Evaluation of an Ensemble Dispersion Calculation

ROLAND R. DRAXLER

NOAA/Air Resources Laboratory, Silver Spring, Maryland

(Manuscript received 30 March 2002, in final form 12 August 2002)

ABSTRACT

A Lagrangian transport and dispersion model was modified to generate multiple simulations from a single meteorological dataset. Each member of the simulation was computed by assuming a ±1-gridpoint shift in the horizontal direction and a ±250-m shift in the vertical direction of the particle position, with respect to the meteorological data. The configuration resulted in 27 ensemble members. Each member was assumed to have an equal probability. The model was tested by creating an ensemble of daily average air concentrations for 3 months at 75 measurement locations over the eastern half of the United States during the Across North America Tracer Experiment (ANATEX). Two generic graphical displays were developed to summarize the ensemble prediction and the resulting concentration probabilities for a specific event: a probability-exceed plot and a concentration-probability plot. Although a cumulative distribution of the ensemble probabilities compared favorably with the measurement data, the resulting distribution was not uniform. This result was attributed to release height sensitivity. The trajectory ensemble approach accounts for about 41%–47% of the variance in the measurement data. This residual uncertainty is caused by other model and data errors that are not included in the ensemble design.

1. Introduction

Deterministic model predictions provide a sense of confidence that is not always supported by the underlying model assumptions. These issues are well known among the model developers and other experts but are frequently overlooked when the model results are transmitted to the consumer of these products. Model pollutant prediction uncertainty information may be especially critical for economic decisions, regulatory applications, and human health issues, especially for the atmospheric release of hazardous materials. Increasingly faster computational platforms can now support multiple simulations quickly enough to permit enhancing deterministic model predictions with additional measures of model prediction uncertainty. Uncertainties in air-quality model results can occur through the misrepresentation of the meteorological conditions because of inadequate spatial or temporal data resolution, sampling errors, or perhaps the inclusion of unrepresentative observations. Further errors or excessive simplification in the model's description and representation of various physical processes can lead to further uncertainties in the results.

Ensemble in its current common usage applies to the development of multiple meteorological forecasts. As indicated by Toth (2001), ''ensemble techniques involve the perturbation of the initial conditions (and possibly the model) to an extent representative of initial (or model) uncertainties.'' The application of ensemble methods to meteorological forecasts was reviewed by Molteni et al. (1996), Sivillo et al. (1997), and Toth et al. (1997) for synoptic-scale weather prediction. Some of the success of the ensemble approach was attributed to the fact that the model errors were small when compared with errors from the initial conditions. Errors in defining the meteorological fields in data-void regions are attributed to the use of model-generated first-guess fields in the analysis. These are fields used for retrospective air-quality simulations. It is clear that air-quality forecasts would be confined by the same limitations as the underlying meteorological forecast. An air-quality model ensemble computation cannot always rely upon the multiple meteorological data fields produced by the meteorological ensemble because the fields are typically saved at reduced resolution in time and with fewer vertical levels. For example, see Toth et al. (1997) for a description of the typical suite of National Centers for Environmental Prediction (NCEP) meteorological ensemble products.

The coarse-resolution ensemble data can be used for applications that only need to represent variations in the large-scale flow fields. For instance, Scheele and Sigmund (2001) used the forecast meteorological fields from each member to compute an ensemble of trajectories that were compared with the analysis trajectory to estimate potential trajectory error of the forecast.

Corresponding author address: Roland R. Draxler, R/ARL Rm. 3350, 1315 East–West Hwy., Silver Spring, MD 20910.
E-mail: roland.draxler@noaa.gov

However, more-complex applications, such as boundary layer pollutant dispersion, can be very sensitive to data resolution because of large variations in vertical mixing and vertical wind shears. The problem of using coarse-resolution meteorological ensemble fields for dispersion calculations was approached by Straume et al. (1998) and Straume (2001) by using the ensemble fields as initial and boundary conditions for a high-resolution atmospheric dynamic model to create fields with finer spatial and temporal resolution for the dispersion calculation. The ensemble dispersion results were compared with tracer measurements made during the European Tracer Experiment (Nodop et al. 1998) and were used to relate the dispersion-model concentration variations to the ensemble meteorological input fields. Straume (2001) demonstrated that the dynamic model could provide higher-resolution data to resolve boundary layer processes. Straume (2001) also noted that the selection of ensembles that are based upon the singular vectors that show the greatest growth at longer forecast times may not be the most appropriate for shorter-range dispersion forecasts.

In retrospective air-quality analyses, the term ''Monte Carlo models'' is more frequently used for ensemblelike applications such as model sensitivity analyses to determine the likely range of concentrations that may occur. In one approach, Venkatram (1979) used the term ''ensemble'' to differentiate short-term concentrations from their long-term, or ensemble, mean. Using Monte Carlo methods, Hanna et al. (1998) randomly sampled a large selection of input variables with an assumed distribution to perform 50 simulations of an urban-scale photochemical model. One limitation of this approach was that the range of the input parameters was determined through a survey of experts. The sensitivity coefficients (response to input change) may in fact be a function of the meteorological conditions during that particular simulation. Dabberdt and Miller (2000) reviewed the current status of ensemble dispersion modeling and presented the results from their simulation of an accidental pollutant release in which a priori uncertainties with a uniform distribution were assigned to various meteorological and dispersion parameters. Results were presented as contour patterns and receptor-oriented concentration histograms. They identified future research challenges in quantifying model uncertainties and information display. In Europe, the Real-Time Model Intercomparison (RTMOD) program (Galmarini et al. 2001) has been designed to collect the individual model results from the participating national modeling centers and then to provide various ensemble dispersion products based upon the collective results.

The Monte Carlo approach was further refined by Warner et al. (2002) by generating ensemble meteorological fields using an atmospheric dynamic model with different data sources for initial and boundary conditions and various combinations of boundary layer and surface physics parameterizations. In addition to computing the dispersion independently for each ensemble member, Warner et al. (2002) used the ensemble fields to generate the wind field variances, which were then used directly by the dispersion model to compute the air concentration probability distribution.

In one application of an ensemble dispersion model, the underlying concept is that of the ensemble trajectory as outlined by Stohl (1998). The approach, which parallels the meteorological ensemble, is that the starting trajectory position is not always exactly known because of uncertainties in source height and model-versus-actual terrain heights. These initially small errors may grow quickly in divergent flows. The potential trajectory divergence can be accounted for by starting an ensemble of trajectories with slightly different starting positions (Merrill et al. 1985). Another approach was taken by Kahl (1996) in generating a trajectory ensemble by adding a random perturbation to each wind component. The standard deviation of the perturbation was determined from previous meteorological data interpolation error studies. Stohl et al. (1995) found that trajectory temporal interpolation errors were greater than spatial interpolation errors because of the variability in the vertical velocity field.

The ensemble approach in this study follows that of Baumann and Stohl (1997), in which they computed a trajectory ensemble to account for trajectory error and interpolation error. In the former, trajectories were started in a circle with a radius of approximately one grid distance about the starting position. The interpolation error was accounted for by adding a perturbation to the velocity components. In this study, only the sensitivity to trajectory errors is addressed in the ensemble dispersion calculation. Ensemble products can be generated for individual events and also over a wide range of events to examine their statistical properties. Both issues will be addressed in this analysis by comparing the results with measured air concentration data.

2. Ensemble dispersion model setup

The ensemble dispersion calculation is composed of two components: the meteorological data and the plume dispersion computation. In the current configuration, only one set of meteorological data is used to create an ensemble of dispersion calculations. The transport and dispersion calculations are made using a modified version of the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model (Draxler and Hess 1998). The ensemble model results will be compared with air concentration data collected during a 3-month tracer experiment.

a. Dispersion model

The HYSPLIT model (version 4) is a complete system that can compute simple air parcel trajectories or complex dispersion and deposition scenarios; see Draxler

and Hess (1997) for a more detailed description of the model. Version 4 is a result of a joint effort between the National Oceanic and Atmospheric Administration (NOAA) Air Resources Laboratory and Australia's Bureau of Meteorology. New features include improved advection algorithms, updated stability and dispersion equations, a new graphical user interface, and the option to include chemical transformation modules. All simulations used the default configuration of modeling the horizontal dispersion through the growth of pollutant ''puffs'' and the vertical dispersion by a Monte Carlo particle approach. The meteorological data are vertically interpolated to the dispersion model's internal terrain-following coordinate system, which is on the same horizontal grid as the input data. Particles can be released from any location or height within the model domain. A fixed number of particles (750) is released at a constant rate over each 3-h emission period. The particles split with time after the horizontal puff dimension exceeds the meteorological grid size. After a duration of 5 days, a single release is typically tracked by almost 5000 particles. The particles are transported about the model domain by the mean wind field and a turbulent component computed from the horizontal and vertical diffusion coefficients. Air concentrations are computed on a user-defined grid by summing the masses of all particles in each concentration grid cell divided by the cell's volume, averaged over 24 h, which corresponds to the sample collection period. The concentration grid was defined with a horizontal resolution of 0.5° and a depth of 10 m. Surface gridcell concentrations represent the layer averaged from the ground to 10 m.

b. Meteorological input data

Tracer transport and dispersion was calculated using meteorological data fields produced by the NCEP–National Center for Atmospheric Research reanalysis project (Kalnay et al. 1996). The reanalysis data were obtained from NCEP 4 times per day on the model sigma surfaces on a global Gaussian grid of 1.875° resolution. More detailed information on the reanalysis data archives can be found at the NCEP Web site (available online at http://wesley.wwb.noaa.gov/reanalysis.html). To reduce the data volume and because the dispersion model requires the data to be on a conformal grid, the data points from the global grid were bilinearly interpolated to a regional Mercator grid. No vertical interpolation was required. The Mercator grid parameters were selected to ensure that the Mercator grid points were in almost identical locations with the Gaussian grid points, which minimizes the importance of the interpolation technique. The 57×41 regional Mercator grid has a 136.5-km spacing and covers the continental United States and Canada (from 21.9°N, 127.5°W to 60.0°N, 52.5°W). Each final processed data file contains 1 month of data (surface pressure, air temperature, true-direction wind components, and vertical velocity) on 28 model sigma surfaces, with the first 7 levels within 150 hPa of the surface.

c. Creation of the dispersion ensemble

The assumption behind the dispersion ensemble is that errors in the downwind plume position are primarily a function of the accumulation of initial errors in the particle trajectories. These errors are especially important just downwind of the source, when the size of the pollutant plume is much smaller than the meteorological grid size. This fact is analogous to the basis of meteorological ensemble forecasts, in which differences in initial conditions may lead to very different results (Molteni et al. 1996; Sivillo et al. 1997). Each ensemble member of the dispersion forecast is computed from the same pollutant source location, but during the calculation the meteorological grid is offset by ± 1 grid point in the horizontal direction and ± 250 m in the vertical direction (about 1 grid point), thereby resulting in 27 members of the ensemble. The exact computational procedure is that the particle position is shifted relative to the meteorological grid prior to the advection and dispersion calculation, then it is adjusted back, prior to computing the particle's contribution to the concentration grid. The same shift is applied for the entire ensemble-member simulation. This procedure is identical to the more traditional approach of shifting the tracer (or trajectory) source location, the effect of which would also be felt during the entire simulation. The reason the data grid needs to be shifted relative to the particle, rather than just adjusting the source location, is to ensure that the concentration predictions at the measurement locations are not influenced by differences in alignment between the source and receptor, but only by the variation in meteorological conditions between grid points.

The ensemble members are identified in Table 1 by their gridpoint offsets. The control member is listed as number 1 (no horizontal shift and a -1 vertical shift) because the tracer used for verification was released near ground level. To maintain a more symmetric ensemble, the computations for all ensemble members used a 250-m release elevation; hence, the -1 vertical shift would be most representative of the actual verification data. The shift of the data in the computational grid to create the ensemble members and the height of the tracer release are independent and are linked only through the identification of which ensemble member is most representative of the actual tracer release height.

The rationale of shifting the meteorological grid to determine initial transport errors is that the meteorological data field, limited in spatial and temporal resolution, is only an approximation of the true flow field, which is continuous in space and time. Only features that are several times as large as the grid spacing are resolved by the reanalysis meteorological data fields.

Although there are many possible approaches to the creation of an ensemble dispersion calculation, the

TABLE 1. Gridpoint offsets of each ensemble member with a horizontal shift equal to about 136 km and a vertical shift equal to 250 m. The central point in the horizontal and vertical directions is indicated by a zero offset in all directions. Member 1 is considered to be the control case because the verification tracer release was near ground level. The normalized-mean-square error (pg m^{-3}) for the GGW and STC releases is shown in the last two columns.

Member	ΔX	ΔY	ΔZ	GGW	STC
1 (control)	0	0	−1	91	108
2	0	+1	−1	98	109
3	0	−1	−1	99	201
4	+1	0	−1	97	1203
5	+1	+1	−1	99	271
6	+1	−1	−1	94	112
7	−1	0	−1	92	113
8	−1	+1	−1	104	99
9	−1	−1	−1	105	261
10	0	0	0	81	113
11	0	+1	0	75	108
12	0	−1	0	97	113
13	+1	0	0	89	134
14	+1	+1	0	81	105
15	+1	−1	0	119	114
16	−1	0	0	70	103
17	−1	+1	0	70	120
18	−1	−1	0	97	105
19	0	0	+1	285	290
20	0	+1	+1	312	245
21	0	−1	+1	285	313
22	+1	0	+1	356	285
23	+1	+1	+1	304	257
24	+1	−1	+1	292	330
25	−1	0	+1	264	307
26	−1	+1	+1	303	284
27	−1	−1	+1	287	303

method chosen for this analysis was designed to illustrate the uncertainties that can be introduced by limitations in the spatial resolution of the meteorological data. The grid resolution of the meteorological data driving the ensemble calculation determines the scale of the dispersion that can be represented by the model. The approach in this paper is an illustration of the method, and the results only apply to this specific combination of meteorological conditions, dispersion model, and measurement data.

3. Tracer measurement data used for evaluation

Inert atmospheric tracers provide excellent benchmarks for testing atmospheric dispersion models. Most long-range experiments have been conducted under very controlled conditions (preselected meteorological conditions), leading to the question of how this practice might be biasing the statistical evaluation of dispersion models. Further, verification of a dispersion probability forecast requires experimental data with a sufficient number of independent events. Few long-range tracer dispersion experiments match that criterion. One notable exception is the Across North America Tracer Experiment (ANATEX; Draxler et al. 1991).

ANATEX was conducted for a 3-month period from 5 January to 29 March 1987 and consisted of a series of simultaneous 3-h-duration near-ground-level tracer releases from Glasgow, Montana (GGW at 48.4°N, 106.5°W; average release rate of 27.9 kg h^{-1}), and St. Cloud, Minnesota (STC at 45.6°N, 94.2°W; average release rate of 16.7 kg h^{-1}), using a different perfluorocarbon tracer at each location. At both sites, the tracer was vented from a pipe just above the roof of a one-story building, an effective release height of about 10 m. GGW was devoid of any significant vegetation, and the STC site was in a suburban location surrounded by 10–20-m-tall trees. The inert nondepositing tracer was released every 60 h, starting at 1700 UTC 5 January 1987, regardless of the meteorological conditions. Seventy-seven air samplers were placed primarily at rawinsonde stations in the eastern half of the United States and southern Canada and were programmed to accumulate daily samples starting at 1400 UTC 5 January 1987. Considering that the measurements were of 24-h duration, it would not be uncommon for a sample to contain tracer from two releases. The experiment consisted of 66 independent tracer releases (33 at each location) and, in conjunction with the 77 measurement locations, over 85 days resulted in the potential collection of about 13 000 samples. However, missing data, laboratory errors, instrument failures, and other reasons left about 10 700 samples.

Note that the bulk of the samples, around 80%, had air concentrations near ambient background, which has already been subtracted from the measurement data. Many near-background concentrations are an inevitable result of having tracer emissions that are independent of the meteorological conditions. Some releases may not have been captured by the measurement network. The background tracer concentrations are very stable (7 $\pm$ 0.5 pg m^{-3} for STC and 9 $\pm$ 5.6 pg m^{-3} for GGW), but the total uncertainty determined from duplicate measurements (Draxler et al. 1991) approached 8 pg m^{-3} at background levels for either tracer. Electronic versions of the tracer data and related technical reports were available from NOAA Air Resources Laboratory (see information online at http://www.arl.noaa.gov/datem/).

4. Example results from the ensemble calculation

The ensemble dispersion model was designed to run in a serial or multiprocessor computational environment as an ''embarrassingly parallel problem'' (Foster 1994). As noted above, the dispersion ensemble is composed of 27 members. Each process defined a different ensemble member (see Table 1). Although the tracer release was near ground level, the simulation release height was set to 250 m, thereby resulting in three vertical variation sets (0, 250, and 500 m) of nine horizontal variation ensemble members. The lowest level was most representative of the actual tracer release height. The model, run independently for GGW and STC for a 2040-h simulation time, required about 5 CPU h (on an IBM

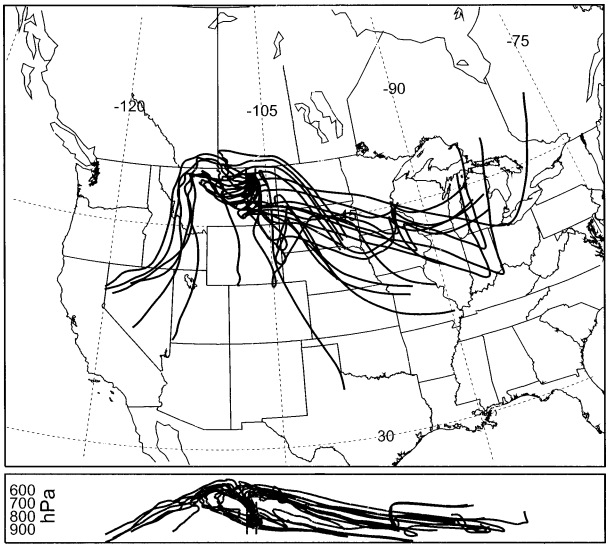

FIG. 1. One trajectory of 120-h duration for each ensemble member started at the time of the first tracer release from Glasgow, MT, at 1700 UTC 5 Jan 1987.

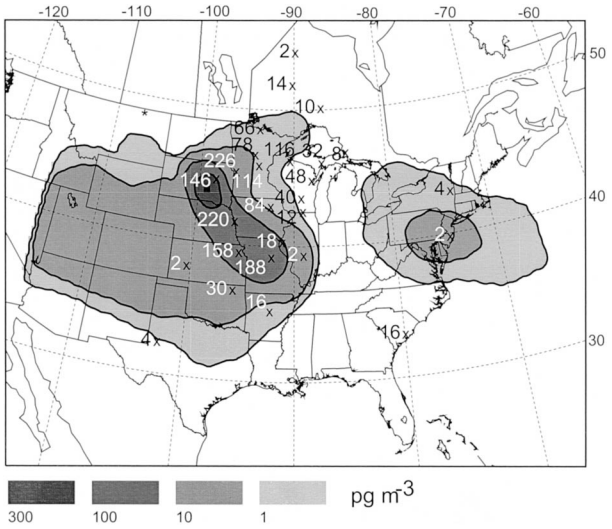

FIG. 2. Twenty-four-hour-average measured air concentrations (pg m^{-3}) starting 1400 UTC 9 Jan 1987 indicated by the numbers at the measurement locations (×) with above-background values, and contours for the control ensemble member computation results for the first two GGW (*) releases (5 and 8 Jan 1987).

43P 260 computer) per member. Files with gridded daily average air concentrations were written for each member, were processed by grid point, and were sorted in ascending order to create various probability output files on the same grid projection as the concentration data.

As an example to illustrate the ensemble calculation, a trajectory for each member corresponding to the path of one particle released at the start of the first tracer release from GGW (1700 UTC 5 January) is shown in Fig. 1. Each trajectory is of 120-h duration. Their spread is indicative of the divergence that could be expected in the flow. At first, many of the trajectories lingered near the source location. A small group of trajectories went to the west while the majority showed transport to the east. These 27 trajectories only represent ''one particle'' each at the start of the release; a total of 750 particles were released over the 3-h emission period. Each of the 750 particles will have a different trajectory, and their distribution is captured in the concentration fields produced by the full ensemble dispersion calculation. The figure illustrates the sensitivity of the results to simple gridpoint shifts of the source location within the meteorological data field.

The intent of the development of an ensemble dispersion product is to determine the uncertainty associated with an individual event. The calculation for that event may be based upon archived meteorological data, in which case the uncertainty is associated with how well the meteorological data and the associated dispersion model can represent the measured data, or it may be a prediction based upon forecast meteorological data, which contains the additional uncertainty related to ''verification of the forecast.'' The probability products that are routinely produced from the ensemble dispersion calculation include 1) air concentrations for the control case—the calculation most representative of the actual emission location and height, 2) the probability that a specific concentration value will be exceeded, and 3) the concentration contours at different standard probability levels such as the 50th, 75th, and 90th percentiles. The following examples, for a single measurement day, are shown to illustrate the interpretation of different kinds of dispersion probability products.

The measured above-background ANATEX concentrations at all stations are shown in Fig. 2 for the GGW tracer, together with contours of the control dispersion simulation concentrations. The measured air concentrations (pg m^{-3}) are averaged over a 24-h period from 1400 UTC 9 January to 1400 UTC 10 January. The measurements consist of a mix of tracer from the first two releases (1700 UTC 5 January and 0500 UTC 8 January) representing travel times of 2–5 days from the release starts. A more detailed evaluation of the model results suggests that the measurements to the south were from the first release, those to the north were from the second release, and those in the center contained a mix of the two releases. There is good correspondence between the highest measured and modeled concentrations (>100 pg m^{-3}). However, a number of high measurements in Minnesota and Wisconsin do not correspond well with the control model prediction. Part of this underpredicted region over Minnesota is covered by the area between the 1–10 pg m^{-3} isopleths. A large predicted region of low concentrations over the East Coast may be undersampled.

A ''probability-exceed'' (PE) plot, the probability of exceeding a specific concentration, in this example 10 pg m^{-3}, is shown in Fig. 3. The innermost contour, 75%, centered about the South Dakota–Nebraska–Iowa bor-

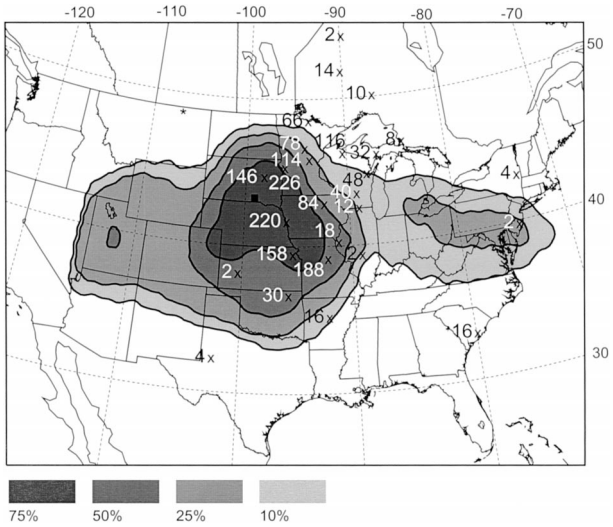

FIG. 3. The PE plot for 10 pg m^{-3} for the first two GGW releases. The probability contours represent the percentage of the ensemble members that exceed an air concentration of 10 pg m^{-3}. The measured air concentrations shown in Fig. 2 are also shown for comparison.

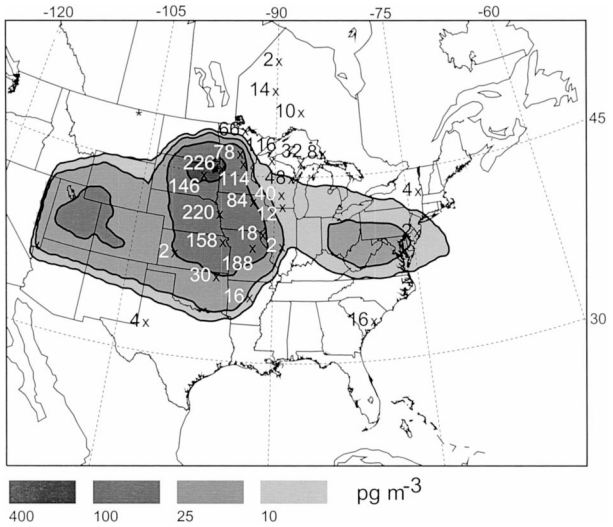

FIG. 4. The CP plot for the 90th percentile for the 24-h period starting 1400 UTC 9 Jan 1987. Within a contour, 10% of the members predict a higher concentration than the one given by the contour (pg m^{-3}), and, outside of it, 90% of the ensemble members predict a lower concentration. The measured air concentrations shown in Fig. 2 are also shown for comparison.

der, indicates that 75% of the ensemble members had a higher concentration within that region. Along the outermost contour, only 10% of the members had concentrations at that level or higher. For instance, between Wisconsin and Michigan, where the control calculation showed no concentrations greater than 1 pg m^{-3}, there were several high measurements. In this region, between 10% and 25% of the ensemble members showed concentrations greater than 10 pg m^{-3}. There are other regions that show similar probability levels but have no measured concentrations. For instance, the large number of members showing contributions as far west as Utah may be a valid result (also shown in the trajectories in Fig. 1) but cannot be confirmed because no measurement data were collected west of 105° longitude.

A ''concentration-probability'' (CP) plot is a transformation of the same information by showing the concentration contours for a specific probability level. The concentration contours are shown in Fig. 4 for the 90th-percentile concentration. For instance, along the innermost contour over South Dakota, 90% of the ensemble members have a concentration of 400 pg m^{-3} or lower outside of the contour and 10% of the members have a higher concentration within the contour. A similar interpretation applies to the other contours. For the situation from Wisconsin to Michigan, there is a 10% chance that concentrations will be greater than 10 pg m^{-3}. The 10 pg m^{-3} contours are identical in Figs. 3 and 4.

There is a wide choice in how the results from a complex series of simulations can or should be displayed. This area is evolving (Dabberdt and Miller 2000; Straume 2001) and depends in great part on the specific application and end-user requirements. Although, to a large degree, the measured tracer was encompassed by the ensemble members, the illustration of the results for one event does not ''verify'' the method. Measured data from the 3-month experimental period provide an opportunity to test the ensemble prediction.

5. Statistical verification of the ensemble

The overall performance of the control dispersion calculation can be illustrated most clearly by a simple scatter diagram (Fig. 5) that shows the measured and calculated air concentrations at all stations for all tracer releases conducted from GGW and STC for the entire 3-month experimental period. About 8000 samples are represented in the diagram. Those data points with a zero concentration for both the measured and control values were excluded. The large scatter between model results and measurements is consistent with the performance of many similar regional transport and dispersion models (van Dop et al. 1998). The results do not provide much guidance in terms of the model's uncertainty for a specific release when the predictions cover a large range of different weather types. As noted earlier in the discussion of the experimental data, a large fraction of the measurements were near the ambient background. The measurement detection limit is obvious from the ''step'' nature of the lowest air concentration measurements. Many of the model calculations were also very low: essentially undetectable. Although the background concentration was subtracted from the measurements, there can be considerable ambiguity in the concentrations when they are of comparable magnitude to the subtracted background (about 8 pg m^{-3}; see section 3). In the subsequent statistical analyses, measured con-

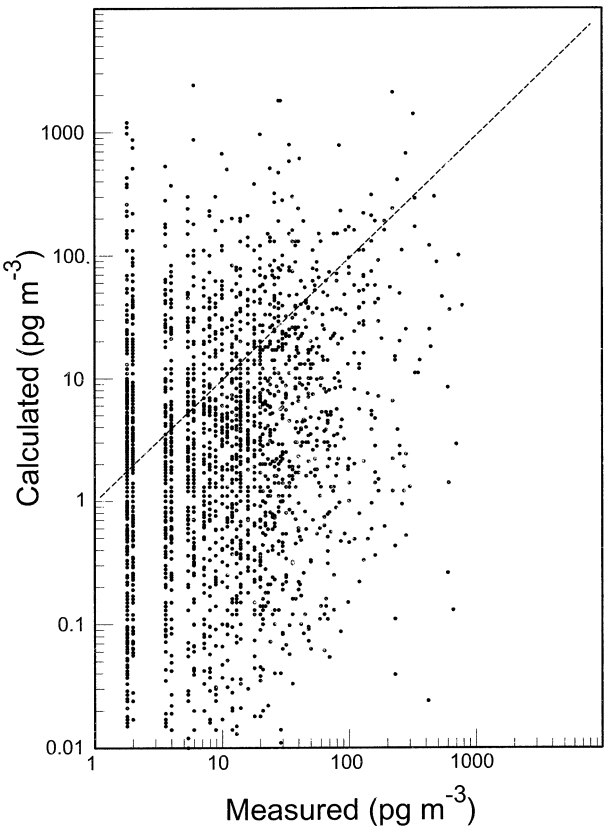

FIG. 5. Scatter diagram for daily average measured concentrations vs calculated concentrations of the ensemble control member for the entire 3-month experimental period for GGW and STC tracer releases. The dashed line represents the one-to-one correspondence line.

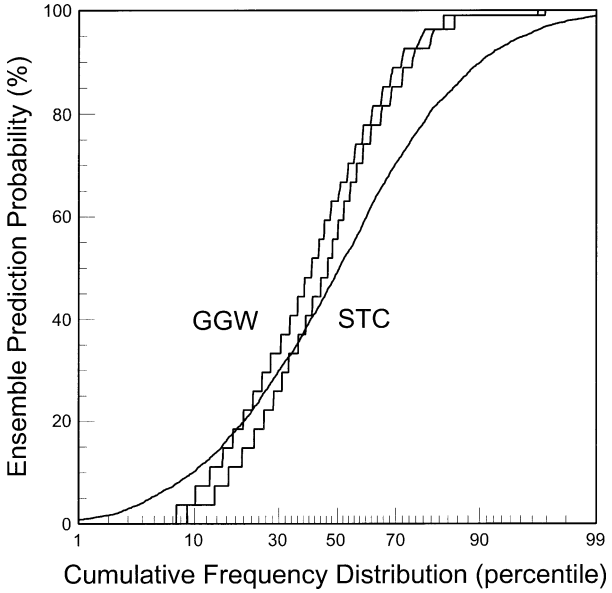

FIG. 6. Cumulative frequency distribution of the GGW and STC measurements in terms of their probability within the ensemble distribution. The ordinate indicates where each measurement falls within the ensemble (%), and the abscissa indicates the cumulative distribution of the measurements. The solid line is an example uniform distribution. Values of less than 10 pg m^{-3} are considered to be at or below background.

centration values below 10 pg m^{-3} were considered to be indistinguishable from zero. This criterion left about 7000 samples for which either the measurement or the ensemble prediction had an above-background concentration.

A quantitative measure of the quality of the ensemble prediction can be determined from the measurement data. For example, when the ensemble predictions indicate that a particular measured concentration lies on the median of the ensemble, how often does that actually occur? Or stated simply, how often do the measurements fall on the ensemble median concentration? For each measurement, there is a corresponding probability to indicate where that value lies among the ensemble predictions. Assume that for each measurement M, the ensemble concentrations C_i are sorted, such that $C_{i+1} > C_i$ for k members. Then the ensemble prediction probability is $100nk^{-1}$, and n is the ensemble member C_n that has the largest value less than or equal to M. The probability for each measurement was thus determined, and the cumulative distribution of these ensemble prediction probabilities is shown in Fig. 6. As an example, measured concentrations fell at the 50th percentile of the ensemble (ordinate) 40% of the time (abscissa) or less for GGW releases and 46% of the time or less for STC releases. The distribution has a ''steplike'' appearance because the probability increment cannot be less than 100% divided by 27 members. The smooth solid line indicates the shape of a uniform distribution. The underlying assumption in the configuration of the ensemble model is that any member is equally likely to occur. This clearly is not the case.

The ensemble member that provides the best fit with each measurement from the GGW release is shown in Fig. 7 as a histogram of the number of times in which that ensemble member had a concentration that was

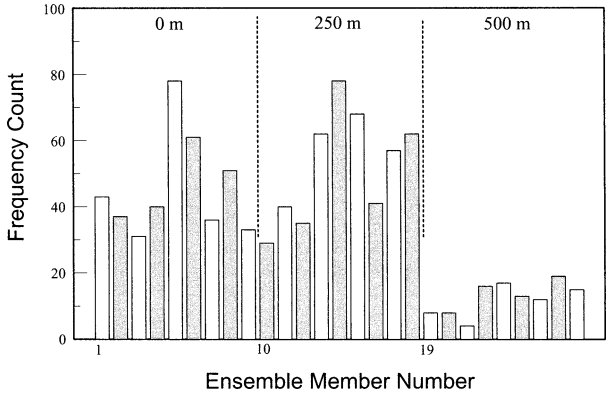

FIG. 7. Histogram of the ensemble members that had a concentration nearest to the measured concentration for the GGW release by member number.

nearest in value to the measured value. Ensemble members are numbered from 1 to 27, as indicated in Table 1. The control member has ΔX (west–east), ΔY (south–north), and ΔZ (vertical) offsets of 0, 0, and -1, respectively. The control member vertical offset is defined at -1 rather than at 0 because the actual tracer release height is at ground level. One reason that the measurement distribution was shown to be nonuniform in Fig. 6 can be attributed to the unrepresentativeness of the uppermost release height, most likely because of the effects of typical wintertime vertical stratification. The sums of the nine member contributions for 0, 250, and 500 m were 420, 472, and 112, respectively. The middle-release-height members provided a slightly greater contribution to the total than any of the other release heights, indicating a better fit with the measurements. A similar pattern was observed for the STC release histogram (not shown), with a sum of member contributions at the three heights of 207, 300, and 123, respectively. Within the middle-release-height group, the three ΔY members with $\Delta X = 1$ provided the greatest contribution to the best fit.

An explanation for the better results with ensemble combinations of $\Delta X = 1$ is not apparent, although it may be a function of grid resolution, local terrain, wind direction, and vertical mixing. For instance, because of local factors, the actual tracer may not mix upward as quickly as the model predicts, and when it does there may be a better correspondence with a horizontal meteorological grid offset in a preferential direction based upon synoptic conditions. Even with a more detailed evaluation of the computed concentrations from each member as compared with measurements, it may be difficult to diagnose the cause of the ensemble member's performance. For instance, Table 1 also lists the normalized-mean-square error (NMSE, as defined by Mosca et al. 1998) by member for the GGW and STC releases. In this case, each measurement is paired in space and time with the calculation from the same ensemble member for the entire experimental period. The overall results are comparable to the ''best-fit'' approach shown in Fig. 7, in that the two lowest vertical offsets showed the lowest error. However, the $\Delta X = 1$ members no longer show much difference when compared with other horizontal offsets. It is difficult to draw conclusions about a model's physical processes from statistical results, especially when model rankings change with the statistical parameter (Straume 2001; Mosca et al. 1998).

The NMSE results shown in Table 1 suggest an explanation for the similarities and differences between the 250-m release height versus the ground-level release results. The similarities are due in part to the fact that the dispersion model represented transport in the lowest level of the boundary layer by the first sigma-level wind (about 80 m above the surface). Therefore, transport differences between 250 and 500 m are expected to be greater than those between 250 m and the surface. The NMSEs for the ground-level STC results show considerable variations among ensemble members. An examination of the concentration predictions shows that two of the nearby sampling locations occasionally had very high predicted values. NMSE is very sensitive to extremes. These high values were not present in the upper-level-release-height calculations. This result indicates that either the low-level model winds were not as representative of the tracer transport winds or the actual vertical mixing was faster than that predicted by the model. This sensitivity to initial conditions is one of the assumptions in the design of the ensemble method.

The ensemble approach in this study only accounts for variations in the meteorological fields related to grid spacing, but those uncertainties are not the only ones in the calculation. When each measurement is paired with the best-fit ensemble member concentration, as was shown in Fig. 7, the correlation with the measurements indicates that the current grid offset approach can explain 47% and 41% of the variance at GGW and STC, respectively. The control simulation correlation was essentially zero. There may be other ensemble designs that could explain more of the variance in the measurements and perhaps provide a better fit with the probability distribution. For instance, errors in the model's physical parameterizations were not included and clearly are a factor. It remains for future investigations to determine whether other ensemble designs (such as Warner et al. 2002; Straume 2001) can provide a greater explanation of the variance in the measurements. The intent of this study is only to show that the range of ensemble predictions provides a realistic approximation to the measurements.

6. Summary and conclusions

A Lagrangian transport and dispersion model was modified to generate multiple simulations from a single meteorological dataset. Each member of the simulation was computed by assuming a ± 1-horizontal-gridpoint shift and a ± 250-m vertical shift in the meteorological data fields. The configuration resulted in 27 ensemble members. The ensemble dispersion model was run in either a serial computational environment or as an ''embarrassingly parallel problem.'' Each member was assumed to have an equal probability. The model was tested by creating an ensemble of daily average air concentrations at 77 locations over the eastern half of the United States during the Across North America Tracer Experiment during January–March 1987.

Unlike the ensemble dispersion model results of Straume (2001) and Warner et al. (2002), both of which used an atmospheric dynamic model to compute an ensemble of meteorological data fields for use by the dispersion model, the results shown here used only one meteorological data source and internally generated the ensemble dispersion product. The results are very computationally efficient and can be created quickly enough

for real-time applications. Furthermore, the ensemble predictions were compared with the results of 66 tracer releases in a variety of synoptic meteorological conditions, rather than with the single ETEX release used by Straume (2001) or the hypothetical release simulated by Warner et al. (2002).

Two graphical displays were developed to summarize the ensemble prediction and the resulting concentration probabilities for a specific event. The probability-exceed plot showed the probability of not exceeding a specific concentration and the concentration-probability plot showed the concentration contours for a specific probability level. Both the PE and CP analyses were developed from a cumulative sort of the ensemble concentration predictions at each grid location and can be used to view the spatial distribution of the ensemble model results for any number of concentration or probability levels. These displays are independent of the number of members in the ensemble and, therefore, provide a simple tool to summarize a number of complex simulations.

An analysis of the ensemble prediction for each measurement showed that the simple ensemble approach using the trajectory-error method to define the ensemble provided a reasonably close match to the measured distributions. However, there was not a uniform probability that each ensemble member was equally likely to occur. This result was attributed in part to the fact that some groups of ensemble members were more likely to have a concentration closer to the measurement than others. In particular, the middle-release-height computation was shown to have a better fit with the measurements than the computations from the surface, corresponding to the actual tracer release height. This result was most likely due to the fact that the lowest-level model winds were not always most representative of the tracer transport levels, perhaps because of the model's underestimate of the vertical mixing. Some of the results may have been affected by the noncentered (tracer release at the surface) design of the ensemble. However, the ensemble approach is only intended as a demonstration of the method, and the application of this particular configuration to other synoptic situations (summer), with different meteorological datasets (perhaps higher-resolution forecasts) and even using other dispersion models, needs further evaluation.

The results of this study have shown that an ensemble dispersion model can generate a reasonable approximation of a measured concentration probability distribution by assuming that the sources of model uncertainty are only attributed to the uncertainty in the pollutant particle trajectories. The approach uses a single meteorological dataset to create the ensemble members as compared with the multiple meteorological simulations used by Straume (2001) and Warner et al. (2002). Therefore, the approach is computationally fast enough that it could be considered for forecasting applications. The trajectory ensemble approach accounts for about 41%–47% of the variance in the measurement data. The ensemble computation does not model all of the uncertainty in the calculation (or in the measurements or meteorological data), and therefore there is still some residual uncertainty in the modeling system. More of the variance may be explained with different ensemble designs, which can provide a starting point for future investigations.

Acknowledgments. This research was sponsored by the NASA Solid Earth and Natural Hazards, Research, and Applications program: NRA 98-OES-13.

REFERENCES

Baumann, K., and A. Stohl, 1997: Validation of a long-range trajectory model using gas balloon tracks from the Gordon Bennett Cup 95. *J. Appl. Meteor.,* **36,** 711–720.

Dabberdt, W. F., and E. Miller, 2000: Uncertainty, ensembles and air quality dispersion modeling: Applications and challenges. *Atmos. Environ.,* **34,** 4667–4673.

Draxler, R. R., and G. D. Hess, 1997: Description of the HYSPLIT_4 modeling system. NOAA Tech. Memo. ERL ARL-224, 24 pp. [NTIS PB98-116593.]

——, and ——, 1998: An overview of the HYSPLIT_4 modeling system for trajectories, dispersion, and deposition. *Aust. Meteor. Mag.,* **47,** 295–308.

——, R. Dietz, R. J. Lagomarsino, and G. Start, 1991: Across North America Tracer Experiment (ANATEX): Sampling and analysis. *Atmos. Environ.,* **25A,** 2815–2836.

Foster, I. T., 1994: *Designing and Building Parallel Programs.* Addison-Wesley, 381 pp.

Galmarini, S., R. Bianconi, R. Bellasio, and G. Graziani, 2001: Forecasting the consequences of accidental releases of radionuclides in the atmosphere from ensemble dispersion modelling. *J. Environ. Radioact.,* **57,** 203–219.

Hanna, S., J. C. Chang, and M. E. Fernau, 1998: Monte Carlo estimates of uncertainties in predictions by a photochemical grid model (UAM-IV) due to uncertainties in input variables. *Atmos. Environ.,* **32,** 3619–3628.

Kahl, J. D. W., 1996: On the prediction of trajectory model error. *Atmos. Environ.,* **30,** 2945–2957.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Merrill, J. T., R. Bleck, and L. Avila, 1985: Modeling atmospheric transport to the Marshall Islands. *J. Geophys. Res.,* **90,** 12 927–12 936.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.,* **122,** 73–119.

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.

Nodop, K., R. Connolly, and G. Girardi, 1998: The field campaigns of the European Tracer Experiment (ETEX): Overview and results. *Atmos. Environ.,* **32,** 4095–4108.

Scheele, M. P., and P. C. Siegmund, 2001: Estimating errors in trajectory forecasts using ensemble predictions. *J. Appl. Meteor.,* **40,** 1223–1232.

Sivillo, J. K., J. E. Ahlquist, and Z. Toth, 1997: An ensemble forecasting primer. *Wea. Forecasting,* **12,** 809–818.

Stohl, A., 1998. Computation, accuracy and applications of trajectories—a review and bibliography. *Atmos. Environ.,* **32,** 947–966.

——, G. Wotawa, P. Seibert, and H. Kromp-Kolb, 1995: Interpolation errors in wind fields as a function of spatial and temporal res-

olution and their impact on different types of kinematic trajectories. *J. Appl. Meteor.*, **34,** 2149–2165.

Straume, A. G., 2001: A more extensive investigation of the use of ensemble forecasts for dispersion model evaluation. *J. Appl. Meteor.*, **40,** 425–445.

——, E. Koffi, and K. Nodop, 1998: Dispersion modeling using ensemble forecasts compared to ETEX measurements. *J. Appl. Meteor.*, **37,** 1444–1456.

Toth, Z., 2001, Meeting summary: Ensemble forecasting in WRF. *Bull. Amer. Meteor. Soc.*, **82,** 695–697.

——, E. Kalnay, S. M. Tracton, R. Wobus, and J. Irwin, 1997: A synoptic evaluation of the NCEP ensemble. *Wea. Forecasting*, **12,** 140–153.

van Dop, H., and Coauthors, 1998: ETEX: A European Tracer Experiment; observations, dispersion modelling and emergency response. *Atmos. Environ.*, **32,** 4089–4094.

Venkatram, A., 1979: The expected deviation of observed concentrations from predicted ensemble means. *Atmos. Environ.*, **13,** 1547–1549.

Warner, T. T., R.-S. Sheu, J. F. Bowers, R. I. Sykes, G. C. Dodd, and D. S. Henn, 2002: Ensemble simulations with coupled atmospheric dynamic and dispersion models: Illustrating uncertainties in dosage simulations. *J. Appl. Meteor.*, **41,** 488–504.