1.1 AN EVALUATION OF THE NATIONAL AIR QUALITY FORECAST CAPABILITY

FOR THE SUMMER OF 2007

Brian Eder[1*], Daiwen Kang[2], Rohit Mathur[1], Jon Pleim[1] and Shaocai Yu[2]

ASMD, ARL, NOAA, Mail Drop E-243-01, Research Triangle Park, NC 27711

[1]In partnership with the NERL, U. S. Environmental Protection Agency, RTP, NC 27711
[2]On assignment from Science and Technology Corporation, Hampton, VA 23666

1. INTRODUCTION

The real-time National Air Quality Forecast Capability (NAQFC), developed through a collaborative effort between the National Oceanic and Atmospheric Administration (NOAA) and Environmental Protection Agency (EPA), has expanded in both its scope and spatial coverage since its deployment in 2004. Initially, the NAQFC provided forecasts of maximum 1- and 8-hour ozone (O_3) concentrations for the northeast quadrant of the United States utilizing the National Center for Environmental Prediction's (NCEP) Eta meteorological model (Black, 1994; Rogers et al., 1996) coupled with EPA's Community Multiscale Air Quality (CMAQ) modeling system (Byun and Schere, 2006). After a successful "experimental" phase (where success was determined by the System's ability to correctly predict exceedance and nonexceedance of O_3 concentration threshold metrics on 90% or more days), these forecasts were transitioned into operations during September of that year, with official dissemination to the public available via a NOAA website: www.nws.noaa.gov/aq/.

In June of 2005 an updated version of the NAQFC, incorporating numerous refinements developed by NOAA's Air Resources Lab, began providing experimental simulations over an expanded domain that included all States east of the Rocky Mountains. This version, which was similarly transitioned into operations during September of 2005, provided O_3 forecasts through the summers of 2006 and 2007. During this two year period, major changes were taking place with the experimental phase of the NAQFC. First and foremost, was the replacement of NCEP's Eta meteorological model with the North American Mesoscale (NAM) meteorological model run of the Weather Research and Forecast (WRF) model (Skamarock, et al., 2007). The spatial domain of this new NAM-CMAQ configuration, which had been running in parallel with the operational Eta-CMAQ runs, was also expanded to encompass the contiguous United States (CONUS) and bordering areas (Figure 1).

The purpose of this paper is to provide a brief synopsis of this new system's performance during the summer of 2007 that supported, in part, its transition into operational status on September 18th. This evaluation uses a suite of metrics similar to those used in evaluation of the earlier configurations of the NAQFC (Kang et al. 2006; Eder, et al., 2006a), focusing primarily on the performance of *discrete forecasts* for the maximum 8-hour O_3 concentrations covering a four month period (June 1 through September 30) using O_3 measurements obtained from EPA's AIRNow network.

2. DESCRIPTION OF THE NAQFC

As discussed above, the NAQFC now uses the NAM model run of the WRF model configured with EPA's CMAQ Modeling System. A brief summary of their linkage, relevant to this study, is presented below. For a more in-depth description, readers are referred to Otte et al. (2005). The NAM consists of the Nonhydrostatic Mesoscale Model (NMM) (Janjic, 2003) version of the WRF model. The WRF-NMM model is used to prepare the meteorological fields for input to the CMAQ. The NCEP Product Generator software is used to perform bilinear interpolations and nearest-neighbor mappings of the WRF Post-processor output from NAM forecast domain to the CMAQ forecast domain. The processing of the emission data for various pollutant sources has been

* *Corresponding author address:* Brian K. Eder, ASMD, NERL, EPA. MDE-243-01, RTP, NC 27711; phone: 919.541.3994; e-mail: eder.brian@epa.gov.

adapted from the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system (Houyoux et al., 2000) on the basis of the EPA's 2001 national emission inventory. The Carbon Bond chemical mechanism (version 4.2) is used to represent the photochemical simulations. Detailed information on transport and cloud processes in the CMAQ is described in Byun and Schere (2006).

For this application, O_3 concentrations are forecast over the CONUS using a 12-km horizontal grid spacing on a Lambert Conformal map projection. There are 22 layers extending from the surface to 100 hPa using the same hybrid pressure/sigma vertical coordinate used by the WRF-NMM. CMAQ layers are interpolated from the 60 WRF-NMM layers, such that CMAQ's layer interfaces are coincident with WRF-NMM layer interfaces. The initial condition chemical fields for CMAQ are specified using the previous forecast cycle. The NAM 06 and 12 UTC cycles are used for the forecast cycles. The primary NAM–CMAQ model forecast for next day surface-layer O_3 is based on the current day's 12 UTC NAM cycle. The target forecast period is local midnight through local midnight. An additional eight hours are required beyond midnight to calculate peak 8-h average O_3 concentrations. As a result, a 48-h NAM–CMAQ forecast is needed (based on the 12 UTC initialization) to obtain the desired 24-h forecast period.

3. OZONE DATA

Hourly, near real-time, O_3 (ppb) data obtained from EPA's AIRNow program are used in the evaluation (www.epa.gov/airnow). Nearly 1100 monitors were available Nationwide (and from bordering Canadian Provinces) resulting in over three million hourly O_3 observations for the study period. The maximum 8-hour concentrations are calculated for each station and each day over the four month evaluation period using the same forward calculation method used in the model. Concentrations are considered missing if half of the hourly observation data are missing for the day. This evaluation uses a stringent, direct matching between monitor location and grid cell (i.e. no interpolation or smoothing of observed or model concentrations is used). If two or more monitors are located within a model grid cell, their average value is used as the representative measurement for that grid cell.

4. STATISTICS

A suite of statistical metrics was calculated, focusing primarily on *discrete* (observed versus modeled concentrations), though including some *categorical* (observed versus modeled exceedances / non-exceedances) forecasts of the maximum 8-hour O_3 simulations. For the *discrete* forecast evaluation, the metrics included: Root Mean Square Error (RMSE), Normalized Mean Error (NME), Mean Bias (MB), Normalized Mean Bias (NMB) and correlation coefficient (r). Further detail describing the discrete metrics is provided in Eder et al. (2006a).

For the *categorical* forecast evaluation, the model's Accuracy (A), Bias (B), False Alarm Rates (F), and Critical Success Index (CSI) were calculated (Jolliffe and Stephenson, 2003). An additional categorical metric, called the Weighted Success Index (WSI) is also calculated. The WSI, recently developed by Kang et al. (2007), is similar to, though less stringent than the CSI. Each of these categorical metrics are based on the observed exceedances, non-exceedances *versus* forecast exceedance, non-exceedances for the 8-hour O_3 standard (85 ppb). All of the metrics were calculated daily throughout of the evaluation period and are summarized for the four month period June through September.

5.0 METEOROLOGICAL CONDITIONS

The meteorological conditions during the O_3 season of 2007 were abnormally warm and dry across most of the CONUS, resulting in conditions often conducive to the formation of O_3. As a result, over 1600 exceedances of the NAAQS 8-hour standard were recorded. For the three month period (June-August), the average temperature for the contiguous U.S. was 23.2℃, which was 1.0℃ above the 20th century mean and the sixth warmest summer since national records began in 1895 (www.ncdc.noaa.gov/). Of the 48 contiguous states covered by the NAQFC domain, all except 10 experienced above or much above normal temperatures. Exceptions included the Northeast States, which were near normal and several lower Midwest States, primarily Texas and Oklahoma, which were below normal due primarily to record precipitation amounts. The summer was drier than average for the nation as a whole, (32^{nd} driest since 1895) especially in the Southeast, mid-Atlantic, and Ohio Valley as well as the northern Plains and Northern Rockies.

6.0 EVALUATION

Because of the major differences in the NAQFC configurations employed over the last few years (i.e. different meteorological drivers, different domains, etc), a direct comparison with earlier performances is not feasible. This should not; however, preclude their use in establishment of a benchmark or "frame of reference" against which the latest results can judged. Accordingly, comparisons to earlier performance will occasionally be included in the subsequent discussion.

6.1. *Discrete evaluation.*

Examination of Table 1, which provides a summary of discrete statistics for the maximum 8-hr O_3 forecasts, reveals that the 2007 NAQFC performed as well or better than previous configurations, despite the expansion of the forecast domain into numerous areas comprised of complex terrain. The CONUS-wide mean correlations were $\geq$ 0.70 for each of the four months. These values represent a marked improvement over the inaugural forecasts of the NAQFC, where evaluation of a comparable time period (June-September, 2004) produced a mean correlation of 0.59 over a much smaller domain (Eder et al., 2006a); and are comparable to those associated with the 2005 season forecasts, where the mean correlation was 0.69 over a comparable time period (Eder et al., 2006b).

The NAM-CMAQ configuration systematically overpredicted the 8-hr O_3 concentrations, continuing a trend established by the two earlier NAQFC configurations, though to a lesser degree. The summer-long mean forecast value of 53.21 ppb was 4.25 ppb higher than the observed value (48.96 ppb), producing a CONUS-wide NMB of 8.67%. This value is considerably lower than those resulting from the 2004 (21.6%) and 2005 (20.0%) evaluations. As with the bias, the error associated with the configuration was also lower. The summer-long RMSE of 13.00 ppb (NME = 20.43%) represented marked improvements over earlier forecast (NME = 24.0% for 2004; 28.1 for 2005).

6.1.1 *Temporal*

In order to investigate the performance of the NAQFC over time, several of the discrete statistics discussed above were calculated (CONUS-wide) and plotted as a daily time series. As seen in Figure 2, the correlations exhibit a fairly consistent nature, with values exceeding 0.60 (0.70) over 90% (55%) of the days. Examination of the meteorological conditions associated with the few days when the correlations fell below 0.60 revealed that much of the domain (especially the eastern half, where a majority of the monitors are located) was subjected to extensive cloud cover and precipitation, resulting in low observed O_3 concentrations that the model was unable to replicate.

Unlike the correlations, the NMB and NME associated with the forecast were not temporally consistent, as they exhibited a increasing trend as the forecast season progressed (Figure 3). During the month of June, the CONUS NMB was essentially zero; however this is somewhat misleading as biases were slightly negative during the first three weeks of the month and greatly positive during the last week. This positive trend continued during the remainder of the evaluation period with the NMB rising to nearly 12% during July and over 15% during August and September. The NME likewise increased from an average of less than 18% during June, to over 22% during the remaining three months, occasionally exceeding 30% on individual days during August and September.

Reasons for these systematic increases in NMB and NME are being investigated. In addition to changes in atmospheric conditions, several corrections and modifications were made to the experimental NAQFC configuration, with special emphasis on improving performance over the western US.. The June 12^{th}modification involved changing the approach used in CMAQ for Planetary Boundary Layer simulations (PBL) from the direct use of the NAM eddy diffusion scheme to an approach developed and recently modified by Pleim (2007) called the Asymmetric Convective Model-2. The next two changes, which were associated with the WRF-NMM, involved modifications to roughness length and canopy resistance calculations. The roughness length for heat was modified so that under stable conditions, it is simply a function of the surface-layer bulk Richardson number, hence eliminating its dependence on surface elevation. The minimum canopy resistance was increased for evergreen (from 125.0 to 250.0 sm^{-1}) and mixed forest (from 125.0 to 150.0 sm^{-1}). The impact of the modifications, made on June 19^{th} are still being investigated.

The next changes were made on July 18^{th} and involved minor corrections to CMAQ. The first correction, removed vegetative canopy uptake for several species not involved in plant respiration:

NO, NO_2, and CO, The impact of the slight reduction in their deposition velocity, was to increase their concentrations, and in turn, the additional NOx produced slightly higher (2 – 5 ppb) O_3 concentrations as determined by recent sensitivity tests. The second correction, which had minimal impact on concentrations, fixed a minor numerical error in the plume-rise calculations.

6.1.2 *Spatial*

When examined over space, the experimental test configuration generally performed better, in terms of correlation, over the eastern half of the CONUS (Figure 4a). Correlations were generally lower in the western States, especially California, where values were generally less than 0.50 in southern and coastal locations. Reasons for these spatial discrepancies include: 1) the NAQFC has been in operation over the eastern CONUS for several years, allowing iterative refinement of the system in this area; and 2) the complexity of the terrain can be problematic, considering that the NAQFC uses a 12 km grid.

Examination of the spatial distribution of both the NMB and NME (which are similar due to the majority of biases being in one direction, i.e. positive, after the middle of June) reveals that the NAQFC was generally within 25% for the NME and ± 25% for the NMB over a majority of the domain. Several areas of poorer performance, where the NMB (NME) often exceed 25% and in some cases 50% are, however; noted. These areas include: coastal and southern California where several factors are being scrutinized, including inadequate representation of the complex terrain (discussed above); the marine boundary layer and emissions. Poor performance is also seen across the extreme northeastern portion of the domain, stretching from southern Quebec, eastward through Maine into New Brunswick. Here the system tends to overpredict, especially on days when clean, near-pristine air is advected into the area following the passages of cold fronts. This likely indicates that the northern boundary conditions need to be adjusted lower. And finally, the third area experiencing systematic overprediction is along the southeast Atlantic and Gulf coasts, stretching from Georgia to Texas. This is likely due to the system's inability to accurately resolve the sea breeze, given its 12 km resolution. Case studies are being developed for these regions in which each major component of the NAQFC will be closely scrutinized in order to determine the source of error.

6.2 *Categorical evaluation*

Table 2 provides both the monthly and summer categorical statistics that are based on observed exceedances, non-exceedances *versus* forecast exceedance, non-exceedances associated with the 8-hour O_3 standard (85 ppb). As seen in the table, values of the Accuracy (A), which measure the percentage of forecasts that correctly predict an exceedance or non-exceedance are high (> 95%). These high values, which are comparable to earlier NAQFC configurations, must be interpreted with care, as they are greatly influenced by the overwhelming number of correctly forecast non-exceedances (Eder et al. 2006b). The Critical Success Index (CSI), which is not unduly influenced by the number of correctly forecast non-exceedances, provides a more (perhaps too) stringent test of the forecast system. Values of the CSI indicate that just over 15% of the 1155 actual exceedances observed during the summer of 2007 were successfully predicted. This CSI, while disappointingly low, is comparable to earlier configurations of the NAQFC.

Because of the especially strict nature of the CSI, Kang et al. (2007) recently developed a similar, though less stringent, metric called the Weighted Success Index (WSI). The WSI "provides partial credit" to observation-forecast pairs that fall just outside limits established by the CSI but within a designated factor (set to 1.5 for this application), thereby providing a more representative measure of the model's performance. As seen in Table 2, values of the WSI averaged over 54% for the summer of 2007, which is considerably higher than the WSI value of 37.5% associated with the inaugural NAQFC configuration.

Values of categorical Bias (B), which indicate if the number of exceedances of the O_3 standard are overpredicted (B>1 indicates "false positive") or underpredicted (B<1 indicates "false negative") mirror values of the discrete bias (MB and NMB). They indicate an under prediction of exceedances during June (0.43) but over prediction during the remaining months, especially August (2.2). The 2007 summer averaged B (1.44) is considerable lower than those associated with either the 2004 (2.6) or 2005 (2.0) NAQFC configurations, indicating a general decrease in the number of "false positives" forecasts. The False Alarm Ratio, which indicates the percentage of times that an exceedance was forecast but did not occur, averaged 77.6%, which is an improvement over earlier configurations.

7. SUMMARY

The purpose of this paper has been to provide a brief synopsis of an ongoing performance evaluation of the new WRF-CMAQ NAQFC developed collaboratively between NOAA and EPA. This evaluation, which covered the four month period from June through September, used O_3 observations obtained from EPA's AIRNow program and a suite of statistical metrics for both discrete and categorical forecasts.

Domain-wide results revealed that the expanded 2007 NAQFC performed as well or better than previous configurations, despite the expansion of the forecast domain into many areas dominated by complex terrain. The CONUS-wide mean correlations were $\geq$ 0.70 for each of the four months. These values represent a marked improvement over the inaugural forecasts of the NAQFC. This latest configuration still overpredicts 8-hr O_3 concentrations (NMB = 8.67%), though to a lesser degree than earlier NAQFC configurations. When examined over time, the evaluation revealed that this overprediction systematically increased as the summer progressed. This was attributable to numerous changes and/or corrections made to the system during its experimental phase.

Examination of the spatial distribution of both the NMB and NME reveals that the NAQFC was generally within 25% for the NME and $\pm$ 25% for the NMB over a majority of the domain. Several areas of poorer performance, where the NMB (NME) often exceed 25% and in some cases 50% were noted. These areas include coastal and southern California and the southeast Atlantic and Gulf coasts. Reasons for the poorer performance in these areas are likely attributable to the system's inability to correctly resolve sea-breezes and complex terrain, given its 12 km resolution. The extreme northeastern portion of the domain, stretching from southern Quebec, eastward through Maine into New Brunswick was another problem area. Here the NAQFC systematically overpredicts, likely indicating that the northern boundary concentrations need to be lower. Case studies are being developed for these regions in which each major component of the NAQFC will be closely scrutinized in order to determine the source of bias (error).

8. REFERENCES

Black, T. 1994: The new NMC meso-scale Eta Model: description and forecast examples. *Weather and. Forecasting*, 9, 265-278.

Byun, D. and K. L. Schere, 2006: Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality modeling system. *Applied Mechanics Review*, **59**, 51-77.

Eder, B. K., D. Kang, R. Mathur, S. Yu and K. Schere, 2006a: An Operational evaluation of the Eta-CMAQ air quality forecast model. *Atmos. Environ*. **40**: 4894 - 4905.

Eder, B., D. Kang, R. Mathur, S. Yu and K. Schere, 2006b, An evaluation of the Eta-CMAQ air quality forecast model for the summer of 2005. *AWMA Specialty Conference Environmental Data Analysis*, Oct. 5, 2005, Oak Brook, IL.

Houyoux, M.R., J.M. Vukovich, C.J. Coats Jr., N.M. Wheeler, and P.S. Kasibhatla, 2000: Emission inventory development and processing for the seasonal model for regional air quality (SMRAQ) project. *J. of Geophysical Res.* 105, 9079-9090.

Janjic, Z. I., 2003: A Nonhydrostatic Model Based on a New Approach. *Meteorology and Atmospheric Physics*, 82, 271-285.

Jolliffe, I.R. and D.B. Stephenson, 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science.* John Wiley & Sons, Ltd., West Sussex, England, 240 pp.

Kang, D., R. Mathur, K. Schere, S. Yu and B. Eder, 2007: New categorical metrics for air quality model evaluation. *J. Appl. Met and Clim*., 46, 549-555.

Otte, T.L., and coauthors, 2005: Linking the Eta model with the Community Multi-scale Air Quality (CMAQ) modeling system to build a national air quality forecasting system. *Weather and Forecasting,* 20, 367-384.

Rogers, E., T. Black, D. Deaven, G. DiMego, Q. Zhao, M. Baldwin, N. Junker, and Y. Lin, 1996: Changes to the operational "early" Eta Analysis/Forecast System at the National Centers for Environmental Prediction. *Weather and Forecasting* 11, 391-413.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang and J. G. Powers, 2007: A Description of the Advanced Research WRF Version 2. NCAR/TN–468+STR, NCAR Technical Note.

Disclaimer The research presented here was performed under the Memorandum of Understanding between the U.S. Environmental Protection Agency (EPA) and the U.S. Department of Commerce's National Oceanic and Atmospheric Administration (NOAA) and under agreement number DW13921548. This work constitutes a contribution to the NOAA Air Quality Program. Although it has been reviewed by EPA and NOAA and approved for publication, it does not necessarily reflect their policies or views.

Period	Obs. (ppb)	Mod. (ppb)	RMSE (ppb)	NME (%)	MB (ppb)	NMB (%)	r
June	50.32	50.29	11.36	17.58	0.03	0.22	0.71
July	46.90	52.36	12.71	20.95	5.46	11.87	0.70
August	49.08	56.29	14.51	23.61	7.21	15.23	0.70
September	45.45	52.24	12.58	22.45	6.80	15.39	0.71
Summer (J, J, A)	**48.96**	**53.21**	**13.00**	**20.43**	**4.25**	**8.67**	**0.70**

Table 1. Summary of discrete statistics for maximum 8-hr O_3 forecast for CONUS.

Period	A	B	FAR	CSI	WSI	> 85 ppb
June	98.02	0.43	50.18	17.78	56.24	499
July	97.38	1.87	80.95	14.16	50.85	263
August	95.37	2.15	81.23	14.7	55.12	393
September	98.4	1.46	83.42	10.92	58.04	197
Summer (J, J, A)	**96.91**	**1.44**	**77.64**	**15.21**	**54.15**	**1155**

Table 2. Summary of categorical statistics for maximum 8-hr O_3 forecast for CONUS.

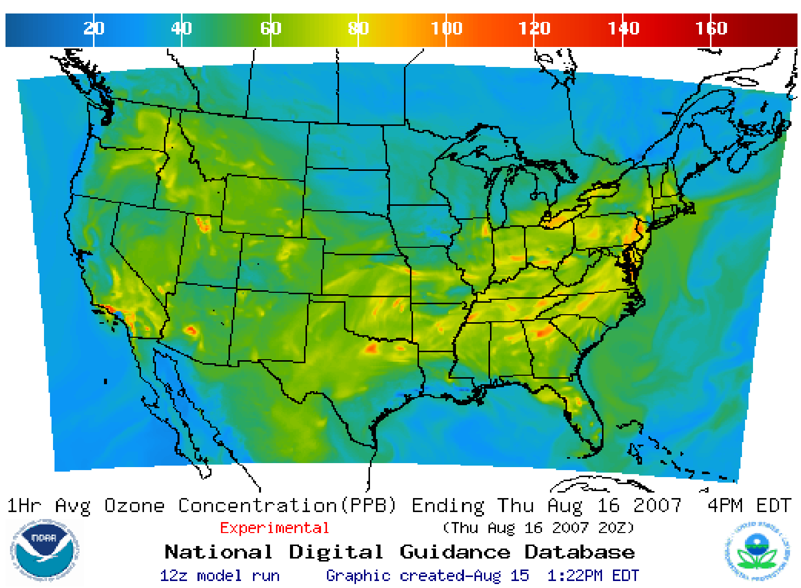

Fig. 1. An example of NOAA's Air Quality Forecast National Digital Guidance Database for the maximum 1-hour O_3 concentrations, depicting the WRF-CMAQ modeling domain, available at www.weather.gov/aq/. (Note that as of Aug 16, this configuration was still "Experimental".)

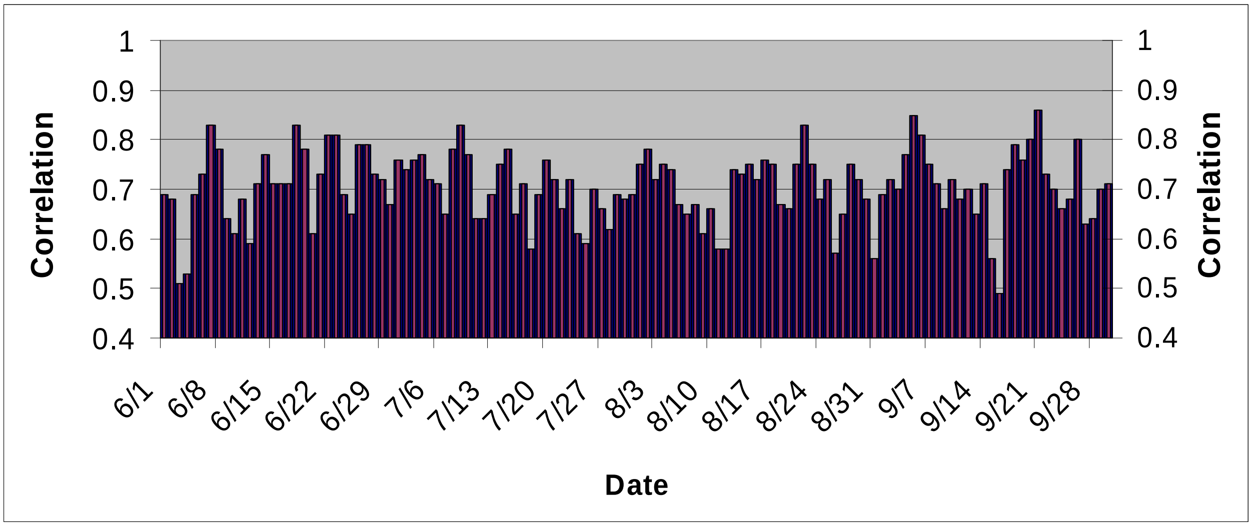

Fig. 2. Daily distribution of the mean CONUS-wide correlation coefficients from June 1 through September 30, 2007.

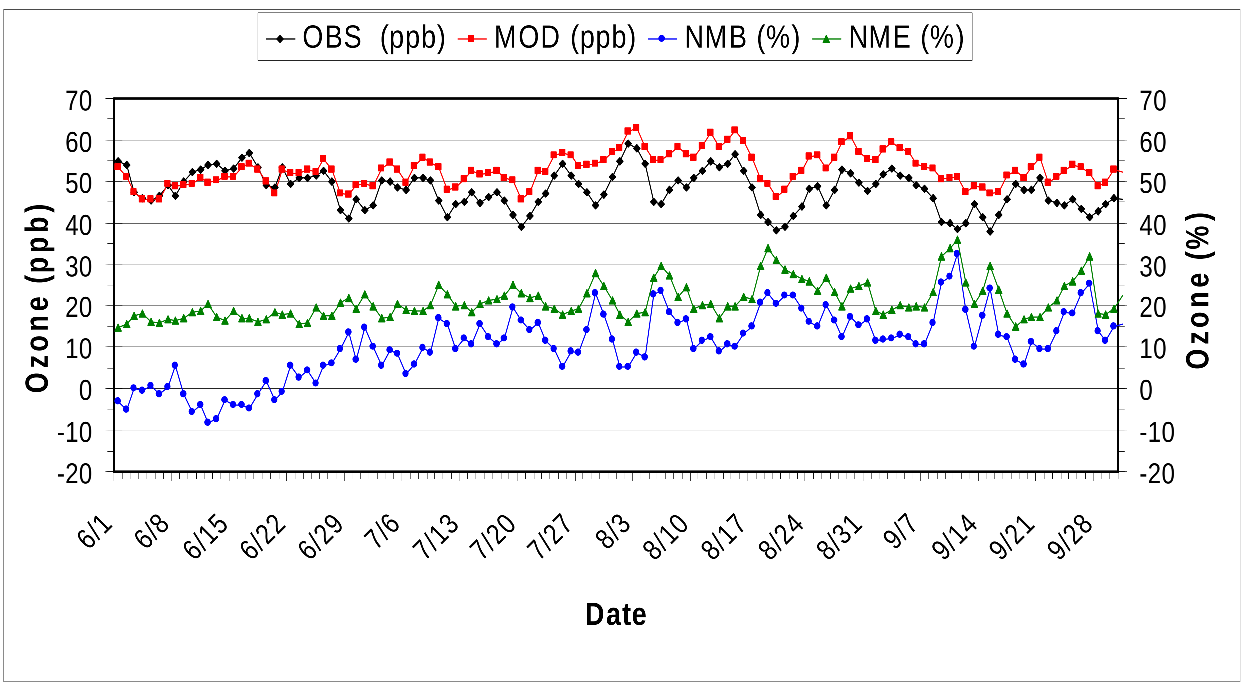

Fig. 3. Daily distribution of the observed and modeled O_3 concentrations along with the Normalized Mean Bias (NMB) and Normalized Mean Error (NME) from June 1 through September 30, 2007.

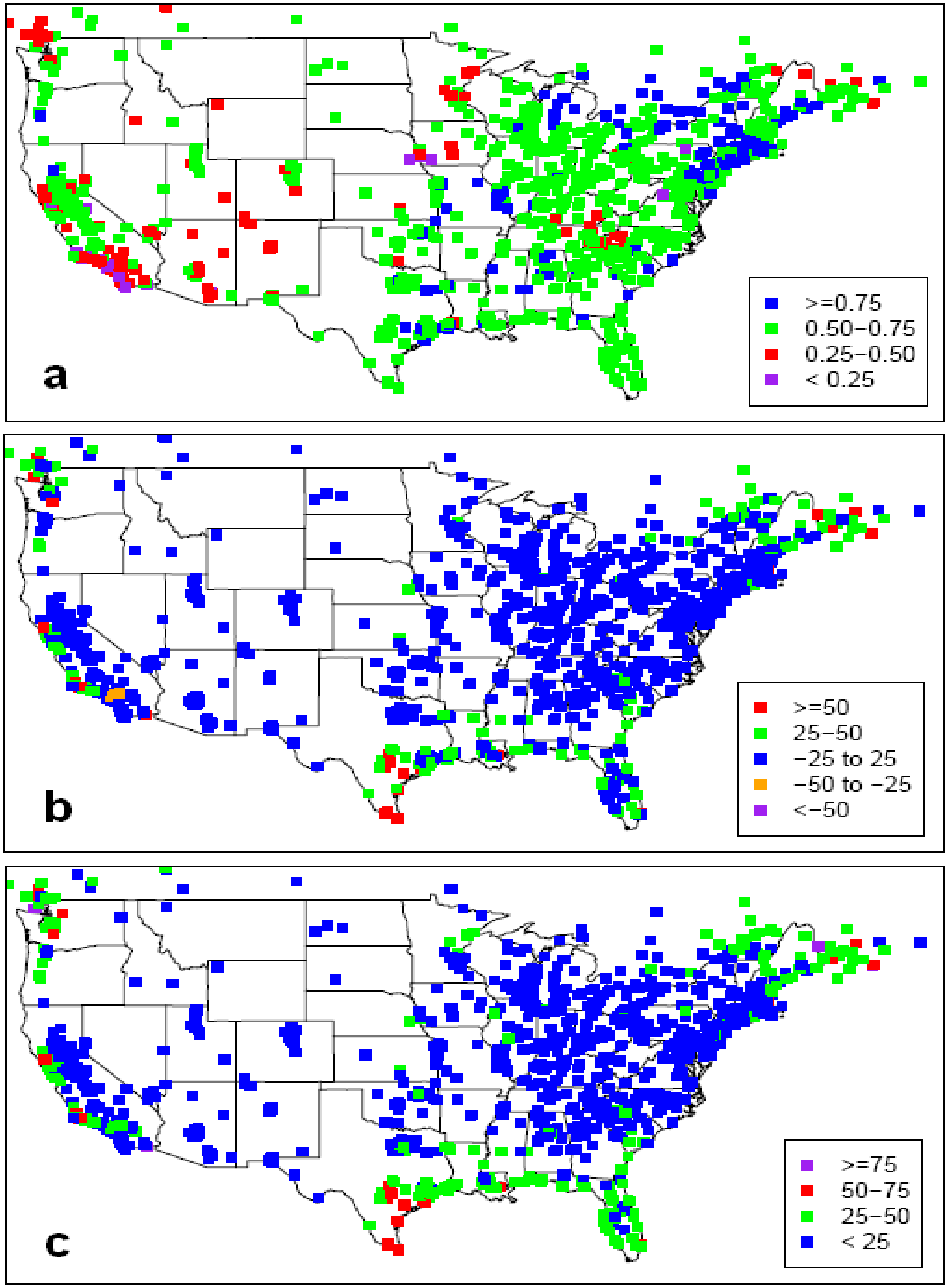

Fig. 4. Spatial distribution of the Correlation coefficient (a), Normalized Mean Bias(%) (b) and Normalized Mean Error (%) (c) associated with the max. 8-hr O_3 concentrations for the summer (June, July and August) of 2007.