Statistical Models for Monitoring and Regulating Ground-level Ozone

Eric Gilleland[1] and Douglas Nychka[2]

Abstract

The application of statistical techniques to environmental problems often involves a tradeoff between simple methods that are easily implemented and interpreted and more complicated methods that may have smaller errors. In this paper we compare simple and complicated statistical models for interpolating the U.S. Environmental Protection Agency (EPA) National Ambient Air Quality Standard (NAAQS) for ground-level ozone off of a network of monitoring sites. A recent change in the NAAQS for ground-level ozone is based on the fourth-highest value from the daily sequence of maximum 8-hour average ozone (FHDA). In particular, two models are given special attention: a daily model that uses an autoregressive model to obtain many Monte Carlo samples of FHDA and a seasonal model that assumes the FHDA is Gaussian and employs standard statistical techniques to model the FHDA. We find that the daily model is superior enough to the seasonal model to warrant the added complexity. We also suggest other ways to model FHDA.

1 Introduction

The application of statistical techniques to environmental problems often involves a tradeoff between simple methods that are easily implemented and interpreted and more complicated methods that may have smaller errors. In this paper we compare simple and complicated statistical models for interpolating the U.S. Environmental Protection Agency (EPA) National Ambient Air Quality Standard (NAAQS) for ground-level ozone off of a network of monitoring sites. A recent change in the NAAQS for ground-level ozone is based on the fourth-highest value from the daily sequence of maximum 8-hour average ozone (FHDA). The standard requires that the average over three ozone seasons be below 80 parts per billion (ppb) in order for a location to be in attainment. Understanding the spatial distribution for the FHDA for a given ozone season presents a new statistical problem for infering regions of attainment or nonattainment because it is not clear that the FHDA field (a field of order statistics) is Gaussian–a fundamental

[1] *Corresponding author address*: Eric Gilleland, National Center for Atmospheric Research, Research Applications Program, Boulder, CO 80307-3000; email: ericg@ucar.edu

[2] National Center for Atmospheric Research, Geophysical Research Project

assumption of most standard spatial statistical techniques. This problem is addressed in this work.

Although the use of spatial statistics for interpreting air quality measurements would not be disputed by a statistical audience, surprisingly the use of monitoring data in a regulatory context is often limited to point locations. Accordingly, Holland *et al.* [6] argue for the introduction of modern statistical methods to understand the spatial and temporal extent of pollution fields based on monitoring data. Given the range of statistical backgrounds associated with the regulatory community, it is appropriate to propose statistical methods that are simple and understandable to a broad group when such methods provide an accurate and defensible analysis. In particular, for interpreting the FHDA standard, it is useful to ascertain the feasibility of approximate statistical methods that treat the FHDA statistics directly. From this perspective we compare two statistical models. The first, a fairly complex model, uses a spatial AR(1) model for daily ozone measurements and samples the FHDA field conditional on the data for the entire season. This approach will be referred to as the *daily model*. The second model, referred to the *seasonal* model, is a geostatistical model that predicts the FHDA field from the network values using standard best linear unbiased estimation, or kriging (see Cressie [1] or Stein [9] for more details on kriging). This seasonal model is similar to the model proposed by Fuentes [2], except that the region of interest here is much smaller and so can be assumed to be spatially and temporally stationary. A third approach that will be used as a benchmark estimates the FHDA field by way of a thin plate spline (see Green and Silverman [4] or Hastie and Tibshirani [5] for details on thin plate splines). This last method is generic and uses the least amount of information concerning the actual air quality context.

The paper is organized as follows. Section 2 describes the ozone monitoring data. Section 3 describes how the ozone data is standardized and presents the daily and seasonal models. Section 4 presents a comparison of these models along with thin plate spline interpolation as a benchmark. One interesting aspect is that the daily model implies a reasonable covariance function for the FHDA field, and this choice is included in the results for the seasonal approach. Our results suggest that the daily model is worth the extra effort and complexity. Section 5 discusses the results and suggests some future research.

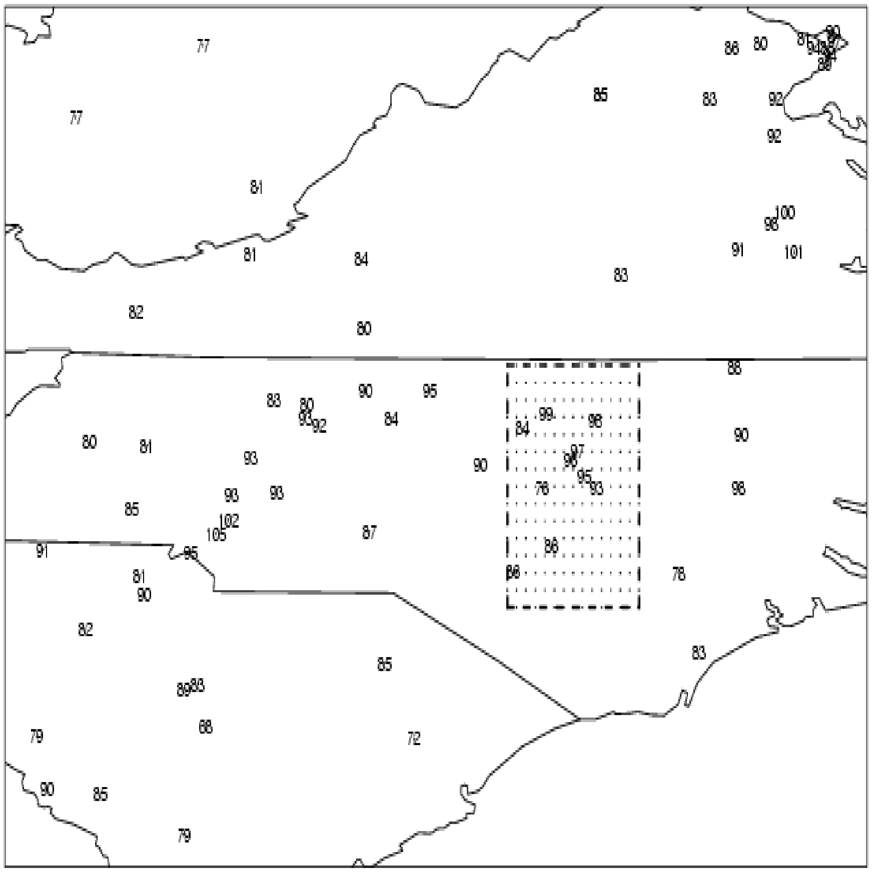

Figure 1: Network of ozone monitoring locations with RTP region (dashed rectangle (with 15×15 grid)). The numbers represent the fourth-highest value for 1997.

2 Ozone Monitoring Data

Data used for these analyses consist of maximum daily 8-hour average ozone levels measured in parts per billion (ppb) for the 72 monitoring stations in a study region centered on North Carolina (Figure 1). The dashed rectangle in Figure 1 shows a region around the Research Triangle Park (RTP), N.C., in which the models will be interpolated onto a 15×15 grid. These data are a subset of the 513 stations covering the eastern United States used by Fuentes [2] and can be obtained from the web through `www.cgd.ucar.edu/stats/Data`.

Partly because of the high cost of operation, ozone monitoring occurs only during the hotter months when weather conditions are most conducive to forming ozone; this "ozone season" essentially spans the months from April through October. For these analyses, the data cover five seasons (1995-1999), with each season consisting of 184 days. Summary statistics for the

Table 1: Observed fourth-highest daily 8-hour maximum (FHDA) ozone.

	1995	1996	1997	1998	1999
No. of stations	66	69	72	71	70
mean	85.68	83.12	87.64	95.87	92.70
Std.Dev.	7.527	6.736	7.495	5.974	7.864
min	67	70	68	80	67
Q1	81	80	82	92	88
median	86	82	87	96	93
Q3	91	87	93	100	98
max	102	101	105	110	108
Stations with missing values	6	3	0	1	2

FHDA statistics for the stations in the study region over this period are given in Table 1.

There are some missing values and; in particular, there are seasons where some stations have no values reported. The fact that 1995 has six missing values means that there were six stations that had less than four observations for that season. The 1997 season is the only season where all of the stations have a nearly compete record; of the 13,248 data points, only 189 total data points across all of the 72 stations are missing for this season. In particular, the stations with the most missing values for this season have 50, 45, 25 and 16 missing values respectively, and 57 of the stations have no missing values at all. Given that the data record is nearly complete, our analysis will not handle missing observations explicitly, and no effort will be made to inpute the few missing measurements.

3 Daily model for ozone and a seasonal model for FHDA

Ideally one would like to know the complete multivariate distribution of the FHDA field. This would enable us to not only interpolate to values off of the monitoring network, but to better understand the error associated with such an interpolation. Because the FHDA statistic represents an order statistic based on a serially correlated sample, it is difficult to derive a form for its distribution. As an alternative, we suggest a model for daily ozone measurements and then infer the distribution of the FHDA through aggregating the daily model over the season.

The daily model uses a spatial AR(1) model to model the daily maximum 8-hour averages and then uses Monte Carlo sampling to approximate the conditional distribution of the FHDA field given the observed network data. One important practical issue is whether this constructive and rigorous approach has any benefits over a simple spatial interpolation of the FHDA sample statistics. We divide the presentation of the model into a preliminary standardization, autoregressive time series for individual stations and spatial model for the innovations. The last part of this section describes the algorithm to simulate a process from the daily model.

3.1 Standardizing the Data

Ozone can have a seasonal effect even during the relatively short ozone season described in Section 2. It is useful to account for this seasonality as a fixed effect before modeling space-time structure.

Let $O(\mathbf{x}, t)$ denote the maximum 8-hour average ozone at location $\mathbf{x}$ and day t. The following standardization is used for the daily maximum 8-hour ozone measurement

$$O(\mathbf{x}, t) = \mu(\mathbf{x}, t) + \sigma(\mathbf{x})u(\mathbf{x}, t) \tag{1}$$

and we assume that $u(\mathbf{x}, t)$ for any given location and time has mean zero and variance 1. Note that μ is a function of both time and space in order to remove any seasonality.

The seasonal means are smoothed over space using a single value decomposition approach. The m individual station time series are regressed on an intercept and three sine and cosine pairs with periods 365, 365/2 and 365/3 and let $\mathbf{B}$ denote the $m \times p$ matrix of regression coefficients across all the stations. Next, $\mathbf{B}$ can be decomposed as $\mathbf{B} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices and $\mathbf{D}$ is a diagonal matrix of the singular values of $\mathbf{B}$. By setting some of the singular values of $\mathbf{D}$ to zero (call the resulting matrix $\mathbf{D}^*$), the multiplication $\mathbf{B}^* = \mathbf{U}\mathbf{D}^*\mathbf{V}^T$ yields a matrix of the original regression parameters, but having reduced the variability across stations. For the analyses here, the first three principle components were retained (i.e., the last four singular values were set to zero); and, in this case, results smooth the estimated parameters over space. Finally, the estimates of μ and σ based on station locations are extrapolated to unobserved locations using thin plate spline interpolation.

3.2 Daily Model

Given the standardized process, $u(\mathbf{x}, \mathbf{t})$, we consider a spatial AR(1) model.

$$u(\mathbf{x}, t) = \rho(\mathbf{x})u(\mathbf{x}, t-1) + \varepsilon(\mathbf{x}, t) \tag{2}$$

The shocks, $\varepsilon(\mathbf{x}, t)$, are assumed to be independent over time and be a mean zero Gaussian process over space with spatial covariance

$$C(\varepsilon(\mathbf{x}, t), \varepsilon(\mathbf{x}', t)) = k(\mathbf{x}, \mathbf{x}') \tag{3}$$

Here, the covariance (3) is considered to be isotropic and stationary so that $k(\mathbf{x}, \mathbf{x}') = \psi(|\mathbf{x}-\mathbf{x}'|)$. For this application the great circle distance is used as a metric to measure separation between locations. Equation (2) implies a space-time covariance function

$$C(u(\mathbf{x}, t), u(\mathbf{x}', t-\tau)) = \frac{(\rho(\mathbf{x})\rho(\mathbf{x}'))^{\tau}\psi(|\mathbf{x}-\mathbf{x}'|)}{1-\rho(\mathbf{x})\rho(\mathbf{x}')}, \tau \geq 0 \tag{4}$$

Thus, if the AR(1) parameters are not constant over space, then (i) the spatial process $u(\mathbf{x}, t)$ is not stationary even if the shocks are stationary in space and (ii) covariance (4) is not space-time separable. Ma [7] further shows that there are circumstances where Equation (4) is not positive (or even nonnegative) definite. This is not a concern for the daily model presented here because this covariance is not used and we sample directly from the daily model (2).

3.3 Sampling the distribution of FHDA conditioned by the monitoring data

Under the assumption that all the components of the data model are known, there is a straightforward algorithm for sampling the FHDA field conditional on the observed data. This algorithm is quite efficient and uses the autoregressive structure over time to recursively generate the daily process. Let $\mathbf{x}_0$ be a location where ozone is unobserved. A spatial prediction for the FHDA at this location involves two steps. One first obtains a sample of the time series of daily ozone measurements at this location conditional on the observed data (for all locations and all times). Next one calculates the FHDA for this series. By elementary probability, the resulting FHDA statistics will be a sample of the FHDA field at $\mathbf{x}_0$ conditional on the data. Repeating these two steps, one can generate a random sample that approximates the FHDA conditional distribution; and, of course, the sample mean is a point estimate for the conditional expectation of the FHDA at $\mathbf{x}_0$. The conditional variance can be used as a measure of uncertainty.

Sampling from the conditional distribution of the ozone is simplified by the autoregressive structure over time and the restriction of spatial dependence to the shocks in the AR(1) innovation. In this section we will assume that all parameters $(\mu(\mathbf{x}, t), \sigma(\mathbf{x}), \rho(\mathbf{x}), \psi)$ are fixed quantities and known, but more will be said about this assumption in Section 5. Also let $\{\mathbf{x}_k, \text{ for } 1 \leq k \leq m\}$ be the station locations. Based on these assumptions it is sufficient to find the conditional distribution of $\{u(\mathbf{x}_0, t), 1 \leq t \leq T\}$ given $\{u(\mathbf{x}_k, t),\ 1 \leq t \leq T, \text{ and } 1 \leq k \leq m\}$ because the standardized random variables can always be transformed back to the raw scale of the measurements. Moreover, knowledge of $\{u(\mathbf{x}, t),\ 1 \leq t \leq T\}$, for any $\mathbf{x}$ is equivalent, through the autoregressive relationship, to $\{u(\mathbf{x}, 1), \varepsilon(\mathbf{x}, t),\ 2 \leq t \leq T\}$. Recall that the AR shocks are temporally independent so that the conditional distribution for the ozone fields at $\mathbf{x}_0$ can be found based on the much simpler conditional distribution of $\varepsilon(\mathbf{x_0}, t)$ given $\{\varepsilon(\mathbf{x_k}, t),\ 1 \leq k \leq m\}$. Thus we can easily generate a conditional ozone field by considering the conditional field of the AR(1) shocks and then transform these results to the original scale of measurements.

The algorithm for conditional sampling of the FHDA field is now summarized below.

1. Initialize the time series by sampling $[u(\mathbf{x}_0, 1)|u(\mathbf{x}_k, 1), 1 \leq k \leq m]$
2. For t in 2 to T sample the spatial shocks from $[\varepsilon(\mathbf{x_0}, t)|\{\varepsilon(\mathbf{x_k}, t),\ 1 \leq k \leq m\}]$.
3. Accumulate the sampled shocks and initial values using the autoregressive relationship (2) to obtain a conditional realization of the standardized process $u(\mathbf{x}_0, t)$.
4. Unstandardized and compute the FHDA at $\mathbf{x}_0$ based on this series to obtain the ozone series in the raw scale.

Note that the shocks at a station location are based on the actual daily observations and so the sample is tied explicitly to the data. If in fact $\mathbf{x_0}$ is at a station location and the spatial process has a zero nugget variance, then the resulting conditional sample will just be the observed data. Thus the "conditional realization of the FHDA field" will be the FHDA statistic for that station's measurement. It should be noted that this algorithm works because we assume complete observations at the station locations. It would be more complicated if observations were sparse over time. For these data there are no instances where there are missing observations at a given time point at every station. Therefore, when shocks are sampled from the conditional distribution, locations that have missing values are simply not used in the calculation for that time point. Although it is possible to sample in Step 1 exactly, we have found that sampling from a geostatistical model fit to the standardized fields is adequate.

In this algorithm it is straightforward to replace the conditional sampling of a single location with a vector, or grid of locations. Thus one obtains a conditional field with spatial and temporal dependence among the grid points consistent with the space-time model. In addition, this algorithm can be modified simply to simulate a space-time process that follows this model. In this case one does not condition on observed data, and one substitutes an unconditional sample for the conditional sample of the shocks at Step 2. This unconditional sampling is used in the next section to identify an approximate Gaussian model for the FHDA field.

3.4 Seasonal Model

For the seasonal model we posit that the FHDA field is approximately Gaussian distributed, so the main modeling issue is to derive a suitable covariance function. The Gaussian assumption can be justified by simulations of bivariate data. Based on the results of other research (see Gilleland *et al* [3]), the joint distribution of the fourth-highest order statistics for bivariate samples of size 184 were found to be again approximately bivariate normal. In the next section we present results based on a standard geostatistical technique for estimating a stationary covariance from the FHDA values. However, with access to a daily model, one can also compute a covariance model for the FHDA field by Monte Carlo simulation. One generates many realizations of the space-time ozone process and accumulates independent realizations of the FHDA field. One then can use the sample covariances among the FHDA realizations to identify and fit a covariance function. This is a purely computational exercise; and by increasing the Monte Carlo sample size, one can obtain arbitrarily accurate estimates of the second moments for the FHDA spatial process implied by the daily model.

4 Results

4.1 Daily Model Results

Individual autoregressive models were fit by maximum likelihood to the standardized station data to estimate the parameters of the AR(1) parameter; an AR(2) model was also tried, but did not show significant improvement to warrant the added complexity. The AR(1) parameters vary across stations and so suggest that the daily ozone observations are nonstationary even in this small homogeneous region. To assess stationarity of the AR shocks, a local correlogram is fit for each station location using an exponential covariance function. The nugget variance and the range parameters do not vary significantly across the domain, and each has a small range and standard deviation; from 0.83 to 1.03 ppb (0.04 ppb) for the nugget and 164 to 328 miles (33 miles)–suggesting that the spatial shocks field can be approximated by a stationary process. The general shape of the empirical correlations suggested fitting a mixture of exponentials function to these correlations.

$$\psi(h) = \alpha \exp(-h/\theta_1) + (1-\alpha) \exp(-h/\theta_2) \tag{5}$$

where θ_1 represents short range correlation and θ_2 the long range correlation and h is the great circle distance between two locations. Correlation model (5) allows the spatial field to be interpreted as the sum of two independent spatial processes with possibly different correlation scales without changing the smoothness of ψ zero, but the shape will be modified for short distances in a similar fashion to the Matérn family. The reader should note that unlike a geostatistical analysis for a single field, the correlations associated with the shocks are statistics based on a large ($n > 500$) sample size, which enables enough accuracy to facilitate modeling detailed features such as the mixture component. Figure 2 plots the fitted correlogram for the spatial shocks as a function of distance of separation.

The fitted function is shown as the solid line in Figure 2. The shocks are tightly correlated within a distance of about 100 miles and are still somewhat correlated as far as about 200 miles. The fitted parameter estimates are $\hat{\alpha} \approx 0.13$ (0.017), $\hat{\theta}_1 \approx 11$ miles (3.373 miles) and $\hat{\theta}_2 \approx 272$ miles (16.887 miles) with bootstrap standard errors in parentheses.

The fitted model was used to generate conditional fields for the FHDA for each year and the rectangular RTP subregion in Figure 1. One thousand Monte Carlo realizations were used to approximate the distribution. Standard errors of prediction are summarized in Table 4, and on average these prediction errors are about 3 ppb. Results are similar for other years.

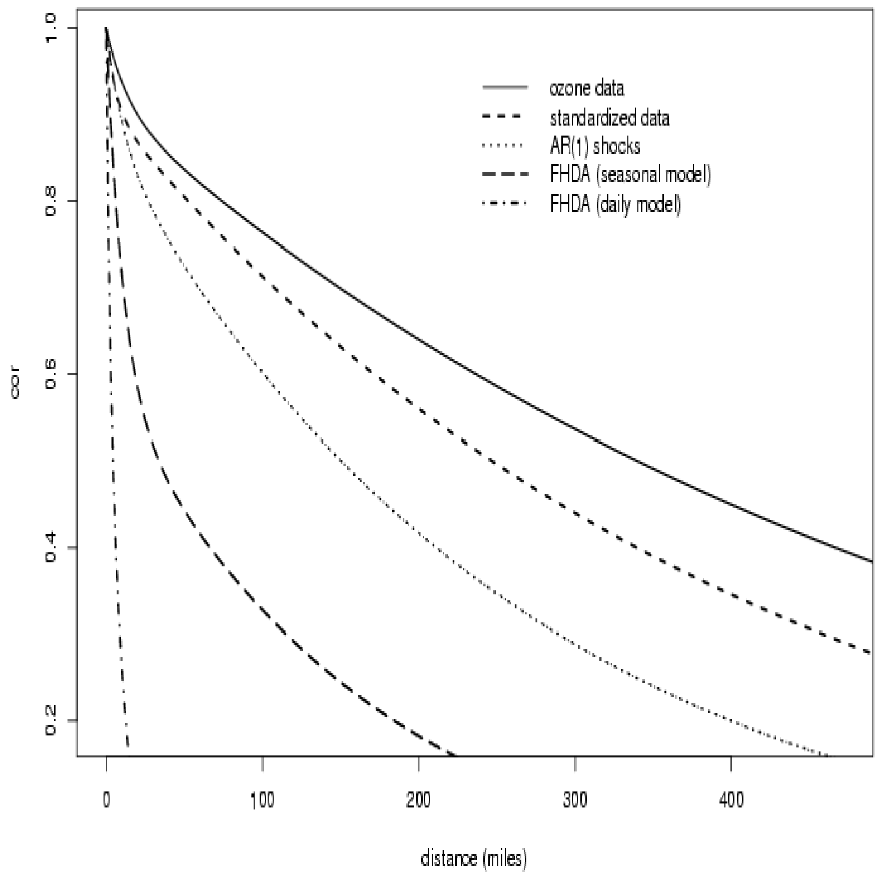

Figure 2: Fitted empirical correlation functions for original daily maximum 8-hour average ozone measurements, the standardized daily values, the spatial AR(1) shocks and unconditional (seasonal model) and conditional (daily model) simulations of the FHDA field.

4.2 Seasonal Model Results

The seasonal approach applies a spatial model directly to the FHDA values, so a key step is to estimate a covariance function for this field. Empirical variograms for each of the 5 seasons indicate that almost all of the spatial dependence in the FHDA field appears to be limited to a very short range less than 100 miles. A mixture of exponentials variogram

$$\gamma(h) = \sigma^2(1 - \alpha \exp(-h/\theta_1) - (1-\alpha)\exp(-h/\theta_2))$$

was fit using all five years of data with parameter estimates: $\hat{\sigma} \approx 7.37$ ppb, $\hat{\alpha} \approx 0.38$, $\hat{\theta}_1 \approx 0.62$ miles and $\hat{\theta}_2 \approx 48.61$ miles and subsequently converted to a covariance function, ψ_v. Standard errors of prediction for using this model for the RTP grid are summarized in Table 4. On average, these prediction errors are about 6 ppb, slightly greater than that of the daily model approach.

For comparison to estimating the covariance from the FHDA variogram, a covariance function was estimated from unconditional simulations of the daily model. Based on a Monte Carlo sample of 600 FHDA simulated fields, a mixture of exponentials (5) was fit to the empirical correlations, call it ψ_m. The estimated parameters are $\hat{\alpha} \approx 0.51$, $\hat{\theta}_1 \approx 8.66$ and $\hat{\theta}_2 \approx 128.76$. The spatial prediction errors using this covariance are summarized in Table 4 and are, on average, 5.6 ppb, which is comparable to the seasonal model prediction errors using ψ_v. Regardless of covariance used, the seasonal model is fit using the fields package [8] in R.

The thin plate spline model was fit (also using the fields package [8] in R) with $m = 2$ and the smoothing parameter chosen by generalized cross-validation. A thin plate spline of this order includes a linear spatial drift and is the limiting case of a Matern covariance function as the range becomes large. Moreover, the smoothing parameter for the spline is directly equivalent to estimating the nugget variance. Standard errors of prediction (Table 4) are, on average, about 3 ppb, considerably greater than all of the other models.

4.3 Model Comparison

Table 2 compares the mean predicted FHDA between the models. The root mean squared error for the fields is with respect to the daily model predictions and indicates that the differences in the spatial predictions among the three methods are consistently small across all years. The

Table 2: RMSE of mean predicted FHDA (ppb) between the daily model approach and the seasonal and thin plate spline approaches.

	Thin Plate Spline	Seasonal Model (ψ_m)
1995	3.54	2.66
1996	3.01	2.24
1997	2.82	2.39
1998	2.55	2.20
1999	6.07	1.78

relatively large value between the daily model and the thin plate spline may be caused by higher variability of FHDA for this year.

Daily model standard errors (Table 4) are generally smaller than the seasonal models; particularly away from station locations. The spline method tends to have similar prediction standard errors as the daily model, but there is less prediction precision away from the monitoring network than the seasonal model.

Model-based standard errors can either be realiable or misleading depending on the adequacy of the spatial model. It is also of interest to use cross-validation (CV) to evaluate the average prediction error of these methods. The standard leave-one-out procedure was applied to each monitoring location and method, and Table 3 reports for each year the CV RMSE for the differences between the predicted FHDA and the actual station values. The seasonal model CV RMSE for either choice of covariance and thin plate spline are very similar for each year. The daily model CV RMSE is consistently lower than the other models, but only slightly.

5 Discussion

Although care is needed in generalizing results from a specific data set to other cases, this work has shown a preference to analyze the FHDA standard using a daily model for ozone and then aggregating over the season to infer the FHDA field. The results for the North Carolina study region show that the seasonal model is reasonable, but the daily model is generally more accurate, based on the CV measures of RMSE in addition to having lower model standard

Table 3: Leave-one-out cross-validation RMSE (ppb) and root median squared error (ppb) (in parentheses) for predicting FHDA.

	Thin Plate Spline	Seasonal Model (ψ_v)	Seasonal Model (ψ_m)	Daily Model
1995	5.34 (2.74)	5.19 (2.65)	5.33 (2.92)	4.74 (2.68)
1996	5.61 (3.46)	5.51 (3.32)	5.68 (4.31)	4.82 (3.20)
1997	6.27 (4.37)	6.03 (3.93)	6.05 (3.75)	4.61 (2.65)
1998	5.00 (3.79)	4.98 (3.87)	4.93 (3.59)	3.22 (2.57)
1999	6.25 (4.76)	6.47 (3.97)	6.30 (3.62)	4.93 (2.16)

Table 4: Mean standard errors of prediction (ppb).

	Thin Plate Spline	Seasonal Model (ψ_v)	Seasonal Model (ψ_m)	Daily Model
1995	2.23	5.68	5.27	2.67
1996	2.49	5.96	5.90	2.87
1997	2.91	6.41	6.02	2.98
1998	2.75	5.35	4.85	2.93
1999	4.34	6.76	6.22	2.97

errors of prediction.

Conceptually, the daily model has advantages in using fairly simple statistical components on a daily scale that can produce relatively complicated seasonal statistics. For example, as long as the AR(1) shocks are stationary over space, the entire daily model can be fit using standard geostatistical and regression methods even if the original field (in this case standardized maximum 8-hour ozone levels) is nonstationary. We believe that part of the success of the daily model is that much of the spatial correlation and the nonstationarity of the raw measurements can be accounted for by standardizing the process and building in a temporal evolution. While the seasonal model is much simpler and easier to employ in general, it can actually be more complicated if the FHDA field is not stationary.

The lack of long-range correlation structure in the FHDA field simulated by the daily model approach (conditional on the data) and reaffirmed by empirical variograms of the observed FHDA field suggest that standard spatial techniques may not be very effective at predicting the FHDA at locations relatively far from any monitoring station. Figure 2 contrasts the different correlation scales among different transformations of the ozone field, and we note the marked difference between daily fields and the seasonal FHDA. This is further justified by the greater standard errors of prediction found by both the seasonal model and the thin plate spline at locations away from the monitoring network. Additionally, the apparent correlation structure in the FHDA field, found from using the daily model approach without conditioning on the data, may be an artifact of the model. One source of model bias may be the lack of a more heavy tail distribution associated with the AR shocks.

The data used here are from a relatively small, homogeneous subset of a much larger network of monitoring stations, and the more practical application of our work is to extend the analysis to the Eastern United States. Stationarity for this field cannot be assumed (see, for example, Fuentes [2]) and, indeed, some heterogeneity might be modeled using additional covariates. Generally, meteorological data, such as temperature, might be difficult to use for the daily model approach because at any time point the temperatures at two locations will vary and may or may not be similar. Also, determining meteorological covariates where there are not weather stations is problematic. Two possible covariates, however, that would be easy to incorporate into the daily model are elevation and aspect, and preliminary results have suggested these are useful.

In this work we have considered some parameter uncertainty in parts of the models, but have not propagated the uncertainty into the FHDA fields. A fully Bayesian model could perhaps synthesize covariates, model parameters, and any uncertainty associated with them in an efficient manner. Note that by varying the model parameters in the algorithm, one can include uncertainty into the daily model analysis resulting from uncertainty in the parameters. Although a fully Bayesian approach may be the most elegant solution, bootstrapping is a good compromise in terms of less demands for new software and computing resources. For example, one could use a parameteric bootstrap to generate a sample of parameters that reflect the uncertainty (in a frequentist sense!) with respect to the MLE. These values would then be used to generate the conditional FHDA fields.

We understand that the use of an extreme order statistic (4^{th} largest) suggests a standard sensitive to the tail of the ozone distribution. For this reason, a productive extension of our models is to incorporate methods for extreme value theory to explicitly model the frequency of large, but rare, ozone events. Part of the benefit of this approach is a statistical description for the entire tail of the distribution, rather than just a particular order statistic or quantile. The biggest challenge would be to incorporate spatial correlations into such a model. This might be accomplished by fitting Generalized Pareto (GP) distributions at each location with values from other stations used as covariates in the scale parameter. Another idea would be to fit GP distributions without any spatial covariates and introduce correlations through a penalty based on the scale parameter.

In closing, although this problem suggests ample areas of new research, we also believe the daily model provides a substantial improvement in interpreting monitoring data with respect to the regulatory standard. Moreover, our methods are easily implemented with supporting packages in the R environment and so can be used by a broad group of scientists beyond statistical research.

6 Acknowledgements

This research was supported by National Science Foundation grant DMS 9815344. The authors thank William Cox for making the ozone data sets available.

References

[1] Cressie, Noel A.C. *Statistics for Spatial Data (Revised Edition).* Wiley Interscience, New York, 1993.

[2] Fuentes, Montserrat. Statistical assessment of geographic areas of compliance with air quality. *Journal of Geographic Research–Atmosphere*, 108(D24), 2003.

[3] Gilleland, Eric, Nychka, D. (Major Advisor) and Lee, T. (Major Professor) Ph.D. dissertation: Statistical Models for Monitoring and Regulating Ground-level Ozone. Colorado State University, Fort Collins, CO, (in progress).

[4] Green, P.J. and Silverman, B.W. *Nonparametric Regression and Generalized Linear Models.* Chapman and Hall, 2-6 Boundary Row, London SE1 8HN, UK, 1994.

[5] Hastie, T.J. and Tibshirani, R.J. *Generalized Additive Models.* Chapman and Hall, London; New York, 1990.

[6] Holland, David M., Cox, William M., Scheffe, Rich, Cimorelli, Alan J., Nychka, Douglas and Hopke, Philip K. Spatial prediction of air quality data. *EM: A White Paper*, pages 31–35, 2003.

[7] Ma, Chunsheng. Stationary spatio-temporal covariance models. *Journal of Multivariate Analysis*, in press, 2002.

[8] Nychka, D., Meiring, W., Royle, J.A., Fuentes, M. and Gilleland, E. Fields: R Tools for Spatial Data. `http://www.cgd.ucar.edu/stats/software`, 2002.

[9] Stein, Michael L. *Interpolation of spatial data: some theory for kriging.* Springer-Verlag, 175 Fifth Ave., New York, N.Y. 10010, 1999.