Comparisons of Transport and Dispersion Model Predictions of the Mock Urban Setting Test Field Experiment

STEVE WARNER, NATHAN PLATT, AND JAMES F. HEAGY
Institute for Defense Analyses, Alexandria, Virginia

JASON E. JORDAN
Northrop Grumman Corporation, Alexandria, Virginia

GEORGE BIEBERBACH
National Center for Atmospheric Research,* Boulder, Colorado

(Manuscript received 8 October 2005, in final form 10 February 2006)

ABSTRACT

The potential effects of a terrorist attack involving the atmospheric release of chemical, biological, radiological, nuclear, or other hazardous materials continue to be of concern to the United States. The Defense Threat Reduction Agency has developed a Hazard Prediction Assessment Capability (HPAC) that includes initial features to address hazardous releases within an urban environment. Improved characterization and understanding of urban transport and dispersion are required to allow for more robust modeling. In 2001, a scaled urban setting was created in the desert of Utah using shipping containers, and tracer gases were released. This atmospheric tracer and meteorological study is known as the Mock Urban Setting Test (MUST). This paper describes the creation of sets of HPAC predictions and comparisons with the MUST field experiment. Strong consistency between the conclusions of this study and a previously reported HPAC evaluation that relied on urban tracer observations within the downtown area of Salt Lake City was found. For example, in both cases, improved predictions were associated with the inclusion of a simple empirically based urban dispersion model within HPAC, whereas improvements associated with the inclusion of a more computationally intensive wind field module were not found. The use of meteorological observations closest to the array and well above the obstacle array—the sonic anemometer measurements 16 m above ground level—resulted in predictions with the best fit to the observed tracer concentrations. The authors speculate that including meteorological observations or vertical wind profiles above or upwind of an urban region might be a sufficient input to create reasonable HPAC hazard-area predictions.

1. Introduction

The potential effects of a terrorist attack involving the atmospheric release of chemical, biological, radiological, nuclear, or other hazardous materials continue to be of concern to the United States. The U.S. Departments of Defense and Homeland Security need to be able to estimate the effects resulting from hazardous releases within an urban environment on the underlying population to aid planning, emergency response, and recovery efforts. These estimates require accurate knowledge of the concentrations of dispersed material in time and space. In particular, estimates of where and when relatively low level human effects thresholds are exceeded are required.

Urban transport and dispersion modeling has been a subject of study since the late 1970s (Britter and Hanna 2003). Until recently, urban transport and dispersion models have been divided into two main categories: 1) low-fidelity models (e.g., urban canopy models) that account for the large-scale effects of urban terrain, such as drag from buildings and boundary layer perturbations, and 2) high-fidelity models (e.g., computational fluid dynamics models) that include detailed represen-
tations of buildings, streets, and other urban features as well as the effects of traffic-induced turbulence, heat-island formation, flows associated with the deep canyons of some large cities, the relative lack of moisture, and differential heating on building faces. Between those categories lies a recent class of urban transport and dispersion models that takes into account detailed urban features but employs empirical turbulence and wind profile parameterizations derived from these details in the propagation of contaminants. Examples are the particle-based “MESO”/Realistic Urban Spread and Transport of Intrusive Contaminants (RUSTIC) model (Diehl et al. 1982; Hendricks et al. 2004; Burrows et al. 2004) and the Defense Threat Reduction Agency (DTRA) puff-based Hazard Prediction Assessment Capability (HPAC; DTRA 2001) model with urban features (Urban HPAC). In this paper we concentrate on the HPAC model because we have been involved in validation efforts with HPAC for several years.

A few recent field experiments have included the release of environmentally safe, inert, tracer gases in urban environments. For example, tracer gases were released in Salt Lake City, Utah, in 2000 (Allwine et al. 2002) and in Oklahoma City, Oklahoma, in 2003 (Allwine et al. 2004). An important use of the data collected during these field experiments is to provide support for the evaluation of transport and dispersion models. Data collected during the 2000 Salt Lake City atmospheric tracer and meteorological study, which is referred to as “Urban 2000,” have been used to aid assessments of the validity of the HPAC (Warner et al. 2004a; Chang et al. 2005). The information associated with the 2003 Oklahoma City field experiment has only recently been released and will be the subject of future analyses.

In 2001, a well-documented baseline, idealized, scaled, urban setting was created by releasing tracer gases in an array of shipping containers positioned in the desert of Utah. This atmospheric tracer and meteorological study is known as the Mock Urban Setting Test (MUST; Biltoft 2002). The goal of MUST was to acquire meteorological and dispersion datasets at near-full scale for use in urban dispersion model development and evaluation. The array of shipping containers that were used during this field experiment did not represent an actual, or even notional, scaled-down version of a city. Rather, the array represented an ideal scaled setting that allowed for the careful investigation of the effects of multiple, identical, buildings on atmospheric transport and dispersion. Therefore, the focus of the MUST experiments was to better characterize transport and dispersion of plumes around buildings.

This paper describes comparisons of the tracer gas observations of MUST with HPAC predictions. Several sets of predictions based on different configurations of HPAC and varying meteorological inputs are examined in this paper. Details associated with the preparation of these predictions, for example, the protocols required to create plausible predictions, are described elsewhere (Warner et al. 2005). This paper presents comparisons between sets of predictions and provides assessments of relative performance among HPAC configurations.

Table 1 provides a list of acronyms and shorthand notations used throughout this paper.

a. Brief description of the MUST field experiment

A $12 \times 10$ array of shipping containers was positioned at Dugway Proving Ground in Utah to form an approximate $200$-m square, which was meant to represent a scaled version of an ideal, that is, simple, urban environment. The (nominally identical) containers had dimensions of $12.2$ m in length, $2.42$ m in width, and $2.54$ m in height. This 200-m square area is smaller than a real urban setting (e.g., a typical city) but larger than a wind tunnel experiment. Figure 1 shows a schematic drawing of the entire array of containers. In this planar view, the containers are shown as green rectangles at their actual locations. The figure also shows four rows of ground samplers.

The tracer gas propylene was released 68 times between 11 and 27 September 2001. Propylene concentrations were sampled at 50 Hz. There were five puff trials that involved a series of instantaneous releases, and there were 63 continuous release trials. The puff-based releases can provide information useful for the evaluation of probabilistic model outputs. The relatively high frequency of tracer gas sampling can allow for examination of the underlying statistical characteristics of the observed concentration fluctuations (Yee and Biltoft 2004). In this study, we consider the 63 continuous release trials. Propylene concentrations were not available for the five continuous releases associated with 11 September 2001 and for single releases on 14 and 15 September 2001. For the 56 remaining continuous releases, release locations within and just outside the array were varied and release heights were varied from 0.15 to 5.2 m. The MUST releases were conducted at night or in the early morning hours during stable conditions. Additional information on the MUST experiment can be found in Biltoft (2002).

The following meteorological observations associated with MUST were included:

- Sonic anemometer/thermometers, referred to here simply as sonics, were placed on a 32-m tower at the center of the grid [at 4, 8, 16, 24, and 32 m above
ground level (AGL) on pneumatic masts just outside the array (at 4, 8, and 16 m AGL), and on four 6-m towers within the array (at 2.4 and 6 m AGL) as shown in Fig. 1. Measurements were not collected on all sonics during all releases.

- Six portable weather information and display systems (PWIDS) collected wind speed, wind direction, temperature, and humidity data at 2 m AGL and were located within 5 km of the array.
- Four surface atmospheric measurement systems (SAMS) measured wind speed and direction at 10 m AGL and temperature, pressure, and humidity at 2 m AGL, with one SAMS located about 1.5 km south-southeast of the array and the other three located several kilometers away.
- A helium-filled balloon equipped with sensors to measure wind speed and direction (cup anemometer and vane) and temperature, pressure, and humidity, all as a function of altitude AGL, was used. This device is referred to as a tethersonde. Tethersonde “flights” occurred within 1 h of some MUST releases, and operation was from a position 330 m east of the northeast corner of the MUST array.
- A Doppler acoustic sounder (or sodar: “sound detection and ranging”), used to measure vertical profiles of wind and turbulence, was operated during MUST. This system provided 10-min-average wind profiles between 15 and 200 m AGL in 5-m intervals.
- A 924-MHz wind profiling radar was located 1.5 km south-southeast of the array. This system provided wind profiles in 100-m increments (gates) through several kilometers AGL.

Figure 2 shows the relative locations of the meteorological instruments that were outside of the array and used in this study—SAMS, PWIDS, tethersonde, sodar, and radar profiler.

b. Brief description of HPAC urban configurations

DTRA’s HPAC is composed of a suite of software modules that can generate source terms for hazardous releases, retrieve and prepare meteorological information for use in a prediction, model the transport and dispersion of the hazardous release over time, and plot and report the results of these calculations. HPAC has been applied to various national defense problems, including military studies and operational planning.

For hazardous material transport and dispersion,
HPAC uses the Second-Order Closure Integrated Puff (SCIPUFF) model and an associated mean wind field model (Sykes et al. 1996). SCIPUFF is a Lagrangian model for atmospheric dispersion that uses the Gaussian puff numerical method—an arbitrary time-dependent concentration field represented by three-dimensional Gaussian distributions—and bases its turbulent diffusion parameterization on second-order closure theories.

For urban applications of HPAC, the vertical wind and turbulence profiles can be modified to account for urban effects. The basic exponential profile shape that is used is described in Cionco (1972) and has been shown to fit a variety of canopy types, including plant canopies as well as discrete objects. In using HPAC 4.04 (4.04 was the designation of the Urban HPAC version available at the time of our study) to provide predictions in an urban environment, one can conveniently capture some of the effects of the urban canopy on transport and dispersion by setting the surface type to “urban.” For the baseline Urban HPAC predictions examined in this study, the surface type was indeed set to urban.

In addition to the baseline Urban HPAC predictive capability described above, HPAC offers an urban dispersion model (UDM) and an urban wind field module (UWM), either or both of which can be invoked. To use UDM and UWM, HPAC requires a building database that provides the locations, planar geometries, and heights of buildings to support the calculation of flows in the urban regime. The U.K. Defence Science and Technology Laboratory created the UDM (Hall et al. 2002) component of HPAC. The UDM is designed to compute the transport and dispersion of an instantaneous discharge [“a puff” or train(s) of puffs] of pollutant released over a surface containing a mixture of pollutants.

Fig. 1. Planar schematic view of the 12 × 10 array of containers at the MUST site, with close-in sonics sites (towers and masts) shown. The black dots correspond to the samplers, and the green rectangles correspond to the shipping containers.
open and urban areas. UDM considers variations in dispersion rates as a function of surface changes and direct interaction of pollutant cloud with surface obstacles. The UWM component of HPAC predicts steady-state winds inside the urban boundary layer using a canopy parameterization (Lim et al. 2003). UWM is designed to provide a computationally fast solution, within a computational fluid dynamics framework, by considering spatially averaged obstacle effects. Therefore, the predicted winds of UWM represent spatiotemporal averages. UWM-generated average winds can then be used by HPAC (e.g., with or without UDM) to drive material transport and dispersion. Initial conditions (e.g., wind vectors for the outer domain) used for the UWM computations within HPAC can be set by including a mass-consistent, three-dimensional, gridded wind field based on observations or by providing a numerical weather prediction. For these predictions, the Stationary Wind Fit and Turbulence (SWIFT) diagnostic wind field model, which is resident within HPAC, was used to generate this outer gridded wind field.

SWIFT is a mass-consistent wind model that ensures that air flows around and over (but not through) terrain features. For example, SWIFT can model wind direction deflections resulting from a localized terrain barrier or wind speed increases through a terrain pass. For this study, conducted over relatively flat terrain, SWIFT served the purpose of generating a gridded wind field from the relatively sparse meteorological observations (through interpolation), and this outer grid served as the initial condition for the more detailed UWM wind field computations. Additional details associated with the UDM and UWM components of HPAC can be found in Warner et al. (2004a) and the references therein.

2. Study method

a. HPAC predictions that were considered

The first goal of this study was to create plausible predictions of MUST using the HPAC software. HPAC predictions of the MUST field experiment have not
previously been reported. For this study, four types of Urban HPAC predictions were created: 1) surface type entered as “urban,” denoted baseline or “UC” (for urban canopy); 2) UDM, denoted “DM”; 3) UWM, denoted “WM”; and 4) both UDM and UWM, denoted “DW.” In general, HPAC default settings were used to create the predictions. Further details of the protocols used to create these predictions are provided in Warner et al. (2005).

1) BRIEF DESCRIPTION OF METEOROLOGICAL INPUT OPTIONS THAT WERE EXAMINED

For each model type (UC, DM, WM, and DW), four meteorological input options were examined. The inclusion of four different sets of meteorological input options was carried out to assess how transport and dispersion model performance might vary using different sets of available meteorological information. A brief description of the four input options that were used for this study is provided below.

First, all of the 5-min-averaged surface observations associated with the six PWIDS and four SAMS sites were used to create predictions denoted with the prefix “SUR,” as in SUR_UC. Next, a site 1.5 km south-southeast of the MUST array was chosen as a single site for meteorological observations. This site was generally upwind of the release and included observations from a single SAMS, a single PWIDS, and the radar wind profiler (described earlier) and is referred to as the “SPP” option (for SAMS, PWIDS, and profiler). For both the SUR and SPP options, the diagnostic wind field model SWIFT, which is resident within HPAC, was used to create gridded wind fields. SWIFT does not account for wind field profiles within an urban canopy (i.e., an environment with many buildings) and, in particular, the MUST array.

A third set of predictions used a meteorological input option referred to as “ALL.” This option included observations from the six PWIDS, the four SAMS, the radar wind profiler (mentioned previously), the sodar, and the tethersonde. Last, we examined the observations associated with the sonics mounted on the five towers within the array and two pneumatic masts just outside the array. These meteorological observations had the advantage of being closest to the array. To ensure that the sonics observations would be relatively unperturbed by the container array, we chose to use only the 16-m observations, which were available on the 32-m central tower and the two pneumatic masts. However, sonics data did not exist for three of the release days, and therefore for this meteorological input option, referred to as “SON,” only 37 releases were modeled. SWIFT was used to process the observations associated with the ALL and SON options.

The choice of these four meteorological input options can be understood in the following ways. First, SUR roughly corresponds to surface observations that might be available for a typical city. Surface observation stations usually include civilian and military airports and selected sites within and around cities. There are two possible justifications for the SPP meteorological option. This option roughly corresponds to surface and upper-air meteorological observations that might be available for a city from a major airport. In addition, the SPP option corresponds to a single station that provides vertical meteorological profile information that can be easily used to drive different urban transport and dispersion model predictions. Because of the somewhat limited spatial variability of the single-site winds (i.e., SPP), it is an ideal candidate for diagnostic analysis of the resulting hazard plume. The ALL meteorological option tries to simulate a situation with an abundance of meteorological information available to drive the urban model. In general, larger amounts of meteorological information are available during a field experiment than would be for a typical model application. As such, there is a possibility that too much meteorological information can be given to the model, particularly if the model is not specifically designed to ingest some types of data. A previous study of Urban 2000 has shown that using all available meteorological information as input to Urban HPAC did not result in improved prediction of hazards (Warner et al. 2004a). The SON meteorological option simulates a situation in which detailed information is available close to the release. There is an intuitive belief that providing detailed meteorological information near the release should necessarily lead to improved hazard-area predictions. For the previous Urban 2000 study (Warner et al. 2004a), one of the meteorological options selected included measurements from the top of a tall building in the downtown area of Salt Lake City. Urban HPAC evaluations using Urban 2000 showed that the resulting predictions were generally inferior to predictions obtained using the other meteorological options that were examined (Warner et al. 2004a). This apparent contradiction can be plausibly explained by postulating that the building-top measurements were affected by the nearby urban structures, resulting in variable winds, and that the models examined were not designed to use such variable, perhaps nonrepresentative, winds. By using sonic anemometer measurements at 16 m AGL, which was approximately 6 times the height of the container tops, it was expected that wind fields would not be strongly influenced by the containers. We chose the
16-m measurements because of their availability and because, at 6.3 times the shipping container height, they were expected to be less influenced by the containers than the 8- or 4-m sonics measurements. One study (Leone et al. 2001) suggests that the height of the surface roughness sublayer—a function of building height, building density, and stability—can vary between 2 and 5 representative building heights. Nelson et al. (2004) suggest that for nighttime wind measurements at the MUST site the constant stress layer occurs above about 6 building heights and above about 3 building heights in some cases. Measurements inside the “urban canopy” could be affected by local conditions and thus might not be representative of the flow over the region. It is this representative flow over the urban region that is required by model tools such as UWM and UDM. Additional details associated with the chosen meteorological input options are provided in Warner et al. (2005).

2) **SUMMARY OF HPAC PREDICTIONS OF MUST THAT WERE CREATED**

During our initial attempts to create HPAC predictions of MUST, we observed some peculiar model behavior that led us to suspect that there were problems with HPAC software, at least for these urban applications. For example, an integration error associated with the use of UWM was identified, and, working together with the model developer, a “workaround” was created (Warner et al. 2005). The proposed workaround for this problem required turning off the digital terrain feature in the SWIFT calculation. It was expected that, in the case of MUST, the exclusion of terrain in the outer SWIFT domain would have limited impact because of the relatively flat desert terrain surrounding the MUST array. This procedure was followed for the MUST WM and DW predictions. In a similar way, a separate set of predictions with UDM turned on was created with the terrain feature disabled. This was done so that a strictly comparable set of UDM-based predictions (at least for comparisons with WM and DW) would be available for analysis. This set of predictions was labeled “D1.” The results that are described in this paper are associated with predictions that were created with the currently available version of HPAC (4.04 SP3) that incorporated the software corrections associated with the errors discovered during the early part of this study. Additional details associated with the creation of HPAC predictions and the identification of software errors can be found in Warner et al. (2005).

Twenty sets of HPAC predictions were generated as described in Table 2. Nomenclature associated with the 112 project folders is as follows: xxx_yyy_zz, where xxx describes the yearday of the MUST release, yyy describes the meteorological input option, and zz describes the transport and dispersion model mode (e.g., 256_SUR_UC).

**b. Protocol for point-to-point comparisons of predictions and observations**

For this analysis, we compared predictions and observations paired in space and time; we refer to this kind of analysis as “point to point” comparisons. For MUST, one can consider 40 surface sampler locations and 56 independent releases. For each release, we compared predictions and observations for 10-s, 1-min, 5-min, and “whole duration” concentrations. Whole duration means that the averaging time was set to the release duration plus 120 s to cover the maximum offset of sampler operating time after the gas was turned off and was estimated by reviewing sampler “on” and gas “off” times.

The methods used for the comparisons of this study have been previously described (Warner et al. 2004a,b). A variety of statistical metrics to examine bias, scatter, and correlation were examined, as was a user-oriented measure of effectiveness (MOE; Warner et al. 2004b) that allowed for assessments of the ability of the model to predict the “hazardous” region—that is, the region above a concentration threshold of interest. The metrics discussed in this paper are fractional bias (FB), normalized absolute difference (NAD), bounded normalized mean-square error (BNMSE), fraction of predictions within a factor of x (FACx), and the MOE. These metrics are defined as

---

**Table 2. Shorthand notations for the 20 model combinations (5 transport and dispersion modes × 4 meteorological input options) that were examined.**

<table>
<thead>
<tr>
<th>Meteorological input options</th>
<th>UC</th>
<th>DM</th>
<th>D1</th>
<th>WM</th>
<th>DW</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUR</td>
<td>SUR_UC</td>
<td>SUR_DM</td>
<td>SUR_D1</td>
<td>SUR_WM</td>
<td>SUR_DW</td>
</tr>
<tr>
<td>SPP</td>
<td>SPP_UC</td>
<td>SPP_DM</td>
<td>SPP_D1</td>
<td>SPP_WM</td>
<td>SPP_DW</td>
</tr>
<tr>
<td>ALL</td>
<td>ALL_UC</td>
<td>ALL_DM</td>
<td>ALL_D1</td>
<td>ALL_WM</td>
<td>ALL_DW</td>
</tr>
<tr>
<td>SON</td>
<td>SON_UC</td>
<td>SON_DM</td>
<td>SON_D1</td>
<td>SON_WM</td>
<td>SON_DW</td>
</tr>
</tbody>
</table>
\[ FB = \frac{(C_p - C_o) - (\bar{C}_o + C_p)}{2C_o + C_p}, \quad (1) \]

\[ NAD = \frac{\sum_{i=1}^{n} |C_p^o - C_o^i|}{\sum_{i=1}^{n} [C_p^o + C_p^i]}, \quad (2) \]

\[ BNMSE = \frac{\sum_{i=1}^{n} [C_p^o - C_o^i]^2}{\sum_{i=1}^{n} [C_p^o + C_p^i]^2}, \quad (3) \]

\[ FAC_x = \text{fraction of data for which } 1 \leq \frac{C_p}{C_o} \leq x, \quad \text{and} \]

\[ MOE = (x, y) = \left( \frac{A_{OV}}{A_{OB}}, \frac{A_{OV}}{A_{PR}} \right) = \left( 1 - \frac{A_{FN}}{A_{OB}}, 1 - \frac{A_{FP}}{A_{PR}} \right), \quad (4) \]

where \( C \) = observation/prediction of interest (e.g., concentration), \( C_p \) corresponds to model predictions, \( C_o \) corresponds to observations, a bar above the quantity (e.g., \( \bar{C} \)) denotes the average, \( n \) = number of data points used in the comparisons, \( C_o^i \) refers to the \( i \)th observed concentration, \( C_p^i \) refers to the \( i \)th predicted concentration, \( x = 2, 5, \) and 10, \( A_{FN} \) = region of false negative, \( A_{FP} \) = region of false positive, \( A_{OV} \) = the region of overlap, \( A_{PR} \) = region of the prediction, and \( A_{OB} \) = region of the observation.

The MOE described above has two dimensions, with the \( x \) axis corresponding to the ratio of the overlap region to the observed region and the \( y \) axis corresponding to the ratio of the overlap region to the predicted region. When these mathematical definitions are algebraically rearranged, one recognizes that the \( x \) axis corresponds to \( 1 \) minus the false negative fraction (e.g., the decreasing false negative axis in Figs. 5 and 6, which are described below) and the \( y \) axis corresponds to \( 1 \) minus the false positive fraction (e.g., the decreasing false positive axis in Figs. 5 and 6, which are described below). This two-dimensional MOE includes directional effects; that is, the prediction of the location of a hazard and not just the shape and size of the plume are critical to obtaining a high MOE “score.” The perfect MOE score is \((1, 1)\), implying complete “overlap” of the predictions and the observations. From Eq. (5) it can be seen that MOE values along the “diagonal” of the two-dimensional MOE space indicate equal sizes (e.g., areas or amounts of material) of the prediction and the observation (i.e., \( A_{PR} = A_{OB} \)), even if the locations differ. The quantities \( A_{FN}, A_{FP}, \) and \( A_{OV} \) can be computed directly from the predictions and field trial observations paired in space and time and can be defined as:

\[ A_{OV} = \sum_{i=1}^{N} A_{OV}(i), \]

\[ A_{FN} = \sum_{i=1}^{N} A_{FN}(i), \quad \text{and} \]

\[ A_{FP} = \sum_{i=1}^{N} A_{FP}(i). \quad (6) \]

For average concentration–based calculations, at index \( i \), we set:

\[ A_{OV}(i) = \min\left[ C_p^o, C_o^i \right], \]

\[ A_{FN}(i) = \begin{cases} C_o^i - C_p^i & \text{if } C_o^i > C_p^i \\ 0 & \text{otherwise} \end{cases} \]

\[ A_{FP}(i) = \begin{cases} C_p^i - C_o^i & \text{if } C_p^i > C_o^i \\ 0 & \text{otherwise} \end{cases} \quad (7) \]

For “threshold exceedance quantities” at index \( i \), given threshold of interest \( T \), we set:

\[ A_{OV}(i) = \begin{cases} 1 & \text{if } C_o^i \geq T \text{ and } C_p^i \geq T \\ 0 & \text{otherwise} \end{cases} \]

\[ A_{FN}(i) = \begin{cases} 1 & \text{if } C_o^i \geq T \text{ and } C_p^i < T \\ 0 & \text{otherwise} \end{cases} \]

\[ A_{FP}(i) = \begin{cases} 1 & \text{if } C_p^i < T \text{ and } C_p^i \geq T \\ 0 & \text{otherwise} \end{cases} \quad (8) \]

Additional discussions associated with the computation of average concentration–based and threshold-based MOE values can be found in Warner et al. (2004a,b).

Nonparametric hypothesis test methods for detecting statistically significant differences between sets of predictions for a given metric and procedures for estimating confidence intervals are described in Warner et al. (2004a,b, 2005). In addition, and as important as any metric, comparative plots of model predictions and observations were created and scrutinized for all releases.

3. Results and discussion

Figure 3 provides an example comparison of predictions and observations on a logarithmic scale. These example comparisons are shown for the 40 surface-based sampler locations and are based on a single
MUST release. Predictions from four HPAC urban modes—UC, DM, WM, and DW—are shown. Green shading in the figure corresponds to “overlap,” which indicates levels of concentration that were both predicted and observed. Red shading corresponds to “false negative” and indicates marginal levels of concentration observed but not predicted. Yellow shading corresponds to “false positive,” indicating marginal levels of concentration predicted but not observed. This figure indicates that, for this single release, the WM mode led to substantial underprediction (considerable false negative at several locations) and the DM mode led to the
least underprediction. The rest of the discussions in this paper are based all 56 MUST releases that were examined.

a. Results of HPAC–MUST comparisons are consistent with previous HPAC–Urban 2000 comparisons

Several of the general findings of this study were consistent with those of previous Urban 2000 analyses. For example, when considering the surface samplers, the FAC2, FAC5, and FAC10 results for the MUST comparisons ranged from 0.21 to 0.39, from 0.37 to 0.65, and from 0.47 to 0.77, respectively. The corresponding Urban 2000 FAC2, FAC5, and FAC10 ranges were 0.28–0.35, 0.41–0.52, and 0.50–0.67, respectively. Next, we found that comparative results (based on hypothesis testing) of different HPAC configurations led to similar conclusions regardless of the time resolution that was examined. For MUST, we compared average observed and predicted concentrations for 10-, 60-, and 300-s time periods, as well as for the entire release duration, which was typically about 15 min. Previous Urban 2000 studies found an identical result—that is, similar relative results among HPAC configuration regardless of the time resolution examined, albeit with time resolutions of 30 min and 2 h. Last, our studies of both MUST and Urban 2000 revealed that predictions of low-threshold hazard regions (i.e., sampler locations and times at which a low threshold is exceeded) are much more improved in terms of the MOE than are predictions of average concentrations at specific locations and times. The implication is that, even in an urban environment, low-threshold hazard regions may be relatively well predicted by HPAC. Of course, this conclusion rests on field-trial-quality meteorological and source-term information and is therefore limited for some operational applications for which such information may not be available.

b. Model bias, with respect to over- and underprediction of the average concentration, varies by HPAC mode

In general, HPAC predictions completed in the UC and WM modes resulted in less material being predicted at the surface samplers than was observed (underprediction). On the other hand, for the HPAC predictions completed with UDM turned on (DM, D1, and DW), more material was predicted at the surface samplers than was observed (overprediction). Underpredictions of the average concentration for a resolution equal to the whole release duration time at the surface samplers for UC predictions varied from about a factor of 1.2 to 3.0. For the WM mode, underpredictions varied from about a factor of 1 (i.e., no bias) to 2.0. In contrast, the DM and D1 modes resulted in overpredictions that ranged from about a factor of 1.6 to 3.0, and the DW mode similarly led to an overprediction range of about factors of 1.3 to 2.2. The above conclusions were based on examinations of bias, fractional bias, and MOE values. Examination of plots of predicted and observed values suggests that the inclusion of UDM led to a broader, more dispersed plume that tended to reside in the array longer than the plumes generated by the UC or WM modes. Such behavior is in general agreement with our understanding of UDM modeling features. This plume behavior (with respect to broadness and residence time) could have contributed to the variability in relative bias associated with these different HPAC modes that was observed. Therefore, the relative under- versus overpredicting nature of the UC (or WM) and DM modes can be simply understood.

c. Close-in sonics meteorological observations resulted in the best predictions

Of the four meteorological input options examined, the SON option resulted in the best predictions; for example, the SON_DM configuration resulted in the lowest overall scatter between observations and predictions as measured by NAD. Figure 4 compares NAD values (where a value of 0.0 implies a perfect prediction) for several HPAC configurations. This figure shows that the predictions with the SON option resulted in the lower NAD for all HPAC modes (UC, DM, D1, WM, and DW) when compared with the ALL option (for the 22 releases that these sets of predictions had in common). In a similar way, HPAC predictions that used the SON option always resulted in lower (or in one case, the same) NAD values than those that used the SUR option (for the 37 releases that these two sets of predictions had in common). Comparisons of NAD values for predictions completed with the SON and SPP options led to mixed results. The SON_DM and SON_D1 predictions resulted in the lowest overall NAD values. However, the SPP_WM predictions resulted in a lower NAD value than the SON_WM predictions. A similar, although marginal, difference was found for the corresponding DW predictions, and no difference was seen for the UC predictions. We speculate that some compensating errors may have caused the relatively improved performance associated with including the SPP option with the WM HPAC mode.

Overall, however, the conclusion is that predictions completed with the SON meteorological option resulted in the least scatter and typically the best MOE.
values—that is, those closest to the perfect (1, 1). We find this result particularly plausible because the SON option includes meteorological observations very close to and “within” the array but, as noted earlier, at a height corresponding to 6.3 times the container height, likely placing these observations out of the perturbed flow. For comparison, the closest meteorological observation associated with the SUR and SPP options was about 1.5 km from the array. We also conclude here that predictions completed with the SON meteorological input option are best for carrying forward comparisons of various HPAC dispersion modes (UC, DM, D1, WM, and DW) because this option appeared to provide for the most accurate representation of the wind field based on the previously described relative model performance.

The previous Urban 2000 comparisons (Warner et al. 2004a) found that the building-top measurements showed significant temporal variability in wind direction. For those Urban 2000 predictions, the inclusion of a building-top measurement did not improve model performance. We speculated that this was because the measurements were too close to the actual building and perhaps not high enough above the nearby buildings (e.g., less than 2 times the height of the 25 tallest buildings in the downtown Salt Lake City area); thus the measurements were perturbed by the local conditions and hence were not representative of the larger-scale flows.

d. UDM improved predictions, UWM did not

The discussion of this section focuses on the results obtained when the close-in meteorological observations associated with the SON input option were used. The inclusion of UDM improved predictions of the MUST observations relative to the baseline (UC) predictions. A review of Fig. 4 shows that the SON_DM and SON_D1 configurations resulted in lower NAD values than the corresponding UC configuration.

Figure 5 compares approximate 0.95 confidence regions (the colored clusters of points) for MOE values associated with SON_DM and SON_UC. MOE values at three thresholds—0.01, 0.1, and 1 ppm—as well as for average concentration are shown. The perfect MOE value, complete overlap of the predictions and observation, occurs at (1, 1). Values above the diagonal indicate underprediction, and values below the diagonal indicate overprediction. Figure 5 shows that the UC mode led to underpredictions of the number of samplers that exceeded the threshold (increased false negative) for all thresholds (over two orders of magnitude). This conclusion holds for all four meteorological input options that were considered. In contrast, the DM predictions resulted in about the right number of samplers that exceeded the threshold (increased false negative) for all thresholds (over two orders of magnitude). We conclude that the DM predictions correspond to an improvement relative to the baseline (UC) in terms of predicting the hazard area, at least at these threshold levels; that is, the DM predictions resulted in a substantially lower false-negative fraction.

For the MOE values based on average concentration, underprediction is seen for the UC mode and overprediction is seen for the DM mode. Comparisons of bias and fractional bias values for UC and DM predictions confirm this finding. These threshold-based and aver-
age concentration trends—“UC underpredicting and DM about right or overpredicting”—are consistent with the results of past Urban 2000 studies (Warner et al. 2004a).

Overall, the DM mode appears to be an improvement over the UC mode for predictions of MUST, especially when the “best” meteorological option, SON, is used. The improvement of DM predictions relative to UC is also particularly evident when hazard regions (e.g., number of samplers exceeding a low threshold) are considered or when minimization of the false-negative fraction is considered as especially important.

The inclusion of UWM did not improve HPAC predictions of MUST with respect to measures of scatter or MOE values. In fact, there was some evidence that UC predictions were improved relative to WM. Figure 6 compares MOE confidence regions for the UC and WM predictions for the SON option. With respect to the average concentration and 1.0-ppm threshold-based MOE values, the UC and WM confidence regions over-

---

**Fig. 5.** Comparisons of MOE values for HPAC SON_DM and SON_UC predictions of MUST observations at 40 surface samplers. Values are computed for the predictions of whole-duration average concentrations. Colored clusters (red = DM; black = UC) correspond to approximate 0.95 confidence regions, with the point estimate located near the center of the cluster. Values are shown for three thresholds—(upper left) 0.01, (upper right) 0.1, and (lower left) 1 ppm—and for (lower right) average concentration.
lap considerably, with both locations in the two-dimensional MOE space indicating underpredictions. For the lower threshold-based values (0.01 and 0.1 ppm), the UC predictions appear to be an improvement over the WM predictions, having somewhat reduced false-negative fractions. Overall, the conclusion is that the addition of UWM did not improve upon the baseline (UC) HPAC predictions. The addition of UWM to UDM (to create DW) did not lead to consistent improvement over the UDM-only mode (D1) for predictions of MUST. In fact, the UDM-only mode D1 resulted in better threshold-based MOE values for the SON and ALL meteorological input options than did the DW mode.

For the SON option, the ranking of HPAC modes shown in Fig. 7 is consistent with hypothesis testing that examined measures of scatter and MOE values. The above conclusions with respect to HPAC-mode relative performance are consistent with previous studies of Urban 2000.

4. Conclusions

A critical result of this study is the relative consistency of conclusions found between comparisons of
HPAC predictions of MUST observations and those previously described for predictions of Urban 2000. For example, for both MUST and Urban 2000, improved predictions were associated with the inclusion of UDM; the HPAC baseline mode (UC) typically led to larger false-negative fractions with respect to hazard-area (i.e., number of samplers exceeding a threshold) predictions, and improvements associated with the inclusion of UWM were not found. We note that the MUST array is not simply a scaled-down version of a typical city and, as such, corresponds to a relatively unique environment (e.g., relative to the Urban 2000 Salt Lake City experiment). One might view the MUST array scale as being most consistent with a small industrial facility. The above findings of relative consistency are therefore suggested to be somewhat robust.

An important part of future studies, including the planned evaluation of HPAC with information collected during the Joint Urban 2003 field experiment, will be to confirm and expand upon these previous findings. Urban 2000 studies (Warner et al. 2004a) found that using close-in meteorological measurements—for example, from a downtown building top—did not result in improved HPAC predictions. The current study finds that the use of meteorological observations well above the obstacle array (the sonic-anemometer measurements at 16 m AGL) resulted in predictions with the best fit to the observed tracer concentrations. We speculate that including meteorological observations or vertical wind profiles above or upwind of an urban region might be sufficient input to create reasonable HPAC hazard-area predictions. It has been suggested recently that real-time wind field information from lidar-based systems can be used to construct vertical profiles near and above the city of interest to generate improved hazard predictions (Allwine et al. 2004). These ideas will be explored in future research.

In addition, a wider range of meteorological conditions, perhaps higher wind speeds or more variable wind speeds and directions than were observed in Salt Lake City during Urban 2000 or during MUST, might create situations wherein the hoped-for benefits of UWM can be demonstrated. To this end, we are currently investigating relatively high resolution UWM predictions of a recent set of tracer gas releases conducted in Oklahoma City (Allwine et al. 2004). Questions associated with what represents the most appropriate operational meteorological inputs for the prediction of transport and dispersion within an urban environment will also, again, need to be addressed in future studies.

The MUST field experiment, in large part because of its scale, served as a substantial challenge for HPAC and it has forced a close look at urban capabilities in ways that would not have been illuminated at actual urban scales. A few software integration errors associated with UDM and UWM were uncovered during this study that likely would not have been discovered at full scale. For example, if the too-fast predicted wind speeds in the UWM domain that led to the discovery of a software integration error had been associated with a full-scale experiment (with the concomitant lower computational resolution), one might easily have accepted such behavior as plausible, especially for lower wind speeds (e.g., as were present in Salt Lake City during much of Urban 2000).

Acknowledgments. This effort was supported by the Defense Threat Reduction Agency, with Mr. Richard N. Fry Jr. as project monitor, and the Central Research Program of the Institute for Defense Analyses. The authors thank R. Ian Sykes and Doug Henn of Titan Corporation, David R. Brook of the U.K. Defence Science and Technology Laboratory, and David Stewart of RiskAware for valuable discussions and assistance. The views expressed in this paper are solely those of the authors. No official endorsement by the U.S. Department of Defense is intended or should be inferred.

REFERENCES


