LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-JRNL-218640

Application of dynamic linear regression to improve the skill of ensemble-based deterministic ozone forecasts

M. O. Pagowski, G. A. Grell, D. Devenyi, S. E. Peckham, S. A. McKeen, W. Gong, L. Delle Monache, J. N. McHenry, J. McQueen, P. Lee

February 3, 2006

Atmospheric Environment

Disclaimer

Application of dynamic linear regression to improve the skill of ensemble-based deterministic ozone forecasts

M. Pagowski*

Global Systems Division, Earth System Research Laboratory, R/GSD1,

325 Broadway, Boulder, Colorado 80305-3328, USA

**[In collaboration with the Cooperative Institute for Research in the Atmosphere (CIRA),*

Colorado State University, Fort Collins, Colorado]

G. A. Grell, D. Devenyi, S. E. Peckham

Cooperative Institute for Research in Environmental Sciences (CIRES),

University of Colorado,

Global Systems Division, Earth System Research Laboratory, R/GSD1,

325 Broadway, Boulder, Colorado 80305-3328, USA

---

* *Corresponding author.* Fax: (303) 497-7262.

E-mail address: Mariusz.Pagowski@noaa.gov

S. A. McKeen

Cooperative Institute for Research in Environmental Sciences (CIRES),

University of Colorado,

Chemical Sciences Division, Earth System Research Laboratory,

325 Broadway, Boulder, Colorado 80305-3328, USA

W. Gong

Meteorological Service of Canada, Downsview, Ontario, Canada

L. Delle Monache

Atmospheric Science Programme, Earth Ocean Sciences Dept.,

University of British Columbia, 6339 Stores Road, Vancouver, BC V6T1Z4, Canada

J. N. McHenry

Baron Advanced Meteorological Systems, c/o North Carolina State University, Raleigh,

North Carolina, USA

J. McQueen, and P. Lee

NWS/NCEP/NOAA, Camp Springs, Maryland, USA

Manuscript Submitted to Atmospheric Environment

Abstract

Forecasts from seven air quality models and surface ozone data collected over the eastern USA and southern Canada during July and August 2004 provide a unique opportunity to assess benefits of ensemble-based ozone forecasting and devise methods to improve ozone forecasts. In this investigation, past forecasts from the ensemble of models and hourly surface ozone measurements at over 350 sites are used to issue deterministic 24-h forecasts using a method based on dynamic linear regression. Forecasts of hourly ozone concentrations as well as maximum daily 8-h and 1-h averaged concentrations are considered. It is shown that the forecasts issued with the application of this method have reduced bias and root mean square error and better overall performance scores than any of the ensemble members and the ensemble average. Performance of the method is similar to another method based on linear regression described previously by Pagowski et al., but unlike the latter, the current method does not require measurements from multiple monitors since it operates on individual time series. Improvement in the forecasts can be easily implemented and requires minimal computational cost.

Keywords: Photochemical air quality; Modeling; Ozone; Ensemble forecast

1. Introduction

Future behavior of nonlinear dynamical systems, such as the atmosphere, is crucially dependent on the initial state and forcings, and for this reason, its prediction with numerical models is inherently uncertain due to the errors in initial and boundary conditions, physical parameterizations, and numerics (e.g. Kalnay, 2003). To account for the range of possible solutions using ensembles in weather forecasting is a common practice at meteorological centers around the world. However, this practice is rare in air quality modeling despite the fact that the rationale for using ensembles in air quality forecasting is at least as strong as for weather forecasting. In addition to inaccuracies in reproducing physical states of the atmosphere provided by atmospheric models, uncertainty is enhanced by the complexity of chemical reactions and equations which attempt to describe them, and the questionable quality of emission inventories (Russell and Dennis, 2000).

Commonly, ensemble forecasts are used to provide probabilistic distribution of the future scenario (probabilistic forecasts). They can also be used to provide best estimates of the future state of the atmosphere (deterministic forecasts). The latter approach is of sole interest in this paper.

In meteorology, Krishnamurti et al. (1999) used multiple linear regression to improve forecasts by specifying weights for members of ensemble of models (superensemble). Hou et al. (2001) observed that averaged ensemble short-term forecasts generally verify better against observations than any ensemble member. Shin and Krishnamurti (2003) demonstrated that the dynamic linear regression (DLR, hereafter) provides superior forecasts of precipitation compared with the regular regression. In air quality, Vautard et al. (2001) and Delle Monache and Stull (2003) adopted the idea of ensemble-based air quality forecasting to predict ozone over Europe, employing different atmospheric and air quality models, chemical mechanisms, and initial as well as lateral boundary conditions. Ozone forecasts obtained by ensemble averaging were shown to perform better than any single ensemble member. Recently, Pagowski et al. (2005), using ensemble forecasts of surface ozone over the eastern U.S.A. and southern Canada, showed that overall statistics of the ensemble forecasts can be improved compared to averaging with the application of linear regression, while McKeen et al. (2005) discussed advantages of bias removal methods for the same data. Also, Delle Monache et al. (2005a and b) applied the Kalman filter to ensemble members and ensemble average to reduce bias of the ozone forecasts over the Southwest British Columbia, Canada.

During summer 2004, seven air quality models participated in the International Consortium for Atmospheric Research on Transport and Transformation/New England Air Quality Study (ICARTT/NEAQS) conducted over the eastern U.S.A. and southern Canada. The models included AURAMS (A Unified Regional Air-quality Modeling System), BAMS (Baron Advanced Meteorological System) MAQSIP (Multiscale Air Quality Simulation Platform) at 45-km and 15-km resolution, CHRONOS (Canadian Hemispheric and Regional Ozone and NOx System), CMAQ/ETA (Community Multiscale Air Quality Model/ETA), STEM-2K3 (Sulphur Transport and Emissions Model – 2003), and WRF/Chem (Weather Research and Forecast model/Chemistry version). These models provided daily forecasts of ozone from 6 July to 30 August 2004. Hourly averaged ozone surface ozone concentration at multiple locations in the eastern U.S.A. and Canada are stored in the Aerometric Information Retrieval Now (AIRNow) database. From this database, measurements at over 350 locations could be used for verification of models.

The ozone observations, their processing for verification, and models are described in section 2. Subsequently, a method to obtain a deterministic forecast with DLR is described with details given in the appendix. Next, results of the application method and its verification and comparison with the performance of single models and the ensemble mean are discussed. Summary and conclusions are provided in the final section.

2. Description of data

a. Observations

Surface ozone data used in this study comprises hourly measurements at over 350 sites from 0000 UTC 6 July to 0000 UTC 30 August 2004 (56 days). Figure 1 shows locations of the sites, the AIRNow site classification, and outline of the domain of model overlap, as well as the latitude/longitude belt for sites used in plotting Fig. 2. Figure 2 shows the temporal variability of the maximum daily 8-h averaged ozone concentrations for the sites confined within the latitude/longitude belt in Fig. 1. It can be seen that summer 2004 was characterized by few high ozone episodes, a result of cool and rainy weather. Out of 16480 observations, only 87 exceeded the EPA mandated 85 ppbv threshold for the maximum daily 8-h averaged ozone concentration, and there were no observations above the 125 ppbv threshold for the maximum daily 1-h averaged ozone concentration. The small number of observations of the elevated ozone during this summer limits our ability to assess performance of the models and the proposed method with respect to the above thresholds.

b. Models

Below, a brief description of the models and data processing for evaluation is given. Detailed information is available in McKeen et al. (2005).

AURAMS (Moran et al., 1998) is a Canadian off-line air quality model with gas-phase chemistry based on the ADOM II mechanism and size- and chemically-resolved representation of aerosol microphysics and gas-aerosol interaction (Gong et al., 2003). Meteorological fields are supplied by the Global Environmental Model (GEM, Côtè et al., 1998a, b), which is the operational weather forecasting model of the Meteorological Service of Canada. National inventories for Canada for 1990 and national inventories for U.S.A. for 1990 projected to 1995 were processed with the Canadian Emission Processing System (CEPS, Moran et al., 1997). The horizontal grid resolution of this model is equal to 42 km.

MAQSIP (McHenry et al., 2004) is a chemical model that uses the modified Carbon-Bond 4 chemistry mechanism (Gery et al., 1989) with additions and modifications by DeMore et al. (1994), Chang et al. (1996), Carter (1996), Kasibhatla et al. (1997), and McHenry and Coats (2003). It is driven off-line by the Penn State/NCAR MM5 mesoscale model (Grell et al., 1994). Emissions are provided by Sparse Matrix Operator Kernel Emissions system (SMOKE, Coats, 1996). MAQSIP provided forecasts at 45-km and 15-km resolutions.

CHRONOS (Pudykiewicz et al., 1997) is a Canadian operational air quality model with gas-phase chemistry very similar to AURAMS, and aerosol chemistry based on bulk aerosol representation for PM2.5 (primary and secondary) and coarse PM (primary only). Meteorological input comes from GEM, and emissions inventories processed similarly as for AURAMS. CHRONOS provided forecasts at 21-km resolution.

CMAQ (Byun and Ching, 1999) in the current configuration uses the Carbon-Bond 4 gas-phase chemistry mechanism and tri-modal size distribution for aerosol chemistry (Binkowski and Shankar, 1995). The chemical model is driven off-line by NWS/NCEP Eta forecasts. Emissions are supplied by SMOKE. CMAQ was run at 12-km resolution.

STEM-2K3 (Carmichael et al., 2003) employs the SAPRC-99 gas-phase chemistry mechanism (Carter, 2000) with detailed treatment of aerosols (Tang et al., 2004). Meteorological fields are supplied by the MM5 model in the off-line mode. Emissions are processed using EPA NEI-99 (EPA, 2004) inventory and the IGAC-GEIA archive (Guenther et al., 1995). Resolution of the model was equal to 12 km.

WRF/Chem (Grell et al., 2005) is an on-line air quality model in which chemistry is fully coupled with meteorology, and concentrations of gaseous and aerosol species are calculated simultaneously with physics. Gas-phase chemistry is based on the RADM2 mechanism (Stockwell et al., 1995) and parameterization of aerosols is based on Binkowski and Shankar (1995), Ackermann et al. (1998), and Schell et al. (2001). As in the STEM-2K3, EPA NEI-99 (EPA, 2004) emissions inventories were used. Model was run at 27-km resolution. Results from this model differ from those previously analyzed by Pagowski et al. (2005) and McKeen et al. (2005) in that the current simulations were performed with an improved emission inventory.

In the current study, only 24-h forecasts of the chemical models issued at 0000 UTC (0600 UTC for CMAQ) were used. No interpolation was applied for verification, but model values were matched with the site measurements located within the grid cell. For consistency with observations, results from models output at the top of the hour were averaged temporally.

It can be noted that the models use different emission inventories and employ different emission modeling systems. Also, horizontal resolution varies from 12 km to 42 km, and the vertical resolution of the models is quite different. Since the focus of this paper is on the application of the technique and not on model evaluation, individual models are simply labeled with random capital letters.

3. Description of the method

In contrast to static linear models, which assume that properties of investigated processes do not change in time, dynamic linear models allow for temporal evolution of the characteristics of these processes. Following West and Harrison (1989), the dynamic linear model is defined as

$$Y(t) = \mathbf{F}(t)\mathbf{\Theta}(t) + \nu(t), \qquad \nu(t) \sim \mathcal{N}(0, V(t)) \,, \tag{1a}$$

$$\mathbf{\Theta}(t) = \mathbf{\Theta}(t-1) + \mathbf{w}(t), \qquad \mathbf{w(t)} \sim \mathcal{N}(\mathbf{0}, \mathbf{W}(t)) \,, \tag{1b}$$

where Y is an observation, vector $\mathbf{F}$ is a regressor, and $\mathbf{\Theta}$ is a state vector. It is assumed that observation ($\nu(t)$) and model ($\mathbf{w(t)}$) errors are uncorrelated, and have normal distribution with zero mean and variances V and $\mathbf{W}$, respectively. In the case of the ensemble, vector $\mathbf{F}$ contains ensemble member forecasts, and vector $\mathbf{\Theta}$ contains weights for the ensemble members. West and Harrisson (1989, pp. 60–71, pp. 117–121) show that if the observational error variance is constant but unknown, estimate of its inverse has a Gamma distribution and can be found recursively. Error variance of the weights ($\mathbf{W}$) may be assigned based on the discount factor approach in which the initial assumed variance is multiplied by a constant factor at each forecast time (*ibid*, pp. 57–60). Shin and Krishnamurti (2003) applied this algorithm to predict precipitation with several atmospheric models. Here, we follow the same method. Further details on the algorithm are given in the appendix.

The adaptive nature of the algorithm, which requires very little training, presents an alternative to Model Output Statistics (MOS), where a longer time series (not available for ensemble ozone forecasts) is necessary to establish regression coefficients.

4. Application and performance of the method

To obtain the deterministic ozone forecast at time t+24h, the DLR is used to calculate weights which are subseqently applied to ensemble members' forecasts at time t and then using observations and past forecasts at times $t-24$h, $t-48$h, $\ldots$, $mod(t, 24)$h. If a forecast of any of the ensemble members or an observation at a site is not available at time t weights computed for the latest of the previous times, $t - k \times 24$h are used to issue a deterministic forecast. Alternatively, a forecast can be issued by applying DLR to a smaller ensemble of available members. The above procedure is repeated for each measurement site. A typical filter performance is illustrated in Fig. 3a, showing a 5-day time series of observations and 24-h forecasts with DLR and ensemble average (AVE) at a selected location. In Fig. 3b the variability of weights during this period is shown. It can be noted that the forecast with the equal weights has a systematic positive bias, which is largely removed by the application of the weights. Analysis of this and other time series shows that the magnitude of the weights varies strongly during the day, but much weaker on a day-to-day basis. This observation justify our preference to use in the DLR only the past forecasts at the same time of a day, rather than all the available past forecasts.

Overall statistics of the forecasts for the DLR ensemble, AVE ensemble, and individual models are given in Table 1. Bias, root mean square error ($RMSE$), and correlation

($Corr$) were calculated for time series of observations and models at every monitor location and averaged over all the sites (in calculating $RMSE$, MSE was averaged over all the sites and subsequently its root square was calculated to obtain the $RMSE$). Confidence intervals at 95% for $RMSE$ and $Corr$ for DLR are (10.55, 10.60) and (0.809, 0.812), respectively, and for AVE are (17.96, 18.06) and (0.764, 0.767), respectively.

It is noteworthy that the DLR forecasts evaluate better against the observations than any model or ensemble average. Especially significant is the decrease in bias and RMSE, while the improvement in correlation is more modest. As pointed out previously for this data, the ensemble average generally has better correlation than the individual models, but its bias and RMSE can suffer from large errors of some ensemble members (McKeen et al., 2005). Application of the DLR eliminates this negative effect of the individual model biases.

The above forecasts were used to calculate daily maxima of 8-h and 1-h averaged ozone concentrations and the corresponding model performance scores. To assess the skill of the models, persistence, understood as the matching maximum at the same location on the previous day, was also included. Scatterplots for the DLR ensemble, ensemble average, and persistence for daily maxima of 8-h and 1-h averaged ozone concentrations are shown in Fig. 4. In both cases the plots have similar characteristics. DLR has the best alignment along $y = x$ line but it underestimates observed high ozone concentrations for the 1-h maximum. The ensemble average has stronger bias for the whole range of concentration values, somewhat decreased for the elevated ozone concentrations. It is apparent that the ensemble average does not reflect the amplitude of the observed maximum ozone variation. Persistence is very dispersive, suggesting that it is a poor forecast and is clearly deficient compared with either of the two previous models. Bias, $RMSE$, and $Corr$ for the daily maxima were also calculated for DLR and AVE ensembles. In addition, these statistics were calculated for results obtained with a method described by Pagowski et al. (2005) which is also based on linear regression (labeled SVD for Singular Value Decomposition, a method to solve linear system of equations, hereafter). The results are given in Table 2. Confidence intervals at 95% for the daily maximum of 8-haveraged ozone concentrations for $RMSE$ and $Corr$ for DLR are (9.70, 9.91) and (0.674, 0.691), respectively, for AVE are (13.69, 13.99) and (0.708, 0.723), respectively, and for SVD are (10.00, 10.22) and (0.717, 0.731), respectively. Corresponding 95% confidence intervals for the same models for the daily maximum of 1-h averaged ozone concentrations are for $RMSE$ (10.89, 11.13), (13.37, 13.67), (11.62, 11.88), and for $Corr$ (0.707, 0.722), (0.703, 0.718), and (0.696, 0.712), respectively. It can be seen that DLR and SVD methods have similar performance. While bias and $RMSE$ are substantially reduced for both daily maxima, correlations for all the methods differ little. An advantage of the DLR method over SVD is that the latter is only applicable when observations are available from a large number of monitoring sites so that the system of linear equations remains significantly overdetermined (Pagowski et al., 2005) while the DLR method can be applied even to a time series from a single monitoring site after a very short (one-day or so) training period.

Categorical forecasts consist of a statement that a certain event will happen or not, such as that ozone concentration at a given location will be higher than a certain threshold. They are very useful in air quality for informing the public about the predicted health risks associated with atmospheric pollution by associating it with colors that represent health-exposure risk (McHenry et al., 2004). Categorical forecasts form a basis for a contingency table, given in Table 3, and are used to define performance measures such as bias ratio (BR, Wilks, 1995)

$$BR \equiv \frac{a+b}{a+c} \ , \tag{2}$$

and equitable threat score (ETS, also called Gilbert's score, Schaefer, 1990)

$$ETS \equiv \frac{a-ch}{a+b+c-ch} \ , \tag{3}$$

where ch is given by

$$ch \equiv \frac{(a+b)(a+c)}{a+b+c+d} \ ,$$

where a, b, c, and d stand for a number of events as defined in Table 2. BR is a ratio of the area forecast to exceed a threshold to the area observed to exceed the threshold. ETS is a ratio of the area correctly forecast to exceed a threshold to the sum of the area observed, and the area incorrectly forecast to exceed the threshold. Random chance ch is subtracted from these areas. Optimal value for both BR and ETS is 1. Negative values of ETS signal no forecast skill. These two measures are plotted in Fig. 5 for four thresholds equal to 30, 50, 70 and 85 ppbv for the maximum daily 8-h and 1-h averaged ozone concentration. Unfortunately, skill of the forecasts cannot be adequately assessed for higher thresholds, since elevated ozone was rarely observed.

Assessment of the skill of forecasts should account for the fact that larger BR tends to increase ETS (Hamill, 1999). As one would expect, the forecast area of persistence is very close to the observed ($BR \sim 1$). These forecasts are, nevertheless, poor, as is evident from the scatterplots in Fig. 4 and low ETS's in Fig. 5. An analysis of Fig. 5 reveals that for the daily maximum of 8-h averaged ozone concentration, the ensemble average has very large BR's for the higher thresholds. Even despite that, its ETSs are lower than those for the DLR ensemble. For the 30-ppbv threshold, its skill is even smaller than for the persistence forecasts. The DLR ensemble is clearly superior to either of the other two. The assessment of the relative performance of ensembles is more complicated for the daily maximum of 1-h averaged ozone concentration. Here, slightly smaller, but still large, values of BR for the ensemble average can be seen in Fig. 5. Compared to the DLR ensemble, its ETSs are lower for the two lower thresholds and higher for the higher thresholds, undoubtedly due to the effect of increased BR on ETS. BR and ETSs of the DLR ensemble are comparable to the same measures for the daily maximum of 8-h averaged ozone concentration except for the highest threshold, where the elevated ozone is clearly underpredicted. It is possible that due to the rarity of high ozone episodes in summer 2004, no sufficient training data for DLR existed to calculate weights adequate for short-term elevated ozone concentrations.

5. Summary and conclusions

A method based on dynamic regression fitting of 24-h forecasts from an ensemble of air quality models to observations was applied to predict surface ozone concentrations. The resulting forecasts had significantly smaller bias and RMSE, in comparison with the same statistics computed for the ensemble average or single models. Performance of the method is similar to that described by Pagowski et al. (2005). Advantage of the DLR method over the latter is that the latter method is only applicable when observations are available from a large number of monitoring sites so that the system of linear equations remains significantly overdetermined (Pagowski et al., 2005) while the DLR method can be applied even to a time series from a single monitoring site after a short training period.

Also, with the application of the DLR method, the equitable threat score of the 24-h forecasts calculated for the maximum daily 8-h and 1-h averaged ozone concentration had consistently higher values over a range of thresholds. The most appealing feature of the method is its fast adaptability, giving it an advantage over static regression, which requires a much larger data series for training.

While deterministic forecasts based on an ensemble of models are issued to provide the best estimate of a future state, they inevitably result in a loss of information by reducing a probabilistic solution to a categorical one. The advantage, in terms of an economic value of certain probabilistic over deterministic forecasts in meteorology, was shown by Zhu et al. (2002). A future study to address the economic benefits of probabilistic forecasts in air quality using the current dataset is being considered.

APPENDIX

Dynamic linear regression algorithm

Below an algorithm from West and Harrison (1989) is given with some rudimentary explanations without proofs which can be found in the original work.

It can be shown that for a process defined in Eq. 1, when the observational error variance (V) is constant, but unknown, its inverse ($\phi = V^{-1}$) has a Gamma distribution. Furthermore, the error variance of the state vector has Student T distribution with zero mean and variable variance. Below the new formulation of the process is given.

$$Y(t) = \mathbf{F}(t)\boldsymbol{\Theta}(t) + \nu(t), \qquad \nu(t) \sim \mathcal{N}(0, V)\ , \tag{A.1a}$$

$$\boldsymbol{\Theta}(t) = \boldsymbol{\Theta}(t-1) + \mathbf{w}(t), \qquad \mathbf{w(t)} \sim \mathbf{T}(\mathbf{0}, \mathbf{W}(t))\ . \tag{A.1b}$$

With the model specified as above, the following distributional results are valid at each time t:

$$(\boldsymbol{\Theta}(t-1) \mid D(t-1)) \sim \mathbf{T}_{n(t-1)}\left[\mathbf{m}(t-1), \mathbf{C(t-1)}\right]\ , \tag{A.2a}$$

$$(\mathbf{\Theta}(t) \mid D(t-1)) \sim \mathbf{T}_{n(t-1)}\left[\mathbf{m}(t), \mathbf{R}(t)\right] \ , \tag{A.2b}$$

$$(\phi \mid D(t-1)) \sim \mathbf{G}\left[n(t-1)/2, d(t-1)/2\right] \ , \tag{A.2c}$$

$$(Y(t) \mid D(t-1)) \sim \mathbf{T}_{n(t-1)}\left[f(t), Q(t)\right] \ , \tag{A.2d}$$

where the probablity is conditional on the existing information about the time series up to and including $t-1$ ($D(t-1)$ stands for the information available at time $t-1$), $\mathbf{m}$ is the mean value of the state vector used to issue forecast $f = \mathbf{Fm}$, n is the time counter, parameter d is given below, and $\mathbf{C}$, $\mathbf{R}$ and $\mathbf{Q}$ are state vector and observational variances about the mean, respectively. Updated state vector variance $\mathbf{C}$ is calculated from a recurrence relationship given below, while variances $\mathbf{R}$ and $\mathbf{Q}$ are calculated from

$$\mathbf{R}(\mathbf{t}) = \mathbf{C}(t-1) + \mathbf{W}(t) \ , \tag{A.3a}$$

$$\mathbf{Q}(\mathbf{t}) = \mathbf{F}(t)\mathbf{R}(t)\mathbf{F}^{\mathbf{T}}(t) + S(t-1) \ , \tag{A.3b}$$

with S being a prior estimate of the variance V. The following recurrence relationships provide the method to update variance and the mean of the state vector and to issue the forecast f:

$$(\mathbf{\Theta}(t) \mid D(t)) \sim \mathbf{T}_{n(t)}\left[\mathbf{m}(t), \mathbf{C}(\mathbf{t})\right] \ , \tag{A.4a}$$

$$(\phi \mid D(t)) \sim \mathbf{G}\left[n(t)/2, d(t)/2\right] \ , \tag{A.4b}$$

where

$$\mathbf{m}(t) = \mathbf{m}(t-1) + \mathbf{A}(t)e(t) \ , \tag{A.5a}$$

$$\mathbf{C}(t) = (S(t)/S(t-1))\left[\mathbf{R}(t) - \mathbf{m}(t)\mathbf{m}^{\mathbf{T}}(t)Q(t)\right] \ , \tag{A.5b}$$

$$n(t) = n(t-1) + 1 \ , \tag{A.5c}$$

$$d(t) = d(t-1) + S(t-1)e^2(t)/Q(t) \ , \tag{A.5d}$$

$$S(t) = d(t)/n(t) \ , \tag{A.5e}$$

$$e(t) = Y(t) - f(t) \ , \tag{A.5f}$$

$$\mathbf{A}(t) = \mathbf{R}(t)\mathbf{F}(t)/Q(t) \ . \tag{A.5g}$$

The only remaining unknown in the above recursion is the process error variance $\mathbf{W}$. It is calculated assuming a constant rate of increase of uncertainty of the state vector variance $\mathbf{R}$ by assigning

$$\mathbf{R}(t) = \mathbf{C}(t-1)/\delta \ , \tag{A.6a}$$

and from Eq. A.3a

$$\mathbf{W}(t) = \mathbf{C}(t-1)(1-\delta)/\delta \ , \tag{A.6b}$$

where δ is a discount factor and $0 < \delta \leq 1$.

The evolution of the process is little dependent on the initial estimates of the observational error variance S, state vector variance $\mathbf{C}$, and the state vector mean $\mathbf{m}$ which need to be specified on entry.

Acknowledgements. This research was funded by Early Start Funding from NOAA/NWS Office of Science and Technology and would not be possible without the participation of the AIRNow program and participating stakeholders. Credit for program and management is given to Paula Davidson (NOAA/NWS/OST), Steve Fine (NOAA/ARL), and Jim Meagher (NOAA/AL). The authors thank V. Bouchet, R. Moffet (from Canadian Meteorological Centre), and Y. Tang (from University of Iowa) for providing model data, Dr. Yuanfu Xie for his comments on the manuscript and Nita Fullerton and Susan Carsten for editorial review.

This work was performed under the auspices of the U. S. Department of Energy by University of California, Lawrence Livermore National Laboratory under contract W-7405-Eng-48.

References

Ackermann, I.J., Hass, H., Memmesheimer, M., Ebel, A., Binkowski, F.S., and Shankar, U., 1998. Modal aerosol dynamics model for Europe: Development and first applications. Atmospheric Environment, 32, 2981-2999.

Binkowski, F.S., and Shankar, U., 1995. The regional particulate model. I. Model description and preliminary results. Journal of Geophysical Research, 100, D12, 26191–26209.

Byun, D.W., and Ching, J.K.S. (editors), 1999. Science Algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) Modeling System, EPA-600/R-99/030, Office of Research and Development, U.S. EPA, Washington, D.C..

Carmichael, G. R., et al., 2003. Regional-scale chemical transport modeling in support of intensive field experiments: Overview and analysis of the TRACE-P observations. Journal of Geophysical Research, 10.1029/2002JD003117.

Carter, W. P. L., 1996. Condensed atmospheric photooxidation mechanism for isoprene. Atmospheric Environment, 30(24), 4275-4290.

Carter, W., 2000. Documentation of the SAPRC-99 chemical mechanism for VOC reactivity assessment. Final Report to California Air Resources Board Contract No. 92–329, University of California Riverside.

Chang, J.S., Li, Y., Beauharnois, M., Huang, H.-C., Lu, C.-H., and Wojcik G., 1996. SAQM user's guide. California Air Resources Board, Sacramento, CA, 500 pp.

Coats, C.J., 1996. High-performance algorithms in the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system. Proc. Ninth AMS Joint Conference on Applications of Air Pollution Meteorology with AWMA, Amer. Meteor. Soc., Atlanta, GA, 584–588.

Côtè, J., Gravel, S., Méthot, A., Patoine, A., Roch, M., Staniforth, A., 1998a. The operational CMC-MRB Global Environmental Multiscale (GEM) Model. Part I: Design considerations and formulation. Monthly Weather Review, 126, 1373–1395.

Côtè, J., Desmarais, J.-G., Gravel, S., Méthot, A., Patoine, A., Roch, M., Staniforth, A., 1998b. The operational CMC-MRB Global Environmental Multiscale (GEM) Model. Part II: Results. Monthly Weather Review, bf 126, 1397–1418.

Delle Monache, L., and Stull, R.B., 2003. An ensemble air-quality forecast over western Europe during an ozone episode. Atmospheric Environment, 37, 3469–3474.

Delle Monache, L., Deng, X., Zhou, Y., and Stull, R., 2005a. Ozone ensemble forecasts. Part I: A new ensemble design. Submitted to Journal of Geophysical Research.

Delle Monache, L., Nipen, T., Deng, X., Zhou, Y., and Stull, R., 2005b. Ozone ensemble forecasts. Part II: A Kalman-filter predictor bias correction. Accepted for publication in Journal of Geophysical Research.

DeMore, W.B., et al., 1994. Chemical kinetics and photochemical data for use in stratospheric modeling, Evaluation Number 11. NASA Jet Propulsion Laboratory Publ. 84-26, Pasadena, CA, 273 pp.

EPA, EPA Clearinghouse for inventories and emissions factors, 2004. 1999 National Emission Inventory Documentation and Data – Final Version 3.0, http://www.epa.gov/ttn/chief/net/1999inventory.html.

Gery, M. W., et al., 1989. A photochemical kinetics mechanism for urban and regional scale computer models. Journal of Geophysical Research, 94, 12295–12356.

Gong, S.L. et al., 2003. Canadian Aerosol Module: A size segregated simulation of atmospheric aerosol processes for climate and air quality models: Part 1. Module development. Journal of Geophysical Research, 108(D1), 4007, doi:10.1029/2001JD002002

Grell, G.A., Dudhia, J., and Stauffer, D.R., 1994. A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5). NCAR/TN-398+STR, 122 pp.

Grell, G.A., Peckham, S.E., Schmitz, R., McKeen, S.A., Frost, G., Skamarock, W., and Eder, B., 2005. Fully coupled "online" chemistry within the WRF model. Atmospheric Environment, 39, 6957–6975.

Guenther, A. B., et al., 1995. A global model of natural volatile organic compound emissions. Journal of Geophysical Research, 100D, 8873–8892.

Hamill, T. M., 1999. Hypothesis tests for evaluating numerical precipitation forecasts. Weather and Forecasting, 14, 155–167.

Hou, D., Kalnay, E., and Droegemeier, K.K., 2001. Objective verification of the SAMEX'98 ensemble forecast. Monthly Weather Review, 129, 73–91.

Kalnay, E., 2003. Atmospheric modeling, data assimilation and predictability. Cambridge Univ. Press, New York, 341 pp.

Kasibhatla, P., Chameides, W.L., Duncan, B., Houyoux, M., Jang, J., Mathur, R., Odman, T., and Xiu, A., 1997. Impact of inert organic nitrate formation on ground-level ozone in a regional air quality model using the Carbon Bond Mechanism 4. Geophysical Research Letters, 24(24), 3205–3208.

Krishnamurti, T.N., Kishtawal, C.M., LaRow, T., Bachiocchi, D., Zhang, Z., Williford, C.E., Gadhil, S., and Surendran, S., 1999. Improved weather and seasonal climate forecasts from multi-model superensemble. Science, 285, 1548–1550.

McHenry, J.N., Ryan, W.F., Seaman, N.L., Coates, C.J., Pudykiewicz, J., Arunachalam, S., and Vukovich, J.M., 2004. A real-time Eulerian photochemical model forecast system. Bulletin of the American Meteorological Society, 85, 525–548.

McHenry, J.N., and Coats, C.J., 2003. Improved representation of cloud/actinic flux interaction in multiscale photochemical models. Combined Preprints CD-ROM, Proc. Fifth Conference on Atmospheric Chemistry: Gases, Aerosols, and Clouds, Paper 2.1 Amer. Meteor. Soc., Long Beach, CA.

McKeen, S.A. et al., 2005. Assessment of an ensemble of seven real-time ozone forecasts over Eastern North America during the summer of 2004. Journal of Geophysical Research, 110(D21), D21307.

Moran, M.D., Scholtz, M.T., Slama, C.F., Dorkalam, A., Taylor, A., Ting, N.S., Davies, D., Makar, P.A., and Venkatesh, S., 1997. An overview of CEPS1.0: Version 1.0 of the Canadian Emissions Processing System for regional-scale air quality models. Proc. 7th AWMA Emission Inventory Symp., Oct. 28–30, Research Triangle Park, North Carolina, Air Waste Management Association, Pittsburgh.

Moran, M.D., A.P. Dastoor, S.-L. Gong, W. Gong, and P.A. Makar, 1998. Conceptual design for the AES unified regional air quality modelling system (AURAMS), Internal Report, 100 pp., Air Quality Research Branch, Atmospheric Environment Service, (Now Meteorological Service of Canada; available from the first author).

Pagowski, M., et al., 2005. A simple method to improve ensemble-based ozone forecasts, Geophysical Research Letters, 32, L07814, doi:10.1029/2004GL022305.

Pudykiewicz, J., Kallaur, A., and Smolarkiewicz, P.K., 1997. Semi-Lagrangian modeling of tropospheric ozone, Tellus, 49B, 231–258.

Russell, A., and R. Dennis, 2000. NARSTO critical review of photochemical models and modeling. Atmospheric Environment, 34, 2283–2324.

Schaefer, J. T., 1990. The critical success index as an indicator of warning skill. Weather and Forecasting, 5, 570–575.

Schell, B., Ackermann, I.J., Hass, H., Binkowski, F.S., Ebel, A., 2001. Modeling the formation of secondary organic aerosol within a comprehensive air quality model system. Journal of Geophysical Research, 106(D22), 28275-28294, 10.1029/2001JD000384.

Shin, D. W., and Krishnamurti T.N., 2003. Short- to medium-range superensemble precipitation forecasts using satellite products: 1. Deterministic forecasting. Journal of Geophysical Research, 108(D6), 8383, doi:10.1029/2001JD001510.

Stockwell, W. R., Middleton, P., Chang, J.S., and Tang, X., 1995. The effect of acetyl peroxy-peroxy radical reactions on peroxyacetyl nitrate and ozone concentrations. Atmospheric Environment, 29, 1591–1599.

Tang, Y., et al., 2004. Three-dimensional simulations of inorganic aerosol distributions in East Asia during spring 2001. Journal of Geophysical Research, 109, D19S23,doi:10.1029/2003JD004201.

Vautard, R., Blond, N., Schmidt, H., Derognat, C., and Beekmann, M., 2001. Multi-model ensemble ozone forecast over Europe: Analysis of uncertainty. Paper presented at 26th EGS General Assembly, Nice, France.

West, M, and Harrison, J., 1989. Bayesian forecasting and dynamic models. Springer, New York, 680 pp.

Wilks, D.S., 1995. Statistical methods in the atmospheric sciences. Academic Press, San Diego, CA, 467 pp.

Zhu, Y., Toth, Z., Wobus, R., Richardson, D., Mylne, K., 2002. The economic value of ensemble-based weather forecasts. Bulletin of the American Meteorological Society, 83, 73–83.

Tables and Captions

TABLE 1. Bias (ppb), RMSE (ppb), and correlation (unitless) for hourly forecasts of surface ozone concentrations with DLR ensemble, averaged ensemble, and single models.

Model	Bias	RMSE	Corr
DLR	0.91	10.57	0.810
AVE	12.70	18.01	0.765
A	5.32	17.50	0.668
B	8.01	18.13	0.616
C	9.33	18.81	0.631
D	11.53	24.16	0.696
E	14.83	21.40	0.651
F	31.16	37.45	0.549
G	8.77	16.52	0.721

TABLE 2. Bias (ppb), RMSE (ppb), and correlation (unitless) for maximum daily 8-h and 1-h averaged surface ozone concentrations calculated from hourly forecasts with DLR ensemble, averaged ensemble, and single models.

Model		8-h max			1-h max	
	Bias	RMSE	Corr	Bias	RMSE	Corr
DLR	0.36	9.80	0.683	-2.96	11.01	0.715
AVE	7.98	13.12	0.715	6.91	13.52	0.710
SVD	0.51	10.11	0.724	0.95	11.75	0.704

TABLE 3. Contingency table as a basis for categorical forecasts.

	Observed YES	Observed NO
Forecast YES	a	b
Forecast NO	c	d

Figure Captions

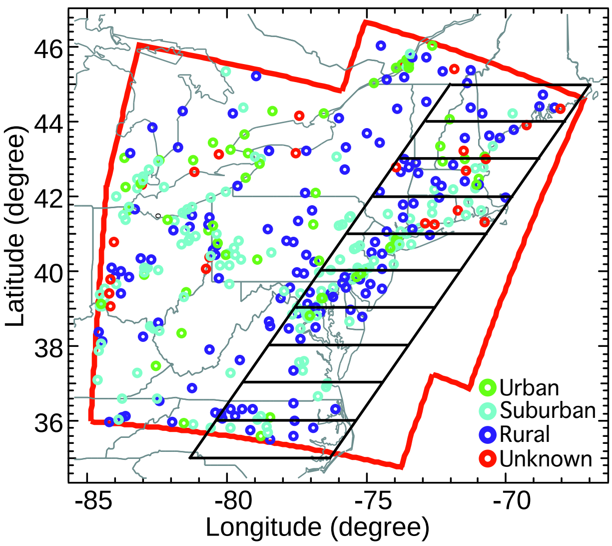
FIG. 1. Locations of the sites, the AIRNow site classification, and outline of the domain of model overlap. Also, latitude/longitude hatching for sites used in plotting Fig. 2 is shown.

FIG. 2. Temporal variability of the maximum daily 8-h averaged ozone concentrations for the bin-averaged sites confined within the shown latitude/longitude belt. Contours of temperature at 900 mb are overlaid.

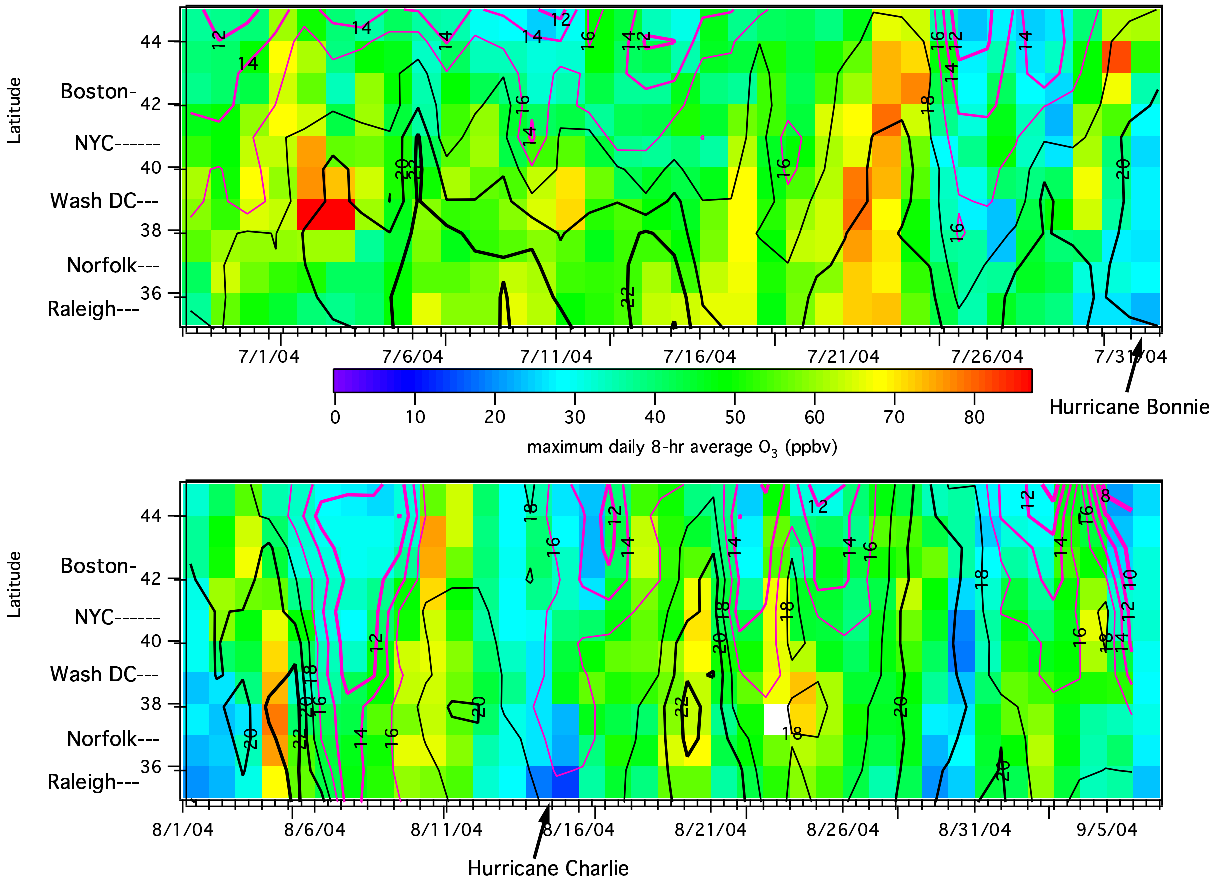
FIG. 3. Comparison of observations (red), DLR ensemble (purple), and ensemble average (blue) for a time series at a particular ozone monitor (a) and weights (b). The different colors in the figure b signify different models.

FIG. 4. Scatterplots for DLR ensemble (a,d), ensemble average (b,e), and persistence (c,f) for the daily maximum of 8-h (top) and 1-h (bottom) averaged ozone concentration.

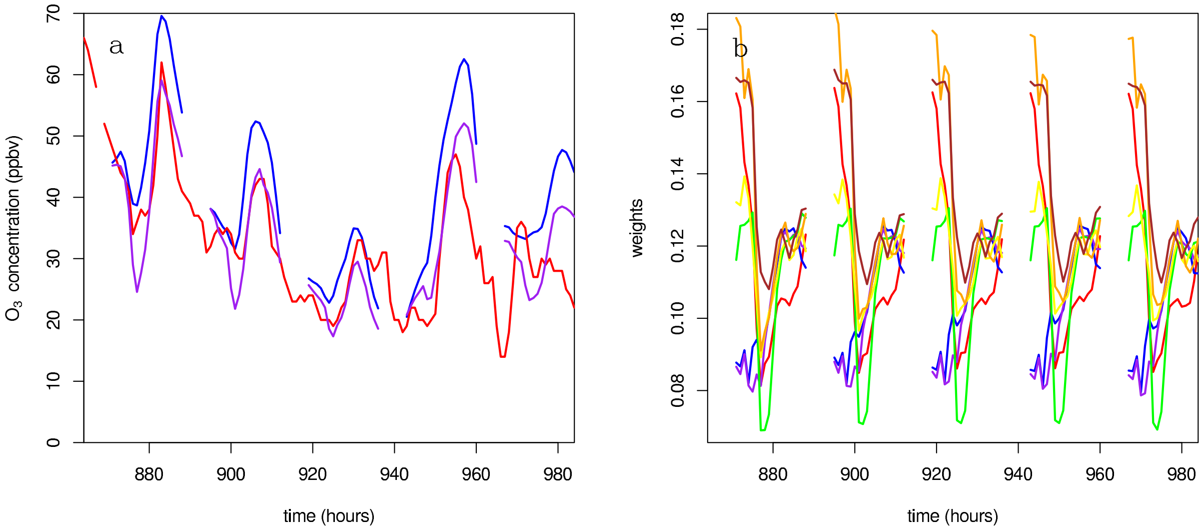
FIG. 5. Bias ratio (a,c) and equitable threat score (b,d) for DLR ensemble, ensemble average, persistence or the daily maximum of 8-h (top) and 1-h (bottom) averaged ozone concentration.



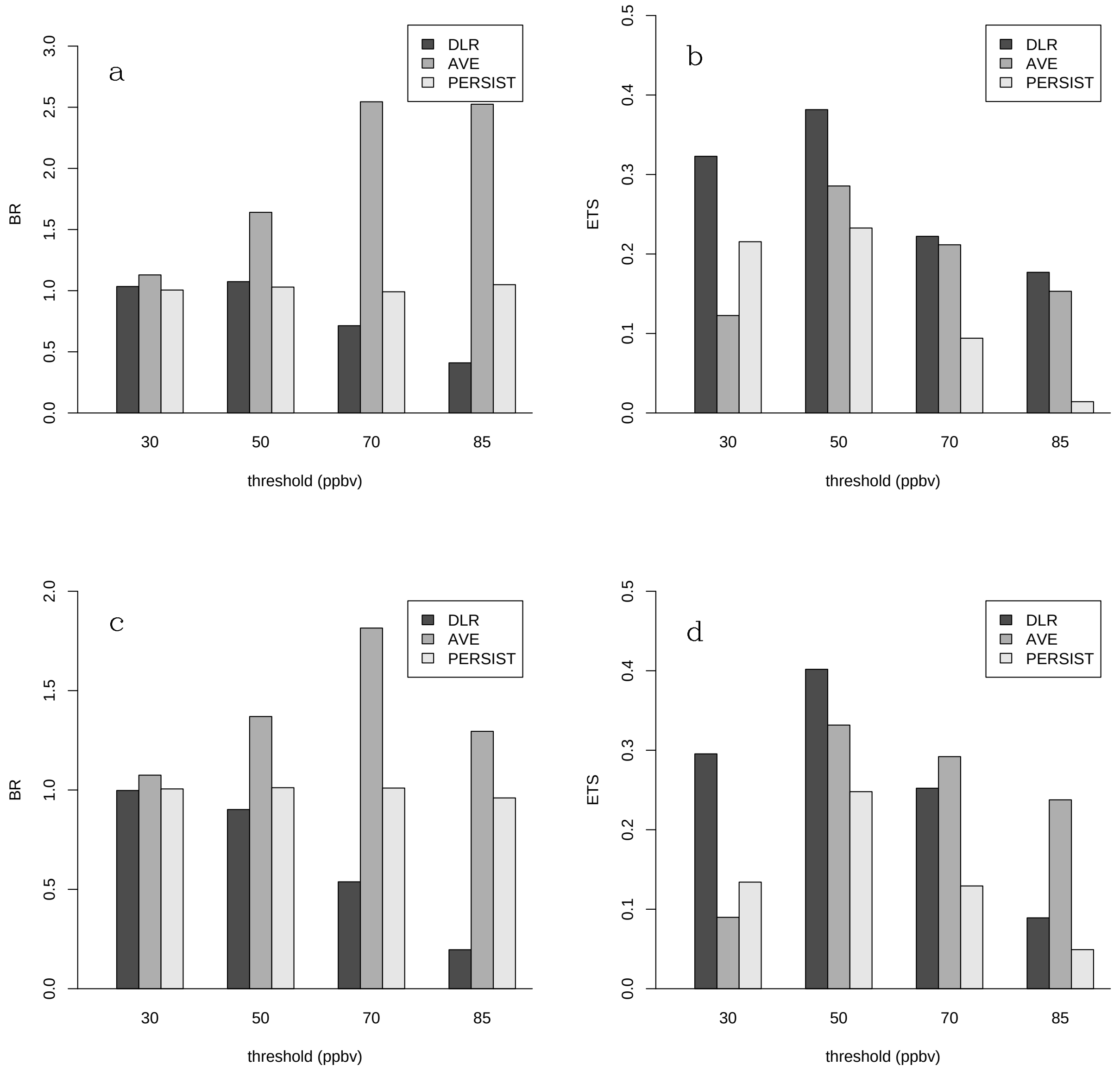
46
44
42
40
38
36
Latitude (degree)
-85
-80
-75
-70
Longitude (degree)
Urban
Suburban
Rural
Unknown



Latitude
Boston-
NYC------
Wash DC---
Norfolk---
Raleigh---
44
42
40
38
36
7/1/04
7/6/04
7/11/04
7/16/04
7/21/04
7/26/04
7/31/04
Hurricane Bonnie
0
10
20
30
40
50
60
70
80
maximum daily 8-hr average O_3 (ppbv)
8/1/04
8/6/04
8/11/04
8/16/04
8/21/04
8/26/04
8/31/04
9/5/04
Hurricane Charlie



a
O_3 concentration (ppbv)
time (hours)
b
weights
time (hours)


a
b
c
d
DLR
AVE
PERSIST
BR
ETS
threshold (ppbv)
30
50
70
85
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0.1
0.2
0.3
0.4