

Statistical correction and downscaling of chemical transport model ozone forecasts over Atlanta

Serge Guillas^{a,*}, Jinghui Bao^a, Yunsoo Choi^b, Yuhang Wang^b

^a*School of Mathematics, Georgia Institute of Technology, 686 Cherry Street, Atlanta, GA 30332-0160, USA*

^b*School of Earth and Atmospheric Sciences, Georgia Institute of Technology, USA*

Received 4 January 2007; received in revised form 14 September 2007; accepted 18 October 2007

Abstract

The Regional Air Quality forecAST (RAQAST) model is a regional chemical transport modeling system for ozone and its precursors over the United States. Since the grid size is 70 by 70 km, forecasts cannot be made for a specific surface site. We use EPA monitoring stations from the Atlanta area to downscale and improve local forecasts using RAQAST outputs. We use the Model Diagnostic and Correction (MDC) approach. First, we regress the observations on the model outputs with an autoregressive noise component. Second, we regress the residuals of this first regression on variables associated with wind speed, precipitation amounts and the diurnal cycle. Deficiencies of 3-D model results are identified and corrected. Evaluation using measurements for a different period confirms that the statistically adjusted outputs reduce forecast errors by up to 25%.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Surface ozone; Model evaluation; Statistical adjustment

1. Introduction

Surface ozone is one of EPA criteria pollutants that can be potentially hazardous to human health and biological ecosystems (NRC, 1991) and is therefore regulated under the National Ambient Air Quality Standards. Ozone is a key precursor for the main oxidant, hydroxyl radicals (OH) in the troposphere. It is also a greenhouse gas, particularly in the upper troposphere. Accurate surface or tropospheric O₃ forecast is challenging because ozone photochemical production is nonlinear. Previously, reasonable forecasting performances have

been obtained using either statistical or 3-D regional chemical transport models (CTMs) (Ryan and Piety, 2000; McHenry et al., 2004; Vaughan et al., 2004; Otte et al., 2005; Jing et al., 2006).

Direct statistical methods for ozone prediction have given reasonably good results. A thorough inter-comparison of statistical techniques (Schlink et al., 2003) shows that these methods can perform very well compared to deterministic CTMs. The best statistical techniques are the ones that can handle nonlinearities, e.g. generalized additive models or neural networks. Statistical methods based on functional data analysis, i.e. treating the collection of data points as curves, exhibited very promising results (Damon and Guillas, 2002; Aneiros-Perez et al., 2004). However, the nature of

*Corresponding author.

E-mail address: guillas@math.gatech.edu (S. Guillas).

statistical modeling does not enable better understanding of chemical and physical processes, which are simulated by CTMs.

Statistical methods can help improve the forecasts when systematic deficiencies can be identified. Both the coarse grid resolution and not-well understood chemical and physical processes are the two major factors explaining the differences. We will correct the 3-D model outputs directly (Guillas et al., 2006). In this work, we investigate if statistical methods can be used to correct and downscale ozone forecasts by a 3-D regional CTM. Our main assumption is that the deficiencies of a CTM can be linearly explained by explanatory variables. This stems from the idea that usually first approximations can be easily modeled linearly. Hence, the nonlinear effects and feedbacks are modeled by the CTM, and some of the model deficiencies are corrected using statistical comparisons between observations and forecasts from the past. Therefore, we combine the strengths of chemical and physical modeling of the CTM with the use of historical data via statistical methods to enable better forecasts. Our approach is novel because of this combination of strengths. Previous studies of ozone forecasts using CTMs have not taken advantage of statistical methods to directly correct the outputs. We show that the CTM forecasts are often greatly improved.

2. Description of the RAQAST modeling system

The Regional Air Quality forecAST (RAQAST) is a fully automated 48-h forecasting system available online (<http://apollo.eas.gatech.edu/forecast>). The RAQAST system has two components, the Penn State/National Center for Atmospheric Research mesoscale model MM5 and a regional CTM. The models have a horizontal spatial resolution of 70 km with 23 vertical layers from the surface to 10 hPa (with 20 layers below 100 hPa). The forecasting system predicts meteorological variables and the concentrations of ozone and its precursors over North America.

Our goal in this work is to improve the chemical forecasts. To minimize the error in meteorological forecasts, we use four-dimensional data assimilation (FDDA) (Stauffer et al., 1991) with the 6-h National Center for Environmental Prediction (NCEP) reanalysis data, surface and rawinsonde observations. Most meteorological variables are archived every 30 min except those related to convection processes, which are archived every

2.5 min because of the highly temporally varying nature of convection. The horizontal domain of MM5 has five extra grids beyond that of CTM on every side to minimize potential transport anomalies near the boundaries. We use the ETA Mellor–Yamada–Janjic (MYJ) 2.5-order closure scheme (Black, 1994) for turbulent calculation.

The regional CTM (Zeng et al., 2003, 2006; Choi et al., 2005; Wang et al., 2006; Jing et al., 2006) has been used and updated from the previous model by McKeen et al. (1991), and the photochemical, dry and wet deposition modules are adopted from the GEOS-CHEM model (Bey et al., 2001 and references therein). The chemistry module incorporates a fast and numerically accurate Gear-type solver (Jacobson and Turco, 1994). The model includes a detailed photochemical mechanism with about 200 reactions and 120 concentration-varying chemical tracers; 24 tracers (family or species) are transported to describe O_3 – NO_x –hydrogen chemistry. The photolysis rate is calculated with an efficient Fast-J algorithm (Wild et al., 2000), which accounts for cloud, aerosol and Mie scattering. The altitude dependent cloud optical depth is calculated using MM5 liquid water content (Stephens et al., 1978). The UV surface albedo for the photolysis rate calculation is obtained from TOMS observations (Herman and Celarier, 1997). The tracer transport scheme is that by Walcek (2000). The cumulus convective scheme by Grell (1993) is implemented in the CTM to be consistent with the meteorological model; sub-grid scale updraft and downdraft processes and large-scale subsidence are considered. The convective wet scavenging for soluble tracers follows that by Liu et al. (2001). The top and bottom layers of convection are determined by MM5 simulations; cloud fraction is determined using the scheme described by Geleyn (1981).

We apply the simulation results of the global GEOS-CHEM model to specify boundary conditions of trace gas concentrations in the regional model. Fossil fuel emissions of NO_x and CO over the US are obtained from the 1999 US Environmental Protection Agency National Emission Inventory version 2. For regions in Canada and Mexico not covered by the EPA NEI inventory, the GEIA global emission inventory (Benkovitz et al., 1996) is used. Other inventories for combustion and industrial sources, and algorithms for emissions from vegetation and soils are taken from the global GEOS-CHEM model (Bey et al., 2001). The monthly mean leaf area index (LAI) distribution is

derived from 1 km Advanced Very High Resolution Radiometer (AVHRR) data by Bonan et al. (2002). The lightning NO_x algorithm in the model is described by Choi et al. (2005). Cloud-to-ground lightning flashes in the model are constrained by the observations from the National Lightning Detection Network (NLDN). The forecasted chemical concentrations from the model are archived every hour.

3. Observations

The observed data set for hourly ozone in Atlanta, Georgia covers the period of June–August 2005. The total number of EPA AIRNow network of measure sites in Atlanta is 12. They are: Confederate Ave., Conyers, Douglasville, Fayetteville, Gwinnett, Kennesaw, McDonough, Newnan, South Dekalb, Tucker, Waleska, Yorkville. Because some sites did not report data during that period and some sites had too many missing values, finally seven sites were selected (see Fig. 1). They are (1) Confederate Ave. (33.721°N , 84.35778°W), (2) Douglasville (33.743°N , 84.779°W), (3) Fayetteville (33.456°N , 84.420°W), (4) Kennesaw (34.014°N , 84.608°W), (5) McDonough (33.435°N , 84.162°W), (6) Newnan (33.404°N , 84.7468°W) and (7) Yorkville (33.928°N , 85.045°W).

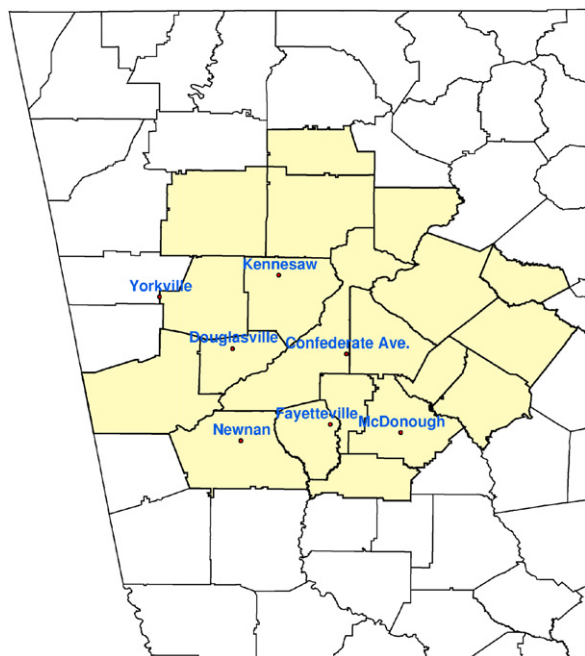


Fig. 1. EPA stations in the Atlanta area used in the study.

The hourly forecasts for the period of June–August 2005 were given by the RAQAST system. The choice of the grid cell is as follows: we pick the grid cell that includes the measurement station, and the grid cells are internally coded through their lower left corners. In the analysis, we use the period of June 1–August 19, 2005 as “historical” and the period of August 20–29 for evaluation. Fig. 2 displays the measurements taken at the seven EPA monitoring stations in the metropolitan Atlanta area used in our study and the RAQAST hindcast for August 21–28, 2005. We can see that the model provides forecasts that are higher or lower than observations. For instance, the peak on August 22 is overestimated by RAQAST, whereas the peak on August 26 is underestimated. Minimum values at night time are systematically underestimated by RAQAST throughout August 21–28, 2005. The coarse resolution of RAQAST is the reason why the model does not capture local variations.

A number of parameters were considered in the statistical forecasts since they can strongly affect ozone concentrations. Temperature, wind speed, boundary layer height and precipitation are included because these variables characterize the meteorological environment for mixing, transport and biogenic emissions. After some preliminary correlation studies, we find that temperature has high correlation with the daily cycle, and wind speed has correlation with boundary layer height. Temperature seems to be well modeled by MM5 data assimilation; and the effect of temperature on ozone, as modeled in RAQAST, seems to be well understood. Indeed, in preliminary analyzes, temperature never played a significant role in explaining the discrepancies between RAQAST and observations. Furthermore, the high correlation with the daily cycle triggers near-collinearities in the regressions and thus deteriorates the diagnostics and correction (Stewart, 1987). Besides, by removing temperature and boundary layer height, we avoid some potential overfitting. So, wind speed and precipitation were selected to statistically improve the RAQAST outputs.

4. Correction and downscaling of model outputs

We follow the Model Diagnostic and Correction (MDC) approach (Guillas et al., 2006). The gist of the method is to explain the deficiencies of the model by a linear combination of a set of explanatory variables. Hence, we assume that, in

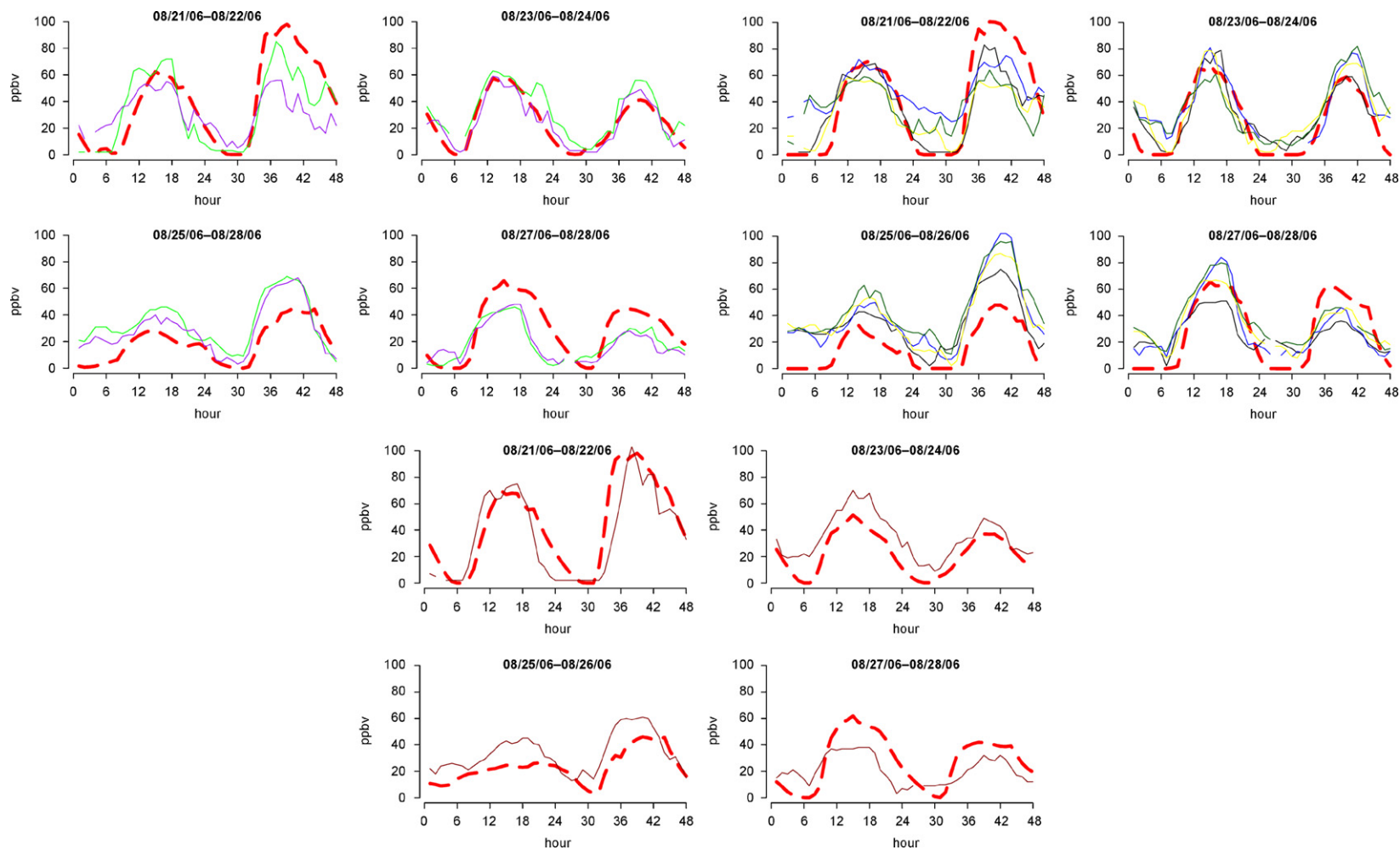


Fig. 2. RAQAST outputs (dashed red line) and EPA measurements (solid lines) from stations located in the corresponding RAQAST grid cell, August 21–28, 2005. Top left panel: cell 1 and stations Fayetteville and Newnan. Top right panel: cell 2 and stations Confederate Ave., Douglasville, Kennesaw, Yorkville. Lower panel: cell 3 and station McDonough.

first approximation, the model errors are linear in the variables (but we do not assume that ozone itself is linearly related to the explanatory variables). First, the deficiencies are estimated in a linear regression of ozone measurements on the model outputs. Second, the deficiencies are attributed to specific ancillary variables (wind speed and precipitation) in a linear regression.

Thus the two steps proceed as follows:

- *Step 1*: Regress the measurements on model outputs, with an autoregressive model of order one (AR(1)) component. We denote by $O(t)$ the measured hourly ozone, and by $M(t)$ the model outputs:

$$O(t) = c + aM(t) + N_t, \quad (1)$$

where N_t is AR(1), i.e. $N_t = \rho N_{t-1} + \varepsilon_t$ with (ε_t) white noise.

- *Step 2*: Regress the estimated residuals $\hat{\varepsilon}_t$ from the first step regression on the indicators of the hours $h_i(t) = 1$ for the hour i and 0 otherwise, precipitation amounts $r(t)$, and wind speed $w(t)$. We included diurnal harmonics of $w(t)$ since the wind speed effect on ozone, even for model deficiencies, can be very different for different hours of the day. We denote by $wc(t)$ the wind speed multiplied by $\cos(2\pi(t-6)/24)$, and $ws(t)$ the wind speed multiplied by $\sin(2\pi(t-6)/24)$:

$$\hat{\varepsilon}_t = \sum_{i=1}^{12} \alpha_i h_i(t) + c_r r(t) + c_{wc} wc(t) + c_{ws} ws(t) + \varepsilon'_t, \quad (2)$$

where (ε'_t) is a white noise.

The AR term in the first step is included to account for unexplained variability. Fig. 3 shows that if we carry out the MDC method without including the AR term, the residuals still present strong correlations, whereas it is not the case when an AR(1) term is included. However, there are still some significant autocorrelations, typically around 0.2, in the residuals from the second regression, so we improve step 1 by including an AR(2) noise:

Step 1: AR(2). Regress the measurements on model outputs, with an autoregressive model of order two (AR(2)) component. We denote by $O(t)$ the measured hourly ozone, and by $M(t)$ the model outputs:

$$O(t) = c + aM(t) + N_t, \quad (3)$$

where N_t is AR(2), i.e. $N_t = \rho_1 N_{t-1} + \rho_2 N_{t-2} + \varepsilon_t$ with (ε_t) white noise.

This choice of an AR(2) noise is better. Fig. 3 displays the autocorrelation functions (ACFs) of the (ε'_t) in step 2 for three cases: step 1 with N_t white noise (left column), step 1 with N_t AR(1) (middle column) and step 1 with N_t AR(2) (right column). It shows that the temporal correlations are high for N_t white noise. It means that (ε'_t) cannot be considered white noise, and thus justifies a modeling of N_t by an autoregressive process. With N_t AR(1), the correlations are much smaller, and with N_t AR(2) they are close to zero. Therefore, the AR(2) approach seems beneficial. Furthermore, comparing N_t AR(2) to N_t AR(1), the Residual Standard Errors in the regressions are lower, the uncertainties in estimates are lower, and the predictions are slightly better. It would be interesting to build a statistical technique that would help us determine the optimal structure for the residuals in the first regression, but this goes beyond the scope of this paper. Moreover, we believe that only a marginal improvement would be gained from fitting a more complicated model than AR(2) in the first step because the ACFs are almost flat.

We illustrate our diagnostics by showing the results of the two regressions for station 6. Throughout the analysis, we use the term significant for statistically significant, under the normal assumption that an estimate which is twice as large in absolute value than its standard deviation is significant at the 95% level. Table 1 gives the results of the first regression in the MDC method. Significant additive and multiplicative biases underscore the overall deficiencies in RAQAST. The multiplicative bias of 0.52 indicates that the amplitude of typical daily variations in RAQAST are larger than the observed ones: The relationship between observations and model outputs is optimal (in the least squares sense) when a factor of 0.52 is applied to the model outputs. Note that this does not mean that the amplitude of the model outputs is twice as large as the amplitude of observations because the term N_t includes such variations as well. Thus, the scalar a ought to be interpreted more as a index for how much information the model can explain in the data. Both autocorrelation parameters ρ_1 and ρ_2 are highly significant. It illustrates the fact that there is a need, in the fitting procedure, for such short-term autocorrelations. The hour-to-hour chemical and physical variations that RAQAST was unable to model are captured by the AR(2) time series.

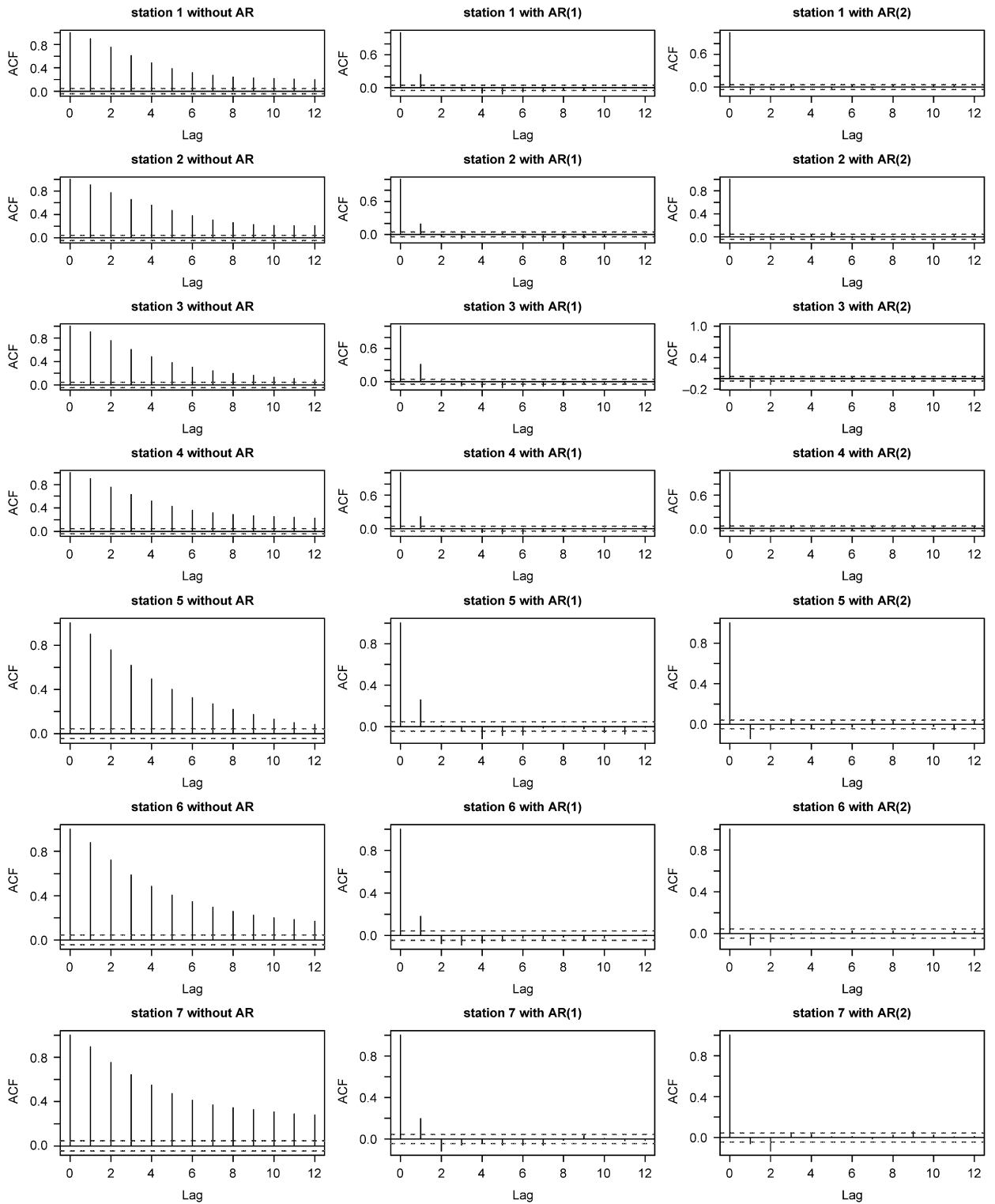


Fig. 3. Sample autocorrelation functions (ACF) of the residuals in step 2, after carrying out the two steps of the MDC method. The ACF are displayed for the cases without or with an AR term in the first regression.

Table 1
Coefficient estimates, with standard errors in parentheses

Step 1, station 6	
<i>Coefficient estimates</i>	
<i>c</i>	11.06 (1.35)
<i>a</i>	0.52 (0.03)
ρ_1	1.21 (0.022)
ρ_2	-0.33 (0.022)

Step 1 regression in the MDC method (1). Estimation is carried out using data over the period June 1 (05:00, hour 1) to August 20 (04:00, hour 24).

Table 2
Coefficient estimates for the hourly indicators, with standard errors in parentheses

Step 2, station 6	
<i>Coefficient estimates</i>	
1	-1.29 (0.69)
2	-2.07 (0.69)
3	-0.94 (0.69)
4	-0.74 (0.68)
5	-1.08 (0.68)
6	-1.28 (0.65)
7	-0.26 (0.65)
8	3.59 (0.66)
9	2.98 (0.66)
10	2.85 (0.67)
11	4.32 (0.68)
12	4.27 (0.69)
13	2.92 (0.69)
14	2.86 (0.70)
15	1.82 (0.70)
16	1.11 (0.70)
17	1.11 (0.69)
18	-1.14 (0.68)
19	-0.99 (0.69)
20	-6.09 (0.69)
21	-5.67 (0.69)
22	-1.46 (0.69)
23	-1.55 (0.69)
24	-1.33 (0.69)

Step 2 regression in the MDC method (2). Estimation is carried out using data over the period June 1 (05:00) to August 20 (04:00). Hence the numbers 1–24 correspond to the 24 h in 05:00–04:00.

Table 2 shows the coefficient estimates for the indicators of the hours, from 05:00 (hour 1) to 04:00 (hour 24). For the morning hours, the coefficients are not significant. This indicates that RAQAST represents well the morning variations of ozone for this location. For the hours 12:00–19:00, the coefficients are significantly positive. It means that,

after the adjustment for additive and multiplicative bias in step 1, and the correction for the wind and precipitation effect in step 2, RAQAST provides predictions that are significantly lower than the observations on average, for this location and for these hours (12:00–19:00). Note that the actual RAQAST predictions, not adjusted for biases, wind and precipitation, are usually higher than observations, since the wind and precipitation effects can override the hourly effect. These hours are the peak hours for ozone, and are considered the most difficult to predict. For the evening hours (19:00–21:00), the coefficients are not significant. For the night (00:00–03:00), the coefficients are significantly negative, which indicate adjusted forecasts higher than observations, when all the other influences are taken into account. As a result, the detection of such hourly deficiencies illustrates the shortcomings of the model in terms of daily amplitude.

Table 3 provides information on the treatments of two meteorological variables by RAQAST for the prediction of ozone: wind speed and precipitation. We used the aforementioned wind speed first harmonics. The only significant wind speed coefficient is the one multiplied by $\sin(2\pi(t-6)/24)$, and is positive. The values of this index are negative in the afternoon. This demonstrates that RAQAST does not fully reflect the influence of wind speed on ozone. More precisely, the positive coefficient underscores the fact that observations are more positively correlated to these wind speed coefficients, and RAQAST forecasts ought to be more related with wind speed first harmonics. In particular, in the afternoon, with large wind speed, the model tends to overpredict ozone concentrations. For precipitation, the coefficient is significantly negative, with a non-negative index. RAQAST is

Table 3
Coefficient estimates for the wind speed multiplied by $\cos(2\pi(t-6)/24)$, the wind speed multiplied by $\sin(2\pi(t-6)/24)$, with standard errors in parentheses, and for the precipitation index

Step 2, station 6	
<i>Coefficient estimates</i>	
Wind speed (cos)	0.14 (0.11)
Wind speed (sin)	0.38 (0.11)
Precipitation	-9.61 (3.12)

Step 2 regression in the MDC method (2). Estimation is carried out using data over the period June 1 (05:00) to August 20 (04:00).

currently not representing well enough the decrease in ozone due to precipitation. RAQAST should predict lower forecasts than it does now, when precipitation occurs. All of these corrections are carried out in the following statistical adjustment.

The corrected model outputs $\tilde{M}(t)$ corresponding to the model outputs $M(t)$ for the time period August 20 (05:00)–August 30 (04:00), for a specific station, is given by

$$\tilde{M}(t) = \hat{c} + \hat{a}M(t) + \hat{N}'_t, \tag{4}$$

where $M(t)$ is the model output for the period August 20 (05:00)–August 30 (04:00) and

$$\hat{N}'_t = \hat{\rho}\hat{N}'_{t-1} + \sum_{i=1}^{12} \hat{\alpha}_i h_i(t) + \hat{c}_r r(t) + \hat{c}_{wc} wc(t) + \hat{c}_{ws} ws(t), \tag{5}$$

where the hats indicate that the quantities are the estimates of the coefficients from the first two steps.

Table 4 provides a comparison of the root mean squared errors (RMSE), when comparing forecasts made by RAQAST or by its statistically corrected version, for the seven stations. It turns out that the RMSE are typically improved by 10–25%. When we compare the daily maximum 8-h average forecasts with the observed daily maximum 8-h averages, used by the EPA, the improvements are typically in the 10–30% range. The RMSEs for the daily maximum 8-h averages should be computed over a larger period than 10 days, and they are shown here for illustration purposes only. A careful study of the average improvement should be done over large

samples of days where peaks are observed. This goes beyond the scope of this paper.

To examine the spread of the errors, Fig. 4 displays the boxplots of errors for station 7. It shows that the errors for the statistically downscaled model are more centered towards 0 and that the spread is largely reduced.

Fig. 5 displays the observations and the forecasts given by RAQAST and its statistically corrected version (upper panel) for station 6, as well as the adjustments due to explanatory variables and hours (lower panel). This is the station where the improvements are the second largest (see Table 4). It illustrates the effect statistical correction. We can see that over this time period the adjustment provided forecasts that are much closer to measurements when the peak was overestimated by the model (August 21, 26 and 27). However, when the model underpredicts ozone concentration, the statistical adjustment tends to make predictions a bit worse (August 23 and 25). Night time ozone concentrations are better predicted by the adjusted model. Moreover, the adjusted model shows smaller variations than the original model.

These adjustments seem to be mainly attributed to the step 1 correction and to a lesser extend to the step 2. Indeed, the smaller amplitude can be attributed to $a = 0.52(0.03)$, and the larger values at night can be explained by $c = 11.06(1.35)$. The wind and precipitation effects are a bit more difficult to discern. The lower panel of Fig. 5 display these influences. The negative index of the wind, multiplied by a sine function, with a positive coefficient of 0.38 (0.11), may explain partly the

Table 4
Root mean squared errors (RMSE) for the forecasts made by the RAQAST model and the corrected RAQAST model, over the validation period August 20 (05:00)–August 30 (04:00)

RMSE	Station						
	1	2	3	4	5	6	7
<i>Hourly</i>							
RAQAST	16.74	18.76	16.32	18.64	15.20	16.73	20.80
Corrected RAQAST	13.24	15.85	14.51	14.25	13.69	12.64	15.64
Improvement (%)	20.90	15.50	11.10	23.60	9.90	24.40	24.80
<i>Daily 8-h maximum average</i>							
RAQAST	2.37	2.83	3.31	2.28	3.24	2.89	2.98
Corrected RAQAST	1.62	2.63	2.82	2.16	2.66	2.28	2.24
Improvement (%)	31.60	7.10	14.80	5.30	17.90	21.10	24.80

Hourly errors and daily 8-h maximum average errors. Respective improvements in percentages are displayed.
Note: RMSEs for daily 8-h maximum average errors are only shown for illustration purposes, since they are based on a small sample of 10 days.

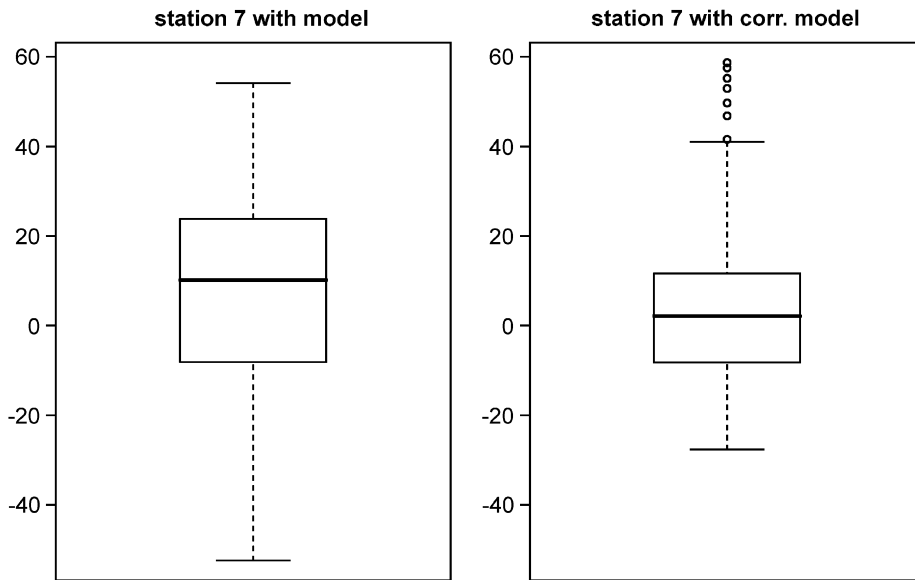


Fig. 4. Station 7. Boxplots of hourly prediction errors, August 20 (05:00)–August 30 (04:00). Left: RAQAST predictions, right: corrected RAQAST predictions.

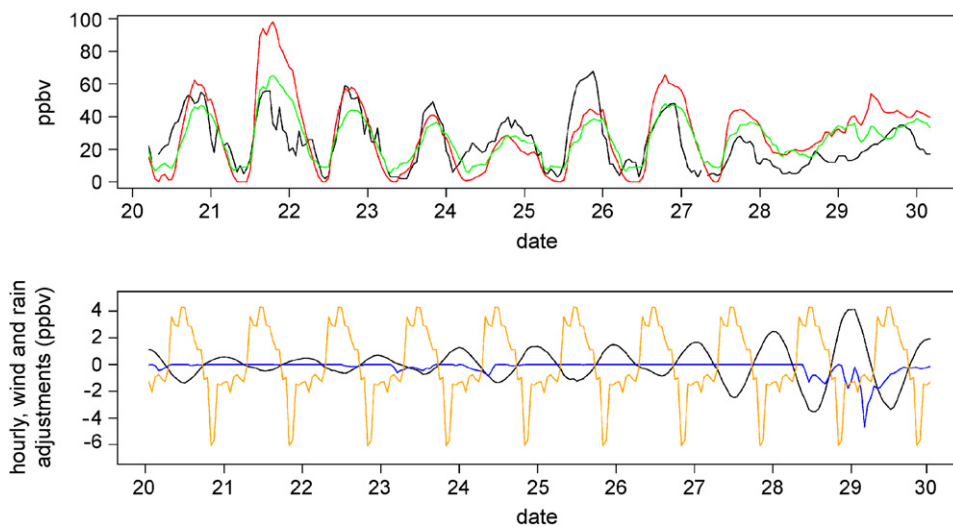


Fig. 5. Station 6. August 20 (05:00)–August 30 (04:00). Upper panel: Observations (black), RAQAST predictions (red), corrected RAQAST predictions (green). Lower panel: Daily cycle (orange), wind speed (black) and precipitation (blue) effects in step 2. *Note:* Different scales.

better fit to observations on August 21, 26 and 27. The precipitation amounts are close to zero almost all the time, except on August 29. During that day the precipitation index is positive, with a corresponding coefficient of -9.61 (3.12). The precipitation index decreases the values of RAQAST outputs, which improves the predictions. Similar results are obtained for the other stations.

5. Conclusions and perspectives

MDC improvements of RAQAST outputs improve forecast accuracy compared with the original RAQAST model. In particular, the RAQAST model sometimes overestimates peak ozone concentrations during the daytime and generally underestimates night time ozone concentrations compared with EPA observations in Atlanta. The

statistical adjustments of RAQAST with wind speed and precipitation generally reduce the overestimation/underestimation of the surface ozone in RAQAST. In the afternoon, when wind or precipitation is present, RAQAST should provide lower forecasts. The daily cycle in RAQAST is still not yet perfectly represented, with systematic overestimation/underestimation depending on the hour of the day.

Hourly RMSEs of RAQAST predictions (15–21 ppbv) over the 10-day evaluation period are much larger than RMSEs of 8-h daily maxima (2–3 ppbv). Peaks of 2–3 h can be very difficult to forecast precisely. However, averages of eight consecutive hours are easier to predict. A potential continuation of our work would focus on days with high concentrations only. An adjustment based on this training sample (which might include several years of data) could considerably improve the prediction of pollution episodes. Longer validation periods would also be necessary to confirm such results.

The statistical correction of RAQAST model outputs sometimes makes the discrepancies of high-peaked ozone from the model and observation larger. Further study needs to be performed to find other meteorological variables to improve the forecasts of the maximum ozone peaks which are important in the public health warning system.

It would be interesting separate the errors due to a lack of chemical and physical representations from the errors due to the large resolution. Running a high resolution model would help examine the source of the errors. The computer model is extremely time consuming at a high resolution so this experiment has to be carried out carefully. In particular, misalignments for local weather features ought to be modeled.

Choosing a grid cell corresponding to a particular station with more advance ideas, such as distance or wind direction, is an interesting direction. We were able to carry out some experiment where we chose alternatively the three grid cells, or jointly the three of them in the regression (step 1 of MDC). The results give residual mean square errors that are extremely close to the ones using the grid cell which includes the station. The inclusion of the three grid cells naturally weights them according to the distance, but the fit is only slightly better—with RMSE lower by roughly 1.5%—with a risk of overfitting. The natural extension would be to consider weather patterns, as wind direction. How-

ever, the effective sample size will be reduced, especially for rare weather patterns. With a larger data set, we suggest to use empirical orthogonal functions to carry out statistical downscaling. Benestad (2002) presented empirically downscaled temperature scenarios for a number of sites in northern Europe, and used a method involving common EOFs which account for several meteorological variables at once. Also, Spak et al. (2007) compared the dynamical downscaling for temperatures with statistical regression based on EOFs and concluded that they have similar skills.

In order to explore structural deficiencies in RAQAST, we need to calibrate the model by choosing the best tuning parameters. These parameters are inputs in the computer model such as boundary conditions or unknown constants in chemical or transport equations. Statistical calibration of complex computer models is a difficult and computationally intensive task (Bayarri et al., 2002). We expect better forecasts after calibration of RAQAST. Nevertheless, downscaling and correction of the outputs, after calibration, will still help give even more precise forecasts.

A fruitful extension of this work would include spatial modeling. It would be interesting to analyze the deficiencies in terms of spatial distributions, thus enabling the modelers to carry out some meaningful adjustments, say if the patterns of deficiencies match some meteorological conditions. In particular, we need to compare urban, suburban and rural sites and see if there are any systematic trends. Also, downscaling at every location in space, and not only where monitors are situated, is of great interest. We are currently investigating space–time approaches for this purpose, either with space–time covariance models or with a Bayesian space–time approach. Both of these methods will yield measures of spatial uncertainties as well.

References

- Aneiros-Perez, G., Cardot, H., Estevez-Perez, G., Vieu, P., 2004. Maximum ozone concentration forecasting by functional non-parametric approaches. *Environmetrics* 15, 675–685.
- Bayarri, M., Berger, J., Higdon, D., Kennedy, M., Kottas, A., Paulo, R., Sacks, J., Cafeo, J., Cavendish, J., Lin, C., Tu, J., 2002. A framework for validation of computer models. Technical Report 128, National Institute of Statistical Sciences.
- Benestad, R.E., 2002. Empirically downscaled multimodel ensemble temperature and precipitation scenarios for Norway. *Journal of Climate* 15 (21), 3008–3027.

- Benkovitz, C.M., Schultz, M.T., Pacyna, J., Tarrason, L., Dignon, J., Voldner, E.C., Spiro, P.A., Logan, J.A., Graedel, T.E., 1996. Global gridded inventories for anthropogenic emissions of sulfur and nitrogen. *Journal of Geophysical Research* 101, 29239–29253.
- Bey, I., Jacob, D.J., Yantosca, R.M., Logan, J.A., Field, B., Fiore, A.M., Li, Q., Liu, H., Mickley, L.J., Schultz, M., 2001. Global modeling of tropospheric chemistry with assimilated meteorology: model description and evaluation. *Journal of Geophysical Research* 106, 23,073–23,096.
- Black, T.L., 1994. The new NMC mesoscale ETA-model—description and forecast examples. *Weather and Forecasting* 9, 265–278.
- Bonan, G.B., Levis, S., Kergoat, L., et al., 2002. Landscapes as patches of plant functional types: an integrating concept for climate and ecosystem models. *Global Biogeochemical Cycles* 16 (2), in press, doi:10.1029/2005GL025080.
- Choi, Y., Wang, Y., Zeng, T., Martin, R., Kurosu, T., Chance, K., 2005. Evidence of lightning NO_x and convective transport of pollutants in satellite observations over North America. *Geophysical Research Letters* 32, L02805.
- Damon, J., Guillas, S., 2002. The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics* 13, 759–774.
- Geleyn, J.F., 1981. Some diagnostics of the cloud/radiation interaction in the ECMWF forecast model. In: *Proceedings of Workshop on Radiation and Cloud–Radiation Interaction in Numerical Modeling*, pp. 135–162.
- Grell, G.A., 1993. Prognostic evaluation of assumptions used by cumulus parameterization. *Monthly Weather Review* 121, 764–787.
- Guillas, S., Tiao, G.C., Wuebbles, D.J., Zubrow, A., 2006. Statistical diagnostic and correction of a chemistry-transport model for the prediction of total column ozone. *Atmospheric Chemistry and Physics* 6, 525–537.
- Herman, J.R., Celarier, E.A., 1997. Earth surface reflectivity climatology at 340–380 nm from TOMS data. *Journal of Geophysical Research* 102, 28003–28011.
- Jacobson, M.Z., Turco, R.P., 1994. A sparse-matrix, vectorized GEAR code for atmospheric transport models. *Atmospheric Environment* 33, 273–284.
- Jing, P., Cunnold, D., Choi, Y., Wang, Y., 2006. Summertime tropospheric ozone columns from aura OMI/MLS measurements versus regional model results over the United States. *Geophysical Research Letters* 33, L17817.
- Liu, H., Jacob, D.J., Bey, I., Yantosca, R.M., 2001. Constraints from ²¹⁰Pb and ⁷Be on wet deposition and transport in a global three-dimensional chemical transport model driven by assimilated meteorological fields. *Journal of Geophysical Research* 106 (D11), 12109–12128.
- McHenry, J.N., Ryan, W.F., Seaman, N.L., Coats, C.J., Pudykiewicz, J., Arunachalam, S., Vukovich, J.M., 2004. A real-time eulerian photochemical model forecast system—Overview and initial ozone forecast performance in the Northeast US orridor, *Bulletin of the American Meteorological Society*, 85 (4), 525–548.
- McKeen, S., Hsie, E., Trainer, M., Tallamraju, R., Liu, S., 1991. A regional model study of the ozone budget in the Eastern United States. *Journal of Geophysical Research* 96, 10809–10846.
- NRC, 1991. Rethinking the ozone problem in urban and regional air pollution. Technical Report, National Academy, Washington, DC.
- Otte, T.L., Pouliot, G., Pleim, J.E., Young, J.O., Schere, K.L., Wong, D.C., Lee, P.C.S., Tsidulko, M., McQueen, J.T., Davidson, P., Mathur, R., Chuang, H.-Y., DiMego, G., Seaman, N.L., 2005. Linking the Eta model with the community multiscale air quality (CMAQ) modeling system to build a national air quality forecasting system. Technical Report, NCEP NOTES.
- Ryan, W.F., Pietry, C.A., 2000. Air quality forecasts in the mid-Atlantic region: current practice and benchmark skill. *Weather and Forecasting* 15, 46–60.
- Schlink, U., Dorling, S., Pelikan, E., Nunnari, G., Cawley, G., Junninen, H., Greig, A., Foxall, R., Eben, K., Chatterton, T., Vondracek, J., Richter, M., Dostal, M., Bertuccio, L., Kolehmainen, M., Doyle, M., 2003. A rigorous inter-comparison of ground-level ozone predictions. *Atmospheric Environment* 37, 3237–3253.
- Spak, S., Holloway, T., Lynn, B., Goldberg, R., 2007. A comparison of statistical and dynamical downscaling for surface temperature in North America. *Journal of Geophysical Research—Atmospheres* 112.
- Stauffer, D.R., Seaman, N., Binkowski, F.S., 1991. Use of four-dimensional data assimilation in a limited-area mesoscale model part II. Effects of data assimilation within the planetary boundary layer. *Monthly Weather Review* 119, 734–754.
- Stephens, G., Paltridge, G., Platt, C., 1978. Radiation profiles in extended water cloud III: observations. *Journal of the Atmospheric Sciences* 35, 2133–2141.
- Stewart, G.W., 1987. Collinearity and least squares regression (with discussion). *Statistical Science* 2, 68–100.
- Vaughan, J., Lamb, B., Frei, C., Wilson, R., Bowman, C., Kaminsky, C.F., Otterson, S., Boyer, M., Mass, C., Albright, M., Koenig, J., Collingwood, A., Gilroy, M., Maykut, N., 2004. A numerical daily air quality forecast system for the pacific northwest. *Bulletin of the American Meteorological Society*, 549–561.
- Walcek, C.J., 2000. Minor flux adjustment near mixing ratio extremes for simplified yet highly accurate monotonic calculation of tracer advection. *Journal of Geophysical Research* 105, 9335–9348.
- Wang, Y., Choi, Y., Zengand, T., Ridleyand, B., Blake, N., Blakeand, D., Flocke, F., 2006. Late-Spring increase of trans-pacific pollution transport in the upper troposphere. *Geophysical Research Letters* 33.
- Wild, O., Zhu, Q., Prather, M.J., 2000. Fast-J: accurate simulation of in- and below-cloud photolysis in global chemical models. *Journal of Atmospheric Chemistry* 37, 245–282.
- Zeng, T., Wang, Y., Chance, K., Browell, E.V., Ridley, B.A., Atlas, E.L., 2003. Widespread persistent near-surface ozone depletion at northern high latitudes in Spring. *Geophysical Research Letters* 30, 2298.
- Zeng, T., Wang, Y., Chance, K., Blake, N., Blake, D., Ridley, B., 2006. Halogen-driven low altitude O₃ and hydrocarbon losses in Spring at Northern high latitudes. *Journal of Geophysical Research*, submitted for publication, 111, D17313, doi:10.1029/2005JD006706.