

The Use of Global and Mesoscale Meteorological Model Data to Predict the Transport and Dispersion of Tracer Plumes over Washington, D.C.

ROLAND R. DRAXLER

National Oceanic and Atmospheric Administration/Air Resources Laboratory, Silver Spring, Maryland

(Manuscript received 11 August 2005, in final form 8 November 2005)

ABSTRACT

The data from a yearlong tracer dispersion experiment over Washington, D.C., in 1984 were used to evaluate Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) dispersion model calculations using coarse global meteorological reanalysis data [NCEP–NCAR and 40-Yr ECMWF Re-Analysis (ERA-40)] and calculations using meteorological data fields created by running a high-resolution meteorological model [fifth-generation Pennsylvania State University–NCAR Mesoscale Model (MM5)]. None of the meteorological models were optimized for urban environments. The dispersion calculation using the ERA-40 data showed better performance than those using the NCEP–NCAR data and comparable performance to those using MM5 data fields. Calculations with MM5 data that used shorter-period forecasts were superior to calculations that used forecast data that extended beyond 24 h. Daytime dispersion model calculations using the MM5 data showed an underprediction bias not evident in calculations using the ERA-40 data or for nighttime calculations using either meteorological dataset. It was found that small changes in the wind direction for all meteorological model data resulted in dramatic improvements in dispersion model performance. All meteorological data modeled plume directions were biased 10°–20° clockwise to the measured plume direction. This bias was greatest when using the global meteorological data. A detailed analysis of the wind observations during the November intensive, which had the greatest difference between the model and measured plume directions, showed that only the very lowest level of observed winds could account for the transport direction of the measured plume. In the Northern Hemisphere, winds tend to turn clockwise with height resulting in positive directional transport bias if the lowest-level winds are not represented in sufficient detail by the meteorological model.

1. Introduction

There has been increasing interest in providing more accurate pollutant plume dispersion predictions at distances of a few to tens of kilometers from the source. At this range plumes interact with the local environment in more complex ways that limit the traditional approach of using a single meteorological observation near the pollutant source location. At the same time the plumes are not yet large enough to span the domain of even a single grid cell from routine global or even regional forecast model outputs. The interpretation of dispersion model plumes, their interaction with the local environment, and the appropriateness of regional model guidance are starting to become part of the routine

duties of the local meteorological forecaster. Current National Weather Service instructions to their local forecast offices on providing non-weather-related emergency products (NWS 2004) outline in some detail the dispersion model products that are routinely available or could be requested and for which the local forecast office would be expected to provide interpretation. Online dispersion model training is now available for NWS forecasters (information online at http://meted.ucar.edu/dispersion/cam_hys/noflash.htm).

The goal of this analysis is to use some historical experimental dispersion data appropriate over an urban area to evaluate how well routine meteorological model products in conjunction with a dispersion model could simulate pollutant releases. There are many different dispersion models that can be used to evaluate experimental tracer data. Many countries with a national meteorological service have some dispersion modeling capability available as demonstrated by the 28 models that participated in the European Tracer Ex-

Corresponding author address: Roland R. Draxler, National Oceanic and Atmospheric Administration/Air Resources Laboratory, R/ARL-SSMC3, Rm. 3350, 1315 East–West Highway, Silver Spring, MD 20910.
E-mail: roland.draxler@noaa.gov

periment (Graziani et al. 1998). Chang et al. (2003) compared the performance of several of the models supported by the U.S. Environmental Protection Agency and U.S. Department of Defense to the data collected from a series of tracer experiments in Utah and Nevada and they found that the dispersion model's performance was highly dependent upon the method used to interpolate the meteorological observations. However, special meteorological observations may not always be available in emergency response applications. The Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT; Draxler and Hess 1998) model is currently being used by the NWS to support dispersion predictions for large-scale nuclear incidents (Draxler et al. 1997), volcanic eruptions, wild fire smoke transport, and for emergency response accidents, using routine NWS forecast model products. The NWS guidance document suggests that HYSPLIT could be applied at distances of 10 km and greater from a pollutant source. The model's parameterizations are not limited to just long-range simulations. Typically, the prediction of any dispersion model that uses an external source for its meteorological data is limited by the assumptions made in deriving those data, for example, how the meteorological model's resolution and definition of the local terrain features influence the calculation of the low-level wind field.

Issues surrounding urban-scale dispersion are again of great interest and experiments have recently been conducted in New York City, New York; Oklahoma City, Oklahoma; and Salt Lake City, Utah (SLC). The SLC experiments have been the subject of several model evaluation studies (Chang et al. 2005; Warner et al. 2004). Although they were conducted in the central urban area, the SLC experiments consisted of only a few tracer releases with the most distant sampling arc 6 km downwind. Generally, these recent experiments are limited to a few releases, during carefully selected scenarios, determined by criteria such as no rain, steady moderate winds, and certain wind directions. In terms of experimental data, there is only one experiment that contains a sufficient number of tracer releases to cover a broad range of weather situations. The Metropolitan Tracer Experiment (METREX; Draxler 1987) was conducted for more than 1 yr in 1984 around Washington, D.C., and its suburbs and consisted of hundreds of tracer releases with air samples collected from a few kilometers to about 75 km from the tracer source locations.

During the time METREX was conducted, few meteorological models were available that could be applied at a scale of less than 10 km. Computational resources today easily permit the METREX period to be

rerun with more modern mesoscale meteorological models. Initial and boundary conditions (ICs and BCs) for a mesoscale model can easily be obtained from the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR; Kalnay et al. 1996) or the European Centre for Medium-Range Weather Forecasts (ECMWF; Kållberg et al. 2004) reanalysis projects. Although improved urbanized mesoscale models are under development (Otte et al. 2004), for this study the standard version of the fifth-generation Pennsylvania State University–NCAR Mesoscale Model (MM5; Grell et al. 1994) was run to provide data fields at a horizontal resolution down to 4 km. Although the global meteorological data are based upon analyses, the resulting high-resolution MM5 simulations are in effect very short forecasts and are very comparable to what would be available in a real emergency situation today.

Evaluating the meteorological and dispersion components of a calculation requires model evaluation tools that work well with air concentration data and their inherent uncertainties. Chang and Hanna (2004) provide a detailed review of the advantages and disadvantages of various statistical approaches to dispersion model evaluation. Dispersion model simulations have been run for the entire year of 1984 using different meteorological data. Statistical results have been used to draw some initial conclusions about the meteorological and dispersion models' performance. These statistical analyses will be further refined to enhance the differences between the simulations. Finally, a detailed analysis has been conducted for one of the intensive experiments, in which there are sufficient measurements to show the plume in the sampling network, providing a visual representation of the conclusions derived from the statistical analysis.

2. Tracer experimental data: METREX

The tracer experiment started in December of 1983 and ran through December 1984. Two inert perfluorocarbon [100 g h^{-1} perfluoromonomethylcyclohexane (C_7F_{14}), PMCH; 300 g h^{-1} perfluordimethylcyclohexane (C_8F_{16}), PDCH] tracers (PFTs) were released for 6-h durations from two locations simultaneously about 20 km outside of Washington, D.C., at regular 36-h intervals alternating between nighttime (starting at 0300 UTC) and daytime releases (starting at 1500 UTC). Continuous air samples were collected as 8-h averages at one urban and two suburban sites and as 30-day averages (samples were changed the first of each month) at 93 sites throughout the region. Meteorological measurements of wind and temperature were made

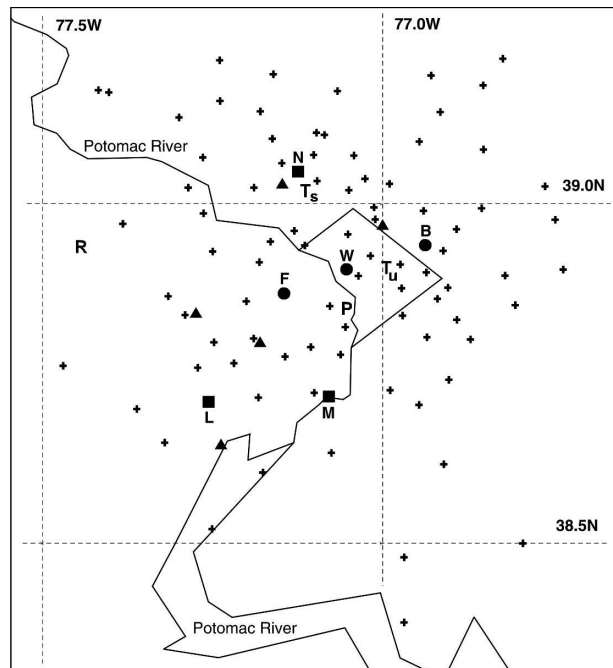


FIG. 1. The locations of the tracer release (squares) at Lorton, VA (L), Mount Vernon, VA (M), and the NOAA building in Rockville, MD (N); 8-h sampling (circles) in Fairfax, VA (F), Washington, DC (W), and Bladensburg, MD (B); 30-day sampling (pluses); meteorological towers (triangles); urban (u) and suburban (s) tethered (T) sites; and the pibal (P) and rawinsonde (R) sites. The approximate distance between the 8-h sampling locations is 10 km. The solid lines show the Potomac River and the boundaries of the District of Columbia.

at two levels (about 10 and 60 m) on five existing towers instrumented for this experiment. Once each month starting in April 1984, about 4–5 kg of a third PFT was released over 4 h at various locations depending upon the forecast wind direction. The tracer was collected and analyzed in the 30-day network so that any of the third PFT measured that month would have been from that one release. A general summary of the experiment and results can be found in Draxler (1987) and a detailed description is available in Draxler (1985).

The METREX tracer release and air sampling network is shown in Fig. 1. The PMCH was released from the northern location (N—Rockville, Maryland) through the end of May 1984 and then moved to the southwest location (L—Lorton, Virginia) from June through December 1984. The southeast release location (M—Mount Vernon, Virginia) released PDCH for the entire period. All releases were near ground level, except at Rockville, which was made from the rooftop of a National Oceanic and Atmospheric Administration (NOAA) office building about 30 m above ground level. The sequential 8-h average air samples were col-

lected at three locations in the middle of the sampling and release network. The central location was in the downtown Washington, D.C., urban area (W), while the two other sites (F and B) were in the suburbs. The 30-day samplers (pluses) were located at NWS cooperative weather observer sites.

Because tracer releases occurred on a regular schedule, regardless of whether the wind direction was expected to carry the tracer over the 8-h sampling network, over 50% of the 8-h samples showed no tracer concentration and about 10% of the samples had significant amounts of tracer. The remaining 40% of the samples showed very low tracer amounts. Although ambient background concentrations had been subtracted from the published measured PMCH and PDCH air concentrations, at low concentration levels, the measured data may still have considerable uncertainty due to various uncontrolled factors in the sampling and analysis procedures. As was done in previous analyses of these data, the 8-h air concentration values less than 30 pg m^{-3} for PMCH and 150 pg m^{-3} for PDCH are considered to be zero ($\text{pg} = \text{picograms}$). The remaining nonzero concentrations represent 10% of the samples. Due to the longer averaging times, all 30-day concentration values were used in the analysis as reported. Duplicate samples collected during the experiment showed that 95% of the 8-h duplicate samples were within 50% of each other and 85% of the 30-day duplicate samples were within 50% of each other.

3. Dispersion model and meteorological data

a. Overview of HYSPLIT

In HYSPLIT, the computation is composed of three components: particle transport by the mean wind, a turbulent transport component, and the computation of air concentration. Pollutant particles are released at the source location and passively follow the wind. The mean particle trajectory is the integration of the particle position vector in space and time. The turbulent component of the motion defines the dispersion of the pollutant cloud and it is computed by adding a random component to the mean advection velocity in each of the three-dimensional wind component directions. The vertical turbulence is computed from the wind and temperature profiles and the horizontal turbulence is computed from the velocity deformation. Air concentrations are computed by summing each particle's mass as it passes over the concentration grid. The concentration grid is treated as a matrix of cells, each with a volume defined by its dimensions. At a minimum, the model requires gridded three-dimensional fields of the vector wind components and temperature, which are linearly

interpolated in space and time to the pollutant particle's position. Additional parameters for the mixing computation are computed internally if not provided in the input meteorological data file. Detailed descriptions of the model can be found in Draxler and Hess (1997, 1998). No special model modifications were made to the model for this study to account for the urban environment.

b. Coarse-grid global meteorological data

Because of the historical nature of the METREX data, only two meteorological datasets were freely available to support the analysis. One was the NCEP–NCAR reanalysis, obtained from the National Oceanic Atmospheric Administration's Climate Diagnostics Center (information available online at <http://www.cdc.noaa.gov>) and the ECMWF's 40-Yr Re-Analysis (ERA-40) obtained from their data server (online at <http://data.ecmwf.int/data>). Reanalysis data are created in a series of steps in which observations are combined with a short-range forecast from the previous analysis in a way to minimize the deviations between the observations and the short-range forecast. For both the NCEP–NCAR and ERA-40 data, model output data fields were available on a 2.5° grid at 6-h intervals (synoptic times) on pressure surfaces. Key variables in the NCEP–NCAR data were used in HYSPLIT at 2 and 10 m and 1000, 925, 850, 700, 600, and 500 hPa. The ERA-40 contained the same data plus an additional level at 775 hPa. The ERA-40 reanalysis was created by a 60-level spectral model (T159L60), which corresponds to a horizontal grid spacing of about 100 km. The NCEP–NCAR reanalysis was created by a 28-level spectral model (T62L28), which corresponds to a horizontal grid spacing of about 200 km.

c. High-resolution meteorological data

To generate the high-resolution meteorological fields, MM5 (version 3–6) was run in a series of 36-h forecasts using the ERA-40 data for initial and boundary conditions. MM5 was configured with four nested grids (108, 36, 12, and 4 km), each with 34 levels and where the lowest layer has a depth of about 38 m. The 108 km grid was centered at 39°N, 80°W. The finest nested grid, a 52 by 52 gridpoint 4-km grid, was centered at about 39°N, 77.25°W. All grids were one-way nested except the 36-km grid, which was configured as a two-way nest. The data fields from the four grids were output every 3 h, 2 h, 1 h, and 30 min from the 108- to 4-km grids, respectively.

The 36-h simulations were restarted every day at 0000 UTC. Twenty-four forecast hours were extracted

from each of the daily 36-h MM5 output files, one set including and the other excluding the first 12 h, which were then packed and reformatted into monthly data files. The two sets of data files represent a short (to +24 h) or slightly longer duration (+12 to +24 h) forecast.

The version of MM5 used here is the one available from NCAR with only one urban land-use category and it has not been optimized for urban environments (e.g., Otte et al. 2004). The other physics options for the simulation were set to Kain–Fritsch for the cumulus parameterization, the Dudhia simple ice scheme for explicit moisture, the Eta Mellor–Yamada boundary layer scheme, the cloud–radiation scheme and the standard five-layer soil model. See Grell et al. (1994) for more detailed explanations of these parameterizations and other model options.

4. Statistical evaluation methods for HYSPLIT

Procedures for evaluating dispersion model calculations have a long history (Fox 1984; Hanna 1989, 1993; Chang and Hanna 2004). The problem eludes simple solutions because the variability in atmospheric motions cannot be deterministically represented in any model, resulting in the inevitable mismatches between predicted and measured concentrations paired in space and time. In some respects the METREX evaluation is a little easier because the relatively long duration samples (8 h and 30 days) average some of the variability.

Virtually all of the previously mentioned evaluation procedures were designed for shorter-range experiments conducted under well-controlled conditions such that the tracer was released only when certain meteorological criteria were satisfied. Typically in these experiments, many measurements were made in a single observed plume. In contrast, during METREX, the 8-h average concentrations are composed mostly of near-zero values with typically only one or two samplers showing a tracer plume if the wind directions were favorable for the preceding release.

The dispersion model evaluation protocol used here follows the procedures used by Mosca et al. (1998) and Stohl et al. (1998). However, only five statistical parameters were selected from their broad list to represent well-defined evaluation categories. Each of these is defined in the appendix. Both Mosca et al. (1998) and Stohl et al. (1998) recognized the problem in dealing with the uncertainties of “near background” measurement data and avoiding statistical parameters that may be too sensitive to small variations in the measurement values such as ratios between measured and calculated concentration. For a quick evaluation comparison, it is

TABLE 1. Results for the 8-h and 30-day sampling networks from the HYSPLIT calculation using various meteorological input data. The correlation coefficient R , fractional bias (FB), figure-of-merit in space (FMS; %), percent within a factor of five (5X; %), and the Kolomogorov–Smirnov parameter (KSP; %) are defined in the appendix. An asterisk alongside R or FB indicates that the statistic is significantly different from zero at the 99% confidence level.

Network	Input data	R	FB	FMS	5X	KSP	Rank
8 h	NCEP–NCAR	0.05	0.46*	23.2	11.3	51	1.50
	ERA-40	0.15*	0.12	19.4	11.2	53	1.63
	MM5	0.15*	−0.23	22.9	13.3	38	1.75
30 day	NCEP–NCAR	0.25*	−0.10	78.0	59.2	20	2.59
	ERA-40	0.31*	−0.26*	78.1	59.2	20	2.55
	MM5	0.41*	−0.71*	76.9	49.8	28	2.30

desirable to have a single parameter, which could be used to determine an overall degree of model performance. Stohl et al. (1998) found that the ratio-based statistics are the most sensitive to measurement errors while the correlation coefficient is one of the most robust. Although Chang and Hanna (2004) are more critical of the correlation coefficient due to its sensitivity to high concentrations, the measured concentrations in METREX span a relatively narrow range. Chang and Hanna (2004) also summarized attempts by several different researchers to define a single model evaluation parameter, such as ranking models by each statistic and then ordering by the total rank. In the following evaluation the rank was defined by giving equal weight to the normalized (0 to 1) sum of the correlation coefficient R , the fractional bias (FB), the figure of merit in space (FMS), and the Kolomogorov–Smirnov parameter (KSP), such that the total model rank would range from 0 to 4 (from worst to best):

$$\text{Rank} = R^2 + 1 - |\text{FB}/2| + \text{FMS}/100 + (1 - \text{KSP}/100). \quad (1)$$

5. HYSPLIT model configuration for the numerical experiments

The dispersion calculation was configured to run a one-full-year simulation (8784 h) at each of the tracer release sites from 1 January 1984 through 31 December 1984. Ten thousand tracer particles were released over 6 h every 36 h to correspond with each tracer release. The maximum number of particles permitted on the computational domain was always twice the number released for each emission to permit two consecutive releases to stay on the concentration grid. Particles older than 72 h were dropped from the calculation. Winds were never so light to cause particles to stay on the concentration grid for 72 h. A 1° square concentration grid of 0.005° horizontal resolution (about 500 m) and 100-m vertical resolution was centered at 39°N , 77°W to cover the METREX sampling region.

The concentration grid was defined to be at a higher resolution than the sampling network to minimize interpolation of the grid cell average concentrations to represent values at the actual sampling locations. Eight-hour-average concentrations, corresponding to the sequential sampling periods, were computed by HYSPLIT directly on the model's concentration grid. The air concentrations were then bilinearly interpolated to the location of the 8-h sampling sites and the 30-day sampling sites. The 8-h model concentrations at the 30-day locations were then averaged to obtain the 30-day air concentrations. To simplify the analysis and presentation of the results, in all subsequent discussions, unless otherwise noted, the PMCH and PDCH statistics are shown together after normalizing the PDCH concentrations to the ratio of the average PMCH/PDCH release rate (0.324).

Dispersion calculations using three different meteorological data sources are summarized in Table 1 and somewhat unexpectedly, considering the differences in resolution, indicate that the ERA-40 and MM5 results are quite comparable. However, the 8-h samples show a marked improvement in correlation and bias when using the ERA-40 data compared with calculations using only the NCEP–NCAR data. Although the 30-day samples show increasing correlation with increasing data resolution, the improved performance was more than offset by increases in bias. These results support the initial choice to use the ERA-40 data over the NCEP–NCAR data for ICs and BCs for MM5 based upon the underlying higher spatial resolution of the ERA-40 data assimilation system.

Running MM5 for 36 h every 24 h provided an opportunity to quantify the degradation of the dispersion calculation with increased forecast time. In the METREX simulations, two different analysis files were produced. One that consisted of only the first 24 h of the daily 36-h simulation (0000–2400 UTC) and one that consisted of only the last 24 h of the 36-h simulation (1200–1200 UTC). The results given in Table 2

TABLE 2. Results using the 4-km MM5 with the initial period (+0 to +24 h) and excluding the initial period (+12 to +36 h) for the 8-h and 30-day sampling networks. Statistical parameters are defined in the appendix. An asterisk alongside R or FB indicates that the statistic is significantly different from zero at the 99% confidence level.

Network	Period	R	FB	FMS	5X	KSP	Rank
8 h	0–24	0.15*	-0.23	22.9	13.3	38	1.75
	12–36	0.06	-0.66*	22.2	11.1	43	1.46
30 day	0–24	0.41*	-0.71*	76.9	49.8	28	2.30
	12–36	0.30*	-0.90*	76.8	43.1	33	2.08

indicate better performance over all the statistical parameters when using the initial meteorological data rather than the extended forecast data. The model's drift from reality with increased simulation time is an important consideration and it relates in large part to the timing of the tracer release with respect to the MM5 initialization time. Tracer releases alternated between starting at 0300 and 1500 UTC simultaneously at both release locations. The 1500 UTC release would always be 15 h after MM5 initialization, regardless of whether the initial period was included or excluded. However, the 0300 UTC release would be 3 h after initialization when including the initial period but 27 h after initialization excluding the initial period. Therefore, the two simulations represent, on average, the performance degradation with an increase in forecast time of about 24 h.

6. Model sensitivity results

The base dispersion model configuration for the subsequent sensitivity calculations will use the 4-km resolution MM5 data using the initial 24-h forecast period. Both the 8-h and 30-day networks had a sufficient number of data points upon which to draw conclusions. For instance, the base MM5 calculation indicated that the 8-h network had 835 samples, of which 583 showed some measured tracer above the threshold concentration. The other samples showed only a calculated concentration with no corresponding measurement. Measured and calculated concentration pairs that were both zero were not included in the statistics. In the 30-day network, 1771 samples were available, of which 1408 showed some measured tracer values.

a. Distance dependency

A cursory visual inspection of Fig. 1 shows that the 30-day sampling network consists of a variety of source to receptor distances, from just a few kilometers to over 75 km. However, the 8-h sampling network consists of

a limited number of source–receptor distances that ranged from 20 to 40 km with an average distance of 24 km. In the previous discussion of the results shown in Table 1, the dispersion model showed better performance for most of the statistical parameters for the 30-day samples than for the 8-h samples except that the 30-day samples showed a significant increase in the fractional bias. Some of these results can be explained by differences in model performance with distance.

The model's correlation with the 30-day sampling data was always better than with the 8-h data. This is primarily because the 30-day sampling averages the results of about 20 individual tracer releases, thereby averaging over the inevitable under- and overpredictions associated with random errors in both meteorological and dispersion model performance during each tracer release.

The differences in model bias between the 8-h and 30-day samples are subtler. The model's performance with distance in terms of mean concentrations for all the 30-day samples (including zeros) is shown in Fig. 2 for calculations using the ERA-40 and MM5 data. The monthly data have been aggregated into 5- and 10-km source-to-receptor distance bins. All model simulations show comparable performance to about 20 km. At farther downwind distances the calculations using MM5 resulted in a statistically significant (at the 99% level) underprediction of air concentration. The model's performance at 25 km with the 30-day samples is consistent with the previous 8-h sampling results and the overall large bias shown in the 30-day samples is primarily caused by much larger underpredictions at the most distant samplers.

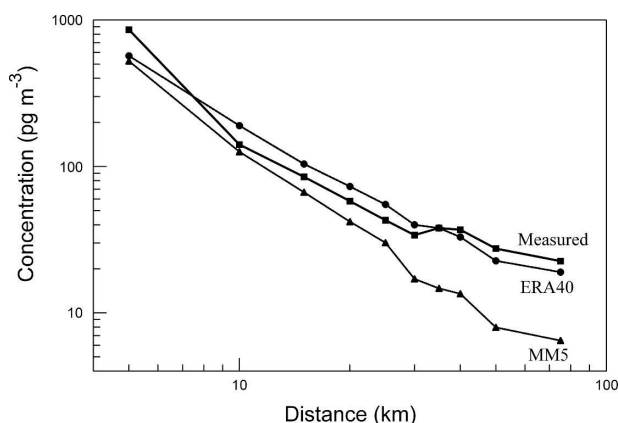


FIG. 2. The mean annual tracer concentration with distance from the release points computed using two different meteorological data sources (dashed lines) and aggregated into 5- and 10-km bins and with the measured data (solid line) from the 30-day sampling network.

b. Diurnal variations

The 8-h samples can be divided into those representing daytime and nighttime releases. The nighttime release started at 0300 UTC and consisted of three 8-h sampling periods during the same day beginning at 0500, 1300, and 2100 UTC. A daytime release started at 1500 UTC and consisted of the same three sampling periods, with the tracer release occurring in the middle period. The results are summarized in Table 3 and show that the ERA-40 has comparable performance both the daytime and nighttime samples, but the MM5 calculation shows less bias at night than during the day. The previous section indicated the bias was only an issue for the most distant samplers, suggesting that excessive vertical mixing in the daytime boundary layer may be one cause of the bias. Hanna and Yang (2001) came to a similar conclusion, where they determined that in the daytime mesoscale models predict a weaker inversion at the top of the mixed layer that could result in too much pollutant mixing out of the boundary layer.

c. Mixing depths

The underprediction of daytime air concentrations with MM5 as compared with using the ERA-40 data can in part be explained by the computation of the afternoon mixed layer depth. The annual average time series of mixing depth is shown in Fig. 3 at the central 8-h sampling location. Note that MM5 values are available every 30 min while the ERA-40 values are only available at 6-h intervals and must therefore be interpolated by the model to the intermediate times. The difference between the dashed and solid lines shown in Fig. 3 illustrates the limits of linear interpolation in representing boundary layer dynamics. For both datasets, mixing depths are computed in HYSPLIT as the first height above ground at which the potential temperature exceeds the surface temperature by 2°. Initially the 1500 UTC tracer release would see the same mixed layer depth, but at farther downwind distances

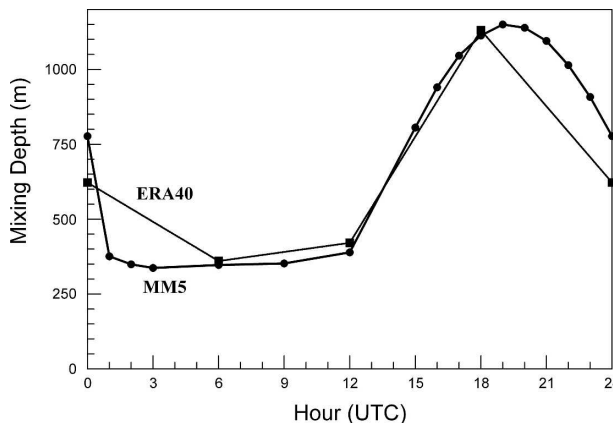


FIG. 3. Time series of annual average HYSPLIT calculated mixing depths over the central 8-h sampling location using the ERA-40 (dashed line) and MM5 (solid line) meteorological data.

and hence at later times the MM5 mixed layer depths are 20% larger than those derived from the ERA-40 data. The opposite situation does not occur at night for the 0300 UTC release, because the vertical mixing is weak and the plume would never reach the height of the mixed layer.

There is no evidence that one mixing depth is more correct than the other as there may be compensating errors in the dispersion calculation. Clearly mixing depth heights and a weaker mixed layer inversion can only explain part of the MM5 daytime underprediction. One possible remaining explanation is that the vertical plume structure using the finer-resolution MM5 data resulted in greater concentration gradients near the ground, with most of the tracer aloft, while the ERA-40-calculated plume was well mixed near the ground. Unfortunately, there were no concentration measurements above the surface and the corresponding model predictions were not saved above the surface samplers.

7. Confidence range for emergency response

The analysis of a model’s performance for emergency response applications is best evaluated using the 8-h samples. However, when short-duration samples are paired in space and time, the correlation tends to be low and little significance can be attributed to small differences in the statistical results. This is primarily due to the episodic nature of each release and the somewhat chaotic nature of atmospheric flow on the space and time scales used to simulate the transport and dispersion of the tracer plume. This is certainly not a new issue. It can be addressed through averaging or adjusting the calculated plume position to create an artificial overlap with the measurements. In their model evaluation Weil et al. (1992) matched the maximum concen-

TABLE 3. Results of the dispersion calculation with ERA-40 and MM5 meteorological data using the 8-h sampling data divided into daytime and nighttime samples. Statistical parameters are defined in the appendix. An asterisk alongside *R* or *FB* indicates that the statistic is significantly different from zero at the 99% confidence level.

Network	Data	<i>R</i>	<i>FB</i>	FMS	5X	KSP	Rank
8-h day	ERA-40	0.02	-0.16	19.7	12.8	54	1.58
	MM5	-0.10	-0.53*	23.7	13.6	34	1.64
8-h night	ERA-40	0.11*	0.23	22.4	12.4	58	1.54
	MM5	0.11	-0.11	28.4	18.5	34	1.90

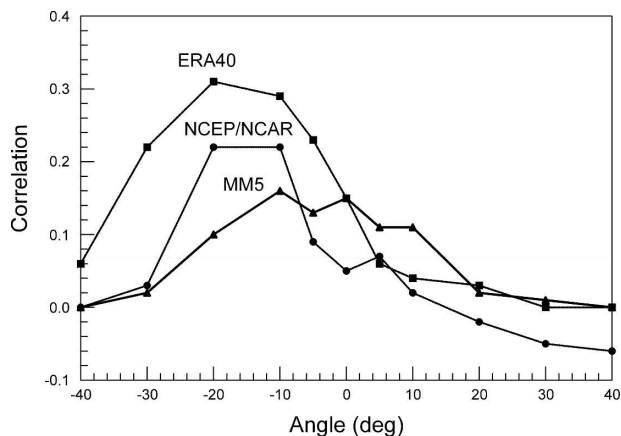


FIG. 4. The correlation coefficient for measured and calculated concentrations using different meteorological data and with various constant angular wind direction adjustments using the data from the 8-h sampling network. Positive angles indicate that the calculation required a clockwise adjustment. The dotted horizontal line indicates the minimum correlation that would be significantly different from zero at the 99% level.

tration of the calculated plume to the measured plume prior to computing the statistics.

For the sparse 8-h sampling network, if large errors in the concentration predictions are due to very small errors in the transport direction, then a simple test procedure can be applied to rotate the predicted plume direction to determine the angular adjustment required for a best fit to the measured data. Computationally, this was accomplished by applying an angular rotation to the position of the samplers with respect to the release so the new sampler location becomes

$$\phi' = d \sin(\theta + \Delta\theta) + \phi \quad \text{and} \quad (2)$$

$$\lambda' = fd \cos(\theta + \Delta\theta) + \lambda, \quad (3)$$

where θ is the source to receptor angle, $\Delta\theta$ represents the wind direction change, ϕ and λ are the latitude and longitude of the original sampler position, d is the distance from the source to the sampler, and f is the map factor. Air concentrations are then interpolated to the new sampler locations from the original model simulation results. Moving the sampler has the same effect as adjusting the input wind directions by a fixed amount.

Figure 4 shows the correlation coefficient between the measured and calculated 8-h concentrations as a function of rotation angle for calculations using various meteorological data. The center point, angle zero, represents the no-offset base calculation. A positive angle is equivalent to moving the plume clockwise about the release point. Angular deviations ($\Delta\theta$) were evaluated in rotation increments of 5° – 10° with a maximum forecast error of 40° . That range is within the uncertainty

found by Hanna and Yang (2001) in their evaluation of mesoscale meteorological models, which found that the uncertainty of model-predicted wind direction was on the order of 50° .

The most obvious feature shown in Fig. 4 is the apparent sensitivity of all the model results to negative rotation angles. This means that the “real” plume tended to be counterclockwise to the modeled plume. This result can be due to the fact that limited vertical resolution in the meteorological models overestimates the mean boundary layer pollutant transport velocity because there are an insufficient number of levels near the ground. In the Northern Hemisphere winds tend to turn clockwise with height. The previous interpretation is supported by the fact that the higher-resolution MM5 data, entirely based upon the ERA-40, showed very little angular sensitivity. A similar bias was found previously by Draxler (1990) when comparing winds at a tower in rural South Carolina to winds computed by the Nested Grid Model, NOAA’s forecast model at that time.

A low correlation between measured and calculated data paired in both space and time makes it difficult to decide which model simulation is performing better. However, the different response of each model simulation to rotation angle suggests that we can assign confidence sectors to each calculation based upon how much of an angular correction is required to improve the fit with the measurements. For each averaging time period, the mean square deviation,

$$\text{MSD}(k) = \sum (P_j - M_j)^2, \quad (4)$$

is computed over all three sampling locations (j) for each positive and negative angular offset (k) within the preselected range. The model offset with the minimum deviation then represents the model prediction for all samplers during that time period. The correlation of the resulting minimum deviations is shown in Fig. 5. All models show substantially improved performance up to angular deviations of $\pm 10^\circ$. One interpretation of this approach is to decide that a model calculation should have a certain level of performance, such as a correlation coefficient in the range of 0.5. Then the dispersion calculation confidence sector (the angular region that would contain the plume) would be 50% larger when using the ERA-40 data (12°) than when using the MM5 data (8°).

8. Intensive experiments

In addition to the special release of a third tracer during the monthly intensive experiments, two tether-

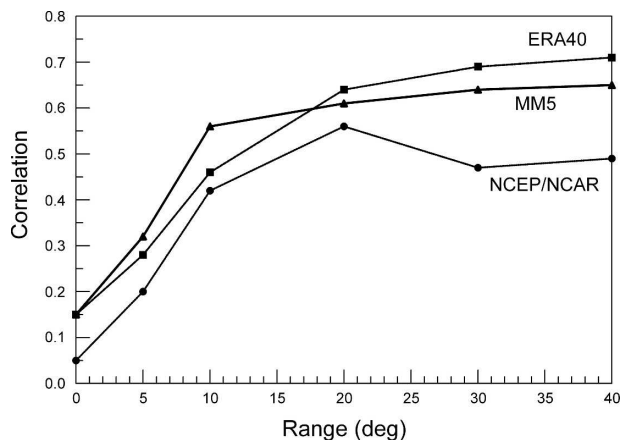


FIG. 5. The correlation coefficient for measured and calculated concentrations after making the optimal wind direction adjustment within a specified confidence sector range independently for each 8-h sampling period.

sonde systems were operated (except September and October) from May through December from about 2100 to 0400 LST. One site (T_s) was near the suburban tracer release point and the other, an urban site (T_u), was just north of the U.S. Capitol building (see Fig. 1). The instrument package measures wind speed, direction, and wet- and dry-bulb temperatures and was carried aloft by a blimp-shaped balloon inflated with helium. During each intensive, hourly single theodolite pibals were taken at National Airport, just south of the U.S. Capitol building, from 2200 to 0300 LST, and rawinsondes were taken at 2100, 0000, and 0300 LST at the normal reporting site, a rural area, near Dulles International Airport (see Fig. 1). Out of the nine intensive experiments, six were conducted at night and three during the day. No supplemental soundings were collected for the daytime experiments. Tracer was released during six experiments, but measurements on the 30-day network were only available for three experiments (Draxler 1986).

The simulation for the November intensive, a 4-hour release, started at 0300 UTC on the 8 November, is shown in Fig. 6. Although the model-predicted plume is to the south-southwest, multiple high measurements suggest an initial southeast transport direction later turning to the southwest. This demonstrates that a single angular adjustment cannot always account for systematic errors in model transport and that the transport bias is clearly clockwise.

Wind direction profiles, at various locations, near the end of the tracer release, are shown in Fig. 7. Only the suburban tethersonde site about 20 km to the northwest of the tracer release and the downtown pibal show any westerly component to the wind direction and then only

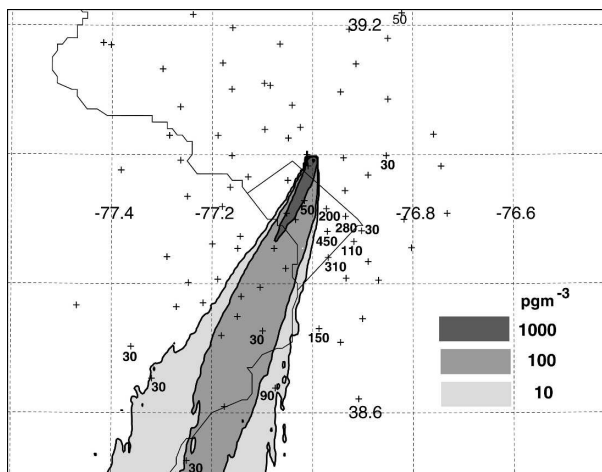


FIG. 6. Measured (at plus sign locations) and calculated (using MM5) surface tracer concentration (pg m^{-3}) contours resulting from the November intensive experiment tracer release just north of the central urban area. Note that the concentrations represent a time-integrated sample and only nonzero values are shown.

at the lowest levels. The upper levels of the rawinsonde and pibal observations do not show as much of an easterly component as the MM5 soundings and could account for some of the high tracer measurements due south of the release location. The tower wind directions at 0600 UTC near the release site (349° at 10 m and 022° at 60 m) are consistent with other wind observations at those heights. There is no measured or modeled wind direction that could account for the initial southeast direction of the measured tracer pattern. The implied plume transport direction to intersect those sam-

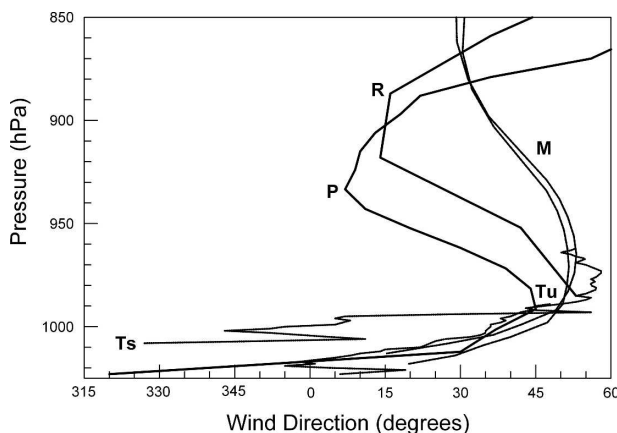


FIG. 7. The wind direction profiles at 0600 UTC 8 November for various data sources and locations; where M = MM5 (dashed), R = rawinsonde (solid), T = tethersonde (dotted) at the urban (u) and suburban (s) locations, and P = pibal (solid). The MM5 profiles are shown for the grid point nearest to each tethersonde location. Sounding locations are shown in Fig. 1.

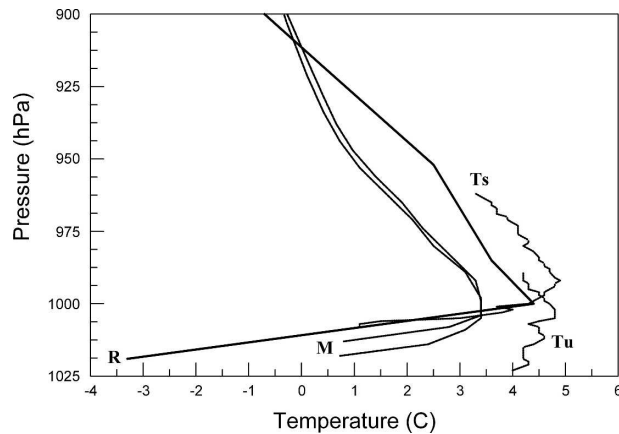


FIG. 8. Temperature profiles at 0600 UTC 8 November from various data sources at different locations, where M = MM5 (dashed), R = rawinsonde (solid), and T = tethersonde (dotted). The MM5 profiles are shown for the grid point nearest to each tethersonde location.

plers would require wind directions of around 330° – 345° . This suggests a tracer plume that consistently stayed below the canopy and then later mixed out to be transported with the more northeasterly winds.

The vertical temperature profiles downtown and in the adjacent suburban area are shown in Fig. 8 and indicate some of the modeling limitations. As expected, the model's representation of the low-level temperature profile was more characteristic of the suburban and rural areas than the urban downtown. The warmer low-level temperatures downtown clearly did not translate into more vertical mixing, as the wind profile (Fig. 7) would have showed a more easterly component. In fact the urban and suburban low-level wind directions were almost identical. In contrast, Hanna and Yang (2001) found that mesoscale models such as MM5 underestimated the near-ground nighttime vertical temperature gradients, which could be the result of too much vertical mixing. In any event, these issues require further study.

9. Summary and conclusions

The performance of HYSPLIT, a dispersion model originally developed for longer-range simulations, was evaluated over the Washington, D.C., area. Model-predicted air concentrations were compared with tracer measurements made at distances of 5–75 km from the source locations. The air concentration data for model evaluation were collected during METREX, a year-long study that consisted of over 500 individual tracer releases. In general, the results suggest that the dispersion model parameterizations were just as valid over

the shorter-distance scales of METREX as in previous longer-range evaluation studies. Although the use of global-scale meteorological data over the Washington, D.C., area provided realistic concentrations for longer-duration simulations, the use of a higher spatial- and temporal-resolution modeling tool, such as MM5, to create data fields more appropriate for the METREX domain, improved the performance of the dispersion model calculations more for the shorter-duration samples than the longer-duration samples when the results were expressed in terms of a smaller confidence sector. In addition, the following was found:

- 1) In terms of gross statistical results, if only global data were used for the dispersion calculation, the ERA-40 data showed better performance than using the NCEP–NCAR data for the 8-h samples. The results were comparable for the 30-day sampling network.
- 2) Dispersion calculations with MM5 data that included the initial period were clearly superior to calculations with MM5 data that excluded the initial period for both the 8-h and 30-day samples due to an increased drift from reality with increased forecast time.
- 3) All meteorological data provided good and comparable performance with distance from the source to about 25 km downwind. After that point, calculations using the MM5 data showed increased bias toward the underprediction of concentration.
- 4) Dividing the 8-h samples into day and night showed that the concentration underprediction was much greater for the MM5 calculation during the day than at night.
- 5) It was found that small changes in the wind direction with respect to the tracer release location provided dramatic improvements in dispersion model performance. Results using MM5 showed the least directional bias and the greatest sensitivity to directional changes compared to calculations using the other meteorological data.
- 6) The wind direction adjustment showed negative bias meaning that the model plume direction tended to be clockwise to the measured plume direction. This bias was greatest when using the global meteorological data.
- 7) A detailed analysis of the wind observations during the November intensive showed that only the very lowest level winds could account for the transport direction of the measured plume.

It can be concluded that the provision of operational dispersion model products for local applications using only global or mesoscale meteorological data for the

calculation can be justified if the dispersion model results are expressed within a confidence range based upon the resolution of the meteorological data driving the calculation. Although the results here are based upon global analysis data to provide initial and boundary conditions for the mesoscale model, the computation without the inclusion of local observations is comparable to what would be used in an emergency response application, where no observational data would be immediately available to compute a plume forecast.

APPENDIX

A Summary of Statistical Performance Measures

The correlation coefficient R is used to represent the scatter among paired measured M and predicted P values:

$$R = \frac{\sum(M_i - \bar{M})(P_i - \bar{P})}{\sqrt{\sum(M_i - \bar{M})^2 \sum(P_i - \bar{P})^2}}, \quad (\text{A1})$$

where the summation is taken over the number of samples and the overbar represents a mean value.

Although strongly influenced by both the correlation and bias, the percentage of calculations within a certain factor of the measured value is a popular statistic, such as within a factor of five ($5\times$) defined as the percentage of values that satisfy

$$0.2 \leq \frac{P}{M} \leq 5.0. \quad (\text{A2})$$

A normalized measure of bias is the fractional bias (FB). Positive values indicate overprediction and FB ranges in value from -2 to $+2$ and it is defined by

$$\text{FB} = 2 \frac{(\bar{P} - \bar{M})}{(\bar{P} + \bar{M})}. \quad (\text{A3})$$

The spatial distribution of the calculation relative to the measurements can be determined from the figure of merit in space (FMS), which is defined as the percentage of overlap between measured and predicted areas. Rather than trying to contour sparse measurement data, the FMS is calculated as the intersection over the union of predicted p and measured m concentrations in terms of the number N of samplers with concentrations greater than zero:

$$\text{FMS} = 100 \frac{N_p \cap N_m}{N_p \cup N_m}. \quad (\text{A4})$$

Differences between the distribution of unpaired measured and predicted values are represented by the Kolomogorov–Smirnov parameter, which is defined as the maximum difference between two cumulative distributions when $M_k = P_k$, where

$$\text{KSP} = \text{Max}|D(M_k) - D(P_k)| \quad (\text{A5})$$

and D is the cumulative distribution of the measured and predicted concentrations over the range of k values such that D is the probability that the concentration will not exceed M_k or P_k . It is a measure of how well the model reproduces the measured concentration distribution regardless of when or where it occurred. The maximum difference between any two distributions cannot be more than 100%.

REFERENCES

- Chang, J. C., and S. R. Hanna, 2004: Air quality model performance evaluation. *Meteor. Atmos. Phys.*, **87**, 167–196.
- , P. Franzese, K. Chayantrakom, and S. R. Hanna, 2003: Evaluations of CALPUFF, HPAC, and VLSTRACK with two mesoscale field datasets. *J. Appl. Meteor.*, **42**, 453–466.
- , —, Z. Boybeyi, and P. Franzese, 2005: Use of Salt Lake City URBAN 2000 field data to evaluate the Urban Hazard Prediction Assessment Capability (HPAC) dispersion model. *J. Appl. Meteor.*, **44**, 485–501.
- Draxler, R. R., 1985: Metropolitan Tracer Experiment (METREX). NOAA Tech. Memo. ERL ARL-140, 102 pp.
- , 1986: Measured influence of the nocturnal urban heat island on pollutant transport. Preprints, *Fifth Joint Conf. on Applications of Air Pollution Meteorology*, Chapel Hill, NC, Amer. Meteor. Soc., 171–174.
- , 1987: One year of tracer dispersion measurements over Washington, D.C. *Atmos. Environ.*, **21**, 69–77.
- , 1990: The calculation of low-level winds from the archived data of a regional primitive equation forecast model. *J. Appl. Meteor.*, **29**, 240–248.
- , and G. D. Hess, 1997: Description of the HYSPLIT_4 modeling system. NOAA Tech. Memo. ERL ARL-224, 24 pp.
- , and —, 1998: An overview of the HYSPLIT_4 modelling system for trajectories, dispersion, and deposition. *Aust. Meteor. Mag.*, **47**, 295–308.
- , M. Jean, B. Hicks, and D. Randerson, 1997: Emergency preparedness: Regional Specialized Meteorological Centers at Washington and Montreal. *Radiat. Prot. Dosimetry*, **73**, 27–30.
- Fox, D. G., 1984: Uncertainty in air quality modeling. *Bull. Amer. Meteor. Soc.*, **65**, 27–36.
- Graziani, G., W. Klug, and S. Mosca, 1998: Real-time long-range dispersion model evaluation of the ETEX first release. EUR-17754, Joint Research Center, European Commission, 216 pp. [Available from Office for Official Publications of the European Communities, Luxembourg.]
- Grell, G. A., J. Dudhia, and D. R. Stauffer, 1994: A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5). NCAR TN-398+STR, 138 pp. [Available from NCAR, P.O. Box 3000, Boulder, CO 80307.]
- Hanna, S. R., 1989: Confidence limits for air quality model evalu-

- ations, as estimated by bootstrap and jackknife resampling methods. *Atmos. Environ.*, **23**, 1385–1398.
- , 1993: Uncertainties in air quality model predictions. *Bound.-Layer Meteor.*, **62**, 3–20.
- , and R. Yang, 2001: Evaluations of mesoscale model's simulations of near-surface winds, temperature gradients, and mixing depths. *J. Appl. Meteor.*, **40**, 1095–1104.
- Källberg, K., A. Simmons, S. Uppala, and M. Fuentes, 2004: The ERA-40 archive. ECMWF ERA-40 Project Report Series, No. 17, 31 pp. [Available from European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, United Kingdom.]
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Mosca, S., G. Graziani, W. Klug, R. Bellasio, and R. Bianconi, 1998: A statistical methodology for the evaluation of long-range dispersion models: An application to the ETEX exercise. *Atmos. Environ.*, **32**, 4307–4324.
- NWS, 2004: Non-weather related emergency products specification. National Weather Service Instruction 10-518, Operations and Services, Public Weather Services, NWS, 28 pp. [Available online at <http://www.nws.noaa.gov/directives>; also from Operations and Services, Public Weather Services, 1325 East-West Highway, Silver Spring, MD 20910.]
- Otte, T. L., A. Lacer, S. Dupont, and J. K. S. Ching, 2004: Implementation of an urban canopy parameterization in a mesoscale meteorological model. *J. Appl. Meteor.*, **43**, 1648–1665.
- Stohl, A., M. Hittenberger, and G. Wotawa, 1998: Validation of the Lagrangian particle dispersion model FLEXPART against large-scale tracer experiment data. *Atmos. Environ.*, **32**, 4245–4264.
- Warner, S., N. Platt, and J. F. Heagy, 2004: Comparisons of transport and dispersion model predictions of the URBAN 2000 field experiment. *J. Appl. Meteor.*, **43**, 829–846.
- Weil, J. C., R. I. Sykes, and A. Venkatram, 1992: Evaluating air-quality models: Review and outlook. *J. Appl. Meteor.*, **31**, 1121–1145.