The University of British Columbia

Department of Statistics

Technical Report #239

Combining Measurements and Physical Model Outputs for the Spatial Prediction of Hourly Ozone Space - Time Fields

Zhong Liu, Nhu Le, James V. Zidek

July 8, 2008

Combining Measurements and Physical Model Outputs for the Spatial Prediction of Hourly Ozone Space - Time Fields*

Zhong Liu[†], Nhu D. Le[‡], James V. Zidek[§]

July 9, 2008

Abstract.This technical report extends to a spatial setting, an existing temporal two-step linear regression recalibration procedure designed to make the outputs from a deterministic time series simulator comparable with measurements of that series. The result, which Kriges the site specific coefficients of that procedure, enables both temporal forecasting and spatial prediction. Although the extension is somewhat ad hoc, unlike an alternative in work now in preparation, it is computationally simple and fairly transparent. Moreover, in an application where the procedure is used to combing measurements of and simulated model outputs for an hourly ozone field over the eastern and central regions of the United States, we find that it outperforms the model output alone and Kriging, a purely spatial method.

*The work was supported in part by grants from the Natural Science and Engineering Research Council of Canada (NSERC) and the Pacific Institute for the Mathematical Sciences (PIMS).

[†]Department of Statistics, U. of British Columbia. Email: zliu@stat.ubc.ca

[‡]British Columbia Cancer Research Centre. Email: nle@bccrc.ca

[§]Department of Statistics, U. of British Columbia. Email: jim@stat.ubc.ca

Keywords: ozone; deterministic model output; Kriging, spatial-temporal model; spatial prediction; MAQSIP; AQM; AQS

1 Introduction

This technical report presents a method we call space - time model regression (STMReg) for combining the outputs from a numerical, chemical transport model (CTM) for fields of hourly ozone concentrations with measurements made at fixed monitoring sites. However, STMReg could be adapted for other environmental fields for which deterministic process simulator outputs are available as they are in this case.

Ozone is of particular interest as it has long been associated with damage both to human health and welfare, the latter referring to such things as reduced crop yields and damage to vegetation (EPA (2005)). Hence in the United States (US), the Clean Air Act requires that National Ambient Air Quality Standards (NAAQS) be set, periodically reviewed, and enforced to protect both. To afford that protection, metrics computed from hourly concentrations of ozone must be regulated according to the evidence.

However characterizing the hourly ozone field in rural areas proves challenging because of the sparsity of monitors there. This necessitates the use of a spatial predictor to characterize that field and in the most recent review of the NAAQS for ozone, kriging was used (EPA (2005)). The well - known deficiencies of that method, even the monitoring network is reasonably dense (Le and Zidek (2006)), points to the desirability of an better method.

But no spatial prediction method can be expected to do better when the network is sparse because the statistical dependence between sites declines with distant meaning that the monitored sites will generally provide little information about distant unmonitored sites. That suggests the need to provide an improved foundation on which to build the predictor. In this technical report we do just that and use the output from a CTM to bring in knowledge from the physical sciences as a supplement the noisy measurements.

Fortunately such models for air pollution have long been available owing to the importance of the so - called criteria pollutants like ozone. Although deterministic, these numerical computer models (CTMs) simulate the processes involved in its formation and transportation of the pollutants. We refer to their outputs as "model outputs." They have been constructed for a variety of purposes, one been to characterize the "Policy Related Background Level", the subject of another paper now under preparation (Liu et al. (2008b)).

The present technical report uses the AQM (airs quality model), the non-hydrostatic version of the MAQSIP (Multiscale Air Quality Simulation Platform Odman and Ingram (1996)) for its foundation. The AQM requires as inputs the outputs of another deterministic model, the MM5 (Mesoscale Meteorological 5) described by Grell et al. (1995), and simulates mesoscale atmospheric circulation. In this report, ground level ozone concentrations are measured in parts-per-billion in volume (ppbv) and the CTM outputs are simulated at a resolution of 6×6 km^2 grid cells. However we could in principle, set other resolutions for those grid cells. In general, it is more computationally expensive to simulate outputs on grid cells with higher resolution.

The measurements of hourly ozone concentrations used in this technical report, come from the monitoring stations in the AQS (air quality system) monitoring network. That network was set up by US Environmental Protection Agency to monitor criteria pollutants and insure compliance with the NAAQS to meet the requirements of the USA's Clean Air Act.

Although the model outputs are simulated at a different spatial resolution (the mesocale) from that on which the measurements are taken (the microscale), the two series of hourly values are linearly correlated. In fact, Kasibhatla and Chameides (2000) shows there are linear correlations between different quantiles of measurements and model outputs. Also Hogrefe et al. (2001) conclude that measurements and model outputs have a stronger correlation at longer time scales. Fiore et al. (2003, 2004) show that the first two principal components of model outputs and measurements are strongly correlated. This correlation can be exploited and we show that the model outputs can enhance our spatial predictions of ozone.

At the same time, these model outputs need to be recalibrated (Liu et al. (2008b)). For one thing, they are produced on a coarser grid cell than the one on which responses are measured. Hence they may fail to incorporate some local geographical or meteorological information which affects the local ozone concentration level. That discrepancy points to the need for recalibration to achieve better predictive performance.

Before describing next what we do in this technical report, we note that Bayesian melding proposed by Fuentes and Raftery (2005) provides another approach for combining measurement with model output for spatial prediction and it works quite well in some respects as shown by Liu et al. (2008a). That approach has the advantage of recognizing at the fundamental difference between the meso- and micro - scales that are intrinsic to the simulated output and measurements respectively. In fact, it resembles Reynolds's averaging in its approach. However, it proves to be very computationally challenging. As well, like kriging, it is a spatial (only) method and hence does allow strength to be borrowed time by exploiting the temporal autocorrelation that can be quite strong for hourly ozone concentrations.

Our approach builds on the one Guillas et al. (2006) proposes, a two-step linear regression approach to forecast the total column ozone level by adjusting the model outputs from a CTM according to the satellite measurements. In this technical report, we extend this approach in an ad-hoc way to a spatial-temporal model to get STMReg, which not only forecasts future ozone but also spatially predicts those concentrations at unmonitored sites.

We organize this technical report as follows. Section 2 reviews the approach proposed by Guillas et al. (2006). Section 3 presents our extension to a spatial temporal context the result, STMReg. Section 4 presents our application to the substantive problem of characterizing hourly ozone concentration fields. Finally Section 5 concludes this technical report with our findings and some discussion.

2 Two-Step Linear Regression

This section reviews the two-step linear regression approach proposed by Guillas et al. (2006) on which our work in ensuring sections builds. As mentioned in Section 1, ozone measurements are obtained at each of a number of monitored sites (or "stations" in a common terminology). At the same time, model outputs are simulated on grid cells, some containing stations and some not. For spatial prediction, we choose grid cells that contain the monitored sites and as well grid cells that contain the unmonitored sites to be predicted to maximize prediction accuracy by borrowing as much strength as possible from the CTM output.

Both the model outputs and measurements of ozone represent hourly concentrations and at each station, we have two hourly time series, measurements and model outputs. The two-step linear regression approach fits two models at each station where for simplicity, station subscripts are suppressed:

$$
\begin{aligned}
O_t &= c + aM_t + N_t \\
N_t &= \sum_{i=1}^{p} \rho_i N_{t-i} + \epsilon_t,
\end{aligned} \tag{1}
$$

and

$$
\begin{aligned}
\epsilon_t &= \boldsymbol{\alpha} \boldsymbol{Z}_t + e_t \\
e_t &\sim N(0, \sigma_e^2),
\end{aligned} \tag{2}
$$

O_t and M_t being the measurement and model output, respectively, at time t. The parameters in Models (1) and (2) are station specific, that is, they vary from station - to - station. To incorporate the temporal correlation of the measurements, we assume the residuals $\{N_t\}$ form an auto-regressive (AR) process as defined by Box et al. (1994). In practice, to decide the order of the AR process $\{N_t\}$, we first fit a simple linear regression model with $\{O_t\}$ as the responses and $\{M_t\}$, the covariates. Then we identify the order of the AR process $\{N_t\}$ from the PACF (partial auto-correlation function) plot of the residuals. In Section 4, we show that we may reasonably take $\{N_t\}$ to be an AR(1) process in our application.

To implement the two-step linear regression approach, we first fit Model (1) using generalized least squares (GLS) method of Box et al. (1994) with $\{N_t\}$ an AR(1) process. However, unlike ordinary least squares (OLS), the GLS does not assume the residuals are identically or independently distributed. We use the function "gls" in the R package "nlme" developed by Pineiro et al. (2008) to implement the GLS method. If the model outputs capture the temporal correlation of the measurements well, then Model (1) should be enough. However, the model outputs may fail to capture some seasonal trend in the measurements, for example, the 24-hour periodic cycles in the hourly measurements. In that case, the residuals $\{\epsilon_t\}$ may have non-zero means. So we need to fit Model (2) in the second step. In the paper by Guillas et al. (2006), the covariates $\{\boldsymbol{Z}_t\}$ include the 12-month indicator functions and other meteorological or chemical variables which are related to the formation of ozone. In Section 4, we simply use the indicator functions for each of the 24 hours as the covariates $\{\boldsymbol{Z}_t\}$ because no other meteorological or chemical variables are available in our dataset.

The residuals $\{e_t\}$ of Model (2) follow normal distributions $N(0, \sigma_e^2)$ independently and identically. So we simply use the OLS method to fit Model (2).

Suppose at each station, we have measurements and model outputs from time $t = 1$ to T, these being hours, days, months or even years. From time $t = T+1$ to $T+T^{'}$, only the model outputs are available and we want to forecast the measurements for that time period. Without loss of generality, we assume that the $\{N_t\}$ form an AR(1) process. First, we fit Models (1) and (2) as described previously to obtain the estimates $\hat{c}$, $\hat{a}$, $\hat{\rho}$ and $\hat{\boldsymbol{\alpha}}$. Second, we obtain forecasts by using the following equations iteratively

$$\begin{aligned} \hat{O}_t &= \hat{c} + \hat{a} M_t + \hat{N}_t \\ \hat{N}_t &= \hat{\rho} \hat{N}_{t-1} + \hat{\boldsymbol{\alpha}} \boldsymbol{Z}_t, \end{aligned} \tag{3}$$

where $t = T+1, \cdots, T+T^{'}$ and $\{\hat{O}_t\}$ are the forecasts of the future measurements. We have $\hat{N}_T$ from fitting Models (1) and (2).

The disadvantage of the above two-step regression approach is that it is a purely temporal model. That is, it can only forecast the future measurements for stations with a historic series of measurements. But it cannot spatially predict unmeasured values at sites without such a series. Also, since we fit Models (1) and (2) and forecast future measurements at each stations independently, we cannot borrow strength from neighboring stations to impress forecast accuracy. That is, we totally ignore the spatial correlations between the measurements across stations. To incorporate that spatial correlation, we propose an simple spatial-temporal approach that uses Kriging, a classical spatial interpolation approach, to exploit the spatial correlation.

3 Including Spatial Dependence

This section extends the two-step linear regression approach of the previous section, to a spatial-temporal procedure we called STMReg. First, we apply the two-step linear regression approach at each of the stations where historic measurements are available. Then we spatially interpolate the parameter estimates (a, c, ρ, $\boldsymbol{\alpha}$) to locations without such measurements using Kriging. Finally we plug the interpolated parameter estimates into Equation (3) to predict measurements. The underlying assumption in this extension is that the spatial inter - site correlations can be captured by a combination of the model output series that are available at all sites, and the parameter correlations across sites.

For completeness, we now describe the Kriging methodology, a classical approach to spatial interpolation that starts with measurements of a random field $\boldsymbol{Y} = [Y(\boldsymbol{s}_1), ..., Y(\boldsymbol{s}_n)]^T$. To predict the random response Y at a new site $\boldsymbol{s}_0$, the simplest Kriging approach, ordinary Kriging assumes constant means and variances over the spatial domain. That is $\mathrm{E}[\boldsymbol{Y}(\boldsymbol{s}_i)] = \mu$ and $\mathrm{Var}[\boldsymbol{Y}(\boldsymbol{s}_i)] = \sigma^2$ for all $i = 1, \cdots, n$. Universal Kriging allows instead, the mean of the random field $\{Y(\boldsymbol{s})\}$ to depend on the site specific coordinates indexed by the vector $\boldsymbol{s}$. Section 4 shows that ordinary Kriging is adequate for our application.

The interpolated value produced by Kriging at $\boldsymbol{s}_0$ is

$$Y^*(\boldsymbol{s}_0) = \sum_{i=1}^{n} w_i Y(\boldsymbol{s}_i),$$

where w_i is the weight assigned to $Y(\boldsymbol{s}_i)$ under the constraint $\sum_{i=1}^{n} w_i = 1$. The weights $\{w_i\}$ depend on the larger the spatial correlation between $Y(\boldsymbol{s}_i)$ and $Y(\boldsymbol{s}_0)$ the bigger the weights. We find them by minimizing the mean square error (MSE) of the interpolator $Y^*(\boldsymbol{s}_0)$. Obviously $Y^*(\boldsymbol{s}_0)$ is an unbiased estimator of $Y(\boldsymbol{s}_0)$ so we can minimize that MSE by minimizing its variance:

$$\begin{aligned}\sigma_E^2 &= E\left[(Y^*(\boldsymbol{s}_0) - Y(\boldsymbol{s}_0))^2\right] \\ &= -\gamma(\boldsymbol{s}_0 - \boldsymbol{s}_0) - \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) + 2\sum_{i=1}^{n} w_i \gamma(\boldsymbol{s}_i - \boldsymbol{s}_0). \qquad (4)\end{aligned}$$

Note that the so-called semi-variogram function, $\gamma(\cdot)$, which plays an important role in Kriging, is defined by

$$\gamma(\boldsymbol{h}) = \frac{1}{2}C(\boldsymbol{h}) = \frac{1}{2}\mathrm{Var}\left[Y(\boldsymbol{s}+\boldsymbol{h}) - Y(\boldsymbol{s})\right].$$

This technical report relies on the exponential covariance function,

$$C(\boldsymbol{h}) = \sigma^2 \exp(-|\boldsymbol{h}|/\lambda) \text{ if } |\boldsymbol{h}| > 0, \text{ otherwise } C(\boldsymbol{h}) = \sigma^2,$$

σ^2 being the spatial variance and λ being the range parameter, to generate its inter - site covariance field. It seems adequate for our application. However, we note that many other choices are available, the most common being based on the Gaussian kernel, the double exponential kernel, or the Matern kernel. In fact, any function that can generate a positive definite matrix yields a valid covariance function. Moreover Bochner's celebrated theorem states that any valid covariance function must be the Fourier transforms of non-negative Borel measures (Stein (1999)). We refer interested readers to the excellent introduction to this topic given by Wackernagel (1998) and to Cressie (1993); Stein (1999) for some asymptotic theory. Finally, to implement Kriging, we use the R package "geoR" developed by Ribeiro Jr. and Diggle (2001).

In summary, the spatial-temporal model we propose here involves the following steps:

1. Fit Models (1) and (2) at the stations where historic measurements are available to obtain estimates $\hat{a}$, $\hat{c}$, $\hat{\rho}$, $\hat{\boldsymbol{\alpha}}$ at each of the stations.

2. Use simple Kriging to interpolate the estimates from Step 1, to the stations where no historic measurements available.

3. Forecast unmeasured responses at other sites as in Section 2.

Thus this spatial-temporal method is essentially the two-step linear regression with Kriging used to interpolate the estimates.

4 Application to Ozone Fields

This section addresses a problem of great practical importance, the characterization of the field of hourly ozone concentrations over large regions of the United States. We show how STMReg can be applied and compare the result of using that procedure with alternatives.

The measurements and model outputs of hourly ozone data range from May 15 to September 11, 1995 over eastern and central regions of the United States. The model outputs are generated by AQM and the measurements are obtained from the AQS monitoring network. The model output data are complete but some measurements are missing. To avoid extraneous side issues that would arise from using sites with a large amount of missing data in the assessment of this technical report's new methodology, we just choose the 78 of the 375 monitoring stations in the AQS network, having less than 100 missing measurements. Since in those 78 stations, relatively few measurements are missing, we used for simplicity, the 24-hour mean to fill in the missing measurements. For example, if measurements are missing at 10AM for some of the days, then we use the average of measurements at 10AM on other available days to fill in those missing values. To match the measurements with model outputs, we also chose 78 grid cells each of which contains one of those 78 stations. Finally although more than 2800 hours' data are available, we use just the first 480 hours, enough for our assessment of STMReg.

To investigate the performance of STMReg, we conducted an empirical study, using only the measurements and model outputs for the first 240 hours at 15 stations, leaving the balance in a hold - out sample to serve as a validation set. Joining these data were the measurements from the remaining 63 stations. The validation process consisted then of imputing the measurements in that hold - out sample. To avoid confusion, we use the term "forecast" to refer to the imputation of future measurements in the hold - out sample from the 15 stations with historical measurements and "prediction", to refer to the imputation of all measurements in the hold - out sample from the remaining 63 stations. Figure 1 shows the locations of the chosen 78 stations. That figure reveals that the 78 stations are sparsely scattered over a vast area, thus challenging our interpolation method.

As first step we computed the Pearson's correlation coefficients between measurements and model outputs over time at the 78 stations, displayed in Figure 2. We can see that the correlation coefficients are larger than 0.5 at most stations, which proves strong linear relationship between measurements and model outputs. In the two-step linear regression approach, we assume that the process $\{N_t\}$ is a AR(1) process. To justify that assumption, Figure 3 shows the ACF(autocorrelation function) and PACF(partial autocorrelation function) plots of the residuals from a simple linear regression model and the two-step linear regression approach for Station 1. We can see that there is almost no temporal correlation left in the residuals from the two-step linear regression approach.

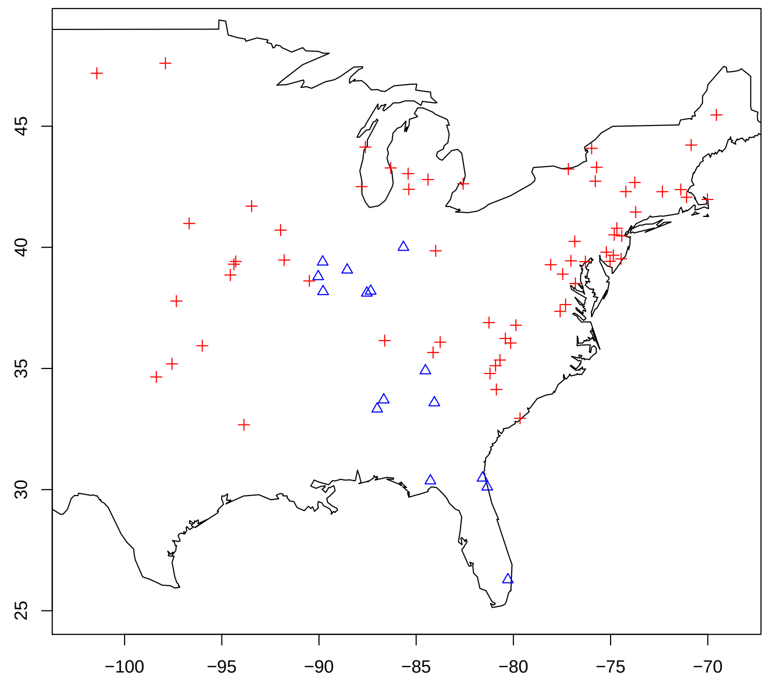

Figure 1: The x-axis is the longitude in degrees, y-axis, the latitude in degrees. Δ: 15 stations with historic measurements; +: 63 stations "without" measurements.

Such plots for the other stations show similar results, thus validating our decision to treat $\{N_t\}$ as an AR(1) process. In this data analysis, the covariates $\{\boldsymbol{Z}_t\}$ in Model (2) are the 24-hour indicator functions since we have hourly measurements and model outputs.

In the data analysis, we first apply the two-step linear regression approach to the measurements and model outputs at each of the 15 stations. Table 1 presents the estimates of parameters c, a and ρ at the 15 stations. The estimates and their standard errors show that the parameters c and a differ significantly from 0 and 1 at most of the stations (10 out of 15). We use the RMSPE (root mean square prediction error) or RMSFE (root mean square forecast error) to evaluate the prediction and forecast results. At each station $\boldsymbol{s}$, we define RMSPE or RMSFE as

$$\sqrt{\frac{1}{T'}\sum_{t=T+1}^{T+T'}\left(O_{\boldsymbol{s},t}-\hat{O}_{\boldsymbol{s},t}\right)^2}, \qquad (5)$$

$O_{\boldsymbol{s},t}$ being the measurement in the hold - out sample at station $\boldsymbol{s}$ and time t while $\hat{O}_{\boldsymbol{s},t}$ is the corresponding imputed value. We calculate the RMSFE at each of the 15 stations with historic measurements and the RMSPE at the remaining 63 stations in the hold - out sample.

Now we compare the forecast and prediction results between the raw model outputs and the adjusted model outputs obtained by applying the STMReg procedure. Table 2 presents the RMSFE results at each of the 15 stations. Tables 3 and 4 present the RMSPEs of three competitive prediction approaches: adjusted model outputs, raw model outputs, and Kriging. Here Kriging means we only use the measurements at the 15 stations to spatially interpolate them at the remaining 63 stations. Because Kriging is a purely spatial approach, we apply Kriging to each of the 240 hours respectively.

To illustrate better the forecasts and predictions obtained from the adjusted model output, we use a pair-wise T-test to test the hypothesis that the adjusted and raw outputs have the same RMSPE and RMSFE. The p-value for RMSFE is $0.00012 < 0.05$, RMSPE, $0.001 < 0.05$. The small p-values show that we can significantly improve the prediction and forecast results by using the two-step linear regression approach and its extension to adjust the model outputs.

We are now in a position to present our conclusions and we do so in the next section.

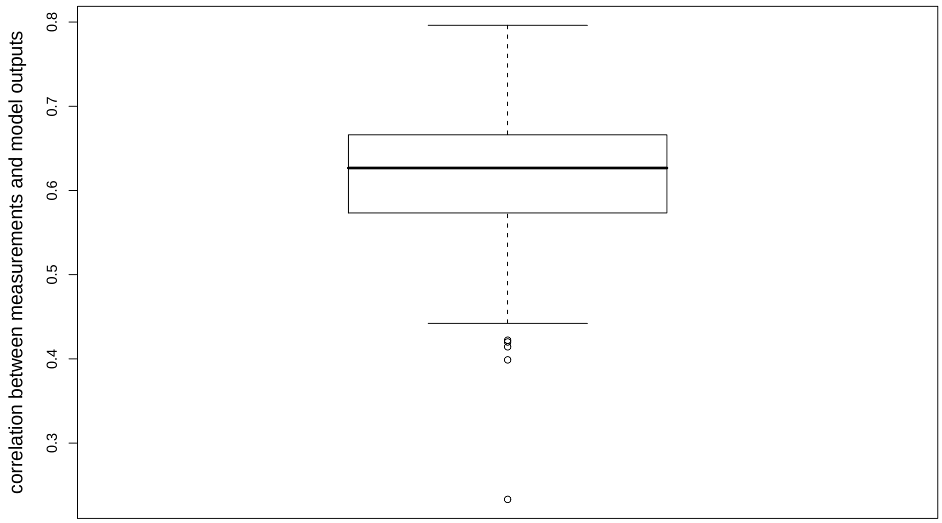

Figure 2: Boxplot of Pearson's correlation coefficients between hourly measurements and model outputs at the 78 stations in our study.

5 Discussion and Conclusions

The application made in the previous section, of the STMReg procedure proposed in this report leads to the following conclusions.

- The two-step linear regression approach can only forecast the unmeasured ozone concentrations, Kriging can only spatially predict them, while STMReg can both forecast and predict them at non-monitored sites and future times in the geographical domain of interest.

- Tables 2, 3 and 4 show that in averaging over the stations, STMReg gives the smallest RMSFE and RMSPE compared with the model outputs alone and with Kriging. For forecasting, STMReg can reduce the mean RMSFE by about 30% compared with model outputs, the mean RMSPE, by 8%. Both the RMSFE and RMSPE for the new approach are significantly smaller than those for the model

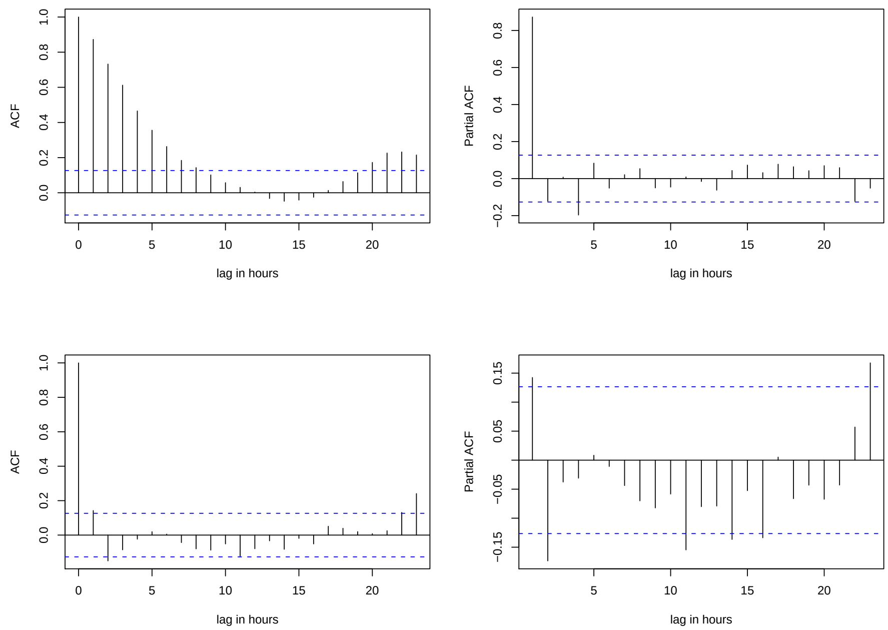

Figure 3: ACF and PACF plots of the residuals of a simple linear regression model and the two-step linear regression approach for Station 1. The top two plots are the ACF and PACF plots of the residuals of a simple linear regression. The two plots below are the ACF and PACF plots of the residuals of the two-step linear regression approach.

outputs alone. STMReg has a smaller RMSFE than model outputs at all the 15 stations and a smaller RMSPE at 47 out of 63 stations.

- In averaging over 63 stations, Kriging with measurements approach has the biggest RMSPE. This is not surprising because we only use 15 stations to predict measurements at 63 stations in a relatively big region. Moreover, Kriging is a purely spatial approach that does not take advantage of the temporal correlation in the data.

- Table 1 shows the estimates $\hat{\rho}$ to be around 0.9 at the 15 stations, which shows that $\{N_t\}$ has strong auto-correlation.

To create STMReg, we fit Models (1) and (2) independently. We then use Kriging to interpolate the parameter estimates across space. However the model underlying the new approach seems "self-contradictory" in that it entails fitting Models (1) and (2), assuming the ozone concentrations are conditionally independent across space given the site specific parameters, and then using Kriging to capture their spatial correlation. So strictly speaking, that model is not a spatial - temporal one. Moreover, this ad-hoc model does not provide estimates of uncertainties for either the forecasts or predictions. So in work currently in preparation, we present an alternative approach that overcomes some of these deficiencies, where spatial correlation plays a more fundamental role.

However, the latter comes with the price of complexity, along with added computing costs and a lack of transparency. Thus, despite its somewhat ad hoc nature, we believe STMReg will serve as an attractive option in many situations.

Acknowledgements

We are indebted to Dr Prasad Kasibhatla for supplying in convenient form, the data used in our analysis and for suggesting to the third author, while he was on leave at the Statistical and Mathematical Sciences Institute (SAMSI), the problem of finding a way to use AQM output in the imputation of AQS ozone concentrations.

References

Box, G., Jenkins, G. and Reinsel, G. (1994). *Time Series Analysis: Forecasting and Control; 3rd Edition*, Holden-Day.

Cressie, N. A. (1993). *Statistics for Spatial Data*, Wiley-Interscience.

EPA (2005). Air quality criteria for ozone and related photochemical oxidants.

Fiore, A. M., Jacob, D., Liu, H., Yantosca, R. M., Fairlie, T. D. and Li, Q. (2004). Variability in surface ozone background over the united states: Implications for air quality policy, *Journal of Geophysical Research*.

Fiore, A. M., Jacob, D., Mathur, R. and Martin, R. V. (2003). Application of empirical orthorgonal functions to evaluate ozone simulation with rgional and global models, *Journal of Geophysical Research*.

Fuentes, M. and Raftery, A. (2005). Model evaluation and spatial interpolation by bayesian combination of observations with outputs from numerical models, *Biometrics* **61**: 36–45.

Grell, G., Dudhia, J. and Stauffer, D. (1995). A description of the fifth-generation penn state/ncar mesoscale model (mm5), *Technical report*, National Center for Atmospheric Research, Boulder, Colorado.

Guillas, S., Tiao, G., Wuebbles, D. and Zubrow, A. (2006). Statistical diagnostic and correction of a chemistry-transport model for the prediction of total column ozone, *Atmospheric Chemistry and Physics* **6**: 527–537.

Hogrefe, C., Rao, S., Kasibhatla, P., Hao, W., Sistla, G., R., M. and McHenry, J. (2001). Evaluating the performance of regional-scale photochemical modeling systems: Part ii-ozone predictions, *Atmos. Environ.* **35**: 4175–4188.

Kasibhatla, P. and Chameides, W. L. (2000). Seasonal modeling of regional ozone pollution in the eastern united states, *Geophys. Res. Letter* **27**: 1415–1418.

Le, N. D. and Zidek, J. V. (2006). *Statistical Analysis of Environmental Space-Time Process*, Springer.

Liu, Z., Le, N. and Zidek, J. (2008a). An appraisal of bayesian melding for physical - statistical modeling, *In preparation*.

Liu, Z., Le, N. and Zidek, J. (2008b). Calibration of deterministic models for ozone fields, *In preparation*.

Odman, M. and Ingram, C. (1996). Multiscale air quality simulation platform (maqsip) source code documentation and validation, *Technical report*, Environmental Programs, MCNC-North Carolina Supercomputing Center.

Pineiro, J., Bates, D., DebRoy, S. and Sarkar, D. (2008). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-88.

Ribeiro Jr., P. and Diggle, P. (2001). geoR: a package for geostatistical analysis, *R-NEWS* **1**(2): 15–18.

Stein, M. (1999). *Interpolation of spatial data : some theory for kriging*, New York : Springer.

Wackernagel, H. (1998). *Multivariate Geostatistics - An Introduction with Applications, 2nd Edition*, Springer-Verlag.

Table 1: The number in bold font indicates that the estimates c and a significantly differ from 0 and 1 at that station at significance level, 0.05.

station	$\hat{c}$	$sd(\hat{c})$	$\hat{a}$	$sd(\hat{a})$	$\hat{\rho}$
1	10.41	6.62	0.47	0.09	0.91
2	12.55	7.44	0.40	0.10	0.91
3	26.36	3.01	0.36	0.06	0.85
4	20.75	4.70	0.40	0.07	0.90
5	13.07	4.47	0.65	0.09	0.87
6	27.83	4.58	0.14	0.06	0.93
7	4.01	10.54	0.67	0.17	0.93
8	6.23	7.94	0.70	0.09	0.93
9	18.96	4.76	0.45	0.09	0.89
10	25.30	3.78	0.23	0.07	0.91
11	24.09	3.30	0.41	0.07	0.85
12	24.40	4.67	0.38	0.07	0.92
13	25.97	4.39	0.41	0.07	0.91
14	35.20	5.09	0.25	0.07	0.91
15	7.85	5.60	0.56	0.09	0.89

Table 2: The second and third columns are the RMSFE (root mean square forecast error) of the forecasts by STMReg and the model outputs. The number in bold font indicates that STMReg has the smaller RMSFE at that station. The unit of measurements for the numbers in this table is ppbv.

station	ad-hoc	model outputs
1	16.55	26.30
2	15.40	31.13
3	13.65	16.29
4	16.04	20.06
5	13.92	14.36
6	11.97	16.51
7	14.96	26.90
8	16.28	18.98
9	12.03	16.74
10	9.08	11.58
11	11.35	16.64
12	13.31	15.89
13	10.01	15.04
14	9.87	14.66
15	12.19	17.51
average	13.11	18.57

Table 3: RMSPE at stations 1-35. Column 2: ad-hoc approach; column 3: model outputs; column 4: Kriging approach. The number in bold font indicates the ad-hoc has the smallest RMSPE at that station. The unit of measurement for the numbers in this table is ppbv.

station	ad-hoc	model outputs	Kriging
1	13.64	16.32	16.51
2	10.12	13.48	17.05
3	13.19	13.38	24.15
4	13.51	22.65	19.17
5	25.35	19.59	43.40
6	17.31	11.85	40.89
7	12.14	14.55	22.80
8	12.15	17.35	23.10
9	15.91	21.85	27.81
10	14.11	13.31	31.90
11	15.65	11.49	35.59
12	16.42	15.02	29.32
13	19.16	11.77	36.71
14	14.78	16.46	20.84
15	12.35	15.05	16.70
16	11.62	14.24	17.11
17	16.92	18.35	19.14
18	15.65	16.15	23.07
19	13.00	14.76	18.33
20	12.17	15.38	17.70
21	10.57	12.65	18.79
22	9.48	11.21	14.83
23	13.51	16.11	10.38
24	9.29	14.17	19.64
25	12.51	15.47	24.78
26	11.21	13.33	25.74
27	15.95	15.94	27.33
28	12.46	14.97	26.94
29	18.69	18.02	32.11
30	17.13	13.94	32.67
31	17.09	21.82	25.94
32	20.92	16.28	34.19
33	20.83	16.47	32.95
34	15.19	15.39	27.05
35	20.63	16.01	24.43

Table 4: RMSPE at stations 36-63. Column 2: ad-hoc approach; column 3: model outputs; column 4: Kriging approach. The number in bold font indicates the ad-hoc has the smallest RMSPE at that station. The unit of measurement for the numbers in this table is ppbv.

station	ad-hoc	model outputs	Kriging
36	20.63	16.01	35.43
37	20.89	18.30	33.09
38	12.84	13.91	26.54
39	20.18	24.27	20.20
40	15.18	18.99	16.91
41	16.70	20.00	16.57
42	16.00	18.46	18.77
43	13.03	19.99	26.53
44	11.38	17.69	23.84
45	13.26	17.79	14.22
46	13.79	13.75	31.88
47	14.56	15.21	28.34
48	15.41	18.91	19.62
49	19.41	19.65	25.71
50	15.48	13.42	13.38
51	12.64	12.86	14.31
52	12.90	16.20	13.79
53	15.73	19.14	16.47
54	13.53	18.63	13.77
55	17.23	17.42	15.31
56	15.20	18.32	21.02
57	16.50	20.09	25.79
58	16.52	17.01	25.05
59	14.50	18.97	23.48
60	14.75	18.36	18.03
61	13.58	16.15	19.83
62	18.75	24.94	14.62
63	14.32	15.92	19.99
average	15.05	16.51	23.36