

USE OF AIR QUALITY FORECASTING AS A DIAGNOSTIC TOOL OVER THE NORTHEASTERN US

Prakash Doraiswamy^{1,§,*}, Christian Hogrefe^{1,2}, Winston Hao², Kevin Civerolo², J.-Y. Ku², Robert Henry² and Gopal Sistla²

¹Atmospheric Sciences Research Center, University at Albany, Albany, NY

²Bureau of Air Quality Analysis and Research (BAQAR), New York State Department of Environmental Conservation (NYSDEC), Albany, NY

[§]On assignment to BAQAR, NYSDEC

1. INTRODUCTION

Typical air quality simulations involve modeling selected episodes retrospectively based on time periods of interest. This limits the ability of evaluating the performance of a model beyond the episodic periods or select years, and to identify issues that may not be apparent. Since June 2005, the research group at the Bureau of Air Quality Analysis and Research (BAQAR) at New York State Department of Environmental Conservation (NYSDEC) has been conducting daily air quality forecasting of ozone and fine particles (PM_{2.5}) using the Community Multi-scale Air Quality Model (CMAQ) on a pilot-scale, in collaboration with EPA and NOAA. These simulations have been (and continue to be) archived providing a dataset for varied air quality applications. As an example, this study focuses on the archived hourly concentration fields for isoprene during June through August 2005 and compares its performance to measured concentrations on a diurnal scale at locations in the northeastern part of the US. The overall objective is to assess the utility of the forecast based simulations as a diagnostic tool by using a combination of analyses on the response of the model to biogenic, anthropogenic and secondary species.

2. MODEL AND OBSERVATIONAL DATABASE

2.1. *Model Set-Up and Archived Database*

The data used in this study were from the archived air quality forecast simulations conducted

by NYSDEC for the summer of 2005. The modeling system used during that time period has been described elsewhere (Hogrefe et al. 2006; 2007). Briefly, it consisted of the ETA meteorological model, the PREMAQ (Otte et al. 2004; 2005) emissions and meteorology preprocessor, and the CMAQ (Byun and Ching 1999) photochemical model (v4.4), operated in a forecasting mode. The PREMAQ processor included BEIS3.12 for the biogenic emissions processing. The simulation was started on June 1, 2005, which was initialized using default conditions in CMAQ. Each CMAQ simulation was performed for 48 hours starting at 12:00 Greenwich Mean Time (GMT). Simulations for every day thereafter were initialized using modeled concentration fields from the previous day. Time-invariant boundary conditions were used. In this analysis, results for June 1-11, 2005 are excluded to eliminate the effects of initial conditions. The modeling domain covered almost the entire Eastern U.S with a 12 km horizontal grid resolution. All emission and meteorological inputs processed by PREMAQ, and the CMAQ-processed concentration fields of all model species for layer 1 (i.e., ground level surface layer, ~35 m height above ground level [agl]) and selected species for the lowest 14 layers (~2750 m agl) are archived regularly, among other custom analyses. The predicted concentration fields from these files are used in this analysis.

This study focuses on diurnal predictions of isoprene concentrations during June-August of 2005, the dominant time period in 2005 for which observational data were available for most sites.

2.2. *Observational Database*

Measured concentrations of isoprene were obtained from the EPA Air Quality System (AQS) for all monitors in the New England and Mid-Atlantic State region comprising of Maine, Vermont, New Hampshire, New York, Massachusetts, Connecticut, Rhode Island, Pennsylvania, New Jersey, Maryland, Delaware and District of Columbia. Table 1 lists the sites with hourly

* *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

isoprene measurements that were used in this analysis. In addition to isoprene, hourly concentrations of ethylene (anthropogenic and biogenic origin), ozone (product of photochemical reaction) and nitrogen dioxide (anthropogenic and chemical production) were also obtained, where available, for the same sites, to serve as supplemental information in the evaluation of model performance.

3. DATA ANALYSIS, RESULTS AND DISCUSSION

Hourly model predictions were compared with measured concentrations of isoprene. Various statistical measures of performance, including the correlation coefficient (r), root mean square error (RMSE), normalized mean gross error (NMGE) and normalized mean bias (NMB) were calculated for each site utilizing hourly data. In addition to the traditional statistical measures, diurnal concentrations averaged by hour for each month were calculated and plotted for both observed and model-predicted concentrations.

Table 2 lists the statistical summary for the 2.5-month period (June 12-August 31, 2005) for each site. Correlations for that period ranged from 0.25 to 0.63. Sites in NJ, PA, MD, DC and one site in CT showed a net over-prediction as shown by a positive NMB, while other sites showed a net under-prediction. While the results presented in Table 2 provide a good overall summary of the model behavior over the entire summer period, an examination of observed and simulated diurnal profiles is presented as further focus of this study. First, the correlation between hourly observed and simulated concentrations were calculated separately for each day utilizing all available data within each 24-hour period, and box plots were constructed displaying the distribution of these correlations at all sites (Figure 1). As seen, the distribution of daily correlations differed between sites. Some sites exhibit zero to negative correlations indicating poor to inverse agreement. Median daily correlations ranged from 0.24 to 0.72. The upper end of whiskers of all sites was typically above 0.8, indicating days of excellent agreement of diurnal profiles. Since the focus of the study is the Mid-Atlantic Region, the following sites are further examined in this study: an urban site in NY (SI:360050083) located at NY Botanical Gardens (NYBG), Bronx, NY, a suburban site in NJ (SI:340210005) located at Rider College (NJRC) and a rural site in PA (SI:420010001), which is a NARSTO site located at Arendtsville, PA (PANARSTO).

Figure 2 presents the distribution of the hourly bias (CMAQ-predicted minus Observed) at each of the three sites. At the NYBG, the predictions differ from observations during mid-day hours. The bias (negative, implying an under-prediction) starts to deviate from zero at about 8 am local standard time, reaches a maximum around 2 pm, and then decreases back to zero around 7 pm. The suburban NJRC site showed zero bias until 5am, then a constant positive bias until 5 pm, and a peak between 6 pm to 9 pm. The rural PANARSTO site also shows similar pattern, with zero bias till 5 am, then near-zero bias until 1 pm, followed by a pronounced positive bias between 5 pm and 9 pm. Figure 3 (left panel) shows the corresponding average diurnal plots for the three sites. The urban and suburban sites show a pronounced dual peak in the predicted concentrations, which is present in measured concentrations at the suburban site, but not at the urban site. The rural site shows a less pronounced morning peak, while still possessing the evening peak. The questions of interest are: why do we see dual peaks in predicted concentration at an urban site, when the measurements do not; and is it possible to identify causes for agreement/disagreement seen at the sites?

The answers to the above questions lie in analyzing the possible sources and sinks of isoprene, which will determine the ultimate shape of the profile. To aid in this analysis, CMAQ-predicted nitrogen dioxide (NO_2) (top half of each chart in left panel, Figure 3), ozone (O_3) (middle panel) and ethylene (C_2H_4) (right panel) concentrations were also compared to observations at these sites. Model-predicted HO_2 radical concentrations are also plotted as an indicator of products of isoprene reactions. In addition, estimated planetary boundary layer (PBL) heights were analyzed. All these plots are displayed in Figure 3. Based on these plots, the following hypotheses can be made: For the NYBG site, the NO_2 concentrations are over-predicted, which is consistent with predicted night-time ozone concentrations being lower than observed (due to scavenging by oxides of nitrogen, NO_x). It is possible that the over-prediction in NO_2 concentrations (possibly resulting from overestimation of emissions of NO_x) resulted in more OH radical formation which consumed isoprene. For the NJRC and PANARSTO sites, it could be hypothesized that the emission rate of isoprene is overestimated. The predicted ethylene concentrations (plotted on right y-axis, right panel of Figure 3), while being overestimated by more than 2 times, have a profile similar to observed

concentrations, except at PANARSTO between midnight and 5 am.

To further test the hypotheses, the time period under consideration was re-run using CMAQv4.6 and enabling the process analysis feature. Figure 4 presents the diurnal average of integrated process rates (IPR) for the first layer (surface) at NYBG as reported by the process analysis module. While the diurnal profile of the predicted concentrations varied between the three sites, the IPR profiles were similar at all three sites. For the surface layer, the dominating processes contributing to an increase in the isoprene concentrations are emissions followed by horizontal advection, while those contributing to a depletion of isoprene are losses by diffusion in the vertical direction followed by consumption by chemical reactions. From this, it is obvious that the resultant profile is primarily a balance between the emissions, advection, chemistry and vertical diffusion processes. The previously shown differences in predicted versus observed diurnal profiles at NYBG, that is, a depression in predicted profiles during mid-day hours, could imply that the predictions are either overestimating the chemical or diffusion losses or underestimating additional sources. For upper layers (not shown), chemical reactions appear to be the major and dominant pathway for depletion of isoprene, while the source of isoprene was diffusion from lower layers. Although the chemistry effect is smaller than the diffusion loss, it is potentially large enough (~ -0.5 ppb/hr) to partially explain the differences between observed and simulated diurnal profiles. For the other two sites, the process analyses confirmed emissions to be the major source of isoprene. Given the reasonable agreement in the shape of the predicted versus observed diurnal profiles of ozone and ethylene during mid-day hours at these two sites, it appears that the mixing processes (i.e., PBL height, vertical mixing) are modeled satisfactorily, and could not be solely responsible for the over-prediction of isoprene. It is likely that the emissions are overestimated, resulting in a positive bias. However, measurements at the individual process level (emissions, isoprene flux, etc.) would be needed to corroborate these findings.

The predicted peak in the evening is around 8-9 pm for the urban site, while it is around 6pm for the suburban and rural sites. At NYBG, an analysis of the IPRs, shows that horizontal advection of isoprene is the dominant process that contributes to the addition of isoprene in upper layers, and approximately 40% of the source term at the surface layer at that 8-9 PM. However, at

NJRC and PANARSTO, while horizontal advection contributed to addition of isoprene, it was not the major source at 6 PM when the peak was observed. Although process analyses parameters for ethylene were not computed, the fact that the shape of the predicted diurnal profile of ethylene closely resembles measured profiles suggests that the horizontal advection process is treated adequately, and is unlikely to be the sole cause of the secondary peak in CMAQ-predicted isoprene profiles, which is either absent (at NYBG) or less pronounced (at NJRC and PANARSTO) in observations. At NYBG, this could mean a slight underestimation of night-time isoprene consumption in reactions. The physical and chemical mechanisms that result in loss of isoprene at night are highly uncertain and have been the subject of extensive research (Faloon et al. 2001; Goldan et al. 1995; Sillman et al. 2002; Starn et al. 1998). In CMAQ, the reaction between isoprene and nitrate radical is the only major reaction that contributes to nighttime depletion of isoprene. This in turn could imply an underestimation of production of nitrate radical concentrations. Starn et al. (1998) found that the rapid decrease in observed isoprene concentrations was associated with conditions when the product, $[O_3][NO_2] > 300$ ppb². At NYBG, while the observed $[O_3][NO_2]$ product dropped from 657 to 382 ppb² between 7 pm and 11 pm, the CMAQ predicted value was 395 ppb² at 7 pm, which dropped to 152-145 ppb² between 8 pm and 11pm. This is consistent with the theory that overestimated NO emissions resulted in scavenging of ozone, and over-prediction of NO₂. The cause of the peak at the other two sites could be related to a combination of overestimated isoprene emissions and possibly lack of surface emissions/sinks; however more research is needed to verify this.

4. CONCLUSIONS

Analyses of archived air quality forecasting simulations present a unique opportunity to evaluate the model under varied atmospheric conditions. This study presented an example of isoprene analysis based on a combination of techniques. Overall, the model appeared to track the diurnal profile as demonstrated by median correlations typically greater than 0.5 at most sites. Causes of model performance discrepancies do not seem to be universal, but appear to differ by site. The analyses revealed possible overestimation of isoprene emissions at NJRC and PANARSTO sites, and an over-prediction of NO_x emissions at NYBG site. Process analysis enabled confirming

initial hypotheses derived from analyzing isoprene and other supporting species.

5. DISCLAIMER

This research was funded in part by the U.S. Environmental Protection Agency (EPA) and the National Oceanic Atmospheric Administration (NOAA). 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.

6. REFERENCES

Byun, D. W. and J. K. S. Ching (1999). Science algorithms of the EPA MODELS-3 community multiscale air quality (CMAQ) modeling system. Washington, DC, Office of Research and Development, U.S. Environmental Protection Agency.

Faloona, I., D. Tan, W. Brune, J. Hurst, D. Barket, T. L. Couch, P. Shepson, E. Apel, D. Riemer, T. Thornberry, M. A. Carroll, S. Sillman, G. J. Keeler, J. Sagady, D. Hooper and K. Paterson (2001). "Nighttime observations of anomalously high levels of hydroxyl radicals above a deciduous forest canopy." *Journal of Geophysical Research* **106**(D20): 24315-24333.

Goldan, P. D., W. C. Kuster, F. C. Fehsenfeld and S. A. Montzka (1995). "Hydrocarbon measurements in the southern United States: The Rural Oxidants in the Southern Environment (ROSE) Program 1990." *Journal of Geophysical Research* **100**(25): 945-963.

Hogrefe, C., W. Hao, K. Civerolo, J.-Y. Ku and G. Sistla (2006). Exploring Approaches to Integrate Observations and CMAQ Simulations for Improved Air Quality Forecasts. *Models-3 Users' Workshop*. Chapel Hill, NC.

Hogrefe, C., W. Hao, K. Civerolo, J.-Y. Ku, G. Sistla, R. S. Gaza, L. Sedefian, K. Schere, A. Gilliland and R. Mathur (2007). "Daily Simulation of Ozone and Fine Particulates over New York State: Findings and Challenges." *Journal of Applied Meteorology* **46**(NOAA/EPA Golden Jubilee Special Issue): 961-979.

Otte, T. L., G. Pouliot and J. E. Pleim (2004). PREMAQ: A New Pre-Processor to CMAQ for Air Quality Forecasting. *Models-3 Conference*. Chapel Hill, NC.

Otte, T. L., G. Pouliot, J. E. Pleim, J. O. Young, K. L. Schere, D. C. Wong, P. C. S. Lee, M. Tsidulko,

J. T. McQueen, P. Davidson, R. Mathur, H. Y. Chuang, G. DiMego and N. L. Seaman (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** (3): 367-384.

Sillman, S., M. A. Carroll, T. Thornberry, B. K. Lamb, H. Westberg, W. H. Brune, I. Faloona, D. Tan, P. B. Shepson, A. L. Sumner, D. R. Hastie, C. M. Mihele, E. C. Apel, D. D. Riemer and R. G. Zika (2002). "Loss of isoprene and sources of nighttime OH radicals at a rural site in the United States: Results from photochemical models." *Journal of Geophysical Research* **107**(D5).

Starn, T. K., P. B. Shepson, S. B. Bertman, D. D. Riemer, R. G. Zika and K. Olszyna (1998). "Nighttime isoprene chemistry at an urban-impacted forest site." *Journal of Geophysical Research* **103**(D17): 22437-22448.

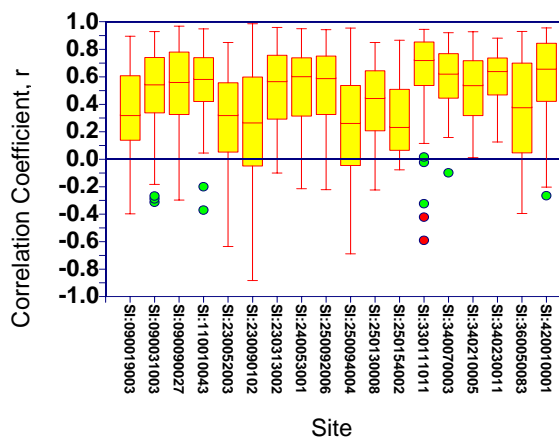


Figure 1. Daily Correlation Coefficients by Site during June 12-August 31, 2005. Only days with valid data for 18 hours or more (i.e., 75% of 24 hours) are included in the above plot. The box indicates the 25th, 50th and 75th percentiles. The upper (or lower) whiskers represent the largest (or the lowest) observation that is less than or equal to the 75th percentile plus 1.5 times the inter-quartile range [IQR] (or greater than or equal to the 25th percentile minus 1.5 times IQR). Data that fall outside the whiskers are outliers. Green dots are mild outliers (within 3 times IQR from the end of whiskers), while red dots are "severe" outliers (> 3 times IQR from end of whiskers).

Table 1. Sites Used in Model Evaluation

State	County	Site ID	Latitude (degrees)	Longitude (degrees)	Site Description	Land Use	Location Setting
Connecticut	Fairfield	090019003	41.1183	-73.3367	Sherwood Island State Park	Forest	Rural
Connecticut	Hartford	090031003	41.7847	-72.6317	McAuliffe Park	Residential	Suburban
Connecticut	New Haven	090090027	41.3011	-72.9028	1, James Street	Commercial	Urban city center
<i>District of Columbia</i>		110010043	38.9189	-77.0125	S.E. End McMillian Reservoir	Commercial	Urban city center
Maine	Cumberland	230052003	43.5608	-70.2078	Two Lights State Park	Residential	Rural
Maine	Hancock	230090102	44.3517	-68.2272	Top of Cadillac Mountain	Mobile	Rural
Maine	York	230313002	43.0833	-70.7500	Frisbee School, Goodsoe Rd	Residential	Suburban
Maryland	Baltimore	240053001	39.3108	-76.4744	Woodward and Franklin Roads, Essex	Residential	Suburban
Massachusetts	Essex	250092006	42.4744	-70.9725	390, Parkland	Commercial	Urban city center
Massachusetts	Essex	250094004	42.7894	-70.8092	Sunset Blvd	Residential	Suburban
Massachusetts	Hampden	250130008	42.1945	-72.5557	Anderson Rd, AFB	Commercial	Suburban
Massachusetts	Hampshire	250154002	42.2983	-72.3347	Quabbin Summit	Forest	Rural
New Hampshire	Hillsborough	330111011	42.7204	-71.5231	Gilson Road	Residential	Suburban
New Jersey	Camden	340070003	39.9228	-75.0972	Copewood & E. Davis Streets, Trailer	Residential	Suburban
New Jersey	Mercer	340210005	40.2828	-74.7467	Rider College, Lawrence Township	Residential	Suburban
New Jersey	Middlesex	340230011	40.4619	-74.4298	R.U. Veg Research Farm, 3 Ryders Ln	Agricultural	Rural
New York	Bronx	360050083	40.8659	-73.8808	200th Street and Southern Blvd	Commercial	Urban city center
Pennsylvania	Adams	420010001	39.9200	-77.3100	NARSTO Site, Arendtsville	Residential	Rural

Table 2. Statistical Summary of Model Performance for Isoprene Based On Hourly Data from June 12 – August 31, 2005*

Site ID	No. of Pairs	Obs. Mean (ppb)	Pred. Mean (ppb)	Correlation, r	RMSE ^a (ppb)	No. of Pairs, where Obs. > 0	NMGE ^b (%)	NMB ^c (%)
090019003	1616	0.477	0.231	0.446	0.718	1334	75.7	-53.1
090031003	1729	0.495	0.263	0.433	0.589	1714	73.8	-46.9
090090027	1865	0.260	0.284	0.525	0.433	1691	82.6	7.7
110010043	1520	0.402	0.959	0.404	1.210	1348	165.3	128.4
230052003	1831	0.183	0.219	0.255	0.516	1113	102.9	-11.6
230090102	1630	0.311	0.106	0.318	0.420	1337	81.2	-67.4
230313002	1602	0.435	0.383	0.441	0.644	1437	76.1	-13.2
240053001	803	0.429	0.524	0.544	0.575	794	75.1	22.0
250092006	1454	1.001	0.468	0.418	1.158	1428	72.1	-53.3
250094004	1606	0.491	0.196	0.351	0.805	1501	85.6	-60.2
250130008	1625	0.818	0.446	0.393	0.917	1615	71.5	-45.5
250154002	824	1.989	0.526	0.380	2.943	820	79.3	-73.6
330111011	1515	0.885	0.765	0.580	1.022	1504	63.8	-13.9
340070003	1810	0.397	1.180	0.524	1.323	1803	214.3	196.9
340210005	1796	0.439	0.697	0.502	0.716	1783	102.8	58.4
340230011	1583	0.783	1.141	0.542	1.236	1575	88.4	45.3
360050083	1526	0.769	0.300	0.272	0.922	1525	78.2	-61.0
420010001	1722	0.479	1.370	0.634	1.995	1644	211.2	185.2

* Model predictions were not available for June 1, 2005. June 2-11, 2005 were not considered to avoid effect of initial conditions. Note that there were a total of 1944 possible hours. Sites with number of pairs less than 1458 imply fewer than 75% of total possible dataset.

^a RMSE: Root Mean Square Error = $\sqrt{\frac{1}{n} \sum_{i=1}^n (C_{pred} - C_{obs})^2}$

^b NMGE: Normalized Mean Gross Error = $\frac{100 * \frac{1}{n} \sum ABS(C_{pred} - C_{obs})}{\frac{1}{n} \sum C_{obs}}$

^c NMB: Normalized Mean Bias = $\frac{100 * \frac{1}{n} \sum (C_{pred} - C_{obs})}{\frac{1}{n} \sum C_{obs}}$

where, C_{pred} is the CMAQ predicted concentration; C_{obs} is the measured concentration; ABS is the absolute value of the term; and n is the number of pairs of valid data for RMSE, while for NMGE and NMB, it is the number of pairs of valid data for which $C_{obs} > 0$.

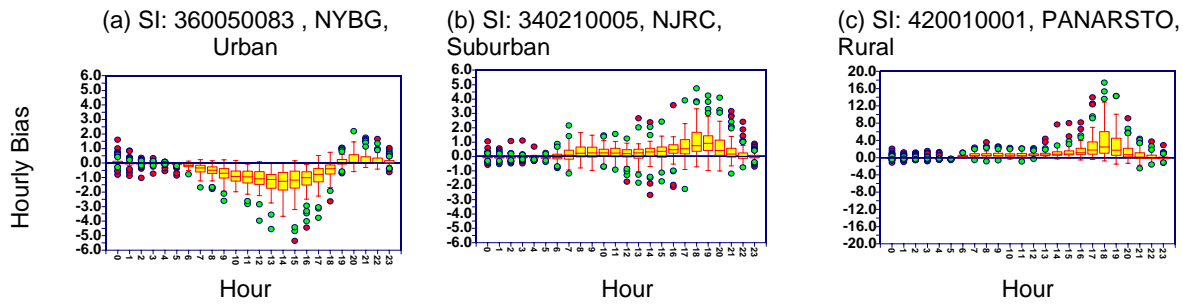


Figure 2. Hourly Bias (Predicted minus Observed) of Isoprene at Selected Sites during June 12-August 31, 2005

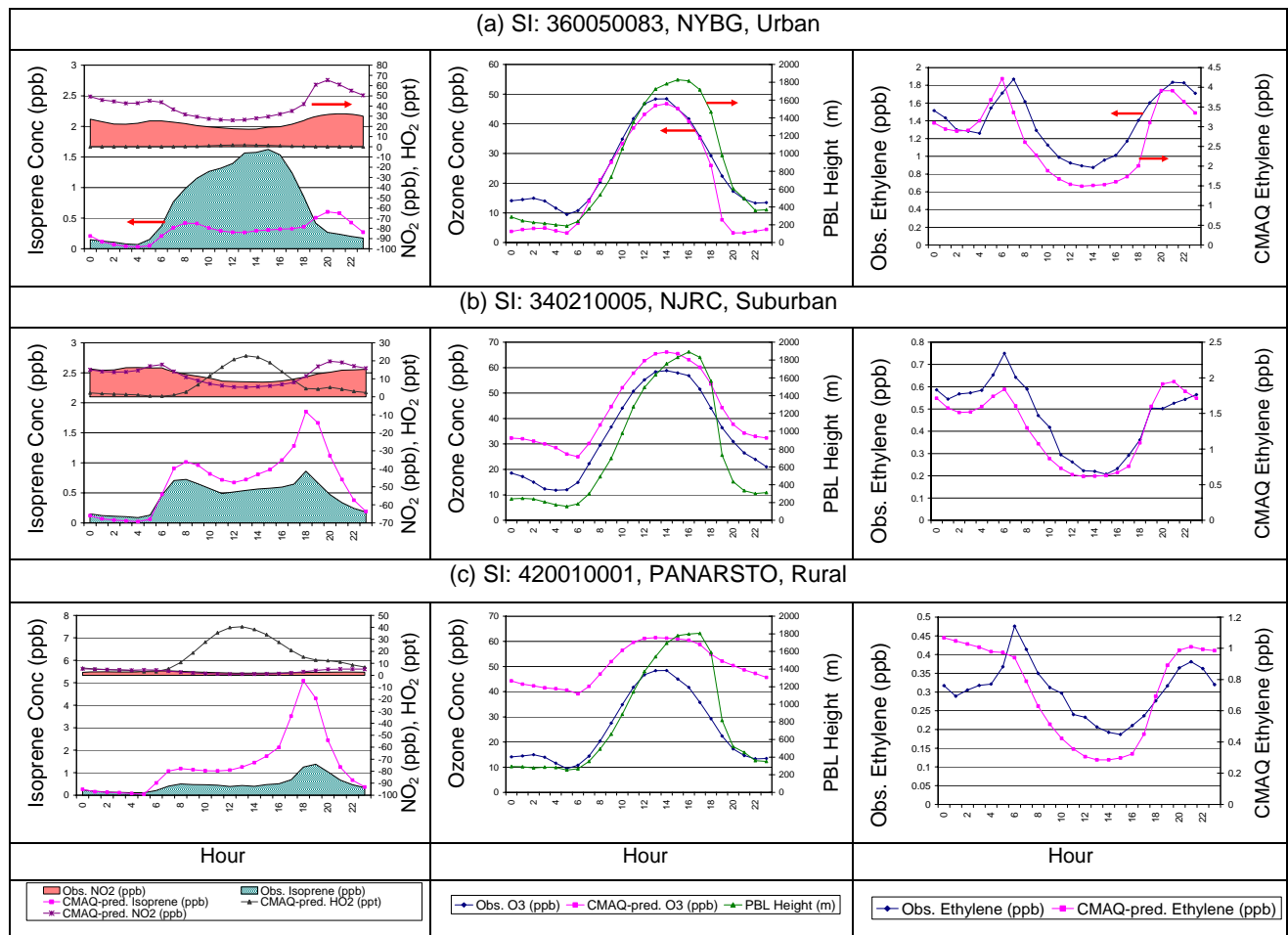


Figure 3. Average Diurnal Profile of Isoprene and Related Species during June 12-August 31, 2005. Note that certain species/parameters are plotted on the right y-axis, as indicated in the axis caption.

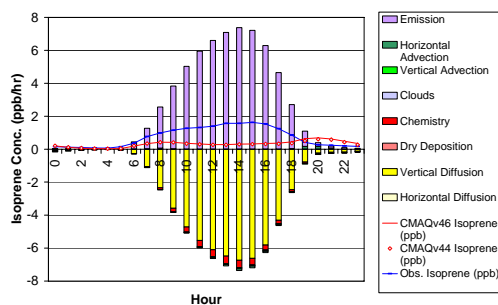


Figure 4. Diurnal Average of Integrated Process Rates Contributing to Predicted Isoprene Concentrations at NYBG during June 12 - August 31, 2005