NOAA Technical Memorandum OAR ARL-251

WHERE WILL IT GO: ATMOSPHERIC DISPERSION MODELING OUT TO 30 KM TO GUIDE RESPONDERS DURING ACCIDENTAL OR HOSTILE RELEASE OF HAZARDOUS MATERIALS

Ronald. J. Dobosy

Atmospheric Turbulence and Diffusion Division
Oak Ridge, Tennessee

ATDD Contribution File No. 03/05

TABLE OF CONTENTS

NOTATION

ADAPT	Atmospheric Data Assimilation and Parameterization Tools (NARAC)
ADPIC	Advection Diffusion Particle in Cell
ALOHA	Areal Locations of Hazardous Atmospheres (CAMEO®)
ANATEX	Across-North-America Tracer Experiment
ARAC	Atmospheric Release Advisory Capability (now NARAC)
ARL	Air Resources Laboratory (NOAA)
CAPARS	Computer Assisted Protective Action Recommendation System (RARC)
CAMEO®	Computer-aided management of emergency operations
CFD	Computational Fluid Dynamics
COAMPS	Coupled Ocean-Atmosphere Mesoscale Prediction System (US Navy)
CP	"Concentration Probability," probability of exceeding a specified concentration
DOE	US Department of Energy
DP26	Dipole Pride Experiment 26 (Nevada Test Site)
DSTL	Defence Scientific and Technical Laboratory (UK)
DTRA	Defense Threat Reduction Agency (USA)
ECMWF	European Centre for Medium-Range Weather Forecasting
EPA	US Environmental Protection Agency
FDDA	Four-Dimensional Data Assimilation
FOM	Figure of Merit
GIS	Geographical Information System
GUI	Graphical User Interface
HOTMAC	Higher-Order Turbulence Model for Atmospheric Circulation (YSA)
HPAC	Hazard Prediction and Assessment Capability
HVAC	Heating, Ventilating, and Air Cooling
HYSPLIT	Hybrid Single-particle Lagrangian Integrated Trajectory
IEM	Innovative Emergency Management, Inc.
LES	Large-Eddy Simulation
LODI	Lagrangian Operational Dispersion Integrator (NARAC)
LPM	Lagrangian Particle Model
MC-SCIPUFF	Mass-Consistent (wind model) for SCIPUFF
MM5	Fifth generation PSU/NCAR Mesoscale Model
MOE	Measure of Effectiveness
MRF	Medium-Range Forecast (an NCEP boundary-layer scheme)
NARAC	National Atmospheric Release Advisory Center (DOE)
NCAR	National Center for Atmospheric Research
NCEP	National Center for Environmental Prediction (NOAA)
NOAA	National Oceanic and Atmospheric Administration (USA)
NORAPS	Naval Operational Regional Atmospheric Prediction System
NWS	National Weather Service (NOAA)
OMEGA	Operational Multiscale Environmental Model with Grid Adaptivity (SAIC)
PE	Probability of Exceeding (a concentration or dose)
PIC	Particle in Cell
PSU	Pennsylvania State University

RAMS	Regional Atmospheric Modeling System
RANS	Reynolds-Averaged Navier-Stokes
RARC	Regional Atmospheric Response Center
RHC	Robust Highest Concentration
SAIC	Science Applications International Corporation
SCIPUFF	Second-order Closure Integrated Puff (HPAC)
SWIFT	Stationary Wind Fit and Turbulence (HPAC)
TKE	Turbulent Kinetic Energy
TRAC	Terrain-Responsive Atmospheric Code (CAPARS)
UDM	Urban Dispersion Model (DSTL)
VLSTRACK	Vapor, Liquid, and Solid Tracking (US Navy)
YSA	Yamada Science and Art Inc.

LIST OF FIGURES

LIST OF TABLES

WHERE WILL IT GO: ATMOSPHERIC DISPERSION MODELING OUT TO 30 KM TO GUIDE RESPONDERS DURING ACCIDENTAL OR HOSTILE RELEASE OF HAZADOUS MATERIALS

Ronald Dobosy

ABSTRACT.

This report presents a critical survey of the state of dispersion modeling for emergency response to atmospheric release of hazardous materials. The focus is primarily on scales from 0 km to 30 km. Major advances in sensors and computing power over the past two decades have enabled impressive advances in the whole system of guidance for emergency responders. Tools available include both databases and, models. Friendly user interfaces guide the process from initial discovery, through evacuation or shelter, to medical treatment and cleanup. Transport and dispersion models have grown in sophistication along with the rest of the tools. Distributions of contaminants' concentration and dose can now be forecast, in principle, in the most complex situations of weather, terrain, and urban geometry. Estimates of uncertainty can be provided with these forecasts. The store of increasingly sophisticated datasets for testing and further development of models continues to grow. Tradeoffs between model complexity and run speed, and between measurement detail and affordable cost continue to loosen through increased power of computing and communication.

For all this, dispersion remains a stochastic problem with large irreducible variance, especially within the time and resource constraints imposed by real-world emergency response. Responders must apply the guidance with judgment based on experience and training, preferably supported by direct, any-time access to people expert in dispersion modeling and interpretation of its results. Though no specific recommendations are given here, this document is intended as a resource from which such recommendations can come.

1. INTRODUCTION

The increased risk of hostile activity in the United States has revived general interest in tools to guide responders in emergencies involving atmospheric release of hazardous materials. A full suite of such tools includes databases, models, and a user interface. The data describe properties of potential contaminants, configurations of potential sources, current weather conditions, and the like. Models give guidance on the wind and turbulence, the transport and dispersion through the environment, and the level of harm to people and assets. User interfaces provide rapidly grasped graphical representation of the guidance provided by the system.

This report presents a detailed survey of the state of atmospheric dispersion modeling for emergency response on scales from 0 km to 30 km. Such modeling is useful in three modes: pre-event planning, response to an event, and post-event assessment and recovery. Though some

methods surveyed here apply more readily to pre- and post-event activities, the viewpoint taken for this survey is primarily that of real-time response to an actual emergency. A survey of model validation techniques is also presented with some results from validation studies. Major advances in dispersion modeling, supported by similar advances in modeling of the atmospheric environment have accompanied the explosion in computing power over the last twenty years.

2. THE EMERGENCY RESPONSE MODELING ENVIRONMENT

It is important from the start to recognize these tools as guides, not as black boxes that give *the* answer. The model consists of a chain of forecasts of stochastic processes, each of high variability. It can best hope to estimate the ensemble mean, along with the variance due to natural variation and to error. The actual outcome will be an individual realization, subject to the high variability characteristic of the natural system being simulated. Furthermore, the contamination will follow myriad complex pathways on scales too fine to be resolved in the limited time available. Intuition must be built from experience comparing real and simulated situations. Judgment remains necessary.

The requirements for emergency response modeling on 10 km scales are severe. The short travel times over such limited distances in ordinary winds demand rapid deployment. Flexibility is also vital since it is impossible to foresee when and where an emergency will strike. The model must use whatever information is immediately available for the site. If the release is small and the terrain simple, a model such as ALOHA (USEPA, 1998) provides the fastest response. Beyond several kilometers, however, wind patterns and surface conditions are likely to be heterogeneous. Terrain influence is likely to be strong. A model must account for these complexities, including the change of relevant physical processes as the dispersion progresses to larger and larger scales.

The importance of such a modeling capability has long been recognized. Numerous private companies have high-quality response systems available by license or subscription. (http://www.ieminc.com/, http://www.ssesco.com/, and http://www.rarc.org/, among others) Science Applications International Corporation (SAIC) has the Operational Multiscale Environmental Model with Grid Adaptivity (OMEGA). Government agencies also participate. The Department of Energy has its National Atmospheric Release Advisory Center (NARAC), which has developed over the past 25 years a system that has responded to many of the celebrated releases, beginning with the accident at Three-Mile Island. The Defense Threat Reduction Agency (DTRA) maintains a similar capability for military and related US Government applications. They use the Hazard Prediction and Assessment Capability (HPAC), which features the SCIPUFF dispersion model. A list of some primary groups making dispersion estimates for emergency response is given in Table 1.

2

Table 1: Emergency-Response Modeling Activities

Organization	Atmosphere Model	Dispersion Model	Reference
NARAC	ADAPT, NORAPS, COAMPS	LODI	Lee *et al.* (1998); Leone *et al.* (1997)
DTRA	SWIFT, MC-SCIPUFF	SCIPUFF	Sykes *et al.* (1998); Sykes and Gabruk (1997)
NAVY	COAMPS	VLSTRACK	Bauer and Gibbs (1998)
IEM		D2-Puff™	Innovative Emergency Management, Inc (2000).
NWS	ETA	ALOHA HYSPLIT	USEPA (1998) Draxler and Hess (1998)
RARC CAPARS	NUATMOS	TRAC	Ciolek and Magtutu (1998). AlphaTRAC (2002).
Earth Tech Inc.	CALMET	CALPUFF	Scire *et al.* (2000a,b)
SAIC	OMEGA	OMEGA	Bacon *et al.* (2000)

These groups maintain aggressive research efforts that have evolved sophisticated simulations of contaminant spread in the atmosphere. Generally such sophistication requires prior commitment of major computing and sensing resources. Organizations having valuable assets at significant risk to accident or attack will normally have relations in place with such a group as listed above. However, an incident somewhere among the great number of "ordinary" places is also likely. Responders in such places need ready access to whatever can be brought rapidly to their support.

3. ISSUES IN EMERGENCY RESPONSE MODELING

A DTRA workshop (Beriwal and Merkle, 2001) discussed the range of issues involved with response to hostile or accidental atmospheric releases. They viewed the problem in multiple parts:

Source specification: Successful specification of the source of the contaminants is difficult in an emergency, particularly under hostile intent. Clearly the consequences will differ depending on the type of material released. Compare accidental release of CO_2 to hostile release of aerosolized pneumonic plague. Less obvious, but also serious is the effective height of the source, especially in complex terrain, where the wind direction may vary strongly with height. There have been celebrated cases where the forecast plume's path was directed 90° from the actual path. Even given a perfectly specified wind field, the forecast can be spectacularly wrong through errors in source height.

Transport and fate in atmosphere: Successful forecast of the dispersion of a cloud of contaminants requires detailed and accurate representation of the transporting wind and the diffusing turbulence. A simple assumption of constant wind direction and speed along the contaminant cloud's path is inadequate beyond the first few kilometers in most real cases. The fate of contaminants in the atmosphere is influenced by the full range of phenomena found there:

rain, surface deposition, resuspension, secondary circulations, thunderstorms and other mesoscale weather patterns, landforms, vegetation, and many more.

Dose Response: Given a forecast dose of a specified contaminant, the response of the population adds a further level of complexity involving invasion pathways and physiological response. Invasion pathways can be simple, such as direct inhalation, or complex, such as through a food chain. Since the current scope is dispersion modeling, there will be no further discussion of this topic. Its importance is however recognized here.

Epidemiology: When a bio-pathogen is involved, the mechanism of propagation expands beyond meteorology to include human behavior. This area of active development is also beyond the scope of the present discussion, but important to keep in mind. The initial dispersion of pathogens may be a meteorological problem, but the primary difficulty is likely to be the travel and behavior of human carriers.

Agriculture and biota: This category was included in the DTRA workshop to consider attack against crops. As with the other events, responders depend on a good dose forecast.

Recognizing the large uncertainties inherent in emergency-response modeling, we believe nevertheless that it can be intelligently applied to yield useful assessments of risk. In the remainder of this report we will discuss the state of modeling to produce a good dose forecast. "Good" in this context includes at least timeliness, ease in interpreting results, physical defensibility, and provision for uncertainty.

4. COMPUTING AN ESTIMATED DOSE

4.1 Source Specification

Emergency response model systems normally include a library of source configurations, covering stationary or moving sources of multiple shapes, sizes, and directions of release. Submodels covering initial plume rise, explosive spread, and downwash (from a stack or the roof of a building), are available for application as required by the situation (e.g. Lee *et al.*, 1997, Leone *et al.*, 1997, Sykes *et al.*, 1998). Important details, however, will usually be missing. Such information as exit velocity (including direction) ambient and effluent temperature, chemical constitution, atmospheric stability, wind, and turbulence may only be rough estimates. A Monte Carlo approach is commonly used to assess the significance of this uncertainty, but its full implementation is time-consuming in an actual emergency. Generally something is known about the source, even if only as bounds on conditions. If the model framework provides access to rapid execution of several test scenarios the knowledge can be applied. The risk of some undesirable outcome can then be usefully assessed in the time available.

Source specification can be critical when the wind changes strongly with height. In such a case the primary transport direction and perhaps the nature of the cloud's spread can depend strongly on the effective release height. If there is significant puff rise due to buoyancy, explosion, or

4

high exit velocity, the effective release height may be uncertain. It is likewise important in such situations to have good knowledge of the wind field.

For "large enough" sources, if the sources' location, size, and shape are known, their strength can be computed to considerable accuracy using backward (or receptor-oriented) dispersion methods (Flesch *et al.*, 1995; Uliasz, 1993). These methods determine the history of transport and dispersion that produced the concentration found at the receptor point at the time of observation. They work best if the sources' linear dimensions are nontrivial fractions (a few percent) of the travel distance to the receptor. Such sources include evaporating pools and airbursts. A "hybrid receptor" model is another option. This is an ordinary source-oriented model with unit source strength. The actual source strength is then estimated as the ratio between the observed concentration at a receptor and the concentration forecast for that receptor.

4.2 Wind Fields and Meteorological Environment

A NARAC demonstration at Los Alamos, New Mexico (Bowen *et al.*, 2000) illustrated the importance of considering the entire wind and turbulence field in computing dispersion of contaminants. Coastal regions and complex terrain share a propensity for strong changes in wind direction with height. In this case, wind measurements showed direction sheer greater than 45° from 12 m to 92 m above ground. To isolate the wind-field's influences from those of the release conditions, the authors specified a hypothetical release, only in the model, at a height of 45 m. This altitude maximized the effect of the wind variations. ADPIC, the Advection-Diffusion Particle-in-Cell model of NARAC simulated the dispersion using the real wind field. Wind was measured at 12m, 36m, and 45m above ground at each of four towers, one of which also sampled wind at 92 m AGL. Results using only the wind measured at 12 m were strikingly different from those using all available winds (Figure 1). In complex terrain, failure to consider the full three-dimensional structure of the wind can lead to spectacular errors, especially under stable conditions where vertical mixing is suppressed.

Yet the quality of the wind data and the nature of the release must also be considered. Chang *et al.* (2000) tested three emergency-response models against the Dipole Pride 26 (DP26) set of field data. This experiment of November 1996 was run by DTRA (Watson *et al.*, 1998; Biltoft, 1998). The experiment was in a valley in Nevada, about 40 km wide and 800 m deep. There were three 10 km sampler lines perpendicular to the valley axis, roughly 7 km apart, having 30 samplers each for the SF_6 tracer. Release height was 6 m above ground. Samples were taken 1.5 m above ground. Eight meteorological stations reported 15-minute averages of surface conditions. Upper winds were sampled from two sites, one launching pilot balloons, the other launching both pilot balloons and rawinsondes. The Hazard Prediction and Assessment Capability's (HPAC) model, SCIPUFF, was run with three different wind inputs: a sophisticated diagnostic model giving wind and turbulence in three dimensions, a mass-
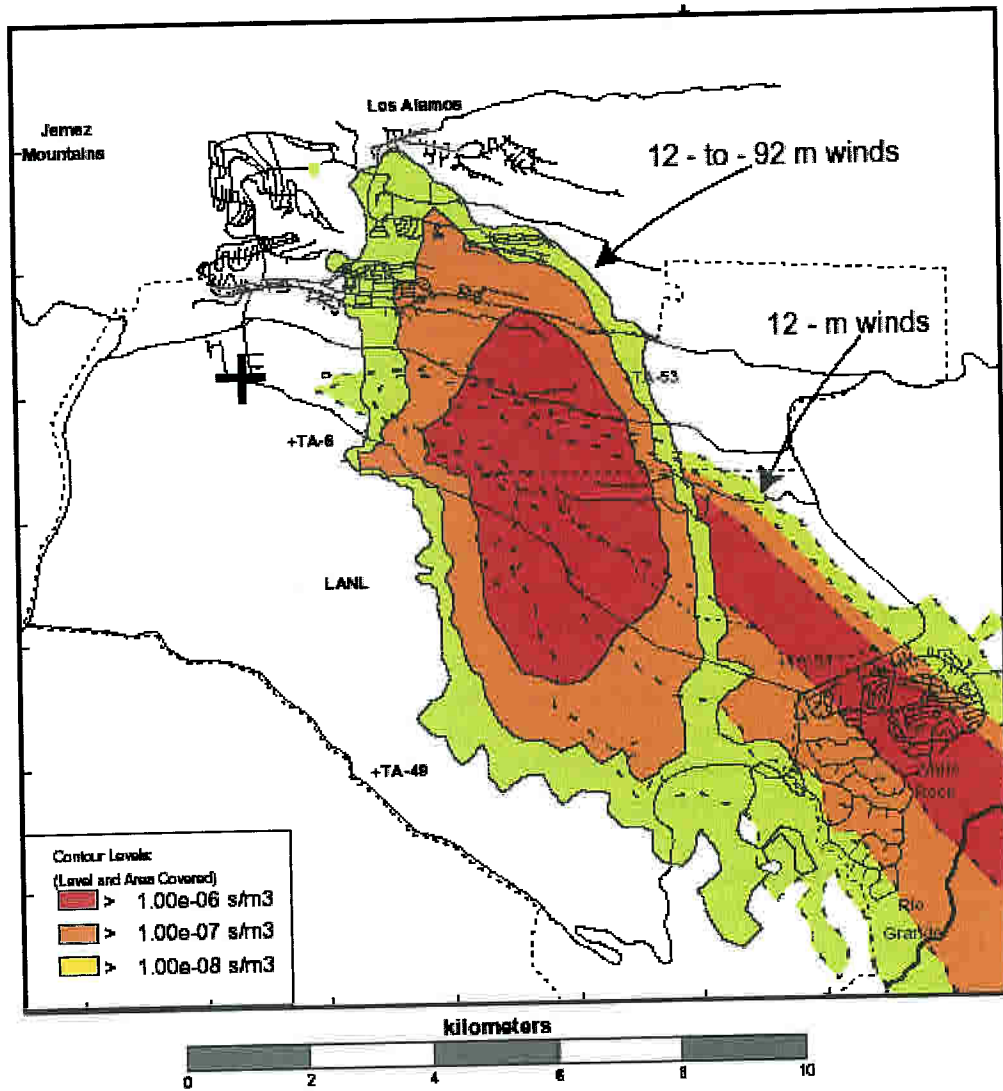
Figure 1: Dispersion forecast using only surface winds for an elevated release in a sheared environment can be spectacularly wrong (from Bowen, et al., 2000)

consistent wind model also using the upper wind observations, and the same mass-consistent model using only surface observations. The investigators unexpectedly found the least-sophisticated wind input to provide nearly as good a forecast as the most sophisticated. Both the release and the receptors were near the ground, along with the densest meteorological measurements in time and space. The upper-air data appear to have been inadequate to give advantage to the greater sophistication of the highest-level diagnosis scheme. Furthermore, the controlling winds were probably at the surface.

The two examples, both from stable conditions in complex terrain, are fairly extreme, the one contrived to maximize the effect of wind sheer aloft, the other apparently having minimal influence from winds aloft. They illustrate some of the complexities in estimating wind fields.

6

There are three primary ways in which the wind pattern can be determined: Interpolation, optimal adjustment, and full four-dimensional modeling (space and time), based on the Navier-Stokes Equations. The Navier-Stokes computations can further be guided toward observations made over the computational domain through the simulation period in a process called four-dimensional data assimilation (FDDA).

In principle, the dispersion environment (wind and turbulence) is best directly measured in sufficient detail to capture its relevant features. If the data provide all necessary constraints, interpolation can fill in the remainder. Interpolation uses a weighted average of the measurements. The weights vary smoothly over the region, summing to unity at any point in the region. The weight given to observations from each measurement site generally decreases with distance from that site, often as the inverse square of that distance. Interpolation procedures can be sophisticated, taking secondary (proxy) information, such as terrain configuration into account. Such techniques include nonlinear least squares and kriging. They compute interpolation estimates through some type of optimization, usually minimizing variance and bias (e.g., see Chilès and Delfiner, 1999). If their assumptions are valid, they can also assess the error of estimation. Such sophisticated schemes are, however, sensitive to violations of their assumptions and complex to compute.

Being purely statistical, interpolation depends on the ability of the underlying set of measurements to represent all the important features of the wind field. Chang et al., (2003) examined the effect of this requirement in a strongly heterogeneous wind field at night in high-relief complex terrain. They used the DP26 Experiment, previously described. Depriving the dense surface network of one site at a time, they computed the dispersion of tracer from the source at 6 m AGL, comparing the results to samples taken at 1.5 m AGL. Some sites were highly important, others less so. The important sites tended to be those sensing strong winds, especially near the source. In some cases, however, a sensor was important for revealing a strong gradient of wind, not otherwise known to the network. DP26 was markedly horizontal and well sampled in that plane. Generally the field must be known in three dimensions, including the three wind components. Since sampling aloft is often inadequate, going beyond mere interpolation is attractive.

A modicum of physical knowledge can be injected by minimally adjusting the interpolated values to satisfy some constraint, such as conservation of mass (e.g. Sugiyama and Chan, 1998; Ross et al., 1988). If temperature profiles are available along with terrain information, the partitioning of flow over and around hills may be varied as a function of local stability. Air in a cold pool will then stay in the pool, while less stable air above can flow over the surrounding elevated terrain. If momentum balance is included in the objective function, such secondary flow features as air drainage down a valley or slope may be represented. Such adjustment schemes cross from interpolation into extrapolation, in that they infer features not explicitly measured in the data. Their inference is based on the assumed physical model from which the constraints are derived. Like all extrapolations, they can give seriously spurious results if the physical constraints, necessarily simplified, are imposed blindly. Using minimal adjustment addresses this limitation, keeping the final result as close as possible to the original data. The adjustment, however, is minimized only globally (usually least squares). Large changes are possible in a small subregion if everywhere else the adjustment is minor. Chang et al. (2003)

used the CALMET scheme deliberately inappropriately to drive the SCIPUFF dispersion model. The appropriate CALPUFF dispersion model, using only the horizontal components of the CALMET output, routinely produced good dispersion forecasts. In extrapolating surface observations to blend with upper-air data, however, CALMET's mass-consistent constraint often developed strong vertical components. SCIPUFF, using full three-dimensional advection, underpredicted concentration at the ground when these spurious vertical components lofted its puffs. Adjustment techniques applied with understanding can overcome some shortcomings of the original measurement deployment, but resilience to inadequate data remains limited.

Dense measurement deployments are normally found only around major installations such as military bases and chemical plants that work with hazardous materials in large quantities. In principle, if mesoscale coverage of a region by measurements is inadequate, a mesoscale model of atmospheric dynamics can compute the winds and turbulence for the area. This crosses fully into the realm of extrapolation, developing fine structure not at all present in the meteorological input. Mesoscale models are at their most useful in situations, such as shorelines or complex terrain, where fine-scale forcing is significant and knowable. High sophistication is required, and many well-known models attempt to provide it, including the US Navy's Coupled Ocean-Atmosphere Mesoscale Prediction System (COAMPS; Hodur, 1997; Doyle and Bond, 2001), the Fifth Generation Mesoscale Model (MM5) of the Pennsylvania State University and the National Center for Atmospheric Research, (NCAR 1999, Grell *et al.*, 1994), the Regional Atmospheric Modeling System, (RAMS, 2002), and the Higher-Order Turbulence Model for Atmospheric Circulation (HOTMAC; YSA, 1998). These models produce detailed representations of turbulence, precipitation, convection, secondary circulations, and other processes important to pollutant dispersion. Detailed digital terrain maps are available for the whole US and for much of the world. The mesoscale mesh can be nested in routine forecast fields coming available at increasingly fine resolution (12 km, as of Rogers, *et al.*, 2002). Standard techniques also exist for assimilation of existing data in four dimensions (*e.g.* Umeda and Martien, 2002)

Recently, adaptive gridding has been introduced into mesoscale modeling for dispersion application. Highly detailed models (*e.g.* Cybyk et al., 1999a,b, 2000) have been used to examine dispersion in urban boundary layers, as will be discussed later. Models such as OMEGA (Bacon *et al.*, 2000) take advantage of adaptive gridding as an alternative to nested grids to improve flexibility. Nesting requires multiple overlapping grids. Each of these should be approximately square in the horizontal for good modeling practice, and each has boundaries to treat. Adaptive gridding incorporates all mesh refinement into a single grid. Only this grid need be square. Its areas of fine resolution may take any shape. Transition of resolution is smooth, eliminating boundaries, though some special requirements remain. Such a grid can in a natural way provide fine resolution in multiple unconnected subregions of any desired shape. For example, a complex shoreline may divide a simple water surface from a simple land surface. If fine resolution is applied only to the serpentine region following the shore, far fewer grid volumes are required. The grid adaptation can even be employed dynamically at run time in response to developing strong gradients in the flow field or the contaminant pattern.

Though adaptive regridding shows promise, there are reasons to delay general adoption. The application is new and only barely explored for assets and liabilities. Existing codes are few and hard to obtain. The digital approximations are more complex than standard finite differences.

An efficient grid generator is required. For dynamic grid adaptation, the grid generator is part of the execution, adding overhead. Even for a static grid, finer spatial resolution requires a finer time step. Furthermore, submodels for turbulence and the like are valid only in a certain range of scales. Where the mesh resolves finer scales, transition to a different submodel is required. For the present, adaptive regridding schemes remain mostly the province of research.

Routine use of ordinary mesoscale modeling for emergency response must also contend with some hurdles. One important issue is the proper balancing of initial conditions to minimize transients. The balance already available from the outer model mitigates this problem, but does not eliminate it. The finer model operates under different assumptions, modeling different physical processes, or the same processes in more detailed ways. The shorter the time available to adjust, the stronger the transients will be, introducing errors into the transport and dispersion calculations. Thus, mesoscale modeling for emergency response applies best if the model is already running at the time of need. With the explosion of computer speed and capacity this has become increasingly practical. Even in a fully adjusted model, however, there remain important sources of irreducible variance (Hanna and Yang, 2001). Moreover, being extrapolations into fine scale, mesoscale models are vulnerable to small variations in specification. A wide ensemble of outcomes may be derived from the same initial and boundary conditions. Consider, for example, the interaction between a plume of contaminants and a thunderstorm. The model may form a thunderstorm in a random location. If its updraft region acquires the plume, the contaminants may be removed from the boundary-layer air altogether, though perhaps deposited in the rain. If not, they may be deflected in quite another direction by downdraft gusts.

One way to treat the uncertainties in the atmospheric environment is to estimate them through use of an ensemble of forecasts slightly varied according to some protocol. There is some interest in the adoption by NWS of ensemble forecasting in preference to forecasting single scenarios at high resolution (Zhu et al., 2002). This benefits estimation of the expected uncertainty in a dispersion forecast, at the expense of having to extrapolate over a greater ratio to achieve the necessary fine scales. The manner of selecting members for the ensemble may furthermore be optimized for some synoptic-scale purpose rather than for mesoscale dispersion forecasting.

Ultimately there is no simple answer to the question of the best source of meteorological data for dispersion calculations. Ideally one has wind observations available at sufficient resolution to describe the wind field in three dimensions and time. That density of observations, if achievable at all, is prohibitive, especially in complex terrain. Lewellen and Sykes (1989) considered a five-dimensional space, adding an ensemble dimension to the usual four. They argue for acceptance of wider spread in the ensemble dimension to relax constraints demanded of measurements in the physical dimensions. Intensive numerical and field experiments can assess the range of possible outcomes, given practically obtainable data. As a side benefit to this effort, some information, such as the variance of horizontal wind, may be found to have particularly strong influence on the outcome of a dispersion calculation. Such information could then be measured more carefully. It is impossible to sample any particular scenario adequately to define which ensemble member is realized in it. For the foreseeable future assessing the hazard from an unexpected release in a random location will retain a large element of judgment and experience.

4.3 Pollutant Dispersion

Dispersion of a trace atmospheric constituent is generally modeled in three primary ways: the Eulerian mass budget, particles in cells (PIC), and Gaussian puffs. The Eulerian mass budget equation is normally implemented as one of the equations of a dynamic model of atmospheric flow directly computing the concentration of its constituent (*e.g.* Uliasz, 1993; Bacon *et al.*, 2000). Being integral to the overall model, such a form handles chemical interactions and other complex physical processes directly. Concentration variance can be treated in a straightforward way through an additional equation. Concentration patterns on scales unresolved by the grid are not represented, a particularly severe limitation close to a source. Conversely, the Eulerian mass budget is quite efficient for widely distributed contaminant.

The (Lagrangian) particle-in-cell (PIC) framework tracks a large number of particles. Its ideal situation is short-range transport from a point source to a volume receptor. It readily handles multiple sources under complex conditions. Each particle is a point, carrying specified mass, released sequentially, and transported by mean and turbulent flow. Concentration is determined by counting the particles in the receptor volume, often a grid volume of the model. Chemical interactions are difficult to represent. Gravitational settling might appear natural, since the dispersant is particles. These particles are, however, only carriers of the contaminant's mass. They have no intrinsic physical counterpart. Correct representation must be provided of the size distribution and other characteristics of the actual particles. The primary limit on PIC is the number of particles required, especially for long-range transport or complex conditions. Determining concentration fluctuation may require even more particles (but see Thompson, 1990).

The Gaussian puff framework tracks a sequence of overlapping puffs. Each puff carries a specified mass of contaminant, all hypothetically released at the same known instant. Its spatial distribution in three dimensions is assumed to match the trivariate Gaussian probability distribution. Puffs, like PIC use the Lagrangian framework, though far fewer puffs, and computer resources, are required to represent the same distribution of contaminant. Turbulent motions on scales small compared to the size of the puff act to disperse (thus enlarge) the puff. Motions on scales large compared to the puff act to move it bodily. Scales of motion comparable to the size of the puff deform the puff. Normally, deformation is avoided by splitting the puff into smaller parts when it gets large enough to be resolved by the wind-field grid. Deformation can be treated explicitly at the price of additional complexity, as can concentration fluctuations (Sykes *et al.*, 1998).

The remainder of this section expands on the two Lagrangian dispersion schemes.

The Gaussian puff models track objects that grow and may change shape with time, requiring appropriate algorithms. These objects (puffs) are generally transported only by the resolved flow, with the turbulence represented by the puffs' changes in shape and size. The scale treated as "turbulence" is thus of puff size or smaller. Empirical relations between travel time and puff growth may be used to estimate the effect of turbulence. Often the spread is assumed isotropic in the horizontal, differing only in the vertical. If a puff expands into the resolved scale, it is customarily split into smaller puffs to retain the explicit influence of the resolved patterns.

The more sophisticated puff models permit full three-dimensional stretching and rotation. Since this complexity is beyond the reach of field experiments, the puffs' growth and deformation are linked theoretically to local properties of atmospheric turbulence. The theory draws on the extensive study of natural and engineering turbulence over the past few decades. In principle it provides automatic applicability to a broad range of conditions. The puffs retain the Gaussian shape, a limit on the degrees of freedom that markedly reduces complexity, at some expense to flexibility. The primary operational model using this concept is the Second-moment Closure Integrated Puff (SCIPUFF) algorithm of Sykes *et al.*, (1998, 1986). This will be discussed at some length here since the argument used in developing the model contributes generally to understanding of the dispersion problem.

In the SCIPUFF model each puff has ellipsoidal shape, its contaminant mass distributed according to the trivariate Gaussian pattern. Ten moments describe this configuration. Angle brackets indicate integration of the given puff's mass density over all space. The zeroth moment, or total mass Q of contaminant in the puff, provides the scale from which to determine the concentration (mass density) c anywhere in the puff.

$$Q = \langle c \rangle$$

The three first moments with components of the location vector x_i give the coordinates of the center of the puff's mass.

$$Q\overline{x_i} = \langle c x_i \rangle; \quad i = 1, 2, 3.$$

The six independent second moments, a generalization of the familiar parameters of puff width, define the spread and orientation of the puff's ellipsoid shape.

$$Q\sigma_{ij} = \left\langle c\left(x_i - \overline{x_i}\right)\left(x_j - \overline{x_j}\right) \right\rangle; \quad i = 1, 2, 3; \quad \sigma_{ij} = \sigma_{ji}$$

In terms of the trivariate Gaussian pattern, they form the covariance matrix. The time tendencies of all ten moments may be derived by integrations involving the budget equation for the contaminant.

$$\frac{\partial c}{\partial t} + \frac{\partial}{\partial x_i}(u_i c) = k\nabla^2 c + S$$

These integrations treat each puff in isolation. The ensemble-average concentration over all puffs is found by superposition (simple addition). Determining the variance of concentration, however, requires evaluation of nonlinear interactions among puffs.

The budget equation, integrated over all space but considering only the concentration of mass belonging to the current puff of interest, reduces to the source term for that puff. This controls the total mass in the puff. Sources include deposition or resuspension at the ground, chemical transformations, and the like. Multiplying the budget equation by the coordinate variables x_i, $i = 1,2,3$, and integrating as before yields the three tendency equations for the center of mass. Multiplying the budget equation by the six dissimilar components of $x_i'x_j'$ and integrating yields the tendency equations for the covariances, which parameterize the puff's spread and shape. The primes indicate displacement from the center of mass.

The complexity of ten equations buys explicit representation of several important physical relations controlling the spread and deformation of the puffs. In particular the vertical shear of horizontal wind explicitly stretches and rotates a puff. This is a useful feature in the atmospheric

boundary layer, especially in complex terrain or on a seacoast. The ten equations, of course, retain unknown turbulent quantities. These primarily describe the in-puff turbulence that spreads the puff over time. They are, however, now formulated in terms of covariances between the turbulent velocity components and concentration fluctuations: $\overline{u'_i c'}$.

These covariances operate on the scale of the puff, intentionally unresolved by the explicit wind field. They are determined through the long-established technique of second-moment closure. The moments involved here are the $\overline{u'_i u'_j}$ and $\overline{u'_i c'}$, not to be confused with the ten moments introduced above to describe the Gaussian puff. Second-moment closure, in principle, explicitly represents the case-dependent aspects of turbulence, such as generation, applying empirical models only to dissipation and other aspects more amenable to universal empirical description. Second-moment closure provides a theoretical link between measurable atmospheric turbulence and the motion, spread and deformation of the Gaussian puff. A hurdle remains. Traditional higher-moment closure considers the various generators of the turbulence, such as buoyancy and shear, to transfer energy between ensemble-average flow and the departure therefrom. Such a formulation convolves all the departure (i.e. turbulent) flow's scales of space and time. Scale is important, however, to contaminant puffs. Turbulent eddies on scales larger than the puff will serve to move or rotate it, while those on scales smaller than the puff will serve to spread it. Using the total measured or modeled atmospheric turbulence will overestimate the rate of spread of a puff unless the characteristic scale of the turbulence just happens to be comparable to the size of the puff. Knowledge of the spectral distribution of turbulent energy is needed. Sykes *et al.* (1998) have developed ways to treat this problem.

The TRAC (Terrain-Responsive Atmospheric Code; Ciolek and Magtutu, 1998; Alpha TRAC, 2002) uses a different approach. Six points define each puff. These are specified initially according to the source algorithm and later allowed to travel independently over limited periods. They travel by mean wind in the resolved field and by turbulent departures, the latter randomly drawn for each point from a distribution determined by the local turbulence statistics. Periodically the locations of these points are used to determine the puff's location, shape, and size. At such times, the points' positions are adjusted to maintain a common point of intersection of the three lines joining pairs of the points. In this way the puffs maintain their regular ellipsoidal shape on which a Gaussian distribution can be defined. Other realizability conditions are also enforced on the points, particularly that the puff grow monotonically, whether there is mean wind or not. This scheme also allows puffs to stretch and rotate in response to mean shears and differential turbulent fluxes. The algorithm rests more on a phenomenological concept than does SCIPUFF. Most likely turbulence statistics obtained from the atmosphere or from a model need to be adjusted in defining the random component of transport. This avoids attributing the total turbulent energy to the scales of puff size and smaller.

The PIC model (Lange, 1978; Legg and Raupach, 1982; Thompson, 1987, 1990; Leone *et al.*, 1997) tracks point masses, rather than puffs. These are far more numerous (hundreds of thousands) and remain unchanged throughout their lifetimes in the model domain. As points, they disperse both by mean and turbulent winds. The mean winds are defined on a grid by one of the methods previously described. The turbulent winds u_i' may be given deterministically as

$$u_i' = -\left(\frac{K}{c}\right)\frac{\partial c}{\partial x_i}$$

where c is the concentration of contaminant, $\frac{\partial c}{\partial x_i}$ is the slope of the concentration of contaminant in the x_i direction, and K is the diffusion coefficient for the contaminant. Thus defined, u_i' always points down the local gradient of c.

A more sophisticated approach, however, is to define u_i' by a generalized Langevin equation, following the assumption that the position and velocity of a particle evolve jointly according to a Markov series expressed by the discrete (finite dt) form of :

$$du_i = a_i(x_k, u_l, t)\,dt + b_{ij}(x_k, u_l, t)\,d\xi_j$$
$$dx_i = u_i\,dt$$

Here the summation convention applies; $i, j, k,$ and $l \ \varepsilon \ \{1, 2, 3\}$; and $d\xi_j$ is a random number from a Gaussian distribution of mean zero and variance (not standard deviation) dt. The Markov series by itself is insufficient constraint to enforce physical realizability. A well-mixed distribution of particles must remain well mixed, a condition satisfied if and only if (Thompson, 1987) coefficients a_i and b_{ij} also satisfy the Fokker-Planck equation:

$$\frac{\partial g_a}{\partial t} = -\frac{\partial}{\partial x_i}(u_i g_a) - \frac{\partial}{\partial u_i}[a_i(x_k, u_l, t)\,g_a] + \frac{\partial^2}{\partial u_i \partial u_j}[B_{ij}(x_k, u_l, t)\,g_a]$$

Here $B_{ij} = 0.5\,b_{ik}\,b_{jk}$. The probability density function $g_a(x_k, u_l, t)$ for Eulerian wind velocity is normally modeled as multivariate Gaussian. It views the atmosphere as a collection of particles, rather than a continuous fluid. It represents the density of probability at time t in position x_k that a fluid particle will be found having velocity u_l. Legg and Raupach (1982) present the physical conditions on a_i and b_{ij} more readably, though perhaps in less generality. Additional developments of particle models enforce the well-mixed condition across interfaces, such as at the top of a mixed layer, where the particle model may experience discontinuous changes in flow properties from one grid volume to the next (Thompson et al., 1997).

To address the expense of tracking a large number of particles over long distances, hybrid models have been used. One form uses particles (or puffs) close to each source. The clouds are allowed to grow until their characteristic scale is resolvable by the model grid. At that time and place the particles (puffs) pass their mass as a volume source to the general mass budget for that constituent. An equation of the Eulerian model keeps track of this general budget. Such a form is useful for slow chemical reactions, such as photochemical smog, some reagents of which originate in point sources. For emergency response this hybrid mode may be rarely invoked.

Another hybrid makes use of the large aspect ratio of atmospheric scales horizontal to vertical. It treats a cloud of contaminant as particles in the vertical and puffs in the horizontal (Hurley, 1994, Draxler and Hess, 1998). In contrast to full particles, which are point masses, these objects (PARTPUFFs) are two-dimensional masses, of zero thickness. Since they cover the horizontal distribution largely as puffs only 1% as many PARTPUFFs need be released as particles. Limiting the spread to horizontal simplifies the puff formulation as well. It has earlier

been noted that atmospheric motions on the scale of a puff deform the puff. Such deformation is treated in three dimensions in a careful, but fairly complex manner in SCIPUFF, as described above. Since the bulk of this complexity comes from puff deformation in the vertical, PARTPUFFs represent considerable simplification. With some justification PARTPUFFS can be assumed circular allowing a univariate Gaussian distribution to describe their horizontal spread. Being particles in the vertical, they simply reflect from the surface or any elevated reflecting layer. The distributed dimensions remain always horizontal, though in terrain-following coordinates this is transformed horizontal, not constant geopotential. In particular, the entire object is assumed to make simultaneous contact with the ground. Concentration is computed at horizontal grid points by summation over all PARTPUFFs within a vertical interval z. The contribution by a PARTPUFF to the concentration at the grid point a horizontal distance r from its center is

$$\Delta c = \frac{\Delta m}{2\pi\sigma_y^2 \Delta z} \exp\left[-0.5\left(\frac{r}{\sigma_y}\right)^2\right],$$

where c is the increment in concentration, m is the total mass of the PARTPUFF, σ_y measures the horizontal spread, and z is the depth of the vertical layer in which the PARTPUFF currently lies. PARTPUFFs contribute to surface concentration if and only if they are within z of the surface.

Tests in the convective boundary layer, following the measurements of Willis and Deardorff (1978, 1981) demonstrated results equivalent to a Lagrangian particle model with 1% as many objects. In a sea breeze, the PARTPUFFs encountered sufficient horizontal shear at the sea breeze front to produce a noticeable departure from the particle simulation. Otherwise the agreement was satisfactory, using again only 1% as many objects. The sea breeze test also demonstrated the effect of resolution too coarse to resolve important features of the contaminant distribution. Counting particles over too large a volume underreported the peak concentration by a factor of two. The horizontal distribution of the PARTPUFFS provided a more realistic result, but improvement for both approaches was achieved by halving the horizontal grid spacing in both directions. Four times as many particles had to be released to accommodate this change, while there was no change in the number of PARTPUFFs. One should, however, note that refining the vertical resolution affects particles and PARTPUFFs alike. A third demonstration involved flow past a hypothetical Gaussian hill in three dimensions simulating diurnal variations between 0000 and 1200 local time. Dispersion from a nighttime source above the nocturnal surface layer at half the height of the hill was simulated almost identically by both methods. About 25 times as many particles as PARTPUFFs were required. The complex situation involved elevated plume impact and local slope flows down the hillside. Both methods also matched in simulating a near-surface release, which exhibited splitting of the pollutant cloud around the base of the hill during the stable nighttime. Daytime patterns were inadequately represented by both schemes. The concentration maximum was very near the source in the convective mixed layer, forming a pattern too fine to be resolved by the horizontal grid. Improving the horizontal resolution, however, would have far less effect on the PARTPUFF simulation than on the particles.

Dispersion at night remains a serious deficiency in all dispersion modeling. This involves highly complex structures on such small scales that turbulence and transport are determined by local

conditions. The structures are decoupled from the ground and from any other larger-scale feature that can be reliably treated by models or observation systems. Organized features appear only intermittently and unpredictably in our present state of knowledge. Numerous studies of such nocturnal features as slope flows in complex terrain provide hope of understanding their role in nocturnal dispersion. A few examples include Atmospheric Studies in Complex Terrain (Clements *et al.*, 1989), and the Vertical Turbulence and Mixing Experiment and URBAN 2000 (Doran *et al.*, 2002; Allwine *et al.*, 2002). New algorithms continue to be developed for the less severely stable conditions (Zilitinkevich, *et al.*, 2002). There may be hope for improved understanding of nocturnal turbulence through computational fluid dynamics and large-eddy simulation. Until computers become considerably more powerful, however, these approaches will be severely limited by computational requirements, even in research mode

4.4 Deposition

Removal mechanisms, by precipitation, by gravitational settling, and by absorption on the surface are important for estimating atmospheric concentration as well as for estimating the amount available to other pathways of harm. Most emergency-response models have some level of treatment. Reflection of Gaussian puffs at the ground can be partial, using deposition velocities for contaminants to determine the deposited fraction. Mass found to be deposited is then removed from the appropriate puff. Wet deposition, a powerful removal mechanism for airborne contaminants, involves complex interactions among droplets, gases, and particles. Simplified algorithms exist (*e.g.* Sykes *et al.*, 1998) which consider the size distribution of both the raindrops and the particles. Much depends, however, on where a mesoscale model locates thunderstorms or ordinary showers.   Since clear mesoscale forcing for such features is unusual, their location is likely to be random. It may be argued that the location and timing of the precipitation is a greater source of uncertainty than is coarse treatment of the wet deposition process itself.

4.5 Convective Storms

The enhancement of dispersion by convective storms is only minimally treated in modeling and observations. As part of the continued merging of contaminant dispersion with meteorology, the results of the considerable study of organized mesoscale phenomena need to be tapped (*e.g.*, Flossmann and Pruppacher, 1988, Respondek *et al.*, 1995). A thunderstorm not only produces washout, but also dramatically alters the wind field. The outcome of transport by such a modified flow depends on the initial location of the pollutants. If they are lifted with the rising warm air, they may be scavenged from a larger area than that on which the rain falls. Furthermore they may be injected into the middle and upper troposphere, not to be returned to the surface locally at all (Lyons *et al.*, 1986; Thompson *et al.*, 1994). This discussion underscores the potential for widely divergent forecasts of dispersion driven by forecasts from a mesoscale model.

4.6 Resuspension and volatilization

Particles and droplets deposited on the surface can be returned to suspension in the atmosphere by wind gusts or simply by evaporation. The surface thus becomes a reservoir from which hazardous material is reintroduced into the air. Though often important for cleanup after an event, resuspension is negligible during the first few hours, compared to direct contamination from the source. Thus, resuspension will not be considered here.

4.7 Dose Calculations

Most emergency emissions have limited duration, and most effects vary with the dose received. Modeling of the dose to which a receptor is exposed involves the same issues as determining the average concentration over a period. The instantaneous situation may be conceptually separated into the usual two regimes. Turbulence on scales large compared to the puff's width will transport the puff as a single object. Turbulence on scales comparable to or smaller than the puff's width will mix and spread the puff. Introducing summation smears the two together. If only small-scale turbulence were present, a given receptor would experience minimal change in concentration over the duration of the release since each emitted puff follows the same path from the source. Large-scale fluctuations, however, nearly always exist, sweeping the puffs over multiple trajectories. A few puffs may pass directly over a receptor, while others miss it entirely. The longer the averaging time, the more of the meandering component of the wind will be included.

Sykes and Gabruk (1997) address this issue with a second-moment model of turbulence. Second-moment closure is not the only valid approach. The empirical relations that estimate puff spread over time can serve if the wind information explicitly resolves fluctuations on the time scale of their characteristic average, about 600 s. It is then possible to release a large number of model puffs, allow them to meander with the resolved wind and sum their concentration over time at the receptor of interest. The puffs must be split when their size exceeds that corresponding to the 600-s time scale.

Sykes and Gabruk (1997), however, provide a framework for understanding the issues of averaging time and dose. Though they considered plumes, their argument is readily adapted to puffs. Consider a puff, horizontally isotropic; of horizontal size l_y and vertical size l_z. The assumption of horizontal isotropy is simply convenient, not necessary. Scales l_y and l_z are related to the standard deviations of a Gaussian puff, but may be larger. For instantaneous dispersion, only that part of the spectrum of atmospheric turbulence on these scales and smaller contributes to the spread of the puff. The remainder of the turbulence spectrum moves the puff bodily. In simulation it must be supplied explicitly by the resolved wind field or by random sampling of wind components from a distribution having variance given by integrating the turbulence spectrum over all scales larger than l_y and l_z.

If an average, or aggregate, cloud is to be considered, however, some of the bodily transfer of the instantaneous puff becomes part of the spread of the aggregate puff. Thus, the effective length scale of the aggregate cloud is larger than the instantaneous scales l_y and l_z. Simulation of the

16

aggregate divides the turbulence spectrum according to this greater length scale. If the aggregation time corresponds to the integral length scale of atmospheric turbulence or larger, the entire spectrum of turbulence acts as a dispersant.

5. RELEATED ISSUES

5.1 Concentration fluctuations

Emergency response models are concerned with the outcome of a single realization. The ensemble average, to be useful, needs to be accompanied by a measure of the range of departure from that average. As already noted, fluctuation of the instantaneous concentration about the ensemble average is usefully divided into two mechanisms. Local entrainment is produced by eddies of size smaller than or comparable to the width of the plume. Meandering of the whole plume is through eddies of much larger size than the width of the plume. The former can be treated using second-order closure of turbulence. The latter may be treated statistically or explicitly in the resolved wind field. One advantage to the theoretical basis of SCIPUFF is the ability to extend naturally to consideration of concentration fluctuation. Data limitations in an actual event, however, are likely to render the advantages of such estimates unrealizable. Meander is of greatest importance near the source, where the plume is narrow. As the width grows, fluctuation within the puff becomes more important, eventually dominant.

Under convective conditions near the source, especially with an elevated source, the fluctuations can be as much as six times the mean (Weil *et al.*, 1992, Deardorff and Willis 1984). Puff width near the source is narrow: comparable to or smaller than the convective plumes in the atmosphere. Therefore, these plumes intermittently bring very high concentrations to the ground, which otherwise has low concentrations. Because of this mechanism, the concentration fluctuations of a narrow plume in strongly convective conditions can be conceived as an exponential stochastic process, which models the waiting time for random events (such as puff strikes) to happen.

With larger clouds and less severe mixing, concentration distributions are better conceived as a series of random dilutions due to entrainment of clean air. Each dilution is a multiplication by a factor less than unity (Csanady, 1972). Thus the logarithm of the concentration is the sum of a large number of random values, the logarithms of the dilution factors. Such a sum tends toward the Gaussian distribution as the number of elements increases. Hence its antilog tends toward the lognormal distribution.

Sykes *et al.* (1998) note that the concentration variation due to meandering by large-scale winds is independent of the fluctuation due to turbulence internal to the plume. Thus they can be treated separately, deriving turbulence within the plume from second-moment closure, and treating meandering explicitly in the wind field, or statistically with a separate meander-induced variance of concentration. The exception is persistent vertical shear of horizontal wind, which produces deformation on space scales short enough to be significant to the puff, but time scales long compared to the internal turbulence scales of the puff. Shear's effect is to stretch the puff in

17

the horizontal while squeezing it in the vertical. Vertical gradients increase, making mixing more efficient and decreasing the dissipation time scale for the fluctuations. This effect can be treated as a special case.

Such computations as given here provide an estimate of the expected departures of the actual concentrations from their predicted means, provided the departures arise from the stochastic nature of atmospheric diffusion. Departures due to erroneous input data and inappropriate model assumptions are not covered. Since data deficiencies are normal in most emergencies, especially over the time scales covered by this report, a "what-if" mechanism is operationally useful. Rapid recomputation under multiple scenarios allows quick assessment of the sensitivity of the forecast to variations in input when data are inadequate for full treatment of concentration fluctuations. Running an ensemble of different models provides the best approach to questions of propriety of model assumptions, though this may require an impractical commitment of resources and time.

Draxler (2003) assessed the effect of one type of input error: mislocation of the source. He used the HYSPLIT model on the continental scale to simulate the ANATEX (Across-North-America Tracer Experiment, Draxler et al., 1991). He used meteorological data from the NCEP/NCAR Reanalysis (Kalnay et al., 1996) on a grid of 136 km horizontal spacing and about 250 m vertical. He formed an ensemble of 27 members by shifting the wind field plus and minus one grid point in each direction. Moving the wind field rather than the source maintained the same relation between source and receptors in all members of the ensemble. Thus, mislocation only of the source and receptors was simulated in an otherwise correctly specified atmospheric environment. The resulting ensemble of computed trajectories spread continent-wide. The majority of the trajectories, however, traveled south and east from the source at Glasgow MT across the Dakotas, Iowa, and Missouri.

Draxler interpreted this outcome using the probability of exceeding (PE) a specified threshold concentration. A related scheme considered the concentration to be exceeded with a specified probability ("concentration probability," CP). He determined distribution of PE by counting the fraction of ensemble members which forecast a concentration exceeding the specified threshold (e.g., 10 pg m^{-3}). The CP he determined by finding the concentration exceeded at each point by the specified fraction (e.g., 10%) of ensemble members' forecasts. He assumed each outcome in the ensemble to be equally likely in this determination. The highest measured concentrations generally fell in the regions of highest PE and CP, though with some strong exceptions.

Warner, et al. (2002) examined the effect of variations in atmospheric inputs and model assumptions. They developed an ensemble of twelve members using boundary and initial conditions from three different large-scale analyses and applying three different models of boundary-layer turbulence and two of soil physics. The MM5 system, configured and parameterized in the twelve different ways, computed the atmospheric environment for SCIPUFF to track a hypothetical release from Khamisiyah, Iraq. SCIPUFF was used in "explicit" and "ensemble" modes. In the explicit mode SCIPUFF was attached to each individual MM5 run to create an ensemble of dispersion outcomes. In the "ensemble" mode, a single run of SCIPUFF estimated both the mean and variance of concentration, taking as input the ensemble mean and variance over MM5's twelve outcomes.

SCIPUFF's dose patterns varied quite noticeably over the ensemble, especially in the spread of the plume. The plume's centerline generally went toward southeast in all simulations but differed considerably in dispersion: some narrow, some wide. The rate of growth of the area receiving a dose above a threshold varied strongly over the members of the ensemble. The narrow-plume dispersions arrived at steady state within ten hours while others, with more meandering and wider plumes, didn't arrive at steady state until after 25 hr. The area above threshold dose varied over the ensemble by more than a factor of four. As with the HYSPLIT ensemble, PE diagrams were fairly compact for higher threshold doses. They spread considerably more with lower thresholds. In emergency response only a few ensemble members at most can be run. It will result in significantly different emergency response to choose one over another, though none appeared to be the best. Ability such as SCIPUFF's to estimate concentration variance might be useful to assess this uncertainty, if provided an estimate of the ensemble's variance. Perhaps such an estimate can be obtained through ensemble studies *a priori*.

5.2 Spread among buildings

A release in an urban area will be spread by very complex patterns. The release may be indoors or underground and may spread through heating, ventilating, and air-cooling systems (HVAC). Some may leak outside. Outdoor releases may follow street canyons or get trapped in narrow spaces. Dispersion among buildings follows very different turbulence patterns from those for which the empirical relations between puff growth and travel time were developed. The importance of individual structures hampers treatment by aggregate statistics. Computational fluid dynamics (CFD) can represent complex flow situations in any detail, subject to the computing resources available. Though the response time of CFD is currently too long to be useful in an emergency, it can guide risk assessment and emergency preparedness. Particular locations at high risk for accident or hostile act can be examined explicitly. More generally CFD can provide high-resolution metadata to calibrate other models and to help develop intuition and rules of thumb for incidence response teams.

Cybyk et al. (1999a,b, 2000) have developed the Three-Dimensional Flux-corrected Transport Algorithm for Computational Transport (FAST3D-CT), from work begun by Boris and Book (1973). A major issue in such modeling is definition of the boundary, which is enormously complex, including explicit representation of structures, trees, detailed terrain features, and the like. The FAST3D-CT uses finite differences with an embedding scheme that can provide very fine mesh where great detail is required while allowing coarser mesh elsewhere. The mesh generator is very flexible, allowing also temporal changes, such as opening a door. Clearly, however, high-performance computer requirements and long lead times preclude routine emergency use.

Two modes are described: simulation of actual releases to evaluate FAST3D-CT, and simulation of hypothetical cases of potential significance. Tracers were released in two experiments at Dugway Proving Ground in Utah. Both were indoors, one in a test building representing a typical apartment complex, the other in an airplane hangar simulating an exhibition hall. The

apartment complex had multiple rooms, connected by doors that were sometimes open, sometimes not. The hangar had a tool crib, HVAC, recesses to park the hangar's doors, and a double row of simulated exhibition booths. Details of the simulated concentration evolution matched those found in the measured evolution. The concentration itself was generally simulated within 50% of the measured values. The interior concentration was found to depend significantly on the wind velocity external to the test buildings.

The hypothetical cases involved simulated releases near the Pentagon and on the Mall in Washington DC. With just the Pentagon immediately downstream of the release, the cloud's pattern was simple, except at the Pentagon itself, where some of the simulated contaminant became trapped in the central courtyard and in the spaces between the concentric rings of offices. A later simulation included trees upstream of the release point and an additional large structure (power plant) beside the Pentagon. The cloud broke into multiple pieces, covered considerably less distance in the simulation time and generally developed a much more complex pattern. Taking at face value the results of such a computation in any real situation is dangerous because of the sensitivity to slight variations in the history of the wind during propagation of the cloud. These qualitatively reasonable simulations clearly show, however, the complexity of dispersion in an urban setting.

The complex geometry of urban dispersion modeling lends itself well to finite-element modeling. Again, great computer resources are required to simulate operational cases. The work, as above, is limited to risk assessments, to testing of simpler models, and to training responders' intuition. Finite-element simulations differ from the finite-difference FAST3D-CT primarily in the discretization scheme used. Otherwise the same equations of fluid dynamics are solved, though turbulence and other parameterized features may be treated differently. Camelli and Löhner (2000) simulated flow around a single L-shaped building achieving qualitatively correct patterns of flow and concentration. The quantitative match to a wind-tunnel simulation showed important deviations, indicating further work. Results were found very sensitive to the specification of the upstream wind profile. Proper specification of initial and boundary conditions is vital to all simulations.

The study of dispersion among urban buildings has revealed at least two important features amenable to rapid-response modeling. The speed of transport of the pollutant cloud is reduced, and the spread of the cloud is increased in all directions, including upwind. Reducing the transport speed delays the arrival of the cloud at downwind receptors. Furthermore, in shifting winds a slower-moving cloud will follow a different trajectory than a faster-moving cloud. A third important feature, unfortunately resistant to accurate simulation, is the trapping of contaminant in narrow spaces. These sites receive a high dose relative to the main cloud and provide a secondary contamination as they are flushed.

Efforts to model the influence of city buildings on dispersion of a contaminant cloud in a way sufficiently simple for emergency response have produced some success. The Urban Dispersion Model of the UK's Defence Science and Technology Laboratory (DSTL, Hall *et al.*, 1997, Griffiths *et al.*, 2000) is a puff model developed from extensive measurements in wind tunnels and in the atmosphere. Individual large objects, generally buildings, have explicit deflection and wake regions that interact with a puff according to its size relative to the obstruction. Hundreds

or thousands of these can be defined in the model field. Smaller objects (*e. g.*, houses) can be treated in bulk considering aggregate characteristics such as street-grid regularity and orientation. Tests with data from Salt Lake City during URBAN 2000 (Allwine *et al.*, 2002) produced encouraging results (Griffiths *et al.,* 2002). An extension of UDM to concentration fluctuations is in progress (Beck *et al.*, 2002). Full implementation of such a model requires a highly detailed survey of the city being represented. Such surveys are becoming increasingly available through airborne laser techniques. Few cities have yet been covered, however.

For cities not yet surveyed in such detail, coarser methods are required. Urban boundary layers are characterized by a thick "canopy." This requires specification of a zero-plane displacement z_d, the effective level of air-surface transfer. Above z_d a simplified urban canopy model defines a roughness sublayer (de Haan *et al.*, 2001; Grimmond and Oke, 2002) up a height z_*, perhaps three times the "typical" building and tree height in the local area. Within this layer, the model's downward momentum flux strengthens as height increases up to z_*. This flux divergence produces a smaller gradient of mean wind than that found in a constant-flux layer. At z_* the usual constant-flux layer begins. Measurements of wind profiles and turbulence should thus be made above z_* to parameterize this model. The surface energy budget favors sensible over latent heating of the air due to low effective soil moisture and rapid rain runoff. Heat exchange with the solid surface exhibits hysteresis due to the thermal characteristics of concrete surfaces, vertical walls, and other features that produce the urban heat island. Grimmond and Oke (2002) provide a recent approach to defining an urban surface energy budget.

Grimmond and Oke (1999) reviewed two means of determining z_d and the surface roughness length z_0. These were called *morphometric* and *anemometric*. The first method derives z_0 and z_d from representative measures of the size and density of buildings and trees in the area. The second method derives the parameters from wind measurements, whether of mean profiles or of turbulence. Unfortunately, the results were discouragingly scattered. The great variety of roughness elements in cities left no clearly superior means of determining aerodynamic properties of any given surface. The authors nevertheless noted some necessary features of good practice. They favored simplified morphometric methods as easier to apply and as yielding results of comparable validity to anemometric methods. The class of cityscape can normally be determined from oblique aerial photographs of representative sections of the city to be modeled. Guidance to this end is provided. A zero-plane displacement is always present over a city, and thus is required in models. Values of roughness length should lie within limits appropriate to the class of cityscape. The authors provided a table of such limits for four classes of cityscape, ranging from detached suburban houses to high-rise urban cores. Additional semiquantitative guidelines were given, accounting for such features as the density of trees and the orientation of major streets relative to the wind direction. The resulting values of z_d and z_0, while not reliable to high precision, can be incorporated into meteorological preprocessors or mesoscale models supporting the dispersion calculations.

Masson *et al.* (2002) tested a scheme called Town Energy Balance (TEB), based on the designs discussed in the previous paragraphs. The TEB represents the partitioning of insolation among the various radiative, turbulent, and surface fluxes at the earth's surface. The investigators placed instruments in commercial areas of Mexico City, Mexico, and Vancouver, British Columbia, Canada. These areas had no significant vegetation cover. Atmospheric turbulent

fluxes were measured at 28 m above ground. Infrared thermometers sensed surface temperatures of representative roads, and of walls and roofs of buildings. A soil flux plate was installed in Vancouver.

The TEB successfully represented several important qualitative features of the surface energy balance including the strong role of heat storage in buildings and streets (urban heat island), the maintenance of upward sensible heat flux at night, the dominance of sensible over latent heat, and the significance of the ratio of building height to street width. Mexico City, with taller stone buildings and narrower streets, retained upward sensible heat flux longer through the night than Vancouver, as expected and measured. Giving streets a roughness length of 0.5 m was required to generate sufficient atmospheric fluxes, especially for the wide streets in Vancouver. The high value represented traffic and other unaccounted roughness. The TEB is forced primarily by insolation, which was estimated well from measurements 8 km to 15 km away. As must be expected with a simple model, however, the simulated fluxes were fairly coarse in quantitative measure. Sensitivity tests showed the ratio of building height to street width to be important, along with the streets' albedo if this ratio is small.

The TEB missed some interesting features which future developments could treat. Sensible heat flux in Mexico City rose significantly just at dawn, when surface storage should theoretically be receiving the majority of the new insolation. Morning traffic was invoked to explain a similar observation in Tokyo, Japan. The tested version of the TEB had no explicit treatment of diurnal traffic patterns. Atmospheric heat fluxes were also poorly represented in the first of two study periods in Vancouver. A sea breeze probably produced advection, also not yet treated in the TEB.

Masson *et al.* (2002) recommended some good measurement practices for verification of urban land-surface models. Their measurement of fluxes, not just temperatures, greatly aided understanding of the energetics of the system. They also noted, however, some deficiencies of the field deployment that produced mismatch between model and measurements in this pioneering effort to verify an urban land-surface model. Principally, the scale of measurement should match the scale of the simulation (tens of kilometers) and should include effects of clouds and other departures from ideal conditions.

For the foreseeable future, emergency-response models on 10 km scales will require simplifications from the full simulation of dispersion among buildings. Two significant features of dispersion in urban boundary layers which are amenable to simplified modeling are the slowing of transport and increase in spread. Work addressing the adequacy of representation of these processes, a sample of which has been cited above, continues to progress.

5.3 User Interface

The high capacity of computing, storage, and display resources currently available even on laptop computers makes a detailed graphical user interface (GUI) attractive for emergency response, especially if the event is likely to cover several hours and tens of kilometers or more. In an activity dependent on rapid response, graphical images allow models and data acquisition

to be quickly implemented and their results readily grasped for rapid decision-making.

Commercially available data bases and geographical information systems (GIS) provide rapid, sophisticated access to a wide range of relevant information. Political boundaries quickly show jurisdictions. Road patterns indicate evacuation routing. Hospitals can be identified. Characteristics of buildings in the area can be known. Such data bases are kept more or less up to date in the course of their use for multiple ongoing commercial and governmental purposes. Boechler (2001) suggests an interesting use of such information: applying infiltration models, based on characteristics of local structures, to determine a phased evacuation. Some people would shelter in place to evacuate later, while others would evacuate immediately. In principle, this would reduce traffic and get more people to safety. Bacon *et al.* (2000) use a GUI to speed grid generation for their adaptive-grid model. Software "wizards" can facilitate quick configuration of a mesoscale model for people not expert in such modeling. Modern communication, coupled with GUI, allows near-real-time access to atmospheric data along with, for example, current contact information for medical personnel or other important people.

It must never be forgotten, however, that the utility of information derived from this wide array of resources is tied to the accuracy of the chain of forecasts on which the derivation relies. Predictions of health effect depend on accurate forecasts of the amount and duration of the contamination and its infiltration into spaces where the people are. Contamination forecasts depend on accurate forecasts of the atmospheric environment, especially wind and mixing, on appropriately fine scales. Forecasts of the local atmospheric environment depend on forecasts of larger-scale patterns, such as the tracks of storms, positions of fronts, and large-scale moisture transport. The potential for surprises is significant, especially for emergencies relevant to this discussion: those affecting more than several square kilometers.

Weather forecasters routinely reduce the likelihood of surprises in developing their forecasts by using output from multiple models. In the field of atmospheric transport and dispersion, multiple agencies, public and private, are capable of producing scientifically defensible forecasts of contaminants' doses and health effects. Having a variety of missions and perspectives, these agencies use multiple models and various sets of modeling parameters. If a protocol for cooperation can be established, a multiplicity of models can become available for use in a given emergency. This protocol would include standardization at least of report formats and communications protocols. It would also provide for the selection of a coordinator from an appropriate location and agency for any given emergency. GUI's based on this standard protocol could then facilitate rapid communication of all information to the appointed coordinator.

6. VERIFICATION

Verification is fundamental to any modeling effort, but must be so construed as to give account to the purpose of the model. For regulatory application the issue is the reliability of a model's forecast of compliance or noncompliance with regulations. Rapid execution is unnecessary, and the highest concentration's location in time or space is of less interest than its value. Emergency

23

applications in contrast require rapid execution, using data of opportunity. The location in time and space of the contaminant cloud from this particular release is the primary issue.

Models are verified by evaluation of the assumptions and simulations, by searching out coding errors, and by comparison with measurements. Comparison with measurements may be direct or indirect. Indirect comparison uses models more sophisticated than the one being tested. The sophisticated model transforms and interpolates the measurements, enhancing the utility of the comparison.

Not all mismatches between models and measurements arise from shortcomings of the models. Certainly, every model is simpler than the scenario it simulates. But measurement programs likewise capture only partially the scenarios they observe. Instruments cannot be placed everywhere, nor will the deployed instruments perfectly represent the phenomena on which they are to report. They may be inappropriately located and exposed, imperfectly calibrated, or insufficiently dense. They may not be measuring the optimal parameters to discern the physical processes being simulated by the model. Like the shortcomings of modeling, these shortcomings of experimental design arise from technological limitations. With care, knowledge, luck, and good judgment, they can be minimized, but not eliminated.

Furthermore, the model operates on ensemble statistics, predicting the mean, and perhaps higher moments, of a distribution of outcomes of identical experiments. Measurements represent only one of these outcomes, though a time or space average may, with appropriate design, estimate an ensemble average. The spread of individual outcomes about the ensemble mean may be large compared to the mean itself, introducing a wide range of irreducible mismatch between the measured and forecast values. Evaluation scores need to be designed to take this situation into account (Weil *et al.*, 1992). Furthermore, emergency-response models are explicitly concerned with estimating the outcome of yet another single realization. This requires explicit consideration of concentration fluctuations in some form in the evaluation and in the operational use of the model.

In the traditional evaluation tests a tracer is released over an array of 100 or more fixed monitors, deployed according to the situation being tested. Often the pollutant cloud will hit only a minority of the sensors, the majority reporting zero. The model domain will have the same result—a perfect forecast, but of no interest. Eliminating all pairs of measured and modeled values, where both are zero, can strengthen the evaluation. The obvious pairing of observed and modeled concentrations matches them by location and time. Such a pairing, however, lumps all sources of variance together, including the irreducible variance noted above. Other comparison schemes have been found more useful.

Moderate errors in transport often have large consequence because of the strong variation of concentration across the plume. Pairing modeled and measured data by rank can mitigate this problem for regulatory applications if the location of high concentration is not important. Emergency response, for which location matters, requires more detailed knowledge of both transport and dispersion of contaminants. These can be separated in the measurement protocol by tracking the contaminant cloud with mobile sensors whose absolute position on earth is always known. The Global Positioning System (GPS) has greatly simplified this activity over

the past decade. Such a protocol allows specification of the measurements' locations relative to the center of the cloud as well as to earth. Pairing measured and modeled results relative to the centers of the measured and simulated clouds tests dispersion alone. Comparing these clouds' centers relative to earth tests the transport.

6.1 Evaluation Criteria

The evaluation criteria for these comparisons are intended to establish the capabilities of models generally and to distinguish between better and poorer models. Many of the criteria presented below were taken from Chang *et al.* (2000). The simplest quantitative measures are purely statistical. The quantities being compared may be the concentrations (C_o, C_p) themselves, or they may be more general statistics (S_o, S_p). The subscripts refer to "observed" and "predicted" values, paired in any of the ways discussed above. The (S_o, S_p) may be a spatial or temporal mean or something more exotic, such as the robust highest concentration (RHC) of Cox and Tikvart (1990).

The simple difference
$$d = S_o - S_p$$
applies best when S has a Gaussian or other symmetrical distribution. Concentrations themselves, being nonnegative, have a skewed distribution, often approximately lognormal. Their ratios are generally more meaningful than their differences. A given value of d under low concentrations may be a large fraction of those concentrations. The same value of d under concentrations one hundred times as large may be insignificant relative to the concentrations. Factors of 100 and more between largest and smallest concentrations are common in evaluation experiments. If a simple average of differences is used, the results will appear to be worse than they are. Pairs (C_o, C_p) with the highest concentrations will dominate the average even if their discrepancy is not particularly large relative to their value. A better use of d is with composite measures S_o and S_p, which themselves subsume all aggregation.

A simple modification, if S is positive and not itself an aggregate, is to normalize by the mean between S_o and S_p to form the Fractional Bias
$$B_F = \frac{2(S_o - S_p)}{S_o + S_p}, \quad S_o, S_p \geq 0$$
An average of individual values of B_F is uniformly influenced by all entries regardless of the magnitude of S. A perfect match gives $B_F = 0$. The fractional bias is bounded by ± 2, which occurs when either of S_p or S_o is zero.

Recognizing the approximately lognormal distribution of the concentrations, the ratio of their geometric means
$$M_G = \exp\left(\overline{\ln C_o} - \overline{\ln C_p}\right)$$
is an attractive aggregate statistic. It measures discrepancies as ratios, rather than arithmetic differences. A perfect result is $M_G = 1$. As measures of the systematic bias of a model, B_F and M_G function similarly. The fractional bias, however, can tolerate a zero value in either S_o or S_p (not both). With M_G all concentrations must be strictly positive. This can be troublesome near

the edge of a cloud. On the other hand, M_G, being unbounded in either direction, is very sensitive to pairs (C_o, C_p), where both are small, but one is much smaller than the other. A plausible approach is to select a threshold concentration such that all quantities below the threshold are assigned the threshold value. The algorithm must, of course, eliminate any pairs of zeros.

The normalized mean square error

$$\varepsilon_{MS} \;=\; \frac{\overline{(S_o - S_p)^2}}{\overline{S_o}\,\overline{S_p}}$$

measures the strength of the spread of predicted values about observed, relative to the square of the geometric mean of the averages. Since the numerator is the mean of d^2, the same comment applies here for lognormally distributed S. The square, in fact, enhances the significance of larger discrepancies. The normalization by the product of the means compensates somewhat for mean discrepancies in the two data sets.

The geometric variance of observed about predicted

$$V_G \;=\; \exp\!\left[\overline{(\ln S_o - \ln S_p)^2}\right]$$

was used by Chang $et\ al.$ (2000) to measure the random scatter of model results. Although this expression is simple, its interpretation is not straightforward. More direct is the geometric variance of the ratio of observed to predicted about its geometric mean:

$$V_{GC} \;=\; \exp\!\left[\overline{\left((\ln S_o - \ln S_p) - \ln M_G\right)^2}\right]$$

This removes the geometric means from the expression, leaving only the random scatter. The scatter may be more readily interpreted from the "geometric standard deviation:"

$$\sigma_G \;=\; \exp\!\left[\ln^{\frac{1}{2}}(V_{GC})\right].$$

If S is lognormally distributed, approximately 68% of the predictions are within a factor σ_G of M_G.

The linear correlation coefficient between predicted and observed values

$$R \;=\; \frac{\overline{\left(S_o - \overline{S_o}\right)\!\left(S_p - \overline{S_p}\right)}}{\sigma_{S_p}\sigma_{S_o}}$$

gives a measure of the interrelation between the two sets of statistics. Again, given the approximate lognormal distribution of concentration, it may be useful to consider computing R on the logarithms of the quantities.

A commonly used evaluation that takes account of the nature of the distribution of the concentration is the fraction within a factor of N

$$facN \;=\; \frac{size\!\left\{A \;\ni\; \frac{1}{N} < \frac{S_p}{S_o} < N\right\}}{size\,\{\Omega\}}.$$

Here Ω is the set of all outcomes of the test. Its subset A is described in the expression, and $size$ $\{A\}$ is the number of members in the set A. Often $N = 2$, though other integers have been used
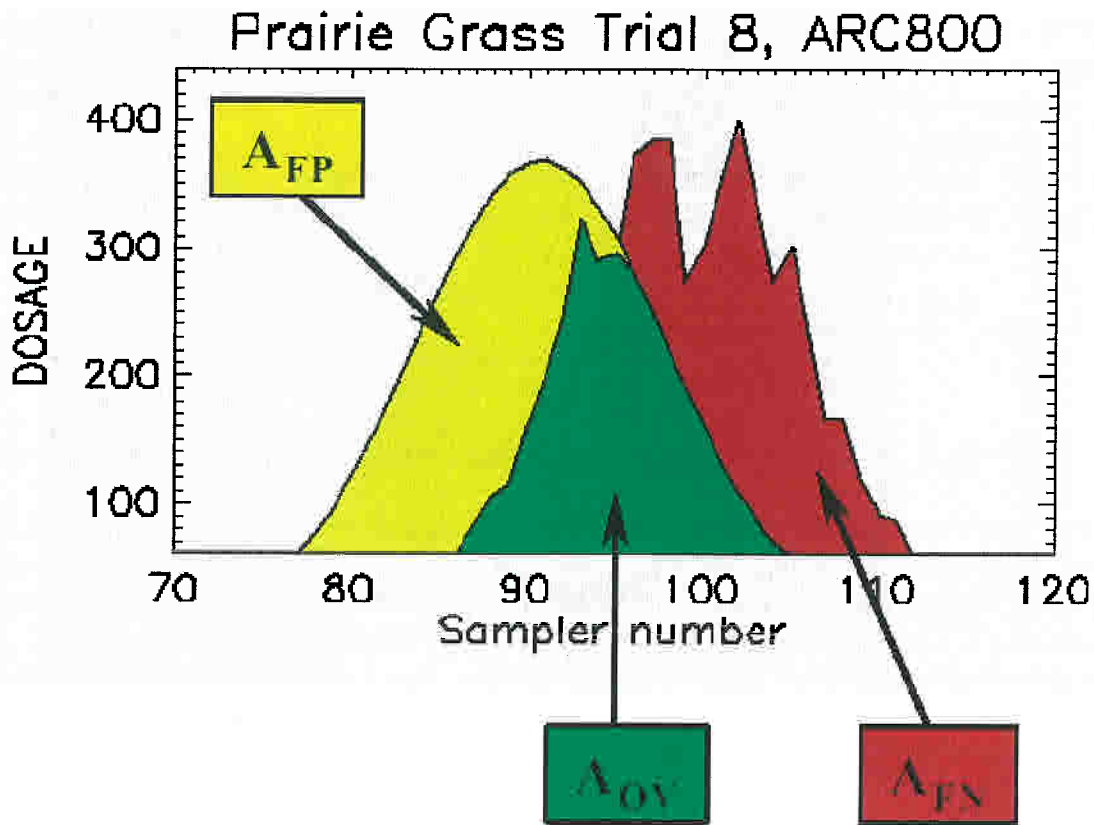
when necessary.

Figure 2: Example of regions used to compute the MOE (from Warner et al., 2001)

6.2 Measure of Effectiveness

Warner *et al.* (2001) have developed a Measure of Effectiveness (MOE) which evaluates a forecast in terms of the regions of false positive A_{FP}, false negative A_{FN}, and overlap A_{OV}. In A_{FP}, contamination was forecast but not experienced. In A_{FN}, contamination was experienced but not forecast. In A_{OV}, contamination was both experienced and forecast, though not necessarily in the same amount. The MOE has two dimensions $\left(1 - \dfrac{A_{FN}}{A_{HIT}}, \quad 1 - \dfrac{A_{FP}}{A_{PR}} \right)$, either of which can range in [0,1], including endpoints. Here A_{HIT} is the region which experienced contamination, while A_{PR} is the region predicted to experience contamination. Thus the two dimensions of the MOE are unity minus the *false negative fraction*, and unity minus the *false positive fraction*. The MOE provides a framework which can incorporate graphically both the model's practical effectiveness and the user's level of risk tolerance.
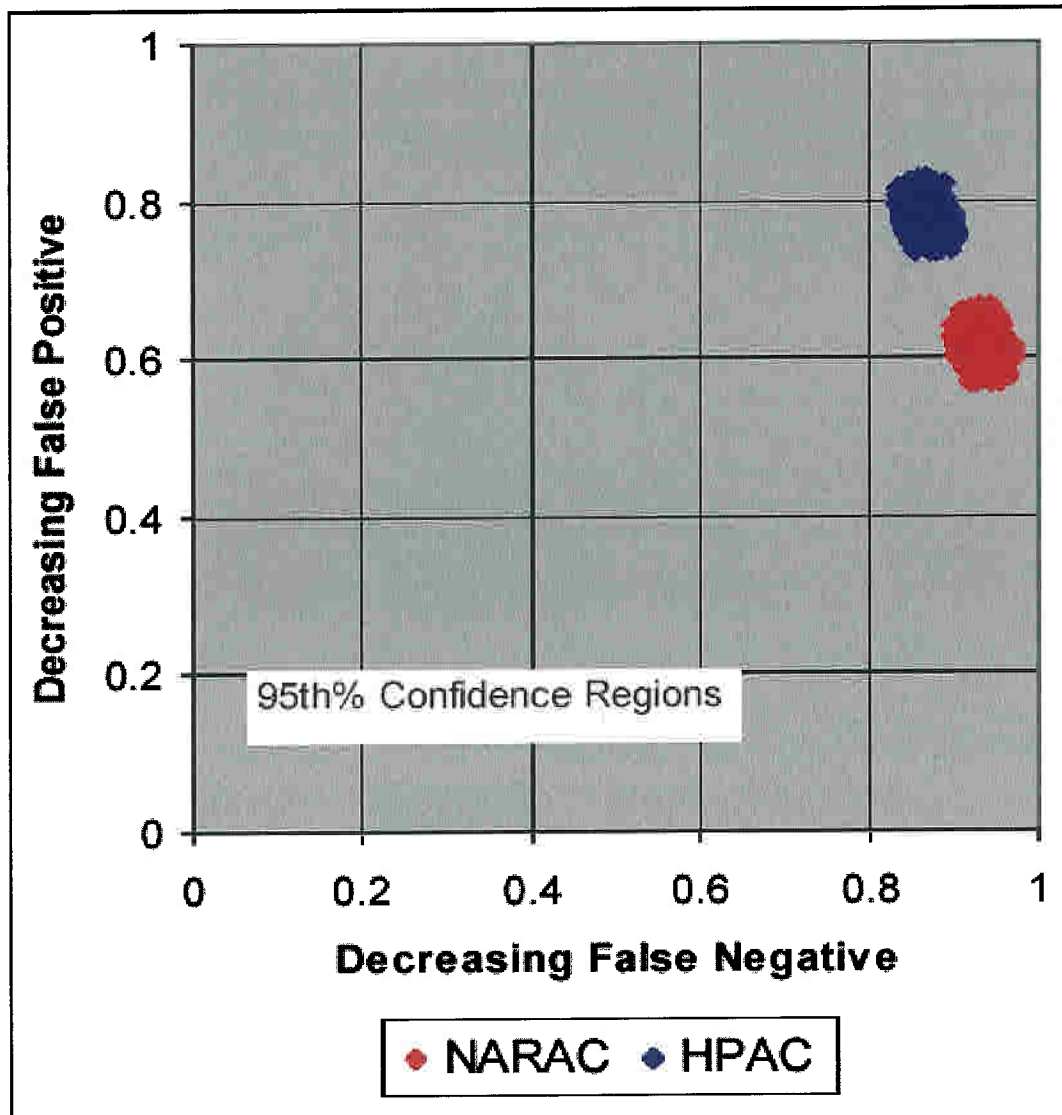
Figure 3: MOE from two models finding the extent of the affected area depicted in Fig 2, where the "effect" is taken to occur when the dose exceeds 60 units. (from Warner et al., 2001)

The index is determined by comparison between data from a dispersion experiment and forecasts from the model being tested. The "areas" may be defined in multiple ways, giving flexibility to the measure. Typical dispersion experiments report the dose of tracer received over a specified sampling time along rings of samplers at several distances from the source. If the extent of an affected area is more interesting than the dose received within, the A_{FN} can be defined as the cumulative length of segments of the sample ring on which dose above some threshold was experienced, but not forecast. The A_{FP} and A_{OV} are defined analogously. When the dose within the affected area is important, integration along a sample ring is used. The A_{FP} is the integral of that forecast dose which exceeds the dose observed. The A_{FN} is the integral of that observed dose which exceeds the dose forecast. The A_{OV} is the integral of that portion of dose both observed and forecast. Fig. 2 shows an example.

The MOE is presented graphically as in Fig. 3. A perfect forecast has MOE = (1,1). When MOE is (0,0), the forecast region is disjoint from that actually affected. Warner *et al.* (2001) used the MOE to compare models of NARAC and HPAC using data from the Prairie Grass Experiment (Barad, 1958). They estimated the 95% confidence interval using bootstrap resampling techniques, to be discussed below. Both models are comparable in overall accuracy, the difference showing the tradeoff between false positive and false negative bias. The NARAC results indicate greater likelihood of false positive than negative. That is, false alarms are more likely than failure to warn. The HPAC is also biased in this direction, but less so.

Different applications may imply different biases in the risk-tolerance tradeoff. Users can represent their bias through a user's figure of merit (FOM)

$$0 \leq FOM = \frac{A_{OV}}{(A_{OV} + C_{FN}A_{FN} + C_{FP}A_{FP})} \leq 1.$$

The coefficients C_{FN} and C_{FP} take nonnegative real values and indicate the user's tolerance of false positive and false negative outcomes. A large coefficient indicates low tolerance. If both coefficients are zero, any outcome is acceptable. A defensive role indicates large C_{FN} and allows smaller C_{FP}, preferring false alarms over failure to warn of a hazard. The FOM, if plotted on Fig. 3, would vary over the region from zero at (0,0) to unity at (1,1). It can be represented as a color spectrum varying from red to green with yellow assigned to 0.5. Models with MOE in the green area are acceptable according to the user's criteria, expressed through the coefficient values.

6.3 Bootstrap Resampling

Given an outcome, consisting of pairs (S_o, S_p), the performance of a model can be assessed using measures such as those just given. This single number however would be different if the test were run again, even under the same conditions. The distribution of such possible outcomes can be estimated from the data on hand using the "bootstrap" resampling technique (Efron and Tibshirani, 1998). The existing collection of pairs (S_o, S_p) is taken to be representative of all possible outcomes of the test. The bootstrap will provide estimates of the underlying population, accurate to the extent that the sample in hand properly represents it. The larger the collection the better is the approximation. The procedure is to form from this set a large collection of alternative outcomes, at least 100, by randomly drawing about as many pairs as in the original collection. The crux of the technique is to draw each successive pair from the full set. That is, the drawing includes replacement. From each alternative outcome, the performance measure of interest is computed. The collection of performance measures estimates a distribution that could have occurred by chance from the test that was run.

Any desired statistic can be determined from this distribution. Central statistics are, however, more accurately estimated than are higher moments and more extreme quantiles (2%-ile or 98%-ile, for example). Such statistics depend on low-probability outcomes, intrinsically unlikely to be present in the underlying outcome set. High-percentile values, however, are interesting as limits to the confidence interval for the performance measure. The accuracy of their estimation can be improved by providing a larger initial set of pairs or by fitting a tail model to the

distribution. Knowledge of the physics of dispersion and of the situation simulated is required to design the resampling scheme and to interpret the results. For example, outcomes from multiple tests may be combined to increase the statistical power. However, tests run under different stability conditions would be inappropriate for combining into a single set.

Another resampling scheme can be applied to assess the likelihood of significant difference between two models by some test statistic, such as one of the performance measures. Test statistics, such as mean or maximum concentration, can also be compared between an individual model and a set of measurements. Each set of outcomes, measurements, model A, model B, *etc.* forms a column in a conceptual table. Each column is ordered by the same protocol. For example, the ordering may be by time and place, or by quantile in their respective sets. Any ordering may be used so long as it is the same for each column.

By the null hypothesis, each column represents a sample from the same population. That is, the underlying "true" value of the test statistic is the same for each column. If so, one can create additional outcomes by drawing from each row a value selected randomly from among the columns. If many new columns are created this way and the test statistic computed for each, a distribution of values of the test statistic can be developed. This distribution gives the likelihood that any particular value of the test statistic occurred by chance, given that all outcome sets came from the same population. If the probability is "small" that the test statistic from one or more of the original sets occurred by chance, the hypothesis is rejected. Following standard statistics practice, some prespecified threshold defines "small."

Optimally for resampling, the departure between simulated and observed values is symmetrically distributed about zero. Pollutant concentrations, being nonnegative, often have approximately a lognormal distribution. Thus, the logarithms of the concentrations are more symmetrically distributed than the concentrations themselves.

6.4 Statistically evaluating model physics

The best model is the one with smallest bias and greatest confidence that error is small. That is, it has a small discrepancy between predictions and measurements, and that discrepancy has a narrow confidence interval. Such pure statistical techniques show one model to perform better than another for the conditions of the sample. The likelihood of good performance under other conditions is greatly enhanced if the model's superior performance reflects a superior simulation of the underlying physics: "did the model get the right answer for the right reason?"

Establishing multiple ensembles of observations can test the ability of the model(s) to represent changes in the physical environment. Membership in an ensemble is defined by values of the model's input parameters ξ_n, which include source configuration, static stability, wind speed, and the like. Subranges of the ξ_n define bins to which measurement runs are classified. All runs in a given bin constitute an ensemble. The width of a bin is determined by the significance of variations in each ξ_i, $i=1, \ldots, n$, which can be tested by resampling.

Sufficient observations to populate each bin with a meaningful ensemble are usually unavailable. A promising alternative for some situations (Weil et $al.$, 1992) is to assume that the dimensionless fluctuation $c'_r \sigma_c^{-1}$ has a universal probability distribution independent of the values of ξ_n. Here σ_c is the standard deviation of the concentration fluctuations, and c'_r is the departure of measured concentration from the ensemble mean due to unaccountable sources of variance. It is the irreducible component of measurement uncertainty. Unfortunately, σ_c is generally unavailable as well, leading to a further compromise. The ratio of predicted to observed concentrations $c_p c_o^{-1}$, paired in any appropriate way is distributed approximately lognormally. Its logarithm, approximately Gaussian, can be plotted as a function of the parameter ξ_i of interest. Given the usual scatter in such a plot, it is preferable to define bins from subranges of ξ_i. If the confidence interval for $\ln(c_p)$-$\ln(c_o)$ excludes zero, the model is judged insufficiently responsive to changes in ξ_i over the affected range.

7. SUMMARY

The impressive increase in computer resources over the past 20 years has enabled major advances in forecasting atmospheric contaminants' dispersion. Highly detailed modeling, based on sophisticated theory, is now available ($e.g.$, Lee et $al.$, 1998; Leone et $al.$, 1997; Sykes et $al.$, 1998; Bacon et $al.$, 2000). Emergency releases are likely to be in urban settings, prompting development of detailed modeling schemes (Cybyk et $al.$, 1999a; 1999b; Griffiths, et $al.$, 2000), along with simpler schemes for more rapid response (Grimmond and Oke, 2002). The more detailed models apply well to the pre- and post-event activities of planning and cleanup. Modeling for response to an event in progress must be simpler to be able to run in the shortest possible time. Furthermore, responders are dealing with one particular outcome of a stochastic process having intrinsically large variance. Continued increases in computer power and communication promise to make the results of increasingly sophisticated modeling available to such responders.

Use of theoretical sophistication in modeling requires accompanying demonstration of improved forecast accuracy and versatility. Increasingly sophisticated field tests and data sets are coming available ($e.g.$, Chang et $al.$, 2003; Masson et $al.$, 2002; Allwine et $al.$, 2002). Forecasts of uncertainty, now possible under current simulation practice ($e.g.$, Sykes et $al.$, 1998), are receiving tests of validity and cost-effectiveness (Warner et al, 2002). In any case a large irreducible uncertainty is characteristic of dispersion forecasting (Hanna and Yang, 2001). The primary source of uncertainty is in representing the atmospheric environment, especially the wind, which determines the contaminants' trajectory. Characterizing this uncertainty after the manner of Warner et $al.$ (2002), Chang et $al.$ (2003), and Draxler (2003) is important and worthy of further work.

Resources, both in computing and in databases, are now sufficient to provide a friendly interface that rapidly displays information in a form readily grasped by emergency decision makers, saving much-needed time. With appropriate cooperation and standardized protocols, output from several agencies' models could be made available to an information coordinator for a given emergency. This coordinator, preferably having knowledge of dispersion in the local area would

choose an appropriate strategy for a given release in the face of the remaining large uncertainties. Having the results from multiple models in standard display format provides some idea of the range of possible outcomes.

Training of those responding to emergencies involving atmospheric releases can benefit from results of model verification and uncertainty assessment applied to decision making in the face of the uncertainty. Warner *et al.* (2002), for example, show that uncertainty is significantly reduced if a higher dose of contaminant can be tolerated. Measures of effectiveness, figures of merit, and other means of expressing risk tolerance (*e.g.*, Warner *et al.*, 2001, discussed in the Verification section above) appear promising for this effort.

8. REFERENCES

Alpha TRAC, 2002: Computer Assisted Protective Action Recommendation System (CAPARS) on web http://www.rarc.org/Overview/CAPARSSummary.HTML, accessed 2002 March.

Allwine, K. J., J.H. Shinn, G. E. Streit, K. L. Clawson, and M. Brown 2002: Overview of URBAN 2000: A multi-scale field study of dispersion through an urban environment. *Bulletin of the American Meteorological Society* **83**, 521 – 536.

Bacon, D. P., N. N. Ahmad, Z. Boybeyi, T. J. Dunn, M. S. Hall, P. C. S. Lee, R. A. Sarma, M. D. Turner, K. T. Waight III, S. H. Young and J. W. Zack 2000: A dynamically adapting weather and dispersion model: The Operational Multiscale Environmental Model with Grid Adaptivity (OMEGA). *Monthly Weather Review* **128,** 2044 – 2076.

Barad, M. L., 1958: Project Prairie Grass: a field program in diffusion. Report **AFCRL-TR-235**.

Boris, J. P. and D. L Book, 1973: Flux-Corrected Transport 1. SHASTA, a fluid transport algorithm that works. *Journal of Computational Physics* **11**, 38 – 69.

Bowen, B. M., J. A. Baars, and G. L. Stone, 2000: Nocturnal wind direction shear and its potential impact on pollutant transport. *Journal of Applied Meteorology* **39**, 437 - 445.

Bower, T. J., and R. L. Gibbs, 1998: Software User's Manual for the Chemical/Biological Agent Vapor, Liquid, and Solid Tracking (VLSTRACK) Computer Model, Version 3.0. Systems Research and Technology Department, Dahlgren Division, Naval Surface Warfare Center, Dahlgren VA.

Beriwal, M., and P. B. Merkle, 2001: Defense Threat Reduction Agency CV modeling and simulation futures workshop. Web file http://www.dtra.mil/about/organization/cbmodeling.pdf, accessed 2002 March 21.

Biltoft, C. A., 1998: *Dipole Pride 26: Phase II of Defense Special Weapons Agency Transport and Dispersion Model Validation.* Prepared for Defense Special Weapons Agency, 6801 Telegraph Road, Alexandria, VA 22310, by Meteorology and Obscurants Divisions, West Desert Test Center, U.S. Army Dugway Proving Ground, Dugway UT.

Camelli, F. E. and R. Löhner, 2000: Flow and dispersion around buildings: an application with FEFLO. *European Congress on Computational Methods in Applied Sciences and Engineering ECOMAS 2000*, 2000 September 11-14, Barcelona, Spain.

Chang, J. C., P. Franzese, and S. R. Hanna, 2000: Evaluation of CALPUFF, HPAC, and VLSTRACK with the Dipole Pride 26 field data. *Proceedings 11th Joint Conference on*

Applications of Air Pollution Meteorology, American Meteorological Society, Air and Waste Management Association, Long Beach CA 2000 January 10 – 14.

Chang, J. C., P. Franzese, K. Chayantrakom, and S. R. Hanna, 2003: Evaluations of CALPUFF, HPAC, and VLSTRACK with Two Mesoscale Field Datasets. *Journal of Applied Meteorology* **42**, 453 – 466.

Chilès, J.-P., and P. Delfiner, 1999: *Geostatistics: Modeling Spatial Uncertainty*. Wiley-Interscience, ISBN 0471083151, 672 pg.

Ciolek, J. T., Jr. and C. S. Magtutu, 1998: Evaluation of the Computer-Assisted Protective Action Recommendation System (CAPARS). From web: http://www.rarc.org/papers/Reacceptance/evaluati.htm accessed 2002 March.

Clemments, W. E., J. A. Archuleta, and P. H. Gudiksen, 1989: Experimental design of the 1984 ASCOT Field Study. *Journal of Applied Meteorology* **28**, 405 – 413.

Csanady, G. T., 1972: *Turbulent Diffusion in the Environment*, D. Reidel, Dordrecht, Netherlands, Boston, MA, 248 pg.

Cybyk, B. Z., J. P. Boris, T. R. Young, Jr., A. M. Landsberg, and C. A. Lind, 1999a: A complex-geometry CFD model for chemical/biological contaminant transport problems. 1999 DoD HPC Users Group Conference, Monterey, CA, 1999 June 7 - 10.

Cybyk, B. Z., J. P. Boris, T. R. Young, C. A. Lind, and A. M. Landsberg, 1999b: A detailed contaminant transport model for facility hazard assessment in urban areas. Paper AIAA 99-3441, *30th Plasmadynamics and Lasers Conference, American Institute of Aeronautics and Astronautics*, 1999 June 28-July 1, Norfolk VA.

Cybyk, B. Z., J. P, Boris, and T. R. Young, Jr. 2000: Coupling of external winds and recirculations with interior contaminant release modeling. Preprints Third Symposium on the Urban Environment, 2000 August 14-18, Davis CA, pg. 58 - 59.

Deardorff, J. W. and G. E. Willis, 1984: Ground-level concentration fluctuations from a buoyant and nonbuoyant source within a laboratory convectively mixed layer. *Atmospheric Environment* **18**, 1297 – 1309.

de Haan, Peter, M. W. Rotach, and M. Werfeli 2001: Modification of an operational dispersion model for urban applications. *Journal of Applied Meteorology* **40** 864 – 879.

Doran, J. C., J. D. Fast, and J. Horel, 2002: The VTMX 2000 Campaign, *Bulletin of the American Meteorological Society* **83**, 537 – 551.

Doyle, J. D. and N. A. Bond, 2001: Research aircraft observations and numerical simulations of a warm front approaching Vancouver Island. *Monthly Weather Review* **129**, 978 – 998.

Draxler, R. R., R. Dietz, R. J. Lagomarsino, and G. Start, 1991: Across North America Tracer Experiment (ANATEX): Sampling and analysis. *Atmospheric Environment* **25A**, 2815-2836.

Draxler, R. R. and G. D. Hess, 1998: Description of the HYSPLIT_4 modeling system. *NOAA Technical Memorandum* **ERL ARL-224**, 27 pg.

Draxler, R. R., 1999: HYSPLIT radiological transport and dispersion model implementation on the NCEP Cray. NOAA Technical Procedures Bulletin **TPB 458**, 13 pg.

Draxler, R.R., 2003: Evaluation of an ensemble dispersion calculation. *Journal of Applied Meteorology* **42**, 308 - 317

Efron, B., and R. J. Tibshirani, 1998: *An Introduction to the Bootstrap*, CRC Press, Boca Ratón FL, 435 pg.

Flesch, T. K., J. D. Wilson, and E. Yee, 1995: Backward-time stochastic dispersion models and their application to estimate gaseous emissions. *Journal of Applied Meteorology* **34**, pg

1320 – 1332.

Flossmann, A. I., and H. R. Pruppacher, 1985: A theoretical study of the wet removal of atmospheric pollutants. Part III: The uptake, redistribution, and deposition of $(NH_4)_2SO_4$ particles by a convective cloud using a two-dimensional cloud dynamics model. *Journal of the Atmospheric Sciences* **45**, 1857 – 1871.

Grell, G.A., J. Dudhia, and D. R. Stauffer, 1994: A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5). NCAR Technical Note **NCAR/TN-398+STR**, 121 pg.

Griffiths, I. H., N. V. Beck, C. John, D. J. Hall, and A. M. Spanton, 2000: The results of an initial validation study of an Urban Dispersion Model. American Meteorological Society Third Symposium on the Urban Environment, Davis CA, 2000 August 14 – 18.

Griffiths, I. H., D. R. Brook, D. J. Hall, A. Berry, R. D. Kingdon, K. L. Clawson, C. Blithoft, J. M. Hargrave, C. M. Clem, D. C. Strickland, and A. M. Spanton, 2002: Urban Dispersion Model (UDM) validation. *Abstracts volume American Meteorological Society Joint Conference* Norfolk VA, 2002 May 20–24, pg. J21-J22.

Grimmond, C. S. B. and T. R. Oke 1999: Aerodynamic properties of urban areas derived from analysis of surface form. *Journal of Applied Meteorology* **38**, 1262 – 1292.

Grimmond, C. S. B., and T. R. Oke, 2002: Turbulent heat fluxes in urban areas: Observations and a local-scale urban meteorological parameterization scheme (LUMPS). *Journal of Applied Meteorology* **41**, 792 – 810.

Hall, D. J., A. M. Spanton, R. W. Macdonald, and S. Walker, 1997: A simple model for estimating dispersion in urban areas. Report **CR 169/97** Prepared for DSTL, Building Research Establishment, Garston, Watford, UK.

Hanna, S. R. and R. Yang, 2001: Evaluations of mesoscale models; simulations of near-surface winds, temperature gradients, and mixing depths. *Journal of Applied Meteorology* **40**, 1095 – 1104.

Hodur, R. M., 1997: The Naval Research Laboratory's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS). *Monthly Weather Review* **125**, pg. 1414-1430.

Hurley, Peter, 1994: PARTPUFF—A Lagrangian particle – puff approach for plume dispersion modeling applications. *Journal of Applied Meteorology* **33**, pg 285 – 294.

Innovative Emergency Management, Inc., 2000: D2-Puff™, an atmospheric dispersion model, on web http://www.ieminc.com/Products_and_Services/Project_descriptions/puff.htm, accessed 2002 March, updated 2000.

Irwin, J. S., 1983: Estimating plume dispersion—A comparison of several sigma schemes. *Journal of Climate and Applied Meteorology* **22**, 92 – 114.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society* **77**, 437 – 471.

Lee, R. L., J. R. Albritton, S. Chan, J. M. Leone, Jr., J. S. Nasstrom, and G. Sugiyama, 1997: ARAC-3, a new generation emergency response modeling system. American Nuclear Society's 6th Topical Meeting, Emergency Preparedness and Response, San Francisco CA, 1997 April 22-25.

Legg, B. J., and M. R. Raupach, 1982: Markov-chain simulation of particle dispersion in inhomogeneous flows: The mean drift velocity induced by a gradient in Eulerian velocity variance.

Leone, J. M., Jr., J. S. Nasstrom, and D. M. Maddix, 1997: A first look at the new ARAC dispersion model. American Nuclear Society's 6th Topical Meeting on Emergency

Preparedness and Response, San Francisco CA, 1997 April 22-25.

Lyons, W. L., R. H. Calby, and C. S. Keen, 1986: The impact of mesoscale convective systems on regional visibility and oxidant distributions during persistent elevated pollution episodes. *Journal of Climate and Applied Meteorology,* **25**, pg 1518-1531.

Masson, V., C. S. B. Grimmond, and T. R. Oke 2002: Evaluation of the Town Energy Balance (TEB) Scheme with direct measurements form dry districts in two cities. *Journal of Applied Meteorology* 41, 1011 – 1026.

NCAR, 1999: MM5 Home Page, on web http://www.mmm.ucar.edu/mm5/mm5.html, accessed 2002 May 29.

ttl, D., R. A. Almbauer, and P. J. Sturm, 2001: A new method to estimate diffusion in stable, low-wind conditions. *Journal of Applied Meteorology* **40**, 259 – 268.

RAMS, 2002: The Regional Atmospheric Modeling System, on web http://rams.atmos.colostate.edu/rams-description.html accessed 2002 May 29.

Respondek, P. S., A. I. Flossmann, R. R. Alheit, and H. R. Pruppacher, 1995: A theoretical study of the wet removal of atmospheric pollutants. Part V: The uptake, redistribution, and deposition of $(NH_4)_2SO_4$ by a convective cloud containing ice. *Journal of the Atmospheric Sciences* **52**, 2121 – 2132.

Rogers, E. T. Black, B. Ferrier, Y Lin, D. Parrish, and G. DiMego, 2002: Changes to the NCEP Meso Eta Analysis and Forecast System: Increase in resolution, new cloud microphysics, modified precipitation assimilation, modified 3DVAR analysis. on web: http://www.emc.ncep.noaa.gov/mmb/mmbpll/eta12tpb/

Ross, D. G., I. N. Smith, P. C. Manins, and D. G. Fox, 1988: Diagnostic wind field modeling for complex terrain: Model development and testing. *Journal of Applied Meteorology* **27**, 785 – 796.

Scire, J. S., F. R. Robe, M. E. Fernau, and R. J. Yarmartino, 2000a: A User's Guide for the CALMET Meteorological Model (Version 5). Available from Earth Tech Incorporated, Long Beach CA: http://www.src.com/calpuff/calpuff1.htm accessed 2002 May 30

Scire, J. S., D. G. Strimiatis, and R. J. Yamartino, 2000b: A User's Guide for the CALPUFF Dispersion Model (Version 5). Available from Earth Tech Incorporated, Long Beach CA: http://www.src.com/calpuff/calpuff1.htm accessed 2002 May 30.

Sugiyama, G. and S. T. Chan, 1998: A new meteorological data assimilation model for real-time emergency response. Paper 7A.1, 10th Joint Conference Applications of Air Pollution Meteorology, American Meteorological Society with Air and Waste Management Association, Phoenix, AZ 11-16 January, 1998.

Sykes, R. I. and D. S. Henn, 1995: Representation of velocity gradient effects in a Gaussian puff model. *Journal of Applied Meteorology* **34**, 2715 – 2723.

Sykes, R. I. and R. S. Gabruk, 1997: A second-order closure model for the effect of averaging time on turbulent plume dispersion. *Journal of Applied Meteorology* **36**, 1038 – 1045.

Sykes, R.I., S. F. Parker, D. S. Henn, C. P. Cerasoli, and L. P. Santos, 1998: PC-SCIPUFF Version 1.2PD: Technical Documentation. **ARAP Report 718**, Titan Corporation, Princeton NJ, 172 pg.

Thompson, A. M., K. E. Pickering, R. R. Dickerson, W. G. Ellis, Jr., D. J. Jacob, J. R. Scala, W.-K. Tao, D. P. McNamara, and J. Simpson, 1994: Convective transport over the central United States and its role in regional CO and ozone budgets. *Journal of Geophysical Research* **99D**, 18703 – 18711.

Thompson, D. J., 1987: Criteria for the selection of stochastic models of particle trajectories in

turbulent flows. *Journal of Fluid Mechanics* **180**, pg 529 – 556.

Thompson, D. J., 1990: A stochastic model for the motion of particle pairs in isotropic high-Reynolds-number turbulence, and its application to the problem of concentration variance. *Journal of Fluid Mechanics* **210**, pg 113 – 153.

Thompson, D. J., W. L. Physick, and R. H. Maryon, 1997: Treatment of interfaces in random walk dispersion models. *Journal of Applied Meteorology* **36**, 1284 – 1295.

Uliasz, M., 1993: The atmospheric mesoscale dispersion modeling system. *Journal of Applied Meteorology* **32**, pg. 139 – 149.

Umeda, T., and P. T. Martien, 2002: Evaluation of a data assimilation technique for a mesoscale meteorological model used for air quality modeling. *Journal of Applied Meteorology* **41**, 12 – 29.

USEPA, 1998: CAMEO® Computer-aided management of emergency operations: factsheet. US Environmental Protection Agency, Chemical Emergency Preparedness and Prevention Office, **550-F-98-003**, 2 pg. http://www.epa.gov/ceppo/cameo/pubs/cameo-fs.pdf, access March 2002.

Warner, S., N. Platt, and J. F. Heagy, 2001: User-oriented measures of effectiveness for the evaluation of transport and dispersion models. *Proceedings of the Seventh International Conference on Harmonisation Within Atmospheric Dispersion Modelling for Regulatory Purposes*, 2001 May 28 – 31, Belgirate (Lake Maggiore) Italy, pg 24 – 29.

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. *Journal of Applied Meteorology* **41**, 488 – 504.

Watson, T. B., R. E. Keislar, B. Reese, D. H. George, and C. A. Biltoft, 1998: *The Defense Special Weapons Agency Dipole Pride 26 Field Experiment*. NOAA Technical Memorandum **ERL ARL-255**, Air Resources Laboratory, Silver Spring Maryland.

Weil, J. C., R. I. Sykes, A. Venkatram, 1992: Evaluating air-quality models: review and outlook. *Journal of Applied Meteorology* **31**, 1121 – 1145.

Willis, G. E. and J. W. Deardorff, 1976: A laboratory study of dispersion from an elevated source within a modeled convective planetary boundary layer. *Atmospheric Environment* **12**, 1305 – 1311.

Willis. G. E. and J. W. Deardorff, 1981: A laboratory study of dispersion from a source in the middle of the convective mixed layer. *Atmospheric Environment* **15**, pg 109 –117.

YSA, 1998: Yamada Science and Art, on web http://www.kmxq.com/ysa/index.html accessed 2002 May 29.

Zhu, Y. Z. Toth, R. Wobus, D. Richardson, and K. Mylne, 2002: The economic value of ensemble-based weather forecasts. *Bulletin of the American Meteorological Society* **83**, 73 – 83.

Zilitinkevich, S., A. Baklanov, J. Rost, A.-S. Smedman, V. Lykosov, and P. Calanca, 2002: Diagnostic and prognostic equations for the depth of the stably stratified Ekman boundary layer. *Quarterly Journal of the Royal Meteorological Society* **128**, 25 – 46.