ARTICLE IN PRESS

Statistical Methodology ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Statistical Methodology

journal homepage: www.elsevier.com/locate/stamet

A comparison of Bayesian models for daily ozone concentration levels

S.K. Sahu *, K.S. Bakar

School of Mathematics, University of Southampton, Southampton SO17 1BJ, UK

ARTICLE INFO

Keywords:
Auto-regressive model
Bayesian inference
Dynamic linear model
Maximum daily eight-hour ozone levels
Space–time modeling

ABSTRACT

Recently, there has been a surge of interest in Bayesian space–time modeling of daily maximum eight-hour average ozone concentration levels. Hierarchical models based on well known time series modeling methods such as the dynamic linear models (DLM) and the auto-regressive (AR) models are often used in the literature. The DLM, developed as a result of the popularity of Kalman filtering methods, provide a dynamical state-space system that is thought to evolve from a pair of state and observation equations. The AR models, on the other hand, cast in a Bayesian hierarchical setting, have recently been developed through a pair of models where a measurement error model is formulated at the top level and an AR model for the true ozone concentration levels is postulated at the next level. Each of the modeling scenarios is set in an appropriate multivariate setting to model the spatial dependence. This paper compares these two methods in hierarchical Bayesian settings. A simplified skeletal version of the DLM taken from Dou et al. (2010) [5] is compared theoretically with a matching hierarchical AR model. The comparisons reveal many important differences in the induced space–time correlation structures. Further comparisons of the variances of the predictive distributions by conditioning on different sets of data for each model show superior performances of the AR models under certain conditions. These theoretical investigations are followed up by a simulation study and a real data example implemented using Markov chain Monte Carlo (MCMC) methods for modeling daily maximum eight-hour average ozone concentration levels observed in the state of New York in the months of July and August, 2006. The hierarchical AR

* Corresponding author.
E-mail address: S.K.Sahu@soton.ac.uk (S.K. Sahu).

1572-3127/$ – see front matter
doi:10.1016/j.stamet.2011.04.009

Please cite this article in press as: S.K. Sahu, K.S. Bakar, A comparison of Bayesian models for daily ozone concentration levels, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.04.009

model is chosen using all the model choice criteria considered in this example.

1. Introduction

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. To evaluate exposure to ozone levels, the United States Environmental Protection Agency (USEPA) collects ozone concentration data continuously from several networks of sparsely and irregularly spaced monitoring sites throughout the United States (US). Data obtained from these sparse networks must be processed using spatio-temporal models to spatially predict and temporally forecast at an unmonitored site in the vast continental land mass of the US.

Space–time modeling of ground level ozone has received much recent attention in the literature; see, e.g., [7,2]. Cox and Chu [3] used a generalized linear model to estimate site specific trends in daily maximum ozone levels. Hierarchical Bayesian approaches for spatial prediction of air pollution have also been developed; see, e.g. [1,8,16,12,11,13,14], and references therein. McMillan et al. [10] propose a regime switching model for ozone level forecasting using meteorological variables as covariates.

The multivariate extension of the DLM for univariate time series [15] to model spatial dependence has been made popular by many authors; see, e.g. [14,8]. Dou et al. [5] compare the modeling approaches for hourly ozone concentration fields based on the DLM with a version of the Bayesian spatial predictor adapted for temporal data; see, e.g., [9]. Their evaluation of a simplified version of the DLM reveal some problematic properties for it, and they propose model alterations to fix some of those; see Section 3 of this paper. They also mention that for computational reasons the DLM's spatial domain has to be restricted, and as a result only data from a limited number of monitoring sites can be analyzed simultaneously. Zheng et al. [17] compare the generalized additive models with the DLM fitted to annually aggregated ozone concentration levels for individual stations separately; they do not compare the space–time covariance structures of the models.

The main focus of this paper is on comparing the DLM with a hierarchical version of the auto-regressive (AR) models developed by Sahu et al. [11,13] for daily maximum eight-hour average ozone concentration data. The AR models can be written as special cases of the DLM when the dimensions of the observation and the state vectors are the same. Obviously, in such a case the comparisons become much easier and the results obtained here will illustrate that. The main objective of this paper is to compare the models when they may differ substantially in many respects. For example, as is often done, at any time point the dimension of the state vector for the DLM is assumed to be much smaller than the same for the AR model where it is equal to the dimension of the observation vector. In our space–time modeling setup, the number of monitoring sites in the data set is the dimension of the observation vector at any time point. Consequently, the 'state vectors' under the two models are very different in nature, although they may be assumed to have the same prior mean and variance. The multivariate DLM and AR models also differ in the way in which spatial correlations are introduced in them; see Section 2. As a result they induce very different space–time correlation structures for the data; see Section 3.1.

This paper proves several inequalities for the variances of the posterior predictive distributions for spatial interpolation and temporal forecasting at unmonitored sites. For both the models, we find the undesirable property that under suitable conditions the variance of the posterior predictive distribution may increase for successive time points conditional on all the data up to that point. We investigate such issues in detail and find conditions under which the reverse result may be true for both of the modeling strategies.

As can be expected, the theoretical investigations can only be done for simpler versions of the models. To compare the performances of the two modeling strategies in practical data

Please cite this article in press as: S.K. Sahu, K.S. Bakar, A comparison of Bayesian models for daily ozone concentration levels, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.04.009

modeling situations we adopt several predictive Bayesian model choice criteria. These show a better performance of the AR models in a simulation study and a real data example on modeling daily maximum eight-hour average ozone concentration levels observed in the state of New York in the months of July and August, 2006.

The remainder of this paper is organized as follows. In Section 2, we briefly describe both the models and their simplified versions. Section 3 discusses the theoretical properties for the induced correlation structures and proves several inequalities involving the variances of the posterior predictive distributions. Section 4 provides a simulation study and the real data example. Finally, a few summary remarks are provided in Section 5.

2. DLM and AR models

2.1. DLM

Let $Z(\mathbf{s}_i, t)$ denote the square root of the observed daily maximum eight-hour average ozone level in parts per billion (ppb) units on day t at the location $\mathbf{s}_i$, with $t = 1, \ldots, T$ and $i = 1, \ldots, n$. We use the square root as the variance stabilizing transformation following the standard practice in the literature; see e.g. [5,11,13].

Let $\mathbf{Z}_t = (Z(\mathbf{s}_1, t), \ldots, Z(\mathbf{s}_n, t))'$ denote the observation vector for any $1 \leq t \leq T$ where T is the total number of days in the data set. The DLM are specified by the following pair of observation and state equations:

$$\mathbf{Z}_t = F_t\boldsymbol{\theta}_t + \boldsymbol{\epsilon}_t, \quad t \geq 1, \tag{1}$$

$$\boldsymbol{\theta}_t = G_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\eta}_t, \quad t \geq 1, \tag{2}$$

where the observation error vector, $\boldsymbol{\epsilon}_t$, is assumed to follow the $N(\mathbf{0}, \Sigma_\epsilon)$ distribution independently and the distribution of $\boldsymbol{\eta}_t$ is specified after we specify the matrices F_t and G_t. To accommodate p (say) known covariate values at time t, denoted by the $n \times p$ matrix X_t, we assume that $F_t = (\mathbf{1}, X_t)$; consequently $\boldsymbol{\theta}_t$ is a $p + 1$-dimensional state vector. We assume the state transfer matrix G_t to be ρI where $|\rho| \leq 1$ and I is the identity matrix. We now assume that $\boldsymbol{\eta}_t \sim N(\mathbf{0}, \Sigma_\eta)$ for $t \geq 1$. For the initial state, we assume that $\boldsymbol{\theta}_0 \sim N(\boldsymbol{\mu}, \sigma_0^2 I)$ for suitable values of the hyperparameters $\boldsymbol{\mu}$ and σ_0^2, where I denotes the identity matrix of appropriate order. Unlike the Dou et al. paper ours does not include any seasonal term in the models for daily data. Seasonal terms are more relevant for modeling the diurnal cyclic components often present in the hourly data.

The observations are spatially correlated; hence a spatially colored covariance matrix must be assumed for Σ_ϵ. For convenience, we assume the exponential covariance function to model spatial dependence and let

$$\Sigma_\epsilon = \sigma_\epsilon^2 \exp(-\phi_\epsilon D)$$

where $\phi_\epsilon > 0$ is a spatial correlation decay parameter assumed to be known, and the $n \times n$ distance matrix D has elements d_{ij}, the distance between $\mathbf{s}_i$ and $\mathbf{s}_j$, $i, j = 1, \ldots, n$.

Details regarding the covariates, X_t will be discussed in the practical examples in Section 4. Following Dou et al. [5] and only for the theoretical comparisons made in Section 3, we consider a simplified version of the above models where we assume that there is no covariate present, i.e. $F_t = \mathbf{1}$ which corresponds to the model that has a site invariant mean. Consequently, $\boldsymbol{\eta}_t$ turns out to be a scalar, and we assume that $\eta_t \sim N(0, \sigma_\eta^2)$. In this case the scalar θ_0 is assumed to follow the $N(\mu, \sigma_0^2)$ distribution.

2.2. Hierarchical AR models

In the following descriptions of the AR models we keep the same notation for the corresponding error vectors, the variance components and the parameters describing the mean structure in the DLM

Please cite this article in press as: S.K. Sahu, K.S. Bakar, A comparison of Bayesian models for daily ozone concentration levels, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.04.009

for comparison purposes, although it is to be noted that parameters have different interpretations under different models. Let $O(\mathbf{s}_i, t)$ denote the true value corresponding to $Z(\mathbf{s}_i, t)$ and $\mathbf{O}_t = (O(\mathbf{s}_1, t), \ldots, O(\mathbf{s}_n, t))'$. The hierarchical AR models are given by

$$\mathbf{Z}_t = \mathbf{O}_t + \boldsymbol{\epsilon}_t \tag{3}$$

$$\mathbf{O}_t = \rho \mathbf{O}_{t-1} + \xi \mathbf{1} + X_t \boldsymbol{\beta} + \boldsymbol{\eta}_t, \tag{4}$$

where ρ is a temporal correlation parameter and the distributions for the error vectors $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$, assumed to be independent, are given below. For the hierarchical error term $\boldsymbol{\epsilon}_t$ we assume the independent $N(\mathbf{0}, \sigma_\epsilon^2 I)$ distribution providing the so called nugget effect, σ_ϵ^2. For the spatial error term, $\boldsymbol{\eta}_t$, we assume the independent $N(\mathbf{0}, \Sigma_\eta)$ distribution where $\Sigma_\eta = \sigma_\eta^2 \exp(-\phi_\eta D)$ provides the spatially colored exponential covariance matrix with site invariant common variance σ_η^2; D continues to be the $n \times n$ distance matrix and $\phi_\eta > 0$ is a known spatial correlation decay parameter. The model specification is completed by assuming the initial condition $\mathbf{O}_0 \sim N(\mu \mathbf{1}, \Sigma_0)$ independently, where $\Sigma_0 = \sigma_0^2 \exp(-\phi_0 D)$; again we assume the exponential covariance function with a known spatial correlation decay parameter $\phi_0 > 0$. The variance component, σ_0^2, for the initial condition is assumed to be the same as for the DLM.

2.3. Differences and similarities between the models

Observe that spatial correlation is introduced at the top level of the DLM. This strategy is adopted following many authors; see e.g., [5] and the references therein. The state variables specified in (2) are usually thought to be free of spatial dependence. It is also not clear how spatial correlation can be introduced through the state variables that are often assumed to belong to a lower dimensional state space than the data vector $\mathbf{Z}_t$. For the AR models, the top level specification (3) providing the nugget effect is advantageous for handling missing data since in an iterative Bayesian computation framework for these models any missing $Z(\mathbf{s}, t)$ is simulated from the simple top level model at each iteration. Moreover, the stationary assumption for the spatial covariance function is likely to be more meaningful for the true underlying process $O(\mathbf{s}, t)$ than the observed noisy process $Z(\mathbf{s}, t)$.

The DLM given by the pair of Eqs. (1) and (2) can coincide with the AR models given by the pair (3) and (4) when the dimension of $\boldsymbol{\theta}_t$, $p + 1$ equals n, the dimension of $\mathbf{O}_t$ and $\mathbf{Z}_t$, and the parameters ξ and $\boldsymbol{\beta}$ are assumed to be zero in Eq. (4). This happens when analyzing univariate time series data, i.e. for $n = 1$, which is not the case when modeling spatio-temporal ozone concentration data. For this modeling problem we assume that n is much larger than $p + 1$, as has been done in the comparison study in [5].

The DLM corresponding to $G_t = \rho I$ in (2) when $|\rho| < 1$ will have many similarities with the AR models. However, there will still be differences in the correlation structures induced by the two models due to the mismatch in the dimensions of the 'state vectors' under the two models; see Section 3.

The hierarchical versions of the AR models match with the DLM in many other respects. First, the marginal mean values of the observations $Z(\mathbf{s}_i, t)$ are matched by assuming $\xi = 0$ in (4) and $\mu = 0$ in both the models. Note that we are not assuming any covariate effect for the theoretical comparisons. Second, both sets of models have three variance components, given by σ_ϵ^2, σ_η^2 and σ_0^2 at the three respective levels of the hierarchical specifications, with broadly similar interpretations: observation error variance, process variance and a variance for the initial condition. Finally, the spatial correlations can also be matched by assuming a common value for the decay parameters ϕ_ϵ and ϕ_η, although these are assumed at different levels for the two models. The additional decay parameter ϕ_0 in the AR models can also be chosen to be the common value corresponding to an assumption of a static spatial correlation field for the true ozone level process $\mathbf{O}_t$. We shall make this assumption for theoretical comparison purposes and use the notation ϕ as the common spatial decay parameter; we will estimate ϕ in the practical data modeling examples in Section 4.

The two sets of models, however, differ in terms of the number of unknown parameters describing the mean structure and the way spatial correlation is introduced in them. The DLM have fewer parameters than the AR models since the mean for each $\mathbf{Z}_t$ under the DLM is described by the

Please cite this article in press as: S.K. Sahu, K.S. Bakar, A comparison of Bayesian models for daily ozone concentration levels, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.04.009

$p+1$-dimensional $\boldsymbol{\theta}_t$ whereas the mean under the AR models is described by the n-dimensional $\mathbf{O}_t$, ρ, ξ and $\boldsymbol{\beta}$. However, as noted above, the marginal means of the observations can easily be matched to be the same by the hierarchical specifications. These facts justify the theoretical comparison study reported in Section 3.

3. Model comparisons

In this section we shall assume that all the variance components: σ_ϵ^2, σ_η^2 and σ_0^2, are known for both the models. However, we shall remove these assumptions in our simulation and real data examples in the next section. Further, to obtain theoretical results for comparisons we shall also assume that there is no covariate present in both the models. That is $F_t = \mathbf{1}$ in the DLM as discussed in the last paragraph of Section 2.1 and $\boldsymbol{\beta} = \mathbf{0}$ and $\xi = 0$ in (4); these assumptions will be removed in the practical examples in Section 4.

For the simplified versions of the DLM specifications in Section 2.1, when $\rho = 1$, direct calculations yield (see also Theorem 1 in [5])

$$\text{Cov}(\mathbf{Z}_t, \mathbf{Z}_{t+k}) = (\sigma_0^2 + t\sigma_\eta^2)\mathbf{1}\mathbf{1}' + 1(k=0)\sigma_\epsilon^2 \exp(-\phi D), \quad k = 0, 1, \ldots, \tag{5}$$

where $1(k = 0) = 1$ if $k = 0$, and 0 otherwise. As expected, here the variance of $\mathbf{Z}_t$ will explode with t. Although this is not a concern when the DLM is used to model non-stationary temporal data observed for a short period of time, it can be stopped by assuming $|\rho| < 1$. In fact, for $|\rho| < 1$ we obtain the following result by direct calculations:

$$\text{Cov}(\mathbf{Z}_t, \mathbf{Z}_{t+k}) = \left(\sigma_0^2\rho^{2t+k} + \sigma_\eta^2\rho^k\frac{1-\rho^{2t}}{1-\rho^2}\right)\mathbf{1}\mathbf{1}' + 1(k=0)\sigma_\epsilon^2\exp(-\phi D),$$
$$k = 0, 1, \ldots. \tag{6}$$

We now obtain a similar result for the AR models as follows. For any positive integer t the AR models imply that

$$Z(\mathbf{s}_i, t) = \epsilon(\mathbf{s}_i, t) + \eta(\mathbf{s}_i, t) + \rho\eta(\mathbf{s}_i, t-1) + \cdots + \rho^{t-1}\eta(\mathbf{s}_i, 1) + \rho^t O(\mathbf{s}_i, 0),$$

and for any integer $k > 0$,

$$\begin{aligned} Z(\mathbf{s}_j, t+k) = {} & \epsilon(\mathbf{s}_j, t+k) + \eta(\mathbf{s}_j, t+k) + \rho\eta(\mathbf{s}_j, t+k-1) + \cdots + \rho^{k-1}\eta(\mathbf{s}_j, t+1) \\ & + \rho^k\eta(\mathbf{s}_j, t) + \rho^{k+1}\eta(\mathbf{s}_j, t-1) + \cdots + \rho^{t+k-1}\eta(\mathbf{s}_j, 1) + \rho^{t+k}O(\mathbf{s}_j, 0). \end{aligned}$$

Now recall that the spatial errors $\boldsymbol{\eta}_t$ and $\boldsymbol{\eta}_{t+k}$ are independent if $k > 0$ and the hierarchical error $\boldsymbol{\epsilon}_t$ is independent of the spatial error $\boldsymbol{\eta}_t$, and the initial random variable $\mathbf{O}_0$ is independent of both $\boldsymbol{\eta}_t$ and $\boldsymbol{\epsilon}_t$. Hence, we have

$$\begin{aligned} \text{Cov}(Z(\mathbf{s}_i, t), Z(\mathbf{s}_j, t+k)) = {} & \text{Cov}(\epsilon(\mathbf{s}_i, t), \epsilon(\mathbf{s}_j, t+k)) + \rho^{2t+k}\text{Cov}(O(\mathbf{s}_i, 0), O(\mathbf{s}_j, 0)) \\ & + \rho^k\text{Cov}(\eta(\mathbf{s}_i, t), \eta(\mathbf{s}_j, t)) + \rho^{k+2}\text{Cov}(\eta(\mathbf{s}_i, t-1), \eta(\mathbf{s}_j, t-1)) \\ & + \cdots + \rho^{k+2t-2}\text{Cov}(\eta(\mathbf{s}_i, 1), \eta(\mathbf{s}_j, 1)) \\ = {} & \rho^{2t+k}\sigma_0^2\exp(-\phi_0 d_{ij}) + \rho^k\frac{1-\rho^{2t}}{1-\rho^2}\sigma_\eta^2\exp(-\phi_\eta d_{ij}), \end{aligned}$$

assuming $|\rho| \neq 1$. Thus we arrive at the following general covariance function:

$$\text{Cov}(\mathbf{Z}_t, \mathbf{Z}_{t+k}) = \rho^{2t+k}\Sigma_0 + \rho^k\frac{1-\rho^{2t}}{1-\rho^2}\Sigma_\eta + 1(k=0)\sigma_\epsilon^2 I, \quad k = 0, 1, \ldots. \tag{7}$$

Clearly, this covariance function will have many similarities to and differences from the earlier ones obtained for the DLM in Eqs. (5) and (6). The following two subsections investigate these properties in more detail.

Please cite this article in press as: S.K. Sahu, K.S. Bakar, A comparison of Bayesian models for daily ozone concentration levels, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.04.009

3.1. Correlation structures

Using the expressions for the general covariance functions in (5) and (6) we follow Dou et al. [5] to establish the following results:

(i) $\text{Cor}(Z(\mathbf{s}_i, t), Z(\mathbf{s}_j, t+k))$ for $i \neq j$ attains its maximum at $k = 0$ and decreases as k increases for both the covariance functions. This can be a reasonable property since the correlation between observations at different locations can be expected to be the maximum at the current time because both of those locations may be influenced similarly by the prevailing meteorological and other conditions, e.g. power station emission volumes, affecting ozone production. The correlation should decrease at different times due to possible mismatches in these conditions at different times.

(ii) $\text{Cor}(Z(\mathbf{s}_i, t), Z(\mathbf{s}_j, t)) \to 1$ as $t \to \infty$ for $i \neq j$ when $\rho = 1$. This seems to be an unreasonable property. The correlation between any two fixed monitors should not increase with time. Indeed, when $|\rho| < 1$ it is straightforward to see from (6) that the limit of this correlation is less than 1.

(iii) $\text{Cor}(Z(\mathbf{s}_i, t), Z(\mathbf{s}_j, t)) \to 1$ as $d_{ij} \to 0$ for $i \neq j$ for both the covariance functions. This is a reasonable property since the observations at two locations close to each other should be very similar.

(iv) $\text{Cor}(Z(\mathbf{s}_i, t), Z(\mathbf{s}_j, t)) \to \frac{\sigma_0^2 + t\sigma_\eta^2}{\sigma_0^2 + t\sigma_\eta^2 + \sigma_\epsilon^2}$ as $d_{ij} \to \infty$ for $i \neq j$ when $\rho = 1$. When $|\rho| < 1$ this limit is given by $\frac{\sigma_0^2\rho^{2t} + \sigma_\eta^2(1-\rho^{2t})/(1-\rho^2)}{\sigma_0^2\rho^{2t} + \sigma_\eta^2(1-\rho^{2t})/(1-\rho^2) + \sigma_\epsilon^2}$. Ideally, this limit should be close to 0 since the observations at two far away locations should tend to be independent of each other. In order to achieve this ideal limit, Dou et al. [5] suggested replacing σ_η^2 by σ_η^2/T and taking σ_0^2 much smaller than σ_ϵ^2 when $\rho = 1$.

Similar properties of the AR models can be derived using the general covariance function (7). We first note the following results for $k > 0$ and $i \neq j$:

$$\text{Var}(Z(\mathbf{s}_i, t)) = \rho^{2t}\sigma_0^2 + \frac{1-\rho^{2t}}{1-\rho^2}\sigma_\eta^2 + \sigma_\epsilon^2$$

$$\text{Cov}(Z(\mathbf{s}_i, t), Z(\mathbf{s}_j, t)) = \rho^{2t}\sigma_0^2 e^{-\phi_0 d_{ij}} + \frac{1-\rho^{2t}}{1-\rho^2}\sigma_\eta^2 e^{-\phi_\eta d_{ij}}$$

$$\text{Cov}(Z(\mathbf{s}_i, t), Z(\mathbf{s}_j, t+k)) = \rho^k\left(\rho^{2t}\sigma_0^2 e^{-\phi_0 d_{ij}} + \frac{1-\rho^{2t}}{1-\rho^2}\sigma_\eta^2 e^{-\phi_\eta d_{ij}}\right).$$

We now have the following results for the AR models.

(i) As in the corresponding DLM case (i), $\text{Cor}(Z(\mathbf{s}_i, t), Z(\mathbf{s}_j, t+k))$ for $i \neq j$ decreases as k increases.

(ii) $\text{Cor}(Z(\mathbf{s}_i, t), Z(\mathbf{s}_j, t)) \to \frac{\sigma_\eta^2 \exp(-\phi d_{ij})}{\sigma_\epsilon^2(1-\rho^2) + \sigma_\eta^2}$ as $t \to \infty$ for $i \neq j$. Unlike in the corresponding case (ii) for the DLM with $\rho = 1$, this correlation does not approach 1.

(iii) $\text{Cor}(Z(\mathbf{s}_i, t), Z(\mathbf{s}_j, t)) \to 1$ as $d_{ij} \to 0$ for $i \neq j$. This is a reasonable property as in the corresponding case (iii) for the DLM.

(iv) $\text{Cor}(Z(\mathbf{s}_i, t), Z(\mathbf{s}_j, t)) \to 0$ as $d_{ij} \to \infty$ for $i \neq j$. Unlike in the corresponding case (iv) for both versions of the DLM, here the ideal limit is reached without any further condition or model adjustments.

3.2. Variance inequalities

The differences in the correlation structures imply very different behaviors in model based predictions and forecasting. Here we investigate the prediction variances by examining five important inequalities capturing various possibilities for predictions. To compare the AR models with the exact simplified version of the DLM proposed by Dou et al. [5] we take $\rho = 1$ in the DLM. Henceforth, we do not consider the $|\rho| < 1$ case for the DLM, although a more fair comparison can be performed in this case.

Please cite this article in press as: S.K. Sahu, K.S. Bakar, A comparison of Bayesian models for daily ozone concentration levels, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.04.009

For simplicity we consider prediction and forecasting at an unmonitored site $\mathbf{s}_1$ given the observations at a monitored site $\mathbf{s}_2$. We also consider data and forecasting for two time points $t = 1$ and 2. We assume that all the parameters $\rho, \phi, \sigma_0^2, \sigma_\eta^2, \sigma_\epsilon^2$ are known. Hence the conditional variance of $Z(\mathbf{s}_1, t)$ given $Z(\mathbf{s}_2, t')$ for any t and t' will be the predictive variance in the Bayesian setting since there is no need to integrate over any unknown parameters to obtain the predictive distributions. The comparisons performed in the simulation study and the real data example in the next section do not make these assumptions.

With four possible space–time random variables $Z(\mathbf{s}_1, 1), Z(\mathbf{s}_1, 2), Z(\mathbf{s}_2, 1)$, and $Z(\mathbf{s}_2, 2)$ we consider the following conditional variances:

$$\mathrm{Var}(Z(\mathbf{s}_1, 1)|Z(\mathbf{s}_2, 1)) = \sigma_\epsilon^2 + \rho^2\sigma_0^2 + \sigma_\eta^2 - \zeta^2 \frac{(\sigma_0^2\rho^2 + \sigma_\eta^2)^2}{\sigma_\epsilon^2 + \rho^2\sigma_0^2 + \sigma_\eta^2}$$

$$\mathrm{Var}(Z(\mathbf{s}_1, 2)|Z(\mathbf{s}_2, 2)) = \sigma_\epsilon^2 + \rho^4\sigma_0^2 + (1+\rho^2)\sigma_\eta^2 - \zeta^2 \frac{\{\rho^4\sigma_0^2 + (1+\rho^2)\sigma_\eta^2\}^2}{\sigma_\epsilon^2 + \rho^4\sigma_0^2 + (1+\rho^2)\sigma_\eta^2},$$

where $\zeta = \exp(-\phi d_{12})$ denotes the spatial correlation between the observations at the two sites at any given time. The general covariance function (7) also allows us to calculate the conditional variances $\mathrm{Var}(Z(\mathbf{s}_1, 1)|Z(\mathbf{s}_2, 1), Z(\mathbf{s}_2, 2))$ and $\mathrm{Var}(Z(\mathbf{s}_1, 2)|Z(\mathbf{s}_2, 1), Z(\mathbf{s}_2, 2))$; the expressions for these are long and hence are omitted for brevity. Instead, we obtain the following results involving them:

$$\mathrm{Var}(Z(\mathbf{s}_1, 1)|Z(\mathbf{s}_2, 1)) - \mathrm{Var}(Z(\mathbf{s}_1, 1)|Z(\mathbf{s}_2, 1), Z(\mathbf{s}_2, 2)) = \frac{N}{\triangle(\sigma_\epsilon^2 + \rho^2\sigma_0^2 + \sigma_\eta^2)}$$

$$\mathrm{Var}(Z(\mathbf{s}_1, 2)|Z(\mathbf{s}_2, 2)) - \mathrm{Var}(Z(\mathbf{s}_1, 2)|Z(\mathbf{s}_2, 1), Z(\mathbf{s}_2, 2)) = \frac{N}{\triangle(\sigma_\epsilon^2 + \rho^4\sigma_0^2 + (1+\rho^2)\sigma_\eta^2)},$$

where

$$N = \zeta^2\rho^2\sigma_\epsilon^4(\rho^2\sigma_0^2 + \sigma_\eta^2)^2$$

and

$$\triangle = \sigma_\epsilon^4 + \sigma_\eta^2(\rho^2\sigma_0^2 + \sigma_\eta^2) + \sigma_\epsilon^2\{\rho^2(1+\rho^2)\sigma_0^2 + (2+\rho^2)\sigma_\eta^2\}.$$

Thus the above two differences in variances are always non-negative. These two variance inequalities ascertain that the variance of the spatial prediction at site $\mathbf{s}_1$ using data from both time points will always be smaller than that when the spatial prediction is done using data from only one time point. Dou et al. [5] prove the exact same inequalities for the DLM with $\rho = 1$.

A striking difference between the two models lies in the expression for N, the numerator. Observe that both differences have a factor ζ^2 in the numerator which implies that the differences increase as the spatial correlation ζ increases. Intuitively, this is a very desirable property since spatial prediction with more observations should become more accurate as the spatial correlation increases. However, the same conclusion cannot be reached for the DLM since the same variance differences involve the spatial correlation ζ only through a factor $(1-\zeta)^2$ in the numerator (see [5]); the details are omitted for brevity. This seems to be an undesirable property of the DLM.

Dou et al. [5] also prove that, for the DLM, conditioned on the same amount of data, the predictive variance of $Z(\mathbf{s}_1, 1)$ would be no greater than that of $Z(\mathbf{s}_1, 2)$, that is,

$$\mathrm{Var}(Z(\mathbf{s}_1, 1)|Z(\mathbf{s}_2, 1), Z(\mathbf{s}_2, 2)) \leq \mathrm{Var}(Z(\mathbf{s}_1, 2)|Z(\mathbf{s}_2, 1), Z(\mathbf{s}_2, 2)).$$

Thus the predictive variance function is a monotonic increasing function of time t based on the same set of data. The same inequality holds for the AR models only under the condition

$$\varsigma \equiv \frac{\sigma_\eta^2}{\sigma_0^2} \geq 1 - \rho^2.$$

Please cite this article in press as: S.K. Sahu, K.S. Bakar, A comparison of Bayesian models for daily ozone concentration levels, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.04.009

Note that this always holds if we set $\rho = 1$ as in the DLM case. For other values of ρ, this condition implies that the ratio of the process and the initial variance, ς, must be bounded below by $1-\rho^2$. This condition holds if we set σ_0^2 to be the limiting variance of η_t given by $\sigma_\eta^2/(1-\rho^2)$ as $t > \infty$.

All four conditional variances discussed so far can be proved to be monotonically decreasing functions of spatial correlation ζ, or equivalently, increasing functions of the distance, d_{12} between the data site, $\mathbf{s}_2$, and the prediction site, $\mathbf{s}_1$. In the time series modeling framework, it is worthwhile to investigate whether or not it is possible to make more accurate spatial prediction by conditioning on additional temporal data, that is, whether inequalities such as

$$\mathrm{Var}(Z(\mathbf{s}_1, 2)|Z(\mathbf{s}_2, 2)) > \mathrm{Var}(Z(\mathbf{s}_1, 2)|Z(\mathbf{s}_2, 1), Z(\mathbf{s}_2, 2)), \tag{8}$$

can be expected to hold. The above inequality, however, is always true due to the fact that the conditional variance decreases as the number of conditioning random variables increases in a nested fashion.

A slight reformulation of the above question is more useful in practical modeling. Would the inequality (8) hold if for the prediction problem in the left hand side we ignore the data at time $t = 1$ completely and apply the model at time $t = 2$ as for the first time? In this case, $\mathrm{Var}(Z(\mathbf{s}_1, 2)|Z(\mathbf{s}_2, 2))$ when the model is applied for the first time at $t = 2$ will be exactly the same as $\mathrm{Var}(Z(\mathbf{s}_1, 1)|Z(\mathbf{s}_2, 1))$. Hence, we need to investigate what conditions will guarantee the inequality

$$\mathrm{Var}(Z(\mathbf{s}_1, 1)|Z(\mathbf{s}_2, 1)) - \mathrm{Var}(Z(\mathbf{s}_1, 2)|Z(\mathbf{s}_2, 1), Z(\mathbf{s}_2, 2)) > 0. \tag{9}$$

For the DLM, Dou et al. [5] show that (9) holds if and only if

$$\frac{\sigma_\epsilon^2}{\sigma_0^2} < \frac{\varsigma + 1}{\varsigma}, \tag{10}$$

where $\varsigma = \frac{\sigma_\eta^2}{\sigma_0^2}$ has been defined above. Note that this condition (10) is free of the spatial correlation parameter ζ. We now investigate the conditions under which (9) holds for the AR models.

The analysis for the AR models is more complicated due to the presence of the extra temporal correlation parameter ρ. We consider the following special and limiting cases. Straightforward calculations yield that the variance difference in (9) is negative if $\sigma_0^2 = 0$. In addition, it goes to ∞ as $\sigma_0^2 \to \infty$; hence, large values of σ_0^2 will guarantee that (9) holds. Indeed, it always holds if we set σ_0^2 to be the limiting variance $\sigma_\eta^2/(1-\rho^2)$.

Now it is interesting to investigate what happens if σ_0^2 takes any other value. We can prove that the inequality (9) holds if

$$\frac{\sigma_\epsilon^2}{\sigma_0^2} < \frac{\varsigma + \rho^2}{\varsigma - (1-\rho^2)} \tag{11}$$

when ζ approaches 1 (i.e. for large spatial correlation). Observe that for $\rho = 1$ the above condition reduces to the one for the DLM case (10); this generalizes the result obtained by Dou et al. [5]. Also note that $\frac{\varsigma+\rho^2}{\varsigma-(1-\rho^2)} \geq \frac{\varsigma+1}{\varsigma}$, always, for any value of $0 < \rho^2 < 1$ which shows that the inequality (9) holds for a wider range of parameter values under the AR models than the DLM. We can also prove that when $\zeta \to 0$, the inequality (9) holds if in addition we have $\sigma_0^2 > \sigma_\eta^2/(1-\rho^2)$.

In summary, the AR models are likely to have better properties if the initial variance σ_0^2 is large compared to the process variance σ_η^2. In practical examples where the models are more complex and parameters are unknown, we will not be able to verify the conditions required for the theoretical results, and we must, therefore, rely on empirical evidence. This is where various Bayesian and non-Bayesian model choice criteria can be used for performing model choice. The following section discusses this with several simulations and a real data example.

Please cite this article in press as: S.K. Sahu, K.S. Bakar, A comparison of Bayesian models for daily ozone concentration levels, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.04.009

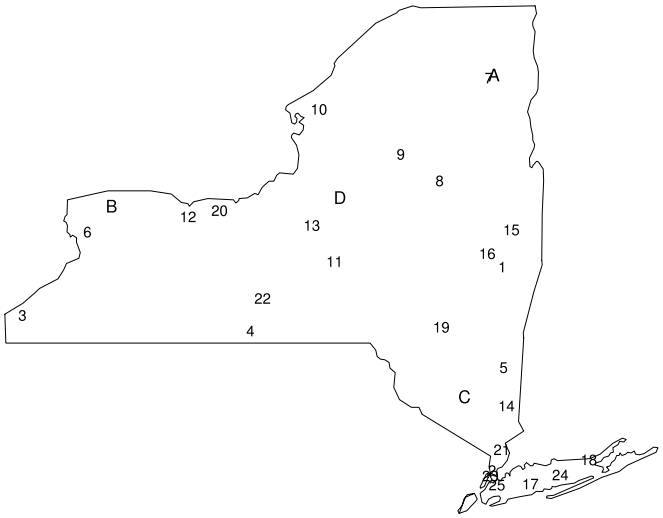

Fig. 1. A map of the 29 ozone monitoring sites in the state of New York. Four randomly chosen sites labeled A, B, C and D are used for validation purposes and the remaining 25 sites (numbered 1–25) are used for modeling.

4. Examples

In this section we compare the DLM and AR models in practical data modeling situations where these models are often implemented. We consider modeling daily maximum eight-hour average ozone concentration data from the 29 ozone monitoring sites in the state of New York for 62 days ($=T$) in the months of July and August in 2006. We shall use data from 25 ($=n$) sites for model fitting, and the data from the remaining 4 ($=m$) sites will be used for model validation purposes. The state of New York is considered as the spatial domain because the ozone monitoring network in this state represents typical practical situations—a cluster of a few sites in and around a big city (the city of New York here) and a moderate number of other sites, situated large distances apart, covering a vast region; see Fig. 1 for a map of New York and the location of the monitoring sites. The data from 62 days in July and August are modeled since these are in the high ozone season (May–August) in the USA. The year 2006 is chosen since that was the latest year for which the output of a computer simulation model (used in our model as a covariate; see below) was available. The spatio-temporal domain considered here represents a moderate computational problem where we can implement the models and obtain results using a reasonable amount of computing time and effort.

In the practical modeling of this section, following Sahu et al. [13], we include as the single covariate the output of a computer simulation model known as the CMAQ (Community Multiscale Air Quality) model. The CMAQ model is based on emission inventories, meteorological information, and land use, and it produces average ozone concentration levels at each cell of a 12 km^2 grid covering the whole of the continental US retrospectively, although there is a version of the model known as Eta-CMAQ which produces forecasts up to two days in advance. In this paper we use the retrospective daily maximum eight-hour average CMAQ ozone concentration for the grid cell covering the monitoring site as the single covariate. The spatial predictions at the unmonitored sites are performed using the CMAQ output at the corresponding grid cells. In our models we have also included other meteorological covariates such as the daily maximum temperature, but none of those turned out to be significant in the presence of the CMAQ output; see Fig. 2 which shows a strong linear relationship between ozone concentration values and the corresponding CMAQ output.

The full Bayesian model is completed by specifying prior distributions for all the unknown parameters. We work with the inverse of the variance components σ^2_ϵ, σ^2_η and σ^2_0 and assume an independent gamma prior distribution with parameters a and b having mean a/b for each of $1/\sigma^2_\epsilon$, $1/\sigma^2_\eta$ and $1/\sigma^2_0$. In our implementation we take $a = 2$ and $b = 1$ implying that these

Please cite this article in press as: S.K. Sahu, K.S. Bakar, A comparison of Bayesian models for daily ozone concentration levels, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.04.009

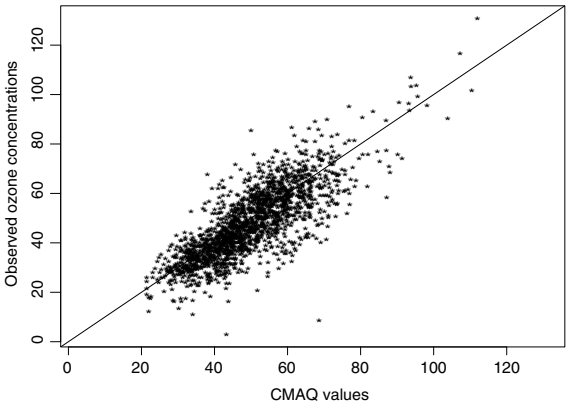

Fig. 2. A scatter plot of daily maximum eight-hour average ozone concentration levels against the CMAQ output for the grid cells covering that monitoring site from 25 sites in New York for 62 days in July and August, 2006. The line $y = x$ is superimposed. The unit of ozone is parts per billion (ppb).

variance components have prior mean 1 and infinite variance. We assign a flat prior $N(0, 10^4)$ for the regression coefficient β. For the common decay parameter, ϕ, we assume an independent uniform prior distribution in (0.001, 1) corresponding to approximate spatial ranges of 3–3000 km. This range adequately covers the state of New York, our study region of interest. We use a Metropolis step to simulate draws from the posterior conditional distribution in our Gibbs sampler implementation. The scale of the proposal distribution in the Metropolis algorithm has been tuned to have a reasonable acceptance rate between 30% and 40%. See [13] for an alternative empirical Bayes method of estimation for ϕ based on minimizing the validation mean square error defined below in (13).

The fully specified Bayesian DLM and AR models cannot be compared using exact analytic methods as was done in Section 3. Hence we use the following practical model selection criteria to compare the models. The predictive model choice criterion (PMCC; see e.g. [6]) is suitable for comparing models with normally distributed error distributions and is given by

$$\text{PMCC} = \sum_{i=1}^{n}\sum_{t=1}^{T} E(Z(\mathbf{s}_i, t)_{\text{rep}} - z(\mathbf{s}_i, t))^2 + \sum_{i=1}^{n} \text{Var}(Z(\mathbf{s}_i, t)_{\text{rep}}), \tag{12}$$

where $Z(\mathbf{s}_i, t)_{\text{rep}}$ denotes a future replicate of the data $Z(\mathbf{s}_i, t)$. The first term in the above is a goodness of fit term (G) while the second is a penalty term (P) for model complexity. The model with the smallest value of PMCC is selected among the competing models. Thus, to be selected a model must strike a good balance between goodness of fit and model complexity. The terms P and G are estimated using composition sampling; at each MCMC iteration k we first draw parameter values from the posterior distribution and then $Z(\mathbf{s}_i, t)^{(k)}_{\text{rep}}$ from the model equations conditional on the drawn parameter values.

To assess the quality of the predictions we define the validation mean square error criterion

$$\text{VMSE} = \frac{1}{mT}\sum_{j=1}^{m}\sum_{t=1}^{T}(\hat{Z}(\mathbf{s}_j, t) - z(\mathbf{s}_j, t))^2 \tag{13}$$

where $\hat{Z}(\mathbf{s}_j, t)$ is the model predicted value of $Z(\mathbf{s}_j, t)$ at time t at the validation site j, and m is the number of validation sites. In the calculations for VMSE, the terms corresponding to the missing observations must be omitted; in such a case the divisor must be adjusted appropriately as well.

We have computed many other model validation criteria such as the mean absolute error. However, the conclusions regarding the model choice and comparison turned to be the same as the ones reported below using PMCC and VMSE.

Please cite this article in press as: S.K. Sahu, K.S. Bakar, A comparison of Bayesian models for daily ozone concentration levels, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.04.009

Table 1
Goodness of fit (G), penalty (P) and VMSE for the DLM and AR models where each model has been fitted to four replicated simulation data sets.

Data set	Simulation model = AR							
	Fitted model							
	AR				DLM			
	G	P	$P+G$	VMSE	G	P	$P+G$	VMSE
1	151.23	556.01	707.24	10.39	472.44	689.03	1161.47	18.57
2	143.68	541.22	684.90	9.82	455.19	642.81	1098.00	17.98
3	164.89	563.09	727.98	11.02	478.33	695.27	1173.60	19.01
4	160.56	557.32	717.88	10.88	474.31	691.48	1165.79	18.80
Data set	Simulation model = DLM							
	Fitted model							
	AR				DLM			
	G	P	$P+G$	VMSE	G	P	$P+G$	VMSE
1	170.22	598.02	768.24	12.33	225.68	482.03	707.71	11.52
2	189.05	611.28	800.33	12.99	243.81	510.78	754.59	12.02
3	155.33	560.75	716.08	11.18	221.03	472.98	694.01	11.05
4	150.78	554.90	705.68	10.92	213.65	459.38	673.03	10.66

4.1. A simulation example

We first provide a simulation example where we test out the two model choice criteria and the MCMC code that we developed for fitting the two sets of models. We simulate four data sets from each of the DLM and AR models. Each data set consists of observations from 29 monitoring sites and 62 days in July and August, 2006. Note that the simulation model includes the CMAQ output as the single covariate. As mentioned above, data from 25 sites will be used for model fitting and the data from the remaining 4 sites will be used for model validation purposes. For both models we set the common value of ϕ at 0.01 for both simulation and fitting. The choice of the simulation model parameters is guided by the practical example provided in the next subsection. For the AR simulation models we set $\rho = 0.2$, $\sigma^2_\epsilon = 0.04$, $\sigma^2_\eta = 0.6$, $\sigma^2_0 = 0.2$, $\mu = 8.0$, $\xi = 1.0$ and $\beta = 0.6$. For the simulation from the DLM we assume $\sigma^2_\epsilon = 0.5$, $\Sigma_\eta = 0.06I$, $\Sigma_0 = 0.2I$ and $\boldsymbol{\mu} = (1.0, 0.6)'$.

Dou et al. [4] have developed a software package *GDLM*.1.0 which is freely available at http://enviro.stat.ubc.ca for implementing the Gibbs sampler for the DLM with $\rho = 1$. To enable the use of this software, and as has been mentioned before, we do not consider the case $|\rho| < 1$ for the DLM in this study. We have developed our own code for implementing the Gibbs sampler for the AR models following Sahu et al. [13]. The details are omitted for brevity. We note that the MCMC chains converge rapidly for both models. We use 15 000 iterates for making inferences after discarding the first 5000 iterations.

Table 1 presents the values of two components, G and P, of PMCC and VMSE for the two models fitted to four replicated simulation data sets from each of the two models. As expected, we see that both PMCC and VMSE (and also G and P) indicate the true simulation model. The two components, G and P, of PMCC also choose the true simulation model. Note also that when data are simulated from the DLM the performance of the incorrectly fitted AR models is not too far away from the DLM. However, when the data are simulated from the AR models the performance of the incorrectly fitted DLM is some distance away from the AR models. Thus the AR models provide reasonably good performance even when data are simulated from the DLM.

We have also calculated some other model choice criteria and model diagnostics such as the mean absolute error criteria and the nominal coverage probability for the 95% prediction intervals. All of those criteria pick the correct simulation model in each case and hence are omitted for brevity. We now proceed to the real data example.

Please cite this article in press as: S.K. Sahu, K.S. Bakar, A comparison of Bayesian models for daily ozone concentration levels, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.04.009

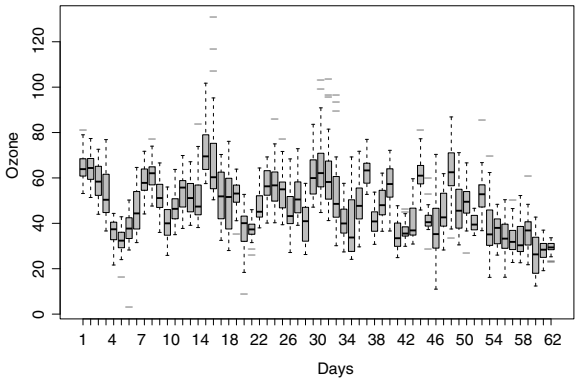

Fig. 3. Boxplot of the daily maximum eight-hour average ozone concentration levels from 25 monitoring sites in New York for 62 days in July and August, 2006.

Table 2
VMSE values for the selected DLM and AR models overall and for the four validation sites.

	A	B	C	D	Overall
AR	34.29	48.47	48.02	53.86	46.16
DLM	51.22	59.67	57.17	65.62	58.42

4.2. The New York data example

We analyze the New York data set obtained from 29 monitoring sites for 62 days in July and August in 2006. Out of these 1798 observations, 80 (4.45%) were found to be missing, which we assume to be at random. In our Bayesian inference setup using MCMC we simply treat these missing values as unknown parameters and simulate from their full conditional distribution at each MCMC iteration.

As mentioned previously, we use data from 25 sites for model fitting and the data from the remaining four sites (labeled A–D in Fig. 1) are used for validation. The boxplot of the data from the 25 monitoring sites is provided in Fig. 3. The plot shows a moderately high level (more than 50 ppb) of ozone concentration values for most days. There is no apparent strong overall trend, although it seems that there is a slight decreasing trend during the last two weeks in August.

The optimal values of the VMSE for the selected DLM and AR models are 58.42 and 46.16, respectively. This shows that the AR models perform much better in model validation than the DLM. In fact, these overall VMSE values are averages of the VMSE values for each of the four validation sites; see Table 2. The VMSE for site D is highest since this is the validation site farthest from its nearest data site; see Fig. 1. This table shows that the AR models outperform the DLM at all four validation sites. Moreover, the values of the PMCC criterion for the selected DLM and AR models are 1066.04 and 735.22, respectively. This also confirms that the AR models are better suited for this particular data set.

Table 3 provides the parameter estimates for the AR model adopted. It shows that the CMAQ output is a significant predictor since β is significant. The temporal correlation parameter ρ is also estimated to be significant. The spatial decay parameter is estimated to be 0.012, corresponding to a range of 250 km. The estimates of the variance components show that on average, the initial variance, σ_0^2, is much larger than the process variance, σ_η^2; hence the theoretical results which required a large initial variance will hold. In particular, we can show that the inequalities (9) and (11) hold for these parameter estimates.

We have also performed forecasting for seven days ahead using both the models. The VMSE for the forecasts was lower for the selected AR model. All these findings provide additional justifications for choosing the AR models for modeling the daily ozone data considered here.

Please cite this article in press as: S.K. Sahu, K.S. Bakar, A comparison of Bayesian models for daily ozone concentration levels, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.04.009

Table 3
Parameter estimates for the selected AR model.

	Mean	95% interval
μ	8.431	(7.582, 8.991)
ξ	1.226	(0.793, 1.811)
ρ	0.198	(0.157, 0.235)
β	0.669	(0.581, 0.734)
σ_ϵ^2	0.048	(0.037, 0.065)
σ_η^2	0.255	(0.198, 0.377)
σ_0^2	0.689	(0.592, 0.768)
ϕ	0.012	(0.009, 0.016)

5. Conclusions

In this paper we have generalized a number of theoretical results obtained by Dou et al. [5] for model comparison purposes. Theoretical results for simple versions of the DLM and AR models show better properties for the AR models under some conditions which have been shown to hold for the practical data example considered in this paper. We have followed the theoretical investigation with a simulation study for a more practical version of the models. As expected, the simulation study shows better performance of the DLM when the data are simulated from it. Similarly, the AR models are seen to be better when the data are simulated from it. These results have been observed for four replicated simulation data sets.

Finally, we have compared the models by fitting them to a real data set for daily maximum eight-hour average ozone concentration levels in the state of New York for 62 days in July and August, 2006. A predictive Bayesian model choice criterion as well as setting aside validation data show that the fitted AR model performs much better than the fitted DLM. These results show that the AR models can be much better than the DLM in practical ozone data modeling situations. Note that these practical results are only valid for the simple version of the DLM considered here. Further investigation comparing the AR models with a more flexible version of the DLM is likely to produce additional useful results.

Acknowledgments

The authors thank Dr. David M. Holland for many helpful suggestions and also for his help in acquiring the New York data set.

References

[1] P.J. Brown, N.D. Le, J.V. Zidek, Multivariate spatial interpolation and exposure to air pollutants, The Canadian Journal of Statistics 22 (1994) 489–510.
[2] R.J. Carroll, R. Chen, E.I. George, T.H. Li, H.J. Newton, H. Schmiediche, N. Wang, Ozone exposure and population density in Harris County, Texas, Journal of the American Statistical Association 92 (1997) 392–404.
[3] W.M. Cox, S.H. Chu, Meteorological adjusted trends in urban areas, a probabilistic approach, Atmospheric Environment 27 (1992) 425–434.
[4] Y. Dou, N.D. Le, J.V. Zidek, A dynamic linear model for hourly ozone concentrations, Technical Report, 228, University of British Columbia, 2007.
[5] Y. Dou, N.D. Le, J.V. Zidek, Modeling hourly ozone concentration fields, Annals of Applied Statistics 4 (2010) 1183–1213.
[6] A.E. Gelfand, S.K. Ghosh, Model choice: a minimum posterior predictive loss approach, Biometrika 85 (1998) 1–11.
[7] P. Guttorp, W. Meiring, P.D. Sampson, A space–time analysis of ground-level ozone data, Environmetrics 5 (1994) 241–254.
[8] G. Huerta, B. Sanso, J.R. Stroud, A spatiotemporal model for Mexico City ozone levels, Journal of the Royal Statistical Society. Series C 53 (2004) 231–248.
[9] N.D. Le, J.Z. Zidek, Statistical Analysis of Environmental Space–Time Processes, Springer, 2006.
[10] N. McMillan, S.M. Bortnick, M.E. Irwin, M. Berliner, A hierarchical Bayesian model to estimate and forecast ozone through space and time, Atmospheric Environment 39 (2005) 1373–1382.
[11] S.K. Sahu, A.E. Gelfand, D.M. Holland, High-resolution space–time ozone modeling for assessing trends, Journal of the American Statistical Association 102 (2007) 1221–1234.
[12] S.K. Sahu, K.V. Mardia, A Bayesian kriged-Kalman model for short-term forecasting of air pollution levels, Journal of the Royal Statistical Society. Series C 54 (2005) 223–244.

Please cite this article in press as: S.K. Sahu, K.S. Bakar, A comparison of Bayesian models for daily ozone concentration levels, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.04.009

[13] S.K. Sahu, S. Yip, D.M. Holland, Improved space–time forecasting of next day ozone concentrations in the eastern US, Atmospheric Environment 43 (2009) 494–501.
[14] J.R. Stroud, P. Müller, B. Sansó, Dynamic models for spatio-temporal data, Journal of the Royal Statistical Society. Series B 63 (2001) 673–689.
[15] M. West, J. Harrison, Bayesian Forecasting and Dynamic Models, Springer, New York, 1997.
[16] C.K. Wikle, Hierarchical models in environmental science, International Statistical Review 71 (2003) 181–199.
[17] J. Zheng, J.L. Swall, W.M. Cox, J.M. Davis, Interannual variation in meteorologically adjusted ozone levels in the eastern United States: a comparison of two approaches, Atmospheric Environment 41 (2007) 705–716.

Please cite this article in press as: S.K. Sahu, K.S. Bakar, A comparison of Bayesian models for daily ozone concentration levels, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.04.009