrix (Stein (2008)). Besides the problem of storage (Cressie and Johannesson 2008), matrix inversion, at each iteration of the model fitting algorithm, such as the EM algorithm, is of $O(n^3)$ computational complexity, which is prohibitive, where n is a large number of modeled spatial locations. This problem also arises in the evaluation of the joint or conditional distributions in Gaussian process based models under a hierarchical Bayesian setup, see e.g., Banerjee et al. (2004). To tackle this problem, we develop a Bayesian forecasting method based on a model recently developed by Sahu and Bakar (2012b), using Gaussian predictive process (GPP) approximation method for the underlying spatial surface, see Banerjee et al. (2008). Throughout this chapter, for convenience, we shall use the acronym GPP to also denote the modelling method based on the GPP approximation.

Forecasting using hierarchical Bayesian models is further limited by the lack of suitable software packages. There are a few available packages for forecasting using variants of the dynamic linear models (West and Harrison 1997), see e.g., Petris et al. (2010). However, these packages do not allow incorporation of rich spatial covariance structure for the modelled data. On the other hand, `spBayes`, a recently developed spatial data analysis package, developed by Finley et al. (2007), can model short-length time series data by treating those as multivariate spatial data, but it is not really intended to handle large volume of spatio-temporal data that can be analyzed using the `spTimer` package developed by Bakar and Sahu (2014).

This chapter develops forecasting methods for three Bayesian hierarchical models that

have been implemented in `spTimer`. The first of these is an independent in time Gaussian process (GP) based regression model that is simple to implement and is often regarded as a starting model. The second is the hierarchical auto-regressive model developed by Sahu et al. (2007), that has been shown to be better in out of sample validation than some versions of dynamic linear models (Sahu and Bakar 2012a) and also a wide class of models (Cameletti et al. 2011). The third and final forecasting method is the one based on the GPP approximation method mentioned above. These methodological developments are then used to augment the `spTimer` package with the forecasting modules that can be used in a wide variety of applications in space-time data analysis.

Another objective of the chapter is to rigorously compare the Bayesian forecasts obtained from the three models. Towards this end we develop Markov chain Monte Carlo (MCMC) implementation methods for several forecast calibration measures and diagnostic plots that have been proposed to compare the skills of the Bayesian forecast distributions, see e.g., Gneiting et al. (2007). The measures include: the continuous ranked probability score which is an integrated distance between the forecasts and the corresponding observations, the hit and false alarm rates and the nominal coverage. The diagnostic plots include the probability integral transform and a marginal calibration plot that is used to calibrate the equality of the forecast and the actual observations, see Section 1.5. These measures and plots enable us to compare the implied Bayesian forecast distributions fully – not just their specific characteristics, e.g., the mean forecast, as would be done by simple measures such as the root mean square error and the mean absolute error.

A substantial application on an air pollutant, ground-level ozone, illustrates the forecasting methods of this chapter. Ground-level ozone is a pollutant that is a significant health risk, especially for children with asthma and vulnerable adults with respiratory problems. It also damages crops, trees and other vegetation. It is a main ingredient of urban smog. Because of these harmful effects, air pollution regulatory authorities are required by law to monitor ozone levels and they also need to forecast in advance, so that at risk population can take necessary precaution in reducing their exposure. In the United States (US), a part of which is our study region in this chapter, the forecasts are issued, often, up to 24-hours in advance by various mass-media, e.g. newspapers and also the website `airnow.gov`. However, ozone concentration levels, and also other air pollutants, are regularly monitored by only a finite number of sites. Data from these sparse network of monitoring sites need to be processed for developing accurate forecasts. In this chapter, we compare the forecasts of ground-level ozone, based on three models using a three-week test data set on daily maximum ozone concentration levels observed over a large region in the Eastern US.

The rest of this chapter is organized as follows: Section 1.3 describes the validation data set we use in this chapter with some summary statistics. In Section 1.4 we develop forecasting methods based on three recently proposed Bayesian spatio-temporal models. Section 1.5 discusses several useful and important forecast calibration methods and develops their MCMC implementation techniques. These are used to compare the forecasting methods with a smaller subset of the full validation data set in Section 1.6. This investigation finds that the GPP model is fast and it performs the best. Subsequently, this model is used in Section 1.7 to analyze and forecast for the full eastern US data set. Finally, Section 1.8 concludes with a few summary remarks.

1.3 Test Data Set

The forecasting models proposed in this chapter will be tested using *daily* ozone concentration data for the 3-week period, June 24 to July 14 in 2010. A daily observation, measured in units of parts per billion (ppb), is the maximum of 24 averages in a day where each average is based on hourly ozone concentration readings from 8 consecutive hours. In this chapter, we use daily data from 639 monitoring sites in the eastern US. We aim to perform forecast validation for completely out of sample data from sites that we do not use for modelling at all. Hence, we set aside data from 62 randomly chosen sites (roughly 10%) for validation purposes. Figure 1.1 provides a map of these validation sites and the remaining 577 sites, data from which are used for modelling.

We perform forecast validation for seven moving windows of data from July 8 to July 14. July 8 is taken to be the earliest day for forecast validation that allows modelling of data for 14 days from June 24 to July 7. We also compare the next day forecasts based on modelling data from just seven previous days that complete a weekly cycle. Thus, for example, for forecasting for July 8 we use data from July 1 to 7.

Often, see, e.g. `airnow.gov`, a deterministic model, known as the community multi-scale air quality (CMAQ) model, is used for forecasting levels of ozone concentration and other air pollutants such as particulate matter. The CMAQ model in forecasting mode, known as Eta CMAQ, is based on emission inventories, meteorological information, and land use, and it produces gridded forecasts, up to two days in advance, for average ozone concentration levels at each cell of a 12 square-kilometer grid covering the whole of the continental US (Ching and Byun, 1999). However, these output are well-known to produce biased forecasts and to reduce this bias, in this chapter, we develop statistical models that are able to improve the Eta CMAQ forecasts by refining those in the light of the observed monitoring data. Incorporation of gridded CMAQ forecasts in a spatial model for point referenced monitoring data poses a spatial misalignment problem that is well known in the literature, see for example, Fuentes and Raftery (2005); Jun and Stein (2004); Lorence (1986). To incorporate the Eta CMAQ output, the hierarchical models are set up as spatio-temporal downscaler models, first implemented by Sahu et al. (2009), and then generalised by Berrocal et al. (2010a,b), and Zidek et al. (2012). We use the forecasts for daily maximum 8-hour average CMAQ ozone concentration for the grid cell covering the monitoring site as the single co-variate, following Sahu et al. (2009).

Many meteorological variables such as the daily maximum temperature are important predictors of ozone levels, see e.g., Sahu et al. (2007). However, the meteorological variables no longer remain significant if the model for ozone levels also includes output of the CMAQ model, see e.g., Sahu and Bakar (2012a). Moreover, direct inclusion of the meteorological variables in an ozone concentration forecasting model will also require forecasting of the meteorological variables in the first place. The models proposed in this chapter avoid this, although we note that the CMAQ forecasts already include future values of the meteorological variables that have been used as model inputs.

Out of the 13,419 observations from 639 sites for 21 days, 299 ($\approx 2.23\%$) are missing. Our Bayesian models automatically estimate those using standard methods. Table 1.1 provides the summary statistics for ozone levels and Eta CMAQ output, where it is seen that the Eta CMAQ forecasts are upwardly biased, although the medians seem to be close. Figure 1.2 investigates this further by providing side by side boxplots for each of 21 days for both the

observed and the Eta CMAQ forecasted ozone levels. This figure also shows that the data set includes an episode of high ozone levels during days 12-16, which corresponds to July 5-9, just after the 4th of July celebrations in the US. This episode of high ozone levels provides an opportunity to model and forecast when their demand is likely to be higher than usual.

1.4 Forecasting Methods

1.4.1 Preliminaries

We first define the generic notations that we need and use throughout the chapter. Let t denote the time where $t = 1, \ldots, T$ and T is the total number of time units. Let $Y(\mathbf{s}_i, t)$ denote the observed point referenced data at location $\mathbf{s}_i$ and at time t for $i = 1, \ldots, n$ where n is the total number of locations. Modelling the data on the original scale, as noted by many authors, see e.g., Sahu et al. (2007) is prohibitive due to the instability in variance that often leads to negative forecasts. In this chapter, we model data on the square-root scale, denoted by $Z(\mathbf{s}_i, t)$, that encourages symmetry and normality, see e.g., Sahu et al. (2007), but report all forecasts and predictions on the original scale, Y, for ease of interpretation by practitioners, although this may increase the mean square error of the forecasts. We also note that other variance stabilizing transformations such as log and the more general Box-Cox transformation can also be adopted depending on the nature of the problem, and finally, the methods we describe below can also be used if a variance stabilising transformation is not needed in the first place. MCMC methods enable us to estimate the uncertainties of the forecasts on the original scale.

Let $O(\mathbf{s}_i, t)$ be the true value corresponding to $Z(\mathbf{s}_i, t)$ at site $\mathbf{s}_i$, $i = 1, ..., n$ at time t. Let $\mathbf{Z}_t = (Z(\mathbf{s}_1, t), ..., Z(\mathbf{s}_n, t))'$ and $\mathbf{O}_t = (O(\mathbf{s}_1, t), ..., O(\mathbf{s}_n, t))'$. We shall denote all the observed data by $\mathbf{z}$, and $\mathbf{z}^*$ will denote all the missing data. Similarly, $\mathbf{O}$ will denote all $\mathbf{O}_t$, for $t = 1, \ldots, T$. Let $N = nT$ be the total number of observations to be modeled.

For forecasting purposes it is of interest to obtain the one-step ahead forecast distribution for noisy data $Y(\mathbf{s}_0, T + 1)$ on the original scale, and not for $O(\mathbf{s}_0, T + 1)$, since our objective is to compare the forecasting methods by validation of the noisy data itself, where $\mathbf{s}_0$ denotes any particular, monitored or un-monitored, site of interest. In the sequel, we shall obtain the marginal one-step ahead forecasts at a number of sites, say m. The joint one-step ahead forecast distribution for the m forecasts can also be developed for the models described below, but are not of interest here.

We also assume that, in general, there are p covariates, including the intercept, denoted by the $n \times p$ matrix $\mathbf{X}_t$. Some of these covariates may vary in both space and time. The notation $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)'$ will be used to denote the $p \times 1$ vector of regression coefficients. We shall use the generic notation $\boldsymbol{\theta}$ to denote all the parameters.

1.4.2 Forecasting Using GP Models

The spatio-temporal linear regression model is defined by:

$$\mathbf{Z}_t = \mathbf{O}_t + \boldsymbol{\epsilon}_t, \tag{1.1}$$

$$\mathbf{O}_t = \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\eta}_t \tag{1.2}$$

where $\boldsymbol{\epsilon}_t = (\epsilon(\mathbf{s}_1, t), ..., \epsilon(\mathbf{s}_n, t))' \sim N(\mathbf{0}, \sigma^2_\epsilon \mathbf{I}_n)$ is the independently distributed white noise error with variance σ^2_ϵ also known as the nugget effect, and $\mathbf{I}_n$ is the $n \times n$ identity matrix. The term $\boldsymbol{\eta}_t = (\eta(\mathbf{s}_1, t), ..., \eta(\mathbf{s}_n, t))'$ is an independent, over time, realization of a spatial Gaussian process with zero mean and the correlation function $\kappa(d; \phi, \nu)$, often assumed to be a member of the Matérn family, see e.g., Banerjee et al. (2004), is allowed to depend on two unknown parameters ϕ and ν describing the correlation at distance d. In effect, this implies that the smooth process, $O(\mathbf{s}, t)$ is assumed to be isotropic and stationary. Note that this does not necessarily imply the same assumptions for the un-transformed noisy data, Y since other hierarchical model components will contribute to the overall space-time correlation function.

Thus we assume that $\boldsymbol{\eta}_t \sim N(\mathbf{0}, \Sigma_\eta)$, where $\Sigma_\eta = \sigma^2_\eta S_\eta$ and $(S_\eta)_{ij} = \kappa(||\mathbf{s}_i - \mathbf{s}_j||; \phi, \nu)$, $i, j = 1, ..., n$; σ^2_η is the site invariant common variance and $\kappa(.; \phi, \nu)$ is the spatial correlation that depends on spatial decay, ϕ, and smoothness, ν, parameters. For convenience, in this chapter we use the exponential covariance function to model spatial dependence as:

$$\Sigma_\eta = \sigma^2_\eta S_\eta = \sigma^2_\eta \exp(-\phi_\eta D)$$

where, $\phi_\eta > 0$ is a spatial correlation decay parameter, and D is the matrix that has elements d_{ij}, that is the distance between sites $\mathbf{s}_i$ and $\mathbf{s}_j$, $i, j = 1, ..., n$. Here, and in the sequel, the matrix exponential is used to mean element-wise exponentiation, i.e. $(\Sigma_\eta)_{ij} = \sigma^2_\eta \exp(-\phi_\eta d_{ij})$, $i, j = 1, ..., n$. The `spTimer` package provides options to implement using the full Matérn family. The error distributions of $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ are assumed to be independent of each other. For future reference, let $\boldsymbol{\theta}$ denote all the parameters, $\boldsymbol{\beta}$, σ^2_ϵ, σ^2_η, and ϕ. We assume independent normal prior distribution with zero mean and a large variance, 10^{10}, to achieve vague prior specification, for the components of $\boldsymbol{\beta}$. The inverse of the variance components σ^2_ϵ σ^2_η are given independent gamma distribution with mean a/b and variance a/b^2. Although any suitable values for a and b can be chosen, we have chosen $a = 2$ and $b = 1$ to have a proper prior distribution for any variance component that will guarantee a proper posterior distribution. We assume uniform prior distribution for the correlation decay parameters ϕ, although many other choices are possible. Full details are provided in the `spTimer` package, see Bakar and Sahu (2014).

To obtain the 1-step ahead forecast distribution of $Z(\mathbf{s}_0, T+1)$ at any un-observed location $\mathbf{s}_0$ at time $T+1$, we first note that:

$$\begin{aligned} Z(\mathbf{s}_0, T+1) &= O(\mathbf{s}_0, T+1) + \epsilon(\mathbf{s}_0, T+1), \\ O(\mathbf{s}_0, T+1) &= \mathbf{x}'(\mathbf{s}_0, T+1)\boldsymbol{\beta} + \eta(\mathbf{s}_0, T+1). \end{aligned}$$

The 1-step ahead forecast distribution is the posterior predictive distribution of $Z(\mathbf{s}_0, T+1)$ given $\mathbf{z}$ and is given by:

$$\begin{aligned} \pi(Z(\mathbf{s}_0, T+1)|\mathbf{z}) = \int \pi(Z(\mathbf{s}_0, T+1)|\boldsymbol{\theta}, \mathbf{O}, O(\mathbf{s}_0, T+1), \mathbf{z}) \pi(O(\mathbf{s}_0, T+1)|\boldsymbol{\theta}, \mathbf{z}) \\ \pi(\boldsymbol{\theta}, \mathbf{O}|\mathbf{z}) dO(\mathbf{s}_0, T+1) d\mathbf{O} d\boldsymbol{\theta}, \end{aligned} \tag{1.3}$$

where $\pi(\boldsymbol{\theta}, \mathbf{O}|\mathbf{z})$ denotes the joint posterior distribution of $\mathbf{O}$ and $\boldsymbol{\theta}$. Note that $\pi(Z(\mathbf{s}_0, T+1)|\boldsymbol{\theta}, \mathbf{O}, O(\mathbf{s}_0, T+1), \mathbf{z}) = \pi(Z(\mathbf{s}_0, T+1)|\boldsymbol{\theta}, \mathbf{O}, O(\mathbf{s}_0, T+1))$ due to the conditional independence of $Z(\mathbf{s}_0, T+1)$ and $\mathbf{Z}$ given $\mathbf{O}$. Similarly, $O(\mathbf{s}_0, T+1)$ does not depend

on $\mathbf{Z}$ given $\boldsymbol{\theta}$, hence in the following development we replace $\pi(O(\mathbf{s}_0, T+1)|\boldsymbol{\theta}, \mathbf{z})$ by $\pi(O(\mathbf{s}_0, T+1)|\boldsymbol{\theta})$.

Now the 1-step ahead forecast distribution (1.3) is constructed by composition sampling as follows. Assume that, at the jth MCMC iteration, we have posterior samples, $\boldsymbol{\theta}^{(j)}$ and $\mathbf{O}^{(j)}$. Then we first draw, $O^{(j)}(\mathbf{s}_0, T+1)$ from $N(\mathbf{x}'_{T+1}\boldsymbol{\beta}^{(j)}, \sigma_\eta^{2(j)})$. Finally, we draw $Z^{(j)}(\mathbf{s}_0, T+1)$ from $N(O^{(j)}(\mathbf{s}_0, T+1), \sigma_\epsilon^{2(j)})$.

Note that in the above paragraph, we use the marginal distribution instead of the conditional distribution because we have already obtained the conditional distribution given observed information up to time T at the observation locations $\mathbf{s}_1, ..., \mathbf{s}_n$, and at the future time $T+1$ there is no further new information to condition on except for the new regressor values $\mathbf{x}(\mathbf{s}_0, T+1)$ in the model. However, the conditional distribution can be used instead if it is so desired. To do this, we note that the joint distribution of $\mathbf{O}_{T+1} = (O(\mathbf{s}_1, T+1), ..., O(\mathbf{s}_n, T+1))'$ is simply given by $N(\mathbf{X}_{T+1}\boldsymbol{\beta}, \Sigma_\eta)$, according to (1.2). Similarly, we construct the joint distribution of $O(\mathbf{s}_0, T+1)$ and $\mathbf{O}_{T+1}$ from which we obtain the conditional distribution $\pi(O(\mathbf{s}_0, T+1)|\mathbf{O}_{T+1})$, that is Gaussian with mean

$$\mathbf{x}(\mathbf{s}_0, T+1)\boldsymbol{\beta} + S_{\eta,12}S_\eta^{-1}(\mathbf{O}_{T+1} - \mathbf{X}_{T+1}\boldsymbol{\beta})$$

and variance

$$\sigma_\eta^2(1 - S_{\eta,12}S_\eta^{-1}S_{\eta,21}),$$

where $S'_{\eta,21} = S_{\eta,12} = e^{-\phi\,\mathbf{d}_{12}}$ and $\mathbf{d}_{12} = (||\mathbf{s}_1 - \mathbf{s}_0||, \ldots, ||\mathbf{s}_n - \mathbf{s}_0||)'$.

For forecasting at any observed site $\mathbf{s}_i$ for any $i = 1, \ldots, n$ at time $T+1$ we note that:

$$Z(\mathbf{s}_i, T+1) = O(\mathbf{s}_i, T+1) + \epsilon(\mathbf{s}_i, T+1),$$
$$O(\mathbf{s}_i, T+1) = \mathbf{x}'(\mathbf{s}_i, T+1)\boldsymbol{\beta} + \eta(\mathbf{s}_i, T+1).$$

These two identities make it clear that the 1-step ahead forecast distribution of $Z(\mathbf{s}_i, T+1)$ given $\mathbf{z}$ can simply be constructed by iteratively sampling from the conditional distribution $O^{(j)}(\mathbf{s}_i, T+1) \sim N(\mathbf{x}'(\mathbf{s}_i, T+1)\boldsymbol{\beta}^{(j)}, \sigma_\eta^{2(j)})$ and then $Z^{(j)}(\mathbf{s}_i, T+1)$ from the normal distribution with mean $O^{(j)}(\mathbf{s}_i, T+1)$ and variance $\sigma_\epsilon^{2(j)}$. Finally, $Z^{(j)}(\mathbf{s}_i, T+1)$ values are transformed back to the original scale giving MCMC samples $Y^{(j)}(\mathbf{s}_i, T+1)$.

1.4.3 Forecasting Using AR Models

Here we briefly describe the forecasting method based on the hierarchical AR models proposed by Sahu et al. (2007, 2009). The model equations are given by:

$$\mathbf{Z}_t = \mathbf{O}_t + \boldsymbol{\epsilon}_t, \tag{1.4}$$

$$\mathbf{O}_t = \rho\mathbf{O}_{t-1} + \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\eta}_t \tag{1.5}$$

where $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ have been previously specified, and ρ is a scalar denoting site-invariant temporal correlation. These auto-regressive models also need an initialization for $\mathbf{O}_0$ which we assume to be independently normally distributed with mean $\boldsymbol{\mu}$ and the covariance matrix $\sigma^2 S_0$ where the correlation matrix S_0 is obtained using the exponential correlation function with a new decay parameter ϕ_0. These additional parameters and initialization random variables are added to $\boldsymbol{\theta}$ and $\mathbf{O}$ respectively.

The temporal correlation, ρ in (1.5), for the smooth process $O(\mathbf{s}, t)$, has been assumed to be site invariant given the effects of the spatially and temporally varying covariates and the spatio-temporal intercepts $\eta(\mathbf{s}, t)$. A site specific temporal correlation will perhaps be needed, though not pursued here, if only the last two terms are omitted from the model. We also assume, for stationarity, that $|\rho| < 1$.

We assume the same set of prior distributions for $\boldsymbol{\beta}$, the variance components σ^2_ϵ and σ^2_η, and the correlation decay parameters ϕ as previously discussed in Section 1.4.2. For the additional ρ parameter we again provide a normal prior distribution with zero mean and a large variance (10^{10} in our implementation), but we restrict the prior distribution in the range $|\rho| < 1$.

Under the AR models the predictive distribution of $Z(\mathbf{s}_0, T+1)$ is determined by $O(\mathbf{s}_0, T+1)$. Following (1.5), we see that $O(\mathbf{s}_0, T+1)$ follows the normal distribution with site invariant variance σ^2_η and mean $\rho O(\mathbf{s}_0, T) + \mathbf{x}'(\mathbf{s}_0, T+1)\boldsymbol{\beta}$. This depends on $O(\mathbf{s}_0, T)$ and as a result, due to this auto-regressive nature, we have to determine all the random variables $O(\mathbf{s}_0, k)$, for $k = 0, \ldots, T$. In order to simulate, all these random variables, we first simulate from the conditional distribution of $O(\mathbf{s}_0, 0)$ given $\mathbf{O}_0$, which is a univariate normal distribution. Then, at the jth MCMC iteration we sequentially simulate $O^{(j)}(\mathbf{s}_0, k)$ given $O^{(j)}(\mathbf{s}_0, k-1)$ for $k = 1, \ldots, T+1$ from the normal distribution with mean $\rho^{(j)} O^{(j)}(\mathbf{s}_0, k-1) + \mathbf{x}'(\mathbf{s}_0, k)\boldsymbol{\beta}^{(j)}$ and variance $\sigma^{2(j)}_\eta$. For forecasting at any observation location $\mathbf{s}_i$ we draw $Z^{(j)}(\mathbf{s}_i, T+1)$ from the normal distribution with mean $\rho^{(j)} O^{(j)}(\mathbf{s}_i, T) + \mathbf{x}'(\mathbf{s}, T+1)\boldsymbol{\beta}^{(j)}$ and variance $\sigma^{2(j)}_\epsilon$. For further details regarding prediction see, Sahu *et al.* (2007). Now these Z values are transformed back to the original scale, Y as in the case of GP models.

1.4.4 Forecasting Using the GPP Model

The models described in Section 1.4.3 assume the AR model for the true values of the modeled response $\mathbf{O}_t$. Sahu and Bakar (2012b) modified this model so that the modified version does not assume a true level $O(\mathbf{s}_i, t)$ for each $Z(\mathbf{s}_i, t)$ but instead assumes a space-time random-effect denoted by $\eta(\mathbf{s}_i, t)$. It then assumes an AR model for these space-time random effects. For a large number of spatial locations the top level space-time random effect term will lead to the estimation problem discussed in the Introduction. Hence, we use the predictive process approximation technique (Sahu and Bakar, 2012b). Here the main idea is to define the random effects $\eta(\mathbf{s}_i, t)$ at a smaller number of locations, m say, where $m << n$, called the knots, and then use kriging to predict those random effects at the data locations.

The top level model is written as:

$$\mathbf{Z}_t = \mathbf{X}_t\boldsymbol{\beta} + \tilde{\boldsymbol{\eta}}_t + \boldsymbol{\epsilon}_t,\ t = 1, ..., T \tag{1.6}$$

where $\boldsymbol{\epsilon}_t$ has been previously specified. The space-time process $\tilde{\boldsymbol{\eta}}_t$ is specified by:

$$\tilde{\boldsymbol{\eta}}_t = A\mathbf{w}_t \tag{1.7}$$

with $A = CS_w^{-1}$ where S_w is the correlation matrix of w_t with ijth element, that corresponds to two locations $\mathbf{s}_i$ and $\mathbf{s}_j$, is given by $\exp(-\phi_w||\mathbf{s}_i - \mathbf{s}_j||)$. The elements of the $n \times m$ matrix C are also calculated using this correlation function.

In the next stage of the modelling hierarchy the AR model is assumed as:

$$\mathbf{w}_t = \rho\, \mathbf{w}_{t-1} + \boldsymbol{\xi}_t, \tag{1.8}$$

where $\boldsymbol{\xi}_t \sim N(\mathbf{0}, \sigma_w^2 S_w)$. Again, we assume that $\mathbf{w}_0 \sim N(\mathbf{0}, \sigma^2 S_0)$, where the elements of the covariance matrix S_0 are obtained using the correlation function, $\exp(-\phi_0 d_{ij})$, which is the same correlation function used previously but with a different decay parameter ϕ_0. The Bayesian model specification here is completed by assuming the same set of prior distributions as noted in the previous two sub-sections.

At an un-observed location $\mathbf{s}_0$, the 1-step ahead Bayesian forecast is given by the predictive distribution of $Z(\mathbf{s}_0, T+1)$, that we determine from equation (1.6) replacing t with $T+1$. Thus, the 1-step ahead forecast distribution has variance σ_ϵ^2 and mean $\mathbf{x}'(\mathbf{s}_0, T+1)\boldsymbol{\beta} + \tilde{\eta}(\mathbf{s}_0, T+1)$, where $\tilde{\eta}(\mathbf{s}_0, T+1)$ is obtained analogous to (1.7) as:

$$\tilde{\eta}(\mathbf{s}_0, T+1) = S_{w,12} S_w^{-1} \mathbf{w}_{T+1}$$

where $S_{w,12} = e^{-\phi_w \mathbf{d}_{12}}$ and $\mathbf{w}_{T+1}$ is obtained from (1.8).

Thus, at each MCMC iteration, we draw a forecast value $Z^{(j)}(\mathbf{s}_0, T+1)$ from this normal distribution. Forecasting at the observation sites $\mathbf{s}_1, \ldots, \mathbf{s}_n$ is performed by noting that, according to (1.6),

$$\mathbf{Z}_{T+1} = \mathbf{X}_{T+1}\boldsymbol{\beta} + \tilde{\boldsymbol{\eta}}_{T+1} + \boldsymbol{\epsilon}_{T+1},$$

with $\tilde{\boldsymbol{\eta}}_{T+1} = A\mathbf{w}_{T+1}$ and $\boldsymbol{\epsilon}_{T+1} \sim N(\mathbf{0}, \sigma_\epsilon^2 I_n)$. Thus, as before $\mathbf{w}_{T+1}$ is obtained from (1.8) and MCMC sampling from the forecast distribution of $Z(\mathbf{s}_i, T+1)$ for $i = 1, \ldots, n$ is straightforward. Again these Z samples are transformed back to the original scale Y, which we use for forecast calibration purposes.

1.5 Forecast Calibration Methods

The three model based forecasting methods discussed in the previous section must be compared using suitable methods. Predictive Bayesian model selection methods are appropriate for comparing Bayesian models, see e.g., Gelfand and Ghosh (1998). However, the main objective of this chapter is forecasting and hence we compare the models on the basis of their forecasting performance. There is a large literature on forecast comparison and calibration methods, see e.g., Gneiting et al. (2007) and the references therein. In the Bayesian context of this chapter, we need to compare the entire forecast predictive distribution, not just summaries like the mean, since forecasting is the primary goal here.

To simplify notation, suppose that $y_i, i = 1, \ldots, m$ denote the m hold-out validation observations that have not been used in model fitting. Note that we use a single indexed notation y_i, instead of the more elaborate $y(\mathbf{s}, t)$ used previously. Clearly, some of these validation observations may be future observations at the modelling sites or completely at new sites – what's important here is that those must not have been used for model fitting. Let $F_i(y)$ denote the model based forecast predictive distribution function of Y_i, the random variable whose realized value is y_i. Thus, $F_i(y)$ is one of the three forecast predictive distributions, corresponding to one of the three models: GP, AR and GPP, described previously in Section 1.4. Let $G_i(y)$ be the true unknown forecast predictive distribution function, which the $F_i(y)$ is trying to estimate. The problem here is to calibrate $F_i(y)$ for $G_i(y)$, $i = 1, \ldots, m$, conditional on the modeled data, $\mathbf{y}$ or equivalently its transformed value

z. Let $\hat{y}_i$ be the intended forecast for y_i, i.e., $\hat{y}_i$ mean or median of the forecast distribution $F_i(y)$, estimated using the mean or median of the MCMC samples $y_i^{(j)}, j = 1, \ldots, J$, where J is a large number. In our implementation in Sections 1.6 and 1.7, we have taken $J = 15,000$ after discarding first 5000 iterations, that was deemed to be adequate, to mitigate the effect of initial values. Below, we describe seven popular forecast calibration and diagnostic methods and develop their computation methods using MCMC.

1. The Root Mean Square Error (RMSE) is defined by:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}.$$

   It is perhaps the most popular forecast comparison criterion and the method with the smallest RMSE value is preferred.

2. Sometimes the Mean Absolute Error, defined by,

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}_i|$$

   is preferred to the RMSE. Both the RMSE and the MAE are on the original unit of the data and they provide a quick check on the magnitude of the errors in the forecasts.

3. The continuous ranked probability score (CRPS) is a proper scoring rule for comparing forecasts, (Gneiting *et al.*, 2007) and is defined by:

$$\text{crps}(F, y) = E_F|Y - y| - \frac{1}{2}E_F|Y - Y'|$$

   where Y and Y' are independent copies of a random variable with distribution function F and finite first moment. With m hold-out observations, we calculate the overall measure, given by

$$\text{CRPS} = \frac{1}{m}\sum_{i=1}^{m}\text{crps}(F_i, y_i).$$

   We estimate the CRPS using J MCMC samples $y_i^{(j)}, j = 1, \ldots, J$, as follows. We first obtain,

$$\hat{\text{crps}}(F_i, y_i) = \frac{1}{J}\sum_{j=1}^{J}|y_i^{(j)} - y_i| - \frac{1}{2J^2}\sum_{j=1}^{J}\sum_{k=1}^{J}|y_i^{(j)} - y_i^{(k)}|,\ i = 1, \ldots, m,$$

   and then the overall average CRPS is estimated as:

$$\hat{\text{CRPS}} = \frac{1}{m}\sum_{i=1}^{m}\hat{\text{crps}}(F_i, y_i).$$

   Again, the model with the smallest CRPS value is the preferred choice.

4. The nominal coverage is defined by

$$NCOV = \frac{1}{m}\sum_{i=1}^{m} 1\left(l_i \leq y_i \leq u_i\right)$$

where l_i and u_i are, respectively, the lower and upper limits of a given predictive interval for y_i and $1(A) = 1$ if A is true and 0 otherwise. Good forecasting methods must have the nominal coverage close to their true value so that the uncertainties in the forecast distributions are correct, not only their central tendencies as measured by the RMSE or the MAE. In practice, the limits l_i and u_i are estimated using the appropriate quantiles of the MCMC samples $y_i^{(j)}, j = 1, \ldots, J$. For example, for 95% prediction intervals, these are estimated to be the 2.5th and 97.5th percentile of $y_i^{(j)}, j = 1, \ldots, J$, respectively.

5. The concentration of the forecast distribution is compared using the sharpness diagram. A sharpness diagram plots the widths of the (m) forecast intervals as side-by-side boxplots where each boxplot is for a particular forecasting method. The forecasting method that produces narrower width forecast intervals, but with good nominal coverages, is preferred.

6. The hit and false alarm rates are also considered by many authors for forecast comparison purposes, see e.g., Sahu et al. (2009). These rates are defined for a given threshold value y_0, which is often the value beyond which the pollutant is considered to be very dangerous. Hit is defined as the event where both the validation observation, y_i and the forecast, $\hat{y}_i$, for it are either both greater or less than the threshold y_0. The false alarm, on the other hand, is defined as the event where the actual observation is less than y_0 but the forecast is greater than y_0. Thus we define:

$$\text{Hit rate}(y_0) = \frac{1}{m}\sum_{i=1}^{m}\left\{1\left(y_i > y_0 \,\&\, \hat{y}_i > y_0\right) + 1\left(y_i < y_0 \,\&\, \hat{y}_i < y_0\right)\right\},$$

$$\text{False alarm}(y_0) = \frac{1}{m}\sum_{i=1}^{m} 1(y_i < y_0 \,\&\, \hat{y}_i > y_0).$$

Forecasting methods with high hit rates and low false alarm rates are preferred.

7. Many authors have proposed the probability integral transform (PIT) diagram as a necessary diagnostic tool for comparing forecasts. For each hold-out observation y_i, the PIT value is calculated as

$$p_i = F_i(y_i), i = 1, \ldots, m.$$

If the forecasts are ideal, and F_i is continuous, then p_i has a uniform distribution. The PIT diagram is simply an histogram of the p_i's, $1, \ldots, m$. Using MCMC samples, p_i is estimated by:

$$\hat{p}_i = \frac{1}{J}\sum_{j=1}^{J} 1\left(y_i^{(j)} \leq y_i\right),\ i = 1, \ldots, m.$$

8. A marginal calibration plot (MCP) is used to calibrate the equality of the forecast and the actual value, and is constructed as follows. First, take a grid, $y_k, k = 1, \ldots, K$, say, covering the domain of the forecast distribution. For each of those y_k values, calculate

$$\hat{G}(y_k) = \frac{1}{m} \sum_{i=1}^{m} 1\left(y_i \leq y_k\right).$$

Now calculate

$$\bar{F}(y_k) = \frac{1}{m} \sum_{i=1}^{m} \hat{F}_i(y_k),$$

where

$$\hat{F}_i(y_k) = \frac{1}{J} \sum_{j=1}^{J} 1\left(y_i^{(j)} \leq y_k\right),\ i = 1, \ldots, m.$$

Now, the plot of the differences $\bar{F}(y_k) - \hat{G}(y_k)$ against y_k, for $k = 1, \ldots, K$ is the desired MCP. If the forecasts are good, only minor fluctuations about 0 are expected. Thus, a forecast distribution whose MCP stays closest to 0 will be the preferred choice.

1.6 Results from a Smaller Data Set

The computation of all the forecast calibration methods for the whole eastern US data set is prohibitive because of the big-n problem as mentioned in the Introduction, see also the next section. Due to this reason, we compare all three forecasting methods using a subset of the whole eastern US data, consisting of four states: Illinois, Indiana, Ohio and Kentucky. There are 147 ozone monitoring sites in these states, see Figure 1.3. We set aside data from 20 randomly selected sites for validation purposes. As mentioned in Section 1.3 we validate for seven days from July 8 to 14.

For the GPP model the knot size is taken as 107, that has been chosen from a sensitivity analysis similar to the ones reported in Sahu and Bakar (2012b). We also have performed a number of different sensitivity analysis with respect to the choice of the hyper-parameter values in the prior distribution, tuning of the MCMC algorithms and also have monitored convergence using trace plots and the package CODA (Plummer et al. 2006). We omit all those details for brevity.

All three models are fitted using the MCMC code developed within the `spTimer` package. As mentioned in Section 1.5, MCMC algorithms are run for a total of 20,000 iterations of which first 5,000 are discarded to mitigate the effect of starting values. The algorithms run very fast taking only about 9, 16 and 3 minutes for the GP, AR and GPP models respectively in a 2.6Ghz personal computer with 4GB of RAM running 32 bit Windows operating system. Thus it is quite fast to fit the models and produce the forecasts using all the models.

The RMSE and the MAE for the seven validation days are plotted in Figure 1.4. As expected, the RMSE and the MAE are very similar (compare the columns). But we do not see a large difference between modelling seven and fourteen days data (compare the rows). The RMSE and MAE of the GP and AR models are very similar and they both have worse performance than the GPP model. This is also confirmed by the CRPS values, see Table 1.2. The nominal coverages, of the 50% and 95% forecast intervals, provided in Table 1.3, however, are not able to compare the forecasting methods; but those show that

all three methods are adequate. The average widths of the forecast intervals, see Table 1.4, clearly shows that the GPP model is the best. This is also confirmed by the sharpness diagram, see Figure 1.5.

The hit and false alarm rates using all seven validation days data are provided in Table 1.5. All three models perform very well. The hit rate increases as the threshold value increases and it is actually 100% when we use the threshold value of 85. The false alarm rate decreases to zero as the threshold value is increased from 65 to 75 ppb. These rates, however, do not discriminate between the three different forecasting methods.

The PIT diagrams for all three forecasting methods for the 14 days data modelling case are provided in Figure 1.6. Here also the GPP model is the preferred choice since its histogram is more uniform than the other two. The same diagrams based on modelling 7-days data showed similar patterns and hence have been omitted.

Figure 1.7 provides the marginal calibration plots of all three models using data for 7 and 14 days. Here also the GPP model performs better than its rivals and the performance is differentiated better in the case of modelling data for 14 days. In addition, calibration improves towards the upper tail of the distribution that assures that the models are able to forecast high levels of ozone concentration quite accurately. In conclusion, we find that the GPP model is best for forecasting among the three methods considered here.

A further remark regarding the performances of the AR and GPP models is appropriate. As with any approximation it can be expected that the approximate GPP model to perform worse than the full AR model. However, the GPP model in Section 1.4.4 cannot be seen as a true approximation for the AR model in Section 1.4.3 due to the inclusion of the auto-regressive term in two very different manners: one at the top level $\mathbf{O}_t$ in (1.5) and the other at the random effect level $\mathbf{w}_t$ in (1.7). Thus the AR and GPP models are very different and it is not surprising that we do not see any strict one-way performance ordering in our examples.

1.7 Analysis of the Full Eastern US Data Set

As mentioned in Section 1.3, we use data from 577 sites to fit our models and the data from 62 sites are set aside for validation purposes. The implementation of the GPP model requires the selection of the number of knots. Using a similar sensitivity study that we have used in Sahu and Bakar (2012b), but with the forecast RMSE, as the criterion we compare the GPP model with 68, 105, 156 and 269 knots which were all inside the land boundary of the United States. The forecast RMSE improved with the increasing knot sizes, but only slightly when the size increased to 269 from 156. Henceforth, we adopt 156 as the knot size that implies a much smaller computational burden.

For the model fitting (a data set with 14 days data) and forecasting using 20,000 iterations, using the same personal computer as in the previous section, we have estimated that the GP model will take about 40 hours, while the AR model will take about 66 hours to run. This excludes the use of GP and AR models for forecasting next day ozone levels, which must be produced within 24 hours of computing time. The GPP model, on the other hand, takes only about 50 minutes to run the same experiment on the same personal computer and is the only feasible method that we henceforth adopt.

We compare the performance of the GPP model based with those obtained from a non-Bayesian linear regression model with the Eta CMAQ output as the only covariate, which is a simple method that does not require advanced modelling and computation techniques. We

also illustrate parameter estimation and maps providing forecast surfaces.

We report the parameter estimates and their standard deviations in Table 1.6 for the model fitting cases with 14 days data. The estimates are broadly similar for different subsets of fitted data. The Eta CMAQ output always remains a significant predictor with very small standard deviation relative to the mean. The temporal correlation remained always near 20%. The random effect variance σ_w^2 is always estimated to be larger than the nugget effect σ_ϵ^2. The estimate of the spatial decay parameter is 0.0024, that corresponds to an effective range of 1250 kilometers. A similar table based on model fitting from 7 days data is omitted for brevity.

We now compare the GPP model based forecasts with those from the linear regression model using the RMSEs based on validation data both from the 62 hold out sites. The RMSE values, provided in Table 1.7, are smaller for the GPP model than the linear regression model. Moreover, the RMSE values are smaller when the forecasting model is trained with 14 days data than the same with 7 days data. The RMSE values for the forecasts made by the Eta CMAQ model are considerably higher, which justifies this additional statistical modelling effort.

The nominal coverages of the 95% forecast intervals, provided in Table 1.8, show that the uncertainty in the forecasts based on the GPP model are about right. However, the nominal coverages for the linear model based forecasts are closer to 100%, which shows that these forecast intervals are too wide and this method fails to reduce uncertainty in the forecasts.

Table 1.8 also provides the CRPS values which turn out to be slightly higher than the values presented in Table 1.2 for the four states data. This is not surprising since it is usually more difficult to extrapolate in larger spatial domains. We have also obtained the false alarm and hit rates of the forecasts from the GPP model which are 0 and 95.33, respectively, when the threshold value is 75 ppb. Clearly, the GPP model is very accurate for forecasting and hence, we do not consider the other diagnostics such as the PIT diagram and the marginal calibration plots. Instead, we proceed to illustrate the forecasts.

Figure 1.8 illustrates the forecast maps based on the GPP model along with their standard deviations for the the 3 days, 8th, 9th and 10th of July. Here, each forecast map has their own color scheme that enables us to show the full spatial variation of the forecasts. In addition, the maps of standard deviations reveal that higher ozone levels are associated with higher uncertainty levels, which is a common phenomenon in ozone concentration modelling.

1.8 Conclusion

This chapter has developed Bayesian forecasting methods using three recently published Bayesian hierarchical models for spatio-temporal data. MCMC methods have been developed to compute the Bayesian forecast distributions based on large space-time data. These methodological developments have enabled us to add the suite of forecasting routines in the contributed `R` software package, `spTimer` which is available from `CRAN` (http://cran.r-project.org/) and allows modelling of large space-time data sets.

The contribution of the chapter also includes development of methods for estimating several forecast calibration measures using output from the implemented Markov chain Monte Carlo algorithms. We have demonstrated that these measures are able to compare different Bayesian forecasting methods rigorously and conclusively. A forecasting method based on a space-time model developed using a Gaussian predictive process approximation

has been shown to be fast and the best for the illustrative ozone concentration forecasting problem of the chapter.

References

Bakar KS and Sahu SK 2014 `spTimer`: Spatio-temporal bayesian modelling using `r`. *Journal of Statistical Software*.

Banerjee S, Carlin BP and Gelfand AE 2004 *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC, Boca Raton.

Banerjee S, Gelfand AE, Finley AO and Sang H 2008 Gaussian predictive process models for large spatial data sets. *Journal of Royal Statistical Society, Series B* **70**, 825–848.

Bauer G, Deistler M and Scherrer W 2001 Time series models for short term forecasting of ozone in the eastern part of austria. *Environmetrics* **12**, 117–130.

Berrocal VJ, Gelfand AE and Holland DM 2010a A bivariate space-time downcsaler under space and time misalignment. *Annals of Applied Statistics* **4**, 1942–1975.

Berrocal VJ, Gelfand AE and Holland DM 2010b A spatio-temporal downcsaler for outputs from numerical models.. *Journal of Agricultural, Biological and Environmental Statistics* **15**, 176–197.

Cameletti M, Ignaccolo R and Bande S 2011 Comparing spatio-temporal models for particulate matter in piemonte. *Environmetrics* **22**, 985996.

Cressie NAC and Johannesson G 2008 Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society, Series: B* **70**, 209–226.

Damon J and Guillas S 2002 The inclusion of exogenous variables in functional autoregressive ozone forecasting.. *Environmetrics* **13**, 759–774.

Feister U and Balzer K 1991 Surface ozone and meteorological predictors on a subregional scale. *Atmospheric Environment* **25**, 1781–1790.

Finley AO, Banerjee S and Carlin BP 2007 `spBayes`: An `r` package for univariate and multivariate hierarchical point-referenced spatial models. *Journal of Statistical Software* **12**(4), 1–24.

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.

Gelfand AE and Ghosh SK 1998 Model choice: A minimum posterior predictive loss approach. *Biometrika* **85**, 1–11.

Gneiting T, Balabdaoui F and Raftery A 2007 Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society, Series B* **69**, 243–268.

Huerta G, Sanso B and Stroud JR 2004 A spatio-temporal model for maxico city ozone levels. *Journal of the Royal Statistical Society, Series C* **53**, 231–248.

Jun M and Stein ML 2004 Statistical comparison of observed and cmaq modeled daily sulfate levels. *Atmospheric Environment* **38**, 4427–4436.

Kumar U and Ridder KD 2010 Garch modelling in association with fft-arima to forecast ozone episodes. *Atmospheric Environment* **44**, 4252–4265.

Lorence AC 1986 Analysis methods for numerical weather prediction. *The Quarterly Journal of the Royal Meteorological Society* **112**, 1177–1194.

McMillan N, Bortnick SM, Irwin ME and Berliner M 2005 A hierarchical bayesian model to estimate and forecast ozone through space and time. *Atmospheric Environment* **39**, 1373–1382.

Petris G, Petrone S and Patrizia C 2010 *Dynamic Linear Models with R*. Springer, Dordrecht.

Plummer M, Best N, K. C and Vines K 2006 CODA: Convergence diagnosis and output analysis for MCMC. *R News* **6**(1), 7–11.

Sahu SK and Bakar KS 2012a A comparison of bayesian models for daily ozone concentration levels. *Statistical Methodology* **9**(1), 144–157.

Sahu SK and Bakar KS 2012b Hierarchical bayesian auto-regressive models for large space time data with applications to ozone concentration modelling. *Applied Stochastic Models in Business and Industry* **28**, 395–415.

Sahu SK and Mardia KV 2005a A bayesian kriged-kalman model for short-term forecasting of air pollution levels. *Journal of the Royal Statistical Society, Series C* **54**, 223–244.

Sahu SK and Mardia KV 2005b Recent trends in modeling spatio-temporal data *Proceedings of the Special meeting on Statistics and Environment*, pp. 69–83. Università Di Messina.

Sahu SK, Gelfand AE and Holland DM 2007 High-resolution space-time ozone modeling for assessing trends. *Journal of the American Statistical Association* **102**, 1221–1234.

Sahu SK, Yip S and Holland DM 2009 Improved space-time forecasting of next day ozone concentrations in the eastern u.s.. *Atmospheric Environment* **43**, 494–501.

Sahu SK, Yip S and Holland DM 2011 A fast bayesian method for updating and forecasting hourly ozone levels. *Environmental and Ecological Statistics* **18**, 185–207.

Sousa SIV, Pires JCM, Martins F, Pereira MC and Alvim-Ferraz MCM 2009 Potentialities of quantile regression to predict ozone concentrations. *Environmetrics* **20**, 147–158.
Stein ML 2008 A modelling approach for large spatial datasets. *Journal of the Korian Statistical Society* **37**, 3–10.
Stroud JR, Muller P and Sanso B 2001 Dynamic models for spatio-temporal data. *Journal of the Royal Statistical Society, Series B* **63**, 673–689.
West M and Harrison J 1997 *Bayesian Forecasting and Dynamic Models* 2nd edn. Springer, New York.
Zidek JV, Le ND and Liu Z 2012 Combining data and simulated data for space-time fields: application to ozone.. *Environmental and Ecological Statistics* **19**, 37–56.

Table 1.1 Summaries of the daily maximum ozone concentration levels and Eta CMAQ output for the test data set described in Section 1.3.

	Minimum	Mean	Median	Maximum
Ozone levels	0.00	50.62	50.99	113.00
CMAQ output	16.50	59.19	60.36	145.50

Table 1.2 CRPS values from modelling data from four states during July 8 (denoted as 7/8) to 14.

	Values from modelling 7 days data							
Models	7/8	7/9	7/10	7/11	7/12	7/13	7/14	7/(8-14)
GP	6.12	10.22	5.04	5.05	4.78	5.70	6.95	6.27
AR	6.19	10.12	4.95	5.31	4.85	4.38	4.31	5.73
GPP	4.95	10.02	4.89	5.33	4.87	4.33	4.13	5.52
	Values from modelling 14 days data							
GP	6.14	9.82	5.33	5.42	5.21	5.64	6.29	6.27
AR	5.91	9.83	4.56	5.27	5.19	4.43	5.90	5.87
GPP	5.32	9.56	4.37	5.30	5.15	4.28	5.26	5.60

Table 1.3 Nominal coverages of the 50% and 95% forecast intervals for the one-step ahead forecasts at the 20 randomly chosen validation sites.

	Intervals			
	Using 7 Days Data		Using 14 Days Data	
	50%	95%	50%	95%
GP	51.43	95.71	55.00	95.71
AR	50.71	94.29	50.71	93.43
GPP	50.71	94.95	49.71	94.00

Table 1.4 Average width of the forecast intervals for the four states data set.

	Using 7 Days Data		Using 14 Days Data	
Models	50%	90%	50%	90%
GP	12.76	30.95	12.57	30.69
AR	13.51	32.95	13.36	32.28
GPP	11.54	28.11	9.58	23.47

Table 1.5 False alarm and hit rates for ozone threshold values of 65 and 75 for the four states data set.

		Using 7 days data		Using 14 days data	
Ozone levels	Model	False alarm	Hit rate	False alarm	Hit rate
65 ppb	GP	0.92	91.67	0.92	91.67
	AR	4.59	92.50	1.83	92.50
	GPP	3.67	91.67	2.75	91.67
75 ppb	GP	0.0	95.83	0.0	95.83
	AR	0.0	95.83	0.0	95.83
	GPP	0.0	96.67	0.0	97.50

Table 1.6 Parameter estimates (mean and sd) for the models based on GPP approximation fitted with 14 days observations for the period June 24 (denoted as 6/24) to July 13, 2010 from the 577 modelling sites in the whole eastern US.

Fitted Days		β_0	β_1	ρ	σ^2_ϵ	σ^2_w	ϕ
6/24–7/7	Mean	4.13	0.37	0.40	0.24	0.49	0.0046
	sd	0.20	0.03	0.04	0.005	0.04	0.0005
6/25–7/8	Mean	4.34	0.36	0.39	0.25	0.53	0.0042
	sd	0.23	0.02	0.03	0.004	0.04	0.0004
6/26–7/9	Mean	4.68	0.33	0.39	0.25	0.57	0.0041
	sd	0.33	0.03	0.04	0.006	0.05	0.0007
6/27–7/10	Mean	3.40	0.33	0.39	0.25	0.52	0.0046
	sd	0.22	0.03	0.04	0.005	0.04	0.0005
6/28–7/11	Mean	4.74	0.31	0.35	0.25	0.60	0.0031
	sd	0.17	0.02	0.04	0.005	0.05	0.0007
6/29–7/12	Mean	4.66	0.31	0.36	0.25	0.54	0.0037
	sd	0.20	0.02	0.04	0.005	0.04	0.0003
6/30–7/13	Mean	4.92	0.29	0.35	0.26	0.60	0.0032
	sd	0.30	0.03	0.04	0.005	0.07	0.0006

Table 1.7 Values of the RMSE of the forecasts at the hold-out sites for the simple linear model and the GPP model based on modelling 7 and 14 days data for the whole of Eastern US. The corresponding RMSE values for the Eta CMAQ output are also shown.

Forecast	CMAQ	7 Days		14 Days	
		Linear	GPP	Linear	GPP
7/8	20.52	12.16	10.34	10.97	10.30
7/9	19.68	12.25	10.79	11.59	10.04
7/10	16.36	9.87	8.59	9.49	8.13
7/11	15.51	8.55	8.17	8.69	7.98
7/12	13.12	8.99	8.67	8.44	8.17
7/13	20.36	12.70	10.85	13.95	9.83
7/14	18.10	9.64	9.20	10.25	9.05

Table 1.8 Nominal coverages of the 95% forecast intervals using the linear and GPP models and the CRPS values for the hold out data for the GPP model for the whole eastern US data set.

Forecast	7/8	7/9	7/10	7/11	7/12	7/13	7/14	7/(8-14)
Nominal coverage of the 95% forecast intervals using the linear model								
7 Days	99.94	99.80	99.44	100.00	100.00	99.07	98.15	99.11
14 Days	99.94	98.50	97.59	100.00	100.00	97.50	98.15	98.64
Nominal coverage of the 95% forecast intervals using the GPP model								
7 Days	93.55	93.75	94.96	95.16	94.96	93.75	95.56	94.53
14 Days	94.62	94.30	94.84	95.05	94.62	94.84	94.84	94.74
CRPS values								
7 Days	10.05	7.98	6.52	6.79	7.12	7.18	7.11	7.54
14 Days	9.43	7.25	5.89	6.80	6.93	6.94	6.74	7.15

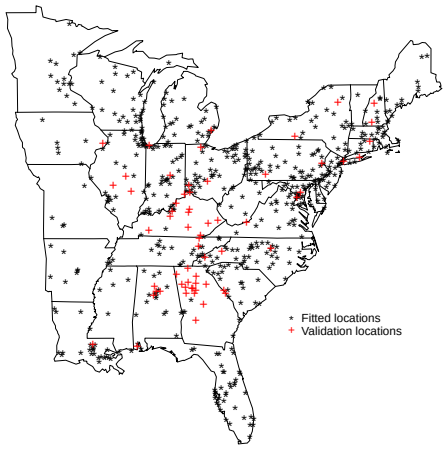


Figure 1.1 A plot of the 639 (62 validation and 577 model fitting) ozone monitoring sites in the eastern US.

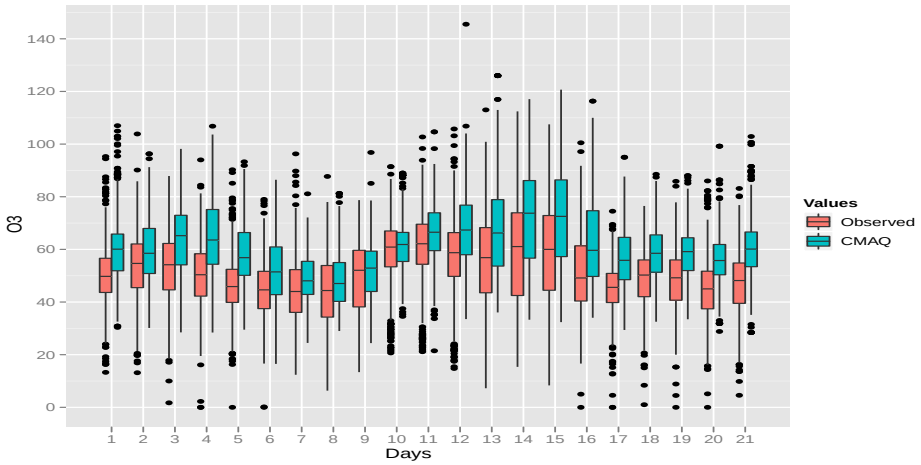


Figure 1.2 Side by side box-plots of the observed daily maximum ozone concentration levels and Eta CMAQ output for 21 days from all 639 sites in the eastern US.

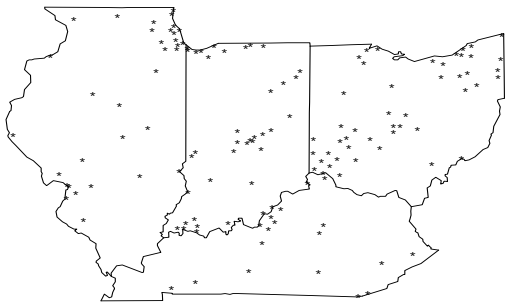


Figure 1.3 Map of the four states, Ohio, Indiana, Illinois and Kentucky. 147 ozone monitoring locations are superimposed.

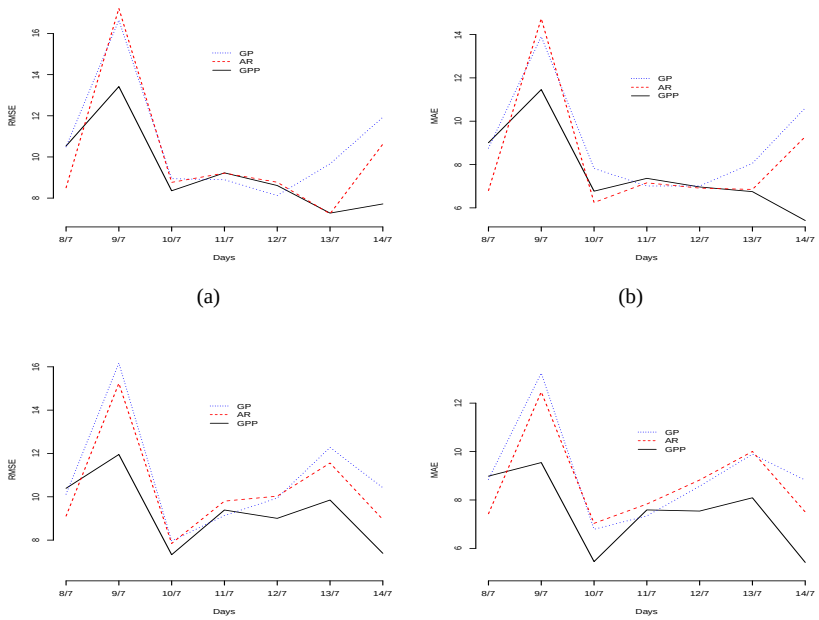
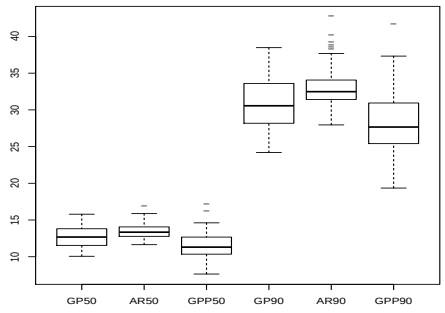
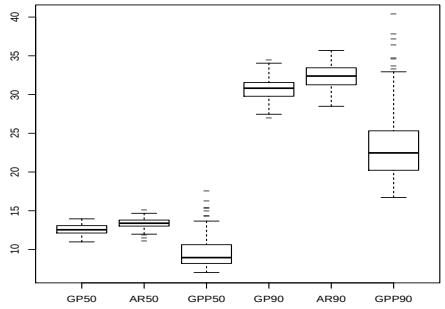


Figure 1.4 Plots of RMSE and MAE based on modelling 7 days data (top row) and 14 days data (bottom row).

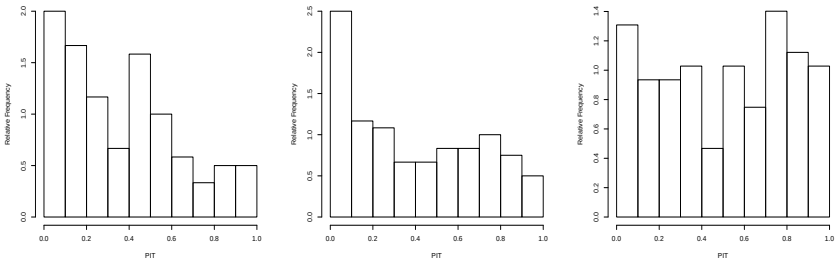


(a)



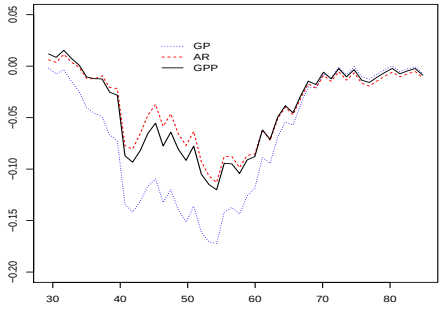
(b)

Figure 1.5 Sharpness diagram using: (a) 7 days data (b) 14 days data.

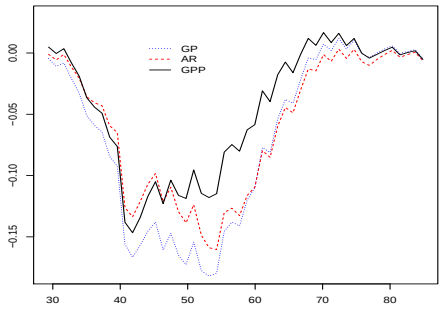


(a) (b) (c)

Figure 1.6 PIT diagrams for (a) GP, (b) AR and (c) GPP models using 14 days data for modelling.

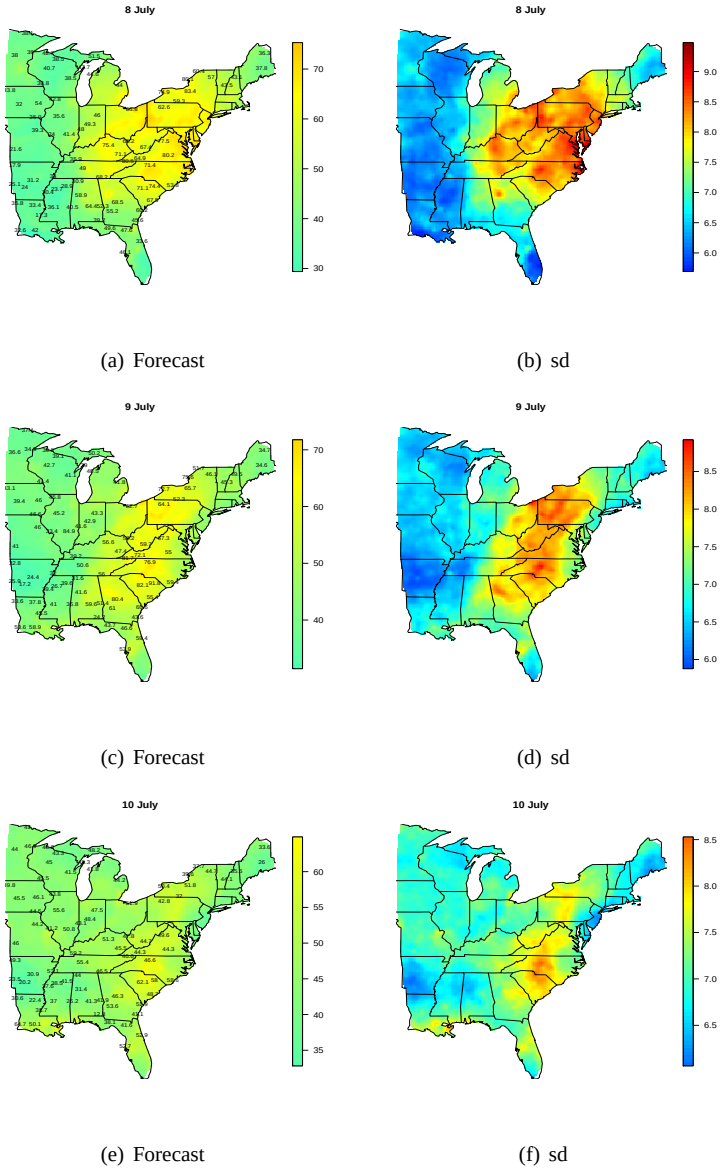


(a)



(b)

Figure 1.7 Marginal calibration plots for all the models using (a) 7 days data (b) 14 days data for modelling.



(a) Forecast (b) sd

(c) Forecast (d) sd

(e) Forecast (f) sd

Figure 1.8 Maps showing the forecasts and their standard deviations for July 8, 9 and 10 in 2010. Observed ozone levels are also superimposed on the forecast maps from a selected number of sites only, to avoid clutter.