Presented at the 8th Annual CMAS Conference, Chapel Hill, NC, October 19-21, 2009

PRELIMINARY EXPERIENCES WITH THE MULTI-MODEL AIR QUALITY FORECASTING SYSTEM FOR NEW YORK STATE

Prakash Doraiswamy[1,†,*], Christian Hogrefe[1,2], Winston Hao[2], Brian Colle[3], Mark Beauharnois[1], Ken Demerjian[1], J.-Y. Ku[2] and Gopal Sistla[2]

[1]Atmospheric Sciences Research Center, University at Albany, Albany, NY
[2]Bureau of Air Quality Analysis and Research (BAQAR), New York State Department of Environmental Conservation (NYSDEC), Albany, NY
[3]School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY
[†] On assignment to BAQAR, NYSDEC

1. INTRODUCTION

In recent years, three dimensional air quality models such as the Community Multiscale Air Quality (CMAQ) model are being used regularly to forecast daily air quality, and offer guidance to official forecasters, who utilize these model-based forecasts along with other forecasting tools. The New York State Department of Environmental Conservation (NYSDEC), in collaboration with the University at Albany and Stony Brook University (SUNY-SB), has implemented an ensemble air quality forecasting system in an attempt to better quantify uncertainties associated with the ozone (O_3) and $PM_{2.5}$ forecasts (Hogrefe et al., 2008). The multi-model system has been in operation since June 2008 with members added over the course of the past year. In its current form, the system consists of six members that are in daily operation. In addition, retrospective simulations were performed in which CMAQ was driven by twelve archived members of the SUNY-SB short range ensemble weather forecasting system (SREF, (Jones et al., 2007); http://chaos.msrc.sunysb.edu/ NEUS/ nwp_graphics.html) for a summer and winter period. This study presents a comparison of the O_3 and total $PM_{2.5}$ forecasts against measurements and official NYSDEC forecasts for summer and winter periods over New York State.

2. MODEL AND OBSERVATIONAL DATABASE

The multi-model ensemble system was implemented in June 2008 initially consisting of 4 members with 2 additional members included over the past year. The ensemble members differ in terms of the meteorological forecasts, the emission inventories and the air quality models that drive the simulations. Table 1 presents a summary of the ensemble members. The NCEP_12z and NCEP_00z simulations utilized the setup described by Otte et al. (2005), Kang et al. (2005), Yu et al. (2008) and Hogrefe et al. (2007). The NYSDEC_3x simulations utilized processed meteorological fields from the NCEP_00z setup, but used NYSDEC emission inventory processed with the Sparse Matrix Operator Kernel Emissions Modeling System (SMOKE). For the SUNYSB and ASRC members, the meteorological fields were processed with the Meteorology-Chemistry Interface Processor (MCIP) and emission inventories were processed with SMOKE prior to the CMAQ (for SUNYSB) or CAMx (for ASRC) simulation. For the retrospective twelve member simulations, CMAQ forecasts were driven by seven daily MM5 and five daily WRF-ARW members from the 00:00 UTC SUNY-SB SREF system. The summer simulation was performed for 6/4/08- 7/22/08, while the winter simulation was conducted from 12/1/08-2/28/09. It should be noted that emission inputs, domain size, and vertical grid structures vary across the members. All the SUNY-SB members use a consistent projection and grid structure.

Hourly measured concentrations of O_3 and total $PM_{2.5}$ were downloaded from the EPA AIRNOW system for monitors in New York State. Daily maximum 8-hr average O_3 and 24-hr average $PM_{2.5}$ concentrations were then determined from the hourly data and used in the subsequent analyses.

3. DATA ANALYSIS, RESULTS AND DISCUSSION

This study focuses on a preliminary evaluation of the O_3 and the total $PM_{2.5}$ forecasts by the multi-model system, including the 12-member retrospective simulation. Model predictions are compared with observations from the AIRNOW network and with official NYSDEC forecasts by air

**Corresponding author:* Prakash Doraiswamy, BAQAR, New York State Department of Environmental Conservation, 625 Broadway, Albany, NY 12233-3259; Phone: (518) 402-8402; Fax: (518) 402-9035; Email: pdorai@gmail.com

quality forecast regions (Figure 1) that are used by the official forecasters. As per that approach, the predicted and observed concentration of a species within a region corresponds to the maximum of the predicted and observed concentration across all monitors within that region. For example, the daily 8-hr maximum O_3 concentration for Region 4 on any day would be equal to the highest of the daily 8-hr maximum O_3 concentrations across the four O_3 monitors within that region. Similarly, ensemble averages and standard deviations are calculated at each monitor. The maximum ensemble average across the monitors within a region (and the standard deviation associated with that mean) represents the average (and standard deviation) for the region. It must be noted that the official DEC forecasts are based upon expert judgment and an examination of a variety of products including the ensemble-based forecast system presented here.

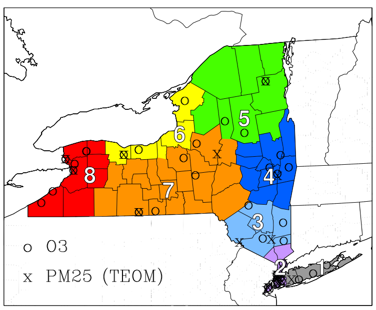

Figure 1. Eight air quality forecast regions in NY State with locations of O_3 and $PM_{2.5}$ monitors

Discrete (bias and root mean squared error [RMSE]) and categorical metrics (False Alarm Ratio [FAR], Probability of Detection [POD], and Critical Success Index [CSI] as per Kang et al. (2005)) were calculated for each region. The categorical metrics used the threshold that corresponds to the transition from the "moderate" to the "unhealthy for sensitive groups" (USG) range of the Air Quality Index (AQI), which is 75 ppb for ozone (U.S. EPA, 2008) and 35.4 µg/m^3 for $PM_{2.5}$ for forecasts issued in New York State (NYSDEC, 2007).

3.1. Summer 2008 (June to September): Daily 4-member system

A time series (not shown) of the observations and the individual model predictions of O_3 and $PM_{2.5}$ showed that the model predictions tracked the observations. Over-predictions were noticed in ozone predictions particularly during September. For $PM_{2.5}$, some under-prediction was noticed, particularly on days when the observed concentrations were near or above 35 µg/m^3. Figure 2 shows the mean bias in O_3 predictions by the 4-member system during June –September 2008. All figures are shown only for selected regions in NY State. Also shown are the performances of the ensemble average, ensemble median, and the official DEC forecasts. O_3 predictions showed a mean bias of ~2 to 7 ppb. Except Regions 1 and 2, it appears that the NCEP_t12z and NCEP_00z models showed lower bias than the SUNYSB members. The SUNYSB_F2 member had the lowest bias in regions 1 and 2. The bias of the ensemble average was not always the lowest, particularly in upstate NY. The DEC forecasts showed similar or lower bias. All forecasts by the ensemble system and the official DEC forecasts showed a RMSE of 9 to 12 ppb. $PM_{2.5}$ was typically under-predicted in all regions except Region 2, with a negative bias reaching -3.6 µg/m^3. For $PM_{2.5}$, the ensemble average appeared to have similar or lower bias than the members. The DEC forecasts were positively biased but within 3 µg/m^3. RMSE of all $PM_{2.5}$ forecasts ranged from ~4 to 11 µg/m^3, with larger variability between the members in Region 2.

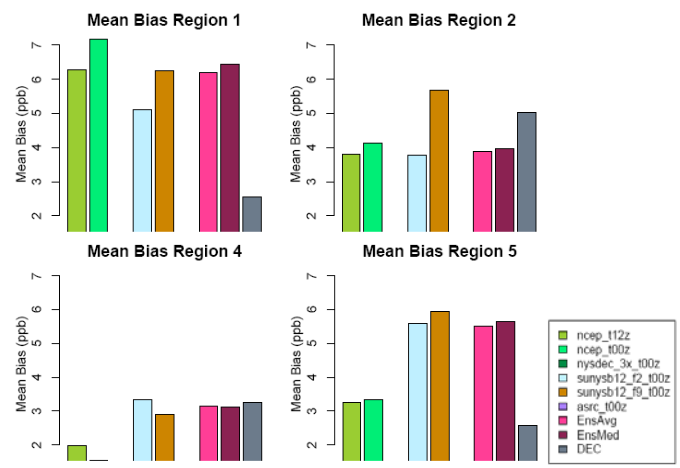

Figure 2. Mean Bias of Ozone Predictions by 4-member ensemble system during June - September 2008 at Selected Regions in NY. From left to right: the NCEP 12z and 00z members are shown in green, SUNYSB_F2 in blue, SUNYSB_F9 in orange, the ensemble average in pink, followed by median and then the official DEC forecasts in grey.

Categorical evaluation for ozone is presented for regions 1 and 2, where more exceedances were observed than other regions in NY. Although the NCEP_00z had 75-80% POD, it was accompanied with 36-56% FAR, resulting in a 40-53% CSI. Overall, the SUNYSB_F2 member appeared to perform better in both Regions 1 and 2 with a CSI of 50 to 67%, although for Region 2, the ensemble average showed a higher (62%) CSI. The official DEC forecasts showed 50-58% POD, but also had high false alarms on the same order as the NCEP members. Very few $PM_{2.5}$ exceedances were predicted, and none occurred on the few days observed. Hence, no categorical metrics are presented for $PM_{2.5}$.

3.2. Winter 2008-2009 (December 2008 to February 2009): Daily 5-member system

In addition to the 4 members simulated during summer, a new member (NYSDEC_3x) was included during the winter simulations. Modeled ozone concentrations during winter were always within an AQI of 50 ("Good" category, O_3 < 60 ppb) with no measurements being available from AIRNOW during this time period. Hence, model performance is presented for $PM_{2.5}$ only. $PM_{2.5}$ predictions tracked the observed time series, with no significant over-predictions, except for Region 2. Figure 3 shows the mean bias of $PM_{2.5}$ predictions during winter. Only region 2 showed an over-prediction by all models with a mean bias of 5 to 9 µg/m^3. At other regions, the models typically under-predicted $PM_{2.5}$ with bias as much as -9 µg/m^3 in Region 7.

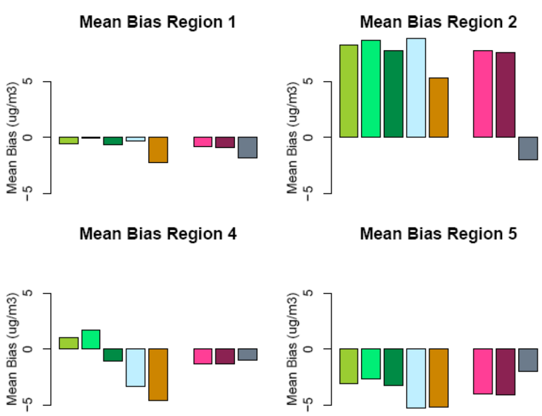

Figure 3. Mean Bias of $PM_{2.5}$ Predictions by 5-member ensemble system during December 2008 - February 2009. The 3rd member is NYSDEC_3x; rest is same as Figure 2.

RMSE ranged from 3.4 to 13 µg/m^3 with larger errors found in Region 2. Exceedances were observed only in Regions 1 and 2. As summarized in Table 3, 100% of the observed exceedances in region 2 were detected by all 5 modeling systems, but none in region 1. The modeling systems were also characterized by 86 to 90% FAR, in region 2 and 100% in region 1. Consequently, the overall CSI was only 10 to 14% in region 2, while being zero in region 1. The official DEC forecasts failed to capture the observed exceedances.

3.3. Retrospective Simulation

In addition to daily simulations with the 4 or 5-member system, retrospective simulations were performed with all 12 weather forecast members of SUNY-SB SREF. Figure 4 presents the mean bias of the 14-member system (12 SREF + 2 NCEP members) during June-July 2008. It is interesting to note that a majority of the MM5-based members showed a negative bias, while the WRF-based members showed a positive bias. Further, with the 14-member system, the ensemble average typically had lower bias than the individual model predictions. The ranges of the categorical metrics for the 14-member system were in general similar to or better than that of the 4-member system. For regions 1 and 2, the ensemble average showed 50-62% CSI, with 62-67% POD.

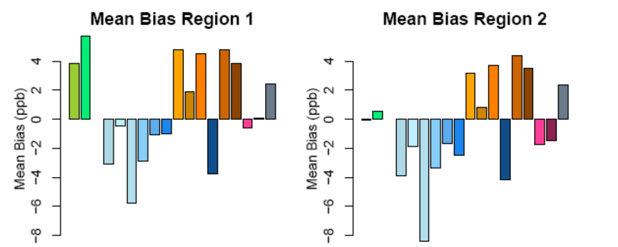

Figure 4. Mean Bias of Ozone Predictions by 14-Member Ensemble System during June-July 2008. From left to right, the NCEP_12z and 00z members are shown in green, the SUNY-SB SREF MM5 members in blue, the SUNY-SB SREF WRF members in orange, ensemble average in pink followed by median and finally, the official DEC forecasts in grey.

Similar to the 4-member system, $PM_{2.5}$ predictions were under-estimated in all regions except Region 2. Contrary to that observed for ozone, no distinction was found in the nature of the bias between the MM5- and WRF-based members. Within the MM5-based members, the 3 members using Mellor-Yamada (MY) scheme for planetary boundary layer predictions showed higher RMSE (18-20 µg/m^3) in Region 2 than the remaining 4 that did not (RMSE were 10-12 µg/m^3), although such differences were not found in other regions. Within WRF, only one member used non-MY scheme which showed similar RMSE (8 to 10 µg/m^3) to other WRF members. For the winter period (Dec 2008-Feb 2009), the 15-member system (12 SREF + 2 NCEP + NYSDEC_3x members) also showed similar performance as the 5-member system. Contrary to that noticed in summer, differences within the MM5 members were not pronounced.

Figure 5 shows time series of the ensemble mean and the absolute standard deviation of ozone predictions by the 14-member system during summer and $PM_{2.5}$ predictions during winter. The standard deviation is shown on the right axis. The absolute standard deviation often, but not always, appeared to increase with increase in concentration, suggesting that a higher absolute uncertainty may be associated with model-based forecasts of episodes. The relative standard deviation among the members provides an estimate of the uncertainty in model predictions arising from differences in model parameters. This was typically around 5 to 15% for ozone in summer, with higher values found in Regions 1 and 2. For $PM_{2.5}$ in winter, it varied widely from day to day, and was around 20 to 30% for Regions 1 and 2, and greater than 30% for other regions, in part due to low predicted concentrations.

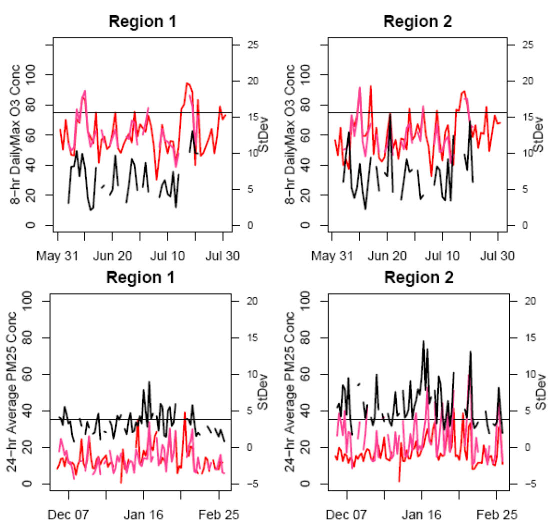

Figure 5. Time Series of observed (red) ensemble mean (pink) and absolute standard deviation (black, right y-axis) of O_3 during June-July 2008 (top) and $PM_{2.5}$ during Dec 2008 – Feb 2009 (bottom) by the 14/15 member system. The horizontal line refers to concentration that marks the transition of AQI to USG.

3.4. Summer 2009: Daily 6-member system

A preliminary analysis of the performance of the multi-model system is presented for the June – August 2009 period. This analysis also presents a comparison against the operational NOAA ozone forecasts that were made available to NYSDEC, which included the same-day forecast from the UTC 06z initialization cycle, and the next-day forecast from UTC 12z initialization. Note that no operational $PM_{2.5}$ forecasts were available from NOAA. Similarities and differences between the NCEP members and the NOAA members are the following: Both use the WRF-NMM meteorological fields. The "NCEP" members utilized those fields in the CMAQ simulation conducted at NYSDEC, while the NOAA members were the operational CMAQ-based forecast simulations conducted by NOAA. Both are based on the same map projection; however, the NOAA simulations are for a slightly larger domain, and hence an interior portion of the domain may be influenced by different boundary conditions. Finally, the NOAA runs utilize an updated emission inventory for the current forecast year, while the NCEP simulations utilize EPA's previous year forecast inventory. The mean bias (not shown) of ozone predictions by all members except the ASRC member ranged from 3.5 to 10 ppb. The ASRC CAMx member showed a larger bias, ~12-17 ppb. The NCEP members and the NOAA members showed similar bias characteristics, except for Regions 2, 5 and 7. The NOAA members showed slightly larger bias than the NCEP members in Region 2, while being lower than NCEP in other regions, probably resulting from differences in emissions. Categorical evaluation of ozone predictions showed that the members showed 50-85% FAR in Regions 1 and 2, in contrast to the 20-60% FAR in summer of 2008.

For the same time period, the $PM_{2.5}$ predictions showed a positive bias (> 2 $\mu g/m^3$) at most regions, in contrast to a negative bias found in summer 2008. Past experience has shown that the models typically under-predict $PM_{2.5}$ concentrations in summer in part due to missing pathways of secondary organic aerosol (SOA) formation. Thus, the over-prediction by the modeling systems may indicate any of the following: lower SOA formation in reality and/or overestimated emissions leading to an over-prediction.

The NYSDEC emissions inventory for the forecast system was developed from different sources. It consisted of the 2009 emission inventory projected from the 2002 base case inventory and developed for air quality planning purposes by the regional planning organizations (RPOs) for all sources except the electric generating units (EGU). For EGU (point), the inventory consisted of a 2005 planning inventory based on the actual 2005 continuous emissions monitoring (CEM) data. A preliminary comparison of the CEM data (*http://camddataandmaps.epa. gov/gdm/*) between 2008 and 2005 showed that the NO_x emissions decreased by an average of ~15% during the ozone season (May-Sep), and by ~20% on an annual emission basis between 2005 and 2008 in the northeast US. Consequently, the NYSDEC emission inventory used in the forecast simulations may be overestimating EGU emissions. In addition, it is possible that the economic recession contributed to decreased industrial activity. For example, a ~12% decline in truck traffic in NY was reported from Jan-June 2009 compared to the same time period in 2008 (Times Union, Albany, NY). In order to evaluate the effect of possible overestimated emissions, a sensitivity analysis was performed in which all anthropogenic emissions of all pollutants were reduced by 20% across the whole domain. The NYSDEC_3x member was rerun with the reduced emissions from August 7 to August 26, 2009, a period that included select days when daytime maximum temperatures were high (>90 °F in downstate NY) and high ozone episodes were observed in Regions 1 and 2. It was found that a 20% decrease in anthropogenic emissions resulted in a maximum of ~7% reduction in the predicted 8-hr daily maximum ozone concentrations in NY regions compared to the base case simulation (4.7 ppb in region 5 to 7.3 ppb in region 1). It reduced the region-wide mean bias by 2-3 ppb, and re-distributed the frequency of AQI categories closer to that of observations. The distribution of AQI frequencies was over-corrected for region 5, closer to observations for region 3, and unaffected for region 2. This may suggest that the percent reduction may be different – lower for rural region such as Region 5, but higher than 20% for an urban region.

As shown on the left panel of Figure 6, the over-prediction was present over the whole domain with normalized mean bias (NMB) typically greater 20-25%. A 20% reduction in emissions lowered the NMB to 10-20%. This analysis suggests that over-prediction in ozone concentrations this summer could be partly related to an over-estimated emissions inventory. This indicates the challenges associated with incorporating up-to-date emissions that are reflective of real-world activity in forecasting applications.

4. CONCLUSIONS

The performance of a multi-model air quality forecasting system during summer and winter periods was presented for the NY region. During summer 2008, the system as a whole appeared to capture the range of observed concentrations. For Long Island and New York City (NYC) regions, the SUNYSB_F2 (MM5-based) member appeared to perform better, while for other regions, the NCEP-based members had a lower bias. The overall success rate, as measured by the critical success index, ranged from 40 to 67%.

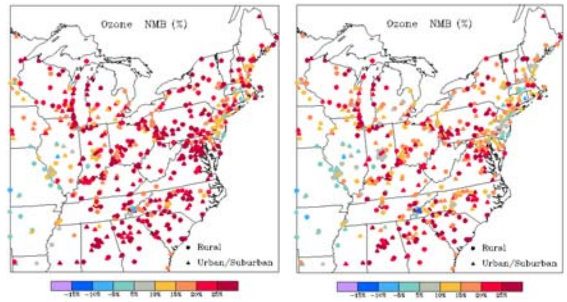

Figure 6. Normalized mean bias of O_3 predictions by NYSDEC_3x member before (left) and after (right) emissions reduction

$PM_{2.5}$ concentrations were under-predicted in most regions during summer 2008. During winter, the models over-predicted $PM_{2.5}$ in NYC, while under-predicting at other regions. Although the probability of detection was 100% for NYC, they were associated with higher rate of false alarms, resulting in an overall CSI of less than 15%. A retrospective simulation of 12-member SUNY-SB SREF presented similar results as the daily members. In this case, the ensemble average showed better performance than the individual models. An examination of the standard deviation between the model predictions indicated that it mostly, but not always, increased with increasing concentrations suggesting possible higher absolute uncertainty on certain episodic days. On a relative basis, this was equivalent to a 5 to 15% variability for ozone in summer, while being 20-30% or greater for $PM_{2.5}$ in winter. Future work will include a probabilistic verification of the system.

An analysis of model predictions for summer 2009 showed significant over-prediction of ozone and $PM_{2.5}$ in contrast to the previous summer. A sensitivity analysis of 20% reduction in emissions suggested that an over-estimated emissions inventory (arising from older electric generating unit emissions and possibly reduced activity associated with economic recession) might partly explain the over-predictions, although the extent of over-estimation may differ from the 20% factor.

Overall, the multi-model system captured the range of observed concentrations for ozone. For $PM_{2.5}$, while the system tracked the observations, it was biased. This may imply that more meteorological/ emission variations may be needed between the members.

5. DISCLAIMER

This work was funded in part by NYSDEC and the New York State Energy Research and Development Authority (NYSERDA) under agreement #10599. The results presented here have not been reviewed by the funding agencies. The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of NYSDEC or the sponsoring agency.

6. REFERENCES

Hogrefe, C. *et al.*, 2008. Multi-Model Air Quality Forecasting Over New York State for Summer 2008. Chapel Hill, NC.

Hogrefe, C. *et al.*, 2007. Daily Simulation of Ozone and Fine Particulates over New York State: Findings and Challenges. Journal of Applied Meteorology 46, 961-979.

Jones, M.S., Colle, B.A. and Tongue, J.S., 2007. Evaluation of a Mesoscale Short-Range Ensemble Forecast System over the Northeast United States. Weather and Forecasting 22, 36-55.

Kang, D. *et al.*, 2005. The New England Air Quality Forecasting Pilot Program: Development of an evaluation protocol and performance benchmark. Journal of the Air & Waste Management Association 55, 1782-1796.

NYSDEC, 2007. DEC Using More Stringent Measurement for Air Quality Forecasts.July 2007, Available online at http://www.dec.ny.gov/environmentdec/35375.html

Otte, T.L. *et al.*, 2005. Linking the Eta Model with the Community Multiscale Air Quality (CMAQ) Modeling System to Build a National Air Quality Forecasting System. Weather and Forecasting 20 367-384.

U.S. EPA, 2008. Revisions to EPA's ozone air quality index fact sheet. Available online at http://www.epa.gov/air/ozonepollution/pdfs/2008_03_aqi_changes.pdf

Yu, S. *et al.*, 2008. Evaluation of real-time $PM_{2.5}$ forecasts and process analysis for $PM_{2.5}$ formation over the eastern United States using the Eta-CMAQ forecast model during the 2004 ICARTT study. Journal of Geophysical Research 113.

Table 1. Summary of daily multi-model forecast system.

Name	Met. Model	Emiss. Inv.	Air Quality Model	Grid Res.	Initialization	Begin Date
NCEP_12z	NCEP WRF-NMM	EPA	CMAQ v4.6, CB4	12-km	12z	June 2005
NCEP_00z	NCEP WRF-NMM	EPA	CMAQ v4.6, CB4	12-km	00z	May 2008
NYSDEC_3x	NCEP WRF-NMM	NYSDEC	CMAQ v4.6, CB4	12-km	00z	November 2008
SUNYSB_F2	SUNY-SB MM5	NYSDEC	CMAQ v4.6, CB4	36-12 km	00z	June 2008
SUNYSB_F9	SUNY-SB WRF-ARW	NYSDEC	CMAQ v4.6, CB4	36-12 km	00z	June 2008
ASRC	ASRC WRF-ARW	NYSDEC	CAMx v4.5.1, CB05	12-km	00z	March 2009

Table 2. Categorical evaluation metrics for O_3 from Jun-Sep 2008. The members are: M1 - NCEP_12z, M2 - NCEP_00z, M3 - SUNYSB_F2, M4 - SUNYSB_F9. Avg refers to the ensemble mean, Med, the ensemble median, and DEC, the official DEC forecasts.

	Region 1							Region 2						
	M1	M2	M3	M4	Avg	Med	DEC	M1	M2	M3	M4	Avg	Med	DEC
POD (%)	60	80	80	60	50	60	50	50	75	58	58	67	50	58
FAR (%)	62	56	20	33	38	33	60	33	36	22	22	11	14	36
CSI (%)	30	40	67	46	38	46	29	40	53	50	50	62	46	44

Table 3. Categorical metrics for $PM_{2.5}$ from Dec 2008 - Feb 2009. The members are: M1 - NCEP_12z, M2 - NCEP_00z, M3 - NYSDEC_3x, M4 - SUNYSB_F2, M5 - SUNYSB_F9.

	Region 1								Region 2							
	M1	M2	M3	M4	M5	Avg	Med	DEC	M1	M2	M3	M4	M5	Avg	Med	DEC
POD (%)	0	0	0	0	0	0	0	0	100	100	100	100	100	100	100	0
FAR (%)	100	100	100	100	NA	100	NA	NA	89	90	88	89	86	88	88	NA
CSI (%)	0	0	0	0	0	0	0	0	11	10	12	11	14	12	12	0