Improving SCIPUFF Dispersion Forecasts with NWP Ensembles

JARED A. LEE
Department of Meteorology, The Pennsylvania State University, University Park, Pennsylvania

L. JOEL PELTIER*
Applied Research Laboratory, The Pennsylvania State University, State College, Pennsylvania

SUE ELLEN HAUPT
Department of Meteorology, The Pennsylvania State University, University Park, and Applied Research Laboratory, The Pennsylvania State University, State College, Pennsylvania

JOHN C. WYNGAARD, DAVID R. STAUFFER, AND AIJUN DENG
Department of Meteorology, The Pennsylvania State University, University Park, Pennsylvania

(Manuscript received 11 December 2008, in final form 30 April 2009)

ABSTRACT

The relationships between atmospheric transport and dispersion (AT&D) plume uncertainty and uncertainties in the transporting wind fields are investigated using the Second-Order Closure, Integrated Puff (SCIPUFF) AT&D model driven by numerical weather prediction (NWP) meteorological fields. Modeled contaminant concentrations for episode 1 of the 1983 Cross-Appalachian Tracer Experiment (CAPTEX-83) are compared with recorded ground-level concentrations of the inert tracer gas C\textsubscript{7}F\textsubscript{14}. This study evaluates a Taylor-diffusion-based parameterization of dispersion uncertainty for SCIPUFF that uses Eulerian meteorological ensemble velocity statistics and a Lagrangian integral time scale as input. These values are diagnosed from NWP ensemble data. Individual simulations of the tracer release fail to reproduce some of the monitored surface concentrations of the tracer. The plumes that are predicted using the uncertainty model in SCIPUFF are broader, improving the overlap between the predicted and observed results. Augmenting the meteorological input to SCIPUFF with meteorological ensemble-uncertainty parameters therefore provides both a better estimate of the expected plume location and the relative uncertainties in the predicted concentrations than single deterministic forecasts. These results suggest that this new parameterization of NWP wind field uncertainty for dispersion may provide more sophisticated information that may benefit emergency response and decision making.

1. Introduction

For homeland and defense security it is necessary to model accurately the atmospheric transport and dispersion (AT&D) of chemical, biological, radiological, or nuclear (CBRN) contaminants from accidental or deliberate releases. Reliable forecasts of both contaminant concentrations and their inherent uncertainty are critical for situational awareness.

Accurate, reliable AT&D forecasts are difficult to make, however. Contaminant dispersion in the atmosphere is a complex process that is dictated by the properties of the turbulent, three-dimensional wind field. In the atmospheric boundary layer (ABL) the details of turbulent flows are, by nature, unpredictable. That is, small perturbations to initial conditions or boundary conditions produce different realizations of the flow and, thus, of contaminant dispersion.

Recognizing the limitations of forecasting a single realization, contemporary numerical weather prediction (NWP) uses ensembles of simulations. The wind field concentrations and their inherent uncertainty are critical for situational awareness.
from a single NWP run represents one member of this ensemble of possible realizations. Ensemble members typically differ by imposed initial and boundary conditions, parameterizations of unresolved physics, and often the choice of NWP modeling system. In addition, ensembles provide an estimate of the spread of possible future states of the atmosphere. For short-range mesoscale ensemble NWP, Grimit and Mass (2002) indicate that the correlation between ensemble spread and forecast error is high for cases of extreme spread; that is, forecast errors are generally smaller for low-spread events and larger for high-spread events. Furthermore, an ensemble mean of weather predictions is expected to be a better forecast than a single prediction, on average (Leith 1974). Similarly, an ensemble of dispersion predictions is also expected to be more useful than a single prediction for the purposes of assessing dispersion uncertainty.

Members of a dispersion ensemble can be generated by making a dispersion forecast for each member of a meteorological ensemble. This dispersion ensemble can then be averaged to obtain a mean prediction. Largely because of computational expense, limited observations, and the need for rapid response, however, a single AT&D model run is often utilized. In such cases the effects of wind field ensemble uncertainty must be modeled.

The use of NWP ensembles in dispersion modeling is rapidly developing (e.g., Straume et al. 1998; Straume 2001; Warner et al. 2002; Galmarini et al. 2004a,b; Mallet and Sportisse 2006; Kolczynski et al. 2009). Warner et al. (2002), for example, created a 12-member NWP ensemble using various combinations of surface physics, boundary layer parameterizations, and large-scale analyses for initial and boundary conditions to determine the sensitivity of the predicted concentration fields to the meteorological inputs.

Other ensemble dispersion modeling efforts include the ENSEMBLE project, a cooperative effort between fifteen European and two North American institutes, primarily for the benefit of European national emergency responders. Each institute provides real-time, long-range dispersion forecasts of hypothetical radioactive releases that are based on different NWP and AT&D models (Galmarini et al. 2004a). Galmarini et al. (2004b) applied the multimodel dispersion method, comparing it to single deterministic forecasts and recorded concentrations from the first episode of the 1994 European Tracer Experiment (ETEX-1). They developed two indicators for comparing their 16-member ensemble’s concentration predictions with the observed contaminant cloud: the agreement in threshold level (ATL) and the agreement in percentile level (APL). These are described in detail and used as metrics in our study. In addition, they defined a median model as the $APL_{50}$ contour, which is the median of all the model member concentration predictions at every grid point at every time step. The spatial coverage of the contaminant predicted using the median model outperformed the best single-member realization when compared with the observed ETEX-1 measurements. Riccio et al. (2007) also demonstrated that an approach using Bayesian model averaging yields nearly identical results to the median model results from Galmarini et al. (2004b), thereby providing a quantitative justification for the median model. These findings further reinforce the expectation that ensemble dispersion forecasting yields better results than deterministic forecasts made at the same resolution.

Another experiment that has been studied using ensemble dispersion methods is the 1983 Cross-Appalachian Tracer Experiment (CAPTEX-83). Deng et al. (2004) simulated the CAPTEX-83 release with the Second-Order Closure, Integrated Puff (SCIPUFF) AT&D model (Sykes et al. 2004) driven by the fifth-generation Pennsylvania State University–National Center for Atmospheric Research Mesoscale Model (MM5) (Grell et al. 1994). Using the MM5–SCIPUFF system they studied the effect of data assimilation and improvements in the horizontal resolution and internal physics parameterizations in MM5 on SCIPUFF dispersion predictions. All of these contribute to improved meteorological accuracy, and the combination of data assimilation and improved physics produced the best SCIPUFF predictions.

In this current study we further analyze the CAPTEX-83 study for the purpose of comparing monitored concentration values to predicted values using a variety of methods to implement the statistics from the ensemble of NWP simulations. As in Deng et al. (2004), we drive SCIPUFF with MM5 data. SCIPUFF models the ensemble mean field of three-dimensional Gaussian puffs of effluent. SCIPUFF treats its time-dependent input wind field (in this case, from MM5) as the ensemble average field and models the statistics of the unresolved turbulence. SCIPUFF also accommodates meteorological ensemble-uncertainty parameters that influence the spread of the mean plume (Sykes et al. 2004; Peltier et al. 2009, manuscript submitted to J. Appl. Meteor. Climatol., hereinafter PWH); parameterizations for some of the uncertainty metrics, however, are not well known.

The goal of this paper is to relate uncertainty in the NWP forecasts to the uncertainty in the concentration forecast. In addition, we seek to connect these uncertainties in ways inspired by Taylor diffusion theory (Taylor 1921). Section 2 discusses various aspects of meteorological and dispersion uncertainty and introduces the velocity variances, ensemble length scale,
2. Modeling meteorological uncertainty and its effect on dispersion uncertainty

a. Overview

There is a large body of literature dealing with uncertainty in dispersion modeling (e.g., Venkatram 1979; Fox 1984; Weil et al. 1992; Dabberdt and Miller 2000; Stein and Wyngaard 2001; Rao 2005; Hanna et al. 2007; Koracin et al. 2007). In our current study, we define uncertainty in terms of variances. Wind uncertainty is the variance, or mean-square error, in the wind components across an ensemble. Similarly, concentration uncertainty is taken to be the variance across an ensemble of concentration field predictions.

Rao (2005) lists three main types of uncertainty that are inherent to dispersion model predictions. The first type is data or parameter uncertainty, which comes from errors or uncertainties in the meteorological data that are driving the dispersion model, or in various parameters or initial and boundary conditions. The second type is model or structural uncertainty, arising from approximations of physical processes and imperfect numerics within the dispersion model. The third type is stochastic uncertainty, which is associated with atmospheric turbulence. He points out that improving the model can minimize the first two types of uncertainty, but not the third type, although it can be quantified statistically. Therefore, AT&D modeling must be treated probabilistically, in terms of the ensemble mean, variance, and probability distribution. It then follows that even an ensemble-mean concentration prediction, which is certainly an improvement over single deterministic forecasts, should be accompanied by variances or probabilities associated with a range of likely events (Dabberdt and Miller 2000).

SCIPUFF provides a probabilistic treatment of dispersion with the capability of predicting concentration probability density functions as functions of time and space (Rao 2005). SCIPUFF’s modeling accounts for the stochastic uncertainty that results from unresolved atmospheric turbulence. Modeling technology is also in place in SCIPUFF, through the ensemble uncertainty modeling, to account for the uncertainty that results from meteorological input data and parameter uncertainties or model physics errors; however, definitions for all of the modeling parameters are not complete. Rao (2005) concludes that uncertainty in the meteorological input data, especially horizontal wind variance, is the main source of total dispersion uncertainty. Other large sources of dispersion uncertainty include transport over complex terrain, transport over long distances, and deep convection with cold pools and outflow boundaries (Deng et al. 2004).

While there are many sources of dispersion uncertainty, uncertainty in the mean wind direction is a leading cause. In a limited-area model, grid-resolved variability leads to an error in the mean wind angle (PWH). While sub-grid variability is often parameterized, grid-resolved variability is usually not parameterized by AT&D codes. SCIPUFF, however, treats grid-resolved variability as a process that augments diffusion. The model introduces additional metrics to describe this variability in terms of ensemble variability. The parameters are the ensemble velocity variances and covariance (UUE, VVE, and UVE, respectively), and a characteristic length scale (SLE; Sykes et al. 2004). We adopt these SCIPUFF terms to make it clear to the reader which variables we are modifying in our use of SCIPUFF within this study. Of these parameters, the physics of SLE is not well defined, and therefore the ensemble variability model within SCIPUFF is seldom activated. Because the ensemble variability model within SCIPUFF is intended to parameterize the effects of ensemble wind field variability, the formulation of models for these grid-resolved variability parameters is an important step toward improved dispersion predictions. This current study assesses such models.

Warner et al. (2002) were the first to set bounds on the ensemble-uncertainty parameters in SCIPUFF. In their case study from the Persian Gulf War in 1991, they used two ad hoc estimates for the horizontal length scale in SCIPUFF, 10 and 50 km, and found that, in general, the hazard area increased with increasing length scale.

PWH apply Taylor’s theory for dispersion of a contaminant (Taylor 1921; Csanady 1973) to the problem of dispersion uncertainty for meteorological model sources. Using the assumption that Lagrangian and Eulerian velocity variances are equal, as in homogeneous, stationary turbulence (Lumley 1962; Corrsin 1963), PWH show that UUE, VVE, and UVE can be diagnosed from Eulerian meteorological-ensemble, deviation-velocity statistics, and that SLE is related to the Lagrangian integral time scale $\tau_L$:

$$\tau_L = \frac{1}{\nu^2} \int_0^\infty \langle v(t) v(t + \tau) \rangle d\tau',$$  

(1)
where \( \nu \) is the horizontal velocity component perpendicular to the ensemble-mean wind. This Lagrangian integral time scale is proportional to the time it takes for the lateral velocity of a particle in a flow to become decorrelated with (or forget) its initial value. When time-persistent deviations in the lateral velocity exist in individual ensemble members, as are caused by differences in wind direction, for example, the lateral velocity autocorrelation function \( \nu(t)\nu(t+\tau) \) decorrelates only at long time separations, yielding large \( \tau_L \) (PWH). In principle, \( \tau_L \) is the parameter that is required for SCIPUFF to estimate the uncertainty of mean concentration that is associated with grid-resolved variability (PWH), although this modeling parameter is implemented in terms of the length scale SLE in SCIPUFF (Sykes et al. 2004).

**b. Methods for modeling meteorological uncertainty**

As a step toward obtaining a parameterization for SLE, we model the release during CAPTEX-83 of a perfluorocarbon tracer gas [perfluoro-monomethylcyclohexane (C\(_7\)F\(_{14}\))] for each member of an MM5 NWP ensemble using SCIPUFF. We collect statistics for the ensemble of predicted contaminant plumes (dispersion uncertainty) and for the ensemble of wind fields (meteorological uncertainty). With this information we can relate the concentration uncertainty to the wind field uncertainty.

The wind field uncertainty is found by computing the single-point ensemble variances of the zonal \((u)\) and meridional \((\nu)\) components of the wind, and the ensemble covariance of \(u\) and \(\nu\) at every grid point over the whole domain across the entire MM5 ensemble. The horizontal wind ensemble variances and covariance are defined as

\[
\begin{align*}
\text{UUE} & = (\overline{u} - \bar{u})^2, \\
\text{VVE} & = (\overline{\nu} - \bar{\nu})^2, \quad \text{and} \\
\text{UVE} & = (\overline{u} - \bar{u})(\overline{\nu} - \bar{\nu}),
\end{align*}
\]

where the overbar indicates an ensemble mean.

Lagrangian particle dispersion models are often used in tracer studies (e.g., Koračin et al. 2007; Straume 2001). To compute \( \tau_L \), we tracked noninteracting Lagrangian particles through the three-dimensional NWP wind fields from each ensemble member, yielding a dataset of particle positions and local particle velocities. A stochastic model is embedded in the Lagrangian particle model to represent the effects of the unresolved subgrid turbulence on the particle trajectories. Therefore, two Lagrangian particles released from the same location at the same time will follow separate tracks (PWH). To track the lateral dispersion of the particles, however, the particle velocities must be rotated from grid-centric Cartesian components to a coordinate system aligned with the ensemble-mean flow. These particle trajectories are statistically processed to give two-point (separated in time) autocorrelation functions of the particle horizontal velocities lateral to the mean flow as a function of time separation \( \tau \):

\[
R_l(\tau) = \left\langle \frac{\overline{\nu(t)}\nu(t+\tau)}{\sqrt{\overline{\nu^2(t)}}\sqrt{\overline{\nu^2(t+\tau)}}} \right\rangle,
\]

where the subscript \( l \) denotes the velocity component lateral to the mean flow, the overbars signify ensemble averages, and the angle brackets denote an average over the particles that are still present in the domain at that time step. Equation (5) is the discrete approximation to the integrand in (1). Only the velocity correlation functions lateral to the ensemble-mean flow need to be considered because the tangential velocity has no effect on the lateral dispersion of a contaminant (Taylor 1921). Equation (5) is normalized by the root-mean-square (RMS) of the time-separated particle velocity functions.

From this average, normalized, lateral particle velocity autocorrelation function, the Lagrangian integral time scale is computed:

\[
\tau_L = \int_0^\infty R_l dt.
\]

As mentioned earlier, SCIPUFF uses a length scale, SLE, as a modeling metric instead of \( \tau_L \). To diagnose a scalar length scale using the scalar \( \tau_L \) from (6) and the tensor deviation-velocity ensemble variances and covariances, a scalar deviation-velocity statistic is needed. Although choice of the velocity parameter is not unique, a prime candidate is the square root of horizontal deviation-velocity energy \( \delta k \):

\[
\delta k = \frac{1}{2} (\text{UUE} + \text{VVE}).
\]

From this deviation-velocity energy and the constant estimate of \( \tau_L \), a time-varying and spatially varying field of the ensemble horizontal decorrelation length scale, SLE, can be calculated:

\[
\text{SLE} = \alpha \tau_L \sqrt{\delta k},
\]

where \( \alpha \) is a constant of order 1. Here we take \( \alpha = 1 \). We implement this SLE in SCIPUFF to model the grid-resolved wind variability.
3. CAPTEX-83 case study

a. CAPTEX-83 experimental setup

The primary goal of the CAPTEX-83 field program was to provide observational data for studies of long-range transport and dispersion of contaminants in the atmosphere (Ferber et al. 1986). Episode 1 of CAPTEX-83 took place on 18–19 September 1983. Over a 3-h period, from 1700 to 2000 UTC 18 September 1983, 208 kg of the tracer gas C$_7$F$_{14}$ were released at ground level at Wright-Patterson Air Force Base (39.80°N, −84.05°E) near Dayton, Ohio. The tracer gas was released in the middle of the day (1200–1500 local standard time). Downwind of the tracer release, a network of 86 monitoring stations in the northeastern United States and southeastern Canada, located in arcs approximately 300–1100 km from the release, recorded surface concentrations of the tracer (Ferber et al. 1986).

b. Meteorological conditions

The region of the experiment includes the eastern Great Lakes, the Mid-Atlantic, and much of New England over a time period from 1200 UTC 18 September to 1800 UTC 19 September 1983. The case was initially influenced by a large high pressure system centered over the Mid-Atlantic coast, with a broad southwesterly wind flow in the Midwest and Northeast. Northwest of the high pressure system, warm and cold fronts associated with a deep 982-hPa low in south-central Canada were moving rapidly through the western Great Lakes. A low-level jet was a prominent nighttime feature associated with this frontal system, and played a large role in transporting the tracer gas. For a more complete description of the meteorology during this episode of CAPTEX-83, including analysis maps, the reader is referred to Deng et al. (2004) and to Deng and Stauffer (2006).

The propagating midlatitude cyclone and its front-driven convection make this an interesting and challenging meteorological case in which to study long-range transport and dispersion. With an unstable ABL, the tracer mixed rapidly throughout the entire depth of the boundary layer (Deng et al. 2004). At the time of the tracer release (1700 UTC 18 September 1983), southwesterly winds were prevalent across Ohio ahead of the advancing cold front and transported the tracer northeastward. As time progressed, the fronts and the convective thunderstorms played a large role in the tracer dispersion.

Uncertainty in the predicted location of the cold front, in particular, and its attendant sharp change in wind direction produced additional uncertainty in the predicted location and concentration of the tracer. The inability in certain NWP ensemble members to represent accurately the mesoscale wind patterns associated with thunderstorms on fine-resolution domains also added uncertainty to concentration estimates (Deng and Stauffer 2006). The complex terrain in the experimental domain, including the Appalachians and Adirondacks, added yet another challenge for AT&D models.

c. Meteorological ensemble

An ensemble of MM5 simulations was constructed to study episode 1 of the CAPTEX-83 release. The experiment designed by Deng et al. (2004) demonstrated how improvements in numerical weather prediction and data assimilation since the 1980s affect long-range transport and dispersion modeling. They created a small number of experiments with three different grid configurations. In their initial study, the 12-km-resolution MM5 runs represented the meteorology better than either the 70- or 4-km-resolution simulations, and hence those 12-km-resolution NWP fields enabled better dispersion forecasts. To improve the MM5 predictions at 4-km resolution, Deng and Stauffer (2006) created a series of experiments with various internal physics and data assimilation schemes. Upon comparing each member of the 4-km ensemble with observations, they found that certain combinations of physics schemes and data assimilation, especially the use of a convective parameterization scheme on the inner 4-km grid, dramatically improve the 4-km forecasts and demonstrate an improvement over the best 12-km-domain MM5 experiment studied in Deng et al. (2004).

Following those studies, this current study uses a 19-member, nested-grid MM5 ensemble with common initial and lateral boundary conditions, but differing internal physics parameterizations and data assimilation schemes. The inner horizontal grid resolution for all the MM5 members is 4 km. There are 32 levels in the vertical, from the surface up to 100 hPa, with 16 levels below 850 hPa. This ensemble includes 14 of the model experiments created for the study by Deng and Stauffer (2006). Table 1 lists each of the ensemble members and the parameterizations used in each member. Four-dimensional data assimilation (FDDA) is used on the outer computational grids for every ensemble member, but only for a select few on the inner 4-km grid. For further discussion of the details of the various parameterizations used in this MM5 ensemble, see Deng and Stauffer (2006).

In this study, after discarding 5 h of model spinup, SCIPUFF is driven by 24 h of MM5 output, beginning at 1700 UTC 18 September 1983. The SCIPUFF domain encompasses much of the Great Lakes region and the Northeast, and is entirely within the 4-km MM5 domain. After converting the MM5 data to a
SCIPUFF-compatible format, the horizontal grid resolution is still 4 km, but there are 25 vertical layers, all of which are in the lowest 5 km above ground level (AGL) and 10 of which are in the lowest 1 km AGL to ensure resolving the ABL.

We wish to determine the variability of the MM5 ensemble to define a “most representative” member (MRM). Therefore, we compare the MM5 output that drives SCIPUFF (over both the SCIPUFF spatial and temporal domains) with meteorological observations. The mean absolute error (MAE) and mean error of each ensemble member are found for wind direction, wind speed, vector wind difference, temperature, and water vapor mixing ratio. Since these errors are not expected to remain constant throughout the entire depth of the troposphere, the statistics are grouped into three layers: the surface layer, 0–58 m AGL, which is the lowest terrain-following full layer in the MM5 simulations; the boundary layer, from 58 to 1000 m AGL; and the lower troposphere, from 1000 to 5000 m AGL. A sample plot of the MAE for each member compared with the observed wind direction, the meteorological parameter having the greatest effect on dispersion uncertainty, is illustrated in Fig. 1. From these computations, we define a most representative member to be the ensemble member that minimizes the mean absolute error on the most parameters. Experiment 12 is determined to be the most representative member of the MM5 ensemble in this study. It combines FDDA observation nudging and a convective parameterization scheme on the inner 4-km grid. It should also be noted that when examining the meteorological error statistics, all five of the MM5 ensemble members that used FDDA observation nudging on the 4-km grid performed markedly better overall than the members that did not use FDDA observation nudging.

d. Experimental methods

The CAPTEX-83 final report (Ferber et al. 1986) did not list details of the exact timing or nature of the tracer observations, so it was necessary to make some assumptions. Our analysis of the tracer data follows the assumption of Haagenson et al. (1987) that each reported sample during the CAPTEX-83 field study is not an instantaneous or accumulated concentration, but rather a 6-h average concentration. It should be noted, however, that the monitoring stations in the arc closest to the release point reported 3-h average concentrations. To make a uniform dataset for our analysis, we again followed the treatment of Haagenson et al. (1987) and averaged consecutive 3-h samples into 6-h samples. As a result, the surface concentrations predicted by SCIPUFF, which were instantaneous concentrations predicted every 15 min, are then averaged over 6-h periods. The 6-h
periods are centered at the approximate reporting times of the observed concentrations at 2200 UTC 18 September 1983, and 0400, 1000, and 1600 UTC 19 September 1983, the same centering times assumed by Deng et al. (2004). As in Draxler (1987), only tracer observations that are at least twice the atmospheric background concentration of $3 \text{ fL L}^{-1}$ (3 parts of the tracer C$_7$F$_{14}$ per $10^{15}$ parts of air) will be used for our analysis because of increased analytic uncertainty at concentrations near the instrument detection limit of $1 \text{ fL L}^{-1}$ (Ferber et al. 1986). Therefore, since the atmospheric background is already subtracted from the reported observations, our analysis of the predicted and observed plumes is based on a threshold concentration of $3 \text{ fL L}^{-1}$.

Several experiments are performed to discern the impact of the ensemble parameterization on concentration uncertainty as compared with the CAPTEX concentration observations. Specifically, the experiments are

1) MRM: the SCIPUFF dispersion prediction driven by the most representative member of the MM5 ensemble (experiment 12);

2) AugMRM: MRM is augmented with the ensemble-uncertainty values of UUE, VVE, UVE [(2)–(4)], and SLE [(8)] at every grid point and every time step, and then processed by SCIPUFF;

3) ModEM: a modeled ensemble-mean dispersion prediction that is created by ingesting the ensemble-mean MM5 data into SCIPUFF;

4) AugModEM: ModEM is augmented with the ensemble-uncertainty values of UUE, VVE, UVE, and SLE at every grid point and every time step, and then processed by SCIPUFF; and

5) ExpEM: an explicit ensemble-mean dispersion prediction that is the mean concentration at every grid point across the dispersion ensemble created by SCIPUFF.

As can be seen from (8), larger values of the ensemble wind variances lead to larger values of SLE, which lead to a broader plume. This broadening occurs because as the differences in the wind direction across the meteorological ensemble increase, so do the ensemble wind variances [(2) and (3)]. With increased uncertainty in the wind direction, there is increased uncertainty in the predicted location of a contaminant plume, and a wider plume reflects this uncertainty, enlarging the potential hazard area. Ideally, this forecast hazard area would be large enough to encompass most areas where contaminant concentrations would occur in a real episode.

The hazard predictions that we create using the augmented experiments (AugMRM and AugModEM), while likely to encompass more locations that observed above-threshold concentration, will necessarily increase false alarms over predictions made from single realizations. The number of false alarms depends on the amount of uncertainty in the meteorological input. A hazard prediction does not necessarily predict where the plume will be, but rather the locations where it could be. In other words, the hazard area is expected to be larger than the observed plume. The ideal balance would be to predict a hazard area that is large enough to encompass the most likely areas where the plume could go, given the chaotic nature of the atmosphere and the uncertainty in meteorological predictions. This is a difficult task, but is necessary to avoid misallocation of emergency response resources in an actual event.

It is not appropriate to compare directly the predictions that were augmented with ensemble-uncertainty parameters with the observed concentrations, which are only a single realization. Traditional metrics like the

---

1 Because the SCIPUFF predictions only extend until 1700 UTC 19 September 1983, however, our analysis for the final period is really a 4-h average centered at 1500 UTC instead of a 6-h average centered at 1600 UTC.
probability of detection (POD), false-alarm rate (FAR), bias score, and threat score are inappropriate for these comparisons. Therefore, we use qualitative measures like the broadening of the predicted plume, and the number of above-threshold concentrations encompassed by the predicted plume or hazard area. For the 19 SCIPUFF ensemble members, however, we do compute POD and FAR, because in that case we compare individual ensemble members, which are single realizations, with a single set of observations, which is also a single realization.

The metrics used to evaluate the performance of the MM5–SCIPUFF ensemble for the CAPTEX-83 case and to examine the effect of meteorological uncertainty on contaminant dispersion are built upon the work of Warner et al. (2002) and Galmarini et al. (2004a,b). Six-hour averages of the agreement in threshold level and agreement in percentile level statistics (Galmarini et al. 2004a,b) across the dispersion ensemble are created for further probabilistic analysis of the ensemble performance, and are described below.

4. Results and discussion

An ensemble of dispersion predictions is generated by SCIPUFF using the ensemble of 19 MM5 NWP simulations from Table 1 as input. The variance of the wind field across the ensemble is generally low, except in the region of fronts propagating across the domain. The signature of a cold front parallel to the major axis of Lake Erie is easily discerned in the velocity variances plotted in Figs. 2 and 3. Frontal signatures are also detected in the variance plots at other times during the simulation (Lee 2007). The variances are highest in these regions because the MM5 ensemble members disagree on the exact positioning and timing of the fronts, and hence also on the wind speed and direction in the vicinity of the fronts. The MM5 wind fields transport the contaminant ahead of the moving cold frontal zone after the release. The low variance in the wind field is likely a result of using the same initial and boundary conditions for each ensemble member. The ensemble was constructed from a set of experiments created for studying the effects of various data assimilation strategies and internal physics parameterizations. Although an ensemble created by varying the internal physics parameterizations captures a good deal of atmospheric variability (Stensrud et al. 2000), it is unlike an operational NWP forecast ensemble.

To determine the appropriate Lagrangian integral time scale, Lagrangian particles (section 2b) are released in the wind field of each of the 19 MM5 ensemble members at several heights at the time of the initial tracer release. The thick black line in Fig. 4 depicts the area average of the normalized lateral velocity autocorrelations (5) across all the particles that were released at an initial height of 100 m AGL that are still in the domain at each time step. This curve is fairly smooth initially, but after \( t \approx 5 \text{ h} \) it becomes jagged because of fewer and fewer particles remaining in the domain and contributing to the statistic. The integrated area under the black curve above the abscissa until the first zero intercept is the Lagrangian integral time scale \( t_L \). This method yields a value of roughly 2 h for \( t_L \) for all of the release heights. These values of \( t_L \approx 10^4 \text{ s} \) are consistent with the expectations of PWH, as well as the findings of Gifford (1982, 1987) and Sørensen (1998).

Figure 5 depicts the ensemble length scale \([8]\) created by using the computed fields of UUE and VVE and assuming a constant value of \( t_L = 2.0 \text{ h} \). The signature of the propagating front is evident. Values of SLE remain

\[ \text{FIG. 2. Zonal wind ensemble variance (m}^2\text{ s}^{-2} \text{) at 1000 UTC 19 Sep 1983 (+17 h). Circles denote CAPTEX monitoring stations; triangle denotes release location near Dayton.} \]

\[ \text{FIG. 3. As in Fig. 2, but for meridional wind ensemble variance.} \]

\[ \text{2 The small pockets of very high ensemble variance values, such as those seen over western Lake Erie in Fig. 2, are a result of a mesoscale modeling artifact that Deng and Stauffer (2006) call a “gridpoint storm.” This spurious and unrealistic convection would have a tremendous impact on the dispersion uncertainty of contaminants being transported through that region. In our case, however, the simulated tracer plume was already transported beyond the region with gridpoint storms, and so was largely unaffected by them.} \]
less than 10 km over most of the domain throughout the simulation, but exceed 25 km in the vicinity of the fronts and 40 km near the spurious mesoscale convection.

An ensemble of SCIPUFF dispersion predictions is created, each member being driven by one member of the ensemble of MM5 meteorological predictions. We calculated the POD and FAR for each of the 19 SCIPUFF ensemble members (not shown). Most members had overall values of between 0.60 and 0.70 for both POD and FAR over the course of the simulation. Interestingly, while experiment 12 (MRM) had the lowest overall FAR (54%), it did not have the best POD (59%). The mean overall POD and FAR were both 63% for the SCIPUFF ensemble.

Probabilistic measures are also used to compare statistically the performance of the dispersion ensemble to the observed tracer concentrations. The statistics used are the agreement in threshold level and agreement in percentile level (Galmarini et al. 2004a,b). The ATL is computed by finding the percentage of dispersion ensemble members at each grid point where the 6-h average concentration exceeds the designated threshold concentration of 3 fL L\(^{-1}\). A graphical representation of the ATL statistic at the four verification times (2200 UTC 18 September 1983 and 0400, 1000, and 1600 UTC 19 September 1983) is displayed in Fig. 6. As seen there and in Table 2, 81% of the total 6-h averaged concentrations fall within the ATL\(_{5}\) contour, which means that at least 2 of the 19 SCIPUFF ensemble members predicted above-threshold 6-h averaged concentrations at those sensor locations. The ATL contours encompass more of the tracer observations with higher percentile levels, as expected, ranging from 62% of the overall observations falling within the ATL\(_{5}\) contour to 81% of the overall observations falling within the ATL\(_{90}\) and ATL\(_{95}\) contours. Also, as expected, the higher percentile level contours encompass a greater fraction of false alarms than do the lower percentile levels because they enclose a larger area (see Table 3). The ATL\(_{5}\) contour, which depicts where half of the SCIPUFF members predicted above-threshold concentrations. More than a third (35%) of the 6-h averaged above-threshold concentration observations are predicted by all 19 SCIPUFF members (APL\(_{95}\) contour).

The APL is computed by finding the 6-h average concentration at every grid point that corresponds to the specified percentile level across the ensemble members. The 90th percentile level concentration contour (APL\(_{90}\)), centered at the four previously mentioned verification times, is displayed in Fig. 7. As summarized in Table 2, 81% of the total 6-h averaged concentrations fall within the APL\(_{90}\) contour, which means that at least 2 of the 19 SCIPUFF ensemble members predicted above-threshold 6-h averaged concentrations at those sensor locations. The APL contours encompass more of the tracer observations with higher percentile levels, as expected, ranging from 62% of the overall observations falling within the APL\(_{5}\) contour to 81% of the overall observations falling within the APL\(_{50}\) and APL\(_{95}\) contours. Also, as expected, the higher percentile level contours encompass a greater fraction of false alarms than do the lower percentile levels because they enclose a larger area (see Table 3).

While recognizing that a wider predicted plume will naturally lead to more false alarms, the use of additional statistics, such as ATL and APL, helps avoid misinterpretation. These statistics give emergency managers additional information about the expected level of risk or likelihood of contaminants being transported to a particular location at a particular time.

The results from these ATL and APL statistics indicate that the MM5 ensemble in this study includes
most of the atmospheric variability necessary to include the observations within the envelope of possible realizations. It is likely that these statistics would be improved further if 22% of monitoring stations had not failed to report a scheduled tracer sample during episode 1 of CAPTEX-83 (Ferber et al. 1986). The failure rate includes stations whose samplers were known to be contaminated and where the possible plume concentrations were not distinguishable from the range of contamination levels. Many of the Canadian samplers were contaminated, resulting in very few nonzero observed concentrations reported from Canada. SCIPUFF, however, predicted nonzero concentrations for many of those stations in several of the ensemble members. Of course, it is also possible that some monitoring stations might have reported zero concentration because detectable amounts of the tracer might not have been present there during the episode.

The meteorological data for MRM and ModEM are augmented with the ensemble-mean statistics UUE, VVE, UVE, and SLE. This allows us to compare model predictions both with and without ensemble-uncertainty parameters, in the case of both the most representative member (MRM and AugMRM) and the modeled

---

**TABLE 2.** Fractions of monitoring stations at which both the predicted and observed 6-h average surface concentrations exceeded the threshold value of 3 fL L$^{-1}$ (after having subtracted the background value of 3 fL L$^{-1}$). The ATL and APL statistics are computed from the 19 members of the SCIPUFF ensemble.

<table>
<thead>
<tr>
<th></th>
<th>2200 UTC</th>
<th>0400 UTC</th>
<th>1000 UTC</th>
<th>1600 UTC</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATL5</td>
<td>1.000</td>
<td>0.800</td>
<td>0.833</td>
<td>0.750</td>
<td>0.811</td>
</tr>
<tr>
<td>ATL25</td>
<td>0.667</td>
<td>0.600</td>
<td>0.833</td>
<td>0.667</td>
<td>0.703</td>
</tr>
<tr>
<td>ATL50</td>
<td>0.667</td>
<td>0.600</td>
<td>0.833</td>
<td>0.500</td>
<td>0.649</td>
</tr>
<tr>
<td>ATL75</td>
<td>0.667</td>
<td>0.500</td>
<td>0.750</td>
<td>0.333</td>
<td>0.541</td>
</tr>
<tr>
<td>ATL95</td>
<td>0.333</td>
<td>0.400</td>
<td>0.500</td>
<td>0.167</td>
<td>0.351</td>
</tr>
<tr>
<td>APL50</td>
<td>0.667</td>
<td>0.600</td>
<td>0.750</td>
<td>0.500</td>
<td>0.622</td>
</tr>
<tr>
<td>APL75</td>
<td>0.667</td>
<td>0.600</td>
<td>0.833</td>
<td>0.667</td>
<td>0.703</td>
</tr>
<tr>
<td>APL90</td>
<td>1.000</td>
<td>0.800</td>
<td>0.833</td>
<td>0.750</td>
<td>0.811</td>
</tr>
<tr>
<td>APL95</td>
<td>1.000</td>
<td>0.800</td>
<td>0.833</td>
<td>0.750</td>
<td>0.811</td>
</tr>
</tbody>
</table>

---

**TABLE 3.** Fractions of monitoring stations at which false alarms were predicted, where the 6-h average surface concentrations exceeded the threshold value of 3 fL L$^{-1}$ (after having subtracted the background value of 3 fL L$^{-1}$). The ATL and APL statistics are computed from the 19 members of the SCIPUFF ensemble.

<table>
<thead>
<tr>
<th></th>
<th>2200 UTC</th>
<th>0400 UTC</th>
<th>1000 UTC</th>
<th>1600 UTC</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATL5</td>
<td>0.800</td>
<td>0.765</td>
<td>0.722</td>
<td>0.679</td>
<td>0.735</td>
</tr>
<tr>
<td>ATL25</td>
<td>0.800</td>
<td>0.769</td>
<td>0.630</td>
<td>0.556</td>
<td>0.679</td>
</tr>
<tr>
<td>ATL50</td>
<td>0.800</td>
<td>0.739</td>
<td>0.583</td>
<td>0.455</td>
<td>0.647</td>
</tr>
<tr>
<td>ATL75</td>
<td>0.714</td>
<td>0.643</td>
<td>0.438</td>
<td>0.429</td>
<td>0.545</td>
</tr>
<tr>
<td>ATL95</td>
<td>0.800</td>
<td>0.429</td>
<td>0.143</td>
<td>0.333</td>
<td>0.409</td>
</tr>
<tr>
<td>APL50</td>
<td>0.800</td>
<td>0.727</td>
<td>0.571</td>
<td>0.455</td>
<td>0.641</td>
</tr>
<tr>
<td>APL75</td>
<td>0.800</td>
<td>0.769</td>
<td>0.630</td>
<td>0.556</td>
<td>0.679</td>
</tr>
<tr>
<td>APL90</td>
<td>0.786</td>
<td>0.742</td>
<td>0.688</td>
<td>0.591</td>
<td>0.697</td>
</tr>
<tr>
<td>APL95</td>
<td>0.800</td>
<td>0.765</td>
<td>0.730</td>
<td>0.679</td>
<td>0.737</td>
</tr>
</tbody>
</table>

---

**FIG. 6.** Six-hour average of the ATL (3 fL L$^{-1}$) values across the dispersion ensemble, centered at (a) 2200 UTC 18 Sep 1983 (+05 h), (b) 0400 UTC 19 Sep 1983 (+11 h), (c) 1000 UTC 19 Sep 1983 (+17 h), and (d) 1600 UTC 19 Sep 1983 (+23 h). Observed 6-h concentrations are reported next to the corresponding monitoring station. Circles denote CAPTEX monitoring stations; triangle denotes release location near Dayton.
ensemble mean (ModEM and AugModEM). Six-hour averaged SCIPUFF predictions at the four verification times are displayed in Figs. 8–12. The control run for MRM is shown in Fig. 8. Figure 9 (AugMRM) illustrates that incorporating these mean statistics results in a slightly larger, smoothed out plume compared to that of Fig. 8. The ModEM plume (Fig. 10), which is the SCIPUFF prediction driven by the ensemble mean of the MM5 members, also becomes larger and less detailed when ensemble-mean statistics are included in the AugModEM plume (Fig. 11). The plume broadening is more noticeable at later times during the simulation, as model member predictions have a tendency to drift apart in time. Figure 12 shows the ExpEM plume, which is similar in shape and area coverage to the AugModEM plume.
Table 4 lists the percentage of the observed above-threshold concentrations that are encompassed within the plumes of the five experiments. It should be noted that qualitative measures such as these are particularly sensitive to the density and locations of the sensor network relative to the plume. For instance, the plume predicted by AugMRM was clearly wider than the plume predicted by MRM, but the AugMRM plume did not encompass any more stations that reported above-threshold concentrations. Throughout the duration of the simulation, the hazard area of the ExpEM plume compares favorably to that of the AugModEM plume, while the area covered by the ExpEM plume encompasses marginally more monitoring stations overall during the simulation (70%) than either the AugMRM plume (59%) or the ModEM plume (62%). Commensurately, the AugModEM and
ExpEM plumes both encompass greater fractions of false alarms than the other experiments (see Table 5), largely because they also encompass more of the observed tracer concentrations than any of our other experiments. Note that in a forecast situation, however, the AugModEM calculation only requires one SCIPUFF run, and thus takes significantly less time to produce a concentration forecast than does an explicit SCIPUFF ensemble. Additionally, the AugModEM plume captures more of the tracer observations made during this field experiment than any single deterministic forecast because the predicted plume is broader.

5. Summary and conclusions

This case study has assessed the impact of including information on wind field uncertainty on the estimation of...
concentration uncertainty. A 19-member ensemble of MM5 simulations drives SCIPUFF, creating an ensemble of predictions of dispersion of the tracer release of episode 1 of CAPTEX-83. The average of this dispersion ensemble is termed the explicit ensemble-mean dispersion forecast (ExpEM). From the MM5 ensemble, the most representative member (MRM, the ensemble member with the smallest differences from the meteorological observations, particularly the wind direction) is chosen as the best meteorological forecast. The ensemble-mean meteorological fields (ModEM) are also computed, along with ensemble-mean statistics such as horizontal wind variances and covariance. A Lagrangian particle dispersion model is used to track hundreds of particles through each member of the ensemble of MM5 wind fields to estimate the Lagrangian integral time scale at \( \tau_L \approx 2 \) h, consistent with Gifford (1982, 1987), Sørensen (1998), and PWH. Based on work by Corrsin (1963), combining the ensemble wind variances and the estimate of \( \tau_L \) produces a three-dimensional, time-varying field for SLE, the decorrelation scale length of the ensemble.

The MRM and ModEM meteorological input data to SCIPUFF are then augmented (AugMRM and AugModEM) with these ensemble-mean statistics, including SLE, which generally results in broader tracer plumes, especially at later times in the simulation. The plume broadening is generally not substantial, however, and the broader plumes resulting from the augmented input files, even from MRM and ModEM, still fail to encompass all of the monitoring stations that reported above-threshold tracer concentrations, although they do encompass the majority of those stations. The ExpEM and AugModEM dispersion predictions both perform well, and both experiments encompass the observed tracer concentrations within the envelope of possible realizations. This finding implies that for this case, there may be value to decision makers in including these statistical uncertainty measures in the plume predictions.

The probabilistic ATL and APL statistics of the SCIPUFF ensemble demonstrate that the MM5 ensemble used in this study accounts for most of the variability inherent in the meteorological conditions during the CAPTEX-83 field study. The fact that most of the dispersion predictions early in the simulation miss some of the monitoring stations that recorded tracer observations is likely a direct result of the way in which the meteorological ensemble is constructed, however. Because each MM5 member uses the same initial and boundary conditions and FDDA on the outer coarse grids, and only differs in the use of FDDA on the inner 4-km grid and the internal physics parameters and parameterizations, the variability in the wind angle is generally small across the CAPTEX-83 domain. If an ensemble similar to an operational NWP ensemble was designed, which varies initial and boundary conditions, the horizontal wind variance across the ensemble might increase, which would then give a higher estimate of SLE. Note that the multiplicative constant in (8) may differ from 1. This issue should be investigated more thoroughly. Differences in the initial wind angle may cause some dispersion ensemble members to predict the early track of the tracer plume more successfully. We postulate that an NWP ensemble that combines both initial condition and physics parameterization variability would be better for dispersion forecasts, since model physics and initial conditions are both imperfect (Stensrud et al. 2000). Combining that variability with a multimodel approach (e.g., Houtekamer et al. 1996; Ebert 2001, 2002) merits consideration as well, because model numerics are also imperfect. This is a topic of future research efforts.

This study demonstrates that incorporating uncertainty in the wind direction into dispersion forecasts expands the hazard area of the predicted contaminant plume. This larger area should, in turn, provide emergency responders with a more accurate representation of the zone of potential impact. This uncertainty information could be valuable in the event of an actual CBRN contaminant release.

Acknowledgments. The Defense Threat Reduction Agency (DTRA) under Contract DTRA01-03-0010 (John Hannan, Contract Monitor) sponsored this project, and Christopher Kiley of Northrop Grumman helped to oversee the project. Thanks also are given to Ian Sykes of Sage Management, and Nelson Seaman,
Walter Kolczynski, Kerrie Long, Jeff Zielonka, Brian Reen, Adam Goss, and Monique Holmes for helpful discussion or other aid given during the course of this project. We also thank the anonymous reviewers for their thoughtful comments. Addressing their comments improved the paper.

REFERENCES


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