CALSUFF, HPAC, and VLSTRACK Evaluation with the Over-Land Alongwind Dispersion (OLAD) Field Data

J. C. Chang, K. Chayantrakom & S. R. Hanna
School of Computational Sciences
George Mason University, Fairfax, Virginia, U.S.A

Abstract

The CALPUFF (California Puff), HPAC (Hazard Prediction and Assessment Capability), and VLSTRACK (Chemical/Biological Agent Vapor, Liquid and Solid Tracking) transport and dispersion models were evaluated with the Over-Land Alongwind Dispersion (OLAD) field data. OLAD involved nearly instantaneous line releases of sulfuric hexafluoride (SF$_6$) from truck or aircraft on a mesoscale field site at Dugway Proving Ground, Utah. Meteorological data were measured at 16 surface sites, at one radiosonde site, and at one pibal site. SF$_6$ concentrations were mainly measured by the whole-air samplers along three sampling lines. In general, all three models underpredicted the maximum dosages across a sampling line. CALPUFF, HPAC, and VLSTRACK predictions indicated a factor of 1.8, 2.0, and 3.3 mean underprediction, respectively; and a factor of two to seven random scatter for all three models, with HPAC having the smallest scatter. The fraction of predictions within a factor of two of observations was about 30% for CALPUFF, 25% for VLSTRACK, and about 50% for HPAC. The underprediction might be due to the model formulations for the vertical dispersion coefficient and the cloud advective speed.

1 Introduction

Transport and dispersion models can predict the fate of toxic substances released in the atmosphere, and the associated impacts on the public and on the environment. A variety of release scenarios is possible. For example, a release
can be routine (e.g., from a smoke stack), accidental (e.g., from a train derailment), or episodic (e.g., from chemical weapons in a battle field). It is important to properly evaluate these models before their predictions can be used to make decisions of grave consequences.

A transport and dispersion model can be evaluated using at least three approaches: statistical, scientific, and operational (Hanna et al. [1]). In a statistical evaluation, the model can be treated as a “black box” where model outputs are examined to see how well they match observations. It is sometimes possible that a model produces the right answers, but as a result of compensating errors. In a scientific evaluation, the model algorithms, physics, and assumptions are examined in detail for their accuracy, efficiency, and sensitivity. In an operational evaluation, issues related to the user-friendliness of the model are considered, such as the user’s guide, the user interface, error checking of input data, internal model diagnostics, and output display. The main focus of this paper is on statistical evaluation. Scientific evaluation of the three models can be found in references such as Pendergrass et al. [2] for VLSTRACK, Allwine et al. [3] for CALPUFF, and Bradley et al. [4] for HPAC.

2 Over-Land Alongwind Dispersion (OLAD) field experiments

The Over-Land Alongwind Dispersion (OLAD) Field Experiments were conducted on 8-25 September 1997 at West Desert Test Center (~40°N, 113°W), U.S. Army Dugway Proving Ground, Utah. Watson et al. [5] and Biltoft et al. [6] provide detailed descriptions of the experiments. The test domain is mostly mud flat, with surrounding mountains (see Figure 1). The typical value of the surface roughness length is 3 cm (Biltoft et al. [6]).

The experiments involved releases of SF$_6$ from a truck or an aircraft in the early morning hours. The field data set bears particular relevance to the growing concern of the use of chemical and biological warfare agents by terrorist groups as weapons of mass destruction. The predominant wind direction was from the southeast. Approximately 10 to 15 kg of SF$_6$ was released by a truck-mounted disseminator traveling at a normal speed of 64 km/hr over a distance of 8 km. About 100 kg of SF$_6$ was released by an aircraft traveling at 200 km/hr over a distance of 16 km. The release heights for the truck and aircraft releases were 3 and 100 m above the ground, respectively. Table 1 shows the OLAD trial dissemination start and end times, positions, and mass; as well as typical wind speeds at 10 m above the ground.

Three lines of whole-air samplers were deployed with 15 samplers at each line to measure SF$_6$ concentrations. Each whole-air sampler measured 15-min average concentration, and was located roughly 100 m from adjacent samplers. The sampling lines were located 2, 5, and 10 km downwind for the truck releases; and were located 10, 15, and 20 km downwind for the aircraft releases. The total sampling period was three hours for each trial. High frequency (i.e., 4 Hz) SF$_6$
Figure 1. Terrain of the Over-Land Alongwind Dispersion (OLAD) test site at West Desert Test Center, U.S. Army Dugway Proving Ground, Utah. Two-meter Portable Weather Information and Display System (PWIDS) instrument masts are listed as P1, P2, etc. Ten-meter Surface Atmospheric Measurement Systems (SAMS) instrument towers are listed as S2, S3, etc. The radiosonde and pibal measuring site is listed as RP. The thin dashed line is the truck line source, and the thin solid lines are the corresponding sampling lines. The thick dashed line is the aircraft line source, and the thick solid lines are the corresponding sampling lines. The southwest corner of the map roughly corresponds to (39.9°N, 113.4°W).
Time is the Mountain Standard Time (MST). Trial 12 was interrupted at six minutes by a disseminator failure, and resumed five minutes later.

<table>
<thead>
<tr>
<th>Trial No.</th>
<th>Date m/dd/yy</th>
<th>Start Time (MST)</th>
<th>End Time (MST)</th>
<th>Mass (kg)</th>
<th>Release Type</th>
<th>Typical 10-m wind speed (m/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9/08/97</td>
<td>0606</td>
<td>0614</td>
<td>12.4</td>
<td>Truck</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>9/09/97</td>
<td>0545</td>
<td>0553</td>
<td>12.9</td>
<td>Truck</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>9/10/97</td>
<td>0629</td>
<td>0632</td>
<td>100.3</td>
<td>Aircraft</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>9/11/97</td>
<td>0556</td>
<td>0559</td>
<td>100.5</td>
<td>Aircraft</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>9/12/97</td>
<td>0558</td>
<td>0608</td>
<td>12.8</td>
<td>Truck</td>
<td>1.5</td>
</tr>
<tr>
<td>6a</td>
<td>9/15/97</td>
<td>0545</td>
<td>0552</td>
<td>12.5</td>
<td>Truck</td>
<td>9</td>
</tr>
<tr>
<td>6b</td>
<td>9/15/97</td>
<td>0646</td>
<td>0658</td>
<td>11.4</td>
<td>Truck</td>
<td>9</td>
</tr>
<tr>
<td>6c</td>
<td>9/15/97</td>
<td>0730</td>
<td>0743</td>
<td>12.6</td>
<td>Truck</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>9/15/97</td>
<td>0945</td>
<td>0956</td>
<td>12.8</td>
<td>Truck</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>9/17/97</td>
<td>0548</td>
<td>0551</td>
<td>96.1</td>
<td>Aircraft</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>9/18/97</td>
<td>0655</td>
<td>0705</td>
<td>7.1</td>
<td>Truck</td>
<td>3.5</td>
</tr>
<tr>
<td>11</td>
<td>9/24/97</td>
<td>0609</td>
<td>0612</td>
<td>99.1</td>
<td>Aircraft</td>
<td>1.5</td>
</tr>
<tr>
<td>12</td>
<td>9/25/97</td>
<td>0300 0311</td>
<td>0306 0319</td>
<td>12.1</td>
<td>Truck</td>
<td>3</td>
</tr>
</tbody>
</table>

Concentrations were also measured by the TGA-4000 (Tracer Gas Analyzers) samplers deployed at both ends of each sampling line.

Winds were measured on the mesoscale domain by eight 2-m PWIDS (Portable Weather Information and Display System) masts, and eight 10-m SAMS (Surface Atmospheric Measurement System) towers (see Figure 1). Upper air winds were measured by pilot balloons (pibals) and radiosondes.

### 3 Overview of dispersion models

The CALPUFF (version 5.0; Scire et al. [7]), HPAC (version 3.2; DTRA [8], and VLSTRACK (version 3.0; Bauer and Gibbs [9]) transport and dispersion models were evaluated with the data collected during the OLAD field experiments. Chang et al. [10] describe in detail how each model was configured. CALPUFF has been traditionally used in U.S. Environmental Protection Agency's (EPA) regulatory applications, where environmental impacts due to routine industrial releases are modeled. HPAC and VLSTRACK, on the other hand, are models widely used by the U.S. Department of Defense.
(DoD) to estimate potential hazards due to dissemination of chemical, biological, and nuclear weapons.

The dispersion model within HPAC is the Second-Order Closure Integrated Puff (SCIPUFF, Sykes et al. [11]) model, which is capable of predicting both the mean and variance of the concentration field. Fundamentally, all three models are based on a Gaussian puff dispersion formulation. However, the three models are different in several ways, such as their boundary layer parameterizations, their treatments of terrain and the transport wind field, their handling of land surface characteristics, and their data ingestion methods and requirements.

Although CALPUFF is a “puff” model, it accepts only hourly average emission and meteorological information, and predicts only hourly average concentrations. As a result, the model cannot directly compare to high-frequency data that are typical of puff experiments such as OLAD. This limitation primarily results from the fact that the model has been traditionally used in regulatory applications where hourly is the predominant time interval of choice. HPAC and VLSTRACK can readily accept and produce higher frequency data. Because all models need to be compared on an equal basis, the higher-frequency data of OLAD were mostly not used.

CALPUFF has its own diagnostic wind-field model called CALMET (Scire et al. [12]). HPAC has two diagnostic wind-field models, the HPAC mass-consistent wind model (MC-SCIPUFF) and the more advanced SWIFT (Stationary Wind Fit and Turbulence) model. SWIFT is adapted from the MINERVE diagnostic model (Perdriel et al. [13]). SWIFT is the default choice for HPAC. VLSTRACK does not have an integrated wind-field model. Observed wind measurements are directly used in the model’s three-point interpolation scheme, with no attempt to maintain mass consistency by adjusting wind fields to the bottom topography.

4 Evaluation methodology

Because the total sampling period for OLAD was only three hours, concentrations from the whole-air samplers integrated over the three-hour measuring period (i.e., dosages) were primarily used for model evaluation. CALPUFF is limited to producing only hourly average concentrations, whereas HPAC and VLSTRACK can produce results at a much shorter time interval (20-s and 60-s, respectively). Therefore, the three-hour sampling period corresponds to only three CALPUFF predictions, but 540 and 180 HPAC and VLSTRACK predictions, respectively. Model evaluation was mainly based on the maximum dosage anywhere along a sampling line.

The model performance was assessed using two basic methodologies. The first methodology involved the use of scatter plots for direct quantitative comparisons of observed and predicted dosages. The second methodology involved the
application of rigorous statistical procedures, which quantify several relevant performance measures (Hanna [14], and Hanna et al. [15]).

For a dataset where both predicted and observed values vary by many orders of magnitude, Hanna et al. [15] consider the following statistical measures to determine quantitative model performance: the geometric mean bias (MG), the geometric mean variance (VG), and the fraction of predictions within a factor of two of observations (FAC2).

\[ MG = \exp \left( \frac{\ln C_o - \ln C_p}{n} \right) \]  
\[ VG = \exp \left[ \frac{(\ln C_o - \ln C_p)^2}{n} \right] \]  
\[ FAC2 = \text{fraction of data which } 0.5 \leq \frac{C_p}{C_o} \leq 2.0 \]

where
\( C \) evaluation objective (the maximum dosage in this paper),
\( C_p \) model predictions,
\( C_o \) observations,
\( \overline{C} \) average, and
\( \sigma_C \) standard deviation.

MG deals with mean biases, whereas VG deals with variances or random scatter. A perfect model would have MG, VG, and FAC2 = 1.0.

Bootstrap resampling (Efron [16]) was used to estimate the mean, \( \mu \), and standard deviation, \( \sigma \), of the above performance measures. The 95% confidence intervals for the performance measures are defined as

\[ \mu \pm t_{95\%} \sigma \left( \frac{n}{n-1} \right)^{1/2} \]

where \( n \) is the number of resamples (e.g., 1000), and \( t_{95\%} \) is the Student’s t value at the 95% confidence limits with \( n-1 \) degrees of freedom.

As suggested by their definitions, MG and VG are strongly influenced by extremely low values, including zeros, which are not uncommon in dispersion modeling. Therefore, when calculating MG and VG for observed or predicted values whose magnitudes may be very low, it is useful to impose a minimum threshold below which the data values are not allowed to drop. The whole-air samplers’ level of detection (LOD) for the SF₆ concentrations was chosen as the basis of this minimum threshold. Watson et al. [5] report that the LOD for OLAD was 3 ppt. Therefore, the corresponding minimum threshold for the 3-hr dosage was around 10 ppt-hr. This lower threshold was used in the scatter plots and in the calculations of MG and VG, to be presented later.
One nice feature of MG and VG is that they already involve the ratio of predicted to observed values. Therefore, for example, MG = 3 would imply a factor of three mean underprediction, MG = 0.3 would imply a factor of three overprediction, and VG = 12 would imply a random scatter that is roughly equivalent to a factor of 3.5 ($\sqrt{12}$) mean bias.

5 Evaluation results

Table 2 summarizes the results of statistical performance evaluation. In general, all three models underpredicted, with CALPUFF having the smallest mean underprediction. For the maximum dosage along a sampling line, CALPUFF, HPAC, and VLSTRACK yielded values of the geometric mean bias (MG) corresponding to a factor of two to three underprediction, and values of the geometric variance (VG) corresponding to a factor of two to seven random scatter. Significance tests show that the values of MG and VG are not significantly different among the three models at the 95% confidence limits. The fraction of predictions within a factor of two of observations (FAC2) was about 30% for CALPUFF, 50% for HPAC, and 25% for VLSTRACK. Figure 2 shows the scatter plots for the maximum dosage, where a threshold dosage of 10 ppt-hr mentioned above was used. There are a few cases where large model errors (up to two orders of magnitude) existed. Further investigation shows that the model performance was appreciably better for the high-wind (≥ 6 m/s) cases. Chang et al. [10] suggest that for a model to systematically underpredict dosage (concentration integrated with time) for an instantaneous line source, the model might overpredict either the vertical dispersion coefficient or the cloud advective speed.

Table 2. Summary of performance measures, including MG (geometric mean bias), VG (geometric mean variance), and FAC2 (fraction of predictions within a factor of two of observations) for the maximum dosage (ppt-hr) along a sampling line for CALPUFF, HPAC, and VLSTRACK at the OLAD site. The average and the highest value of the data are also shown.

<table>
<thead>
<tr>
<th></th>
<th>Observed</th>
<th>CALPUFF</th>
<th>HPAC</th>
<th>VLSTRACK</th>
</tr>
</thead>
<tbody>
<tr>
<td>MG</td>
<td>n/a</td>
<td>1.819</td>
<td>2.057</td>
<td>3.340</td>
</tr>
<tr>
<td>VG</td>
<td>n/a</td>
<td>9.82</td>
<td>3.61</td>
<td>45.68</td>
</tr>
<tr>
<td>FAC2</td>
<td>n/a</td>
<td>0.286</td>
<td>0.476</td>
<td>0.238</td>
</tr>
<tr>
<td>Average (ppt-hr)</td>
<td>2101</td>
<td>748</td>
<td>854</td>
<td>418</td>
</tr>
<tr>
<td>Highest (ppt-hr)</td>
<td>10210</td>
<td>2988</td>
<td>3993</td>
<td>1004</td>
</tr>
</tbody>
</table>
The scatter plots can also be studied to see how well the models predict the highest dosages. Of the five highest observed dosage points, two of these points were included in the five highest predicted dosage points by CALPUFF, four of them by HPAC, and two of them by VLSTRACK. The single highest observed dosage (ppt-hr) on the plots was 10210. The single highest predicted dosage was 2988 for CALPUFF, 3993 for HPAC, and 1004 for VLSTRACK, or underpredictions of a factor of 3.4, 2.6, and 10.2, respectively. HPAC was the only model where the time of the observed highest dosage was the same as the time of the predicted highest dosage.

6 Conclusions and discussions

The performance of three commonly-used transport and dispersion models, CALPUFF, HPAC, and VLSTRACK, was evaluated with the OLAD field data,
which involved instantaneous line sources of SF₆ over mesoscales. All three models underpredicted the maximum dosage along a sampling line by a factor of two to three, on average. The relative scatter was about a factor of two to seven. The fraction of predictions within a factor of two of observations was about 25% for CALPUFF and VLSTRACK, and 50% for HPAC. HPAC was better able to match the absolute observed maximum, although the underprediction was still about a factor of 2.5.

The model evaluation exercise also demonstrates the difficulties in simulating transport and dispersion at mesoscales in mountainous areas. (See Chang et al. [17] for more details.) The network of surface wind monitors suggests much spatial variability, for example, with root-mean-square differences in 10-m wind speeds up to 2 m/s and root-mean-square differences in 10-m wind directions up to 70 deg. The model comparisons were found to be strongly influenced by the diagnostic wind model that was used to generate mass-consistent wind fields from the observed winds. Limited sensitivity study also showed that despite of the dense network of surface wind instruments, diagnostic wind models are not always able to reproduce the spatial variability originally present in all observations when data from just one station were withheld.

Acknowledgements. This research has been sponsored by the U.S. Defense Threat Reduction Agency, with Major Brian Beitler as technical contract representative. The data files and experiment reports were provided by Christopher Biltoft of the U.S. Army’s Dugway Proving Ground.

References


Development and Application of Computer Techniques to Environmental Studies

Dispersion (OLAD) Field Experiment. NOAA Technical Memorandum, Air Resources Laboratory, Silver Spring, MD, 2000.


