

Assessing exceedance of ozone standards: a space-time downscaler for fourth highest ozone concentrations[†]

V. J. Berrocal^a, A. E. Gelfand^{b*} and D. M. Holland^c

The US Environmental Protection Agency is required to monitor, regulate, and set the national ambient air quality standard for ozone. To investigate ozone exposure, the Environmental Protection Agency utilizes monitoring devices along with estimates of gridded ground level ozone concentration produced by a deterministic air quality model, the Community Multiscale Air Quality Model. These two sources of information enable inference regarding spatial exceedance of the national ambient air quality standard (NAAQS) for ozone, which is given in terms of the level of the annual fourth highest ozone concentration.

Here, we extend previous downscaling work to propose a spatial fourth highest extreme value downscaling model to assimilate annual fourth highest ozone concentration data at geo-coded locations with estimates at grid cell level derived from the Community Multiscale Air Quality Model model output. The resulting inference enables us to make probabilistic statements, with associated uncertainty, about the spatial variation in the chance of exceeding the standard. We apply our approach to data in the Eastern USA during years 2001–2008 and compare its predictive performance to that of downscaler models based on Gaussian processes applied to daily data. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: change of support; data fusion; hierarchical modeling; Markov chain Monte Carlo; national ambient air quality standards (NAAQS); r -th largest order statistic distribution

1. INTRODUCTION

Under the Clean Air Act, the US Environmental Protection Agency (US EPA) establishes air quality standards to protect public health and welfare. Through this legislation, the charge to the US EPA is to monitor, set, and revise the national ambient air quality standards (NAAQS) for ozone and five other principal air pollutants that have been found to be harmful to the environment and human health. The preponderance of epidemiological studies have shown that exposure to elevated levels of ground level ozone is associated with increased risks of cardiovascular and respiratory diseases (Bell *et al.*, 2004; Zhu *et al.*, 2003; Zanobetti and Schwartz, 2008). Thus, starting from 1979, the EPA has set the NAAQS for ozone based on extreme levels of ozone concentration. The current NAAQS for ozone, revised in 2008, states that the annual fourth highest daily 8-h maximum ozone concentration, averaged over three consecutive years, should not exceed 75 parts per billion (ppb).

Currently, exceedances of the ozone NAAQS are determined using monitoring sites data. Thus, information on ozone exceedances exists only at monitoring sites sparsely located throughout the USA. To investigate the spatial and temporal variability in exceedance of the NAAQS for ozone over the Eastern USA, in this paper, we extend previously developed methodology (Berrocal *et al.*, 2012) and introduce a spatio-temporal process model for the annual fourth highest ozone concentration. Working within a Bayesian hierarchical modeling framework, we obtain posterior exceedance probabilities with regard to the foregoing standard. In particular, taking advantage of our previous experience with downscaling (Berrocal *et al.*, 2012), we employ a smoothed downscaler to fuse two data sources and obtain improved predictions, with associated uncertainty, of the annual fourth highest ozone concentration. This allows us to generate probabilistic maps of the attainment of the NAAQS for ozone and better identify areas in the Eastern USA that are very likely to exceed the EPA standard. To our knowledge, this is the first model that combines ozone monitoring data with air quality model output to directly predict the variable related to the NAAQS for ozone; previous efforts inferred this variable using models for daily ozone concentration.

The US EPA uses two sources of spatial information to track progress of ozone air pollution: measurements taken at monitoring sites, which essentially provide the true daily ozone concentration but are available only at sparse locations, possibly not every day, and estimates

* Correspondence to: A. E. Gelfand, Department of Statistical Science, Duke University, Durham, NC 27707, U.S.A. E-mail: alan@stat.duke.edu

a Department of Biostatistics, University of Michigan, Ann Arbor, MI 48105, U.S.A.

b Department of Statistical Science, Duke University, Durham, NC 27707, U.S.A.

c National Exposure Research Laboratory, US Environmental Protection Agency, Research Triangle Park, NC 27709, U.S.A.

[†]This article is published in Environmetrics as part of a Special on Issue Physical-statistical modelling; guest-edited by Petra Kuhnert, CSIRO Computational Informatics, PMB 2, Glen Osmond, SA 5064 Australia.

of ozone concentration obtained from deterministic air quality models such as the Community Multiscale Air Quality model (CMAQ) (<http://www.cmaq-model.org/>). By numerically solving systems of partial differential equations representing various diffusion, chemical, and atmospheric processes, air quality models yield estimates of the average daily ozone concentration over grid cells covering the spatial domain with no missingness. However, they might present calibration concerns. To address shortcomings in both data sources, in recent years, considerable effort has been devoted to develop statistical models that combine the two sources of information and explicitly address the difference in spatial scale (grid cells versus point locations) between the two sources.

Two main approaches have emerged to “fuse” the monitoring station data with the air quality model output while solving the “change of support” problem (Cressie (1993); Cressie and Wikle (2011); Gotway and Young (2002); Banerjee *et al.* (2004), chapter 6): (i) Bayesian melding, (Fuentes and Raftery, 2005; McMillan *et al.*, 2010) which is a form of *upscaling* and (ii) *downscaling* through spatial regression modeling approaches (Guillas *et al.*, 2008; Berrocal *et al.*, 2010a, 2010b; Liu *et al.*, 2012). Downscaling has been shown to be computationally feasible for large numbers of grid cells over space and time, and to generate better predictions of daily ozone concentration than other geostatistical interpolation methods (Berrocal *et al.*, 2010b).

Recently, Berrocal *et al.* (2012) have introduced a smoothed downscaler model that regresses the monitoring station data on a derived regressor obtained by smoothing the entire air quality model output with weights that are random and spatially varying. This downscaler provides improved out-of-sample predictions relative to the initial downscaler model at the daily scale and is currently being used to obtain surfaces of daily ozone concentration over the Eastern USA (http://www.epa.gov/esd/land-sci/lcb/lcb_faqs.html).

We see an inherent limitation to extrapolating spatial surfaces associated with the standard using daily data, and the problem is illuminated in the comparison between our model and the daily extrapolation approach. In simple terms, if you want to interpolate a particular variable, why not model it directly? This is particularly true when the variable of interest is *not* a linear function of the daily data such as the fourth highest daily 8-h maximum. More precisely, if we interpolate (average) the daily values and then extrapolate a fourth highest value (like taking a maximum), we achieve something smaller than if we obtain the set of fourth highest values (like taking a maximum) and then interpolate (average) because the maximum of averages is less than or equal to averages of maxima. Furthermore, both the downscaler and the smoothed downscaler model (Berrocal *et al.*, 2010b, 2012) were developed for daily ozone concentration (in fact, daily 8-h maximum ozone concentration) and assume that the monitoring station data, conditional on the air quality model output, is a realization of a Gaussian process. This distributional assumption might be appropriate at daily scales but not for directly modeling the annual fourth highest ozone concentration. If we model this variable, we should work with a process model for the fourth order statistic.

We note recent related work by Reich *et al.* (2013). However, Reich *et al.* (2013) focuses more on calibration of CMAQ model output (in fact, driven by the explicit specification of the so-called reduced form CMAQ model) and on evaluation of emission control programs for ozone. Conversely, we focus directly on fusing observations and computer model output of annual fourth highest daily maxima to study exceedance probabilities. Additionally, while Reich *et al.* (2013) investigate local quantiles of a generalized Pareto distribution, in other words the distribution of daily exceedances using a random local exceedance threshold for a single year, we look at the behavior of exceedance of the NAAQS over an 8-year window. Finally, we consider a region that is three times the size of that in Reich *et al.* (2013) who examined the portion of the Eastern USA delimited by the states of Virginia, Kentucky, Alabama, and Georgia, and we obtain predictions using a single model rather than adopting a computational approximation by fitting in two stages. Other recent work on spatial exceedances appears in French and Sain (2013) who, working with rainfall data, threshold a Gaussian process to obtain an exceedance region of a specified probability.

In this paper, we work with a fourth largest order statistic model that we label as the fourth highest extreme value (FHEV) distribution (drawing on Coles (2001), p. 68-69). Using this model, we follow Sang and Gelfand (2009) and assume that, conditionally on three latent Gaussian processes, the annual fourth highest ozone concentrations at point locations are independent and distributed according to FHEV distributions with spatially varying parameters. The interpretation of this assumption is that the centering/location process for the fourth highest concentrations is smooth and the observed point-level fourth highest are realizations of this process subject to *measurement error* through the FHEV. The FHEV parameters are assumed to depend on the smoothed estimated annual fourth highest ozone concentration derived from the air quality model output, where the smoothing is motivated above and is obtained, as in Berrocal *et al.* (2012), through random and spatially varying weights. We call this model the FHEV smoothed downscaler model.

The paper is organized as follows. In Section 2, we present the ozone concentration data used in our study, while in Section 3, we present the FHEV theory and introduce the FHEV smoothed spatial downscaler drawing from our previous contributions (Berrocal *et al.*, 2010b, 2012). In Section 4, we show results on the predictive performance of the FHEV smoothed downscaler model, and finally, in Section 5, we conclude with a brief discussion.

2. OZONE CONCENTRATION DATA

In the USA, ozone concentration is monitored at over 1000 locations over the conterminous US in a variety of urban, suburban, and rural environments. Daily 8-h maximum ozone concentration (henceforth, daily ozone concentration) is observed mostly during the summer months when ground level ozone concentration tends to be higher and thus more harmful to public health. The EPA supplements the information collected at monitoring sites with the information contained in the outputs of numerical air quality models. In our application, we will use the output of the Models-3/Community Multiscale Air Quality (CMAQ) model (Byun and Schere, 2006), a deterministic numerical model that estimates ozone concentration over grid cells of prespecified dimensions by simulating various chemical and physical processes, such as horizontal and vertical advection, emission injections, deposition, and plume chemistry effects. In particular, we will use CMAQ outputs of average daily ozone concentration over 12 km × 12 km grid cells covering the Eastern USA for every day between 1 January 2001 and 31 December 2008.

The extent of the spatial domain covered by CMAQ has changed over the course of our study time period: in 2001, the CMAQ output covered longitudes $67^{\circ}1'W$ to $100^{\circ}3'W$ and latitudes $25^{\circ}9'N$ to $48^{\circ}7'N$ and consisted of 40,044 $12\text{ km} \times 12\text{ km}$ grid cells arranged in 188 rows and 213 columns. From 2002 to 2006, the CMAQ output rose to 66,960 grid cells, arranged as 240 rows by 279 columns. And in 2007–2008, the domain was extended, covering longitudes from $65^{\circ}4'W$ to $111^{\circ}0'W$ and latitudes from $23^{\circ}0'N$ to $51^{\circ}2'N$, yielding 137,241 grid cells, arranged as 299 rows by 459 columns. In order to employ a common set of grid cells across the years, we have set our spatial domain S to be the 2001 CMAQ domain, and we have utilized daily ozone concentration data only from monitoring stations and CMAQ grid cells that lie in this domain.

More than one thousand monitoring stations, specifically 1035, are located within the boundaries of our spatial domain. Of these, 437 had more than 25% missing data during the summer months in any year between 2001 and 2008 and thus were discarded from the analysis. This proportion is arbitrary, and it may slightly bias downward the fourth highest daily maximum. However, it enables us to retain more than half of the monitoring stations and comparison of CMAQ output on missing versus nonmissing days suggests that missingness is completely at random. For each of the remaining 598 stations, we derived, for each year, the annual fourth highest ozone concentration, obtaining a total of 598 time series, each with 8 time points. We randomly split the monitoring sites into two sets: a training set made of 550 stations and a validation set made of 48 sites. Figure 1 shows the locations of the training and validation sites, while Figure 2(a) and Figure 2(b) display the annual fourth highest ozone concentration observed in years 2001 and 2008, respectively. Figure 2(a) and Figure 2(b) reveal that the average fourth highest ozone concentration decreased from 83.9 ppb in 2001 to 71.2 ppb in 2008. The figures also reveal a great deal of variability across the spatial domain: concentrations are usually lower in the Southeastern and Northwestern parts of the domain while they are elevated in the Northeast. However, the spatial gradient in the annual fourth highest ozone concentration is more striking in 2001 than in 2008. In fact, the standard deviation for the annual fourth highest ozone concentration changed from 9.4 ppb in 2001 to 7.6 ppb in 2008.

Daily ozone concentration is usually modeled on the square root scale using a Gaussian distribution (Carroll *et al.*, 1997; Sahu *et al.*, 2006, 2007; Berrocal *et al.*, 2010a, 2010b, 2012) to encourage stable variance. We also work on the square root scale for the fourth highest in order to provide more stable computation, converting back to the original scale to present inference displays.

For the CMAQ data, we proceeded similarly as for the monitoring data. Hence, for each CMAQ grid cell in S and for each year, we derived the annual fourth highest ozone concentration as output by CMAQ. Spatial maps of the annual fourth highest ozone concentration predicted by CMAQ for years 2001 and 2008 are displayed in Figure 3(a) and Figure 3(b), respectively. In agreement with the monitoring data, CMAQ also estimates higher levels of extreme ozone concentration in the Northeast and lower levels in the Southeast for both years.

There is potential temporal misalignment between the day when the annual fourth highest is observed at a monitoring site and when it occurs in the CMAQ grid cell where the site lies. Evidently, such misalignment will occur among the monitoring stations and among the CMAQ grid cells themselves. This is a well known issue in multivariate modeling of extremes (Resnick and Rubinovitch, 1973). In any event, with interest only in local prediction of the annual fourth highest using CMAQ, this may not be a concern. Moreover, in this regard, we computed the correlation between the square root of the annual fourth highest ozone concentration observed at monitoring sites and the square root of the estimated annual fourth highest ozone concentration at the CMAQ grid cells where the sites lie. We obtained moderate correlation ranging from 0.49 (in 2001) to 0.70 (in 2008), suggesting benefit in using it in our proposed downscaler model. Formal evaluation of the predictive gain obtained by incorporating the CMAQ output is carried out in Section 4 where we compare the predictive performance of our FHEV smoothed downscaler model to that of other interpolation models using only the monitoring station data.

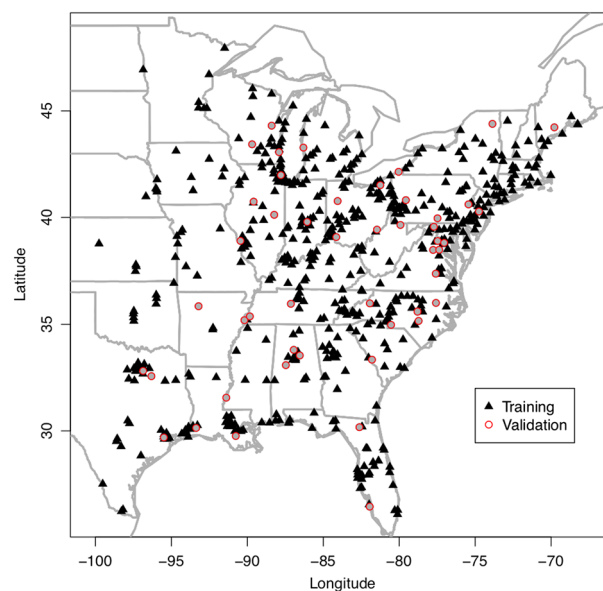


Figure 1. Training and validation sites used to fit the fourth highest extreme value smoothed downscaler model

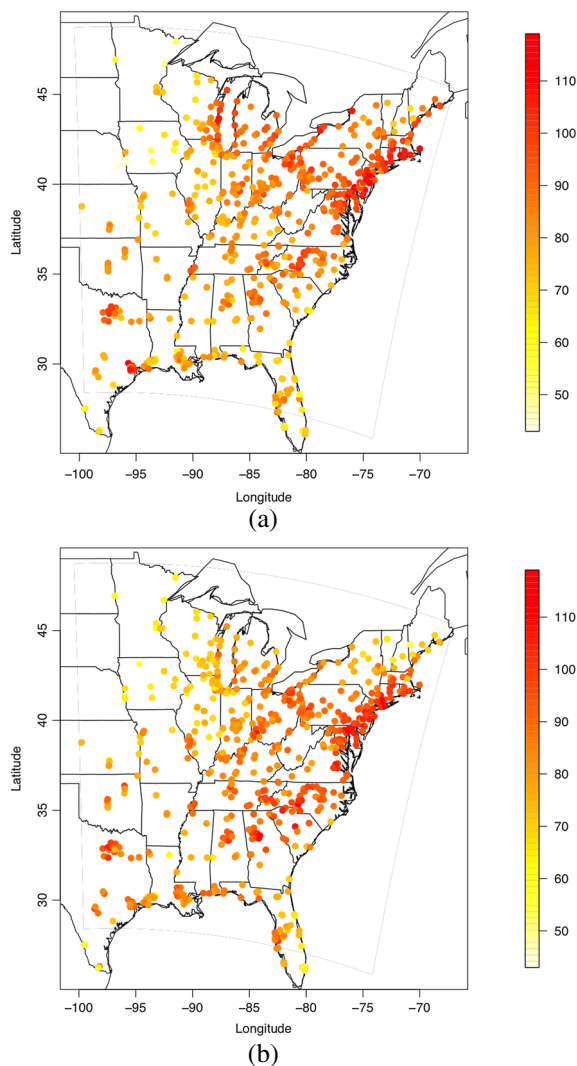


Figure 2. (a, b) Observed annual fourth highest ozone concentration in parts per billion (ppb) at monitoring sites (training and validation sites) in years 2001 and 2008, respectively

3. MODELING

We present our FHEV smoothed downscaler model for the fourth highest under the assumption that a generalized extreme value (GEV) model holds for the maximum. In Section 3.1, we provide the FHEV distribution and its properties. Then, in Section 3.2, we detail our spatial model using the FHEV. Finally, in Section 3.3, we discuss computational issues related to model fitting.

3.1. Fourth highest extreme value distribution

Extreme value theory establishes the limiting form of the distribution of rescaled maxima of independent univariate random variables. Precisely, if Y_1, Y_2, \dots, Y_n are independent and identically distributed univariate random variables with a distribution F (not necessarily known), and there exist two sequences of real numbers $\{a_n\} > 0$ and $\{b_n\}$ such that the limiting distribution G of the rescaled maxima $\frac{\max(Y_1, \dots, Y_n) - b_n}{a_n}$ is non-degenerate for $n \rightarrow \infty$, then G must belong to a very specific family of distributions, the GEV distribution. The GEV distribution depends on three parameters: the location, $\mu \in \mathbf{R}$, the scale, $\sigma > 0$, and the shape parameter, $\xi \in \mathbf{R}$; its general expression is given by

$$G(y) = \begin{cases} \exp \left\{ - \left[1 + \frac{\xi(y-\mu)}{\sigma} \right]_+^{-\frac{1}{\xi}} \right\} & \xi \neq 0 \\ \exp \left\{ - \exp \left[- \frac{(y-\mu)}{\sigma} \right] \right\} & \xi = 0 \end{cases}$$

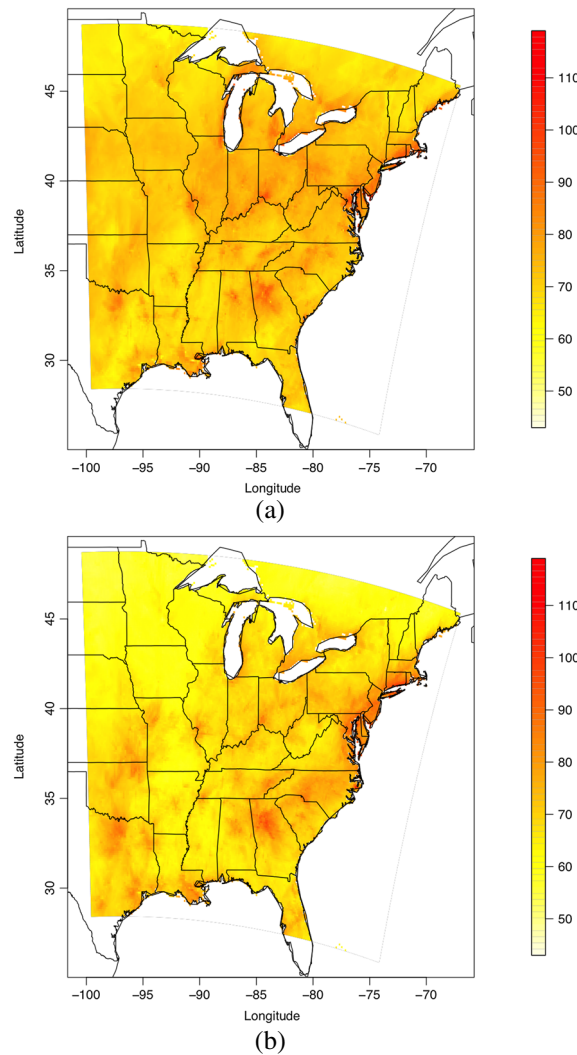


Figure 3. (a, b) Annual fourth highest ozone concentration in parts per billion (ppb) as estimated by CMAQ in years 2001 and 2008, respectively

whence the support for y becomes $\{y : 1 + \frac{\xi(y-\mu)}{\sigma} > 0\}$. Three distributions comprise the GEV family: the reverse Weibull ($\xi < 0$), the Gumbel ($\xi = 0$), and the Fréchet ($\xi > 0$).

Following Coles (Coles, 2001, pp. 66–67), under the conditions that provide an asymptotic GEV distribution for the maximum, the asymptotic distribution of the k th largest order statistic can be obtained. In particular, with $w = \left[1 + \frac{\xi(y-\mu)}{\sigma}\right]$, the cdf of the k th largest order statistic is $F^{(k)}(w) = \exp\left(-w^{-\frac{1}{\xi}}\right) \sum_{j=0}^{k-1} w^{-\frac{j}{\xi}}/j!$. So, when $k = 4$, we have

$$F^{(4)}(w) = \exp\left(-w^{-\frac{1}{\xi}}\right) \left(1 + w^{-\frac{1}{\xi}}/1! + w^{-\frac{2}{\xi}}/2! + w^{-\frac{3}{\xi}}/3!\right)$$

Differentiation and some algebra provide the density (needed for likelihood evaluation in the modeling) as follows:

$$f^{(4)}(w) = \frac{1}{\xi} \exp\left(-w^{-\frac{1}{\xi}}\right) w^{-\left(\frac{4}{\xi}+1\right)}/3!$$

which we refer to as the FHEV distribution. For instance, at $\xi = 1$, we obtain an inverse Gamma, $IG(4, 1)$. More generally, if $z = w^{\frac{1}{\xi}}$, then $z \sim IG(4, 1)$.

3.2. Fourth highest extreme value smoothed downscaler

As noted earlier, following modeling approaches for daily ozone concentration data (Carroll *et al.*, 1997; Sahu *et al.*, 2006, 2007; Berrocal *et al.*, 2010a, 2010b, 2012) and maxima of ozone concentration (Eastoe, 2009; Eastoe and Tawn, 2009), we apply a square root transformation to the annual fourth highest ozone concentration. We denote with $Y_t(\mathbf{s})$, $t = t_0, \dots, T$, the square root of the annual fourth highest ozone concentration observed at site $\mathbf{s} \in \mathcal{S}$ in year t , while for $t = t_0, \dots, T$, we denote with $\{x_t(B_k) : k = 1, \dots, g\}$ the set of $g = 40,044$

values representing the square roots of the annual fourth highest ozone concentrations in year t in grid cells $B_k, k = 1, \dots, g$, as estimated by CMAQ. We model the square roots of the annual fourth highest ozone concentrations at sites $\mathbf{s} \in \mathcal{S}$ as conditionally independent random variables following a FHEV distribution.

A quite general model would take the form,

$$Y_t(\mathbf{s})|\mu_t(\mathbf{s}), \sigma_t(\mathbf{s}), \xi_t(\mathbf{s}) \stackrel{ind}{\sim} \text{FHEV}(\mu_t(\mathbf{s}), \sigma_t(\mathbf{s}), \xi_t(\mathbf{s})) \quad t = t_0, \dots, T$$

where $\mu_t(\mathbf{s}), \sigma_t(\mathbf{s})$, and $\xi_t(\mathbf{s})$ vary in space and time.

First, we relate the location parameter $\mu_t(\mathbf{s})$ to the CMAQ output for years $t = t_0, \dots, T$, by modeling $\mu_t(\mathbf{s})$ as a Gaussian Process, that is,

$$\mu_t(\mathbf{s}) = \beta_{0t} + \beta_{0t}(\mathbf{s}) + \beta_{1t}\tilde{x}_t(\mathbf{s}) \tag{1}$$

where the $\beta_{0t}(\mathbf{s})$ are independent mean zero Gaussian processes with exponential covariance functions having common marginal variance τ_μ^2 and a time-varying decay parameter ϕ_{μ_t} . The regressor $\tilde{x}_t(\mathbf{s})$ in (1) is a smoothed version of the entire set of square roots of annual fourth highest ozone concentrations $\{x_t(B_k) : k = 1, \dots, g\}$ derived from the CMAQ daily output. It is defined for each $t = t_0, \dots, T$, and $\mathbf{s} \in \mathcal{S}$ as a weighted average

$$\tilde{x}_t(\mathbf{s}) = \sum_{k=1}^g w_k(\mathbf{s}, t) \cdot x_t(B_k) \tag{2}$$

where the weights $w_k(\mathbf{s}, t)$ are random, spatially varying, and time-varying. As in Berrocal *et al.* (2012), the weights $w_k(\mathbf{s}, t)$ are in turn constructed by mixing independent copies of a mean-zero, unit-variance latent Gaussian process (GP) provided with an exponential correlation function with decay parameter ϕ_Q , with an exponential kernel.

More specifically, if $\{\mathbf{r}_k : k = 1, \dots, g\}$ denotes the set of CMAQ grid cell centroids for B_1, \dots, B_g , and $\{Q_t(\mathbf{r}_k) : k = 1, \dots, g\}$ are realizations of the latent Gaussian processes $Q_t(\mathbf{r})$ at the CMAQ centroids, then for $k = 1, \dots, g, t = t_0, \dots, T$ and $\mathbf{s} \in \mathcal{S}$,

$$w_k(\mathbf{s}, t) = \frac{K(\mathbf{s} - \mathbf{r}_k; \psi) \cdot \exp(Q_t(\mathbf{r}_k))}{\sum_{l=1}^g K(\mathbf{s} - \mathbf{r}_l; \psi) \cdot \exp(Q_t(\mathbf{r}_l))} \tag{3}$$

where $K(\mathbf{s} - \mathbf{r}_k; \psi) = \exp(-\psi \|\mathbf{s} - \mathbf{r}_k\|)$ is the exponential kernel with decay parameter ψ . The decay parameter ψ determines the magnitude of the neighborhood of grid cells around \mathbf{s} that contribute non-negligibly to the sum in (3). It is possible to estimate it from the data. However, following Berrocal *et al.* (2012), during model fitting, we keep it fixed and equal to 0.08, corresponding to a neighborhood of three grid cells. Essentially, only grid cells that are first, second, and third order neighbors to the grid cell where a site \mathbf{s} lies, receive a nonzero weight $w_k(\mathbf{s}, t)$. In model fitting, we keep ϕ_Q fixed and set a priori to be larger than ψ and equal to 0.12. This translates into a correlation that decays to a value of 0.05 at 24 km, that is, at two 12-km grid cells. We find, in our analysis in Section 4, that this model is best among those we consider. Berrocal *et al.* (2012) offer further discussion of the $Q_t(\mathbf{s})$ surfaces and the induced $w_k(\mathbf{s}, t)$.

We only employ the set of $\{x_t(B_k) : k = 1, \dots, g\}$ to model the location parameter in the FHEV resulting in the GP model for $\mu_t(\mathbf{s})$ above. Exploratory data analysis revealed lack of spatial variability in the estimates of $\sigma_t(\mathbf{s})$ and $\xi_t(\mathbf{s})^\dagger$, and due to known difficulties in estimating the shape parameter in GEV (hence, FHEV) models (Cooley *et al.*, 2007), we hold $\xi_t(\mathbf{s})$ fixed in space and time and equal to ξ , while we let $\sigma_t(\mathbf{s}) > 0$ vary in time but not in space, with a lognormal distribution for each $\sigma_t, t = t_0, \dots, T$. In summary, the first two stages, that is, the data model and the process model, of our hierarchical FHEV smoothed downscaler model are given by, for $\mathbf{s} \in \mathcal{S}, t = t_0, \dots, T$:

$$\begin{aligned} Y_t(\mathbf{s})|\mu_t(\mathbf{s}), \sigma_t(\mathbf{s}), \xi_t(\mathbf{s}) &\stackrel{ind}{\sim} \text{FHEV}(\mu_t(\mathbf{s}), \sigma_t(\mathbf{s}), \xi_t(\mathbf{s})) \\ \mu_t(\mathbf{s})|\beta_{0t}, \beta_{1t}, \tilde{x}_t(\mathbf{s}), \tau_\mu^2, \phi_{\mu_t} &\sim \text{GP}(\beta_{0t} + \beta_{1t}\tilde{x}_t(\mathbf{s}), \tau_\mu^2 \cdot \rho(\cdot, \cdot; \phi_{\mu_t})) \\ \tilde{x}_t(\mathbf{s}) &= \sum_{k=1}^g w_k(\mathbf{s}, t) \cdot x(B_k, t) \\ w_k(\mathbf{s}, t) &= \frac{K(\|\mathbf{s} - \mathbf{r}_k\|; \psi) \cdot \exp(Q(\mathbf{r}_k))}{\sum_{l=1}^g K(\|\mathbf{s} - \mathbf{r}_l\|; \psi) \cdot \exp(Q(\mathbf{r}_l))} \\ Q_t(\mathbf{s})|\phi_Q &\stackrel{ind}{\sim} \text{GP}(0; \rho(\cdot, \cdot; \phi_Q)) \\ \sigma_t(\mathbf{s}) &= \sigma_t \\ \log(\sigma_t) &\stackrel{ind}{\sim} N(\eta_t, v^2) \\ \xi_t(\mathbf{s}) &= \xi \end{aligned}$$

Prior specifications are discussed in Berrocal *et al.* (2012).

If in (1), β_{1t} is set to 0 for all $t = t_0, \dots, T$, and $\mathbf{s} \in \mathcal{S}$, then the FHEV smoothed downscaler model reduces to a geostatistical interpolation model, which ignores the information in CMAQ. It incorporates spatial dependence in the spatially varying parameters of

[†]Since the σ 's and ξ 's in the FHEV are the same as those for the associated GEV, we can use standard software, the R package `ismev` version 1.3.8 developed by J. E. Heffernan and A. C. Stephenson and maintained by E. Gilleland (<http://www.ral.ucar.edu/~ericg/softextreme.php>), on the maxima to look at spatial and temporal variation in the σ 's and ξ 's.

the FHEV distribution but assumes conditional independence in the square roots of the annual fourth highest ozone concentrations. The predictive performance of this latter model, along with that of our smoothed downscaler model, is assessed in Section 4.

3.3. Model fitting

We fit the smoothed FHEV downscaler model using a Markov chain Monte Carlo (MCMC) algorithm. With a first stage likelihood given by products of FHEV distributions, and with Gaussian process priors on the location parameter, the full conditionals for each n -dimensional vector $\boldsymbol{\mu}_t = (\mu_t(\mathbf{s}_1), \dots, \mu_t(\mathbf{s}_n))'$, $t = t_0, \dots, T$, are not available in closed form. Hence, to sample each $\boldsymbol{\mu}_t$, $t = t_0, \dots, T$, from its posterior distribution, we used a Metropolis–Hastings algorithm with a multivariate normal proposal distribution (of dimension n) centered around the current values of $\boldsymbol{\mu}_t$ and with a scaled identity matrix for the covariance where the overall standard deviation was chosen to yield acceptance rates between 20% and 40%. Similarly, to update the scale parameters σ_t , $t = t_0, \dots, T$, we used a Metropolis–Hastings algorithm with a lognormal proposal distribution. Updating the spatially and time-varying weights also requires the use of a Metropolis–Hastings algorithm because the full conditionals of the latent g -dimensional vectors $\mathbf{Q}_t = (Q_t(\mathbf{r}_k) : k = 1, \dots, g)$, $t = t_0, \dots, T$ are not available in closed form.

Due to the large dimensionality of these vectors ($g = 40,044$), we adopted the predictive process framework of Banerjee *et al.* (2008) and Finley *et al.* (2009). Hence, we replaced each g -dimensional vector \mathbf{Q}_t , $t = t_0, \dots, T$, with its projection $\tilde{\mathbf{Q}}_t$, $t = t_0, \dots, T$, onto the lower dimensional space spanned by the g^* -dimensional vector ($g^* = 648$) $\mathbf{Q}_t^* = (Q_t(\mathbf{r}_l^*) : l = 1, \dots, g^*)$ where $\mathbf{r}_1^*, \dots, \mathbf{r}_{g^*}^*$, $g^* \ll g$ are knots systematically selected from the CMAQ grid cell centroids every eight rows and eight columns. Six hundred forty-eight knots selected in a space-filling fashion exceeds the typical number of knots adopted for this approximation.

Updating the parameters of the spatially varying location parameters $\mu_t(\mathbf{s})$, $t = t_0, \dots, T$, of the FHEV distributions is straightforward. With conjugate priors, full conditionals for β_{0t} , β_{1t} , τ_{μ}^2 , τ_{σ}^2 , ϕ_{μ_t} and ϕ_{σ_t} , $t = t_0, \dots, T$, are available in closed form. Thus, sampling from their posterior distributions is carried out through a Gibbs sampling update.

4. RESULTS

4.1. Model comparisons

We evaluate the predictive performance of our FHEV smoothed downscaler model using the annual fourth highest ozone concentration data described in Section 2. Again, we fit the model to $n = 550$ monitoring stations and generate predictions of annual fourth highest ozone concentrations over years 2001–2008 at the $m = 48$ validation sites. We compare the out-of-sample predictive performance of the FHEV smoothed downscaler model with that of other geostatistical interpolation and downscaler models with the goal of addressing the following modeling questions:

1. Is CMAQ useful when predicting annual fourth highest ozone concentrations?
2. Does the information contained in the fourth highest ozone concentrations over a neighborhood of CMAQ grid cells improve prediction at a site \mathbf{s} compared with prediction using only the information contained in the CMAQ grid cell where the site lies?
3. According to ξ , which FHEV distribution (reverse Weibull, Gumbel, or Fréchet) as implied by the associated GEV is more appropriate to model annual fourth highest ozone concentration?
4. As a *straw man*, how well does a Gaussian process model perform in predicting the annual fourth highest ozone concentration?
5. What happens if we predict the annual fourth highest ozone concentrations at sites $\mathbf{s} \in \mathcal{S}$ by modeling daily ozone concentration using a Gaussian process and extrapolate levels of annual fourth highest ozone concentrations from predicted daily ozone concentrations?

To address the first three points above, we consider two additional models: (i) a FHEV downscaler model where the smoothed regressor $\tilde{x}_t(\mathbf{s})$, appearing in the mean function for the location parameter process $\mu_t(\mathbf{s})$, is replaced by $x_t(B_k)$ when $\mathbf{s} \in B_k$, that is,

$$\mu_t(\mathbf{s}) | \beta_{0t}, \beta_{1t} \stackrel{ind}{\sim} GP \left(\beta_{0t} + \beta_{1t} x_t(B_k), C \left(\mathbf{s}, \mathbf{s}'; \tau_{\mu}^2, \phi_{\mu_t} \right) \right) \quad t = t_0, \dots, T$$

where $x_t(B_k)$ denotes the annual fourth highest ozone concentration estimated by CMAQ at grid cell B_k , and $C(\mathbf{s}, \mathbf{s}'; \tau_{\mu}^2, \phi_{\mu_t})$ indicates the exponential covariance function with marginal variance τ_{μ}^2 , decay parameter ϕ_{μ_t} —we call this model simply an FHEV downscaler model; and (ii) an FHEV geostatistical interpolation model obtained from the FHEV smoothed downscaler by setting $\beta_{1t} = 0$ in (1) for each $t = t_0, \dots, T$, that is, ignoring the CMAQ outputs.

Due to the challenges in learning about ξ as a model parameter, we considered selection of ξ as a model choice problem. We fit all three models using different values of the shape parameter ξ . Based on exploratory data analysis and estimates of the shape parameter ξ obtained by fitting a GEV distribution at each location independently, the set of values $\{\xi = -1.0, -0.5, -0.1, 0, 0.1, \text{ and } 0.5\}$ fully spans the range. For each model, we investigate whether the scale parameter, σ_t , should be constant or varying in time.

To address questions 4 and 5, we consider the smoothed downscaler model developed by Berrocal *et al.* (2012). This latter model assumes that the monitoring ozone concentration at sites $\mathbf{s} \in \mathcal{S}$ is a Gaussian process with mean $\mu_t(\mathbf{s})$ as in (1), and with an exponential covariance function along with a nugget effect. We fit this model to both daily ozone concentration data and also to the annual fourth highest ozone concentration data. In the first case, predictions of annual fourth highest ozone concentration are derived from predictions of daily ozone concentration. In the second case, we would not expect a Gaussian distribution to be appropriate for the annual fourth highest. A description of all the models considered is provided in Table 1.

4.2. Assessment

We evaluate the predictive performances of all models using an ensemble of out-of-sample criteria: their respective predictive mean squared error (PMSE), predictive mean absolute error (PMAE), empirical coverage of the 95% predictive intervals, and average width of the 95% predictive intervals. For the FHEV downscaler models and the FHEV geostatistical interpolation model, predictions at validation sites are taken to be the median of the posterior predictive distributions (for all models, predictions are backtransformed to the original scale). We also compare the various models with respect to their continuous ranked probability score (CRPS; Gneiting and Raftery (2007)), averaged across time and validation sites. The CRPS is a strictly proper scoring rule that evaluates the fit of a predictive distribution to observations, rewarding calibration subject to sharpness. It is defined as

$$CRPS(F, x) = \int_{-\infty}^{\infty} (F(y) - \mathbf{1}\{y \geq x\})^2 dy$$

where F is the predictive distribution and x is the observation that materializes. It can be straightforwardly computed through Monte Carlo simulations using the identity provided by Székely and Rizzo (2005)

$$CRPS(F, x) = E_F |X - x| - \frac{1}{2} E_F |X - X'|$$

where X and X' are independent copies of random variables with distribution F and finite first moments. As with the PMAE, the CRPS is reported in the units of the observations and smaller values are better.

We first focus on the choice of ξ . Figure 4 shows plots of the PMSE, PMAE, empirical coverage of the 95% predictive interval, average width of the 95% predictive interval, and average CRPS for the various FHEV models as the shape parameter ξ varies in the set $\{-1.0, -0.5, -0.1, 0, 0.1, 0.5\}$.

As Figure 4 reveals, all models prefer $\xi < 0$, because the PMSE, PMAE, and average CRPS achieve the lowest values for all models when $\xi = -0.5$. Additionally, Figure 4 shows that FHEV models with a time-varying scale parameter yield a better predictive performance than FHEV models which assume a constant scale parameter.

Among the FHEV models, the FHEV smoothed downscaler model with time-varying scale parameter and with $\xi = -0.5$ has the smallest PMSE, PMAE, and average CRPS and yields the shortest 95% predictive intervals with coverage close to nominal. From this first comparison, we conclude that there is useful information contained in the CMAQ output, particularly in the output relative to a neighborhood of grid cells, and exploiting this information leads to improved predictions of the annual fourth highest ozone concentration.

Table 2 compares the PMSE, PMAE, empirical coverage of the 95% predictive interval, average width of the 95% predictive interval, and average CRPS for the FHEV smoothed downscaler with $\xi = -0.5$ and time-varying scale parameter with the analogous summary statistics for the smoothed downscaler models based on a Gaussian distributional assumption. Table 2 reveals that the Gaussian distribution is not appropriate to model the annual fourth highest ozone concentration. Furthermore, applying the smoothed downscaler model to daily

Table 1. Description of the various downscaler and geostatistical interpolation models compared

Model label	Model	First stage likelihood specification	Time-varying scale parameter	CMAQ output	Smoothed CMAQ output
\tilde{M}_1	FHEV smoothed downscaler	FHEV	No	Yes	Yes
\tilde{M}_2	FHEV smoothed downscaler	FHEV	Yes	Yes	Yes
M_1	FHEV downscaler	FHEV	No	Yes	No
M_2	FHEV downscaler	FHEV	Yes	Yes	No
K_1	FHEV geostatistical interpolation model	FHEV	No	No	No
K_2	FHEV geostatistical interpolation model	FHEV	Yes	No	No
D_1	Daily downscaler	Gaussian	—	Yes	Yes
D_2	Downscaler on annual fourth highest	Gaussian	—	Yes	Yes

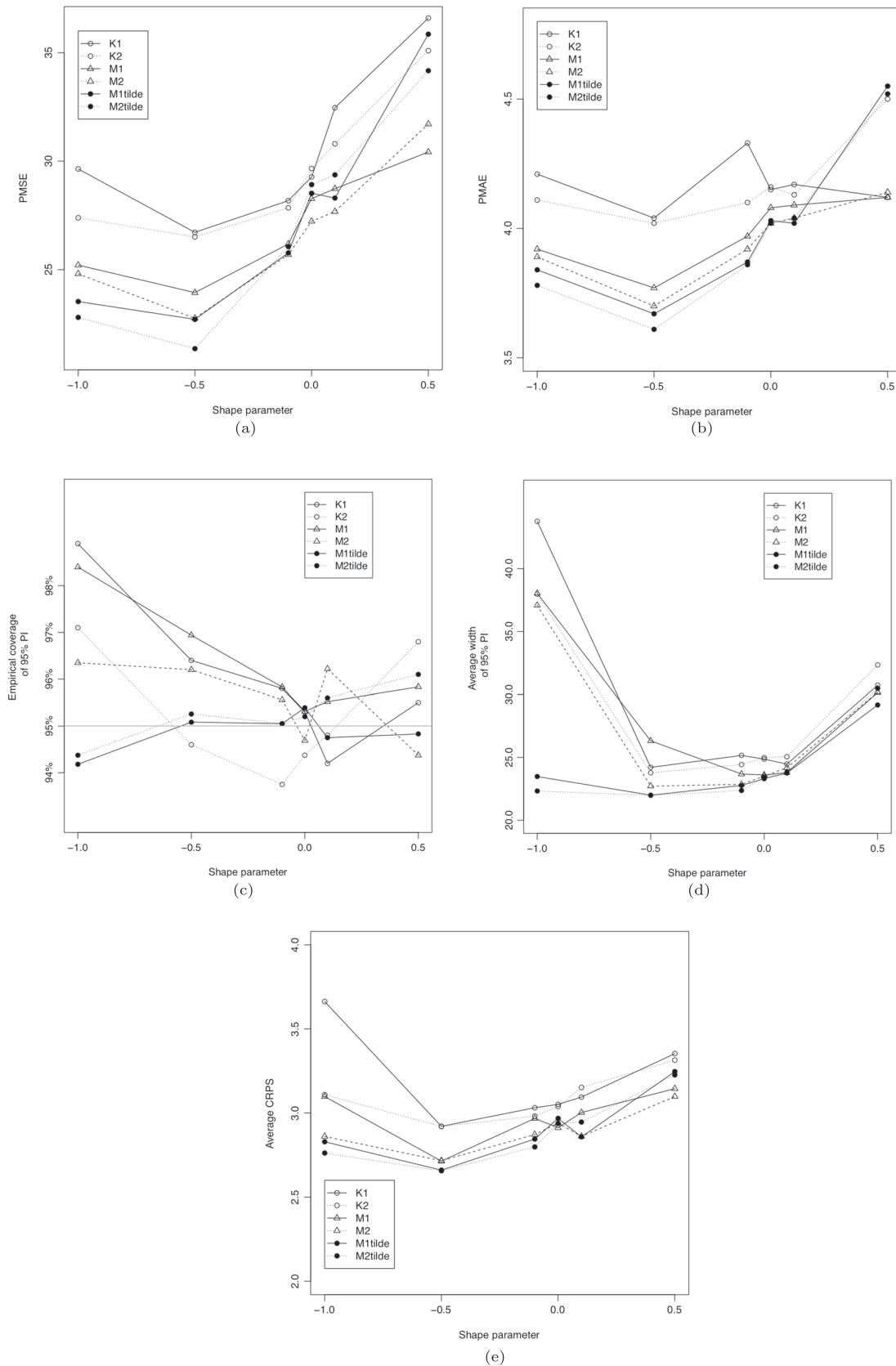


Figure 4. (a) Predictive mean squared error (PMSE), (b) predictive mean absolute error (PMAE), (c) empirical coverage of the 95% predictive interval, (d) average width of the 95% predictive interval, and (e) average continuous ranked probability score (CRPS) for the various FHEV downscaler and FHEV geostatistical interpolation models as the shape parameter ξ varies in the set $\{-1.0, -0.5, -0.1, 0, 0.5\}$

Table 2. Predictive mean squared error (PMSE), predictive mean absolute error (PMAE), empirical coverage of the 95% predictive interval, average width of the 95% predictive interval (PI) at validation sites, and average continuous ranked probability score (CRPS) for the FHEV smoothed downscaler model with time-varying scale parameter and shape parameter $\xi = -0.5$, the daily smoothed downscaler model and the smoothed downscaler model applied to the annual fourth highest ozone concentrations

Model	PMSE	PMAE	Empirical coverage 95% PI (%)	Average width 95% PI	Average CRPS
FHEV smoothed downscaler with time-varying σ_t ($\xi = -0.5$)	21.36	3.63	95.26	21.98	2.66
Daily smoothed downscaler	30.15	4.43	56.25	9.62	16.63
Smoothed downscaler for annual fourth highest ozone	27.65	4.16	97.40	24.61	2.96

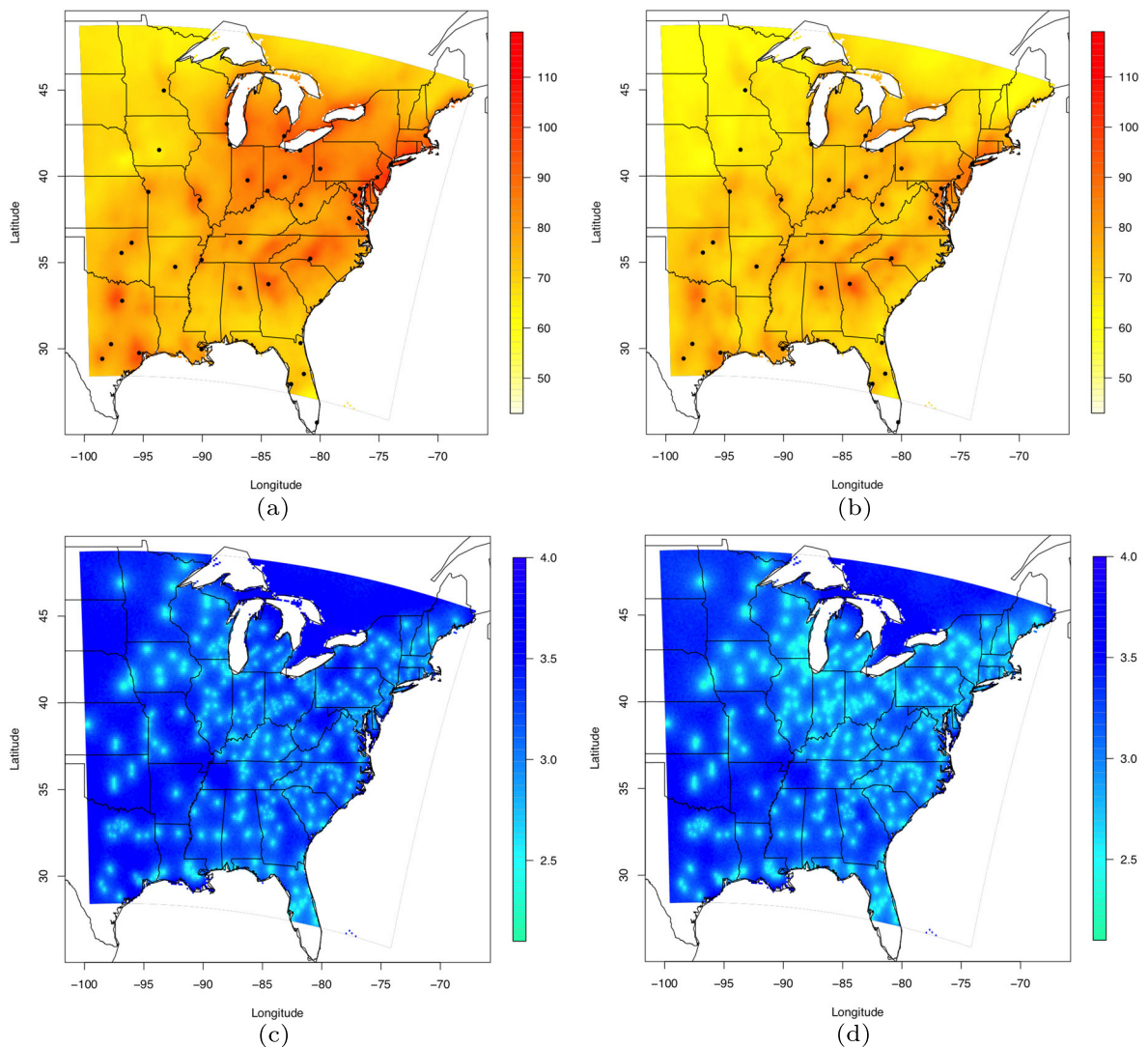


Figure 5. (a, b) Predicted rolling average of the annual fourth highest ozone concentration in parts per billion (ppb) over years (a) 2001–2003 and (b) 2006–2008, as yielded by the FHEV smoothed downscaler with time-varying scale parameter and shape parameter $\xi = -0.5$. (c, d) Posterior predictive standard deviation of the rolling average of the annual fourth highest ozone concentration in parts per billion (ppb) over years (c) 2001–2003 and (d) 2006–2008, as yielded by the FHEV smoothed downscaler with time-varying scale parameter and shape parameter $\xi = -0.5$. In panels (a) and (b), black dots indicate major urban centers in the spatial domain \mathcal{S}

Table 3. Brier score for the predicted probability that a site exceeds the NAAQS standard as predicted at validation sites by the FHEV smoothed downscaler model with time-varying scale parameter and shape parameter $\xi = -0.5$, the daily smoothed downscaler model and the smoothed downscaler model applied to the annual fourth highest ozone concentrations

Model	Brier score
FHEV smoothed downscaler with time-varying σ_t ($\xi = -0.5$)	0.097
Daily smoothed downscaler	0.195
Smoothed downscaler for annual fourth highest ozone	0.114

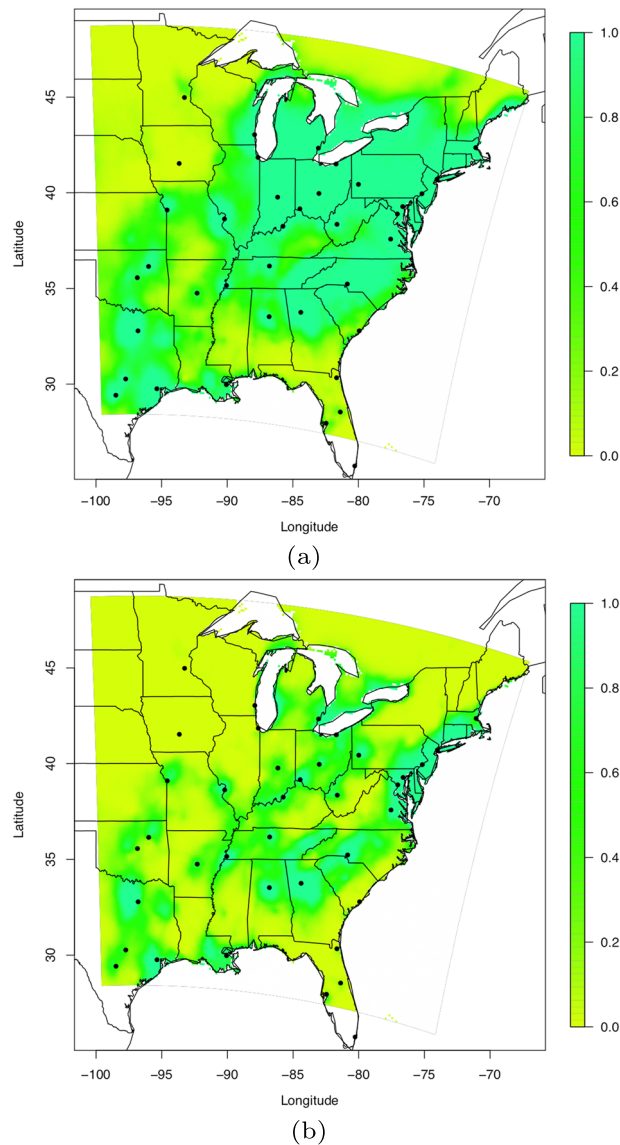


Figure 6. Predicted probability of exceedance relative to the NAAQS for ozone as yielded by the smoothed FHEV downscaler with time-varying scale parameter, shape parameter $\xi = -0.5$, over years (a) 2001–2003 and (b) 2006–2008. In both panels, the black dots indicate major urban centers located in the spatial domain

ozone concentration data to then derive predictions of the annual fourth highest ozone concentration leads to poor centering of the predictive distributions and severe underestimation of prediction uncertainty. Such interpolation proceeds by kriging the daily models to new locations and then taking the annual fourth highest, that is, *averaging* and then taking the fourth highest rather than taking the fourth highest and *averaging*. Evidently, our concerns in Section 1 regarding the former interpolation are borne out.

The FHEV smoothed downscaler model enables direct prediction of the variable related to the NAAQS for ozone. The NAAQS for ozone states that the rolling average of the annual fourth highest ozone concentration over a period of three consecutive years should not exceed 75 ppb. Using our FHEV smoothed downscaler model, we can generate predictions of the rolling average of the annual fourth highest ozone concentration over a 3-year period. We do this simply by averaging the three consecutive annual fourth highest interpolations. We obtain predictions of rolling averages for the periods 2001–2003, 2002–2004, . . . , 2006–2008 at the 48 hold-out sites as well as over the entire spatial domain S . Figure 5 presents the predicted rolling average of the annual fourth highest ozone concentration over years 2001–2003 and years 2006–2008, along with their relative standard errors. As we can clearly see, the rolling averages present a decreasing trend over time with most of the Western region of the domain below the threshold of 75 ppb. The standard error for the predictions is generally between 2 and 3.5 ppb, with significantly lower values at the monitoring sites, as expected.

Using the posterior predictive distribution of the rolling averages of the annual fourth highest ozone concentrations over years 2001–2003, . . . , 2006–2008, we can derive the predicted probability that a site exceeds the NAAQS for ozone. We can evaluate the accuracy and sharpness of the predicted probabilities at validation sites using the Brier score (Brier, 1950; Gneiting and Raftery, 2007), which is a strictly proper scoring rule for binary variables defined as

$$BS = \frac{1}{n} \frac{1}{m} \sum_{l=1}^m \sum_{i=1}^n (p_{il} - o_{il})^2 \tag{4}$$

In (4), p_{il} denotes the predicted exceedance probability for site s_i relative to the NAAQS during the 3-year period l ($l = 2001–2003; \dots, 2006–2008$) while o_{il} indicates the observed event (1 if site s_i exceeds the NAAQS during the same time period, 0 if it does not). A lower Brier score is better. Table 3 presents Brier scores for the FHEV smoothed downscaler model with time-varying scale parameter and $\xi = -0.5$, the daily smoothed downscaler and the smoothed downscaler applied directly to the annual fourth highest ozone concentrations. As observed earlier, the FHEV smoothed downscaler model yields better predictions at hold-out sites also for the probability that a site is not in attainment of the NAAQS.

In an analogous way, we can generate maps of the predicted exceedance probabilities with regard to the NAAQS for sites in the Eastern USA. Figure 6 shows probabilistic maps of non-attainment of the NAAQS for ozone for years 2001–2003 and 2006–2008. As already revealed in Figure 5, most of the major urban centers in our spatial domain are very likely to exceed the NAAQS. However, comparing Figure 6(a) and Figure 6(b), we can discern a noticeable improvement in the air quality over the Eastern USA.

5. DISCUSSION

Currently, assessment of exceedance of the NAAQS for ozone is only carried out at sites with available monitoring data. Thus, there is a need to develop statistical models to provide optimal information for delineating areas of the US that exceed the ozone NAAQS. In this paper, we try to establish a bridge between the regulatory and statistical communities by presenting a data fusion model that allows combination of monitoring station data with estimates provided by a deterministic air quality model in order to obtain better predictions of annual fourth highest levels of ozone concentration. The probabilistic maps of exceedance probabilities that our model allows us to derive could be useful to air quality managers in constructing better emission control programs. We note that although the motivating application is related to air quality and uses the fourth largest order statistic, the methodology proposed here can be applied to any situation where the interest is in combining observational data on extreme order statistics with gridded estimates for these order statistics derived using output from numerical models.

Our FHEV smoothed downscaler model allows us to obtain predictions of the probability that sites exceed the NAAQS for ozone by directly modeling a variable related to the NAAQS. We have shown that other approaches to derive predictions of the annual fourth highest ozone concentration based on Gaussian Process modeling assumptions, lead, in one case, to severe underestimation of prediction uncertainty and, in the other, to predictions with larger biases and wider prediction intervals.

An assumption in our model is the hypothesis of conditional independence among sites, with the spatial dependence between the annual fourth highest ozone concentrations being captured at the second modeling stage through the spatially varying location parameters of the marginal FHEV distributions. A possible extension of this work would be to relax this assumption and allow for dependence among the extreme levels of ozone concentration at various sites by using a copula approach as Sang and Gelfand (2010). Alternatively, one could attempt to adapt max-stable models and approximate the multidimensional likelihood function using a pseudolikelihood approach as in Ribatet *et al.* (2012).

We note that in our model, the location parameter process of the FHEV distributions has a mean function that depends only on the smoothed CMAQ output. However, besides the computer model output, other covariates could be useful in predicting extreme ozone concentration at point level, covariates such as meteorological data (though we would anticipate that the CMAQ model captures these effects), remote sensing data, or land use information.

Finally, we have shown the ability of our modeling approach to identify regions at high risk of exceeding the NAAQS standard with associated uncertainty. However, an important follow up would be to connect these risks to adverse health outcomes, for example, elevation of suitably measured hospital admissions.

Acknowledgements

The US Environmental Protection Agency through its Office of Research and Development partially collaborated in this research. Although it has been reviewed by the Agency and approved for publication, it does not necessarily reflect the Agency's policies or views.

REFERENCES

- Banerjee S, Carlin BP, Gelfand AE. 2004. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC: Boca Raton, Fla.
- Banerjee S, Gelfand AE, Finley AO, Sang H. 2008. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society Series B* **70**:825–848.
- Bell ML, McDermott A, Zeger SL, Samet JM, Dominici F. 2004. Ozone and short-term mortality in 95 US urban communities, 1987–2000. *Journal of the American Medical Association* **292**:2372–2378.
- Berrocal VJ, Gelfand AE, Holland DM. 2010a. A bivariate spatio-temporal downscaler under space and time misalignment. *Annals of Applied Statistics* **4**:1942–1975.
- Berrocal VJ, Gelfand AE, Holland DM. 2010b. A spatio-temporal downscaler for outputs from numerical models. *Journal of Agricultural, Biological and Environmental Statistics* **15**:176–197.
- Berrocal VJ, Gelfand AE, Holland DM. 2012. Space-time data fusion under error in computer model output: an application to modeling air quality. *Biometrics* **68**:837–848.
- Brier GW. 1950. Verification of forecasts expressed in terms of probability. *Monthly Weather Review* **18**:1–3.
- Byun D, Schere KL. 2006. Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System. *Applied Mechanics Reviews* **59**:51–77.
- Carroll RJ, Chen EI, George TH, Li HJ, Newton H, Schmiediche H, Wang N. 1997. Ozone exposure and population density in Harris County, Texas (with discussion). *Journal of the American Statistical Association* **92**:392–415.
- Coles SG. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer: New York.
- Cooley D, Nychka D, Naveau P. 2007. Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association* **102**:824–840.
- Cressie NAC. 1993. *Statistics for Spatial Data*. Wiley: New York.
- Cressie NAC, Wikle CK. 2011. *Statistics for Spatio-Temporal Data*. Wiley: New York.
- Eastoe EF. 2009. A hierarchical model for non-stationary multivariate extremes: a case study of surface-level ozone and NO_x data in the UK. *Environmetrics* **20**:428–444.
- Eastoe EF, Tawn JA. 2009. Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society Series C* **58**:25–45.
- Finley AO, Sang H, Banerjee S, Gelfand AE. 2009. Improving the performance of predictive process modeling for large datasets. *Computational Statistics and Data Analysis* **53**:2873–2884.
- French JP, Sain SR. 2013. Spatio-temporal exceedance locations and confidence regions. *Annals of Applied Statistics* **7**:1421–1449.
- Fuentes M, Raftery AE. 2005. Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models. *Biometrics* **61**:36–45.
- Gneiting T, Raftery AE. 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* **102**:359–378.
- Gotway CA, Young LJ. 2002. Combining incompatible spatial data. *Journal of the American Statistical Association* **97**:632–648.
- Guillas S, Bao J, Choi Y, Wang Y. 2008. Statistical correction and downscaling of chemical transport model ozone forecasts over Atlanta. *Atmospheric Environment* **42**:1338–1348.
- Liu Z, Le N, Zidek JV. 2012. Combining data and simulated data for space-time fields: application to ozone. *Environmental and Ecological Statistics* **19**:37–56.
- McMillan N, Holland DM, Morara M, Feng J. 2010. Combining numerical model output and particulate data using Bayesian space-time modeling. *Environmetrics* **21**:48–65.
- Reich B, Cooley D, Foley K, Napelenok S, Shaby B. 2013. Extreme value analysis for evaluating ozone control strategies. *Annals of Applied Statistics* **7**:739–762.
- Resnick SI, Rubinovitch M. 1973. The structure of extremal processes. *Advances in Applied Probability* **5**:287–307.
- Ribatet M, Cooley D, Davison AC. 2012. Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica* **22**:813–845.
- Sahu SK, Gelfand AE, Holland DM. 2006. Spatio-temporal modeling of fine particulate matter. *Journal of Agricultural, Biological and Environmental Statistics* **11**:61–86.
- Sahu SK, Gelfand AE, Holland DM. 2007. High resolution space-time ozone modeling for assessing trends. *Journal of the American Statistical Association* **102**:1221–1234.
- Sang H, Gelfand AE. 2009. Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics* **16**:407–426.
- Sang H, Gelfand AE. 2010. Continuous spatial process models for spatial extreme values. *Journal of Agricultural, Biological and Environmental Statistics* **15**:49–65.
- Székely GJ, Rizzo ML. 2005. A new test for multivariate normality. *Journal of Multivariate Analysis* **93**:58–80.
- Zanobetti A, Schwartz J. 2008. Mortality displacement in the association of ozone with mortality: An analysis of 48 cities in the United States. *American Journal of Respiratory and Critical Care Medicine* **177**:184–189.
- Zhu L, Carlin BP, Gelfand AE. 2003. Hierarchical regression with misaligned spatial data: relating ambient ozone and pediatric asthma ER visits in Atlanta. *Environmetrics* **14**:537–557.