Documentation of the Evaluation of CALPUFF and Other Long Range Transport Models Using Tracer Field Experiment Data

EPA-454/R-12-003
May 2012

Documentation of the Evaluation of CALPUFF and Other Long Range Transport Models Using Tracer Field Experiment Data

By:
ENVIRON International Corporation
773 San Marin Drive, Suite 2115
Novato, CA 94998

Prepared for:
Tyler Fox, Group Leader
Air Quality Modeling Group

Contract No. EP-D-07-102
Work Order No. 4
Task Order No. 06

U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
Air Quality Assessment Division
Air Quality Modeling Group
Research Triangle Park, NC

FOREWARD

This report documents the evaluation of the CALPUFF and other Long Range Transport (LRT) dispersion models using several inert tracer study field experiment data. The LRT dispersion modeling was performed primarily by the U.S. Environmental Protection Agency (EPA) during the 2008-2010 time period and builds off several previous LRT dispersion modeling studies that evaluated models using tracer study field experiments (EPA, 1986; 1998a; Irwin, 1997). The work was performed primarily by Mr. Bret Anderson while he was with EPA Region VII, EPA/OAQPS and the United States Forest Service (USFS). Mr. Roger Brode and Mr. John Irwin (retired) of the EPA Office of Air Quality Planning and Standards (OAQPS) also assisted in the LRT model evaluation. The LRT modeling results were provided to ENVIRON International Corporation who quality assured and documented the results in this report under Task 4 of Work Assignment No. 4-06 of EPA Contract EP-D-07-102. The report was prepared for the Air Quality Modeling Group (AQMG) at EPA/OAQPS that is led by Mr. Tyler Fox. Dr. Sarav Arunachalam from the University Of North Carolina (UNC) Institute for Environment was the Work Assignment Manager (WAM) for the prime contractor to EPA. The report was prepared by Ralph Morris, Kyle Heitkamp and Lynsey Parker of ENVIRON.

Numerous people provided assistance and guidance to EPA in the data collection, operation and evaluation of the LRT dispersion models. We would like to acknowledge assistance from the following people:

- AJ Deng (Penn State University) – MM5SCIPUFF
- Doug Henn and Ian Sykes (Sage) – SCIPUFF guidance
- Roland Draxler (NOAA ARL) - HYSPLIT
- Petra Siebert (University of Natural Resources – Vienna), Andreas Stohl (NILU) – FLEXPART
- Joseph Scire and Dave Strimaitis (Exponent) – CAPTEX meteorological observations and puff-splitting sensitivity tests guidance
- Mesoscale Model Interface (MMIF) Development Team:
    - EPA OAQPS; EPA Region 7, EPA Region 10; US Department of Interior (USDOI) Fish & Wildlife Service Branch of Air Quality, USDOI National Park Service Air Division and US Department of Agriculture (USDA) Forest Service Air Resources Management Program

Contents

Appendices

EXECUTIVE SUMMARY

ABSTRACT

The CALPUFF Long Range Transport (LRT) air quality dispersion modeling system is evaluated against several atmospheric tracer field experiments. Meteorological inputs for CALPUFF were generated using MM5 prognostic meteorological model processed using the CALMET diagnostic wind model with and without meteorological observations. CALPUFF meteorological inputs were also generated using the Mesoscale Model Interface (MMIF) tool that performs a direct "pass through" of the MM5 meteorological variables to CALPUFF without any adjustments or re-diagnosing of meteorological variables, as is done by CALMET. The effects of alternative options in CALMET on the CALMET meteorological model performance and the performance of the CALPUFF LRT dispersion model for simulating observed atmospheric tracer concentrations was analyzed. The performance of CALPUFF was also compared against past CALPUFF evaluation studies using an earlier version of CALPUFF and some of the same tracer test field experiments as used in this study. In addition, up to five other LRT dispersion models were also evaluated against some of the tracer field experiments. CALPUFF and the other LRT models represent three distinct types of LRT dispersion models: Gaussian puff, particle and Eulerian photochemical grid models. Numerous sensitivity tests were conducted using CALPUFF and the other LRT models to elucidate the effects of alternative meteorological inputs on dispersion model performance for the tracer field studies, as well as to intercompare the performance of the different dispersion models.

INTRODUCTION

Near-Source and Far-Field Dispersion Models

Dispersion models, such as the Industrial Source Complex Short Term (ISCST) or American Meteorological Society/Environmental Protection Agency Regulatory Model (AERMOD) typically assume steady-state, horizontally homogeneous wind fields instantaneously over the entire modeling domain and are usually limited to distances of less than 50 kilometers from a source. However, dispersion model applications of distances of hundreds of kilometers from a source require other models or modeling systems. At these distances, the transport times are sufficiently long that the mean wind fields can no longer be considered steady-state or homogeneous. As part of the Prevention of Significant Deterioration (PSD) program, new sources or proposed modifications to existing sources may be required to assess the air quality and Air Quality Related Value (AQRV) impacts at Class I and sensitive Class II areas that may be far away from the source (e.g., > 50 km). AQRVs include visibility and acid (sulfur and nitrogen) deposition. At these far downwind distances, the steady-state Gaussian plume assumptions of models like ISCST and AERMOD are likely not valid and Long Range Transport (LRT) dispersion models are required.

The Interagency Workgroup on Air Quality Modeling (IWAQM) consists of the U.S. EPA and Federal Land Managers (FLMs; i.e., NPS, USFS and FWS) and was formed to provide a focus for the development of technically sound recommendations regarding assessment of air pollutant source impacts on Federal Class I areas. One objective of the IWAQM is the recommendation of LRT dispersion models for assessing air quality and AQRVs at Class I areas. One such LRT dispersion model is the CALPUFF Gaussian puff modeling system, which includes the CALMET diagnostic wind model and the CALPOST post-processor. In 1998, EPA published a report that evaluated CALPUFF against two short-term tracer test field experiments (EPA, 1998a). Later in 1998 IWAQM released their Phase II recommendations (EPA, 1998b) that included recommendations for using the CALPUFF LRT dispersion model for addressing far-field air quality and AQRV issues at Class I areas. The IWAQM Phase II report did not recommend any specific settings for running CALMET and noted that the required expert judgment to develop a set of recommended CALMET settings would be developed over time.

In 2003, EPA issued revisions to the Guidelines on Air Quality Models (Appendix W) that recommended using the CALPUFF LRT dispersion model to address far-field (> 50 km) air quality issues associated with chemically inert compounds. The EPA Air Quality Modeling Guidelines were revised again in 2005 to include AERMOD as the EPA-recommended dispersion model for near-source (< 50 km) air quality issues.

CALPUFF Modeling Guidance

EPA convened a CALPUFF workgroup starting in 2005 to help identify issues with the 1998 IWAQM Phase II recommendations. The CALPUFF workgroup began to revisit the evaluation of CALPUFF against tracer test field experiments. In May 2009, EPA released a reassessment of the IWAQM Phase II recommendations (EPA, 2009a) that raised issues with settings used in recent CALMET model applications. CALMET is typically applied using prognostic meteorological model (i.e., MM5 or WRF) three-dimensional wind fields as an input first guess and then applying diagnostic wind effects (e.g., blocking, deflection, channeling and slope flows) to produce a STEP1 wind field. CALMET then blends in surface and upper-air meteorological observations into the STEP1 wind field using an objective analysis (OA) procedure to produce the resultant STEP2 to wind field that is provided as input into CALPUFF. CALMET also diagnoses several other meteorological variables (e.g., mixing heights). CALMET contains numerous options that can significantly affect the resultant meteorological fields. The EPA IWAQM reassessment report found that the CALMET STEP1 diagnostic effects and STEP2 OA procedures can degrade the MM5/WRF wind fields. Furthermore, the IWAQM reassessment report noted that options used in some past CALMET applications were selected based on obtaining a desired outcome rather than based on good science. Consequently, the 2009 IWAQM reassessment recommended CALMET settings that would “pass through” MM5/WRF meteorological fields as much as possible for input into CALPUFF. However, further testing of CALMET by the EPA CALPUFF workgroup found that the recommended CALMET settings in the May 2009 IWAQM reassessment report did not achieve the intended desired result to “pass through” as much as possible the MM5/WRF meteorological variables as CALMET still re-diagnosed some and modified other meteorological variables. Based in part on testing by the CALPUFF workgroup using the tracer test field experiments, on August 31, 2009 EPA released a Clarification Memorandum (EPA, 2009b) that contained specific EPA-FLM recommended settings for operating CALMET for regulatory applications.

Mesoscale Model Interface (MMIF) Tool

In the meantime, EPA has developed the Mesoscale Model Interface (MMIF) tool that will “pass through” as much as possible the MM5/WRF meteorological output to CALPUFF without modifying the meteorological fields (Emery and Brashers, 2009; Brashers and Emery 2011; 2012). The CALPUFF Workgroup has been evaluating the CALPUFF model using the CALMET and MMIF meteorological drivers for four tracer test field experiments. For some of the field experiments, additional LRT dispersion models have also been evaluated. This report documents the work performed by the CALPUFF workgroup over the 2009-2011 time frame to evaluate CALPUFF and other LRT dispersion models using four tracer test field experiment databases.

OVERVIEW OF APPROACH

Up to six LRT dispersion models were evaluated using four atmospheric tracer test field experiments.

Tracer Test Field Experiments

LRT dispersion models are evaluated using four atmospheric tracer test field studies as follows:

1980 Great Plains: The 1980 Great Plains (GP80) field study released several tracers from a site near Norman, Oklahoma in July 1980 and measured the tracers at two arcs to the northeast at distances of 100 and 600 km (Ferber et al., 1981).

1975 Savannah River Laboratory: The 1975 Savannah River Laboratory (SRL75) study released tracers from the SRL in South Carolina and measured them at receptors approximately 100 km from the release point (DOE, 1978).

1983 Cross Appalachian Tracer Experiment: The 1983 Cross Appalachian Tracer Experiment (CAPTEX) was a series of five three-hour tracer released from Dayton, OH or Sudbury, Canada during September and October, 1983. Sampling was made in a series of arcs approximately 100 km apart that spanned from 300 to 1,100 km from the Dayton, OH release site.

1994 European Tracer Experiment: The 1994 European Tracer Experiment (ETEX) consisted of two tracer releases from northwest France in October and November 1994 that was measured at 168 monitoring sites in 17 countries.

LRT Dispersion Models Evaluated

The six LRT dispersion models that were evaluated using the tracer test field study data in this study were:

CALPUFF[1]: The California Puff (CALPUFF Version 5.8; Scire et al, 2000b) model is a Lagrangian Gaussian puff model that simulates a continuous plume using overlapping circular puffs. CALPUFF was applied using both the CALMET meteorological processor (Scire et al., 2000a) that includes a diagnostic wind model (DWM) and the Mesoscale Model Interface (MMIF; Emery and Brashers, 2009; Brashers and Emery, 2011; 2012) tool that will “pass through” output from the MM5 or WRF prognostic meteorological models.

SCIPUFF[2]: The Second-order Closure Integrated PUFF (SCIPUFF Version 2.303; Sykes et al., 1998) is a Lagrangian puff dispersion model that uses Gaussian puffs to represent an arbitrary, three-dimensional time-dependent concentration field. The diffusion parameterization is based on turbulence closure theory, which gives a prediction of the dispersion rate in terms of the measurable turbulent velocity statistics of the wind field.

HYSPLIT[3]: The Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT Version 4.8; Draxler, 1997) is a complete system for computing simple air parcel trajectories to complex dispersion and deposition simulations. The dispersion of a pollutant is calculated by assuming either puff or particle dispersion. HYSPLIT was applied primarily in the default particle model where a fixed number of particles are advected about the model domain by the mean wind field and spread by a turbulent component.

1 http://www.src.com/calpuff/calpuff1.htm
2 http://www.sage-mgt.net/services/modeling-and-simulation/scipuff-dispersion-model
3 http://www.arl.noaa.gov/HYSPLIT_info.php

FLEXPART[4]: The FLEXPART (Version 6.2; Siebert, 2006; Stohl et al., 2005[5]) model is a Lagrangian particle dispersion model. FLEXPART was originally designed for calculating the long-range and mesoscale dispersion of air pollutants from point sources, such as after an accident in a nuclear power plant. In the meantime FLEXPART has evolved into a comprehensive tool for atmospheric transport modeling and analysis

CAMx[6]: The Comprehensive Air-quality Model with extensions (CAMx; ENVIRON, 2010) is a photochemical grid model (PGM) that simulates inert or chemical reactive pollutants from the local to continental scale. As a grid model, it simulates transport and dispersion using finite difference techniques on a three-dimensional array of grid cells.

CALGRID: The California Mesoscale Photochemical Grid Model (Yamartino, et al., 1989, Scire et al., 1989; Earth Tech, 2005) is a PGM that simulates chemically reactive pollutants from the local to regional scale. CALGRID was originally designed to utilize meteorological fields produced by the CALMET meteorological processor (Scire et al., 2000a), but was updated in 2006 to utilize meteorology and emissions in UAM format (Earth Tech, 2006).

The six LRT dispersion models represent two non-steady-state Gaussian puff models (CALPUFF and SCIPUFF), two three-dimensional particle dispersion models (HYSPLIT and FLEXPART) and two three-dimensional photochemical grid models (CAMx and CALGRID). HYSPLIT can also be run in a puff and hybrid particle/puff modes, which was investigated in sensitivity tests. All six LRT models were evaluated using the CAPTEX Release 3 and 5 field experiments and five of the six models (except CALGRID) were evaluated using the ETEX field experiment database.

Evaluation Methodology

Two different model performance evaluation methodologies were utilized in this study. The Irwin (1997) fitted Gaussian plume approach, as used in the EPA 1998 CALPUFF evaluation study (EPA, 1998a), was used for the same two tracer test field experiments used in the 1998 EPA study (i.e., GP80 and SRL75). This was done to elucidate how updates to CALPUFF model over the last decade have improved its performance. The second model evaluation approach adopts the spatial, temporal and global statistical evaluation framework of ATMES-II (Mosca et. al., 1998; Draxler et al., 1998). The ATMES-II uses statistical performance metrics of spatial, scatter, bias, correlation and cumulative distribution to describe model performance. An important finding of this study is that the fitted Gaussian plume model evaluation approach is very limited and can be a poor indicator of LRT dispersion model performance, with the ATMES-II approach providing a more comprehensive assessment of LRT model performance.

Fitted Gaussian Plume Evaluation Approach

The fitted Gaussian plume evaluation approach fits a Gaussian plume across the observed and predicted tracer concentrations along an arc of receptors at a specific downwind distance from the tracer release site. The approach focuses on a LRT dispersion model's ability to replicate centerline concentrations and plume widths, modeled/observed plume centerline azimuth, plume arrival time, and plume transit time across the arc. We used the fitted Gaussian plume evaluation approach to evaluate CALPUFF for the GP80 and SRL75 tracer experiments where the tracer concentrations were observed along arcs of receptors, as was done in the EPA 1998 CALPUFF evaluation study (EPA, 1998a).

4 http://transport.nilu.no/flexpart
5 http://www.atmos-chem-phys.net/5/2461/2005/acp-5-2461-2005.html
6 http://www.camx.com/

CALPUFF performance is evaluated by calculating the predicted and observed cross-wind integrated concentration (CWIC), azimuth of plume centerline, and the second moment of tracer concentration (lateral dispersion of the plume [σ_y]). The CWIC is calculated by trapezoidal integration across average monitor concentrations along the arc. By assuming a Gaussian distribution of concentrations along the arc, a fitted plume centerline concentration (Cmax) can be calculated by the following equation:

$$Cmax = CWIC/[(2\pi)^{1/2}\sigma_y]$$

The measure σ_y describes the extent of plume horizontal dispersion. This is important to understanding differences between the various dispersion options available in the CALPUFF modeling system. Additional measures for temporal analysis include plume arrival time and the plume transit time on arc. Table ES-1 summarizes the spatial, temporal and concentration statistical performance metrics used in the fitted Gaussian plume evaluation methodology.

Table ES-1. Model performance metrics used in the fitted Gaussian plume evaluation methodology from Irwin (1997) and 1998 EPA CALPUFF Evaluation (EPA, 1998a).

Statistics	Description
Spatial	
Azimuth of Plume Centerline	Comparison of the predicted angular displacement of the plume centerline from the observed plume centerline on the arc
Plume Sigma-y	Comparison of the predicted and observed fitted plume widths (i.e., dispersion rate)
Temporal	
Plume Arrival Time	Compare the time the predicted and observed tracer clouds arrives on the receptor arc
Transit Time on Arc	Compare the predicted and observed residence time on the receptor arc
Concentration	
Crosswind Integrated Concentration	Compares the predicted and observed average concentrations across the receptor arc
Observed/Calculated Maximum	Comparison of the predicted and observed fitted Gaussian plume centerline (maximum) concentrations (Cmax) and maximum concentration at any receptor along the arc (Omax)

<u>Spatial, Temporal and Global Statistics Evaluation Approach</u>

The model evaluation methodology as employed in ATMES-II (Mosca et al., 1998) and recommended by Draxler et al., (2002) was also used in this study. This approach defines three types of statistical analyses:

- <u>Spatial Analysis</u>: Concentrations at a fixed time are considered over the entire domain. Useful for determining differences spatial differences between predicted and observed concentrations.
- <u>Temporal Analysis</u>: Concentrations at a fixed location are considered for the entire analysis period. This can be useful for determining differences between the timing of predicted and observed tracer concentrations.
- <u>Global Analysis</u>: All concentration values at any time and location are considered in this analysis. The global analysis considers the distribution of the values (probability), overall tendency towards overestimation or underestimation of measured values (bias and error), measures of scatter in the predicted and observed concentrations and measures of correlation.

Table ES-2 defines the twelve ATMES-II spatial and global statistical metrics used in this study, some of the temporal statistics were also calculated but not reported. The RANK model performance statistic is designed to provide an overall score of model performance by combining performance metrics of correlation/scatter (R^2), bias (FB), spatial (FMS) and cumulative distribution (KS). Its use as an overall indication of the rankings of model performance for different models was evaluated and found that it usually was a good indication, but there were some cases where it could lead to misleading results and is not a substitute for examining all performance attributes.

Table ES-2. ATMES-II spatial and global statistical metrics.

Statistical Metric	Definition	Perfect Score
Spatial Statistics		
Figure of Merit in Space (FMS)	$FMS = \frac{A_M \cap A_P}{A_M \cup A_P} \times 100\%$	100%
False Alarm Rate (FAR)	$FAR = \left(\frac{a}{a+b} \right) \times 100\%$	0%
Probability of Detection (POD)	$POD = \left(\frac{b}{b+d} \right) \times 100\%$	100%
Threat Score (TS)	$TS = \left(\frac{b}{a+b+d} \right) \times 100\%$	100%
where,	• "a" represents the number of times a condition that has been forecast, but was not observed (false alarm) • "b" represents the number of times the condition was correctly forecasted (hits) • "c" represents the number of times the nonoccurrence of the condition is correctly forecasted (correct negative); and • "d" represents the number of times that the condition was observed but not forecasted (miss).	
Global Statistics		
Factor of Exceedance (FOEX)	$FOEX = \left[\frac{N_{(P_i > Ni_i)}}{N} - 0.5 \right] \times 100\%$	0%
Factor of α (FA2 and FA5)	$FA\alpha = \left[\frac{N(y - y0 = [x - x0]\alpha)}{N} \right] \times 100$	100%
Normalized Mean Squared Error (NMSE)	$NMSE = \frac{1}{N\overline{PM}} \sum (P_i - M_i)^2$	0%
Pearson's Correlation Coefficient (PCC or R)	$R = \frac{\sum_i (M_i - \overline{M}) \bullet (P_i - \overline{P})}{\left[\sqrt{\sum (M_i - \overline{M})^2} \right] \left[\sqrt{\sum (P_i - \overline{P})^2} \right]}$	1.0
Fraction Bias (FB)	$FB = 2\overline{B} / (\overline{P} + \overline{M})$	0%
Kolmogorov-Smirnov (KS) Parameter	$KS = Max\lvert C(M_k) - C(P_k) \rvert$	0%
RANK	$RANK = \lvert R^2 \rvert + (1 - \lvert FB/2 \rvert) + FMS/100 + (1 - KS/100)$	4.0

MODEL PERFORMANCE EVALUATION OF LRT DISPERSION MODELS

The CALPUFF LRT dispersion model was evaluated using four tracer test field study experiments. Up to five additional LRT models were also evaluated using some of the field experiments.

1980 Great Plains (GP80) Field Experiment

The CALPUFF LRT dispersion model was evaluated against the GP80 July 8, 1980 GP80 tracer release from Norman, Oklahoma. The tracer was measured at two receptor arcs located 100 km and 600 km downwind from the tracer release point. The fitted Gaussian plume approach was used to evaluate the CALPUFF model performance, which was the same approach used in the EPA 1998 CALPUFF evaluation study (EPA, 1998a). CALPUFF was evaluated separately for the 100 km and 600 km arc of receptors.

GP80 CALPUFF Sensitivity Tests

Several different configurations of CALMET and CALPUFF models were used in the evaluation that varied CALMET grid resolution, grid resolution of the MM5 meteorological model used as input to CALMET, and CALMET and CALPUFF model options, including:

- CALMET grid resolution of 4 and 10 km for 100 km and 4 and 20 km for 600 km receptor arc.
- MM5 output grid resolution of 12, 36 and 80 km, plus no MM5 data.
- Use of surface and upper-air meteorological data used as input to CALMET:
  - A = Use surface and upper-air observations;
  - B = Use surface but not upper-air observations; and
  - C = Use no meteorological observations.
- Three CALPUFF dispersion algorithms:
  - CAL = CALPUFF turbulence dispersion;
  - AER = AERMOD turbulence dispersion; and
  - PG = Pasquill-Gifford dispersion.
- MMIF meteorological inputs for CALPUFF using 12 and 36 km MM5 data.

The "BASEA" CALPUFF/CALMET configuration was designed to emulate the configuration used in the 1998 EPA CALPUFF evaluation study, which used only meteorological observations and no MM5 data in the CALMET modeling and ran the CALPUFF CAL and PG dispersion options. However, an investigation of the 1998 EPA evaluation study revealed that the slug near-field option was used in CALPUFF (MSLUG = 1). The slug option is designed to better simulate a continuous plume near the source and is a very non-standard option for CALPUFF LRT dispersion modeling. For the initial CALPUFF simulations, the slug option was used for the 100 km receptor arc, but not for the 600 km receptor arc. However, additional CALPUFF sensitivity tests were performed for the 600 km receptor arc that investigated the use of the slug option, as well as alternative puff splitting options.

Conclusions of GP80 CALPUFF Model Performance Results

For the 100 km receptor arc, there was a wide variation in CALPUFF model performance across the sensitivity tests. The results were consistent with the 1998 EPA study with the following key findings for the GP80 100 km receptor arc evaluation:

- CALPUFF tended to overstate the maximum observed concentrations and understate the plume widths at the 100 km receptor arc.
- The best performing CALPUFF configuration in terms of predicting the maximum observed concentrations and plume width was when CALMET was run with MM5 data and surface meteorological observations but no upper-air meteorological observations.
- The CALPUFF CAL and AER turbulence dispersion options produced nearly identical results and the performance of the CAL/AER turbulence versus PG dispersion options varied by model configuration and statistical performance metric.
- The performance of CALPUFF/MMIF in predicting plume maximum concentrations and plume widths was comparable or better than all of the CALPUFF/CALMET configurations, except when CALMET used MM5 data and surface but no upper-air meteorological observations.
- The modeled plume centerline tended to be offset from the observed centerline location by 0 to 14 degrees.
- Use of CALMET with just surface and upper-air meteorological observations produced the best CALPUFF plume centerline location performance, whereas use of just MM5 data with no meteorological observations, either through CALMET or MMIF, produced the worst plume centerline angular offset performance.
- Different CALMET configurations give the best CALPUFF performance for maximum observed concentration (with MM5 and just surface and no upper-air observations) versus location of the plume centerline (no MM5 and both surface and upper-air observations) along the 100 km receptor arc. For Class I area LRT dispersion modeling it is important for the model to estimate both the location and the magnitudes of concentrations.

The evaluation of the CALPUFF sensitivity tests for the 600 km arc of receptors included both plume arrival, departure and residence time analysis as well as fitted Gaussian plume statistics. The observed residence time of the tracer on the 600 km receptor arc was at least 12 hours. Note that due to the presence of an unexpected low-level jet, the tracer was observed at the 600 km receptor arc for the first sampling period. Thus, the observed 12 hour residence time is a lower bound (i.e., the observed tracer could have arrived before the first sampling period). The 1998 EPA CALPUFF evaluation study estimated tracer plume residence times of 14 and 13 hours, which compares favorably with the observed residence time (12 hours). However, the 1998 EPA study CALPUFF modeling had the tracer arriving at least 1 hour later and leaving 2-3 hours later than observed, probably due to the inability of CALMET to simulate the low-level jet.

Most (~90%) of the current study CALPUFF sensitivity tests underestimated the observed tracer residence time on the 600 km receptor arc by approximately a factor of two. The exception to this was: (1) the BASEA_PG CALPUFF/CALMET sensitivity test (12 hours) that used just meteorological observations in CALMET and the PG dispersion option in CALPUFF; and (2) the CALPUFF/CALMET EXP2C series of experiments (residence time of 11-13 hours) that used 36 km MM5 data and CALMET run at 4 km resolution with no meteorological observations (NOOBS = 2). The remainder of the 28 CALPUFF sensitivity tests had tracer residence time on the 600 km receptor arc of 4-8 hours; that is, almost 90% of the CALPUFF sensitivity tests failed to reproduce the good tracer residence time performance statistics from the 1998 EPA study.

For the 600 km receptor arc, the CALPUFF sensitivity test fitted Gaussian plume statistics were very different than the 100 km receptor arc as follows:

- The maximum observed concentration along the arc or observed fitted centerline plume concentration was underestimated by -42% to -72% and the plume widths overestimated by 47% to 293%.
- The CALPUFF underestimation bias of the observed maximum concentration tends to be improved using CALMET runs with no meteorological observations.
- The use of the PG dispersion option tends to exacerbate the plume width overestimation bias relative to using the CAL or AER turbulence dispersion option.
- The CALPUFF predicted plume centerline tends to be offset from the observed value by 9 to 20 degrees, with the largest centerline offset (> 15 degrees) occurring when no meteorological observations are used with either CALMET or MMIF .
- The 1998 CALPUFF runs overestimated the observed CWIC by 15% and 30% but the current study's BASEA configuration, which was designed to emulate the 1998 EPA study, underestimates the observed CWIC by -14% and -38%.

The inability of most (~90%) of the current study's CALPUFF sensitivity tests to reproduce the 1998 EPA study tracer test residence time on the 600 km receptor arc is a cause for concern. For example, the 1998 EPA study CALPUFF simulation using the CAL dispersion option estimates a tracer residence time on the 600 km receptor arc of 13 hours that compares favorably to what was observed (12 hours). However, the current study CALPUFF BASEA_CAL configuration, which was designed to emulate the 1998 EPA CALPUFF configuration, estimates a residence time of almost half of the 1998 EPA study (7 hours). One notable difference between the 1998 EPA and the current study CALPUFF modeling for the GP80 600 km receptor arc was the use of the slug option in the 1998 EPA study. Another notable difference was the ability of the current version of CALPUFF to perform puff splitting, which EPA has reported likely extends the downwind distance applicability of the CALPUFF model (EPA, 2003). Thus, a series of CALPUFF sensitivity tests were conducted using the BASEA_CAL CALPUFF/CALMET and MMIF_12KM CAL and PG CALPUFF/MMIF configurations that invoked the slug option and performed puff splitting. Two types of puff splitting were analyzed, default puff splitting (DPS) that turns on the vertical puff splitting flag once per day and all hours puff splitting (APS) that turns on the puff splitting flag for every hour of the day. The following are the key findings from the CALPUFF slug and puff splitting sensitivity tests for the GP80 600 km receptor arc:

- Use of puff splitting had no effect on the tracer test residence time (7 hours) in the CALPUFF/CALMET (BASEA_CAL) configuration.
- Use of the slug option with CALPUFF/CALMET increased the tracer residence time on the 600 km receptor arc from 7 to 15 hours, suggesting that the better performance of the 1998 EPA CALPUFF simulations on the 600 km receptor arc was due to invoking the slug option.
- On the other hand, the CALPUFF/MMIF sensitivity tests were more sensitivity to puff splitting than CALPUFF/CALMET with the tracer residence time increasing from 6 to 8 hours using DPS and to 17 hours using APS when the CAL dispersion option was specified.
- The use of the slug option on top of APS has very different effect on the CALPUFF/MMIF residence time along the 600 km receptor depending on which dispersion option is utilized, with slug reducing the residence time from 17 to 15 hours using the CAL and increasing the residence time from 11 to 20 hours using PG dispersion options.

- The best performing CALPUFF configuration from all of the sensitivity tests when looking at the performance across all of the fitted plume performance statistics was use of the slug option with puff splitting in CALPUFF/MMIF.

A key result of the GP80 600 km receptor arc evaluation was the need to invoke the near-source slug option to adequately reproduce the CALPUFF performance from the 1998 EPA CALPUFF evaluation study. Given that the slug option is a very nonstandard option for LRT dispersion modeling, this finding raises concern regarding the previous CALPUFF evaluation. Another important finding of the GP80 CALPUFF sensitivity tests is the wide variation in modeling results that can be obtained using the various options in CALMET and CALPUFF. This is not a desirable attribute for regulatory modeling and emphasizes the need for a standardized set of options for regulatory CALPUFF modeling.

1975 Savannah River Laboratory (SRL75) Field Experiment

The 1975 Savannah River Laboratory (SRL75) field experiment released a tracer on December 10, 1975 and measured it at receptors located approximately 100 km downwind from the tracer release site. The fitted Gaussian plume model evaluation approach was used to evaluate numerous CALPUFF sensitivity tests. Several CALMET sensitivity tests were run to provide meteorological inputs to CALPUFF that varied whether MM5 data was used or not and how meteorological observations were used (surface and upper-air, surface only or no observations). As in the GP80 sensitivity tests, three dispersion options were used in CALPUFF (CAL, AER and PG). In addition, CALPUFF/MMIF sensitivity tests were performed using MM5 output at 36, 12 and 4 km resolution.

Because of the long time integrated sampling period used in the SRL75 experiment, the plume arrival, departure and residence statistics were not available and only the fitted Gaussian plume statistics along the 100 km receptor arc were used in the evaluation. The key findings of the SRL75 CALPUFF evaluation are as follows:

- The maximum plume centerline concentrations from the fitted Gaussian plume to the observed tracer concentrations is approximately half the maximum observed tracer concentration at any monitor along the 100 km receptor arc. As a plume centerline concentration in a Gaussian plume represents the maximum concentration, this indicates that the fitted Gaussian plume is a very poor fit to the observations. Thus, the plume centerline and plume width statistics that depend on the fitted Gaussian plume are a poor indication of model performance for the SRL75 experiment. The observed fitted Gaussian plume statistics were taken from the 1998 EPA study (EPA, 1998a).
- Given that there are many more (~5 times) CALPUFF receptors along the 100 km receptor arc than monitoring sites where the tracer was observed, the predicted maximum concentration along the arc is expected to be greater than the observed maximum concentration. Such is the case with the CALPUFF/MMIF runs, but is not always the case for the CALMET/CALPUFF sensitivity tests using no MM5 data.
- The CALPUFF plume centerline is offset from the observed plume centerline by 8 to 20 degrees. The largest angular offset occurs (17-20 degrees) when CALMET is run with no MM5 data. When MM5 data is used with the surface and upper-air observations the CALPUFF angular offset is essentially unchanged (18-19 degrees) and the removal of the upper-air observations also has little effect on the plume centerline angular offset. However, when only MM5 data are used, in either in CALMET (11-12 degrees) or MMIF (9-10 degrees), the CALPUFF plume centerline offset is improved.

The main conclusion of the SRL75 CALPUFF evaluation is that the fitted Gaussian plume evaluation approach can be a poor and misleading indicator of LRT dispersion model performance. In fact, the whole concept of a well-defined Gaussian plume at far downwind distances (e.g., > 50 km) is questionable since wind variations and shear can destroy the Gaussian distribution. Thus, we recommend that future studies no longer use the fitted Gaussian plume evaluation methodology for evaluating LRT dispersion models and adopt alternate evaluation approaches that are free from a priori assumption regarding the distribution of the observed tracer concentrations.

Cross Appalachian Tracer Experiment (CAPTEX)

The Cross Appalachian Tracer Experiment (CAPTEX) performed five tracer releases from either Dayton, Ohio or Sudbury, Ontario with tracer concentrations measured at hundreds of monitoring sites deployed in the northeastern U.S. and southeastern Canada out to distances of 1000 km downwind of the release sites. Numerous CALPUFF sensitivity tests were performed for the third (CTEX3) and fifth (CTEX5) CAPTEX tracer releases from, respectively, Dayton and Sudbury. The performance of the six LRT models was also intercompared using the CTEX3 and CTEX5 field experiments.

<u>CAPTEX Meteorological Modeling</u>

MM5 meteorological modeling was conducted for the CTEX3 and CTEX5 periods using modeling approaches prevalent in the 1980's (e.g., one 80 km grid with 16 vertical layers) that was sequentially updated to use a more current MM5 modeling approach (e.g., 108/36/12/4 km nested grids with 43 vertical layers). The MM5 experiments also employed various levels of four dimensional data assimilation (FDDA) from none (i.e., forecast mode) to increasing aggressive use of FDDA.

CALMET sensitivity tests were conducted using 80, 36 and 12 km MM5 data as input and using CALMET grid resolutions of 18, 12 and 4 km. For each MM5 and CALMET grid resolution combination, additional CALMET sensitivity tests were performed to investigate the effects of different options for blending the meteorological observations into the CALMET STEP1 wind fields using the STEP2 objective analysis (OA) procedures to produce the wind field that is provided as input to CALPUFF:

- A – RMAX1/RMAX2 = 500/1000
- B – RMAX1/RMAX2 = 100/200
- C – RMAX1/RMAX2 = 10/100
- D – no meteorological observations (NOOBS = 2)

Wind fields estimated by the MM5 and CALMET CTEX3 and CTEX5 sensitivity tests were paired with surface wind observations in space and time, then aggregated by day and then aggregated over the modeling period. The surface wind comparison is not an independent evaluation since many of the surface wind observations in the evaluation database are also provided as input to CALMET. Since the CALMET STEP2 OA procedure is designed to make the CALMET winds at the monitoring sites better match the observed values, one would expect CALMET simulations using observations to perform better than those that do not. However, as EPA points out in their 2009 IWAQM reassessment report, CALMET's OA procedure can also produce discontinuities and artifacts in the wind fields resulting in a degradation of the wind fields even though they may match the observed winds better at the locations of the observations (EPA, 2009a). The key findings from the CTEX5 MM5 and CALMET meteorological evaluation are as follows:

- The MM5 wind speed, and especially wind direction, model performance is better when FDDA is used then when FDDA is not used.
- The "A" and "B" series of CALMET simulations produce wind fields least similar to the MM5 simulation used as input, which is not surprising since CALMET by design is modifying the winds at the location of the monitoring sites to better match the observations.
- CALMET tends to slow down the MM5 wind speeds even when there are no wind observations used as input (i.e., the "D" series).
- For this period and MM5 model configuration, the MM5 and CALMET wind model performance is better when 12 km grid resolution is used compared to coarser resolution.

<u>CAPTEX CALPUFF Model Evaluation and Sensitivity Tests</u>

The CALPUFF model was evaluated against tracer observations from the CTEX3 and CTEX5 field experiments using meteorological inputs from the various CALMET sensitivity tests described above as well as the MMIF tool applied using the 80, 36 and 12 km MM5 databases. The CALPUFF configuration was held fixed in all of these sensitivity tests so that the effects of the meteorological inputs on the CALPUFF tracer model performance could be clearly assessed. The CALPUFF default model options were assumed for most CALPUFF inputs. One exception was for puff splitting where more aggressive vertical puff splitting was allowed to occur throughout the day, rather than the default where vertical puff splitting is only allowed to occur once per day.

The ATMES-II statistical model evaluation approach was used to evaluate CALPUFF for the CAPTEX field experiments. Twelve separate statistical performance metrics were used to evaluate various aspects of the CALPUFF's ability to reproduce the observed tracer concentrations in the two CAPTEX experiments. Below we present the results of the RANK performance statistic that is a composite statistic that represents four aspects of model performance: correlation, bias, spatial and cumulative distribution. Our analysis of all twelve ATMES-II statistics has found that the RANK statistic usually provides a reasonable assessment of the overall performance of dispersion models tracer test evaluations. However, we have also found situations where the RANK statistic can provide misleading indications of the performance of dispersion models and recommend that all model performance attributes be examined to confirm that the RANK metric is providing a valid ranking of the dispersion model performance.

CTEX3 CALPUFF Model Evaluation

Figure ES-1 summarizes the RANK model performance statistics for the CALPUFF sensitivity simulations that used the 12 km MM5 data as input. Using a 4 km CALMET grid resolution, the EXP6B (RMAX1/RMAX2 = 100/200) has the lowest rank of the CALPUFF/CALMET sensitivity tests. Of the CALPUFF sensitivity tests using the 12 km MM5 data as input, the CALPUFF/MMIF (12KM_MMIF) sensitivity test has the highest RANK statistic (1.43) followed closely by EXP4A (1.40; 12 km CALMET and 500/1000), EXP6C (1.38; 4 km CALMET and 10/500) with the lowest

RANK statistic (1.22) exhibited by EXP4B (12 km CALMET and 100/200) and EXP6B (4 km CALMET and 100/200).

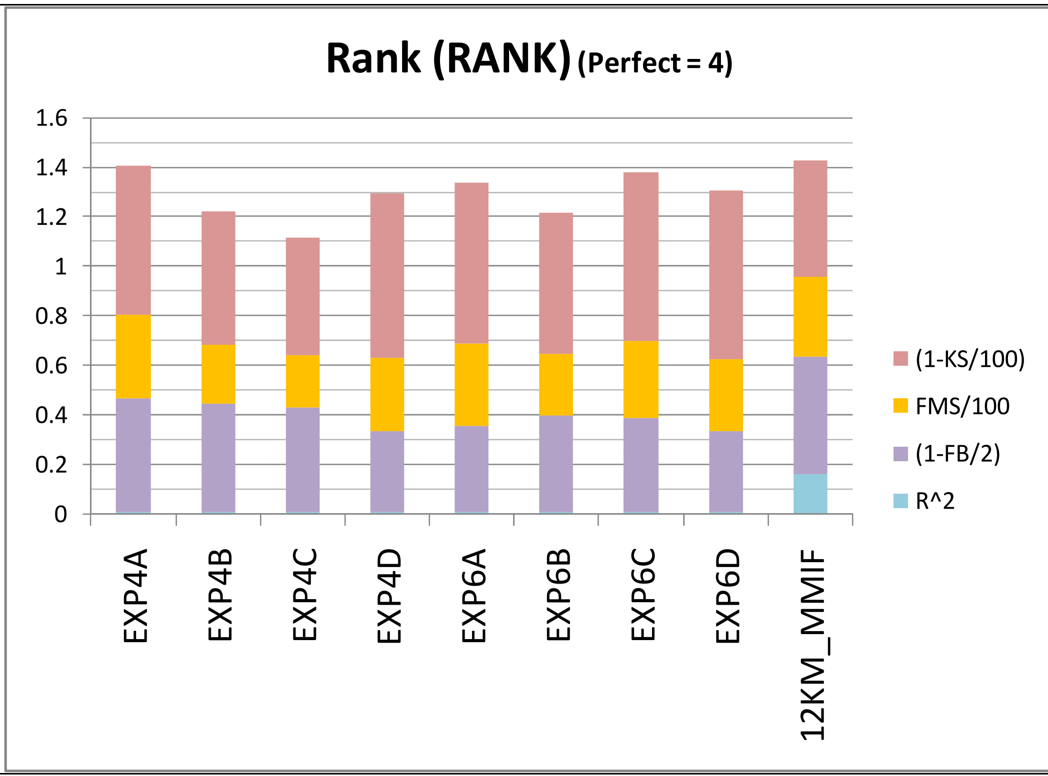


Figure ES-1. RANK performance statistics for CTEX3 CALPUFF sensitivity tests that used 12 km MM5 as input to CALMET or MMIF.

Figure ES-2 compares the RANK model performance statistics for “B” (RMAX1/RMAX2 = 100/200) and “D” (no observations) series of CALPUFF/CALMET sensitivity tests using different CALMET/MM5 grid resolutions of 18/80 (BASEB), 12/80 (EXP1), 12/36 (EXP3), 12/12 (EXP4) 4/36 (EXP5) and 4/12 (EXP6) along with the CALPUFF/MMIF runs using 36 and 12 km MM5 data. The CALPUFF/CALMET sensitivity tests using no observations (“D” series) generally have a higher rank metric than when meteorological observations are used with CALPUFF (“B” series). The CALMET/MMIF sensitivity test using 36 and 12 km MM5 data are the configurations with the highest RANK metric.The CALPUFF/MMIF show a strong relationship between observed and predicted winds than the CALPUFF/CALMET sensitivity tests, which had no to slightly negative correlations with the tracer observations.

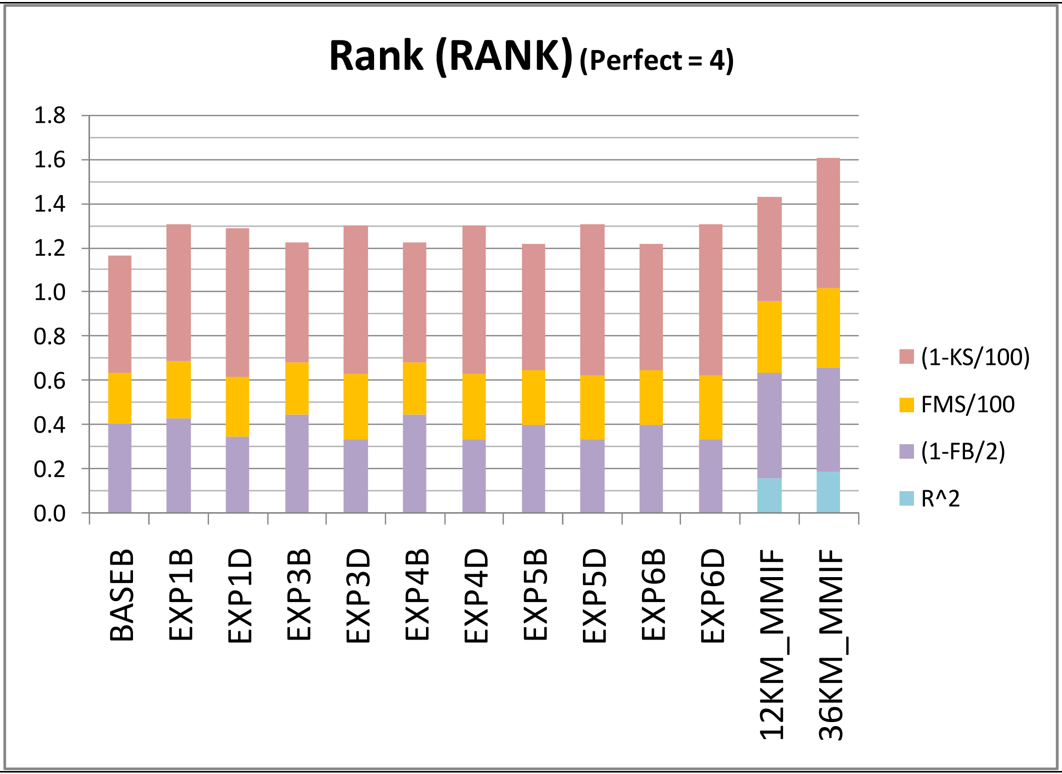


Figure ES-2. RANK performance statistics for CTEX3 CALPUFF sensitivity tests that used RMAX1/RMAX2 = 100/200 ("B" series) or no observations in CALMET ("D" series) and various CALMET/MM5 grid resolutions plus CALMET/MMIF using 36 and 12 km MM5 data.

Table ES-3 ranks all of the CALPUFF CTEX3 sensitivity tests using the RANK statistics. It is interesting to note that the EXP3A and EXP4A CALPUFF/CALMET sensitivity test that uses the, respectively, 36 km and 12 km MM5 data with 12 km CALMET grid resolution and RMAX1/RMAX2 values of 500/1000 have a rank metric that is third highest, but the same model configuration with alternative RMAX1/RMAX2 values of 10/100 (EXP3C and EXP4C) degrades the model performance of the CALPUFF configuration according to the RANK statistic, with a RANK value of 1.12. This is largely due to decreases in the FMS and KS metrics.

Note that the finding that CALPUFF/CALMET model performance using CALMET wind fields based on setting RMAX1/RMAX2 = 100/200 (i.e., the "B" series) produces worse CALPUFF model performance for simulating the observed atmospheric tracer concentrations is in contrast to the CALMET surface wind field comparison that found the "B" series most closely matched observations at surface meteorological stations. Since the CALPUFF tracer evaluation is an independent evaluation of the CALMET/CALPUFF modeling system, whereas the CALMET surface wind evaluation is not, the CALPUFF tracer evaluation may be a better indication of the best performing CALMET configuration. The CALMET "B" series approach for blending the wind observations in the wind fields may just be the best approach for getting the CALMET winds to match the observations at the monitoring sites, but possibly at the expense of degrading the wind fields away from the monitoring sites resulting in worse overall depiction of transport conditions.

Table ES-3. Final Rankings of CALPUFF CTEX3 Sensitivity Tests using the RANK model performance statistics.

Ranking	Sensitivity Test	RANK Statistics	MM5 (km)	CALGRID (km)	RMAX1/RMAX2	Met Obs
1	36KM_MMIF	1.610	36	--	--	--
2	12KM_MMIF	1.430	12	--	--	--
3	EXP3A	1.400	36	12	500/1000	Yes
4	EXP4A	1.400	12	12	500/1000	Yes
5	EXP5C	1.380	36	4	10/100	Yes
6	EXP6C	1.380	12	4	10/100	Yes
7	EXP1C	1.340	36	18	10/100	Yes
8	EXP5A	1.340	36	4	500/1000	Yes
9	EXP6A	1.340	12	4	500/1000	Yes
10	EXP5D	1.310	36	4	--	No
11	EXP6D	1.310	12	4	--	No
12	EXP1B	1.300	36	18	100/200	Yes
13	EXP3D	1.300	36	12	--	No
14	EXP4D	1.300	12	12	--	No
15	BASEA	1.290	80	18	500/1000	Yes
16	EXP1D	1.290	36	18	--	No
17	EXP1A	1.280	36	18	500/1000	Yes
18	EXP3B	1.220	36	12	100/200	Yes
19	EXP5B	1.220	36	4	100/200	Yes
20	EXP4B	1.220	12	12	100/200	Yes
21	EXP6B	1.220	12	4	100/200	Yes
22	BASEC	1.170	80	18	10/100	Yes
23	BASEB	1.160	80	18	100/200	Yes
24	EXP3C	1.120	36	12	10/100	Yes
25	EXP4C	1.120	12	12	10/200	Yes

CTEX5 CALPUFF Model Evaluation

Figure ES-3 summarizes the RANK model performance statistics for the CTEX5 CALPUFF sensitivity simulations that used the 12 km MM5 data as input to CALMET and the 12 and 4 km MM5 data as input to MMIF.

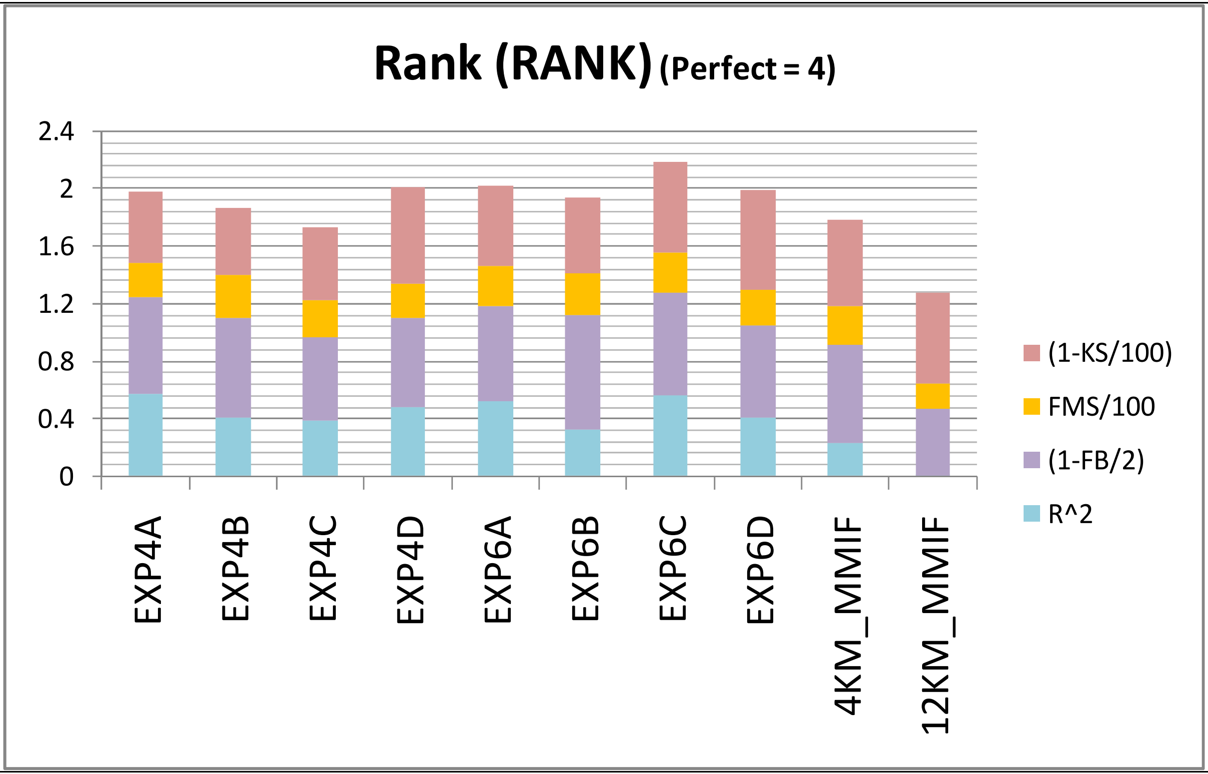


Figure ES-3. RANK performance statistics for CTEX5 CALPUFF sensitivity tests that used 12 km MM5 as input to CALMET or MMIF.

Table ES-4 ranks the model performance of the CTEX5 CALPUFF sensitivity tests using the RANK composite statistic. The 12, 36 and 80 km CALPUFF/MMIF sensitivity tests have the lowest RANK values in the 1.28 to 1.42 range.

Table ES-4. Final Rankings of CALPUFF CTEX5 Sensitivity Tests using the RANK model performance statistic.

Ranking	Sensitivity Test	RANK Statistics	MM5 (km)	CALGRID (km)	RMAX1/RMAX2	Met Obs
1	EXP6C	2.19	12	4	10/100	Yes
2	EXP5D	2.10	36	4	--	No
3	BASEA	2.06	80	18	500/1000	Yes
4	BASEC	2.05	80	18	10/100	Yes
5	EXP5A	2.03	36	4	500/1000	Yes
6	EXP6A	2.02	12	4	500/1000	Yes
7	EXP4D	2.00	12	12	--	No
8	EXP6D	1.99	12	4	--	No
9	EXP4A	1.98	12	12	500/1000	Yes
10	EXP6B	1.94	12	4	100/200	Yes
11	EXP5B	1.89	36	4	100/200	Yes
12	EXP4B	1.86	12	12	100/200	Yes
13	BASEB	1.82	80	18	100/200	Yes
14	EXP5C	1.80	36	4	10/100	Yes

15	BASED	1.79	80	18	--	No
16	EXP3A	1.79	36	12	10/100	Yes
17	EXP3B	1.79	36	12	100/200	Yes
18	EXP3C	1.79	36	12	500/1000	Yes
19	EXP3D	1.79	36	12	--	No
20	4KM_MMIF	1.78	4	--	--	No
21	EXP4C	1.72	12	12	10/100	Yes
22	36KM_MMIF	1.42	36	--	--	No
23	80KM_MMIF	1.42	80	--	--	No
24	12KM_MMIF	1.28	12	--	--	No

Conclusions of the CAPTEX CALPUFF Tracer Sensitivity Tests

There are some differences and similarities in CALPUFF's ability to simulate the observed tracer concentrations in the CTEX3 and CTEX5 field experiments. The overall conclusions of the evaluation of the CALPUFF model using the CAPTEX tracer test field experiment data can be summarized as follows:

- There is a noticeable variability in the CALPUFF model performance depending on the selected input options to CALMET.
    - By varying CALMET inputs and options through their range of plausibility, CALPUFF can produce a wide range of concentrations estimates.
- Regarding the effects of the RMAX1/RMAX2 parameters on CALPUFF/CALMET model performance, the "A" series (500/1000) performed best for CTEX3 but the "C" series (10/100) performed best for CTEX5 with both CTEX3 and CTEX5 agreeing that the "B" series (100/200) is the worst performing setting for RMAX1/RMAX2.
    - This is in contrast to the CALMET wind evaluation that found the "B" series was the CALMET configuration that most closely matched observed surface winds.
    - The CALMET wind evaluation was not an independent evaluation since some of the wind observations used in the model evaluation database were also used as input to CALMET.

Evaluation of Six LRT Dispersion Models using the CTEX3 Database

Six LRT dispersions models were applied for the CTEX3 experiment using common meteorological inputs based solely on MM5. Figure ES-4 displays the RANK model performance statistic for the six LRT dispersion models. The RANK statistical performance metric was proposed by Draxler (2001) as a single model performance metric that equally ranks the combination of performance metrics for correlation (PCC or R^2), bias (FB), spatial analysis (FMS) and unpaired distribution comparisons (KS). The RANK metrics ranges from 0.0 to 4.0 with a perfect model receiving a score of 4.0.

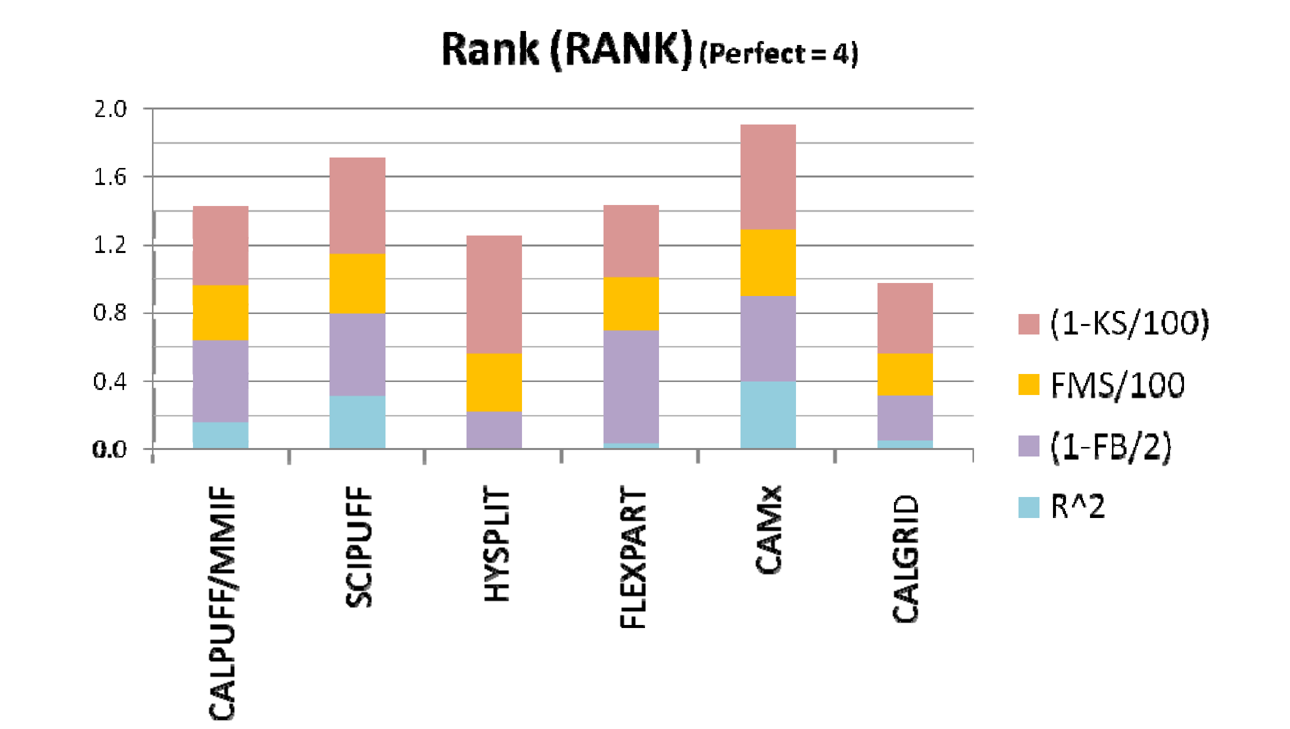


Figure ES-4. RANK statistical performance metric for the six LRT models and CAPTEX Release 3.

Table ES-5 summarizes the rankings between the six LRT models for the 11 performance statistics analyzed and compares them to the rankings obtained using the RANK performance statistic. In testing the efficacy of the RANK statistic for providing an overall ranking of model performance the ranking of the six LRT models using the average rank of the 11 performance statistics (Table ES-5) versus the ranking from the RANK statistical metric (Figure ES-4) are compared as follows:

Ranking	Average of 11 Statistics	RANK
1.	CAMx	CAMx
2.	SCIPUFF	SCIPUFF
3.	FLEXPART	FLEXPART
4.	HYSPLIT	CALPUFF
5.	CALPUFF	HYSPLIT
6.	CALGRID	CALGRID

For the CTEX3 experiment, the average rankings across the 11 statistics is nearly identical to the rankings produced by the RANK integrated statistic that combines the four of the statistics for correlation (PCC), bias (FB), spatial (FMS) and cumulative distribution (KS) with only HYSPLIT and CALPUFF exchanging places. This switch was due to CALPUFF having lower scores in the FA2 and FA5 metrics compared to HYSPLIT. If not for this, the average rank across all 11 metrics would have been the same as Draxler's RANK score. However, the analyst should use discretion in relying too heavily upon RANK score without consideration to which performance metrics are important measures for the particular evaluation goals. For example, if performance goals are not concerned with a model's ability to perform well in space and time, then reliance upon spatial statistics, such as the FMS, in the composite RANK value may not be appropriate.

Table ES-5. Summary of model ranking for the CTEX3 using the ATMES-II statistical performance metrics and comparing their average rankings to the RANK metric.

Statistic	1st	2nd	3rd	4th	5th	6th
FMS	CAMx	SCIPUFF	HYSPLIT	CALPUFF	FLEXPART	CALGRID
FAR	FLEXPART	CAMx	SCIPUFF	CALPUFF	HYSPLIT	CALGRID
POD	CAMx	FLEXPART	SCIPUFF	HYSPLIT	CALPUFF	CALGRID
TS	FLEXPART	CAMx	SCIPUFF	HYSPLIT	CALPUFF	CALGRID
FOEX	HYSPLIT	CAMx	SCIPUFF	CALPUFF	CALGRID	FLEXPART
FA2	CAMx	SCIPUFF	FLEXPART	HYSPLIT	CALGRID	CALPUFF
FA5	CAMx	SCIPUFF	FLEXPART	HYSPLIT	CALPUFF	CALGRID
NMSE	FLEXPART	CAMx	CALPUFF	SCIPUFF	CALGRID	HYSPLIT
PCC or R	CAMx	SCIPUFF	CALPUFF	CALGRID	FLEXPART	HYSPLIT
FB	FLEXPART	CAMx	SCIPUFF	CALPUFF	CALGRID	HYSPLIT
KS	HYSPLIT	CAMx	SCIPUFF	CALPUFF	FLEXPART	CALGRID
Avg. Ranking	CAMx	SCIPUFF	FLEXPART	HYSPLIT	CALPUFF	CALGRID
Avg. Score	1.55	2.72	3.0	4.0	4.27	5.55
RANK Ranking	CAMx	SCIPUFF	FLEXPART	CALPUFF	HYSPLIT	CALGRID
RANK	1.91	1.71	1.44	1.43	1.25	0.98

Evaluation of Six LRT Dispersion Models using the CTEX5 Database

Figure ES-5 displays the RANK model performance statistics for the six LRT models and the CTEX5 field experiment.

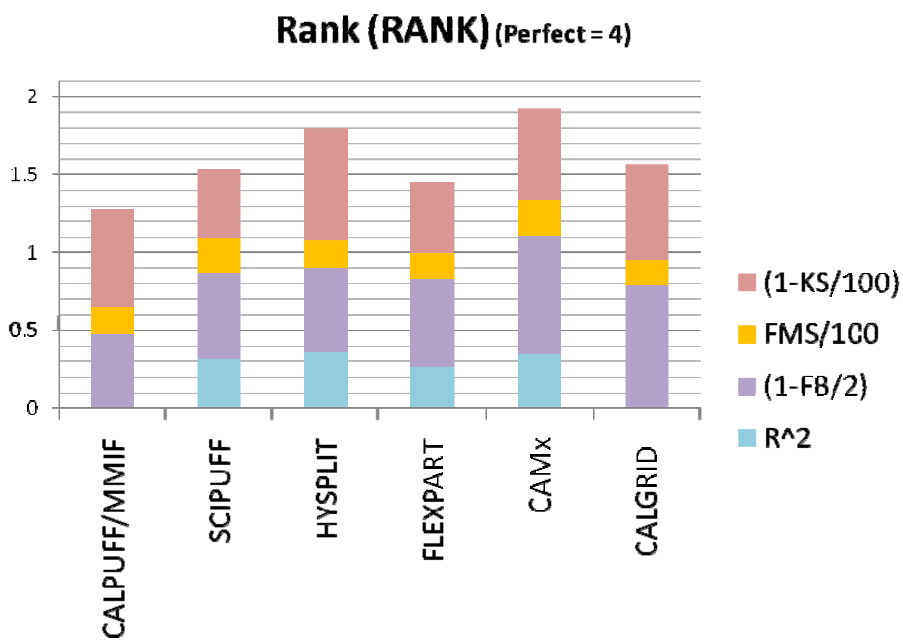


Figure ES-5. RANK statistical performance metric for the six LRT models and CAPTEX Release 5.

Table ES-6 summarizes the rankings of the six LRT models for the 11 performance statistics analyzed for CAPTEX Release 5 and compares the averaging ranking across the 11 statistics against the RANK metric rankings. Unlike the CTEX3 experiment, where CAMx (46%) and FLEXPART (36%) accounted for 82% of the first placed ranked models, there is a wide variation of which model was ranked best performing across the 11 statistical metrics in the CTEX5 experiment. In testing the efficacy of the RANK statistic, overall rankings across all eleven statics were obtained using an average modeled ranking. The average rank across all 11 performance statistics and the RANK model rankings are as follows:

Ranking	Average of 11 Statistics	RANK
1.	CAMx	CAMx
2.	HYSPLIT	HYSPLIT
3.	SCIPUFF	CALGRID
4.	FLEXPART	SCIPUFF
5.	CALPUFF	FLEXPART
6.	CALGRID	CALPUFF

The results from CAPTEX Release 5 present an interesting case study on the use of the RANK metric to characterize overall model performance. As noted in Table ES-6 and given above, the relative ranking of models using the average rankings across the 11 statistical metrics is considerably different than the RANK scores after the two highest ranked models (CAMx and

HYSPLIT). Both approaches show CAMx and HYSPLIT as the highest ranking models for CTEX5 with rankings that are fairly close to each other, however after that the two ranking techniques come to very different conclusions regarding the ability of the models to simulate the observed tracer concentrations for the CTEX5 field experiment.

The most noticeable feature of the RANK metric for ranking models in CTEX5 is the third highest ranking model using RANK, CALGRID (1.57). CALGRID ranks as the worst or second worst performing model in 9 of the 11 performance statistics, so is one of the worst performing model 82% of the time and has an average ranking of 5th best model out of the 6 LRT dispersion models. In examining the contribution to the RANK metric for CALGRID, there is not a consistent contribution from all four broad categories to the composite scores (Figure ES-5). As noted in Table ES-2, the RANK score is defined by the contribution of the four of the 11 statistics that represent measures of correlation/scatter (R^2), bias (FB), spatial (FMS) and cumulative distribution (KS):

$$RANK = \left|R^2\right| + \left(1 - \left|FB/2\right|\right) + FMS/100 + \left(1 - KS/100\right)$$

The majority of CALGRID's 1.57 RANK score comes from the fractional bias (FB) and Kolmogorov-Smirnov (KS) performance statistics with little or no contributions from the correlation (R^2) or spatial (FMS) statistics. As shown in Table ES-6, CALGRID performs very poorly for the FOEX and FA2/FA5 statistics due to a large underestimation bias. The FB component to the RANK composite score for CALGRID is one of the highest among the six models in this study, yet the underlying statistics indicate both marginal spatial skill and a large degree of under-prediction (likely due to the spatial skill of the model).

The current form of the RANK score uses the absolute value of the fractional bias. This approach weights underestimation equally to overestimation. However, in a regulatory context, EPA is most concerned with models not being biased towards under-prediction. Models can produce seemingly good (low) bias metrics through compensating errors by averaging over- and under-predictions. The use of an error statistic (e.g., NMSE) instead of a bias statistic (i.e., FB) in the RANK composite metrics would alleviate this problem.

Adaptation of RANK score for regulatory use will require refinement of the individual components to insure that this situation does not develop and to insure that the regulatory requirement of bias be accounted for when weighting the individual statistical measures to produce a composite score.

Table ES-6. Summary of model rankings using the statistical performance metrics and comparison with the RANK metric.

Statistic	1st	2nd	3rd	4th	5th	6th
FMS	SCIPUFF	CAMx	HYSPLIT	CALPUFF	FLEXPART	CALGRID
FAR	FLEXPART	HYSPLIT	CAMx	SCIPUFF	CALGRID	CALPUFF
POD	SCIPUFF	CAMx	HYSPLIT	FLEXPART	CALPUFF	CALGRID
TS	FLEXPART	HYSPLIT	CAMx	SCIPUFF	CALPUFF	CALGRID
FOEX	CALPUFF	CAMx	HYSPLIT	CALGRID	SCIPUFF	FLEXPART
FA2	HYSPLIT	CAMx	CALPUFF	SCIPUFF	FLEXPART	CALGRID
FA5	HYSPLIT	CAMx	SCIPUFF	CALPUFF	FLEXPART	CALGRID
NMSE	CAMx	SCIPUFF	FLEXPART	HYSPLIT	CALPUFF	CALGRID
PCC or R	HYSPLIT	CAMx	SCIPUFF	FLEXPART	CALGRID	CALPUFF
FB	CAMx	CALGRID	FLEXPART	SCIPUFF	HYSPLIT	CALPUFF
KS	HYSPLIT	CALPUFF	CALGRID	CAMx	FLEXPART	SCIPUFF
Avg. Ranking	CAMx	HYSPLIT	SCIPUFF	FLEXPART	CALPUFF	CALGRID
Avg. Score	2.20	2.4	3.4	3.8	4.3	5.0
RANK Ranking	CAMx	HYSPLIT	CALGRID	SCIPUFF	FLEXPART	CALPUFF
RANK	1.91	1.80	1.57	1.53	1.45	1.28

European Tracer Experiment (ETEX)

The European Tracer Experiment (ETEX) was conducted in 1994 with two tracer releases from northwest France that was measured at 168 samplers located in 17 European countries. Five LRT dispersion models were evaluated for the first (October 23, 1994) ETEX tracer release period (CALPUFF, SCICHEM, HYSPLIT, FLEXPART and CAMx). All five LRT dispersion models were exercised using a common 36 km MM5 database for their meteorological inputs. For CALPUFF, the MMIF tool was used to process the MM5 data. Default model options were mostly selected for the LRT dispersion models. An exception to this is that for CALPUFF puff splitting was allowed to occur throughout the day, instead of once per day which is the default setting. The MM5 simulation was evaluated using surface meteorological variables. The MM5 performance did not always meet the model performance benchmarks and exhibited a wind speed and temperature underestimation bias. However, since all five LRT dispersion models used the same MM5 fields, this did not detract from the LRT model performance intercomparison. The ATMES-II model evaluation approach was used in the evaluation that calculated 12 model performance statistics of spatial, scatter, bias, correlation and cumulative distribution.

ETEX LRT Dispersion Model Performance Evaluation

Figure ES-6 displays the ranking of the five LRT dispersion models using the RANK model performance statistic with Table ES-7 summarizing the rankings for the other 11 ATMES-II performance statistics. Depending on the statistical metric, three different models were ranked as the best performing model for a particular statistic with CAMx being ranked first most of the time (64%) and HYSPLIT ranked first second most (27%). In order to come up with an overall rank across all eleven statistics we average the modeled ranking order to come up with an average ranking that listed CAMx first, HYSPLIT second, SCIPUFF third, FLEXPART fourth and CALPUFF the fifth. This is the same ranking as produced by the RANK integrated statistics that combines the four statistics for correlation (PCC), bias (FB), spatial (FMS) and cumulative distribution (KS), giving credence that the RANK statistic is a potentially useful performance statistic for indicating overall model performance of a LRT dispersion model for the ETEX evaluation.

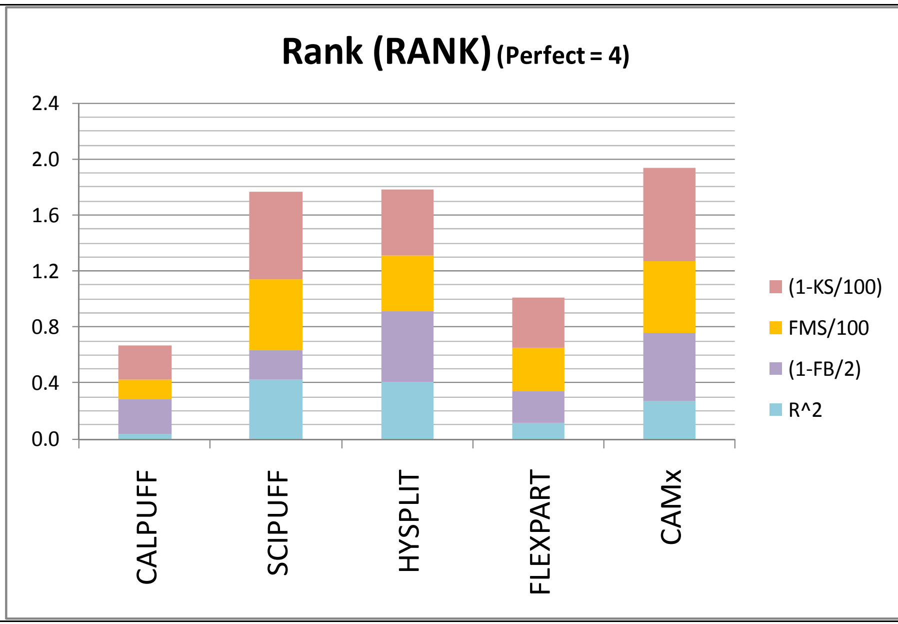


Figure ES-6. RANK statistical performance metric for the five LRT models and the ETEX tracer field experiment.

Table ES-7. Summary of ETEX model ranking using the eleven ATMES-II statistical performance metrics and their average rankings that are compared against the rankings by the RANK composite model performance metric.

Statistic	1st	2nd	3rd	4th	5th
FMS	CAMx	SCIPUFF	HYSPLIT	FLEXPART	CALPUFF
FAR	HYSPLIT	FLEXPART	CAMx	SCIPUFF	CALPUFF
POD	CAMx	SCIPUFF	HYSPLIT	FLEXPART	CALPUFF
TS	CAMx	HYSPLIT	SCIPUFF	FLEXPART	CALPUFF
FOEX	CAMx	SCIPUFF	HYSPLIT	FLEXPART	CALPUFF
FA2	CAMx	SCIPUFF	HYSPLIT	FLEXPART	CALPUFF
FA5	CAMx	SCIPUFF	HYSPLIT	FLEXPART	CALPUFF
NMSE	HYSPLIT	CAMx	CALPUFF	FLEXPART	SCIPUFF
PCC or R	SCIPUFF	HYSPLIT	CAMx	FLEXPART	CALPUFF
FB	HYSPLIT	CAMx	CALPUFF	FLEXPART	SCIPUFF
KS	CAMx	SCIPUFF	HYSPLIT	FLEXPART	CALPUFF
Avg. Ranking	CAMx	HYSPLIT	SCIPUFF	FLEXPART	CALPUFF
Avg. Score	1.55	2.27	2.73	3.82	4.64
RANK Ranking	CAMx	HYSPLIT	SCIPUFF	FLEXPART	CALPUFF
RANK Score	1.9	1.8	1.8	1.0	0.7

<u>Spatial Displays of Model Performance</u>

Figures ES-7 and ES-8 display the spatial distributions of the predicted and observed tracer concentrations 36 and 60 hours after the beginning of the ETEX tracer release. CALPUFF advects the tracer too far north keeping a circular Gaussian plume distribution and fails to reproduce the northwest to southeast diagonal orientation of the observed tracer cloud. The other four LRT dispersion models do a much better job in reproducing the observed tracer cloud spatial distribution. SCIPUFF tends to overestimate the tracer cloud extent and surface concentrations. FLEXPART, on the other hand, underestimates the observed tracer cloud spatial extent and CAMx and HYSPLIT do the best job overall in reproducing the spatial extent of the observed tracer cloud.

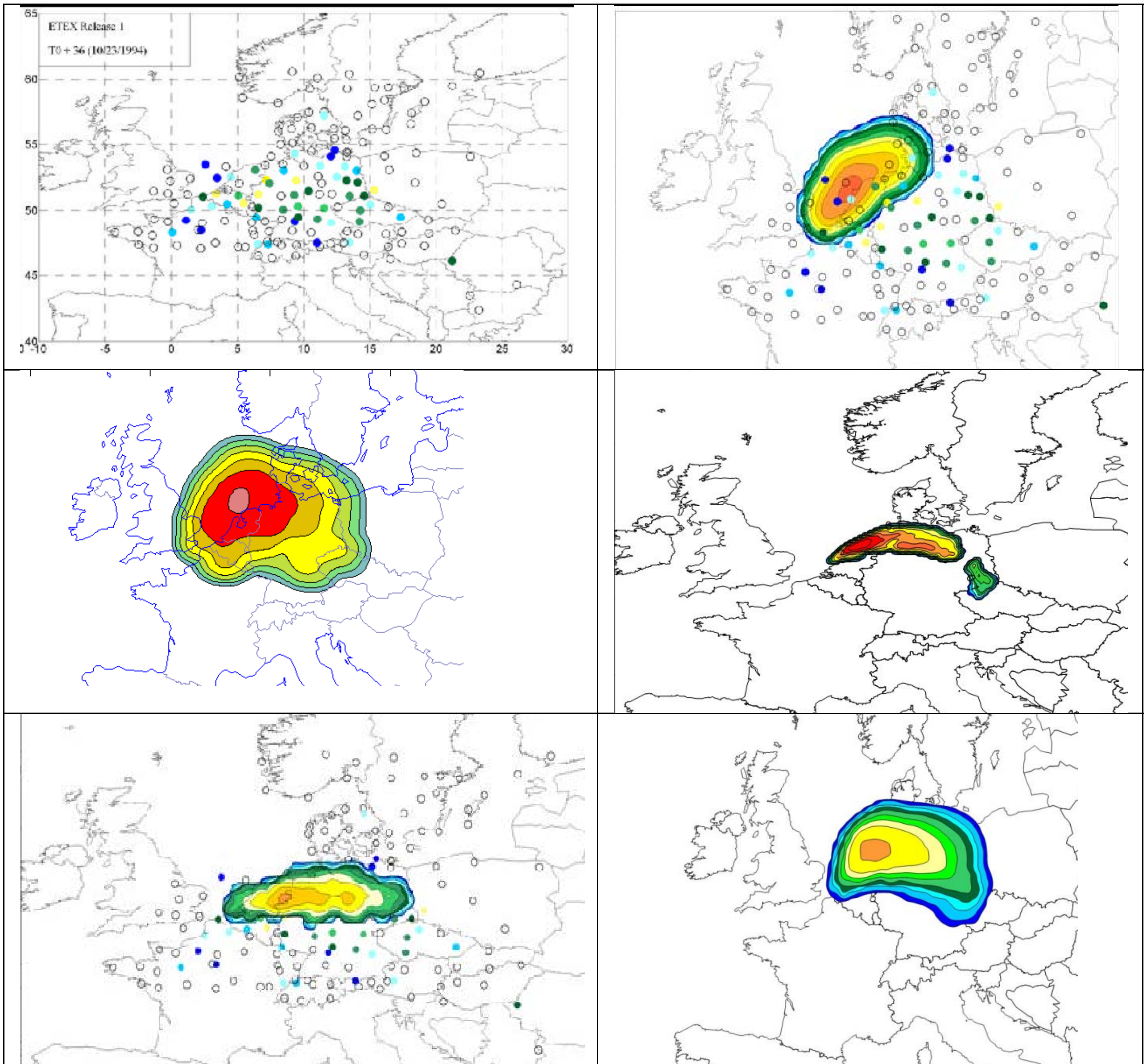


Figure ES-7. Comparison of spatial distribution of the ETEX tracer concentrations 36 hours after release for the observed (top left), CALPUFF (top right), SCIPUFF (middle left), FLEXPART (middle right), HYSPLIT (bottom left) and CAMx (bottom right).

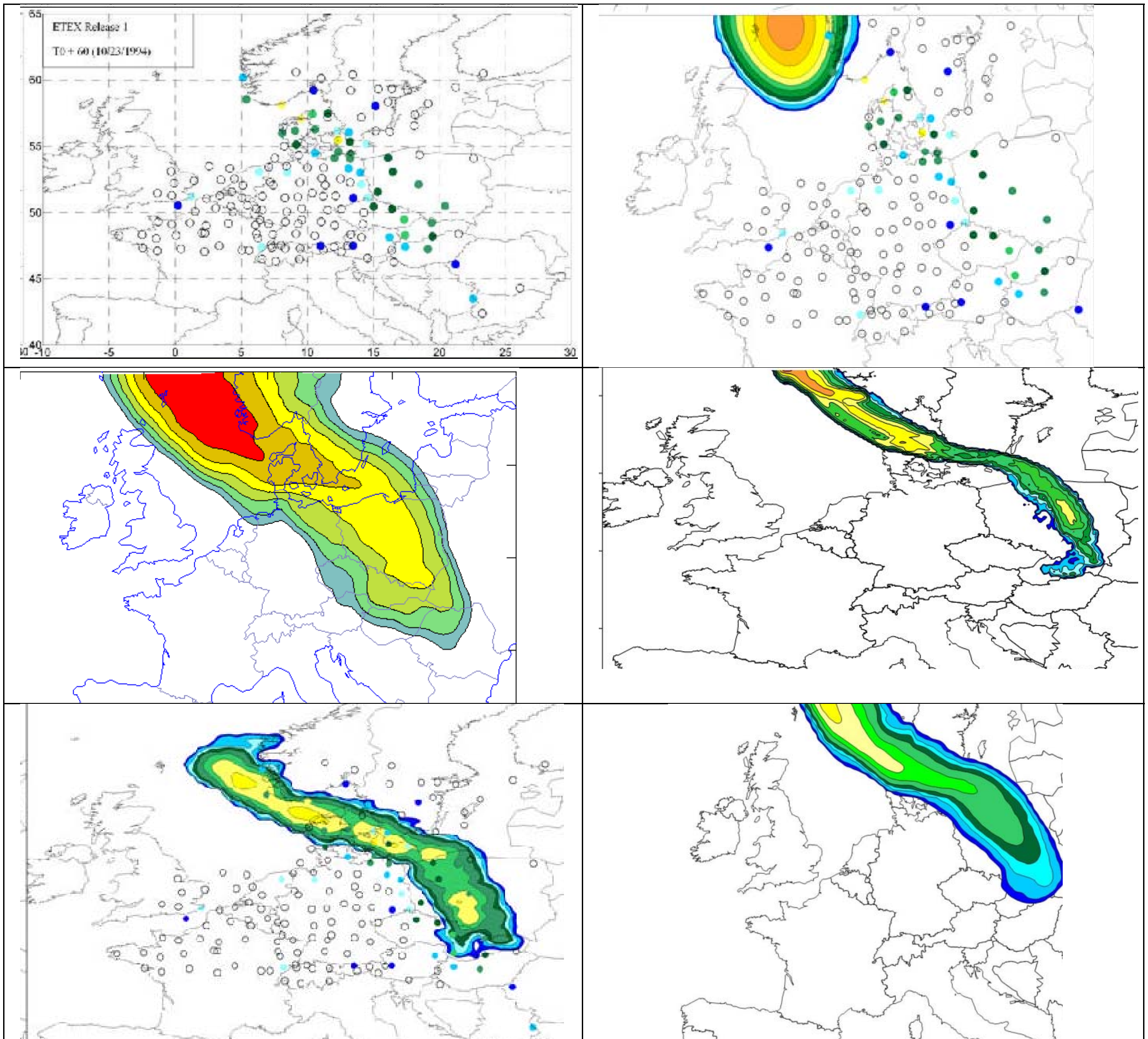


Figure ES-8. Comparison of spatial distribution of the ETEX tracer concentrations 60 hours after release for the observed (top left), CALPUFF (top right), SCIPUFF (middle left), FLEXPART (middle right), HYSPLIT (bottom left) and CAMx (bottom right).

ETEX LRT Dispersion Model Sensitivity Tests

Sensitivity tests were conducted using the CAMx, CALPUFF and HYSPLIT models and the ETEX field study data.

For CAMx, the effects of alternative vertical mixing coefficients (OB70, TKE, ACM2 and CMAQ), horizontal advection solvers (PPM and Bott) and use of the subgrid-scale Plume-in-Grid (PiG) module were evaluated. The key findings from the CAMx ETEX sensitivity tests were as follows:

- The vertical mixing parameter had the biggest effect on model performance, with the CMAQ vertical diffusion coefficients producing the best performing CAMx simulations.
- The horizontal advection solver had a much smaller effect on CAMx model performance with the PPM algorithm performing slightly better than Bott.
- The use of no PiG module produced slightly better performance than use of the PiG module.
- The default CAMx configuration used in the ETEX evaluation (CMAQ/PPM/No PiG) was the best performing CAMx sensitivity test.

CALPUFF sensitivity tests were performed to examine the effects of puff splitting on the CALPUFF model performance for the ETEX field experiment. When EPA listed CALPUFF as the EPA-recommended LRT dispersion model in 2003, they noted that the implementation of puff splitting likely will extend the models applicability beyond 300 km downwind (EPA, 2003). Since many of the ETEX monitoring sites are sited further than 300 km downwind from the release, one potential explanation for the poor CALPUFF model performance is that it is being applied farther downwind than the model is applicable for. Figure ES-9 displays a time series of the Figure of Merit in Space (FMS) performance statistic for the five LRT dispersion models. Although CALPUFF performs reasonably well within the first 12 hours of the tracer release, its performance quickly degrades even within 300 km of the source. Thus, CALPUFF's poor model performance is not due to applying the model to downwind distances beyond its applicability.

Eight CALPUFF puff splitting sensitivity tests were conducted ranging from no puff splitting to aggressive puff splitting for all hours of the day and relaxing some of the puff splitting initiation criteria so that even more puff splitting can occur. The CALPUFF ETEX model performance using no puff splitting and all hour puff splitting was very similar, thus we saw no evidence to support EPA's 2003 statements that puff splitting may extend the downwind applicability of the model. In fact, when some of the puff splitting initiation criteria were relaxed to allow more puff splitting, the CALUFF performance degraded.

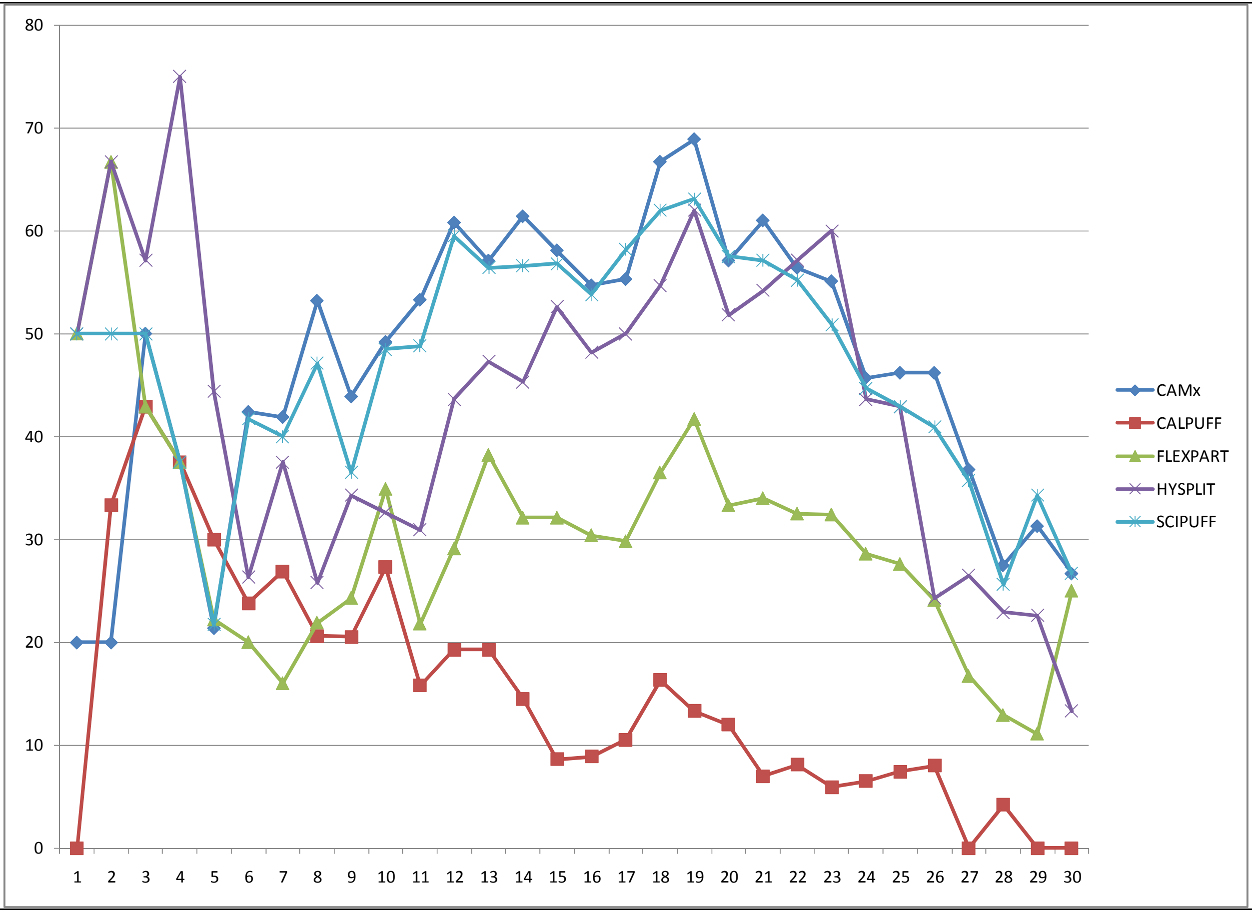


Figure ES-9. Figure of Merit (FMS) spatial model performance statistics as a function of time at three hour increments since the beginning of the tracer release.

The HYSPLIT LRT model was unique among the five LRT dispersion models examined in that it can be run in a particle mode, a Gaussian puff mode or hybrid particle/puff and puff/particle modes. The default configuration used in the HYSPLIT simulations presented previously was the three-dimensional particle mode. Nine HYSPLIT sensitivity tests were performed using different particle and puff formulation combinations. The RANK scores for the HYSPLIT ETEX sensitivity simulations ranged from 1.01 to 2.09, with the fully puff formulation ranked the lowest and hybrid puff/particle combinations ranked highest.

<u>Conclusions of the ETEX LRT Dispersion Model Evaluation</u>

Five LRT dispersion models were evaluated using the 1994 ETEX tracer test field experiment data. The CAMx, HYSPLIT and SCIPUFF models were the highest ranked LRT dispersions models, with CAMx performing slightly better than the other two models. The reasons for the poor performance of CALPUFF appear to be due to its inability to adequately treat horizontal and vertical wind shear. The CALPUFF Gaussian puff formulation retains a well-mixed circular puff despite the presence of wind variations across the puff that would advect tracer concentrations in different directions. Because the puff can only be transported by one wind, CALPUFF is unable to adequately treat such wind variations across the puff. The use of puff splitting, which EPA postulated in 2003 may extend the downwind applicability of the model, failed to have any significant effect on CALPUFF model performance.

CONCLUSIONS OF LRT DISPERSION MODEL TRACER TEST EVALUATION

The following are some of the key conclusions of the LRT dispersion model tracer test field experiment evaluation.

CALPUFF/CALMET Concentration Predictions are Highly Variable: Use of alternative CALMET input options within their range of reasonableness can produce wide variations in the CALPUFF concentration predictions. Given the regulatory use of CALPUFF, this result points toward the need to have a standard set of recommended CALMET settings for regulatory application of CALPUFF to assure consistency and eliminate the potential of selecting CALMET options to obtain a desired outcome in CALPUFF. No one CALMET configuration consistently produced the best CALPUFF model performance, although use of MM5 data with CALMET did tend to improve CALPUFF model performance with 36 and 12 km MM5 data being better than 80 km MM5 data.

Comparison of Current CALPUFF Model Performance with Previous Studies: The comparison of the model performance for current version of CALPUFF with past CALPUFF evaluations from the 1998 EPA study (EPA, 1998a) using the GP80 and SRL75 tracer study field experiments was mixed. For the GP80 100 km receptor arc, the current and past CALPUFF model performance evaluations were consistent with CALPUFF tending to overestimate the plume maximum concentrations and underestimate plume horizontal dispersion. The current version of CALPUFF had difficulty in reproducing the good performance of the past CALPUFF application in estimating the tracer residence time on the GP80 600 km receptor arc. Only by invoking the CALPUFF slug option, as used in the 1998 EPA study, was CALPUFF/CALMET able to reproduce the tracer residence time on the 600 km receptor arc. As the slug option is for near-source modeling and is a very non-standard option for LRT dispersion modeling, this result questions the validity of the 1998 CALPUFF evaluation study as applied for CALPUFF LRT modeling. The CALPUFF/MMIF was less sensitive to the slug option and more sensitive to puff splitting than CALPUFF/CALMET. For consistency, the current and EPA 1998 study CALPUFF evaluation approach both used the fitted Gaussian plume model evaluation methodology, along with angular plume centerline offset and tracer receptor arc timing statistics. The fitted Gaussian plume evaluation approach assumes that the observed and predicted concentration along a receptor arc has a Gaussian distribution. At longer downwind distances such an assumption may not be valid. For the CALPUFF evaluation using the SRL75 tracer field experiment, there was a very poor fit of the Gaussian plume to the observations resulting in some model performance statics that could be misleading. We do not recommend using the fitted Gaussian plume evaluation approach in future studies and instead recommend using approaches like the ATMES-II statistical evaluation approach that is free from any a priori assumption regarding the observed tracer distributions.

EPA-FLM Recommended CALMET Settings from the 2009 Clarification Memorandum: The EPA-FLM recommended CALMET settings in the 2009 Clarification Memorandum (EPA, 2009b) produces wind field estimates closest to surface wind observations based on the CAPTEX CALMET modeling. However, when used as input into CALPUFF, the EPA-FLM recommended CALMET settings produced one of the poorer performing CALPUFF/CALMET configurations when comparing CALPUFF predictions against the observed atmospheric tracer concentrations. Given that the CALMET wind evaluation is not an independent evaluation because some of the wind observations used in the evaluation database are also input into CALMET, the CALPUFF tracer evaluation bears more weight. Other aspects of the EPA-FLM recommended settings generally produced better CALPUFF tracer model performance including use of prognostic meteorological data as input to CALPUFF. The CALPUFF evaluation also found better CALPUFF performance when 12 km grid resolution is used in MM5 or CALMET as opposed to 80 or 36 km.

CALPUFF Model Performance using CALMET versus MMIF: The CALPUFF tracer model performance using meteorological inputs based on the MMIF tool versus CALMET was mixed. The variations of the CALPUFF model predictions using MMIF were much less than when CALMET was used and the CALPUFF/MMIF model performance was usually within the range of the performance exhibited by CALPUFF/CALMET. Specific examples from the tracer tests are as follows:

- For the GP80 100 km receptor arc, the CALPUFF/MMIF exhibited better fitted plume observed tracer model performance statistics than all of the CALPUFF/CALMET configurations except when CALMET was run using MM5 and surface meteorological observations but no upper-air meteorological observations.
- CALPUFF/CALMET using no MM5 data and just meteorological observations exhibited the best plume centerline location on the GP80 100 km receptor arc with CALPUFF/CALMET using just MM5 data and no observations and CALMET/MMIF exhibiting the worst plume centerline location.
- For the GP80 600 km receptor arc, the CALPUFF/MMIF fitted plume model performance statistics are in the middle of the performance statistics for the CALPUFF/CALMET configurations.
- The slug option was needed for CALPUFF/CALMET to produce good 600 km receptor arc tracer residence time statistics but had little effect on CALPUFF/MMIF. However, use of puff splitting greatly improved the CALPUFF/MMIF tracer residence time statistics.
- Of all the CALPUFF sensitivity tests examined, CALPUFF/MMIF using the slug option and puff splitting produced the best CALPUFF fitted plume tracer model performance statistics for the GP80 600 km receptor arc.
- In an opposite fashion to the GP80 100 km receptor arc, for the SRL75 100 km receptor arc the best plume centerline offset was achieved when CALPUFF was run with just MM5 data and no meteorological observations (either with CALMET or MMIF) with performance degraded when meteorological observations are used with CALMET.
- The CALPUFF model performance using the MMIF tool and 36 and 12 km MM5 data performed better than all of the CALPUFF/CALMET sensitivity tests for the CAPTEX CTEX3 experiment. However, the CALPUFF/MMIF using 36 and 12 km MM5 data performed worse than all of the CALPUFF/CALMET sensitivity tests for the CAPTEX CTEX5 experiment.

Comparison of Model Performance of LRT Dispersion Models: Six LRT dispersion modeled were evaluated using the CAPTEX Release 3 and 5 tracer database and five LRT dispersion models were evaluated using the ETEX tracer test field experiment. In each case the same MM5 meteorological data were used as input into all of the dispersion models, although different MM5 configuration options were selected for each tracer experiement.

The CAMx and CALGRID Eulerian photochemical grid models, FLEXPART Lagrangian particle model, HYSPLIT Lagrangian particle, puff and particle/puff hybrid model and CALPUFF and SCIPUFF Gaussian puff models were evaluated. For all three tracer experiments (CTEX3, CTEX5 and ETEX), the CAMx model consistently ranked highest when looking across all of the model performance statistics or when using the RANK composite performance statistic. For the CTEX3 field experiment, the RANK composite performance statistic gave consistent rankings of model performance with the suite of statistical metrics with CAMx being the higheset RANK score (1.91) followed by SCICHEM (1.71).

The rankings of the models using all of the statistics versus the RANK composite statistic were inconsistent for the CTEX5 experiment. Both approaches showed CAMx and HYSPLIT were the highest ranking LRT dispersion model for the CTEX5 field experiment. However, the RANK statistic ranked CALGRID as the 3rd best performing model, whereas when looking at all the performance statistics it was the worst performing model because it exhibited a large spread underestimation bias, had no correlation with the observations and little skill in reproducing the spatial distribution of the observed tracer. The CTEX5 LRT model evaluation points out the need to examine all performance statistics and not rely solely on the RANK composite statistic. It also points out the need to define a RANK-type composite statistic that focuses on the regulatory application of LRT dispersion models where an underestimation bias is undesirable.

Of the three top performing LRT dispersion models, CAMx had the highest RANK composite statistic and scored the highest for most (64%) of the other ATMES-II statistical model performance metrics, with HYSPLIT scoring the highest for 27% of the metrics. Additional findings of the ETEX tracer test evaluation are as follows:

- The model performance rankings were preserved closer to the source (e.g., within 300 km) as well as further downwind.
- CALPUFF puff splitting sensitivity tests had little effect on CALPUFF model performance.
- CAMx vertical mixing and horizontal advection solver sensitivity tests found that use of the MM5CAMx CMAQ-like vertical mixing diffusion coefficients and the PPM advection solver produced the best tracer test model performance. Similar results were seen in the CTEX3 and CTEX5 sensitivity modeling.
- HYSPLIT sensitivity tests using solely particle, solely puff and hybrid particle/puff and puff/particle combinations found that the hybrid configurations performed best and the puff configuration performed worst, with the CTEX3 and CTEX5 sensitivity test producing similar results.

1.0 INTRODUCTION

Dispersion models, such as the Industrial Source Complex Short Term (ISCST; EPA, 1995) or American Meteorological Society/Environmental Protection Agency Regulatory Model (AERMOD; EPA, 2004; 2009c) typically assume steady-state, horizontally homogeneous wind fields instantaneously over the entire modeling domain and are usually limited to distances of less than 50 kilometers from a source. However, dispersion model applications of distances of hundreds of kilometers from a source require other models or modeling systems. At these distances, the transport times are sufficiently long that the mean wind fields cannot be considered steady-state or homogeneous. As part of the Prevention of Significant Deterioration (PSD) program, new sources or proposed modifications to existing sources may be required to assess the air quality and Air Quality Related Values (AQRVs) impacts at Class I and sensitive Class II areas that may be far away from the source. AQRVs include visibility and acid (sulfur and nitrogen) deposition. There are 156 federally mandated Class I areas in the U.S. that consist of National Parks, Wilderness Areas and Wildlife Refuges that are administered by Federal Land Managers (FLMs) from the National Park Service (NPS), United States Forest Service (USFS) and Fish and Wildlife Service (FWS), respectively. Thus, non-steady-state Long Range Transport (LRT) dispersion models are needed to address air quality and AQRVs issues at distances beyond 50 km from a source.

1.1 BACKGROUND

The Interagency Workgroup on Air Quality Modeling (IWAQM) was formed to provide a focus for the development of technically sound recommendations regarding assessment of air pollutant source impacts on Federal Class I areas. Meetings were held with personnel from interested Federal agencies, including the Environmental Protection Agency (EPA), the USFS, NPS and FWS. The purpose of these meetings was to review respective modeling programs, to develop an organizational framework, and to formulate reasonable objectives and plans that could be presented to management for support and commitment. One objective of the IWAQM is the recommendation of LRT dispersion models for assessing air quality and AQRVs at Class I areas.

One such LRT dispersion model is the CALPUFF modeling system (Scire et al., 2000b). The CALPUFF modeling system consists of several components: (1) CALMET (Scire et al., 2000a), a meteorological preprocessor that can use as input surface, upper air, and/or on-site meteorological observations and/or prognostic meteorological model output data to create a three-dimensional wind field and derive boundary layer parameters based on gridded land use data; (2) CALPUFF, a Lagrangian puff dispersion model that can simulate the effects of temporally and spatially varying meteorological conditions on pollutant transport, remove pollutants through dry and wet deposition processes, and includes limited ability to transform pollutant species through chemical reactions; and (3) CALPOST, a postprocessor that takes the hourly estimates from CALPUFF and generates *n*-hr estimates as well as tables of maximum values.

In 1998, EPA published the report entitled “A Comparison of CALPUFF Modeling Results to Two Tracer Field Experiments” (EPA-454/R-98-009) (EPA, 1998a). The 1998 EPA study examined concentration estimates from the CALPUFF dispersion model that were compared to observed tracer concentrations from two short term field experiments. The first experiment was at the Savannah River Laboratory (SRL75) in South Carolina in December 1975 (DOE, 1978) and the second was the Great Plains experiment (GP80) near Norman, Oklahoma (Ferber et al., 1981) in

July 1980. Both experiments examined long-range transport of inert tracer materials to demonstrate the feasibility of using other tracers as alternatives to the more commonly used sulfur hexafluoride (SF_6). Several tracers were released for a short duration (3-4 hours) and the resulting plume concentrations were recorded at an array of monitors downwind from the source. For the SRL75 field experiment, monitors were located approximately 100 kilometers from the source. For the Great Plains experiment, arcs of monitors were located 100 and 600 kilometers from the source.

In 1998, IWAQM released their Phase 2 recommendations in a report "Interagency Workgroup on Air Quality Modeling (IWAQM) Phase 2 Summary Report and Recommendations for Modeling Long Range Transport Impacts" (EPA, 1998b[7]). These recommendations included a screening and refined LRT modeling approach based on the CALPUFF modeling system. The IWAQM recommendations were based in part on the 1998 EPA tracer test CALPUFF evaluation. It was IWAQM's conclusion at the time that it was not possible to prescribe all of the decisions needed in a CALPUFF/CALMET application: *"The control of the CALMET options requires expert understanding of mesoscale and microscale meteorological effects on meteorological conditions, and finesse to adjust the available processing controls within CALMET to develop the desired effects. The IWAQM does not anticipate the lessening in this required expertise in the future"* (EPA, 1998b).

On April 15, 2003, EPA issued a "Revision to the Guideline on Air Quality Models: Adoption of a Preferred Long Range Transport Model and Other Revisions" in the Federal Register (EPA, 2003[8]) that adopted the CALPUFF model as the EPA-recommended (Appendix W) model for assessing the far-field (> 50 km) air quality impacts due to chemically inert pollutants. In 2005, EPA issued another revision to the air quality modeling guidelines that recommended the AERMOD steady-state Gaussian plume model be used for near-source air quality issues. Thus, from 2005 on to present, there are two EPA-recommended models to address air quality issues due to primary pollutants: AERMOD for near-source (< 50 km) assessments; and CALPUFF for far-field (> 50 km) assessments.

In 2005, EPA formed a CALPUFF workgroup to help identify issues with the existing 1998 IWAQM guidance. In response to this, EPA initiated reevaluation of the CALPUFF system to update the 1998 IWAQM Phase 2 Recommendations.

In May 2009, EPA released a draft document entitled the "Reassessment of the Interagency Workgroup on Air Quality Modeling (IWAQM) Phase 2 Summary Report: Revisions to the Phase 2 Recommendations" (EPA, 2009a). In this document, EPA described the developmental status of the CALPUFF modeling system. CALPUFF has evolved continuously since the publication of the original 1998 IWAQM Phase 2 recommendations; however, the status of CALPUFF related guidance has not kept pace with the developmental process. The May 2009 IWAQM Phase 2 Reassessment Report noted that *"The required expertise and collective body of knowledge in mesoscale meteorological models has never fully emerged from within the dispersion modeling community to support the necessary expert judgment on selection of CALMET control options"* (EPA, 2009a). In regards to the 1998 IWAQM Phase 2 lack of prescribing recommended CALMET settings, the May 2009 IWAQM Phase 2 Reassessment Report states: *"In a regulatory context, this situation has often resulted in an 'anything goes' process, whereby model control option selection can be leveraged as an instrument to achieve a desired modeled outcome,*

7 http://www.epa.gov/scram001/7thconf/calpuff/phase2.pdf

8 http://www.federalregister.gov/articles/2003/04/15/03-8542/revision-to-the-guideline-on-air-quality-models-adoption-of-a-preferred-long-range-transport-model

without regard to the scientific legitimacy of the options selected" (EPA, 2009a). The CALPUFF working group noted that when running CALMET with prognostic meteorological model (e.g., WRF and MM5) output as input, the CALMET diagnostic effects and blending of meteorological observations with the WRF/MM5 output degraded the WRF/MM5 meteorological fields. Thus, the 2009 IWAQM Phase 2 Reassessment Report recommended CALMET settings with an objective to try and "pass through" the WRF/MM5 meteorological model output as much as possible for input into CALPUFF.

However, further testing of CALMET and CALPUFF by EPA's CALPUFF workgroup found that the recommended CALMET settings in the May 2009 IWAQM Phase 2 Reassessment Report did not achieve the intended result to "pass through" the WRF/MM5 meteorological variables as CALMET still re-diagnosed some and modified other meteorological variables thereby degrading the WRF/MM5 meteorological fields. Based in part of CALMET evaluations using tracer test field study databases (presented in Appendix B of this report), EPA determined interim CALMET settings that produced the best CALMET performance when compared to observed surface winds and on August 31, 2009 released a Clarification Memorandum "Clarification on EPA-FLM Recommended Settings for CALMET" (EPA, 2009b) with new recommended settings for CALMET. In the August 2009 Clarification Memorandum, EPA reiterated the desire to "pass through" meteorology from the WRF/MM5 prognostic meteorological models to CALPUFF, but the CALMET model at this time was incapable of achieving that objective.

In the meantime, EPA has developed the Mesoscale Model Interface (MMIF) software that where possible directly converts prognostic meteorological output data from the MM5 or WRF models to the parameters and formats required for direct input into the CALPUFF dispersion model thereby bypassing CALMET. Version 1.0 of MMIF was developed in June 2009 (Emery and Brashers, 2009) with versions 2.0 (Brashers and Emery, 2011) and 2.1 (Brashers and Emery, 2012) developed in, respectively, September 2011 and February 2012; we expect that MMIF Version 2.1 will be publicly released in February 2012. MMIF specifically processes geophysical and meteorological output files generated by the fifth generation mesoscale model (MM5) or the Weather Research and Forecasting (WRF) model (Advanced Research WRF [ARW] core, versions 2 and 3) and reformats the MM5/WRF output for input into CALPUFF..

The EPA CALPUFF workgroup has been evaluating CALPUFF using CALMET and MMIF meteorological drivers using data from several historical tracer field studies. In addition to a reevaluation of CALPUFF using CALMET and MMIF for the GP80 and SRL75 tracer studies that were used in the 1998 EPA CALPUF tracer evaluation report (EPA, 1998a), the CALPUFF workgroup has also evaluated CALPUFF using CALMET and MMIF meteorological drivers along with 5 other LRT dispersion models for the 1983 Cross Appalachian Tracer Experiment (CAPTEX). CALPUFF, along with four other LRT dispersion models, were also evaluated using data from the 1994 European Tracer Experiment (ETEX).

1.2 PURPOSE

The purpose of this report is to document the evaluation of the CALPUFF LRT dispersion model using data from four atmospheric tracer experiment field study databases. This includes the comparison of the CALPUFF model performance using meteorological inputs based on the CALMET and MMIF software and comparison of the CALPUFF model performance with other LRT dispersion models.

1.3 ORGANIZATION OF REPORT

Chapter one provides a background and purpose for the study. In Chapter 2, the four tracer field study experiments and LRT dispersion models used in the model performance evaluation are summarized. Chapter 2 also summarizes related previous studies and the approach and methods for the model performance evaluation of the LRT dispersion models.

Chapters 3, 4, 5 and 6 contain the evaluation of the LRT dispersions models using the GP80, SRL75, CAPTEX and ETEX tracer study field experiment data. References are provided in Chapter 7. Appendix A contains an evaluation of the MM5 and CALMET meteorological models using the CAPTEX Release #5 (CTEX5) database. Appendix B presents the evaluation of the CALMET meteorological model using the CAPTEX Release #3 (CTEX3) database that was used in part to formulate the EPA-FLM recommended settings in the 2009 Clarification Memorandum (EPA, 2009b). Results of the evaluation of six LRT dispersion models using the CAPTEX tracer field experiments are presented in Appendix C.

2.0 OVERVIEW OF APPROACH

2.1 SUMMARY OF TRACER TEST FIELD EXPERIMENTS

LRT dispersion models are evaluated using four atmospheric tracer test field studies as follows:

1980 Great Plains: The 1980 Great Plains (GP80) field study released several tracers from a release site near Norman, Oklahoma in July 1980 and measured the tracers at two arcs to the northeast at distances of 100 and 600 km (Ferber et al., 1981).

1975 Savannah River Laboratory: The 1975 Savannah River Laboratory (SRL75) study released tracers from the SRL in South Carolina and measured them at several receptors approximately 100 km from the release point (DOE, 1978).

1983 Cross Appalachian Tracer Experiment: The 1983 Cross Appalachian Tracer Experiment (CAPTEX) was a series of three-hour tracer released from Dayton, OH and Sudbury, Canada during September and October, 1983. Sampling was made in a series of arcs approximately 100 km apart that spanned from 300 to 1,100 km from the Dayton, OH release site (Ferber et al., 1986).

1994 European Tracer Experiment: The 1994 European Tracer Experiment (ETEX) consisted of two tracer releases from northwest France in October and November 1994 that was measured at 168 monitoring sites in 17 countries (Von Dop et al., 1998).

2.2 SUMMARY OF LRT DISPERSION MODELS

Up to six LRT dispersion models were evaluated using the tracer test field study data:

CALPUFF[9]: The California Puff (CALPUFF Version 5.8; Scire et al, 2000b) model is a Lagrangian Gaussian puff model that simulates a continuous plume using overlapping circular puffs. Included with CALPUFF is the CALMET meteorological processor (Scire et al., 2000a) that includes a diagnostic wind model (DWM). The EPA has developed a new Mesoscale Model Interface (MMIF; Emery and Brashers, 2009; Brashers and Emery, 2011; 2012) tool that will "pass through" output from the MM5 or WRF prognostic meteorological models without modifying or rediagnosing the meteorological variables, as is done in CALMET. A major objective of this study was to compare the CALPUFF model performance using CALMET and MMIF meteorological drivers.

SCIPUFF[10]: The Second-order Closure Integrated PUFF (SCIPUFF Version 2.303; Sykes et al., 1998) is a Lagrangian puff dispersion model using Gaussian puffs to represent an arbitrary, three-dimensional time-dependent concentration field. The diffusion parameterization is based on turbulence closure theory, which gives a prediction of the dispersion rate in terms of the measurable turbulent velocity statistics of the wind field. The SCIPUFF contains puff splitting when wind shear is encountered across a puff and puff merging when two puffs occupy the same space.

HYSPLIT[11]: The Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT Version 4.8; Draxler, 1997) is a complete system for computing simple air parcel trajectories to complex dispersion and deposition simulations. The dispersion of a pollutant is calculated by assuming either puff or particle or hybrid puff/particle dispersion. In the puff model,

9 http://www.src.com/calpuff/calpuff1.htm
10 http://www.sage-mgt.net/services/modeling-and-simulation/scipuff-dispersion-model
11 http://www.arl.noaa.gov/HYSPLIT_info.php

puffs expand until they exceed the size of the meteorological grid cell (either horizontally or vertically) and then split into several new puffs, each with its share of the pollutant mass. In the particle model, a fixed number of particles are advected about the model domain by the mean wind field and spread by a turbulent component. The model's default configuration assumes a 3-dimensional particle distribution (horizontal and vertical).

FLEXPART[12]: The FLEXPART (Version 6.2; Siebert, 2006; Stohl et al., 2005[13]) model is a Lagrangian particle dispersion model developed at the Norwegian Institute for Air Research in the Department of Atmospheric and Climate Research. FLEXPART was originally designed for calculating the long-range and mesoscale dispersion of air pollutants from point sources, such as after an accident in a nuclear power plant. In the meantime FLEXPART has evolved into a comprehensive tool for atmospheric transport modeling and analysis

CAMx[14]: The Comprehensive Air-quality Model with extensions (CAMx; ENVIRON, 2010) is a photochemical grid model (PGM) that simulates inert or chemical reactive pollutants from the local to continental scale. As a grid model, it simulates transport and dispersion using finite difference techniques on a three-dimensional array of grid cells. To treat the near-source dispersion of plumes, CAMx includes a subgrid-scale Lagrangian puff Plume-in-Grid (PiG) module whose mass is transferred to the grid model when the plume size is comparable to the grid size.

CALGRID: The California Mesoscale Photochemical Grid Model (Yamartino, et al., 1989, Scire et al., 1989; Earth Tech, 2005) is a PGM that simulates chemically reactive pollutants from the local to regional scale. As with CAMx, it is a grid model that simulates transport and dispersion using finite differencing techniques on a three-dimensional array of grid cells. CALGRID was originally designed to utilize meteorological fields produced by the CALMET meteorological processor (Scire et al., 2000a), but was updated in 2006 to utilize meteorology and emissions in UAM format (Earth Tech, 2006).

Although up to six LRT dispersion models were run for two of the tracer field experiments, a key component of this study was the evaluation of the CALPUFF model and running CALPUFF with various configurations of its meteorological drivers, CALMET and MMIF to help inform regulatory guidance on the operation of the CALPUFF system. Key to developing insight into the performance of any single model is to evaluate other models when configured similarly and using similar meteorological databases. Table 2-1 summarizes which LRT models were run with the four field study tracer experiments presented in this report.

For the GP80 CALPUFF/CALMET application, numerous CALPUFF sensitivity tests were performed using different configurations of CALMET including with and without MM5 data and use of no observations. A limited set of CALPUFF sensitivity tests were also conducted using different dispersion options. The other LRT models (save CALGRID) results were also evaluated for the 600 km distant arc of receptors, but are not presented in the CALPUFF comparison because this evaluation is based upon the NOAA DATEM statistical framework and is not consistent with how CALPUFF was evaluated by EPA for this experiment in 1998.

---

12 http://transport.nilu.no/flexpart
13 http://www.atmos-chem-phys.net/5/2461/2005/acp-5-2461-2005.html
14 http://www.camx.com/

The evaluation of the LRT models using the SRL75 tracer data only has results for CALPUFF. Several CALPUFF/CALMET sensitivity tests were run using only meteorological observations, only MM5 data and hybrid MM5 plus meteorological observations. CALPUFF/MMIF was run using 36, 12 and 4 km MM5 data.

Two tracer releases were evaluated using the CAPTEX database, Releases No. 3 and 5. While all of the models listed in Table 2-1 were run for the CAPTEX database, numerous CALMET sensitivity tests were also conducted, including the evaluation of CALMET using various configurations for CAPTEX Release No. 3 and 5 that helped define the EPA-FLM recommended CALMET settings in the August 2009 Clarification Memorandum (EPA, 2009b).

The LRT model intercomparison using the CAPTEX and ETEX databases was done differently than the other two tracer test evaluations. The objective of the ETEX with CAPTEX LRT model evaluation intercomparison was to evaluate the LRT dispersion models using a common meteorological input database. Thus, all LRT models used the same MM5 meteorological inputs.

Table 2-1. Model availability for the four tracer test field experiments.

Model	GP80	SRL75	CAPTEX	ETEX
CALPUFF/CALMET	Yes	Yes	Yes	No
CALMET/MMIF	Yes	Yes	Yes	Yes
SCIPUFF	No	No	Yes	Yes
HYSPLIT	No	No	Yes	Yes
FLEXPART	No	No	Yes	Yes
CAMx	No	No	Yes	Yes
CALGRID	No	No	Yes	No

2.3 RELEATED PREVIOUS STUDIES

Over the years there have been numerous studies that have evaluated dispersion models using tracer test and other field study databases. In fact, much of the early development of Gaussian plume dispersion formulation was assisted by radioactive ambient field data (Slade, 1968). The development and evaluation of the AERMOD steady-state Gaussian plume model used almost 20 near-source field study datasets[15]. The discussion below is limited to long range transport (LRT) dispersion model evaluations that have been related to the development of the CALPUFF modeling system, which in 2003 was identified as the EPA recommended regulatory LRT model for far-field (> 50 km) air quality modeling of chemically inert compounds (EPA, 2003).

2.3.1 1986 Evaluation of Eight Short-Term Long Range Transport Models

EPA sponsored a study to evaluate 8 LRT models using the GP80 tracer field experiment and Krypton-85 releases from the Savannah River Laboratory (SRL; Telegadas et al., 1980) databases (Policastro et al., 1986). The eight models were MESOPUFF, MESOPLUME, MSPUFF, MESOPUFF-II, MTDDIS, ARRPA, RADM and RTM-II. MESOPUFF, MSPUFF and MESOPUFF-II are Lagrangian puff models that all have their original basis on the MESOPUFF model. MESOPLUME is a Lagrangian plume segment model. MTDDIS is a variable trajectory model that also uses the Gaussian puff formulation. ARRPA is a single-source segmented plume model. RADM and RTM-II are Eulerian grid models. Model performance was evaluated by graphical and statistical methods. The primary means for the evaluation of model performance was the use of the American Meteorological Society (AMS) statistics (Fox, 1981). The AMS statistics recommends

15 http://www.epa.gov/ttn/scram/dispersion_prefrec.htm#aermod

that performance evaluation be based on comparisons of the full set of predicted/observed data pairs as well as the highest predicted and observed values per event and the highest N values (e.g., N=10) unpaired in space or time that represents the highest end of the concentration distribution.

Six of the eight LRT models were applied to both the GP80 and SRL75 experiments. The ARRPA model could only be applied to the GP80 database and the MTDDIS model could only be applied to the SRL75 database. Model performance was generally consistent between the two tracer databases and was characterized by three features:

- A spatial offset of the predicted and observed patterns.
- A time difference between the predicted and observed arrival of the plumes to the receptors.
- A definite angular offset of the predicted and observed plumes that could be as much as 20-45 degrees.

The LRT models tended to underestimate the horizontal spreading of the plume at ground level resulting in too high peak (centerline) concentrations when compared to the observations. For the Lagrangian models this is believed to be due to using sigma-y dispersion (Turner) curves that are representative of near-source and are applied for longer (> 50 km) downwind distances. The spatial and angular offsets resulted in poor correlations and large bias and error between the predicted and observed tracer concentrations when paired by time and location. However, when comparing the maximum predicted and observed concentrations unmatched by time and location, the models performed much better. For example, the average of the highest 25 predicted and observed concentrations (unpaired in location and time) were within a factor of two for six of the eight models evaluated (MESOPUFF, MESOPLUME, MESOPLUME, MTDDIS, ARRPA and RTM-II). The study concluded that the LRT models' observed tendency to over-predict the observed peak concentrations errs on the conservative side for regulatory applications. However, this over-prediction must be weighed against the general tendency of those models to underestimate horizontal spreading and to predict a plume pattern that is spatially offset from the observed data.

2.3.2 Rocky Mountain Acid Deposition Model Assessment Project – Western Atmospheric Deposition Task Force

A second round of LRT model evaluations was conducted as part of the Rocky Mountain Acid Deposition Model Assessment (EPA, 1990). In this study, the eight models from the 1986 evaluation were compared against a newer model, the Acid Rain Mountain Mesoscale Model (ARM3) (EPA, 1988). The statistical evaluation considered data paired in time/space and also unpaired in time/space equally. In this study, it was found that the MESOPUFF-II (Scire et al., 1984a, and 1984b) model performed best when using unpaired data, and that the ARM3 model performed best when using paired data. A final model score was assigned on the basis of a model's performance relative to the others in each of the areas (paired in time/space, unpaired in time/space, and paired in time, not space) for each of two tracer releases considered.

The primary objective was to assemble a mesoscale air quality model based primarily on models or model components available at the time for use by state and federal agencies to assess acid deposition in the complex terrain of the Rocky Mountains.

2.3.3 Comparison of CALPUFF Modeling Results to Two Tracer Field Experiments

The CALPUFF dispersion model (CALPUFF Version 4) was compared against tracer measurements from the GP80 and SRL75 field study experiments in a study conducted by James O. Paumier and Roger W. Brode (EPA, 1998a). The evaluation approach adopted the method used by Irwin (1997) that examined fitted predicted and observed plume concentrations across an arc of receptors. Meteorological inputs for the CALPUFF model were based on CALMET using observed surface and upper-air meteorological data. The study found that for these three tracer releases, there was overall agreement between the observed times and modeled times for both the time required for the plume to reach the receptor arc, as well as the time to pass completely by the arc. However, the transport direction had an angular offset. For the GP80 100 km arc, CALPUFF underestimated the lateral dispersion of the plume and overestimated the plume peak as well as the cross wind integrated concentration (CWIC) average concentrations across the plume; the lateral dispersion and CWIC were within a factor of two of the observed value and the CALPUFF fitted plume centerline concentrations was 2 to 2½ times greater than observed. Very different model performance was seen at the 600 km arc of receptors with simulated maximum and CWIC that were 2 to 2 ½ times lower than observed and lateral dispersion that was 2½ to 3½ times greater than observed.

2.3.4 ETEX and ATMES-II

After the Chernobyl accident in April 1986, the Atmospheric Transport Model Evaluation Study (ATMES) was initiated to compare the evolution of the radioactive cloud from Chernobyl with predictions by mathematical models for atmospheric dispersion, using as input the estimated source term and the meteorological data for the days following the accident. Considerable work was undertaken by ATMES in order to identify and make available the databases of radionuclide concentration in air measured after the Chernobyl accident and of meteorological conditions that occurred. The ATMES LRT dispersion modeling and model evaluation was conducted in the 1989-1990 time period. The performance of the LRT models to predict the observed radionuclides was hampered by the poor characterization of the emissions release from Chernobyl.

In May 1989, it was proposed to carry out a massive tracer experiment in Europe designed to address the weaknesses of ATMES modeling. In the following year the proposal was analyzed and modified to adapt it to the European context, and to take account of the ATMES results, as they became available. The experiment was named ETEX[16], European Tracer Experiment. It was designed to test the readiness of interested services to respond in the case of an emergency, to organize the tracer release and compile a data set of measured air concentrations and to investigate the performance of long range atmospheric transport and dispersion models using that data set.

The period 15 October-15 December 1994 was selected as the possible window for the two tracer experiments as part of ETEX. The first release started at 1600 UTC on October 23, 1994, and lasted 11 hours and 50 minutes. 340 kg of PMCH (perfluoromethylcyclohexane) tracer were released in Monterfil, France (48° 03’ 30’’ N, 2° 00’ 30’’ W) at an average flow rate of 8.0 g/s. The second ETEX tracer experiment started at 1500 UTC on November 14, 1994 and lasted for 9 hours and 45 minutes and released 490 kg of PMCP (perfluromethlcyclopentane) from Monterfil for an average release rate of 11.58 g/s.

16 http://rem.jrc.ec.europa.eu/etex/

The ETEX real-time LRT modeling phase was performed in parallel with the tracer field experiment. When the release started, 28 modeling groups were notified of the starting time, source location, and emission rate. They ran their LRT models in real-time to predict the evolution of the tracer cloud, and their predictions were sent as soon as they were available to the statistical evaluation team at JRC-Ispra. The capability of providing these predictions in real-time was considered to be an important factor, as well as the model performance itself. Therefore, only those institutions that had access to a meteorological model or that received real-time forecasts from a meteorological centre could participate.

The analysis of these calculations could not distinguish the differences between predictions and measurements arising from dispersion model inadequacies as opposed to those arising from the meteorological forecasts used. Almost two years after the ETEX releases, the ATMES-II modeling exercise was launched to evaluate the LRT models in hindcast mode. ATMES-II participants were required to calculate the concentration fields of the first ETEX tracer experiment using ECMWF analyzed meteorological data as input to their own dispersion models. Any institution operating a long-range dispersion model could now participate whether or not it had real-time access to the meteorological data, and the number of participants (49) was increased compared to the ETEX real-time modeling exercise, even though not all of the original ETEX modelers took part in ATMES-II.

Contrary to ETEX, the differences between the measured and modeled concentration fields in ATMES-II could be more directly related to the dispersion simulation, thanks to the use of the same meteorological fields. However, even in this case, discrepancies between models were due not only to the calculation of dispersion, but also to the different ways in which the meteorological information was used. Moreover, ATMES-II modelers could also submit results obtained with a meteorological analysis different from that of ECMWF.

As for the statistical analysis in ETEX real-time modeling exercise, the analysis of ATMES-II model results was divided into time, space and global analyses. The same statistical indices of the first ETEX release were computed in the time analysis, while for the other two analyses some different indices were computed following the requirements of modelers, and the experience gained during the two real-time exercises.

In a general, a substantial improvement in the models' performance in the ATMES-II modeling was seen compared to the ETEX real-time modeling phase for the common statistical indices.

When comparing the results of the ATMES-II statistical analysis with those for the real-time simulation of the first ETEX release, a general improvement of the model performances for those who took part in both exercises is evident. This can be explained by the better resolution of the meteorological fields used, the availability of the measured values of tracer concentration that allowed participants to tune some parameters in their long-range dispersion model and the time elapsed between the two exercises (2 years) during which improvements in model formulation and application procedures took place.

<u>Spatial Analysis</u>: In ATMES-II the spatial analysis consisted of the calculation of the Figure of Merit in Space (FMS) at 12, 24, 36, 48, 60 hours after the release start. The FMS is the ratio of the spatial distribution of the overlap of the predicted and observed tracer pattern to the union of the predicted and observed tracer pattern and is expressed as a percent (note that all statistical metrics are defined in detail in Section 2.4). A big improvement could be observed in the models' FMS compared to the ETEX real-time exercise for the first release. For instance, at 36 hours in ATMES-II all the models had a non-zero FMS, half of the models had FMS>45% and

a quarter of the models had FMS>55%, with a maximum FMS value of 71%. In ETEX, at 36 hours one tenth of the models had a zero FMS (i.e., no overlap of the predicted and observed tracer cloud) and a quarter had an FMS>45%, with a maximum FMS of 67%. At 60 hours in ATMES-II half of the models (against only a quarter of the models of ETEX) had a FMS>30% and the maximum FMS was 58%, while the maximum FMS for ETEX models was 52%.

Temporal Analysis: The temporal analysis was carried out at two arcs of receptors at distances of approximately 600 and 1,200-1,400 km from the release point. In general, the LRT models were better at predicting the time of arrival, duration and peak concentration of the tracer cloud for the central stations of the two arcs, and less satisfactory for the external stations. The Figure of Merit in Time (FMT, see Section 2.4 for definition) the best performances were observed for the central stations of the two arcs. For all the stations selected for the time analysis, FMT of models in ATMES-II improved when compared to the first ETEX release exercise.

Global Statistics: The global statistical indexes also indicate a general improvement of models' performance in ATMES-II compared to the ETEX real time modeling exercise. For instance, only eight models out of 49 (16%) had a bias higher than 0.4 ngm^{-3} (400 pg/m^3) in absolute value; the number of models above the same threshold in ETEX real time was 24 out of the 28 (86%) participants. Almost all models showed a satisfactory agreement with the measured values. However, few models were distinguished by a particularly good (or bad) performance in all respects. More than half of the models showed a relatively small error (NMSE), indicating a limited spread of the predictions around the corresponding measurements. Again, while in the ETEX real-time exercise only four models had an NMSE less than 100, 42 models were below this threshold in ATMES-II. Improvements compared to ETEX could also be seen in the number of predicted and observed pairs within a factor of 2 (FA2) and 5 (FA5) of each other; whereas in ATMES-II half of the models had FA5>45%, in ETEX no model reached that value. There was no negative Pearson correlation coefficient, with the best models showing values slightly less than 0.7.

Conclusions: The three main original objectives of ETEX as follows:

- to test the capability of institutes involved in emergency response to produce predictions of the cloud evolution in real-time;
- to evaluate the validity of their predictions; and
- to assemble a database that allows the evaluation of long-range atmospheric dispersion models.

The ETEX study has formulated the following conclusions:

- The objectives stated in the project design were met.
- ETEX demonstrated the feasibility of conducting a continental scale tracer experiment across Europe using the perfluorocarbon tracer technique.
- There is a large number of institutes that can (and will in the event of a real accident) predict the long-range atmospheric dispersion of a pollutant cloud.
- The rapidity of LRT dispersion modeling groups in predicting the tracer cloud evolution and transmitting the results to a central point was excellent.
- Regarding the quality of the predictions, differences between observations and calculations of 3 to 6 hours in arrival time and a factor of 3 in maximum airborne concentrations at ground level should be viewed as the best achievable with current LRT models.
- The simulation of cloud dispersion at short and mesoscale distances seems to have considerable influence on the long-range cloud development.
- The transition of the dispersion scales from local to long-range modeling should be investigated in more detail.
- ETEX assembled a unique experimental database of tracer concentrations and meteorological data accessible via the Internet.
- ETEX created widespread interest and resulted in considerable dispersion model development as well as the reinforcement of communication and collaboration between national institutes and international organizations.
- The ETEX network of national institutes and international organizations should be maintained and improved to continue model development and demonstrate the technical capability necessary to support emergency management in real cases.
- Further investigations are needed to determine the quality of predictions under complex meteorological conditions, and to quantify the uncertainty of models for emergency management.

2.3.5 Data Archive of Tracer Experiments and Meteorology (DATEM)

The Data Archive of Tracer Experiments and Meteorology (DATEM[17]) is not a single particular study but an archive of tracer experiment and meteorological data and suggested procedures for evaluating LRT dispersion models using atmospheric tracer data (Draxler, Heffter and Rolph, 2002). The DATEM archive currently incorporates data from five long-range dispersion experiments, which represent a collection of more than 19,000 air concentration samples, re-analysis fields from the National Center for Atmospheric Research (NCAR) / National Centers for Environmental Prediction (NCEP) re-analysis project, and statistical analysis programs based upon the ATMES-II evaluation of ETEX. All the emissions and sampling data are in space delimited text files, easily used by FORTRAN programs or imported into any spreadsheet. Meteorological data fields have been reformatted for use by HYSPLIT and are available for download. The statistical programs are all written in FORTRAN and include PC executables with the source code so that they can be compiled on other platforms.

The five long range transport tracer field experiments whose atmospheric and meteorological data reside on the DATEM website are as follows:

17 http://www.arl.noaa.gov/DATEM.php

ACURATE: The Atlantic Coast Unique Regional Atmospheric Tracer Experiment (ACURATE) operating during 1982-1983 and consisted of measuring $Krypton^{85}$ air concentrations from emissions out of the Savannah River Plant in South Carolina (Heffter et al., 1984). 12- and 24-hour average samples were collected for 19 months at five monitoring sites that were 300 to 1,000 km from the release point.

ANATEX: The Across North America Tracer Experiment (ANATEX) consisted of 65 releases of three types of Perflurocarbon Tracers (PFTs) that were released from Glasgow, Montana and St. Cloud, Minnesota over three months (January-March, 1987). The PFTs were measured at 75 monitoring sites covering the eastern U.S. and southeastern Canada (Draxler and Heffter, Eds, 1989).

CAPTEX: The Cross Appalachian Tracer Experiment (CAPTEX) occurred during September and October, 1983 and consisted of 4 PFT releases from Dayton, Ohio and 2 PFT releases from Sudbury, Ontario, Canada (Ferber et al., 1986). Sampling occurred at 84 sites from 300 to 800 km from the PFT release sites.

INEL74: The Idaho National Engineering Laboratory (INEL74) experiment consisted of releases of $Krypton^{85}$ during February-March, 1974 with sampling taken at 11 sites approximately 1,500 km downwind stretching from Oklahoma City to Minneapolis (Ferber et al., 1977; Draxler, 1982).

GP80: The 1980 Oklahoma City Great Plains (GP80) consisted of two releases of PFTs on July 8 and July 11, 1980. The first PFT release was sampled at two arcs at a distance 100 km and 600 km with 10 and 35 monitoring sites on each arc, respectively (Ferber et al., 1981). The second PFT release was only monitored at a distance of 100 km at the corresponding 10 sites from the July 8 release.

The DATEM website also includes a model evaluation protocol for evaluating LRT dispersion models using tracer field experiment that was designed following the procedures by Mosca et al. (1998) for the ATMES-II study and Stohl et al., (1998). The DATEM model evaluation protocol has four broad categories of model evaluation:

1. Scatter among paired measured and calculated values;
2. Bias of the calculations in terms of over- and under-predictions;
3. Spatial distribution of the calculation relative to the measurements; and
4. Differences in the distribution of unpaired measured and calculated values.

A recommended set of statistical performance measures are provided along with a FORTRAN program (statmain) to calculate them. The DATEM recommendations have been adopted in this study and more details on the DATEM recommended ATMES-II model evaluation approach is provided in section 2.4.3.

2.4 MODEL PERFORMANCE EVALUATION APROACHES AND METHODS

2.4.1 Model Evaluation Philosophy

To date, no specific guidance has been developed by the USEPA for evaluating LRT models. According to EPA's *Interim Procedures for Evaluating Air Quality Models (Revised)*, the rationale for selecting a particular data group combination depends upon the objective of the performance evaluation. For this it is necessary to translate the regulatory purposes of the intended use of the model into performance evaluation objectives (EPA, 1984; Britter, et al., 1995). Under the approach for both the 1986 and 1998 EPA LRT model evaluation projects, no particular emphasis was placed on any data group combination or set of statistical measures.

In this study we expand the LRT model performance philosophy to include spatial, correlation/scatter, bias, error and frequency distribution performance metrics.

In their regulatory use within the United States, LRT models are used to predict impacts of criteria pollutants for national ambient air quality standards (NAAQS) and Prevention of Significant Deterioration of Air Quality (PSD) Class I increments. Additionally, Federal Land Management Agencies rely upon the same LRT models in the PSD program for estimates of chemical transformation and removal to assess impacts on air quality related values (AQRV's) such as visibility and acid deposition. The chemistry of aerosol formation is highly dependent upon the spatial and temporal variability of meteorology (e.g., relative humidity and temperature) and precursors (e.g., ammonia).

Recognizing the need for developing an evaluation approach that reflects the intended regulatory uses of LRT models, the model performance evaluation approach of Mosca et al., (1998) and Stohl et al., (1998) used in the ATMES-II study and recommended by DATEM (Draxler, Heffter and Rolph, 2002) was adopted for this study.

We have also included elements of the plume fitting evaluation approach of Irwin (1997) for comparison with the results from the original 1998 tracer evaluation study (EPA, 1998a). The Irwin model evaluation approach is only applicable when you have an arc of receptors at a given distance downwind of the source so that a cross plume distribution and dispersion statistics can be generated. Whereas, the ATMES-II is more applicable when you have receptors spread over a large region and can calculate statistical parameters related to the predicted and observed distribution of the tracer concentrations. Accordingly, we use the Irwin plume fitting statistical evaluation approach for the GP80 and SRL75 tracer experiments whose receptors were defined along arcs at a given distance from the source and we used the ATMES-II statistical evaluation approach for the CAPTEX and ETEX tracer experiments that had receptors that were defined across a broad area.

2.4.2 Irwin Plume Fitting Model Evaluation Approach

Irwin (1997) focused his evaluation of the CALPUFF modeling system on its ability to replicate centerline concentrations and plume widths, with more emphasis placed upon these factors than data such as modeled/observed plume azimuth, plume arrival time, and plume transit time. The Great Plains and Savannah River tracer CALPUFF evaluations (EPA, 1998a) followed the tracer evaluation methodology of the Idaho National Engineering Laboratory (INEL) tracer study conducted on April 19, 1977 near Idaho Falls, Idaho (Irwin, 1997).

Irwin examined CALPUFF performance by calculating the cross-wind integrated concentration (CWIC), azimuth of plume centerline, and the second moment of tracer concentration (lateral dispersion of the plume [σ_y]). The CWIC is calculated by trapezoidal integration across average monitor concentrations along the arc. By assuming a Gaussian distribution of concentrations along the arc, a fitted plume centerline concentration (Cmax) can be calculated by the following equation:

$$\text{Cmax} = \text{CWIC}/[(2\pi)^{1/2}\sigma_y] \qquad (2\text{-}1)$$

The measure σ_y describes the extent of plume horizontal dispersion. This is important to understanding differences between the various dispersion options available in the CALPUFF modeling system. Additional measures for temporal analysis include plume arrival time and the plume transit time on arc. Table 2-2 summarizes the statistical metrics used in the Irwin fitted Gaussian plume evaluation methodology.

Table 2-2. Model performance metrics from Irwin (1997) and 1998 EPA CALPUFF Evaluation (EPA, 1998a).

Statistics	Description
Spatial	
Azimuth of Plume Centerline	Comparison of the predicted angular displacement of the plume centerline from the observed centerline on the arc
Plume Sigma-y	Comparison of the predicted and observed fitted plume widths (i.e., dispersion rate)
Temporal	
Plume Arrival Time	Compare the time the predicted and observed tracer clouds arrives on the receptor arc
Transit Time on Arc	Compare the predicted and observed residence time on the receptor arc
Performance	
Crosswind Integrated Concentration	Compares the predicted and observed average concentrations across the receptor arc (CWIC)
Observed/Calculated Maximum	Comparison of the predicted and observed fitted Gaussian plume centerline (maximum) concentrations (Cmax) and maximum concentration at any receptor along the arc Omax)

The measures employed by Irwin (1997) and EPA (1998a) provide useful diagnostic information about the performance of LRT modeling systems, such as CALPUFF, but they do not always lend themselves easily to spatiotemporal analysis or direct model intercomparison.

For tracer studies such as the Great Plains Tracer Experiment and Savannah River where distinct arcs of monitors were present, the Irwin plume fitting evaluation approach was used in this study.

2.4.3 ATMES-II Model Evaluation Approach

The model evaluation methodology employed for this study was designed following the procedures of Mosca et al. (1998) and Draxler et al. (2002). Mosca et al. (1998) defined three types of statistical analyses:

- Spatial Analysis: Concentrations at a fixed time are considered over the entire domain. Useful for determining differences spatial differences between predicted and observed concentrations.
- Temporal Analysis: Concentrations at a fixed location are considered for the entire analysis period. This can be useful for determining differences between the timing of predicted and observed tracer concentrations.
- Global Analysis: All concentration values at any time and location are considered in this analysis. The global analysis considers the distribution of the values (probability), overall tendency towards overestimation or underestimation of measured values (bias and error), measures of scatter in the predicted and observed concentrations and measures of correlation.

2.4.3.1 Spatial Analysis

To examine similarities between the predicted and observed ground level concentrations, the Figure of Merit in Space (FMS) is calculated at a fixed time and for a fixed concentration level. The FMS is defined as the ratio between the overlap of the measured (A_M) and predicted (A_P) areas above a significant concentration level and their union:

$$FMS = \frac{A_M \cap A_P}{A_M \cup A_P} \times 100\% \qquad (2\text{-}2)$$

The more the predicted and measured tracer clouds overlap one another, the greater the FMS values are. A high FMS value corresponds to better model performance, with a perfect model achieving a 100% FMS score.

Additional spatial performance measures of Probability Of Detection (POD), False Alarm Rate (FAR), and Threat Score (TS) are also used. Typically used as a method for meteorological forecast verification, these three interrelated statistics are useful descriptions of an air quality model's ability to spatially forecast a certain condition. The forecast condition for the model is the predicted concentration above a user-specified threshold (at the 0.1 ngm^{-3} (100 pgm^{-3}) level for ATMES-II study). In these equations:

- "a" represents the number of times a condition that has been forecast, but was not observed (false alarm)
- "b" represents the number of times the condition was correctly forecasted (hits)
- "c" represents the number of times the nonoccurrence of the condition is correctly forecasted (correct negative); and
- "d" represents the number of times that the condition was observed but not forecasted (miss).

The FAR (Equation 2-3) is described as a measure of the percentage of times that a condition was forecast, but was not observed. The range of the score is 0 to 1 or 0% to 100%, with the ideal FAR score of 0 or 0% (i.e., there are observed tracer concentrations at a monitor/time every time the model predicts there is a tracer concentration at that monitor/time).

$$FAR = \left(\frac{a}{a+b}\right) \times 100\% \qquad (2\text{-}3)$$

The POD is a statistical measure which describes the fraction of observed events of the condition forecasted was correctly forecasted. Equation 2-4 shows that POD is defined as the ratio of "hits" to the sum of "hits" and "misses." The range of the POD score is 0 to 1 (or 0%to 100%), with the ideal score of 1 (or 100%).

$$POD = \left(\frac{b}{b+d}\right) \times 100\% \qquad (2\text{-}4)$$

The TS (Equation 2-5) is described as the measure describing how well correct forecasts corresponded to observed conditions. The TS does not consider correctly forecasted negative conditions, but penalizes the score for both false alarms and misses. The range of the TS is the same as the POD, ranging from 0 to 1 (0% to 100%), with the ideal score of 1 (100%).

$$TS = \left(\frac{b}{a+b+d} \right) \times 100\% \quad (2\text{-}5)$$

2.4.3.2 Temporal Analysis

In Section 2.4.1 temporal statistics related to the timing of when the predicted and observed tracer arrives at a monitor or arc of monitors, its residence time over a monitor (or arc) and when the tracer leaves the monitor (or arc) were discussed. Another temporal analysis statistics is the Figure of Merit in Time (FMT), which is analogous to the FMS only it is calculated at a fixed location (*x*) rather than a fixed time as the FMS. The FMT evaluates the overlap between the measures (M) and predicted (P) concentration at location *x* and time t_j. The FMT is normalized to the maximum predicted or measured value at each time interval and is expressed as a percentage value in the same manner as the FMS (Mosca et al., 1998).

$$FMT(\bar{x}) = \frac{\sum_j \min\{M(\bar{x},t_j), P(\bar{x},t_j)\}}{\sum_j \max\{M(\bar{x},t_j), P(\bar{x},t_j)\}} \times 100\% \quad (2\text{-}6)$$

The FMT is sensitive to both differences between measured and predicted and any temporal shifts that may occur.

2.4.3.3 Global Analysis

Following Draxler et al. (2002), four broad categories were used for global analysis of model evaluation. These broad categories are: (1) scatter; (2) bias; (3) spatial distribution of predictions relative to measurements; and (4) differences in the distribution of unpaired measured and predicted values. One or more statistical measures are used from each of the four categories in the global analysis. These include the percent over-prediction, number of calculations within a factor of 2 and 5 of the measurements, normalized mean square error, correlation coefficient, bias, fractional bias, figure of merit in space, and the Kolmogorov-Smirnov parameter representing the differences in cumulative distributions (Draxler et al., 2002).

Factor of Exceedance: In the scatter category, better model performance is observed when the Factor of Exceedance (FOEX) measure is close to zero and FA2 (described next) has a high percentage. A high positive FOEX and high percentage of FA5 would indicate a model's tendency towards over-prediction when compared to observed values.

$$FOEX = \left[\frac{N_{(P_i > Ni_i)}}{N} - 0.5 \right] \times 100\% \quad (2\text{-}7)$$

Where, *N* in the numerator is the number of pairs when the prediction (*P*) exceeds the measurement (*M*) and the *N* in the denominator is the total number of pairs in the evaluation. In FOEX, all 0-0 pairs are excluded from the analysis. FOEX can range from -50% to +50% with a perfect model receiving a 0% value.

Factor of α (FAα): FAα represents the percentage of predicted values that are within a factor of α, where we have used α = 2 or 5. As with FOEX, in FAα all 0-0 pairs are excluded.

$$FA\alpha = \left[\frac{N(y - y0 = [x - x0]\alpha)}{N} \right] \times 100 \tag{2-8}$$

Normalized Mean Squared Error (NMSE): Normalized mean squared error is the average of the square of the differences divided by the product of the means. NMSE gives information about the deviations, but does not yield estimations of model over-prediction or under-prediction.

$$NMSE = \frac{1}{N\overline{P}\overline{M}} \sum (P_i - M_i)^2 \tag{2-9}$$

Pearson's Correlation Coefficient (PCC): Also referred to as the linear correlation coefficient, its value ranges between -1.0 and +1.0. A value of +1.0 indicates "perfect positive correlation" or having all pairings of (M_i, P_i) lay on straight line on a scatter diagram with a positive slope. Conversely, a value of -1.0 indicates "perfect negative correlation" or having all pairings of (M_i, P_i) lie on a straight line with a negative slope. A value of near 0.0 indicates the clear absence of relationship between the model predictions and observed values.

$$R = \frac{\sum_i (M_i - \overline{M}) \bullet (P_i - \overline{P})}{\left[\sqrt{\sum (M_i - \overline{M})^2} \right] \left[\sqrt{\sum (P_i - \overline{P})^2} \right]} \tag{2-10}$$

Fractional Bias (FB): Calculated as the mean difference in prediction minus observation pairings divided by the average of the predicted and observed values.

$$FB = 2\overline{B} / (\overline{P} + \overline{M}) \tag{2-11}$$

Kolmogorov-Smirnov Parameter (KS): The KS parameter is defined as the maximum difference between two cumulative distributions. The KS parameter provides a quantitative estimate where C is the cumulative distribution of the measured and predicted concentrations over the range of k. The KS is a measure of how well the model reproduces the measured concentration distribution regardless of when or where it occurred. The maximum difference between any two distributions cannot be more than 100%.

$$KS = Max |C(M_k) - C(P_k)| \tag{2-12}$$

RANK: Given the large number of metrics, a single measure describing the overall performance of a model could be useful. Stohl et al. (1998) evaluated many of the above measures and discovered ratio based statistics such as FA2 and FA5 were highly susceptible to measurement errors. Draxler proposed a single metric, which he calls RANK, which is the composite of one statistical measure from each of the four broad categories.

$$RANK = \left|R^2\right| + \left(1 - \left|FB/2\right|\right) + FMS/100 + \left(1 - KS/100\right) \qquad (2\text{-}13)$$

The final score, model rank (*RANK*), provides a combined measure to facilitate model intercomparison. RANK is the sum of four of the statistical measures for scatter, bias, spatial coverage, and the unpaired distribution. RANK scores range between 0.0 and 4.0 with 4.0 representing the best model ranking. Using this measure allows for direct intercomparison of models across each of the four broader statistical categories.

2.4.3.4 Treatment of Zero Concentration Data

One issue in the performance evaluation was how to treat zero concentration data. Mosca et al. (1998) filtered the ETEX observational dataset by only retaining non-zero data and zero data within two sample time intervals (6 hours) of the arrival and departure times of the tracer cloud along with any zero observations in between these two time points. Stohl (1998) employed a Monte Carlo approach by adding normally distributed “random errors” to the original values to test the sensitivity of certain statistical measures to zero or near zero values. Stohl (1998) identified that certain statistical parameters may be sensitive to small variations in measurements when using “zero” or near “zero” background concentration data. While the inclusion of “zero” data creates concern about the robustness of certain statistical measures, especially ratio based statistics, there was also concern that only examining model statistics at locations where the tracer cloud was observed provides a limited snapshot of a model’s performance at those locations, and did not offer any insight into a model that may show poorer performance by transporting emissions to incorrect locations or advection to correct locations at incorrect times.

While the arguments for “filtering” of data are valid, it is also important to consider additional statistical measures such as the FAR, POD, and TS where all zero data must be considered. All zero data was retained for inclusion in the spatial analysis, but was filtered for the global statistical analysis. The approach used in this project differs from the approach used by Draxler et al. (2001) in that all zero-zero pairs are considered in their analysis of HYSPLIT performance.

3.0 1980 GREAT PLAINS FIELD STUDY

3.1 DESCRIPTION OF 1980 GREAT PLAINS FIELD STUDY

LRT tracer test experiments were conducted in 1980 with the release of a perfluorocarbon and sulfur hexafluoride tracers from the National Oceanic and Atmospheric Administration (NOAA) National Severe Storms Laboratory (NSSL) in Norman, Oklahoma (Ferber et al., 1981). Two arcs of monitoring sites were used to sample the tracer plumes; an arc of 30 samplers with a 4-5 km spacing located approximately 100 km from the release point that sampled at 45 minute intervals and an arc of 38 samplers through Nebraska and Missouri located approximately 600 km from the release site that sampled at an hourly interval. Figure 3-1 displays the locations of the tracer release site and the monitoring sites on the arcs that are 100 km and 600 km downwind of the source. Two experiments were conducted, one on July 8, 1980 that included both the 100 km and 600 km sampling arcs and one on July 11, 1980 that only included the 100 km sampling arc. The July 8, 1980 tracer field experiment and subsequent Perfluoro-Dimethyl-cyclohexane (PDCH) observed concentrations were used in this model evaluation study. The PDCH tracer was released over a three-hour period from 1900-2200 GMT (1400-1700 CDT) on July 8, 1980 from an open field near the NOAA/NSSL.

3.2 MODEL CONFIGURATION AND APPLICATION

The CALPUFF modeling system uses a grid system consisting of an array of horizontal grid cells and multiple vertical layers. Two grids must be defined in the CALPUFF model, a meteorological grid and a computational grid. The meteorological grid defines the extent over which landuse, winds, and other meteorological variables are defined in the CALMET simulation. The computational grid defines the extent of the concentration calculations in the CALPUFF simulation, and is required to be identical to or a subset of the meteorological grid. For the GP80 simulations, the computational grid is defined to be identical to the meteorological grid. A third grid, the sampling grid, is optional, and is used by CALPUFF to define a rectangular array of receptor locations. The sampling grid must be identical to or a subset of the computational grid. It may also be nested inside the computational grid (i.e., several sampling grid cells per computational grid cell). For the GP80 applications, a sampling grid identical to the computational grid was used with a nesting factor of one (sampling grid cell size equal to the cell size of the computational grid).

To properly characterize the meteorology for the CALPUFF modeling system, a grid that spans, at a minimum, the distance between source and receptor is required. However, to allow for possible recirculation of puffs that may be transported beyond the receptors and to allow for upstream influences on the wind field, the meteorological and computational domains should be larger than this minimum.

The GP80 site is shown in Figure 3-1. Two arcs of monitors were deployed during the field experiment at 100 and 600 kilometers from the source. For this analysis, two separate modeling domains were defined for simulating tracer concentrations on the 100 km and 600 km receptor arcs. For the 100-kilometer arc, a grid extending approximately from 35º N to 36.5º N latitude and from 96º W to 98.5º W longitude was defined.

CALPUFF was operated for the July 8, 1980 GP80 tracer experiment using meteorological inputs based on CALMET and MMIF. For the CALPUFF simulations using CALMET, a UTM coordinate system was used to be consistent with past CALPUFF evaluations (Policastro et al., 1986; EPA, 1998a).

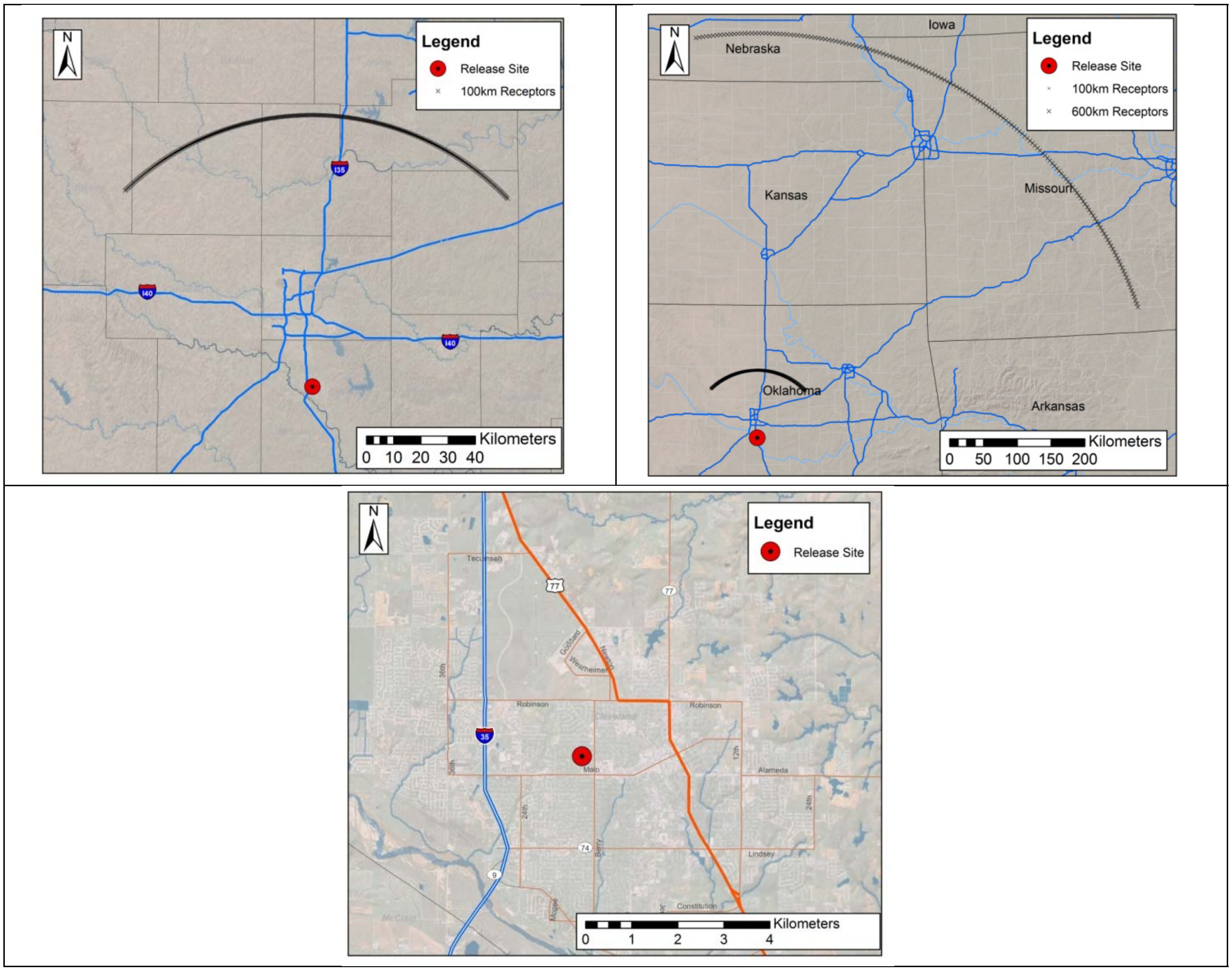


Figure 3-1. Locations of the release site and the 100 km arc (top left) and 600 km arc (top right) of monitoring sites along with a close in view of the release site (bottom) for the GP80 tracer experiment.

3.2.1 CALPUFF/CALMET BASE Case Model Configuration

For the CALPUFF/CALMET 100 km arc BASE case scenario, a 42 by 40 horizontal grid with a 10 km grid resolution was used for the meteorological and computational grids. For the 600 km arc BASE case, the grid extended from approximately 35^o N to 42^o N latitude and from 89^o W to 100^o W longitude using a 44 by 40 horizontal grid with a 20 km grid resolution. In addition, a 220 by 200 horizontal grid with a 4 km grid resolution was also used that encompassed both the 100 km and 600 km arcs.

To adequately characterize the vertical structure of the atmosphere, ten vertical layers were defined corresponding to layer heights at 0, 20, 40, 80, 160, 320, 640, 1,200, 2,000, 3,000 and 4,000 meters above ground level (AGL). The vertical layer structure conforms to the recommendations in EPA's August 2009 Clarification Memorandum on recommended settings for CALMET modeling (EPA, 2009b)

The CALMET preprocessor utilizes National Weather Service (NWS) meteorological data and on-site data to produce temporally and spatially varying three dimensional wind fields for CALPUFF. Only NWS data were used for this effort and came from two compact disc (CD) data sets. The first was the *Solar and Meteorological Surface Observation Network* (SAMSON) compact discs, which were used to obtain the hourly surface observations. The following surface stations were used for each of the field experiments:

Table 3-1. Surface meteorological monitoring sites used in the GP80 CALMET modeling.

State	City
Arkansas	Fort Smith
Illinois	Springfield
Kansas	Dodge City, Topeka, Wichita
Missouri	Columbia, Kansas City, Springfield, St. Louis
Nebraska	Grand Island, Omaha, North Platte
Oklahoma	Oklahoma City, Tulsa
Texas	Amarillo, Dallas-Fort Worth, Lubbock, Wichita Falls

Twice daily upper-air meteorological soundings came from the second set of compact discs, the *Radiosonde Data for North America*. The following stations were used for each of the field experiments:

Table 3-2. Radiosonde monitoring sites used in the GP80 CALMET modeling.

State	City
Arkansas	Little Rock
Illinois	Peoria
Kansas	Dodge City, Topeka
Missouri	Monett
Nebraska	Omaha, North Platte
Oklahoma	Oklahoma City
Texas	Amarillo

Consistent with the August 2009 Clarification Memorandum, some of the CALPUFF/CALMET sensitivity tests utilized CALMET simulations using prognostic meteorological model output as the first-guess wind field for CALMET and then perform the CALMET STEP1 procedures to apply diagnostic effects to the wind fields. CALMET then uses the surface and upper air observations in the objective analysis (OA) phase that blends the meteorological observations with the STEP1 wind field to produce the STEP2 wind field. This method is often referred to as the "hybrid" method.

The terrain and GIS land use data on the original CALPUFF CD were used to define gridded land use data for each field experiment. These data are defined with a resolution of 1/6° latitude and 1/4° longitude. The program PRELND1.EXE, also provided on the CD, was run to extract the data from the GIS data base and map the data to the meteorological domain for each field experiment. The program ELEVAT.EXE (also provided on the CD) was used to process the raw terrain data into average gridded terrain data. The file of terrain and geophysical parameters required by CALMET was constructed from the output files generated by ELEVAT and PRELND1 with additional required records inserted manually to create the final forms of the file for GP80 tracer experiment.

One of the primary purposes of the GP80 experiment was to demonstrate the efficacy of perfluorocarbons as tracers in atmospheric dispersion field studies. Perfluoromonomethylcyclohexane (PMCH) and perfluorodimethylcyclohexane (PDCH) were released during this experiment. For the 1998 EPA CALPUFF evaluation report and the current analyses, the PDCH emission rate was used in the CALPUFF evaluation since the monitoring data appeared to have a more complete record of PDCH concentrations than the other tracers. Table 3-3 displays the source characteristics for the PDCH tracer used in the CALPUFF modeling of the July 8, 1980 GP80 experiment.

Table 3-3. Source characteristics for the CALPUFF modeling of the July 8, 1980 GP80 experiment.

Source	Release height (m)	Stack diameter (m)	Exit velocity (m/s)	Exit temp. (°K)	Total tracer released (kg)	Length of release (hr)	PDCH emission rate (g s^{-1})
Oklahoma	10.0	1.0 [a]	0.001	Ambient [b] (250)	186	3.0	17.22

Notes:
a – The stack diameter was set to 1 meter in diameter to conform to previous tracer evaluation studies.
b – The exit temperature was assumed to be the same as ambient atmospheric temperature. CALPUFF checks the difference between the stack exit temperature and the surface station temperature. If this difference is less than zero, the difference is set to zero. To insure this condition, an exit temperature of 250 K was input to the model.

In the CALPUFF modeling system, each of the three programs (CALMET, CALPUFF, and CALPOST) uses a control file of user-selectable options to control the data processing. There are numerous options in each and several that can result in significant differences. The following model controls for CALMET and CALPUFF were employed for the analyses with the tracer data.

3.2.1.1 CALMET Options

The following CALMET control parameters and options were chosen for the BASE CALPUFF model simulations. The BASE control parameters and options were chosen to be consistent with two previous CALMET/CALPUFF evaluations (Irwin 1997, and EPA 1998a). The most important CALMET options relate to the development of the wind field and were set as follows for the BASE model configuration:

NOOBS	= 0	Use surface, overwater, and upper air station data
IWFCOD	= 1	Use diagnostic wind model to develop the 3-D wind fields
IFRADJ	= 1	Compute Froude number adjustment effects (thermodynamic blocking effects of terrain)
IKINE	= 1	Compute kinematic effects
IOBR	= 0	Do NOT use O'Brien procedure for adjusting vertical velocity
IEXTRP	= 4	Use similarity theory to extrapolate surface winds to upper layers
IPROG	= 0	Do NOT use prognostic wind field model output as input to diagnostic wind field model (for observations only sensitivity test)
ITPROG	= 0	Do NOT use prognostic temperature data output

Mixing heights are important in the estimating ground level concentrations. The CALMET options that affect mixing heights were set as follows:

IAVEZI	= 1	Conduct spatial averaging
MNDAV	= 3	**100km BASE case** – Maximum search radius (in grid cells) in averaging process
	= 1	**600km BASE Case**
HAFANG	= 30.	Half-angle of upwind looking cone for averaging

ILEVZI	= 1	Layer of winds to use in upwind averaging
DPTMIN	= .001	Minimum potential temperature lapse rate (K/m) in stable layer above convective mixing height
DZZI	= 200	Depth of layer (meters) over which the lapse rate is computed
ZIMIN	= 100	**100km BASE case** – Minimum mixing height (meters) over land
	= 50	**600km BASE Case**
ZIMAX	= 3200	**100km BASE case** – Maximum mixing height (meters) over land, defined to be the top of the modeling domain
	= 3000	**600km BASE Case**

A number of CALMET model control options have no default CALMET values, particularly radii of influence values for terrain and surface and upper air observations. The CALMET options that affect radius of influence were set as follows:

RMAX1	= 20	Minimum radius of influence in surface layer (km)
RMAX2	= 50	Minimum radius of influence over land aloft (km)
RMIN	= 2	**100km BASE case** – Minimum radius of influence in wind field interpolation (km)
	= 0.1	**600km BASE Case**
TERRAD	= 10	Radius of influence of terrain features (km)
RPROG	= 0	Weighting factors of prognostic wind field data (km)

A review of the respective CALMET parameters between the 1998 EPA CALMET/CALPUFF evaluation study using CALMET Version 4.0 and the 600 km BASE case scenario in the current CALMET/CALPUFF evaluation using CALMET Version 5.8 indicates differences in some CALMET options. The differences between the two scenarios are presented below in Table 3-4. All other major CALMET options for 600 km BASE case scenario matched the original 1998 EPA analysis. There were no significant differences between the CALMET parameters 100 km BASE case scenarios for the 1998 (CALMET Version 4.0) and the current evaluation (CALMET Version 5.8).

Table 3-4. CALMET Parameters July 8, 1980 GP80 experiment, 1998 and current 600 km analysis.

CALMET Option	Description	1998 EPA Setup	BASE Setup
MNDAV	Maximum search radius for averaging mixing heights (# grid cells)	3	1
ZIMIN	Minimum overland mixing height (in meters)	100	50
ZIMAX	Maximum overland mixing height (in meters)	3200	3000
RMIN	Minimum radius of influence in wind field interpolation (in km)	2.0	0.1

3.2.1.2 CALPUFF Control Options

The following CALPUFF control parameters, which are a subset of the control parameters, were used. These parameters and options were mostly chosen to be consistent with the 1977 INEL study (Irwin 1997) and 1998 EPA CALPUFF evaluation (EPA, 1998a) studies. This includes the use of the slug option (MSLUG = 1) for the 100 km arc CALPUFF simulations. The use of the slug option is very non-standard for LRT modeling and inconsistent with the EPA-FLM recommendations for far-field CALPUFF modeling. As stated on the CALPUFF website[18]:

> "A slug is simply an elongated puff. For most CALPUFF applications, the modeling of emissions as puffs is adequate. The selection of puffs produces very similar results as compared to the slug option, while resulting in significantly faster computer runtimes. However, there are some cases where the slug option may be preferred. One such case is the episodic time-varying emissions, e.g., an accidental release scenario. Another case would be where transport from the source to receptors of interest is very short (possibly involving sub-hourly transport times). These cases generally involve demonstration of causality effects due to specific events in the near- to intermediate-field."

For the farther out 600 km arc, the slug option was not selected (MSLUG = 0) for the initial CALPUFF sensitivity tests even through the slug option was used in the 1997 INEL and 1998 EPA studies. However, we did investigate the use of the slug option, as well as puff splitting, in a set of additional CALPUFF sensitivity tests for the 600 km arc.

CALPUFF options for technical options (group 2):

MCTADJ	= 0	No terrain adjustment
MCTSG	= 0	No subgrid scale complex terrain is modeled
MSLUG	= 1	**For 100 km BASE case** near-field puffs modeled as slugs
MSLUG	= 0	**For 600 km BASE case** modeled as puffs (i.e., no slugs)
MTRANS	= 1	Transitional plume rise is modeled
MTIP	= 1	Stack tip downwash is modeled
MSHEAR	= 0	**100 km BASE case** – Vertical wind shear is NOT modeled above stack top
	= 1	**600km BASE case**
MSPLIT	= 0	No puff splitting

18 **http://www.src.com/calpuff/FAQ-answers.htm**

MCHEM	= 0	No chemical transformations
MWET	= 0	No wet removal processes
MDRY	= 0	No dry removal processes
MPARTL	= 0	**100 km BASE case –** No partial plume penetration
	= 1	**600 km BASE case**
MPDF	= 0	**100 km BASE case –** PDF not used for dispersion under convective conditions
	= 1	**600 km BASE case**
MREG	= 0	No check made to see if options conform to regulatory Options

Two different values were used for the dispersion parameterization option MDISP:

= 2	Dispersion coefficients from internally calculated sigmas
= 3	PG dispersion coefficients for RURAL areas (PG)

In addition, under MDISP = 2 dispersion option, two different options were used for the MCTURB option that defines the method used to compute turbulence sigma-v and sigma-w using micrometeorological variables:

= 1	Standard CALPUFF routines (CAL)
= 2	AERMOD subroutines (AER)

Several miscellaneous dispersion and computational parameters (group 12) were set as follows:

SYTDEP	= 550.	Horizontal puff size beyond which Heffter equations are used for sigma-y and sigma-z
MHFTSZ	= 0	Do not use Heffter equation for sigma-z
XMXLEN	= 0.1	**100 km BASE case –** Maximum length of slug (in grid cells)
	= 1	**600 km BASE case**
XSAMLEN	= 0.1	**100 km BASE case –** Maximum travel distance of puff/slug (in grid cells) during one sampling step
	= 1	**600 km BASE case**
MXNEW	= 199	**100 km BASE case –** Maximum number of slugs/puffs released during one time step
	= 99	**600 km BASE case**
WSCALM	= 1.0	**100 km BASE case –** Minimum wind speed (m/s) for non-calm conditions
	= 0.5	**600 km BASE case**
XMAXZI	= 3300	**100 km BASE case –** Maximum mixing height (meters)
	= 6000	**600 km BASE case**
XMINZI	= 20	**100 km BASE case –** Minimum mixing height (meters)
	= 0	**600 km BASE case**
SL2PF	= 5	**100 km BASE case –** Slug-to-puff transition criterion factor (= sigma-y/slug length)
	= 10	**600 km BASE case**

A review of the respective CALPUFF parameters between the 1998 EPA CALMET/CALPUFF evaluation study using CALMET Version 4.0 and the 600 km BASE case scenario in the current CALMET/CALPUFF evaluation using CALPUFF Version 5.8 indicates differences in some parameters. The differences between the two scenarios are presented below in Table 3-5. All other major CALPUFF options for 600 km BASE case scenario matched the original 1998 EPA analysis. There were no significant differences between the CALPUFF parameters 100 km BASE case scenarios for the 1998 (CALPUFF Version 4.0) and the current evaluation (CALPUFF Version 5.8).

Table 3-5. CALPUFF Parameters July 8, 1980 GP80 experiment, 1998 and Current 600km analysis.

CALPUFF Option	Description	1998 EPA Setup	600KM BASE Setup
MSHEAR	Vertical wind shear is modeled above stack top? (0 = No; 1 = Yes)	0	1
MPARTL	Partial plume penetration of elevated inversion? (0 = No; 1 = Yes)	0	1
WSCALM	Minimum wind speed (m/s) for non-calm conditions	1.0	0.5
XMAXZI	Maximum mixing height (meters)	3300	3000
XMINZI	Minimum mixing height (meters)	20	0
XMXLEN	Maximum length of slug (in grid cells)	0.1	1
XSAMLEN	Maximum travel distance of puff/slug (in grid cells) during one sampling step	0.1	1
MXNEW	Maximum number of slugs/puffs released during one time step	199	99
SL2PF	Slug-to-puff transition criterion factor (= sigma-y/slug length)	5.0	10.0

3.2.2 GP80 CALPUFF/CALMET Sensitivity Tests

Table 3-6 and 3-7 describe the CALMET/CALPUFF sensitivity tests performed for the modeling of the 100 km and 600 km arcs of receptors. The BASEA simulations use the same configuration as used in the 1998 EPA CALPUFF evaluation report for the 100 km arc simulations, only updated from CALPUFF Version 4.0 to CALPUFF Version 5.8. For the 600 km arc simulations, the BASEA used the same configuration as the 1998 EPA study only the near-field slug option was not used. The CALMET and CALPUFF parameters of the BASE case simulations were discussed earlier in this section.

The sensitivity simulations are designed to examine the sensitivity of the CALPUFF model performance to choice of grid resolution in the CALMET meteorological model simulation (10 and 4 km for the 100 km arc of receptors and 20 and 4 km for the 600 km arc of receptors), the use of and resolution of the MM5 output data used as input to CALMET (none, 12 and 36 km) and the use of surface and upper-air meteorological observations in CALMET through NOOBS = 0 ("A" series, use surface and upper-air observation), 1 ("B" series, use only surface observations) and 2 ("C" series, don't use any meteorological observations).

In addition, for each experiment using different CALMET model configurations, three CALPUFF dispersion options were examined as shown in Table 3-8. Two of the CALPUFF dispersion sensitivity tests using dispersion based on sigma-v and sigma-w turbulence values using the CALPUFF (CAL) and AERMOD (AER) algorithms. Whereas the third dispersion test (PG) uses Pasquill-Gifford dispersion coefficients.

Table 3-6. CALPUFF/CALMET experiments for the 100 km arc and GP80 July 8, 1980 tracer experiment.

Experiment	CALMET Grid	MM5 Data	NOOBS	Comment
BASEA	10 km	None	0	Original met observations only configuration (no MM5)
EXP1A	10 km	12 km	0	Aug 2009 IWAQM w/10 km grid using 12 km MM5
EXP1B	10 km	12 km	1	Don't use observed upper-air meteorological data
EXP1C	10 km	12 km	2	Don't use observed surface/upper-air meteorological data
EXP2A	4 km	36 km	0	Aug 2009 IWAQM w/ 4 km grid and 36 km MM5
EXP2B	4 km	36 km	1	No upper-air meteorological data
EXP2C	4 km	36 km	2	No surface or upper-air meteorological data
EXP3A	4 km	12 km	0	Aug 2009 IWAQM w/ 4 km grid and 12 km MM5
EXP3B	4 km	12 km	1	No upper-air meteorological data
EXP3C	4 km	12 km	2	No surface or upper-air meteorological data

Table 3-7. CALPUFF/CALMET experiments for the 600 km arc and GP80 July 8, 1980 tracer experiment.

Experiment	CALMET Grid	MM5 Data	NOOBS	Comment
BASEA	20 km	None	0	Original met observations only configuration (no MM5)
EXP1A	20 km	12 km	0	Aug 2009 IWAQM recommendation using 12 km MM5
EXP1B	20 km	12 km	1	Don't use observed upper-air meteorological data
EXP1C	20 km	12 km	2	Don't use observed surface/upper-air meteorological data
EXP2A	4 km	36 km	0	Aug 2009 IWAQM w/ 4 km grid and 36 km MM5
EXP2B	4 km	36 km	1	No upper-air meteorological data
EXP2C	4 km	36 km	2	No surface or upper-air meteorological data
EXP3A	4 km	12 km	0	Aug 2009 IWAQM w/ 4 km grid and 12 km MM5
EXP3B	4 km	12 km	1	No upper-air meteorological data
EXP3C	4 km	12 km	2	No surface or upper-air meteorological data

Table 3-8. CALPUFF dispersion options examined in the CALPUFF sensitivity tests.

Experiment	MDISP	MCTURB	Comment
CAL	2	1	Dispersion coefficients from internally calculated sigma-v and sigma-w using micrometeorological variables and CALPUFF algorithms
AER	2	2	Dispersion coefficients from internally calculated sigma-v and sigma-w using micrometeorological variables and AERMOD algorithms
PG	3	--	PG dispersion coefficients for rural areas and MP coefficients for urban areas

The CALMET and CALPUFF simulations used for the sensitivity analyses were updated from the BASE case simulations and use the recommended settings for many variables from the EPA August 2009 Clarification Memorandum (EPA, 2009b). A summary of CALMET parameters that changed from the BASE case scenarios for the 100 km and 600 km CALPUFF sensitivity analyses are presented in Tables 3-9 and 3-10. The 100 km CALMET BASE case simulation (BASEA) matched up with the 1998 EPA study CALMET parameters, but did not match up with the EPA-FLM recommendations in the August 2009 Clarification Memorandum. Other than a few CALMET parameters, the 600 km CALMET BASE case simulation (BASEA) matched up well with August 2009 Clarification Memorandum, but not the 1998 EPA study CALMET parameters.

Table 3-9. CALMET wind field parameters for July 8, 1980 GP80 experiment, 100 km analysis.

CALMET Option	2009 EPA-FLM Default	BASEA	EXP1A	EXP1B	EXP1C	EXP2A	EXP2B	EXP2C	EXP3A	EXP3B	EXP3C
NOOBS	0	0	0	1	2	0	1	2	0	1	2
ICLOUD	0	0	0	0	3	0	0	3	0	0	3
IKINE	0	1	0	0	0	0	0	0	0	0	0
IEXTRP	-4	4	-4	-4	1	-4	-4	1	-4	-4	1
IPROG	14	0	14	14	14	14	14	14	14	14	14
ITPROG	0	0	0	1	2	0	1	2	0	1	2
MNDAV	1	3	1	1	1	1	1	1	1	1	1
ZIMIN	50	100	50	50	50	50	50	50	50	50	50
ZIMAX	3000	3200	3000	3000	3000	3000	3000	3000	3000	3000	3000
RMAX1	100	20	100	100	100	100	100	100	100	100	100
RMAX2	200	50	200	200	200	200	200	200	200	200	200
RMIN	0.1	2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
TERRAD	15	10	20	20	20	20	20	20	20	20	20
ZUPWND	1, 1000	1, 2000	1, 1000	1, 1000	1, 1000	1, 1000	1, 1000	1, 1000	1, 1000	1, 1000	1, 1000

Table 3-10. CALMET wind field parameters for July 8, 1980 GP80 experiment, 600 km analysis.

CALMET Option	2009 EPA-FLM Default	BASEA	EXP1A	EXP1B	EXP1C
NOOBS	0	0	0	1	2
ICLOUD	0	0	0	0	3
IKINE	0	1	0	0	0
IEXTRP	-4	4	-4	-4	1
IPROG	14	0	14	14	14
ITPROG	0	0	0	1	2
RMAX1	100	20	100	100	100
RMAX2	200	50	200	200	200
TERRAD	15	10	20	20	20

3.2.3 CALPUFF/MMIF Sensitivity Tests

With the MMIF software tool designed to pass through and reformat the MM5/WRF meteorological model output data for input into CALPUFF, there are not as many options available and hence much fewer sensitivity tests. Note that MMIF adopts the grid resolution and vertical layer structure of the MM5 model and passes through the meteorological variables to CALPUFF so only 36 km and 12 km grid resolutions were examined. The three alternative dispersion options in CALPUFF (CAL, AER and PG) were analyzed using the MMIF 12 km and 36 km CALPUFF inputs. Note that for the 600 km arc CALPUFF/MMIF modeling we found some issues in one of the CALPUFF runs using the AER dispersion option so do not present any AER dispersion results for the 600 km arc modeling; given the similarity in CALPUFF performance using the CAL and AER dispersion options this does not affect the study's results. In addition, 36 km CALPUFF/MMIF results are also not presented for the 600 km arc modeling.

Table 3-11. CALPUFF/MMIF sensitivity tests analyzed with the July 8, 1980 GP80 database.

Grid Resolution	MM5	MDISP	MCTURB	Comment
36 km	36 km	2	1	36 km MM5 with CALPUFF turbulence dispersion (CAL)
36 km	36 km	2	2	36 km MM5 with AERMOD turbulence dispersion (AER)
36 km	36 km	3		36 km MM5 with Pasqual-Gifford dispersion (PG)
12 km	12 km	2	1	12 km MM5 with CALPUFF turbulence dispersion (CAL)
12 km	12 km	2	2	12 km MM5 with AERMOD turbulence dispersion (AER)
12 km	12 km	3		12 km MM5 with Pasqual-Gifford dispersion (PG)

3.3 QUALITY ASSURANCE

The quality assurance (QA) of the CALPUFF modeling system simulations for the GP80 tracer experiment was assessed by analyzing the CALMET and CALPUFF input and output files and the dates they were generated. The input file options were compared against the August 2009 EPA-FLM recommended settings for CALMET and the definitions of the sensitivity tests to assure that the intended parameters were defined. The QA of the MMIF runs was not as complete because no input files or list files were provided to document the MMIF parameters. However, since all the MMIF tool does is pass through the MM5 output to CALPUFF there are not many options available.

The 100 km and 600 km receptor arc CALMET sensitivity simulations used a TERRAD value of 20 km (radius of influence of terrain on wind fields, in kilometers). The 2009 EPA-FLM clarification memorandum recommends that TERRAD = 15. Four CALMET parameters (BIAS, NSMTH, NINTR2, and FEXTR2) require a value for each vertical layer processed in CALMET. The 100 km and 600 km CALMET Base Cases are based on six vertical layers, but the sensitivity simulations are based on ten vertical layers. The CALMET sensitivity simulations were provided with only six values for BIAS, NSMTH, NINTR2, and FEXTR2 even though ten vertical layers were simulated. Therefore, CALMET used default values for the upper four vertical layers (1200 m, 2000 m, 3000 m, and 4000 m).

In addition to the three CALPUFF dispersion options (AERMOD, CALPUFF, and PG), there were other CALPUFF parameters that differed between the 100 km and 600 km CALPUFF/CALMET BASE case and sensitivity cases and CALPUFF/MMIF modeling scenarios. Differences in the CALPUFF parameters used in the 100 km and 600 km receptor arc simulation include:

- All of the CALPUFF 600 km sensitivity runs (CALPUFF/CALMET and CALPUFF/MMIF) and 100 km CALPUFF/MMIF runs were all conducted using only puffs (MSLUG = 0), but the 100 km CALPUFF/CALMET and 1998 CALPUFF simulations assume near-field slug formation (MSLUG = 1).
- CALPUFF 100 km CALPUFF/MMIF runs and all 600 km CALPUFF runs allowed for vertical wind shear (MSHEAR = 1), the 100 km BASE case and 100 km CALPUFF/CALMET sensitivity scenarios assume no vertical wind shear. The IWAQM Phase II (1998) guidance recommends MSHEAR = 0.
- The initial CALPUFF 100 km and 600 km sensitivity tests assumed no puff splitting (MSPLIT = 0), whereas the IWAQM Phase II (1998) recommends that default puff splitting be performed (MSPLIT = 1). This issue was investigated for the 600 km arc using additional CALPUFF sensitivity tests.

- CALPUFF 100 km (all dispersion options) and 600 km PG dispersion simulations, CALPUFF was set-up to not allow for partial plume penetration of inversion layer (MPARTL = 0). The IWAQM Phase II (1998) guidance recommends MPARTL = 1.
- CALPUFF 600 km AERMOD and CALPUFF turbulence dispersion simulations, CALPUFF was set-up to use the Probability Density Function (PDF) option for convective dispersion (MPDF = 1). The IWAQM Phase II guidance does not recommend using PDF for convective dispersion.
- CALPUFF 600 km simulations and 100 km CALPUFF/MMIF simulations use minimum and maximum mixing height values of 0 m and 6000 m, respectively. The CALPUFF 100 km BASE case and sensitivity simulations use minimum and maximum mixing height values of 20 m and 3300 m, respectively. The 1998 IWAQM Phase II guidance recommends the minimum and maximum mixing heights be set equal to 50 m and 3000 m, respectively.
- The CALPUFF 100 km BASE case and sensitivity simulations use a maximum slug length of 0.1 CALMET grid units (XMXLEN = 0.1), whereas the 100 km CALPUFF/MMIF simulations used a maximum length of 1.0 CALMET grid units. The IWAQM Phase II guidance recommends XMXLEN = 1.
- The CALPUFF 100 km BASE case and sensitivity simulations use a maximum slug/puff travel distance of 0.1 grid units per sampling period (XSAMLEN = 0.1), whereas the 100 km CALPUFF/MMIF simulations used a maximum travel distance of 1.0 grid units. The IWAQM Phase II guidance recommends XSAMLEN = 1.
- The CALPUFF 100 km BASE case and sensitivity simulations use a maximum of 199 slugs/puffs released from one source per sampling step (MXNEW = 199), whereas the 100 km CALPUFF/MMIF simulations used a maximum of 99 new slugs/puffs. The IWAQM Phase II guidance recommends MXNEW = 99.
- The CALPUFF 100 km BASE case and sensitivity simulations use a maximum of 5 sampling steps per slug/puff during one time step (MXSAM = 5), whereas the 100 km CALPUFF/MMIF simulations used a maximum of 99 sampling steps per slug/puff. The IWAQM Phase II guidance recommends MXSAM = 99.
- The CALPUFF 100 km BASE case and sensitivity simulations use a minimum sigma-y and sigma-z value of 0.01 m per new slug/puff (SYMIN = 0.01 and SZMIN = 0.01), whereas the 100 m CALPUFF/MMIF simulations used a minimum sigma-y and sigma-z value of 1 m per new slug/puff. The IWAQM Phase II guidance recommends SYMIN = 1 and SZMIN = 1.
- The CALPUFF 100 km BASE case and sensitivity simulations use a minimum wind speed of 1 m/s for non-calm conditions (WSCALM = 1), whereas the 100 km CALPUFF/MMIF simulations used a minimum wind speed of 0.5 m/s. The IWAQM Phase II guidance recommends WSCALM = 0.5.

We noted that the date on the CALMET input control file for the BASEA sensitivity test was later than the date on the CALMET output file for BASEA. We reran the BASEA CALMET and CALPUFF sensitivity tests and got slightly different results.

3.4 GP80 MODEL PERFORMANCE EVALUATION

Previous studies evaluated CALPUFF using the GP80 tracer experiment data using the Irwin plume fitting evaluation approach (EPA, 1998a). Thus, the same approach was adopted in this study so we could compare the performance of the newer version of CALPUFF with past evaluation studies and evaluate whether new options in CALPUFF (e.g., puff splitting) improve CALPUFF's model performance.

3.4.1 CALPUFF GP80 Evaluation for the 100 km Arc of Receptors

Table 3-12 evaluates the CALPUFF sensitivity tests ability to estimate the timing of the plume arrival at the 100 km arc of receptors and the duration of time the plume resides on the 100 km receptor arc. The tracer was observed on the 100 km arc for 5 hours. The 1998 EPA report CALPUFF modeling matched this well using CALPUFF turbulence (CAL) dispersion and estimated the tracer remained on the arc one hour longer than observed using the PG dispersion option. The CALPUFF/CALMET sensitivity tests estimated that the predicted tracer cloud was on the arc the same amount of time as was observed (5 hours) or within one hour of that duration (i.e., within ±20%). With one exception, when the CALPUFF/CALMET estimated that the duration of time on the arc was off by one hour, it was underestimating the amount of time on the arc (i.e., 4 instead of 5 hours). The exception to this was the EXP2A_PG scenario that estimates the tracer plume was on the 100 km arc for 6 hours.

The CALPUFF/MMIF sensitivity tests had the tracer plume arriving at the 100 km arc one hour late and either leaving on time (12 km MMIF) or leaving an hour early. This results in the CALPUF/MMIF sensitivity test underestimating the observed time on the arc by 1 (12 km MMIF) to 2 (36 km MMIF) hours.

Table 3-12. Tracer plume arrival and duration statistics for the GP80 100 km arc.

Scenario	Arrival on Arc		Leave Arc		Duration on Arc	
	Day	Hour	Day	Hour	Hours	Difference
Observed	190	16	190	20	5	
1998 EPA Report						
1998EPA_PG	190	16	190	21	6	20%
1998_CAL	190	16	190	20	5	0%
CALPUFF/CALMET						
BASEA_AER	190	16	190	20	5	0%
BASEA_CAL	190	16	190	20	5	0%
BASEA_PG	190	16	190	20	5	0%
EXP1A_AER	190	16	190	20	5	0%
EXP1A_CAL	190	16	190	20	5	0%
EXP1A_PG	190	16	190	20	5	0%
EXP1B_AER	190	16	190	19	4	-20%
EXP1B_CAL	190	16	190	19	4	-20%
EXP1B_PG	190	16	190	19	4	-20%
EXP1C_AER	190	17	190	20	4	-20%
EXP1C_CAL	190	17	190	20	4	-20%
EXP1C_PG	190	17	190	20	4	-20%
EXP2A_AER	190	16	190	20	5	0%
EXP2A_CAL	190	16	190	20	5	0%
EXP2A_PG	190	16	190	21	6	20%
EXP2B_AER	190	16	190	19	4	-20%
EXP2B_CAL	190	16	190	19	4	-20%
EXP2B_PG	190	16	190	20	5	0%
EXP2C_AER	190	17	190	20	4	-20%
EXP2C_CAL	190	17	190	20	4	-20%
EXP2C_PG	190	17	190	20	4	-20%
EXP3A_AER	190	16	190	20	5	0%
EXP3A_CAL	190	16	190	20	5	0%
EXP3A_PG	190	16	190	20	5	0%
EXP3B_AER	190	16	190	20	5	0%
EXP3B_CAL	190	16	190	20	5	0%
EXP3B_PG	190	16	190	19	4	-20%
EXP3C_AER	190	17	190	20	4	-20%
EXP3C_CAL	190	17	190	20	4	-20%
EXP3C_PG	190	17	190	20	4	-20%
CALPUFF/MMIF						
MMIF12_AER	190	17	190	20	4	-20%
MMIF12_CAL	190	17	190	20	4	-20%
MMIF12_PG	190	17	190	20	4	-20%
MMIF36KM_AER	190	17	190	19	3	-40%
MMIF36KM_CAL	190	17	190	19	3	-40%
MMIF36KM_PG	190	17	190	19	3	-40%

Tables 3-13 and Figures 3-2 through 3-6 display the plume fitting model performance statistics for the various CALPUFF sensitivity tests and the 100 km arc of receptors in the GP80 field experiment and compares them with the previous results as reported by EPA (1998a). The fitted predicted and observed plume centerline concentrations (Cmax) and the percent differences, expressed as a mean normalized bias (MNB), are shown in Table 3-13 with the MNB results reproduced in Figure 3-2. Similar results are seen for the predicted and observed maximum concentrations at any monitoring site along the arc (Omax) that are shown in Table 3-13 and Figure 3-3. The use of either the CALPUFF (CAL) or AERMOD (AER) algorithms for the turbulence dispersion doesn't appear the affect the maximum concentration model performance. Most CALPUFF sensitivity simulations overestimate the observed Cmax value by over 40%, with the 1998EPA_PG and EXP2C_PG simulations overestimating the observed Cmax value by over a factor of 2 (> 100%). The overestimation of the observed Omax value is even greater, exceeding 60% for most of the CALPUFF simulations. The PG dispersion produces much higher maximum concentrations compared to CAL/AER dispersion for experiments EXP2B and EXP2C. But the PG maximum concentrations are comparable or even a little lower than CAL/AER for the other experiments; although in the 1998 EPA study the PG dispersion option produced much higher maximum concentrations. The EXP1B, EXP2B and EXP3B CALPUFF simulations do not exhibit the large overestimation bias of Cmax and Omax as seen in the other experiments and are closest to reproducing the observed maximum concentrations on the 100 km arc, matching the observed values to within ±25%; note that the "B" series of experiments use MM5 data (12, 36 and 12 km for EXP1, EXP2 and EXP3, respectively) but only surface and no upper-air meteorological observations. The CALPUFF/MMIF simulation using the 12 km MM5 data and PG dispersion also reproduced the maximum concentrations to within ±25%.

Most of the CALPUFF sensitivity simulations underestimate the plume spread (σ_y) by 20% to 35% (Figure 3-4), which is consistent with overestimating the observed maximum concentration (i.e., insufficient dispersion leading to overestimation of the maximum concentrations). The exceptions to this are again the "B" series of CALPUFF/CALMET experiments and MMIF12KM_PG. Another exception to this is the EPA1998_PG simulation which agrees with the observed plume spread amount quite well; the explanation for this is unclear and seems inconsistent with the fact that 1998BASE_PG overestimated the observed Cmax/Omax values. The 1998EPA_PG results were taken from the EPA (1998a) report and could not be verified or quality assured so we cannot explain this discrepancy.

The deviations between the observed and predicted plume centerline along the 100 km arc of receptors in degrees is shown in Figure 3-5. The modeled plume centerline tends to be 0 to14 degrees off from the observed plume centerline. The best performing model configuration for the plume centerline location is the BASEA series that uses CALMET with observed surface and upper-air meteorological data but no MM5 data. The CALPUFF/CALMET sensitivity tests that use surface and upper-air ("A" series) and just surface ("B" series) meteorological observations tend to perform best for the plume centerline location, whereas the sensitivity tests that uses no meteorological observations ("C" series) performs the worse, with the plume centerline tending to be 10 to 14 degrees too far west on the 100 km arc for the "C" series of CALPUFF/CALMET sensitivity tests. The CALPUFF/MMIF runs, which also do not include any meteorological observations, also tend to have plume centerlines that are 6 to 12 degrees too far to the west.

Most of the CALPUFF sensitivity tests have cross wind integrated concentrations (CWIC) that are within ±20% of the observed value along the 100 km arc (Figure 3-6 and Table 5-13). The exceptions to this are the EPA1998_PG simulation, the BASEA series of simulations, EXP2A_PG, EXP2B_PG and EXP2C_PG. In general, the CAL and AER CALPUFF dispersion options are performing much better for the CWIC statistics along the 100 km arc than the PG dispersion option.

Table 3-13. CALPUFF model performance statistics using the Irwin plume fitting evaluation approach for the GP80 100 km arc of receptors, the EPA 1998 CALPUFF V4.0 modeling and the CALPUFF sensitivity tests.

CALPUFF Sensitivity Test	Cmax		Omax		Sigma-y		Plume Centerline		CWIC	
	(ppt)	MNB	(ppt)	MNB	(m)	MNB	(degrees)	Diff	(ppt-m)	MNB
Observed	**1.287**		**1.052**		**9,059**		**361.0**		**29,220**	
EPA 1998										
PG	2.700	110%	2.600	147%	9,000	-1%	357.0	-4.0	61,000	109%
Similarity	1.900	48%	1.800	71%	6,900	-24%	360.0	-1.0	33,000	13%
CALPUFF/CALMET										
BASEA_AER	2.221	73%	2.040	94%	7,136	-21%	361.4	0.4	39,720	36%
BASEA_CAL	2.214	72%	2.034	93%	7,165	-21%	361.4	0.4	39,770	36%
BASEA_PG	2.126	65%	1.934	84%	8,827	-3%	359.8	-1.2	47,050	61%
EXP1A_AER	2.086	62%	2.045	94%	5,977	-34%	357.1	-3.9	31,260	7%
EXP1A_CAL	2.088	62%	2.046	94%	5,999	-34%	357.0	-4.0	31,390	7%
EXP1A_PG	1.885	46%	1.839	75%	6,438	-29%	358.3	-2.7	30,420	4%
EXP1B_AER	1.407	9%	1.303	24%	8,492	-6%	358.8	-2.2	29,940	2%
EXP1B_CAL	1.414	10%	1.313	25%	8,478	-6%	358.8	-2.2	30,050	3%
EXP1B_PG	1.291	0%	1.217	16%	8,956	-1%	359.7	-1.3	28,980	-1%
EXP1C_AER	1.979	54%	1.937	84%	6,587	-27%	348.1	-12.9	32,670	12%
EXP1C_CAL	1.988	54%	1.945	85%	6,590	-27%	348.0	-13.0	32,840	12%
EXP1C_PG	2.016	57%	1.983	88%	6,041	-33%	349.4	-11.6	30,530	4%
EXP2A_AER	2.047	59%	1.996	90%	6,209	-31%	357.2	-3.8	31,860	9%
EXP2A_CAL	2.049	59%	1.999	90%	6,236	-31%	357.1	-3.9	32,020	10%
EXP2A_PG	2.013	56%	2.260	115%	11,330	25%	351.2	-9.8	57,180	96%
EXP2B_AER	1.265	-2%	1.145	9%	9,033	0%	359.4	-1.6	28,630	-2%
EXP2B_CAL	1.269	-1%	1.152	10%	9,030	0%	359.4	-1.6	28,710	-2%
EXP2B_PG	1.811	41%	2.034	93%	9,161	1%	357.6	-3.4	41,590	42%
EXP2C_AER	2.138	66%	2.106	100%	6,021	-34%	350.8	-10.2	32,270	10%
EXP2C_CAL	2.144	67%	2.112	101%	6,026	-33%	350.7	-10.3	32,380	11%
EXP2C_PG	2.938	128%	2.897	175%	6,044	-33%	349.4	-11.6	44,510	52%
EXP3A_AER	2.042	59%	1.992	89%	6,212	-31%	356.7	-4.3	31,800	9%
EXP3A_CAL	2.048	59%	1.998	90%	6,238	-31%	356.5	-4.5	32,030	10%
EXP3A_PG	1.827	42%	1.766	68%	6,805	-25%	358.0	-3.0	31,160	7%
EXP3B_AER	1.274	-1%	1.228	17%	8,928	-1%	357.9	-3.1	28,520	-2%
EXP3B_CAL	1.297	1%	1.247	19%	8,828	-3%	357.8	-3.2	28,700	-2%
EXP3B_PG	1.011	-21%	1.140	8%	11,010	22%	359.7	-1.3	27,900	-5%
EXP3C_AER	1.949	51%	1.911	82%	6,612	-27%	347.4	-13.6	32,300	11%
EXP3C_CAL	1.965	53%	1.927	83%	6,615	-27%	347.3	-13.7	32,590	12%
EXP3C_PG	1.999	55%	1.971	87%	6,085	-33%	349.0	-12.0	30,500	4%
CALPUFF/MMIF										
MMIF12KM_AER	1.872	45%	1.836	75%	6,811	-25%	349.5	-11.5	31,970	9%
MMIF12KM_CAL	1.897	47%	1.860	77%	6,805	-25%	349.3	-11.7	32,350	11%
MMIF12KM_PG	1.468	14%	1.318	25%	9,574	6%	350.3	-10.7	35,230	21%
MMIF36KM_AER	1.837	43%	1.811	72%	6,788	-25%	353.2	-7.8	31,250	7%
MMIF36KM_CAL	1.860	45%	1.832	74%	6,768	-25%	353.1	-7.9	31,550	8%
MMIF36KM_PG	1.608	25%	1.567	49%	7,055	-22%	355.1	-5.9	28,440	-3%

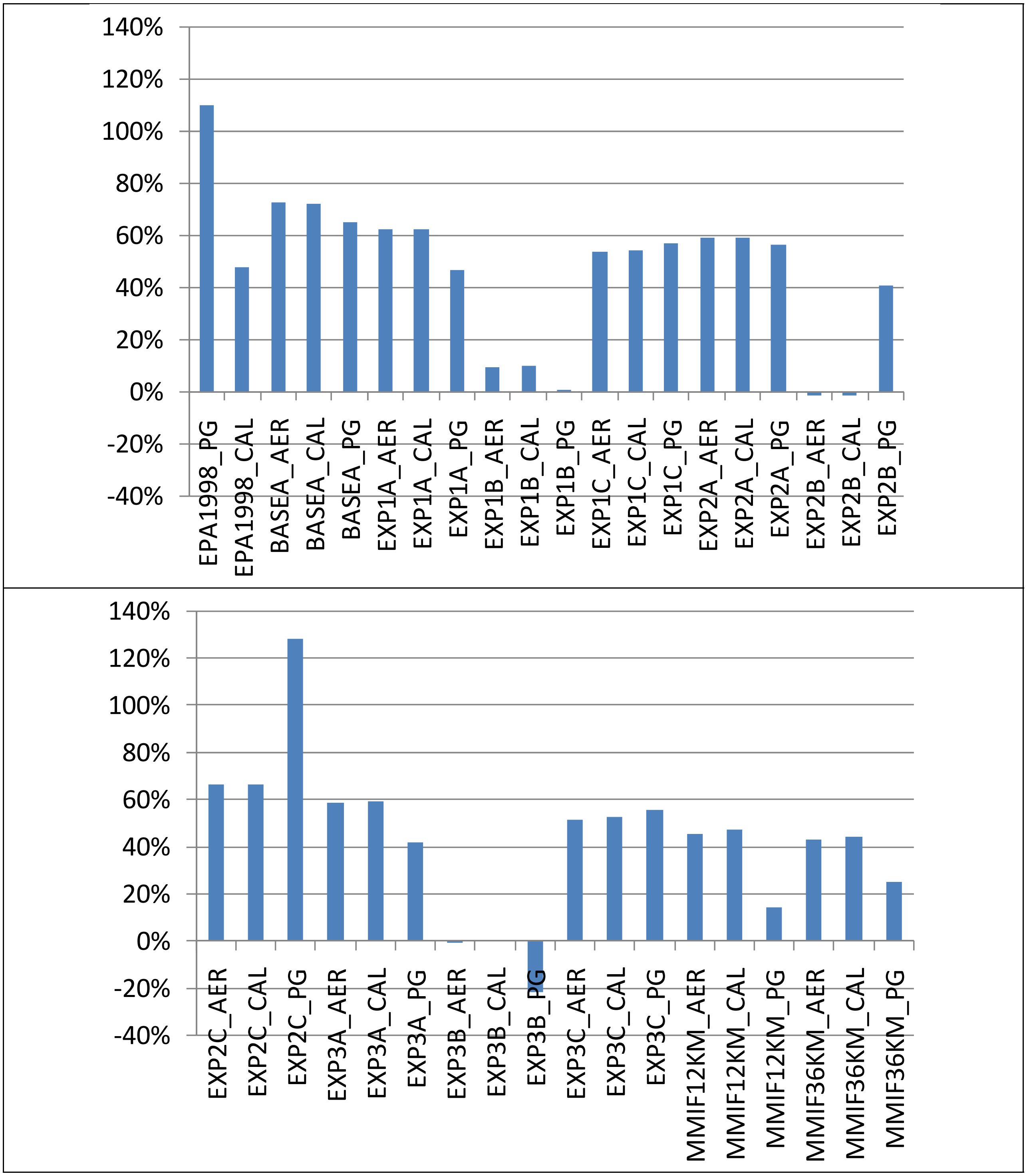


Figure 3-2. Percent difference (mean normalized bias) between the predicted and observed fitted plume centerline concentration (Cmax) for GP80 100 km receptor arc and the CALPUFF sensitivity tests.

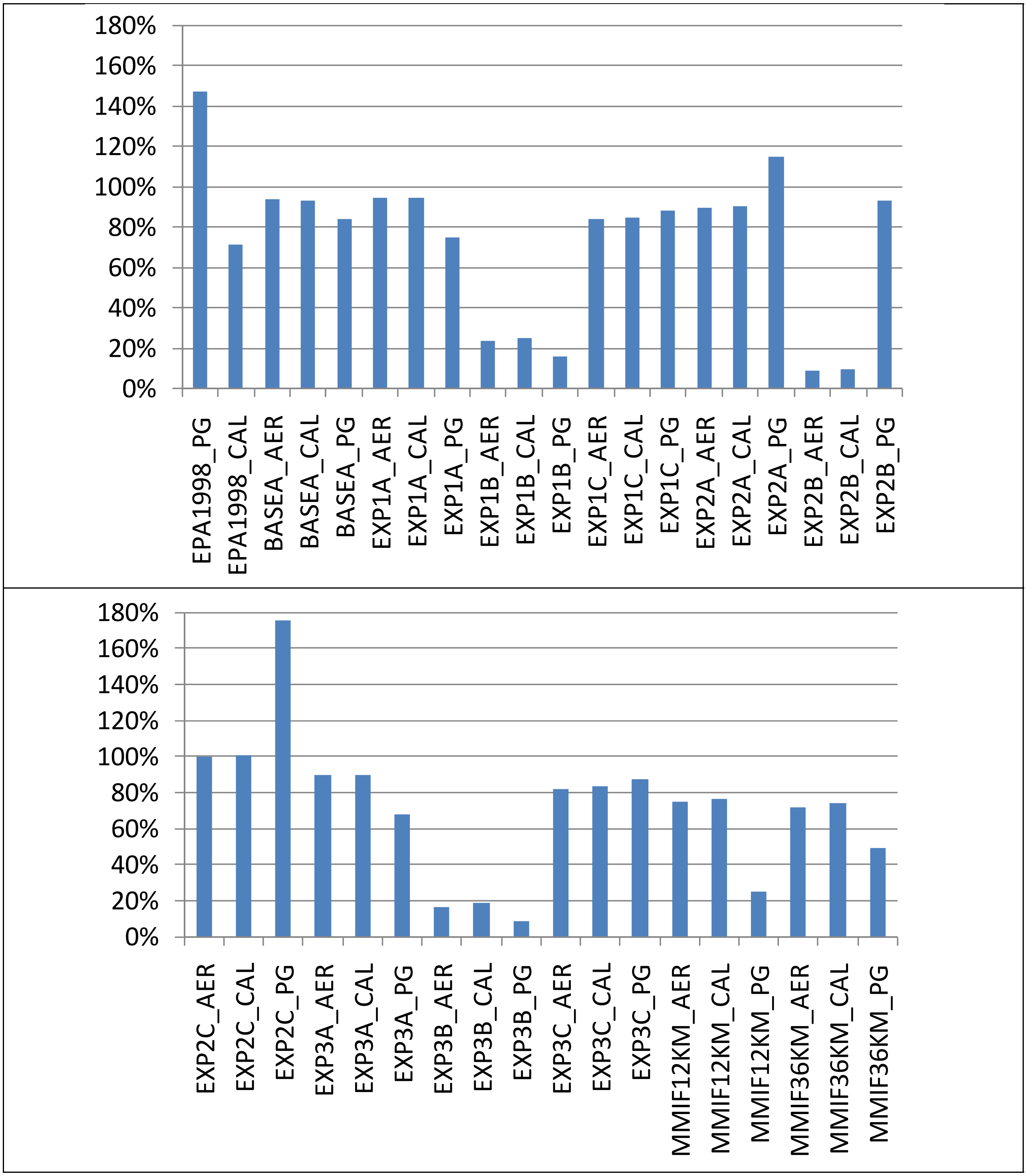


Figure 3-3. Percent difference (mean normalized bias) between the predicted and observed maximum concentration at any receptor/monitor (Omax) for GP80 100 km receptor arc and the CALPUFF sensitivity tests.

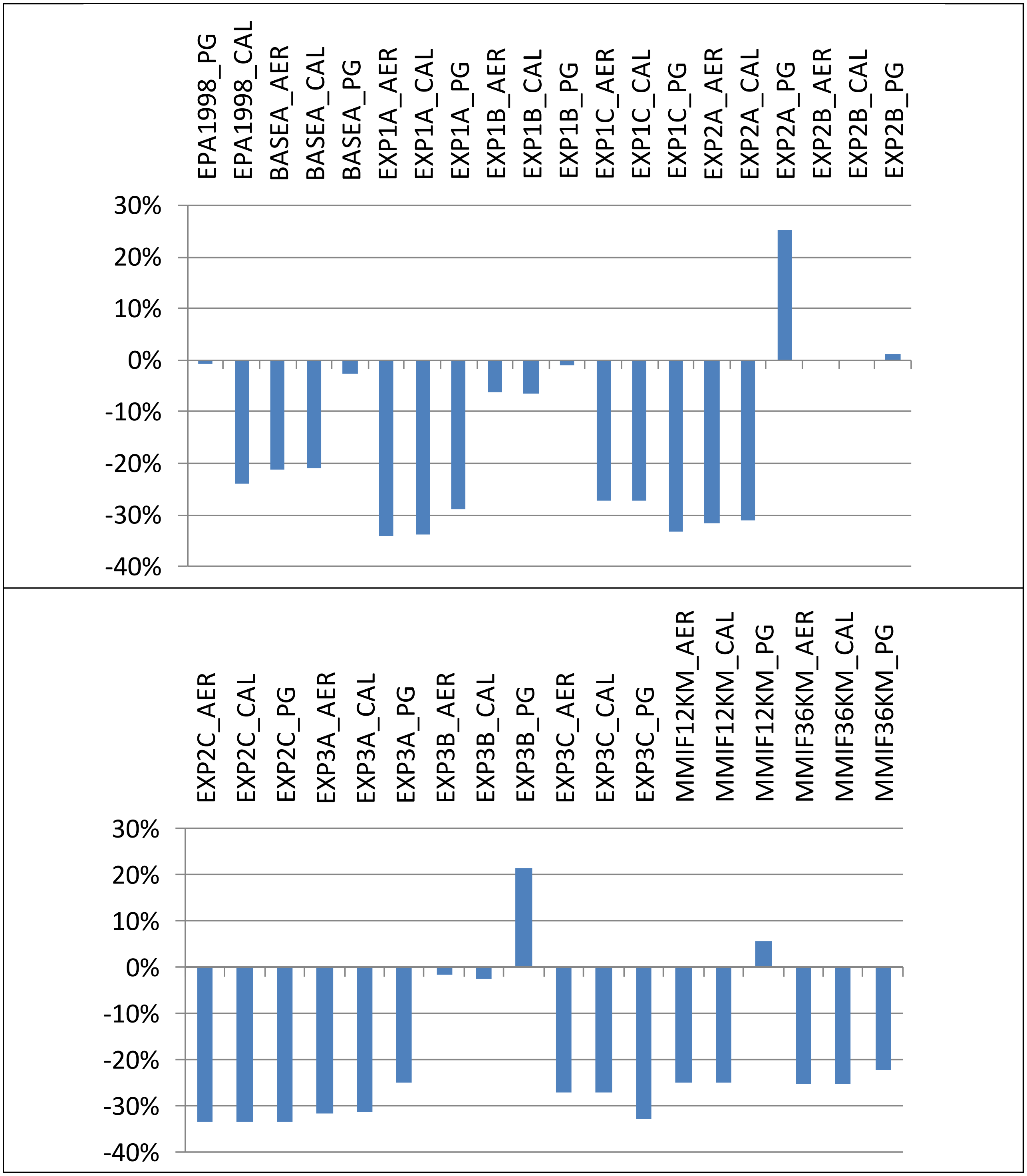


Figure 3-4. Percent difference (mean normalized bias) between the predicted and observed plume spread (σ_y) for GP80 100 km receptor arc and the CALPUFF sensitivity tests.

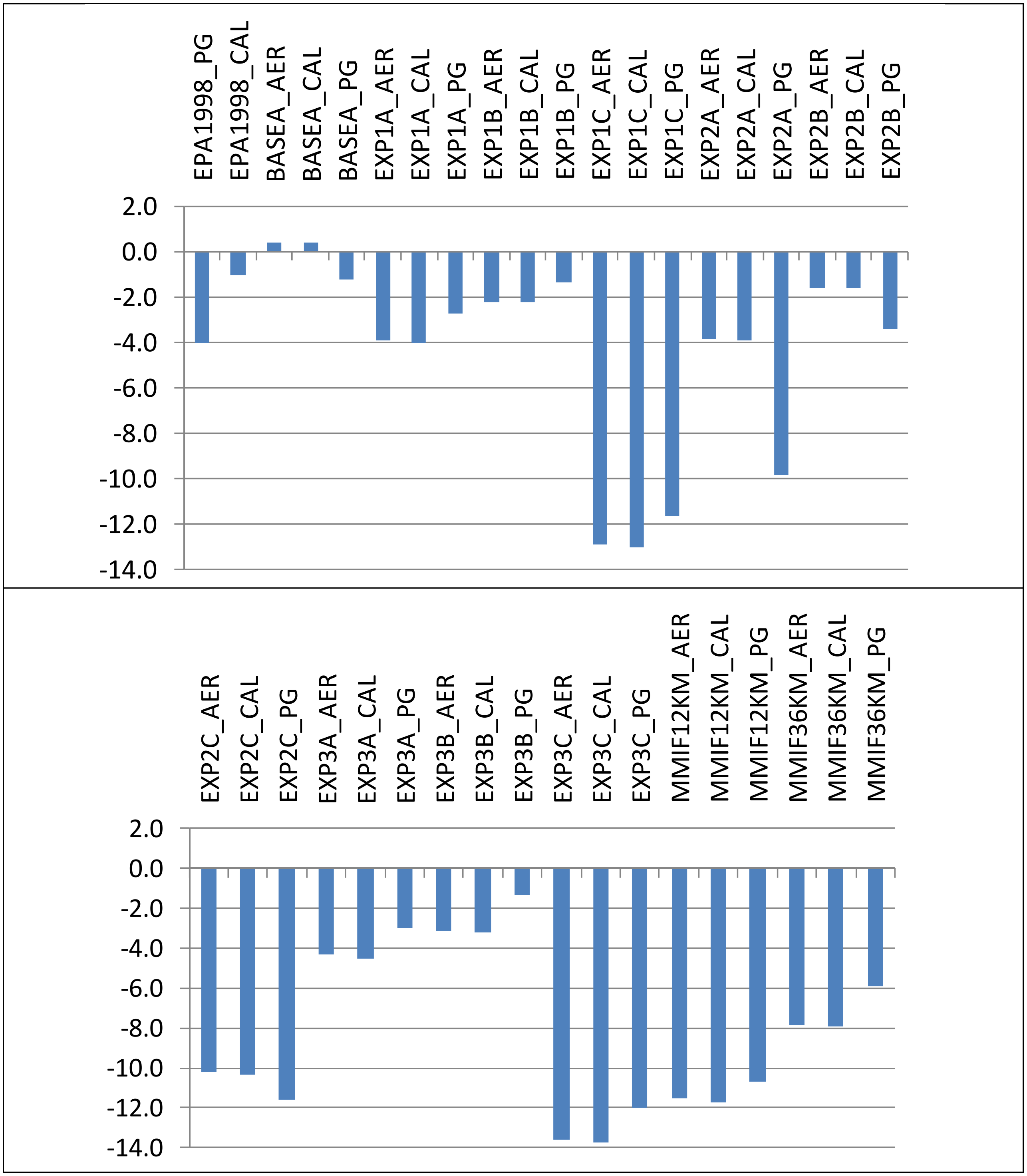


Figure 3-5. Difference in predicted and observed location of plume centerline (degrees) for the GP10 100 km receptor arc and the CALPUFF sensitivity tests.

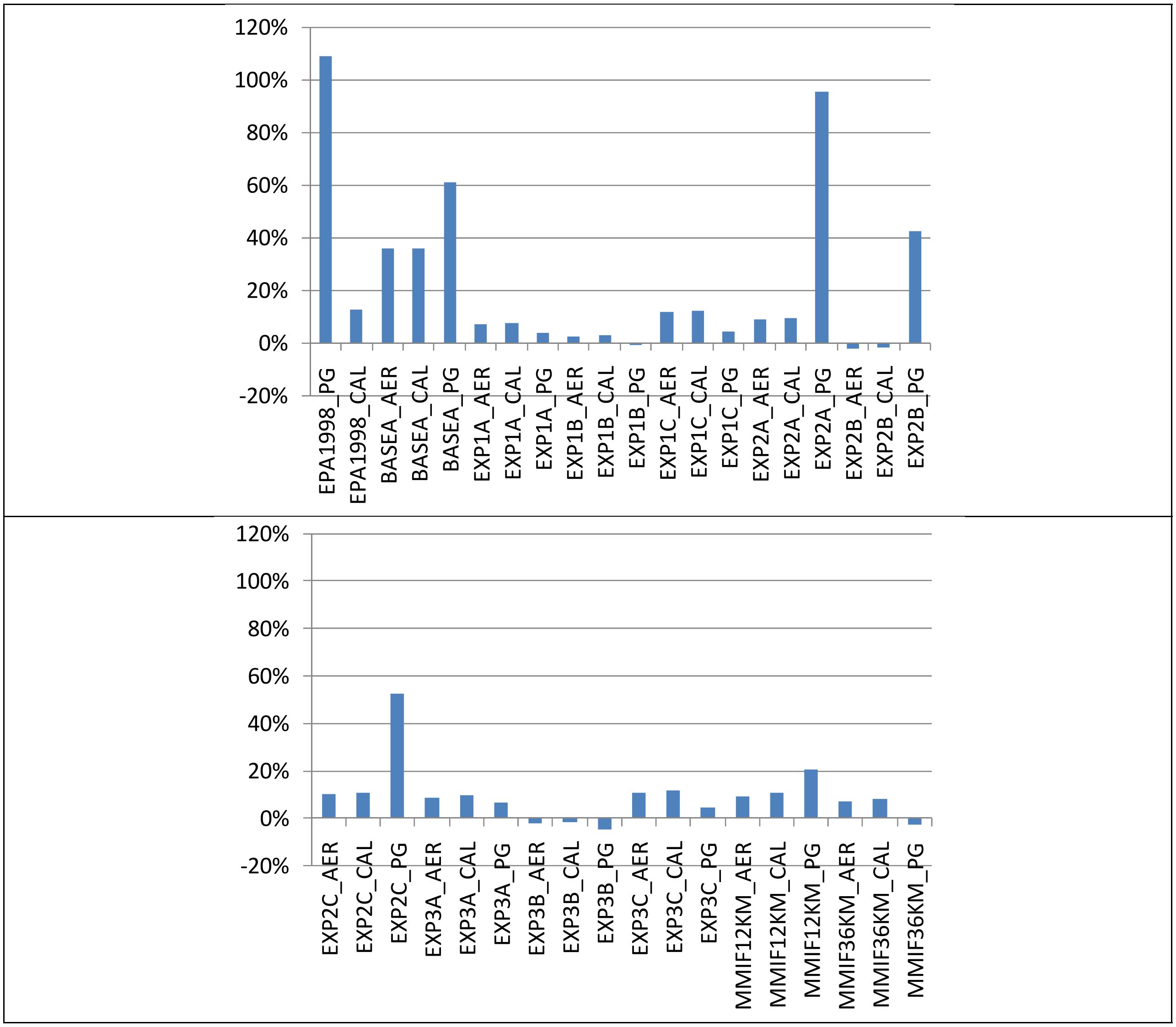


Figure 3-6. Percent difference (mean normalized bias) between the predicted and observed cross wind integrated concentration (CWIC) for the GP10 100 km receptor arc and the CALPUFF sensitivity tests.

3.4.2 CALPUFF GP80 Evaluation for the 600 km Arc of Receptors

Table 3-14 lists the predicted and observed plume arrival and exit time statistics from the 600 km arc of receptors and the duration of time the tracer resides on the 600 km arc for the initial CALPUFF sensitivity tests. Note that the observed tracer was found on the 600 km arc during the first sampling period (hour 2 on Julian Day 191) so the observed tracer may have arrived earlier than that. As explained by EPA (1998a), the observed tracer arrived earlier than expected due to the presence of a low-level jet that was not anticipated. Thus, the observed 12 hour tracer duration on the 600 km receptor arc that assumes it arrived during the first sampling interval at hour 2 could be an underestimate of the actual tracer residence time on the 600 km arc.

Figure 3-7 displays the percent differences in the tracer duration time on the 600 km arc for the initial CALPUFF 600 km sensitivity tests. For most of the initial CALPUFF sensitivity tests, the tracer duration time on the 600 km receptor arc is approximately half (5-6 hours) of what was observed (12 hours). This is in contrast to the 1998 EPA CALPUFF evaluation runs that overstate the duration the tracer resides on the 600 km arc, with values of 14 hours (1998EPA_PG) and 13 hours (1998EPA_CAL). Since the 1998 EPA CALPUFF runs estimated that the tracer arrives after the sampling started (hour 3), then this is a true overstatement of the tracer residence time and not an artifact of the tracer sampling starting after, or at the same time, the observed tracer arrived at the arc. There are a couple exceptions to the initial CALPUFF simulations performed in this study that understated the observed tracer duration on the arc by approximately a factor of 2, which are discussed below.

> The BASEA_PG scenario estimates that the tracer is on the arc for 12 hours, the same as the observed. However, it estimates the tracer leaves three hours earlier (hour 14) than observed (hour 11). Why the BASEA_PG tracer plume time statistics are so different from the two companion turbulence dispersion CALPUFF sensitivity tests (BASEA_CAL and BASEA_AER) is unclear. The same meteorological fields were used in the three BASEA CALPUFF sensitivity tests and the only difference was in the dispersion options. This large difference in the CALPUFF predicted tracer residence time due to use of the PG versus CAL or AER dispersion options (12 hours versus 6-7 hours) was not seen in any of the other CALPUFF sensitivity experiment configurations. Although use of the PG dispersion sometimes increases the estimated tracer residence time on the arc by one hour in some of the CALPUFF sensitivity tests (Table 3-14).

> The EXP2C series of experiments have estimated tracer plume duration times (11-13 hours) that is comparable to what was observed. EXP2C uses 36 km MM5 data and CALMET was run using a 4 km grid resolution with no meteorological observations (NOOBS = 2). When meteorological observations are added, either surface data alone (EXP2B) or surface and upper-air measurements (EXP2A), the tracer duration statistics degrades to only 5 to 8 hours on the arc. It is interesting to note that all of the "C" series of experiments (i.e., use of no meteorological observations in CALMET) exhibit better plume residence time statistics than the experiments that used meteorological observations (with the exception of BASEA_PG discussed previously). But only experiment EXP2C (and BASEA_PG) using 36 km MM5 data and CALMET run with 4 km grid resolution was able to replicate the observed tracer residence time.

Most of the initial CALPUFF sensitivity tests were unable to reproduce the observed tracer residence time on the 600 km arc, as was done in the EPA 1998 study using earlier versions of CALPUFF. Even the BASEA_CAL sensitivity test, which was designed to be mostly consistent with the 1998EPA_CAL simulation, estimated tracer plume residence time that was half of what was observed and estimated by the 1998EPA_CAL simulation. In addition to using difference versions of the CALPUFF model (Version 4.0 versus 5.8), the BASEA_CAL simulation also did not invoke the slug option as was used in 1998EPA_CAL (MSLUG = 1). The use of the slug option is designed for near-source applications and is not typically used in LRT dispersion modeling, so in this study the initial CALPUFF sensitivity tests did not use the slug option for modeling of the 600 km arc. The effect of the slug option is investigated in additional CALPUFF sensitivity tests discussed later in this Chapter.

Figure 3-7. Percent difference in the predicted and observed duration of time tracer is residing on the GP80 600 km arc for the CALPUFF sensitivity tests using puff model formulation and no puff splitting.

Table 3-14. Tracer plume arrival and duration statistics for the GP80 600 km arc and the initial CALPUFF sensitivity tests.

Scenario	Arrival on Arc		Leave Arc		Duration on Arc	
	(Julian Day)	Hour (LST)	(Julian Day)	Hour (LST)	(Hours)	Difference (%)
Observed	191	2	191	14	12	
1998EPA_PG	191	3	191	17	14	17%
1998EPA_CAL	191	3	191	16	13	8%
CALPUFF/CALMET						
BASEA_AER	191	2	191	7	6	-50%
BASEA_CAL	191	2	191	8	7	-42%
BASEA_PG	191	0	191	11	12	0%
EXP1A_AER	191	2	191	6	5	-58%
EXP1A_CAL	191	2	191	6	5	-58%
EXP1A_PG	191	2	191	6	5	-58%
EXP1B_AER	191	1	191	5	5	-58%
EXP1B_CAL	191	1	191	5	5	-58%
EXP1B_PG	191	1	191	4	4	-67%
EXP1C_AER	191	3	191	8	6	-50%
EXP1C_CAL	191	3	191	8	6	-50%
EXP1C_PG	191	2	191	8	7	-42%
EXP2A_AER	191	2	191	6	5	-58%
EXP2A_CAL	191	2	191	6	5	-58%
EXP2A_PG	191	2	191	9	8	-33%
EXP2B_AER	191	1	191	6	6	-50%
EXP2B_CAL	191	1	191	6	6	-50%
EXP2B_PG	191	1	191	6	6	-50%
EXP2C_AER	191	0	191	10	11	-8%
EXP2C_CAL	191	0	191	10	11	-8%
EXP2C_PG	191	0	191	12	13	8%
EXP3A_AER	191	2	191	6	5	-58%
EXP3A_CAL	191	2	191	6	5	-58%
EXP3A_PG	191	2	191	6	5	-58%
EXP3B_AER	191	1	191	5	5	-58%
EXP3B_CAL	191	1	191	5	5	-58%
EXP3B_PG	191	1	191	6	6	-50%
EXP3C_AER	191	2	191	9	8	-33%
EXP3C_CAL	191	2	191	9	8	-33%
EXP3C_PG	191	2	191	9	8	-33%
CALPUFF/MMIF						
MMIF12_CAL	191	3	191	8	6	-50%
MMIF12_PG	191	2	191	8	7	-42%

The fitted Gaussian plume statistics for the GP80 600 receptor arc and the initial CALPUFF sensitivity tests are shown in Table 3-15, with the percent differences (or angular offset for the plume centerline location) between the model predictions and observations also shown graphically in Figures 3-8 through 3-12. Unlike the CALPUFF performance for the 100 km arc that mostly overestimated the fitted plume centerline (Cmax) and observed maximum concentrations at any receptor (Omax), the CALPUFF sensitivity tests under-estimate the Cmax/Omax values for the 600 km arc by 40% to 80% (Table 3-15 and Figures 3-8 and 3-9). The Cmax/Omax underestimation bias is lower (-40% to -60%) with the "C" series (i.e., no meteorological observations in CALMET) of CALPUFF sensitivity tests. The CALPUFF sensitivity tests overstate the amount of plume spread (σ_y) along the 600 km receptor arc compared to the plume that is fitted to the observations (Figure 3-10). The "A" and "B" series of CALPUFF experiments using the turbulence dispersion (CAL and AER) tend to overestimate the plume spread along the 600 km arc by ~50% with the "C" series overestimating plume spread by ~100%. For many of the experiments, use of the PG dispersion option greatly exacerbates the plume spread overestimation bias with overestimation amounts above 250% for EPA1998_PG and its related BASEA_PG scenarios. Given the similarity of the "C" series (CALMET with no meteorological observations) and MMIF CALPUFF sensitivity simulations, it is not surprising that the MMIF runs also overestimate plume spread by ~100%.

The predicted plume centerline angular offset from the observed value has an easterly bias of 9 to 19 degrees (Figure 3-12). The "A" series of CALPUFF/CALMET sensitivity runs tend to have larger (> 15 degrees) plume centerline offsets than the "B" and "C" series of experiments, indicating that using upper-air meteorological observations in CALMET tends to worsen the plume centerline predictions in the CALPUFF sensitivity runs. Surprisingly, the CALPUFF/MMIF sensitivity runs, which also do not use the upper-air meteorological measurements, have angular offsets in excess of 15 degrees.

The observed cross wind integrated concentration (CWIC) across the plume at the 600 km arc is matched better by the CALPUFF sensitivity tests than the maximum (Cmax/Omax) concentrations (Table 3-15 and Figure 3-12). The EPA1998_PG and EPA1998_CAL overestimate the CWIC by 30% and 15%, respectively. However, the BASEA_PG and BASEA_CAL experiments, which are designed to emulate the EPA 1998 CALPUFF runs, underestimate the CWIC by -14% and -38%, respectively. The use of meteorological observations in CALMET appears to have the biggest effect on the CALPUFF CWIC performance with the "A" series (use both surface and upper-air observations) have the largest CWIC underestimation bias and the CALPUFF CWIC performance statistics as upper-air ("B" series) and then surface and upper-air ("C" series) are removed from the CALPUFF modeling. The CALPUFF/MMIF runs underestimated the CWIC by approximately -30%.

Table 3-15. CALPUFF model performance statistics using the Irwin plume fitting evaluation approach for the GP80 600 km arc of receptors for the EPA 1998 CALPUFF V4.0 modeling and the current study CALPUFF V5.8 sensitivity tests.

CALPUFF Sensitivity Test	Cmax		Omax		Sigma-y		Centerline		CWIC	
	(ppt)	MNB	(ppt)	MNB	(m)	MNB	(deg)	Diff	(ppt-m)	MNB
Observed	0.3152		0.3068		16,533		369.06		13,060	
1998EPA_PG	0.1100	-65%	0.1300	-58%	64,900	293%	25.00	15.94	17,000	30%
1998EPA_CAL	0.1400	-56%	0.1300	-58%	42,600	158%	24.00	14.94	15,000	15%
CALPUFF/CALMET										
BASEA_AER	0.1024	-68%	0.1000	-67%	27,780	68%	29.43	20.37	7,133	-45%
BASEA_CAL	0.0875	-72%	0.0817	-73%	36,870	123%	27.55	18.49	8,084	-38%
BASEA_PG	0.0763	-76%	0.0780	-75%	58,780	256%	23.74	14.68	11,240	-14%
EXP1A_AER	0.1004	-68%	0.0985	-68%	25,490	54%	27.39	18.33	6,414	-51%
EXP1A_CAL	0.1020	-68%	0.0997	-68%	25,500	54%	27.30	18.24	6,520	-50%
EXP1A_PG	0.0991	-69%	0.0969	-68%	25,280	53%	28.12	19.06	6,277	-52%
EXP1B_AER	0.1141	-64%	0.1106	-64%	34,040	106%	18.91	9.85	9,739	-25%
EXP1B_CAL	0.1168	-63%	0.1136	-63%	33,600	103%	18.77	9.71	9,840	-25%
EXP1B_PG	0.1117	-65%	0.1085	-65%	29,660	79%	21.76	12.70	8,304	-36%
EXP1C_AER	0.1388	-56%	0.1365	-56%	34,660	110%	19.01	9.95	12,060	-8%
EXP1C_CAL	0.1412	-55%	0.1387	-55%	35,070	112%	18.54	9.48	12,410	-5%
EXP1C_PG	0.1313	-58%	0.1283	-58%	32,400	96%	20.06	11.00	10,660	-18%
EXP2A_AER	0.1068	-66%	0.1046	-66%	24,520	48%	27.72	18.66	6,565	-50%
EXP2A_CAL	0.1073	-66%	0.1052	-66%	24,600	49%	27.57	18.51	6,614	-49%
EXP2A_PG	0.1204	-62%	0.1180	-62%	39,900	141%	24.41	15.35	12,040	-8%
EXP2B_AER	0.1474	-53%	0.1463	-52%	25,520	54%	19.37	10.31	9,426	-28%
EXP2B_CAL	0.1539	-51%	0.1516	-51%	24,230	47%	19.12	10.06	9,346	-28%
EXP2B_PG	0.1007	-68%	0.1149	-63%	42,590	158%	21.27	12.21	10,750	-18%
EXP2C_AER	0.1603	-49%	0.1648	-46%	35,810	117%	21.55	12.49	14,390	10%
EXP2C_CAL	0.1660	-47%	0.1712	-44%	35,330	114%	21.47	12.41	14,700	13%
EXP2C_PG	0.1842	-42%	0.1736	-43%	40,850	147%	19.35	10.29	18,860	44%
EXP3A_AER	0.1075	-66%	0.1048	-66%	24,370	47%	26.82	17.76	6,568	-50%
EXP3A_CAL	0.1079	-66%	0.1057	-66%	24,510	48%	26.70	17.64	6,630	-49%
EXP3A_PG	0.1041	-67%	0.1015	-67%	24,180	46%	27.82	18.76	6,312	-52%
EXP3B_AER	0.1332	-58%	0.1305	-57%	24,030	45%	18.54	9.48	8,025	-39%
EXP3B_CAL	0.1357	-57%	0.1327	-57%	24,050	45%	18.41	9.35	8,179	-37%
EXP3B_PG	0.0733	-77%	0.0655	-79%	38,960	136%	23.12	14.06	7,160	-45%
EXP3C_AER	0.1470	-53%	0.1436	-53%	33,260	101%	18.33	9.27	12,250	-6%
EXP3C_CAL	0.1485	-53%	0.1454	-53%	33,210	101%	18.38	9.32	12,360	-5%
EXP3C_PG	0.1380	-56%	0.1360	-56%	31,260	89%	20.80	11.74	10,820	-17%
CALPUFF/MMIF										
MMIF12KM_CAL	0.1029	-67%	0.1012	-67%	34,290	107%	26.43	17.37	8,842	-32%
MMIF12KM_PG	0.0956	-70%	0.0887	-71%	39,120	137%	24.89	15.83	9,371	-28%

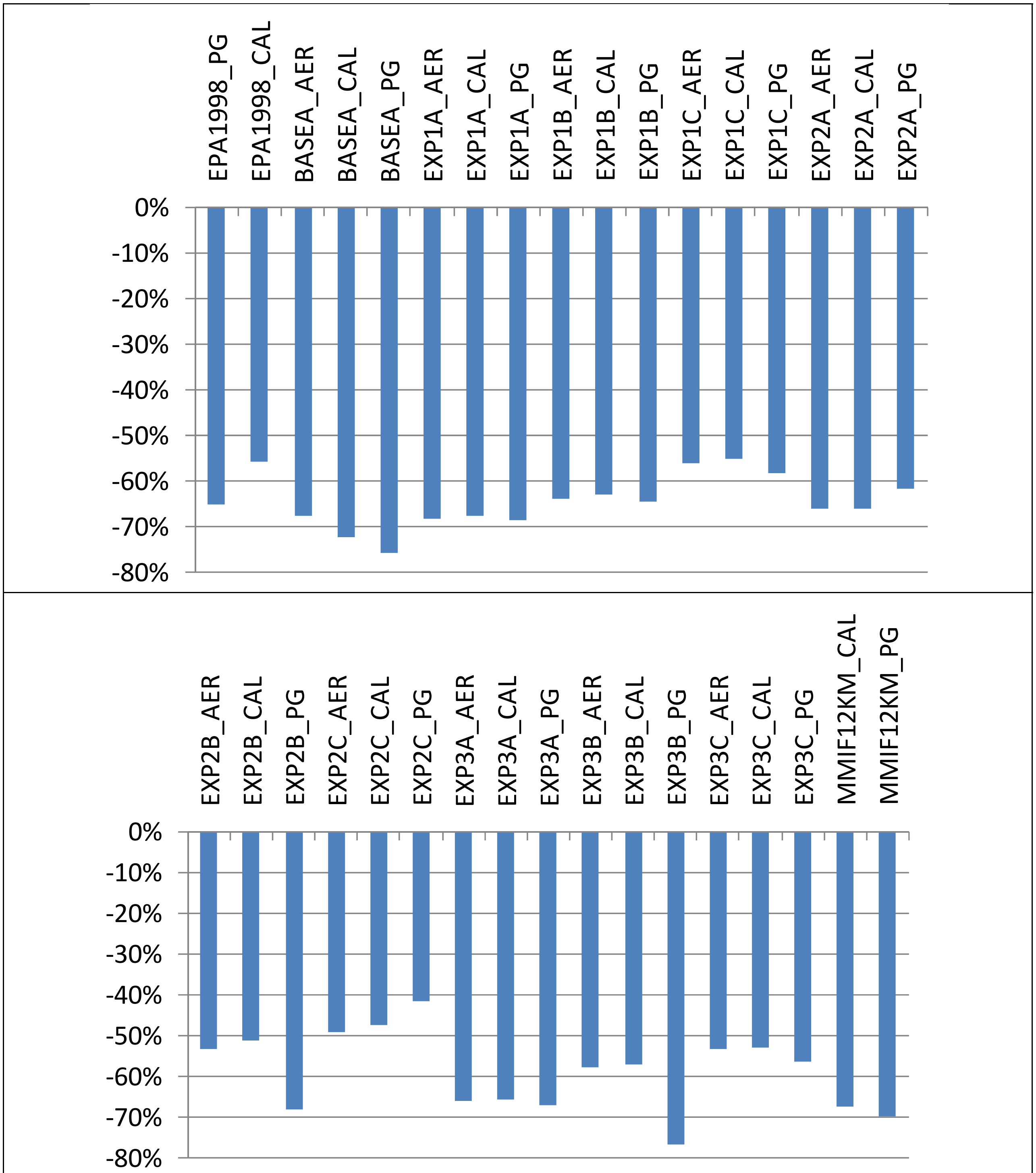


Figure 3-8. Percent difference (mean normalized bias) between the predicted and observed fitted plume centerline concentration (Cmax) for GP80 600 km receptor arc and the CALPUFF sensitivity tests.

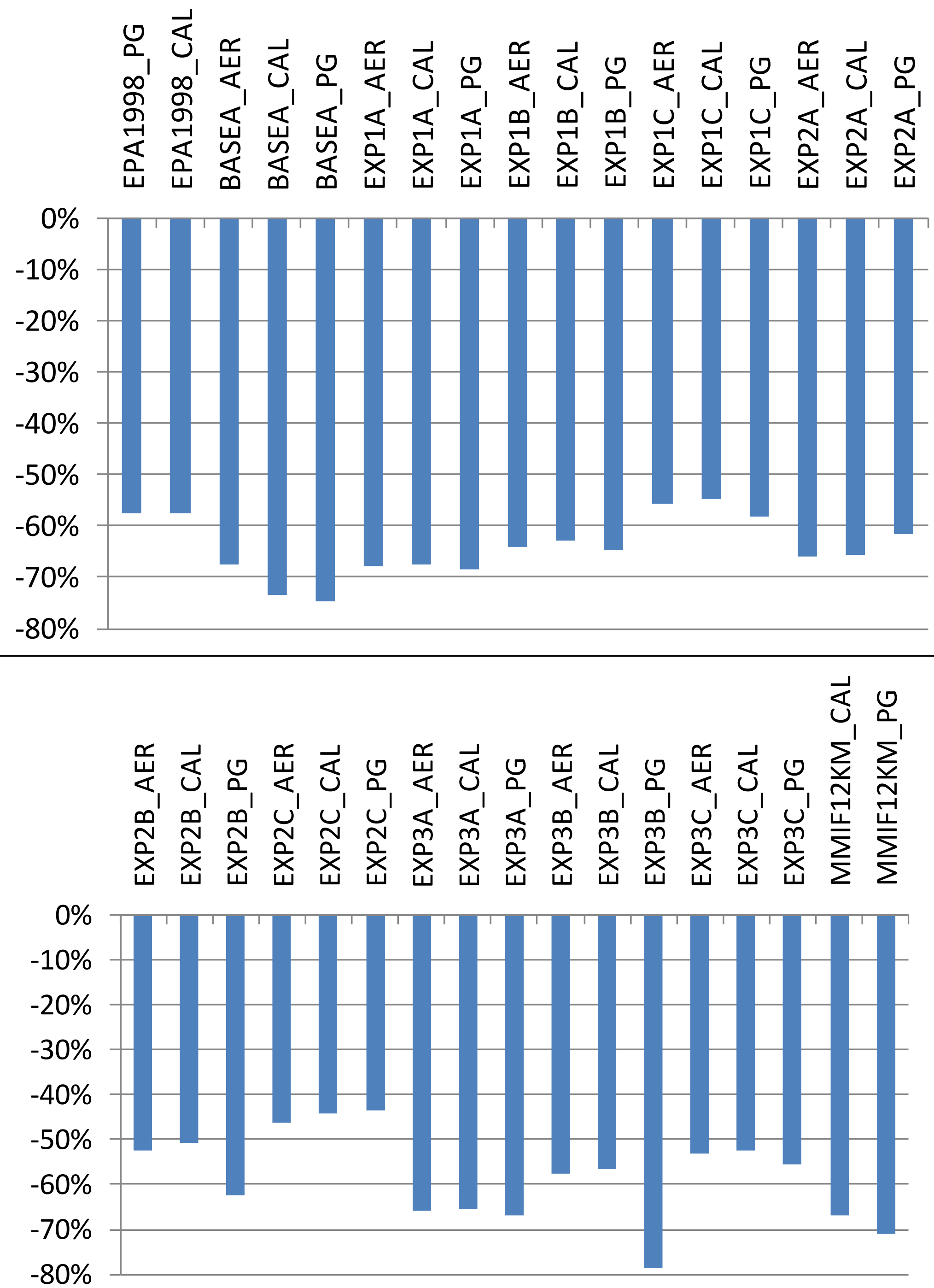


Figure 3-9. Percent difference (mean normalized bias) between the predicted and observed maximum concentration at any receptor/monitor (Omax) for GP80 600 km receptor arc and the CALPUFF sensitivity tests.

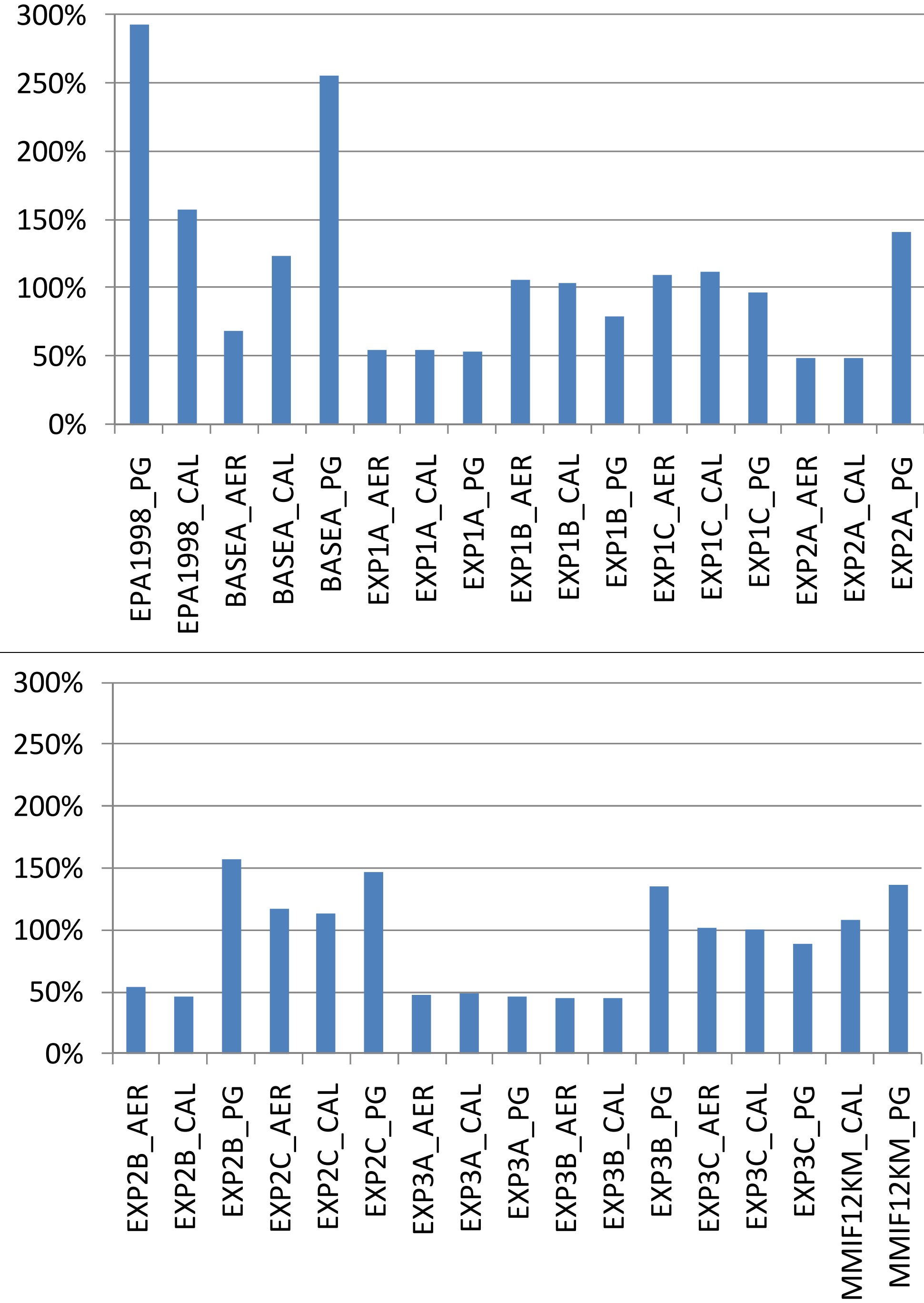


Figure 3-10. Percent difference (mean normalized bias) between the predicted and observed plume spread (σ_y) for GP80 600 km receptor arc and the CALPUFF sensitivity tests.

Figure 3-11. Difference in predicted and observed location of plume centerline (degrees) for the GP10 600 km receptor arc and the CALPUFF sensitivity tests.

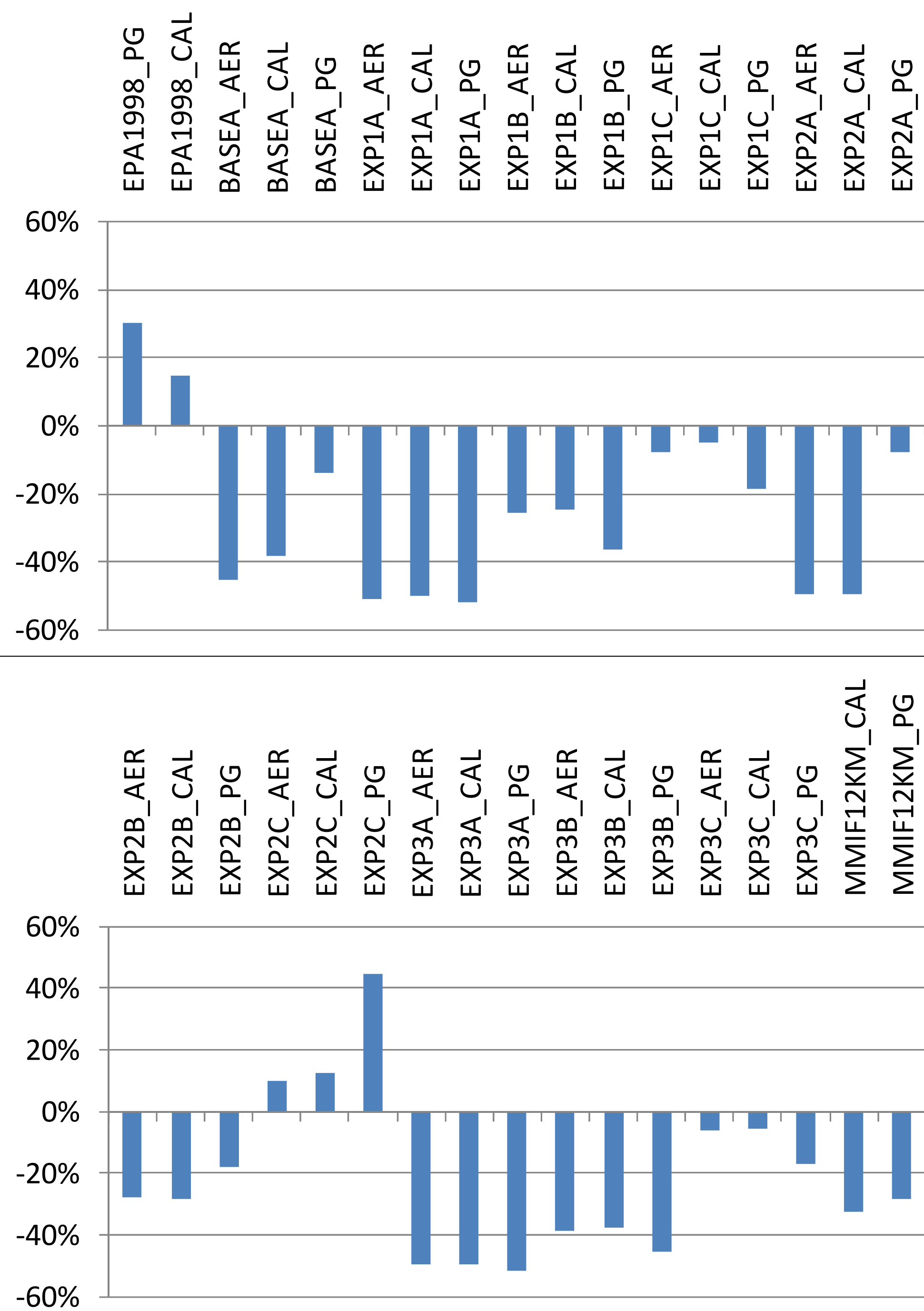


Figure 3-12. Percent difference (mean normalized bias) between the predicted and observed cross wind integrated concentration (CWIC) for the GP10 600 km receptor arc and the CALPUFF sensitivity tests.

3.4.3 SLUG and Puff Splitting Sensitivity Tests for the 600 km Arc

One issue of concern with the initial CALPUFF sensitivity tests was the large differences between the estimated residence time of the tracer on the 600 km receptor arc in the EPA 1998 and current CALPUFF simulations using the CALPUFF (CAL) turbulence dispersion options when the same meteorological observations are used as input into CALPUFF. The 1998EPA_CAL CALPUFF sensitivity simulation estimated that the tracer would remain on the 600 km receptor arc for 13 hours, which compares favorably with what was observed (12 hours) but is almost double what the BASEA_CAL simulation estimated (7 hours). In addition to updates to the CALMET and CALPUFF models that have occurred over the last decade, a major difference in the 1998 EPA and current CALPUFF 600 km arc simulations was that the 1998 EPA CALPUFF modeling used the near-source slug option, whereas the current analysis did not. Another major difference between the version of CALPUFF used in the 1998 EPA and current study was that CALPUFF now has the ability to perform puff splitting. In fact, it was the presence of puff splitting in CALPUFF that caused EPA to comment that CALPUFF may be applicable to distances further downwind than 300 km in the 2003 air quality modeling guideline revision that led to CALPUFF being the recommended long-range transport model for chemically inert pollutants (EPA, 2003).

To investigate this issue, a series of slug and puff splitting sensitivity tests were carried out using the BASEA_CAL CALPUFF/CALMET configuration by incrementally adding the near-source slug option (MSLUG = 1) and puff splitting option (MSPLIT = 1) to the BASEA_CAL model configuration. CALPUFF slug and puff splitting sensitivity tests were also carried out using the MMIF12_CAL and MMIF12_PG model configurations. Two types of puff splitting sensitivity tests were carried out:

- Default Puff Splitting (DPS) whereby the vertical puff splitting flag was turned on for just hour 17 (i.e., IRESPLIT is equal to 1 for just hour 17 and is 0 the other hours); and
- All hours Puff Splitting (APS) that turned on the vertical puff splitting flag for all hours of the day (i.e., IRESPLIT has 24 values of 1).

Table 3-16 displays the tracer residence time statistic on the 600 km receptor arc for the slug and puff splitting sensitivity tests. Using the puff model formulation and no puff splitting (BASEA_CAL), CALPUFF estimates that the tracer resides on the 600 km arc for 7 hours, which is -42% less than observed (12 hours). Using all hours puff splitting in CALPUFF, but still using the puff model formation (BASEA_APS_CAL), does not affect the estimated plume residence time statistic (7 hours). However, when the slug option is used (BASEA_SLUG_CAL) the residence time of the estimate tracer on the 600 km receptor arc more than doubles increasing from 7 to 15 hours. And adding puff splitting (APS) to the slug model formulation increases the estimated tracer duration on the arc by another hour (16 hours).

The sensitivity of the CALPUFF/MMIF model configuration 600 km receptor arc tracer residence time statistic to the specification of the slug and puff splitting options is a little different than the CALPUFF/CALMET BASEA model configuration. Whereas the CALPUFF/CALMET BASEA model configuration saw little sensitivity of the estimated tracer concentration residence time on the arc due to puff splitting, the implementation of default puff splitting increases the tracer residence time from 6 to 8 hours (CAL dispersion) and from 7 to 11 hours (PG dispersion) with all hours puff splitting increasing the residence time even more. The effect of the slug option using the CALPUFF/MMIF modeling platform has a very different effect on the tracer duration time on the arc using the CAL and PG dispersion algorithms. Using the CAL dispersion option with APS, implementing the slug option decreases the tracer residence time of the 600 km arc from 17 to 15 hours. However, using the PG dispersion option with APS, the tracer residence on the 600 km receptor arc increased from 11 to 20 hours when the slug option is invoked using the PG dispersion option.

Table 3-16. Duration of time tracer resides on the GP80 600 km receptor arc (hours) for the CALPUFF slug and puff splitting sensitivity tests.

Scenario	MSLUG	MSPLIT	Duration on 600 km Arc	
			Time (Hours)	Difference (%)
Observed			12	
CALPUFF/CALMET				
BASEA_CAL	0	0	7	-42%
BASEA_APS_CAL	0	1	7	-42%
BASEA_SLUG_CAL	1	0	15	+25%
BASEA_SLUG_APS_CAL	1	1	16	+33%
CALPUFF/MMIF				
MMIF12_CAL	0	0	6	-50%
MMIF12_DPS_CAL	0	1	8	-33%
MMIF12_APS_CAL	0	1	17	+42%
MMIF12_SLUG_APS_CAL	1	1	15	25%
MMIF12_PG	0	0	7	-42%
MMIF12_DPS_PG	0	1	11	-8%
MMMIF12_APS_PG	0	1	11	-8%
MMIF12_SLUG_APS_PG	1	1	20	+67%

Table 3-17 summarizes the plume fitting model performance statistics for the CALPUFF slug and puff splitting sensitivity tests. For the CALPUFF/CALMET BASEA_CAL slug and puff splitting sensitivity tests, the improvements in CALPUFF's estimated tracer residence time on the 600 km receptor arc when the slug option is invoked is accompanied by a further degradation in CALPUFF's ability to estimate the maximum concentrations (Cmax/Omax) as well as increasing CALPUFF's overestimate of the observed plume spread (σ_y) (~16,500 m) from ~120% (~35,000 m) without the slug option to over 250% (~60,000 m) with the slug option. The use of the slug option also improves the angular offset of the plume centerline from off by ~18 degrees to off by ~14 degrees. Finally, without using APS, CALPUFF's CWIC performance is improved from a -38% underestimation to a -12% underestimation, whereas with using APS the improvement in CWIC performance due to using the slug option is less dramatic (-31% to -25%)

Using the CALPUFF/MMIF modeling platform, the changes in the maximum (Cmax/Omax) and plume spread model performance statistics due to the use of the slug option are much less than seen with the BASEA CALPUFF/CALMET modeling platform. Use of the slug option using the CALPUFF/MMIF platform increases the maximum concentrations slightly, whereas with the CALPUFF/CALMET platform the slug option resulting in slight deceases in concentrations. The use of puff splitting had little effect on the CALPUFF/MMIF estimated maximum concentrations and resulted in slightly wider plume widths. The biggest effect puff splitting had on the CALPUFF/MMIF model performance was for the plume centerline angular displacement that improved from 16-17 to 7-8 degrees offset from observed due to the use of puff splitting (DPS or APS). In fact, of all the CALPUFF sensitivity tests examined, CALPUFF/MMIF using puff splitting is the best performing model configuration for estimating plume centerline location. Puff splitting resulted in small improvements in CALPUFF's ability to predict CWIC across the 600 km arc. But the slug option greatly improved CALPUFF/MMIF's ability to reproduce the observed CWIC. For example, using the CAL turbulence dispersion option, CALPUFF/MMIF underestimates the observed CWIC at the 600 km receptor arc by -32% using the puff model configuration and no puff splitting. Using the DPS and APS puff splitting approach reduces the CWIC underestimation bias to -28% and -21%, respectively, And then adding the slug formulation with the APS completely eliminates the CWIC underestimation bias (-2%). In fact, use of the APS and slug options with the CALPUFF/MMIF modeling platform results in the best performing CALPUFF sensitivity test for estimating CWIC across the 600 km arc of all the CALPUFF sensitivity tests analyzed (Tables 3-15 and 3-17).

Table 3-17. Plume fitting statistics for the CALPUFF slug and puff splitting sensitivity tests.

CALPUFF slug and puff splitting sensitivity test	Cmax		Omax		Sigma-y		Centerline		CWIC	
	(ppt)	MNB	(ppt)	MNB	(m)	MNB	(deg)	Diff	(ppt-m)	MNB
Observed	0.3152		0.3068		16,533		369.06		13,060	
CALPUFF/CALMET										
BASEA_CAL	0.0875	-72%	0.0817	-73%	36,870	123%	27.55	18.49	8,084	-38%
BASEA_APS_CAL	0.1014	-68%	0.1029	-66%	35,510	115%	27.19	18.13	9,023	-31%
BASEA_SLUG_CAL	0.0728	-77%	0.0726	-76%	62,650	279%	22.49	13.43	11,430	-12%
BASEA_SLUG_APS_CAL	0.0673	-79%	0.0652	-79%	58,440	253%	23.56	14.50	9,855	-25%
CALPUF/MMIF										
MMIF12KM_CAL	0.1029	-67%	0.1012	-67%	34,290	107%	26.43	17.37	8,842	-32%
MMIF12KM_DPS_CAL	0.1049	-67%	0.1016	-67%	35,960	118%	16.74	7.68	9,454	-28%
MMIF12KM_APS_CAL	0.1108	-65%	0.1076	-65%	37,120	125%	16.30	7.24	10,310	-21%
MMIF12KM_SLUG_CAL	0.1458	-54%	0.1462	-52%	35,190	113%	16.92	7.86	12,860	-2%
MMIF12KM_PG	0.0956	-70%	0.0887	-71%	39,120	137%	24.89	15.83	9,371	-28%
MMIF12KM_DPS_PG	0.1085	-66%	0.1143	-63%	41,610	152%	17.04	7.98	11,310	-13%
MMIF12KM_APS_PG	0.1085	-66%	0.1143	-63%	41,610	152%	17.04	7.98	11,310	-13%
MMIF12KM_SLUG_PG	0.1251	-60%	0.1115	-64%	41,770	153%	17.43	8.37	13,100	0%

3.5 CONCLUSIONS ON GP80 TRACER TEST EVALUATION

For the 100 km receptor arc CALPUFF/CALMET sensitivity simulations, the ability of CALPUFF to simulate the observed tracer concentrations varied among the different CALMET configurations and were not inconsistent with the results of the 1998 EPA CALPUFF evaluation study (EPA, 1998a). The best performing CALPUFF/CALMET configuration was when CALMET was run using MM5 data and just surface meteorological observations and no upper-air meteorological observations. In general, the CAL and AER turbulence dispersion options in CALPUFF performed similarly and performed better than the PG dispersion option. The performance of CALPUFF using the MMIF tool tended to be in the middle of the range of model performance for the CALPUFF/CALMET sensitivity tests; not as good as the performance of CALPUFF/CALMET using MM5 and just surface observations data in CALMET, but better than the performance of CALPUFF using MM5 data and no meteorological observations in CALMET.

The CALPUFF sensitivity modeling results for the GP80 600 km receptor arc were quite variable. With two notable exception (the BASEA_PG and EXP2C configurations), the initial CALPUFF sensitivity tests were unable to duplicate the observed tracer residence time on the 600 km receptor arc as was seen in the 1998 EPA CALPUFF evaluation study (EPA, 1998a). However, when the near-source slug option was used, CALPUFF/CALMET was better able to reproduce the amount of time that the tracer was observed on the 600 km receptor arc. The standard application of CALPUFF for LRT applications is the puff model formulation rather than the slug model formation, which is designed to better simulate a near-source continuous plume. The fact that the slug formulation is needed to produce reasonable CALPUFF model performance for residence time on the 600 km receptor suggests that the findings of the 1998 EPA CALPUFF evaluation study should be re-evaluated.

In general, the CALPUFF/CALMET sensitivity tests that are based on CALMET using MM5 data with no meteorological observations exhibit better plume fitting model performance statistics for the 600 km receptors arc than when meteorological observations are used with CALMET. The use of the slug option with CALMET/CALPUFF, which improved the plume residence time statistics, degrades the maximum concentrations and plume width statistics, but improves the plume centerline and CWIC average plume concentration statistics. Puff splitting had little effect on the CALPUFF/CALMET model predictions on the 600 km receptor arc. However, puff splitting did improve the CALPUFF/MMIF plume centerline and CWIC average plume concentration statistics, as well as the tracer residence time statistics. Puff splitting resulted in a slight degradation of the plume width statistics in CALPUFF/MMIF. Using the slug option with puff splitting in CALPUFF/MMIF results in the best performing CALPUFF model configuration of all the sensitivity tests for the plume centerline and CWIC average plume statistics, although the use of slug and puff splitting does degrade the plume width statistic.

4.0 1975 SAVANNAH RIVER LABORATORY FIELD STUDY

4.1 DESCRIPTION OF THE 1975 SAVANNAH RIVER LABORATORY FIELD STUDY

The 1975 Savannah River Laboratory (SRL75) field experiment was located in South Carolina and occurred in December 1975 (DOE, 1978). A SF_6 tracer was released for four hours between 10:25 and 14:25 LST on December 10, 1975 from a 62 m stack with a diameter of 1.0 m, exit velocity of 0.001 m/s and at ambient temperature. A single monitoring arc was used in the SRL75 experiment that was approximately 100 kilometers from the source with monitoring sites located along I-95 from Mile Post (MP) 76 near St. George, SC in the south to Hwy 36 west of Tillman, SC to the north and along SC 336.

The 1998 EPA CALPUFF evaluation (EPA, 1998a) used the SRL75 SF_6 tracer release in the CALPUFF model evaluation. However, the 1986 8 LRT dispersion model evaluation study (Policastro et al., 1986) used the longer-term SRL Krypton-85 release database (Telegadas et al., 1980). In this study we evaluated CALPUFF using the SRL75 SF_6 database to be consistent with the 1998 EPA study.

4.2 MODEL CONFIGURATION AND APPLICATION

Both the CALMET meteorological model and MMIF tools were used to provide meteorological inputs to CALPUFF. The CALMET modeling was performed using a Universal Trans Mercator (UTM) map projection in order to be consistent with the past CALPUFF applications (EPA, 1998a). The MMIF meteorological processing used a Lambert Conformal Conic (LCC) map projection because in must be consistent with the MM5 coordinate system. Figure 4-1 displays the CALMET/CALPUFF UTM modeling domain and locations of the ~200 receptors used in the CALPUFF modeling that lie along an arc 100 km from the source. The tracer was observed using ~40 monitors that were located along I-95 between MP 24 and 76 that were approximately 100 km from the source. When using the Irwin Gaussian plume fitting model evaluation approach, the tracer observations at the monitoring sites are assumed to be on an arc of receptors 100 km from the source.

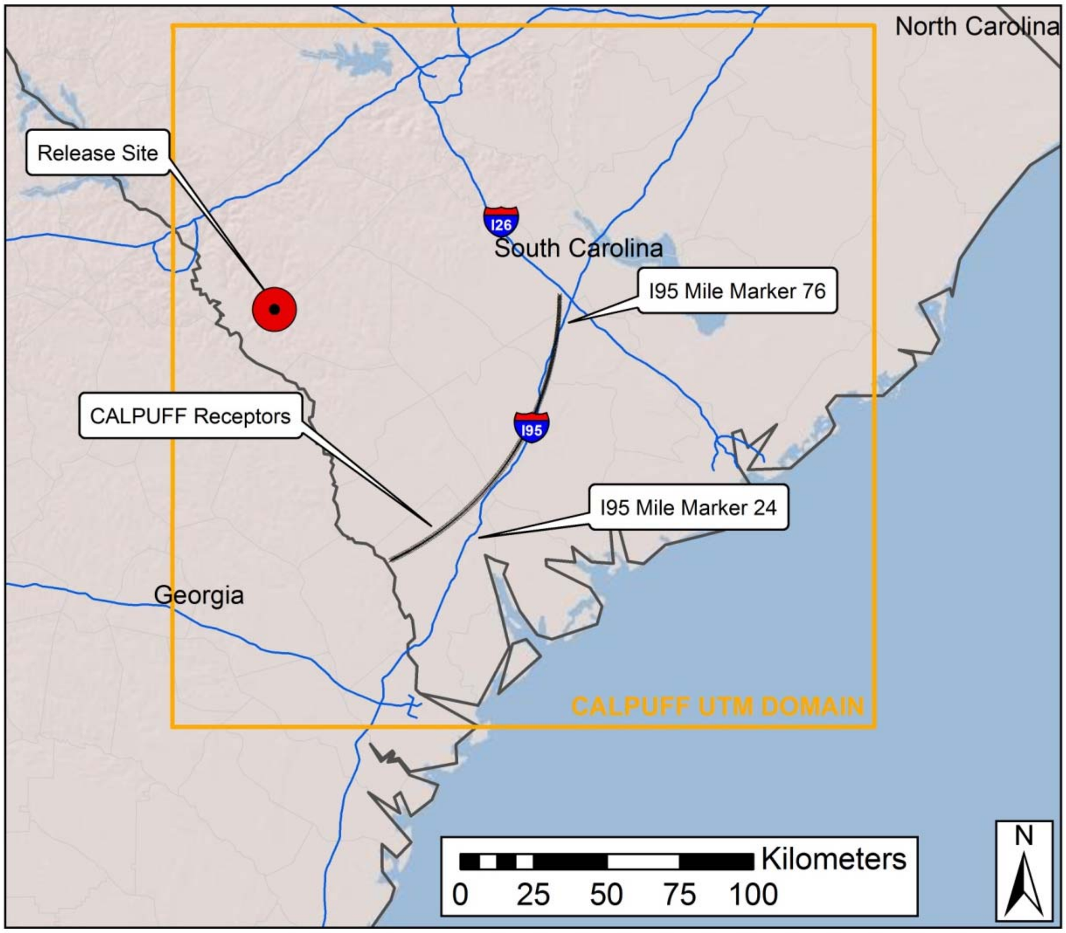


Figure 4-1. CALPUFF/CALMET UTM modeling domain and location of tracer release site and CALPUFF receptors along am arc 100 km from the source for the SRL75 CALPUFF modeling.

In the CALPUFF modeling system, each of the three programs (CALMET, CALPUFF, and CALPOST) uses a control file of user-selectable options to control the data processing. There are numerous options in each and several that can result in significant differences. The following model controls for CALMET and CALPUFF were employed for the analyses with the SRL75 tracer data.

4.2.1 CALMET Options

The following CALMET control parameters and options were chosen for the BASE case model evaluation. The BASE case control parameters and options were chosen to be consistent with two previous CALMET/CALPUFF evaluations (Irwin 1997 and EPA 1998a). The most important CALMET options relate to the development of the wind field and were set as follows:

NOOBS	= 0	Use surface, overwater, and upper air station data
IWFCOD	= 1	Use diagnostic wind model to develop the 3-D wind fields
IFRADJ	= 1	Compute Froude number adjustment effects (thermodynamic blocking effects of terrain)
IKINE	= 0	Do NOT compute kinematic effects
IOBR	= 0	Do NOT use O'Brien procedure for adjusting vertical velocity
IEXTRP	= 4	Use similarity theory to extrapolate surface winds to upper layers

IPROG	= 0	Do NOT use prognostic wind field model output as input to diagnostic wind field model (for observations only sensitivity test)
ITPROG	= 0	Do NOT use prognostic temperature data output

Mixing heights are important in the estimating ground level concentrations. The CALMET options that affect mixing heights were set as follows:

IAVEZI	= 1	Conduct spatial averaging
MNMDAV	= 1	Maximum search radius (in grid cells) in averaging process
HAFANG	= 30.	Half-angle of upwind looking cone for averaging
ILEVZI	= 1	Layer of winds to use in upwind averaging
DPTMIN	= .001	Minimum potential temperature lapse rate (K/m) in stable layer above convective mixing height
DZZI	= 200	Depth of layer (meters) over which the lapse rate is computed
ZIMIN	= 100	Minimum mixing height (meters) over land
ZIMAX	= 3200	Maximum mixing height (meters) over land, defined to be the top of the modeling domain

A number of CALMET model control options have no recommended default values, particularly radii of influence values for terrain and surface and upper air observations. The CALMET options that affect radius of influence were set as follows:

RMAX1	= 20	Minimum radius of influence in surface layer (km)
RMAX2	= 50	Minimum radius of influence over land aloft (km)
RMIN	= 0.1	Minimum radius of influence in wind field interpolation (km)
TERRAD	= 10	Radius of influence of terrain features (km)
RPROG	= 0	Weighting factors of prognostic wind field data (km)

A review of the respective CALMET parameters between the 1998 EPA CALMET/CALPUFF evaluation study using CALMET Version 4.0 and the BASE case scenario in the current CALMET/CALPUFF evaluation using CALMET Version 5.8 indicates differences in some CALMET options. The differences between the two scenarios are presented below in Table 4-1. All other major CALMET options for BASE case scenario matched the original 1998 EPA analysis.

Table 4-1. CALMET parameters for the SRL75 tracer field experiment modeling used in the, 1998 EPA and current BASE case analysis.

CALMET Option	Description	1998 EPA Setup	BASE Setup
IKINE	Adjust winds using Kinematic effects? (yes = 1 and no = 0)	1	0
MNMDAV	Maximum search radius for averaging mixing heights (# grid cells)	3	1
ZUPWND	Bottom and top layer through which domain-scale winds are calculated (in meters)	1,2000	1,1000
RMIN	Minimum radius of influence in wind field interpolation (in km)	2	0.1
RMIN2	Minimum upper air station to surface station extrapolation radius (in km)	-1	4

The CALMET preprocessor can utilize National Weather Service (NWS) meteorological data and on-site data to produce temporally and spatially varying three dimensional wind fields for CALPUFF. Only NWS data were used for this effort and came from two compact disc (CD) data sets. The first was the *Solar and Meteorological Surface Observation Network* (SAMSON) compact discs, which were used to obtain the hourly surface observations. The surface stations used for the SRL75 CALMET modeling are shown in Table 4-2.

Table 4-2. 1975 Savannah River Laboratory surface meteorological stations.

State	Cities
Georgia	Athens, Atlanta, Augusta, Macon, Savannah
North Carolina	Asheville, Charlotte, Greensboro, Raleigh-Durham, Wilmington
South Carolina	Charleston, Columbia, Greer-Spartanburg

Twice daily soundings came from the second set of compact discs, the *Radiosonde Data for North America*. The upper-air rawinsonde meteorological observations used in the SRL75 CALMET modeling are shown in Table 4-3.

Table 4-3. 1975 Savannah River Laboratory tracer experiment rawinsonde sites.

State	Cities
Georgia	Athens, Waycross
North Carolina	Greensboro, Cape Hatteras
South Carolina	Charleston

Six vertical layers were defined for the CALPUFF modeling to be consistent with the Irwin (1997) and EPA (1998a) modeling as follows: surface-20, 20-50, 50-100, 100-500, 500-2000, and 2000-3300 meters.

MM5 prognostic meteorological model simulations were conducted using grid resolutions of 36, 12 and 4 km. The CALMET modeling used the 12 km MM5 data. The MMIF tool was applied using all three MM5 grid resolutions and using the first 27 MM5 vertical layers from the surface to approximately 6,500 m AGL.

4.2.2 CALPUFF Control Options

The following CALPUFF control parameters, which are a subset of the control parameters, were used. These parameters and options were chosen to be consistent with the 1977 INEL study (Irwin 1997) and 1998 EPA CALPUFF evaluation (EPA, 1998a) studies. Note that use of the slug option (MSLUG = 1) is fairly non-standard for LRT modeling. However, that was what was used in the 1997 INEL and 1998 EPA studies so it was also used in this study's CALPUFF evaluation using the SRL75 tracer database.

Technical options (group 2):

MCTADJ	= 0	No terrain adjustment
MCTSG	= 0	No subgrid scale complex terrain is modeled
MSLUG	= 1	Near-field puffs modeled as elongated (i.e., slugs)
MTRANS	= 1	Transitional plume rise is modeled
MTIP	= 1	Stack tip downwash is modeled
MSHEAR	= 0	Vertical wind shear is NOT modeled above stack top
MSPLIT	= 0	No puff splitting
MCHEM	= 0	No chemical transformations
MWET	= 0	No wet removal processes
MDRY	= 0	No dry removal processes
MPARTL	= 0	No partial plume penetration

MPDF	= 0	PDF NOT used for dispersion under convective conditions
MREG	= 0	No check made to see if options conform to regulatory Options

Two different values were used for the dispersion parameterization option MDISP:

= 2	Dispersion coefficients from internally calculated sigmas
= 3	PG dispersion coefficients for RURAL areas (PG)

In addition, under MDISP = 2 dispersion option, two different options were used for the MCTURB option that defines the method used to compute turbulence sigma-v and sigma-w using micrometeorological variables:

= 1	Standard CALPUFF routines (CAL)
= 2	AERMOD subroutines (AER)

Several miscellaneous dispersion and computational parameters (group 12) were set as follows:

SYTDEP	= 550.	Horizontal puff size beyond which Heffter equations are used for sigma-y and sigma-z
MHFTSZ	= 0	Do NOT use Heffter equation for sigma-z
XMXLEN	= 1	Maximum length of slug (in grid cells)
XSAMLEN	= 1	Maximum travel distance of puff/slug (in grid cells) during one sampling step
MXNEW	= 99	Maximum number of slugs/puffs released during one time step
WSCALM	= 0.5	Minimum wind speed (m/s) for non-calm conditions
XMAXZI	= 3000	Maximum mixing height (meters)
XMINZI	= 50	Minimum mixing height (meters)
SL2PF	= 10	Slug-to-puff transition criterion factor (= sigma-y/slug length)

A review of the respective CALPUFF parameters between the 1998 EPA CALMET/CALPUFF evaluation study using CALMET Version 4.0 and the BASE case scenario in the current CALMET/CALPUFF evaluation using CALPUFF Version 5.8 indicates differences in some parameters. The differences between the two scenarios are presented below in Table 4-4. All other major CALPUFF options for current BASE case scenario matched the original 1998 EPA analysis.

Table 4-4. CALPUFF parameters used in the SRL75 tracer field experiment modeling for the 1998 EPA and current BASE case analysis.

CALPUFF Option	Description	1998 EPA Setup	Current Study BASE Setup
SYMIN	Minimum sigma y (meters)	0.01	1
SZMIN	Minimum sigma z (meters)	0.01	1
WSCALM	Minimum wind speed (m/s) for non-calm conditions	1.0	0.5
XMAXZI	Maximum mixing height (meters)	3300	3000
XMINZI	Minimum mixing height (meters)	20	50
XMXLEN	Maximum length of slug (in grid cells)	0.1	1
XSAMLEN	Maximum travel distance of puff/slug (in grid cells) during one sampling step	0.1	1
MXNEW	Maximum number of slugs/puffs released during one time step	199	99
MXSAM	Maximum number of sampling steps per slug/puff during one time step	5	99
SL2PF	Slug-to-puff transition criterion factor (= sigma-y/slug length)	5.0	10.0

4.2.3 SRL75 CALPUFF/CALMET Sensitivity Tests

Table 4-5 describes the CALMET/CALPUFF sensitivity tests performed for the modeling of the 100 km arc of receptors in the SRL75 field study. The BASE simulation uses the same configuration as used in the 1998 EPA CALPUFF evaluation report, only updated from CALPUFF Version 4.0 to CALPUFF Version 5.8. The CALMET and CALPUFF parameters of the BASE case simulations were discussed earlier in this section.

The sensitivity simulations are designed to examine the sensitivity of the CALPUFF model performance to 10 km grid resolution in the CALMET meteorological model simulation, the use of 12 km resolution MM5 output data used as input to CALMET, and the use of surface and upper-air meteorological observations in CALMET through NOOBS = 0 (use surface and upper-air observation), 1 (use only surface observations) and 2 (don't use any observations).

In addition, for each experiment using different CALMET model configurations, three CALPUFF dispersion options were examined as shown in Table 4-6. Two of the CALPUFF dispersion sensitivity tests using dispersion based on sigma-v and sigma-w turbulence values using the CALPUFF (CAL) and AERMOD (AER) algorithms. Whereas the third dispersion option (PG) uses Pasquill-Gifford dispersion coefficients.

Table 4-5. CALPUFF/CALMET experiments for the SRL75 tracer experiment.

Experiment	CALMET Grid	MM5 Data	NOOBS	Comment
BASE	10 km	None	0	Original met observations only configuration
EXP1A	10 km	12 km	0	Aug 2009 IWAQM w/10 km grid using 12 km MM5
EXP1B	10 km	12 km	1	Don't use observed upper-air meteorological data
EXP1C	10 km	12 km	2	Don't use observed surface/upper-air meteorological data

Table 4-6. CALPUFF dispersion options examined in the CALPUFF sensitivity tests.

Experiment	MDISP	MCTURB	Comment
CAL	2	1	Dispersion coefficients from internally calculated sigma-v and sigma-w using micrometeorological variables and CALPUFF algorithms
AER	2	2	Dispersion coefficients from internally calculated sigma-v and sigma-w using micrometeorological variables and AERMOD algorithms
PG	3	--	PG dispersion coefficients for rural areas and MP coefficients for urban areas

The CALMET and CALPUFF simulations used for the sensitivity analyses were updated from the BASE case model configuration that was designed to be consistent with the 1998 EPA study by using recommended settings for many variables from the August 2009 EPA Clarification Memorandum. A summary of CALMET parameters that changed from the BASE case scenarios for the CALPUFF sensitivity tests are presented in Table 4-7.

Table 4-7. CALMET wind field parameters for the SRL75 tracer experiment.

CALMET Option	2009 EPA-FLM Default	BASE	EXP1A	EXP1B	EXP1C
NOOBS	0	0	0	1	2
ICLOUD	0	0	0	0	3
IEXTRP	-4	4	-4	-4	1
IPROG	14	0	14	14	14
ITPROG	0	0	0	1	2
ZIMIN	50	100	50	50	50
ZIMAX	3000	3200	3000	3000	3000
RMAX1	100	20	100	100	50
RMAX2	200	50	200	200	100

4.2.4 CALPUFF/MMIF Sensitivity Tests

With the MMIF software tool designed to reformat the MM5/WRF meteorological model output data for input into CALPUFF, there are much less options available and hence much fewer sensitivity tests as shown in Table 4-8.

Table 4-8. CALPUFF/MMIF sensitivity tests analyzed with the SRL75 tracer experiment.

Grid Resolution	MM5	MDISP	MCTURB	Comment
36 km	36 km	2	1	36 km MM5 with CALPUFF turbulence dispersion
36 km	36 km	2	2	36 km MM5 with AERMOD turbulence dispersion
36 km	36 km	3	--	36 km MM5 with Pasquill-Gifford dispersion
12 km	12 km	2	1	12 km MM5 with CALPUFF turbulence dispersion
12 km	12 km	2	2	12 km MM5 with AERMOD turbulence dispersion
12 km	12 km	3	--	12 km MM5 with Pasquill-Gifford dispersion
4 km	4 km	2	1	4 km MM5 with CALPUFF turbulence dispersion
4 km	4 km	2	2	4 km MM5 with AERMOD turbulence dispersion
4 km	4 km	3	--	4 km MM5 with Pasquill-Gifford dispersion

4.3 QUALITY ASSURANCE

The quality assurance (QA) of the CALPUFF modeling system simulations for the SRL tracer experiment was assessed by analyzing the CALMET and CALPUFF input and output files and the dates they were generated. The input file options were compared against the EPA-FLM recommended settings from the August 2009 Clarification Memorandum (EPA, 2009b) and the definitions of the sensitivity tests to assure that the intended parameters were varied. The QA of the MMIF runs was not completed because no input files or list files were provided to document the MMIF parameters.

The CALMET sensitivity simulations used a radius of influence of terrain on wind fields equal to 10 m (TERRAD = 10). The 2009 EPA Clarification Memorandum recommends TERRAD = 15. The CALMET sensitivity simulations used a minimum extrapolation distance between surface and upper air stations of 4 km (RMIN2 = 4). The 2009 EPA Clarification Memorandum recommends RMIN2 = -1.

Four CALMET parameters (BIAS, NSMTH, NINTR2, and FEXTR2) require a value for each vertical layer processed in CALMET. The CALMET BASE case has six vertical layers, but the sensitivity simulations are based on ten vertical layers. The CALMET sensitivity simulations were provided with only six values for BIAS, NSMTH, NINTR2, and FEXTR2 even though ten vertical layers were simulated. Therefore, CALMET used default values for the upper four vertical layers (i.e., 1200 m, 2000 m, 3000 m, and 4000 m).

In addition to the three CALPUFF dispersion options (AERMOD, CALPUFF, and PG), there were other CALPUFF parameters that differed between the CALPUFF/CALMET (BASE and sensitivity cases) and CALPUFF/MMIF modeling scenarios. The CALPUFF parameter differences include:

- CALPUFF/CALMET sensitivity runs using AERMOD and CALPUFF dispersion were conducted using near-field slug formation (MSLUG = 1), but the CALPUFF/CALMET PG and CALPUFF/MMIF runs were conducted using puffs (MSLUG = 0).
- CALPUFF/CALMET sensitivity runs using AERMOD and CALPUFF dispersion were set-up to not allow for partial plume penetration of inversion layer (MPARTL = 0).

The quality assurance of the post-processing of the SRL75 CALPUFF runs uncovered two errors. The first was that the conversion factor to convert the SF_6 tracer concentrations from mass per volume to ppt was approximately three times too large. The second error was that when calculating the integrated concentrations along the arc, the wrong time period was specified. These two errors were fixed and the CALPUFF results re-processed to generate new plume fitting statistical performance measures.

4.4 MODEL PERFORMANCE EVALUATION FOR THE SRL75 TRACER EXPERIMENT

The Irwin (1997) plume fitting evaluation approach was used to evaluate CALPUFF for the SRL75 field experiment. There are two components to the Irwin plume fitting evaluation approach:

1. A temporal analysis that examines the time the tracer arrives, leaves and resides on the receptor arc; and
2. A plume fitting procedures that compares the predicted observed peak and average plume concentrations and the width of the plume by fitting a Gaussian plume through the predicted or observed concentrations across the arc of receptors or monitors that lie on the 100 km receptor arc.

Because only long-term integrated average observed SF_6 samples were available, the timing component of the evaluation could not be compared against observed values in the SRL75 experiments.

Most of the CALPUFF sensitivity tests estimated that the tracer arrived at the 100 km arc on hour 13 LST, 2½ hours after the beginning to the tracer release. The exceptions to this are the CALPUFF/MMIF simulations using the 4 km MM5 data and CALPUFF/MMIF using the 36 km and PG dispersion that estimated the plume arrives at hour 14 LST. With one exception, the CALPUFF simulations estimated that the tracer resided either 5 or 6 hours on the arc. And with two exceptions, it was the meteorological data rather than the dispersion option that defined the residence time of the estimated tracer on the 100 km receptor arc. The exceptions were for the PG dispersion sensitivity test that in two cases predicted the tracer would remain one less hour on the arc; the CALPUFF/CALMET BASE sensitivity test using the PG dispersion estimated that the tracer would reside only 4 hours on the 100 km receptor arc. Without any observed tracer timing statistics, these results are difficult to interpret.

Table 4-9 displays the model performance evaluation for the various CALPUFF sensitivity tests using the Irwin plume fitting evaluation approach. The observed values were taken from the 1998 EPA CALPUFF tracer test evaluation report data (EPA, 1998a). Also shown in Table 4-4 are the statistics from the 1998 EPA report for the CALPUFF V4.0 modeling using Pasquill-Gifford (PG) and similarity (CAL) dispersion. Note that the EPA 1998 CALPUFF modeling used CALMET with just observations so is analogous to the BASE sensitivity scenario that used CALPUFF V5.8. There are five statistical parameters evaluated using the Irwin plume fitting evaluation approach:

- Cmax, which is the plume fitted centerline concentration.
- Omax, which is the maximum observed value at the ~40 monitoring sites or maximum predicted value across the ~200 receptors along the 100 km arc.
- Sigma-y, which the second moment of the Gaussian distribution and a measure of the plume spread.
- Plume Centerline, which is the angle of the plume centerline from the source to the 100 km arc.
- CWIC, the cross wind integrated concentration (CWIC) across the predicted and observed fitted Gaussian plume.

The first thing we note in Table 4-9 is that the maximum centerline concentration of the fitted Gaussian plume to the observed SF_6 tracer concentrations across the 12 monitors (2.739 ppt) is almost half the observed maximum at any of the monitors (5.07 ppt). As the centerline concentrations in a Gaussian plume represents the maximum concentration, this means that the fitted Gaussian plume is not a very good fit of the observations and the Cmax parameter is not a good indicator of model performance. Comparison of the predicted and observed Omax values that represents the maximum observed concentration across the monitoring sites and the maximum predicted value at any of the 200 receptors along the arc is an apple-orange comparison. We would expect the predicted Omax value to be the same or larger than the observed Omax value given there are ~5 times more samples of the plume in the model predictions compared to the observations. This is the case for all of the CALPUFF/MMIF sensitivity tests. However, when CALPUFF is run using CALMET with no MM5 data (BASE), the predicted Omax value is less than the observed value for both CALPUFF V4.0 and CALPUFF V5.8, which is an undesirable attribute.

The fitted plume width (sigma-y) based on observations is almost doubled the fitted plume width based on the CALPUFF model predictions for all the CALPUFF simulations. However, this is likely due in part to the poor Gaussian plume fit of the observations. Figure 4-2 is reproduced from the 1998 EPA CALPUFF tracer test report and compares the CALPUFF fitted Gaussian plume concentrations with the 13 observed tracer concentrations, where the predicted and observed tracer distributions have been rotated so that their centerlines match up. Of the 13 monitors pictured along the 100 km arc, four have substantial (> 2.0 ppt) concentrations whereas the tracer concentrations at the remaining monitoring sites are mostly <0.2 ppt. Based on this figure, the predicted and observed plume widths match quite well. However, when fitting a Gaussian plume to the observations it appears that the "observed" width is overstated due to the low tracer concentration monitoring sites on the wings of the plume. These results suggest that in the real world the concept of a Gaussian plume may not hold at longer downwind distances, such as the 100 km receptor arc used in the SRL75 field experiment. Consequently, the use of a fitted Gaussian plume as a model evaluation tool may be a poor indicator of model performance for LRT dispersion models.

The plume centerline metric is a useful tool for evaluating the main flow of the center of mass of a plume from the source to receptor arc. The observed plume centerline is at 126 degrees. The CALPUFF/MMIF estimated centerline is off by 8-10 degrees too far south. However, CALPUFF using CALMET and just observations is off by 17 degrees (EPA, 1998a) and 20 degrees (BASE) and it is too far south. Adding the 12 km MM5 data with the observations in CALPUFF (EXP1) only improves the centerline angular offset from 20 to 19 degrees. Removing the upper-air meteorological observations from the CALMET modeling (EXP2) results in no improvements in the CALPUFF/CALPUFF centerline offset (still 19 degrees). However, also removing the surface meteorological observations from the CALMET modeling (EXP3, NOOBS = 2) improves the CALPUFF/CALMET centerline angular offset from 19 to 12 degrees so that it is almost as good as the CALPUFF/MMIF simulations (8 to 10 degrees offset).

Table 4-9. CALPUFF model performance statistics using the Irwin plume fitting evaluation approach using the SRL75 field experiment and the 1998 EPA study and the CALPUFF sensitivity tests.

CALPUFF	Cmax[1]		Omax		Sigma-y[1]		Plume Centerline		CWIC	
Sensitivity Test	(ppt)	MNB	(ppt)	MNB	(meters)	MNB	(degrees)	Diff (deg)	(ppt/m^2)	MNB
Observed	2.739		5.07		11643		125.59		79,940	
EPA 1998										
PG	7.20	163%	6.90	36%	7200	-38%	143	17	129,000	61%
Similarity	5.1	86%	5.00	-1%	6000	-48%	143	17	77,000	-4%
MMMIF										
4KM_AER	8.791	221%	8.625	70%	6810	-42%	135.9	10.31	150,100	88%
4KM_CAL	8.79	221%	8.625	70%	6801	-42%	135.9	10.31	149,800	87%
4KM_PG	8.798	221%	8.656	71%	6844	-41%	135.9	10.31	150,900	89%
12KM_AER	10.63	288%	10.41	105%	6587	-43%	133.8	8.21	175,500	120%
12KM_CAL	10.79	294%	10.42	106%	6492	-44%	133.8	8.21	175,500	120%
12KM_PG	10.7	291%	10.49	107%	6545	-44%	133.8	8.21	175,500	120%
36KM_AER	11.61	324%	11.4	125%	6315	-46%	134.1	8.51	183,800	130%
36KM_CAL	11.62	324%	11.41	125%	6311	-46%	134.1	8.51	183,800	130%
36KM_PG	12.46	355%	12.24	141%	6072	-48%	133.7	8.11	189,700	137%
CALMET										
BASE_AER	3.495	28%	3.241	-36%	6640	-43%	145.8	20.21	58,180	-27%
BASE_CAL	3.505	28%	3.239	-36%	6612	-43%	145.8	20.21	58,100	-27%
BASE_PG	7.322	167%	6.734	33%	6941	-40%	144.8	19.21	127,400	59%
EXP1A_AER	4.849	77%	4.691	-7%	6383	-45%	144.5	18.91	77,580	-3%
EXP1A_CAL	4.849	77%	4.691	-7%	6385	-45%	144.5	18.91	77,600	-3%
EXP1A_PG	7.138	161%	7.337	45%	6307	-46%	143.4	17.81	112,800	41%
EXP1B_AER	5.318	94%	5.289	4%	6132	-47%	145.3	19.71	81,740	2%
EXP1B_CAL	5.303	94%	5.277	4%	6148	-47%	145.3	19.71	81,720	2%
EXP1B_PG	6.468	136%	7.022	39%	6190	-47%	144.7	19.11	100,300	25%
EXP1C_AER	7.892	188%	7.754	53%	5939	-49%	137.4	11.81	117,500	47%
EXP1C_CAL	7.981	191%	7.843	55%	5926	-49%	137.4	11.81	118,600	48%
EXP1C_PG	8.318	204%	8.167	61%	5697	-51%	137.1	11.51	118,800	49%

1. Because of the poor fit of the fitted Gaussian plume with the observed tracer concentrations in the SRL75 experiment, the Cmax and Sigma-y are not meaningful metrics of model performance.

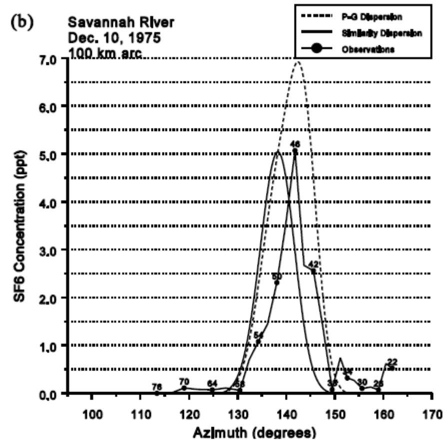


Figure 3. Simulated and observed 7-hour average plume for the Savannah River Laboratory tracer study for a) actual locations and b) observed plume offset 17° to the south.

Figure 4-2. Comparison of predicted fitted plume with observations for the SRL75 tracer experiments (Source: EPA, 1998a). Note that results from this study are not shown.

With the exception of the plume centerline statistic, the Irwin plume fitting evaluation approach was not a very useful evaluation tool for comparing the model predictions and observations using the SRL75 field experiment data. However, it is a useful tool for comparing the CALPUFF simulations using the different versions of CALPUFF/CALMET. The BASE CALPUFF/CALMET sensitivity test in this study was designed to be setup in the same fashion as the 1998 EPA tracer modeling study. Although there are some similarities, there are also some differences. For example, using the PG dispersion results in much higher CWIC in both the 1998 EPA (129,000 ppt/m^2) and BASE (127,400 ppt/m^2) sensitivity tests versus using the CAL turbulence/similarity dispersions options (77,000 ppt/m^2 for 1998 EPA and ~58,000 ppt/m^2 for BASE). The maximum estimated concentration at any of the 200 receptors along the 100 km arc using the PG dispersion are very similar for the 1998 EPA (6.9 ppt) and BASE sensitivity (6.7 ppt) scenario and lower concentrations are estimated using the CAL turbulence dispersion in the 1998 EPA (5.0 ppt) and the BASE (3.2 ppt) sensitivity test.

4.5 CONCLUSIONS OF THE SRL75 MODEL PERFORMANCE EVALUATION

Because the fit of the Gaussian plume to the observed tracer concentrations along the SRL75 100 km receptor arc did not match the observed values well, the fitted plume evaluation approach did not work well using the SRL75 database. Thus, there are few conclusions that can be drawn about the CALPUFF model performance using the SRL75 tracer field experiment data. The plume centerline evaluation is still valid and the use of CALPUFF without using meteorological observations with CALMET either through MMIF or with CALMET using no observations (NOOBS = 2) produces better plume centerline performance than when meteorological observations are used with CALMET. These results are consistent with EPA's thoughts in the 2009 IWAQM Reassessment Report (EPA, 2009a) and August 2009 Clarification Memorandum (EPA, 2009b); it is better to pass through the wind fields and other meteorological field from MM5/WRF to CALPUFF, rather than running them through CALMET, which can introduce artifacts and upset the dynamic balance of the meteorological fields.

5.0 1983 CROSS APPALACHIAN TRACER EXPERIMENT

5.1 DESCRIPTION OF THE 1983 CROSS APPALACHIAN TRACER EXPERIMENT

A series of tracer test field experiments were conducted between September 18 and October 29, 1983 over the northeastern U.S. and southeastern Canada (Ferber et al., 1986; Draxler et al., 1988). The Cross-Appalachian Tracer Experiment (CAPTEX) consisted of 5 tracer releases from Dayton, Ohio and 2 tracer releases from Sudbury, Ontario. Each release was independent of the others and was conducted when the forecast was for the tracer to pass through the center of the sampling network. Samplers were placed at a variety of locations in the northeast U.S. and southeast Canada to distances of about 1,000 km from Dayton. Although synoptic meteorological conditions were similar between releases at each location, there were large differences in the spatial concentration patterns, from narrow to wide. There was even a case of the tracer plume passing over the samplers without mixing to the surface.

The CALPUFF LRT modeling system was evaluated for various model configurations and meteorological inputs using two of the five CAPTEX tracer release experiments:

> CTEX3: The third CAPTEX tracer release occurred on October 2, 1983 where a tracer was released from Dayton, Ohio for two hours between the hours of 1400 and 1600 LST with a release rate of 18.611 g/s.
>
> CTEX5: The fifth CAPTEX tracer release occurred during the end of October with a two hour tracer release from Sudbury, Ontario between hour 23 on October 25, 1983 and hour 01 on October 26, 1983 with a release rate of 16.667 g/s.

Figure 5-1 displays the locations of the two tracer release sites and the tracer sampling network for the CAPTEX tracer field experiments. Also shown in Figure 5-1 are the CALPUFF, CALMET and MMIF modeling domains.

This section describes the evaluation of the CALPUFF LRT dispersion model using the CTEX3 and CTEX5 field experiments using numerous sensitivity tests with alternative meteorological inputs. Appendices A and B present the evaluation of the MM5 and CALMET sensitivity simulations using surface meteorological observations for the, respectively, CTEX5 and CTEX3 experiments. Appendix C presents the evaluation of six LRT dispersion models using the CTEX3 and CTEX5 field studies and common MM5 meteorological inputs.

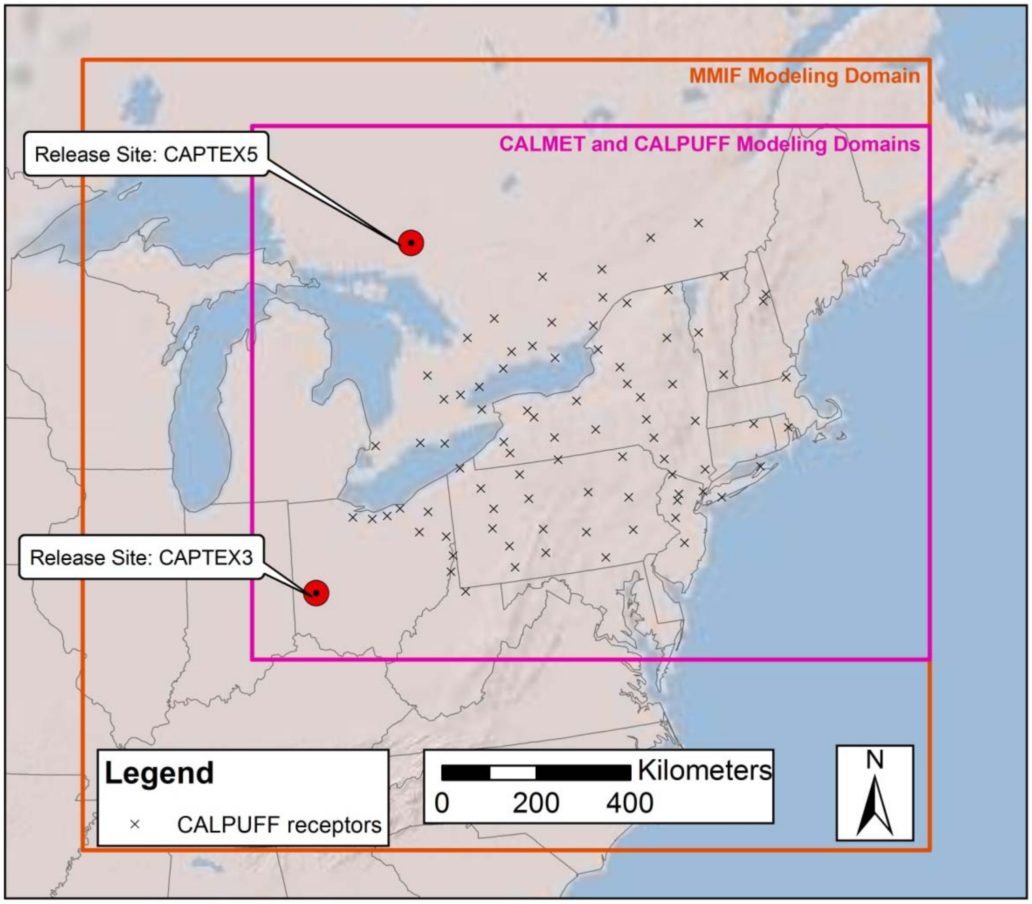


Figure 5-1. Location of Dayton and Sudbury tracer release sites and the tracer sampling network for the CAPTEX tracer field experiments.

5.2 MODEL CONFIGURATION AND APPLICATION

CALPUFF was applied using several different meteorological inputs. The first set was designed to use the same meteorological modeling technology as used in previous years to evaluate CALPUFF V4.0 only using the current regulatory versions of CALPUFF (V5.8) to document the effects of version changes. For the CTEX5 experiment period, the MM5 prognostic meteorological model was applied using grid resolutions of 80, 36 and 12 km to investigate the sensitivity of CALMET and CALPUFF model performance to MM5 grid resolution. For the CTEX3 experiment period, MM5 modeling was performed using grid resolution of 36 and 12 km, for the MM5 80 km sensitivity tests historical 80 km MM4 output data were utilized. CALMET was also run with different grid resolutions (18, 12 and 4 km) using the different MM5/MM4 grid resolution data as input. CALPUFF V5.8 was evaluated using the ATMES-II procedures using the various MM5/CALMET meteorological inputs, as well as inputs from the Mesoscale Model Interface (MMIF) tool that performs a "pass through" of the MM5 meteorological output to provide meteorological inputs to CALPUFF.

5.2.1 MM5 Prognostic Meteorological Modeling

The most recent version of the publicly available non-hydrostatic version of MM5 (version 3.7.4) was used. The MM5 preprocessors pregrid, regrid, little_r, and interpf were used to develop initial and boundary conditions. Nine separate MM5 sensitivity tests were performed for the CTEX5 field experiment period as listed in Table 5-1. As noted previously, for CTEX3 period no 80 km MM5 modeling was performed and historical 80 km MM4 data were used for the CTEX3 CALPUFF sensitivity tests.

The MM5 modeling for this study was based on three vertical structures designed to replicate common vertical structures of meteorological modeling from the 1980's to 2000's with vertical definitions of 16, 33, and 43 layers. The MM5 vertical domain definition for the 33 and 43 layer MM5 sensitivity simulations are presented in both sigma and height coordinates in Tables 5-2 and 5-3. Topographic information for the MM5 system was developed using the NCAR and the United States Geological Survey (USGS) terrain databases. Vegetation type and land use information was developed using the most recent NCAR/PSU databases provided with the MM5 distribution [available at ftp://ftp.ucar.edu/mesouser]. Standard MM5 surface characteristics corresponding to each land use category were used.

Four different grid configurations were defined for the MM5 sensitivity modeling. The first experiment (EXP1) was a baseline run using the horizontal and vertical configuration of MM4 simulations of the late 1980's and early 1990's (similar to the original MM4 dataset published by the EPA). The baseline simulation uses a single domain (no nests) with a horizontal grid resolution of 80 km and 16 vertical levels. The baseline configuration used older physics options more consistent with physics options available at the time of publication of the original EPA MM4 dataset. Physics options include the Blackadar (BLKDR) Planetary Boundary Layer (PBL) parameterization, Anthes-Kuo (AK) convective parameterization, Dudhia Radiation (DRAD), Dudhia Simple Ice Microphysics (SIM), and a 5-layer soil model (5LAYSOIL).

The second MM5 experiment (EXP2) was designed to reflect common grid and physics configurations used in numerical weather modeling for air quality simulations in the late 1990's and early 2000's. EXP2A through EXP2C used three nested domains (108, 36, and 12 km) with a 33 vertical layer vertical structure (Table 5-2). Physics options include the Medium Range Forecast model (MRF) PBL parameterization, Kain-Fritsch (KF) convective parameterization, rapid radiative transfer model (RRTM) radiation, SIM microphysics, and the 5LAYSOIL soil model. EXP2H is a variation of EXP2C, reflecting another common configuration of the period, but using the BLKDR PBL parameterization instead of the MRF PBL.

The third MM5 experiment (EXP3) was designed to reflect the more recent advances in numerical weather modeling for air quality simulations, both in terms of grid configuration and physics options. These options are largely consistent with annual MM5 simulations conducted by the EPA and the Regional Haze Regional Planning Organizations (RPOs). Consistent with EXP2, EXP3 uses three nested domains (108, 36, and 12 km). EXP3 uses the Pleim-Xu (PX) PBL parameterization, the Kain-Fritsch 2 (KF2) convective parameterization, DRAD radiation, and the Pleim-Xu (PX) land surface model (LSM).

A key facet in the MM5 sensitivity modeling was to measure the effectiveness of various four-dimensional data assimilation (FDDA) strategies on meteorological model performance and also determine the importance of assimilated fields in enhancing the performance of long range transport (LRT) model simulations. In EXP1 and EXP2 series, there are a minimum of three

MM5 runs, the first without FDDA (i.e., in forecasting mode), the second with three-dimensional analysis nudging above the PBL only, and the third using both three-dimensional analysis nudging above the PBL and surface analysis nudging below the PBL. Nudging within the PBL was turned off for temperature and mixing ratio. Default nudging strengths were used for both three-dimensional analysis and surface analysis nudging in these scenarios.

In scenarios EXP2I and EXP2J, alternative data assimilation strategies were tested while keeping the three-dimensional and surface analysis nudging. In EXP2I, the nudging strength was doubled. Observational nudging was turned on for EXP2J in addition to the nudging strengths used in EXP2I. The NCAR ds472.0 dataset was used to provide surface observations for the observational nudging.

Although new MM5 meteorological modeling was performed for the scenarios in Table 5-1 for the CTEX5 field experiment, for the CTEX3 field experiment the historical 80 km MM4 data was used for the 80 km MM5/MM4 scenarios and the FDDA sensitivity tests were not performed.

Table 5-1. Summary of CTEX5 MM5 sensitivity tests. design.

Sensitivity Test	Horizontal Grid	Vertical Layers	Physics Options	FDDA Used
EXP1A	80 km	16	BLKDR, AK, DRAD, SIM, 5LAYSOIL	No FDDA
EXP1B	80 km	16	BLKDR, AK, DRAD, SIM, 5LAYSOIL	Analysis Nudging
EXP1C	80 km	16	BLKDR, AK, DRAD, SIM, 5LAYSOIL	Analysis Nudging Surface Analysis Nudging
EXP2A	108/36/12km	33	MRF, KF, RRTM, SIM, 5LAYSOIL	No FDDA
EXP2B	108/36/12km	33	MRF, KF, RRTM, SIM, 5LAYSOIL	Analysis Nudging
EXP2C	108/36/12km	33	MRF, KF, RRTM, SIM, 5LAYSOIL	Analysis Nudging Surface Analysis Nudging
EXP2F	108/36/12km	43	BLKDR, KF, DRAD, SIM, 5LAYSOIL	No FDDA
EXP2G	108/36/12km	43	BLKDR, KF, DRAD, SIM, 5LAYSOIL	Analysis Nudging
EXP2H	108/36/12km	43	BLKDR, KF, DRAD, SIM, 5LAYSOIL	Analysis Nudging Surface Analysis Nudging
EXP2I	108/36/12km	43	BLKDR, KF, DRAD, SIM, 5LAYSOIL	Analysis Nudging Surface Analysis Nudging FDDA x 2 strength
EXP2J	108/36/12km	43	BLKDR, KF, DRAD, SIM, 5LAYSOIL	Analysis Nudging Surface Analysis Nudging FDDA x 2 strength Observational Nudging
EXP4	108/36/12km	43	PXPBL, KF2, DRAD, R2, PXLSM	Analysis Nudging Surface Analysis Nudging
4 km	4 km	43	BLKDR, KF, DRAD, SIM, 5LAYSOIL (EXP2H)	Analysis Nudging Surface Analysis Nudging

Table 5-2. MM5 sensitivity tests EXP2A through EXP2C vertical domain definition using 33 vertical layers.

k(MM5)	sigma	Press. (bar)	height(m)	depth(m)
33	0.0000	10000	14662	1841
32	0.0500	14500	12822	1466
31	0.1000	19000	11356	1228
30	0.1500	23500	10127	1062
29	0.2000	28000	9066	939
28	0.2500	32500	8127	843
27	0.3000	37000	7284	767
26	0.3500	41500	6517	704
25	0.4000	46000	5812	652
24	0.4500	50500	5160	607
23	0.5000	55000	4553	569
22	0.5500	59500	3984	536
21	0.6000	64000	3448	506
20	0.6500	68500	2942	480
19	0.7000	73000	2462	367
18	0.7400	76600	2095	266
17	0.7700	79300	1828	259
16	0.8000	82000	1569	169
15	0.8200	83800	1400	166
14	0.8400	85600	1235	163
13	0.8600	87400	1071	160
12	0.8800	89200	911	236
11	0.9100	91900	675	154
10	0.9200	92800	598	153
9	0.9300	93700	521	152
8	0.9400	94600	445	151
7	0.9500	95500	369	149
6	0.9600	96400	294	74
5	0.9700	97300	220	111
4	0.9800	98200	146	37
3	0.9850	98650	109	37
2	0.9900	99100	73	36
1	0.9950	99550	36	36
0	1.0000	100000	0	0

Table 5-3. MM5 sensitivity tests EXP2F through EXP2H vertical domain definition using 43 vertical layers.

k(MM5)	sigma	Press(mb)	height(m)	depth(m)
43	0.0000	10000	14662	409
42	0.0100	10900	14253	571
41	0.0250	12250	13682	696
40	0.0450	14050	12986	635
39	0.0650	15850	12351	724
38	0.0900	18100	11627	660
37	0.1150	20350	10966	724
36	0.1450	23050	10242	663
35	0.1750	25750	9579	710
34	0.2100	28900	8869	742
33	0.2500	32500	8127	681
32	0.2900	36100	7446	630
31	0.3300	39700	6815	587
30	0.3700	43300	6228	483
29	0.4050	46450	5745	458
28	0.4400	49600	5287	435
27	0.4750	52750	4852	415
26	0.5100	55900	4436	341
25	0.5400	58600	4095	329
24	0.5700	61300	3766	318
23	0.6000	64000	3448	307
22	0.6300	66700	3141	297
21	0.6600	69400	2844	288
20	0.6900	72100	2556	279
19	0.7200	74800	2277	271
18	0.7500	77500	2005	220
17	0.7750	79750	1785	215
16	0.8000	82000	1569	211
15	0.8250	84250	1359	206
14	0.8500	86500	1153	122
13	0.8650	87850	1031	120
12	0.8800	89200	911	119
11	0.8950	90550	792	271
10	0.9100	91900	675	154
9	0.9200	92800	598	153
8	0.9300	93700	521	152
7	0.9400	94600	445	151
6	0.9500	95500	369	149
5	0.9600	96400	294	74
4	0.9700	97300	220	74
3	0.9800	98200	146	73
2	0.9900	99100	73	44
1	0.9960	99640	29	29
0	1.0000	100000	0	0

5.2.2 CALMET Diagnostic Meteorological Modeling

The CALMET (Scire, 2000a) diagnostic meteorological model generates wind fields and other meteorological variables required by the CALPUFF LRT dispersion model in a two-step process.

In STEP 1, an initial first guess wind field is modified through parameterized diagnostic wind field effects due to terrain: blocking and deflection, channeling and slope flows. The first guess wind field can be provided using prognostic meteorological model output (e.g., MM5) or interpolated from observations. The resultant STEP 1 wind field is then modified in STEP 2 by incorporating (blending) surface and upper-air wind observations with the STEP 1 wind field in an Objective Analysis (OA) procedure. CALMET has numerous options on how to generate the STEP 1 wind field as well as how the STEP 2 OA procedure is performed. A series of CALMET sensitivity tests were performed to examine the efficacy of OA, optimal radii of influence for CALMET OA operations, and also to examine the role of horizontal grid resolution on performance of both the diagnostic meteorological model and the performance of the CALPUFF (Scire, 2000b) LRT dispersion model. CALMET was operated at three horizontal grid resolutions (18, 12 and 4 km) with input prognostic meteorological data at horizontal resolutions of 80 km (MM5 EXP1C), 36 km (MM5 EXP2H), and 12 km (MM5 EXP2H). Additionally, the Mesoscale Model Interface (MMIF) tool (Emery and Brashers, 2009) was also applied using MM5 output at 80 km (MM5 EXP1C), 36 km (MM5 EXP2H), and 12 km (MM5 EXP2H) for CTEX5. Since no 80 km MM5 data was available for CTEX3, MMIF was only used using the 36 and 12 km MM5 output for CTEX3. In addition, for CTEX5 MMIF was run using 4 km MM5 output that was generated in a “nest down” simulation from the 12 km MM5 simulation.

33 separate CALMET sensitivity tests were performed using MM5 output from the MM5 sensitivity simulations listed in Table 5-1 and the CALMET sensitivity test experimental configuration design given in Tables 5-4 and 5-5. The definitions of the 33 CALMET sensitivity tests are given in Table 5-6. CALPUFF sensitivity simulations were performed using a subset of the 33 CALMET sensitivity tests for the CTEX3 and CTEX5 tracer test field experiments. For both the CTEX3 and CTEX5 modeling periods, the CALMET EXP2 sensitivity test series was not run with CALPUFF, as well as the EXP1 series for CTEX5. The BASED CALPUFF simulation encountered an error in execution and failed to finish for the CTEX3 modeling period. The 80KM_MMIF was also not run for CTEX3 because MMIF was not designed to use MM4 data. For CTEX5, a 4 km MM5 nest down simulation was performed off of the MM5 EXP2H sensitivity test (see Figure 5-1) so that a 4KM_MMIF CALPUFF sensitivity test could also be performed.

Table 5-4. CALMET sensitivity test experiment configuration for grid resolution.

Experiment	CALMET Resolution (km)	MM5 Resolution (km)
BASE	18	80
EXP1	12	80
EXP2	4	80
EXP3	12	36
EXP4	12	12
EXP5	4	36
EXP6	4	12

Table 5-5. CALMET Objective Analysis (OA) sensitivity test configurations.

Experiment Series	RMAX1 (km)	RMAX2 (km)	NOOBS	Comment
A	500	1000	0	Use surface and upper-air met obs
B	100	200	0	Use surface and upper-air met obs
C	10	100	0	Use surface and upper-air met obs
D	0	0	2	Don’t use surface and upper-air met obs

Table 5-6. Definition of the CALMET sensitivity tests and data sources.

Sensitivity Test	MM5 Experiment and Resolution	CALMET Resolution	RMAX1/RMAX2	NOOBS	CTEX3	CTEX5
BASEA	EXP1C – 80 km	18 km	500/1000	0	Yes	Yes
BASEB	EXP1C – 80 km	18 km	100/200	0	Yes	Yes
BASEC	EXP1C – 80 km	18 km	10/100	0	Yes	Yes
BASED	EXP1C – 80 km	18 km	0/0	2	No	Yes
1A	EXP1C – 80 km	12 km	500/1000	0	Yes	No
1B	EXP1C – 80 km	12 km	100/200	0	Yes	No
1C	EXP1C – 80 km	12 km	10/100	0	Yes	No
1D	EXP1C – 80 km	12 km	0/0	2	Yes	No
2A	EXP1C – 80 km	4 km	500/1000	0	No	No
2B	EXP1C – 80 km	4 km	100/200	0	No	No
2C	EXP1C – 80 km	4 km	10/100	0	No	No
2D	EXP1C – 80 km	4 km	0/0	2	No	No
3A	EXP2H – 36 km	12 km	500/1000	0	Yes	Yes
3B	EXP2H – 36 km	12 km	100/200	0	Yes	Yes
3C	EXP2H – 36 km	12 km	10/100	0	Yes	Yes
3D	EXP2H – 36 km	12 km	0/0	2	Yes	Yes
4A	EXP2H – 12 km	12 km	500/1000	0	Yes	Yes
4B	EXP2H – 12 km	12 km	100/200	0	Yes	Yes
4C	EXP2H – 12 km	12 km	10/100	0	Yes	Yes
4D	EXP2H – 12 km	12 km	0/0	2	Yes	Yes
5A	EXP2H – 36 km	4 km	500/1000	0	Yes	Yes
5B	EXP2H – 36 km	4 km	100/200	0	Yes	Yes
5C	EXP2H – 36 km	4 km	10/100	0	Yes	Yes
5D	EXP2H – 36 km	4 km	0/0	2	Yes	Yes
6A	EXP2H – 12 km	4 km	500/1000	0	Yes	Yes
6B	EXP2H – 12 km	4 km	100/200	0	Yes	Yes
6C	EXP2H – 12 km	4 km	10/100	0	Yes	Yes
6D	EXP2H – 12 km	4 km	0/0	2	Yes	Yes
80KM_MMIF	EXP1C – 80 km	MMIF	NA	NA	No	Yes
36KM_MMIF	EXP2H – 36 km	MMIF	NA	NA	Yes	Yes
12KM_MMIF	EXP2H – 12 km	MMIF	NA	NA	Yes	Yes
4KM_MMIF	4 km EXP2H nest down	MMIF	NA	NA	No	Yes

5.3 QUALITY ASSURANCE

Quality assurance (QA) of the CALMET and CALPUFF sensitivity modeling was performed by analyzing the run control files to confirm that the intended options and inputs of each sensitivity test were used. For the MM5 datasets, performance for meteorological parameters of wind (speed and direction), temperature, and humidity (mixing ratio) are examined. For the CALMET experiments, just model estimated winds (speed and direction) were compared to observations because the two-dimensional temperature and relative humidity fields output are simple interpolated fields of the observations. Therefore, the performance evaluation for CALMET was restricted to winds where the majority of change can be induced by both diagnostic terrain adjustments and varying the OA strategy. Note that except for the NOOBS = 2 CALMET sensitivity tests (experiment K), surface meteorological observations are blended in the wind fields used in the CALMET STEP 2 OA procedure. Thus, this is not a true independent evaluation as the surface meteorological observations used in the evaluation were also used as input into CALMET.

The METSTAT software (Emery et al., 2001) was used to match MM5 output with observation data. The MMIFStat software (McNally, 2010) tool was used to match CALMET output with observation data. Emery and co-workers (2001) have developed a set of "benchmarks" for comparing prognostic meteorological model performance statistics metrics. These benchmarks were developed after examining the performance of the MM5 and RAMS prognostic meteorological models for over 30 applications. The purpose of the benchmarks is not to assign a passing or failing grade, rather it is to put the prognostic meteorological model performance in context. The surface meteorological model performance benchmarks from Emery et al., (2001) are displayed in Table 5-7. Note that the wind speed RMSE benchmark was also used for wind speed MNGE given the similarity of the RMSE and MNGE performance statistics. These benchmarks are not applicable for diagnostic model evaluations.

Table 5-7. Wind speed and wind direction benchmarks used to help judge the performance of *prognostic* meteorological models (Source: Emery et al., 2001).

Wind Speed	Root Mean Squared Error (RMSE) Mean Normalized Bias (NMB) Index of Agreement (IOA)	≤ 2.0 m/s ≤ ±0.5 m/s ≥ 0.6
Wind Direction	Mean Normalized Gross Error (MNGE) Mean Normalized Bias (MNB)	≤ 30° ≤ ±10°
Temperature	Mean Normalized Gross Error (MNGE) Mean Normalized Bias (NMB) Index of Agreement (IOA)	≤ 2.0 K ≤ ±0.5 m/s ≥ 0.8
Humidity	Mean Normalized Gross Error (MNGE) Mean Normalized Bias (NMB) Index of Agreement (IOA)	≤ 2.0 g/kg ≤ ±1.0 g/kg ≥ 0.6

The MM5 and CALMET comparisons to observations for CTEX3 and CTEX5 are provided in the Appendix. The key findings of the CTEX5 MM5 and CALMET model performance evaluation are as follows:

- The MM5 performance using the MRF PBL scheme (EXP2A-C) was extremely poor. For example the temperature exhibited an underestimation bias of over -4 °K, compared to the benchmark of ≤±0.5 °K. Thus, MM5 sensitivity simulations using MRF PBL scheme were discontinued.
- The MM5 wind speed, and especially wind direction, model performance is noticeably better when FDDA was utilized.
- The "A" series of CALMET runs (RMAX1/RMAX2 = 500/1000) always has a wind speed underestimation bias.
- The "C" and "D" series of CALMET sensitivity tests exhibit wind performance that is comparable to the MM5 simulation used as input to CALMET.
- The 36 km and 12 km MM5 simulations exhibit substantially better model performance than the 80 km MM5 simulation.

The CTEX3 and CTEX5 CALMET comparison for wind speed and direction needs to be viewed with the caveat that because the winds are used as input in some of the sensitivity tests, then this is not a true independent evaluation. Thus, it is at all not surprising that the CALMET wind performance at the monitor locations is improved in the CALMET sensitivity tests that used meteorological observations as input compared to those that used no observations. As clearly pointed out in the 2009 Revised IWAQM Guidance (EPA, 2009a), the better wind model performance at the monitors produced when CALMET blends observed surface wind data in the wind fields can produce unrealistic discontinuities and other artifacts in the wind fields.

5.4 CALPUFF MODEL PERFORMANCE EVALUATION FOR CAPTEX

CALPUFF was applied for the CTEX3 and CTEX5 tracer release field experiments using the meteorological inputs corresponding to each of the meteorological sensitivity tests given in Table 5-6. Figure 5-1, presented earlier, displays the locations of the CTEX3 (Dayton, Ohio) and CTEX5 (Sudbury, Ontario) tracer release sites and the tracer monitoring network in northeastern U.S. and southeastern Canada.

A common CALPUFF model configuration was used in all sensitivity tests. This was done to isolate the sensitivity of the model to the different meteorological inputs and not confound the interpretation by changing the CALPUFF model configuration. The CALPUFF model configuration used the options listed in Table 5-8. Mostly default options were utilized for CALPUFF. One parameter that was not the default value was for vertical puff splitting. The default for vertical puff splitting is to turn it on using the vertical puff splitting flag (IRESPLIT) for just hour 17. After the vertical puff splitting flag is turned on a puff performs vertical puff splitting if certain criteria are met based on criteria using the ZISPLIT and ROLDMAX parameters for which default values were specified (see discussion on CALPUFF puff splitting sensitivity tests for the ETEX experiment in Chapter 6 for more details). Once a puff splits in the vertical, the vertical puff splitting is turned off and the puff is not allowed to split until after the puff splitting flag is turned on again at hour 17. In the CTEX3 and CTEX5 CALPUFF sensitivity simulations, the IRESPLIT input was set to turn on the vertical puff splitting flag 24 hours a day so that vertical puff splitting flag for all puffs is always on so vertical puff splitting will always occur whenever the other criteria are met.

Table 5-8. CALPUFF model configuration used in the CTEX3 and CTEX5 sensitivity tests.

Option	Value	Comment
MGAUSS	1	Use Gaussian vertical distribution initially
MCTADJ	0	No terrain adjustment
MSLUG	0	Near-field puffs not modeled as slugs
MTRANS	1	Use transitional plume rise
MTIP	1	Use stack tip downwash
MBDW	1	Use ISC method to simulate building downwash
MSHEAR	1	Model vertical wind shear above stack top
MSPLIT	1	Use puff splitting
MCHEM	0	No chemistry
MWET	0	No wet deposition
MDRY	0	No dry deposition
MDISP	2	Dispersion from internally calculate sigma-y and sigma-z using turbulence
MTURBW	3	Both sigma-y and sigma-z from PROFILE.DAT
MDISP3	3	PG dispersion coefficients for rural areas
MCTURB	2	Use AERMOD subroutine for turbulence variables
MROUGH	0	Don't adjust sigma-y and sigma-z for roughness
MPARTL	1	Use partial plume penetration
MTINV	0	Compute strength of temperature inversion
MPDF	1	Use PDF for dispersion under convective conditions
NSPLIT	3	Split puff into 3 puffs when performing vertical puff splitting
IRESPLIT	24*1	Keep vertical puff splitting flag on all the time (default is just hour 17 = 1, rest 0)
ZISPLIT	100	Vertical splitting is allowed if mixing height exceeds 100 m.
ROLDMAX	0.25	Vertical splitting is allowed if ratio of maximum to current mixing height is > 0.25
NSPLITH	5	Number of puffs that result when horizontal splitting is performed
SYSPLITH	1.0	Minimum width of puff (in grid cells) before horizontal splitting
SHSPLITH	2.0	Minimum puff elongation factor for horizontal splitting
CNSPLITH	1.E-7	Minimum concentrations (g/m^3) in puff for horizontal splitting

5.4.1 CALPUFF CTEX3 Model Performance Evaluation

Because of the large number of CALPUFF sensitivity tests performed for the CTEX3 tracer test field experiment, they are first compared by groups that used a common MM5/MM4 prognostic meteorological grid resolution output as input into CALMET or MMIF. We then compare the CALPUFF sensitivity tests using different MM4/MM5 grid resolutions but common CALMET/MMIF configurations to determine the sensitivity of MM4/MM5 grid resolution on CALPUFF tracer model performance.

5.4.1.1 CALPUFF CTEX3 Model Evaluation using 80 km MM4 Data

Figure 5-2 displays the spatial model performance statistics metrics for the CALPUFF CTEX3 sensitivity tests that used the 80 km MM4 data. There are variations in the rankings across the spatial statistical performance metrics for the CALPUFF sensitivity tests using the 80 km MM4 data. These sensitivity tests use the finest CALMET grid resolution tested in this series (12 km vs. 18 km) and minimizes the influence of the meteorological observations either through the lowest RMAX1/RMAX2 values (EXP1C) or not using meteorological observations at all by running CALMET in the NOOBS = 2 mode (EXP1D).

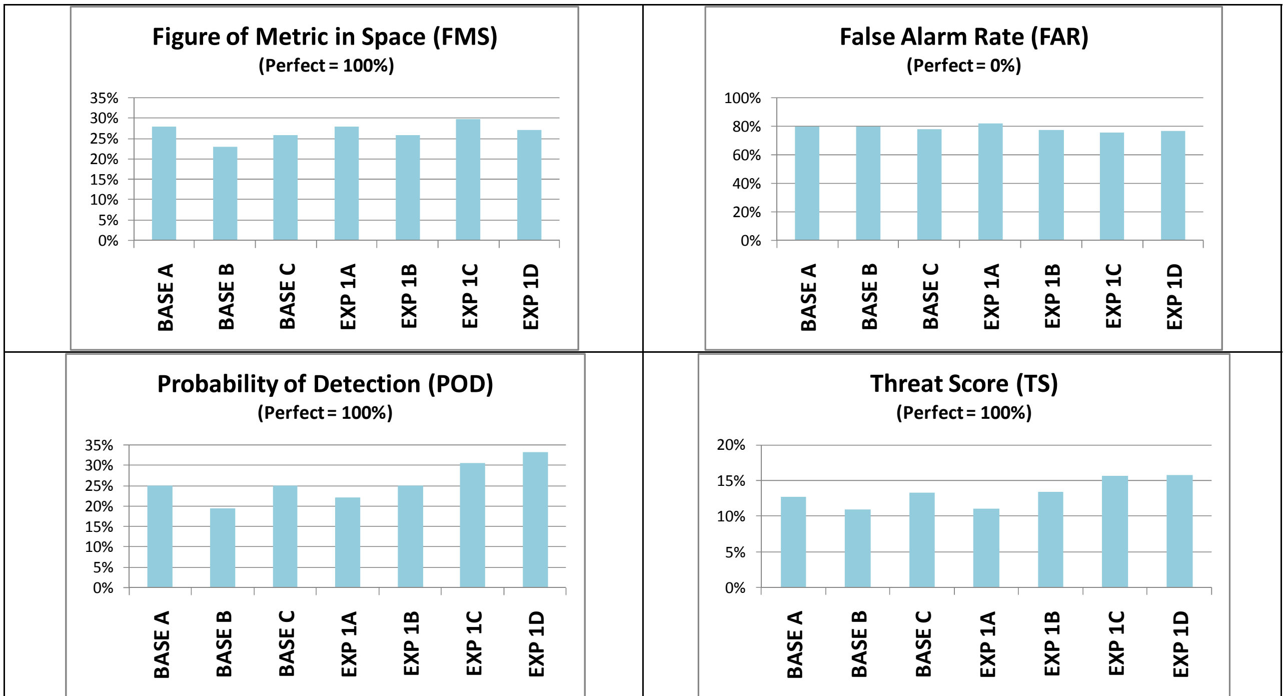


Figure 5-2. Spatial model performance statistics for the CTEX3 CALPUF sensitivity tests that used the 80 km MM4 data.

The global model performance statistics for the CALPUFF sensitivity tests using 80 km MM4 data are compared in Figure 5-3.

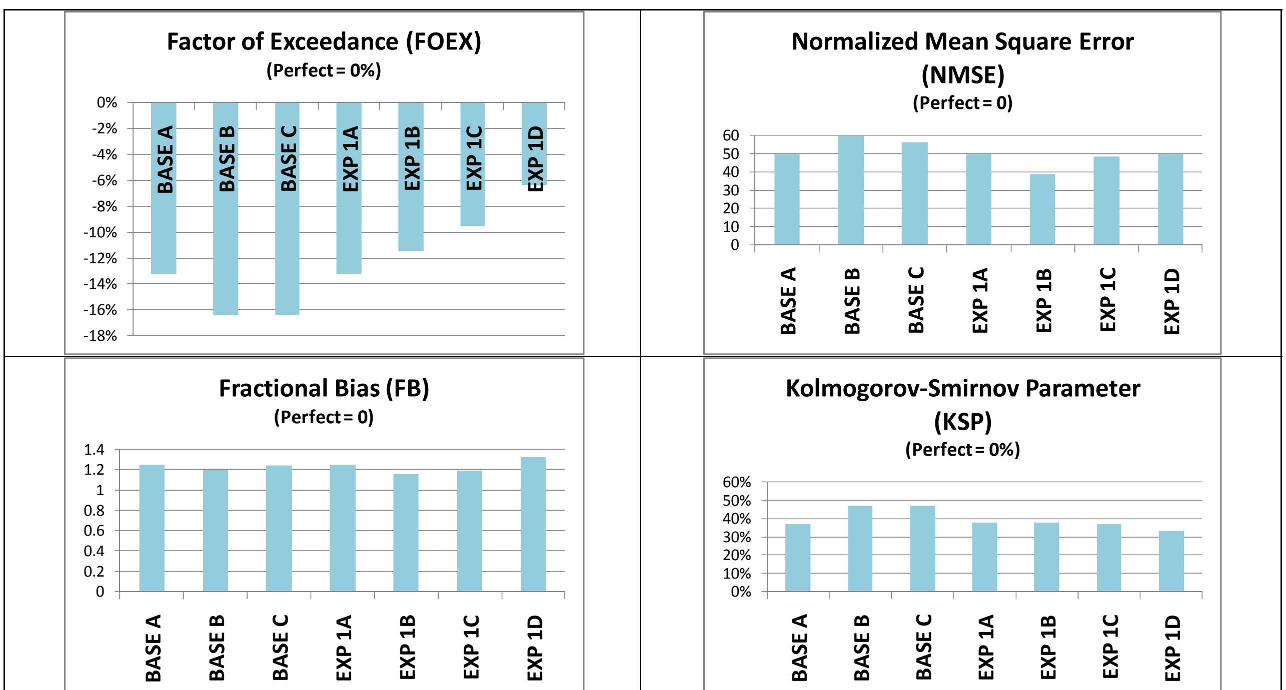
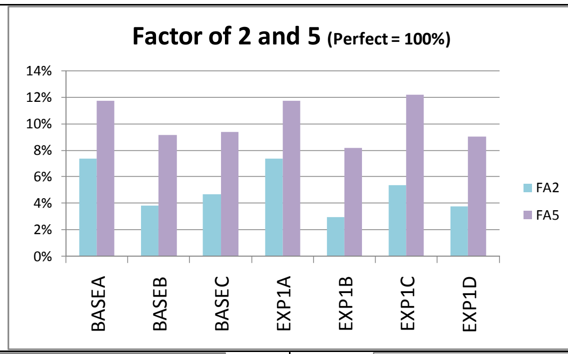
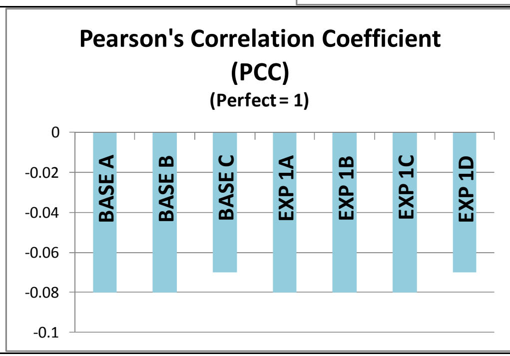


Figure 5-3a. Global model performance statistics for the CTEX3 CALPUFF sensitivity tests using the 80 km MM4 data.





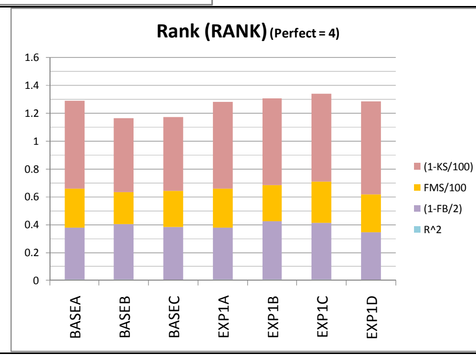


Figure 5-3b. Global model performance statistics for the CTEX3 CALPUFF sensitivity tests using the 80 km MM4 data.

<u>5.4.1.2 CALPUFF CTEX3 Model Evaluation using 36 km MM5 Data</u>

For the CTEX3 CALPUFF sensitivity tests using the 36 km MM5 data, there are 9 CALPUFF sensitivity tests 7 that use CALMET meteorological inputs with 12 and 4 km grid resolution and different OA options and one that uses MMIF meteorological inputs that as a MM5 "pass through" tool uses 36 km grid resolution.

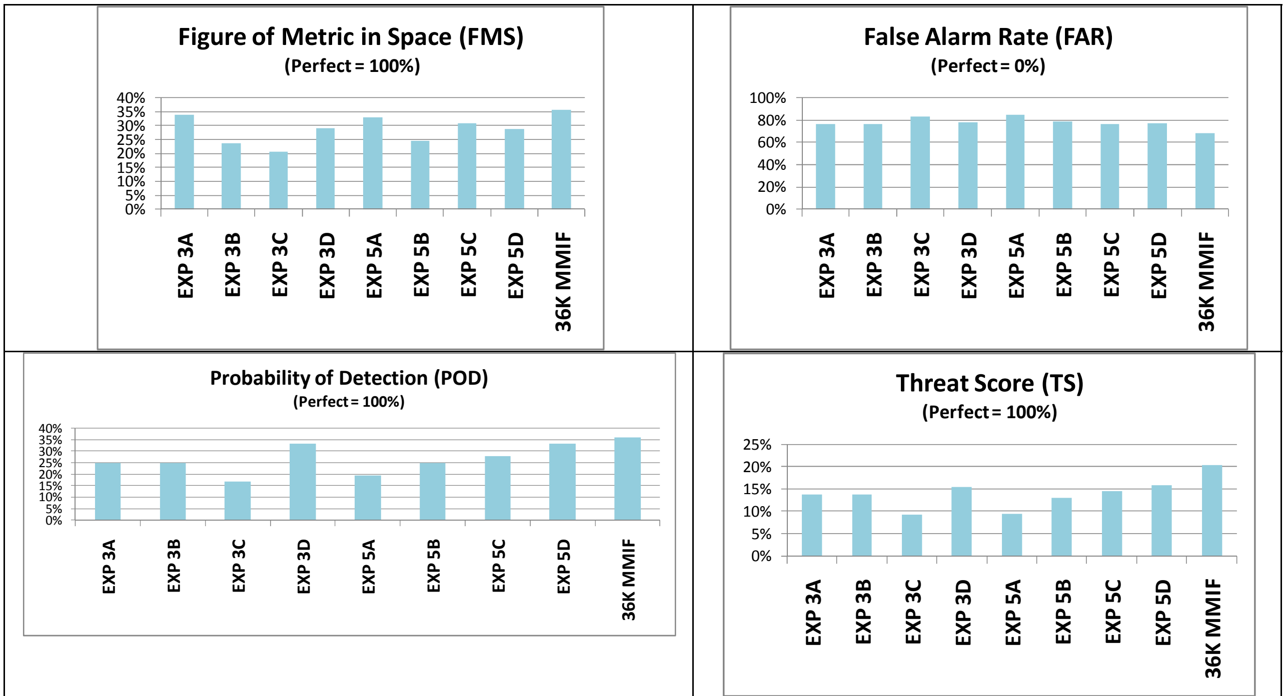


Figure 5-4. Spatial model performance statistics for the CTEX3 CALPUF sensitivity tests that used the 36 km MM5 data.

The global model performance statistics for the CALPUFF sensitivity tests using 36 km MM5 data are shown in Figure 5-5.

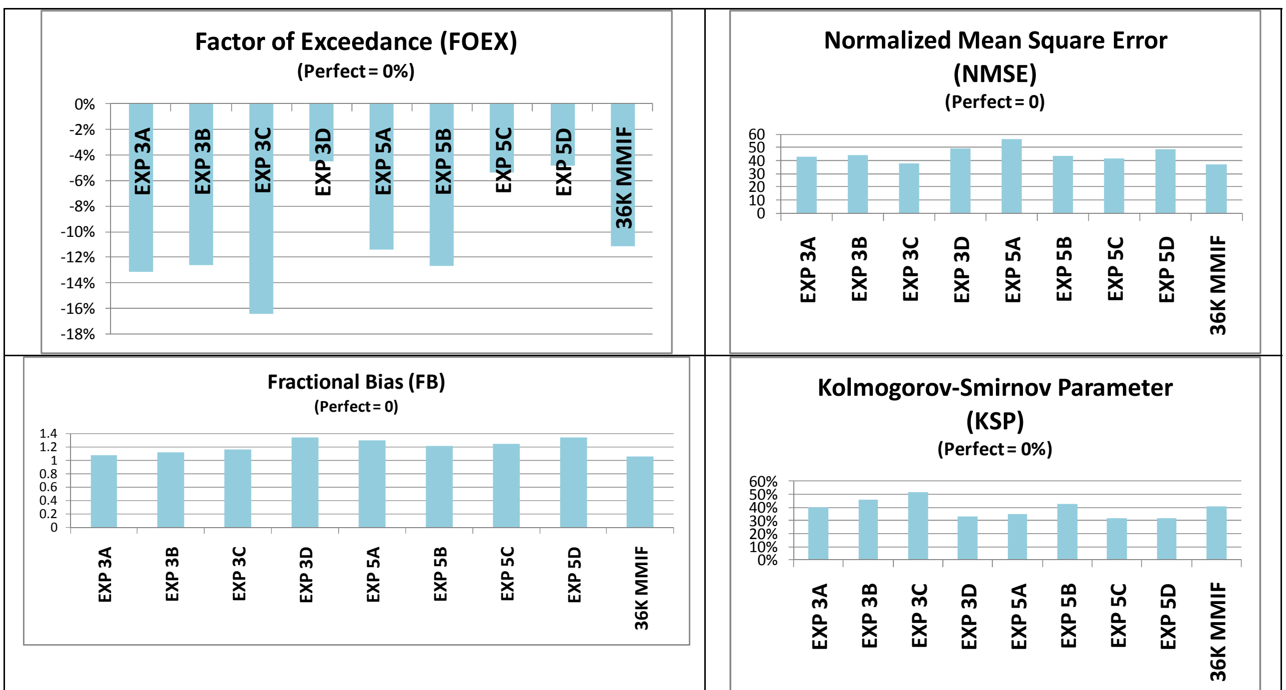


Figure 5-5a. Global model performance statistics for the CTEX3 CALPUFF sensitivity tests using the 36 km MM5 data.

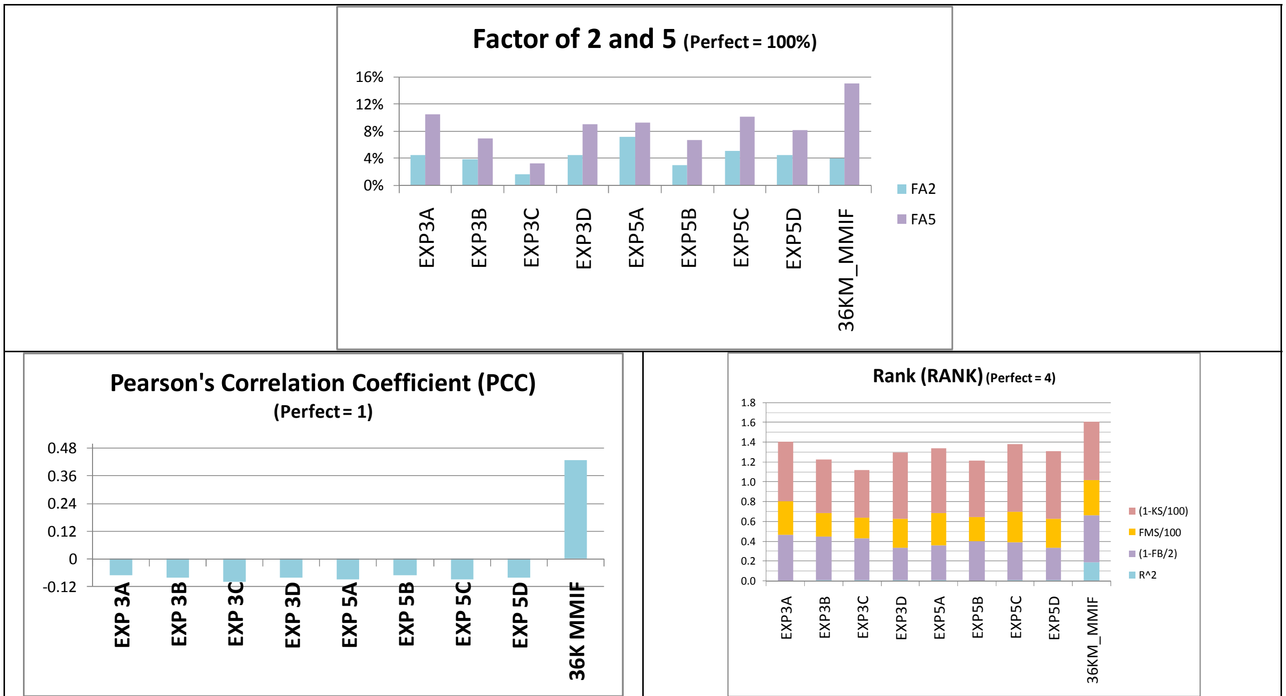


Figure 5-5b. Global model performance statistics for the CTEX3 CALPUFF sensitivity tests using the 36 km MM5 data.

5.4.1.3 CALPUFF CTEX3 Model Evaluation using 12 km MM5 Data

The spatial model performance statistical metrics for the CTEX3 CALPUFF sensitivity tests using 12 km MM5 data are shown in Figure 5-6.

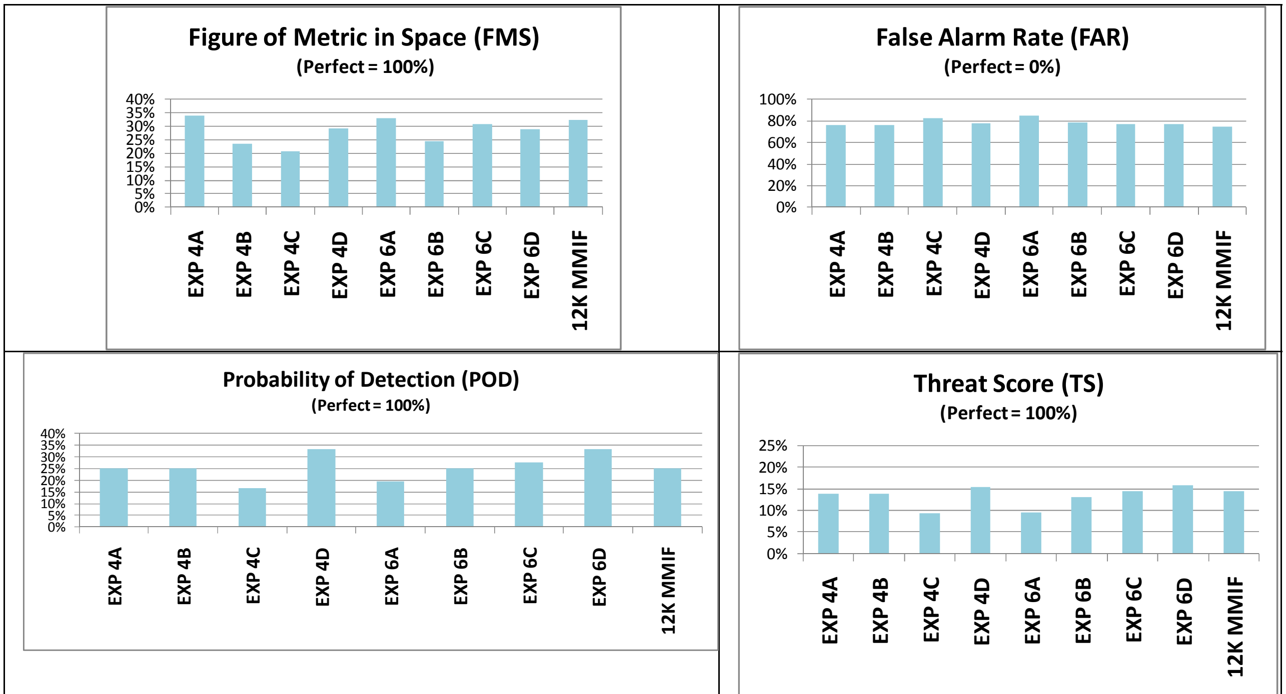


Figure 5-6. Spatial model performance statistics for the CTEX3 CALPUF sensitivity tests that used the 12 km MM5 data.

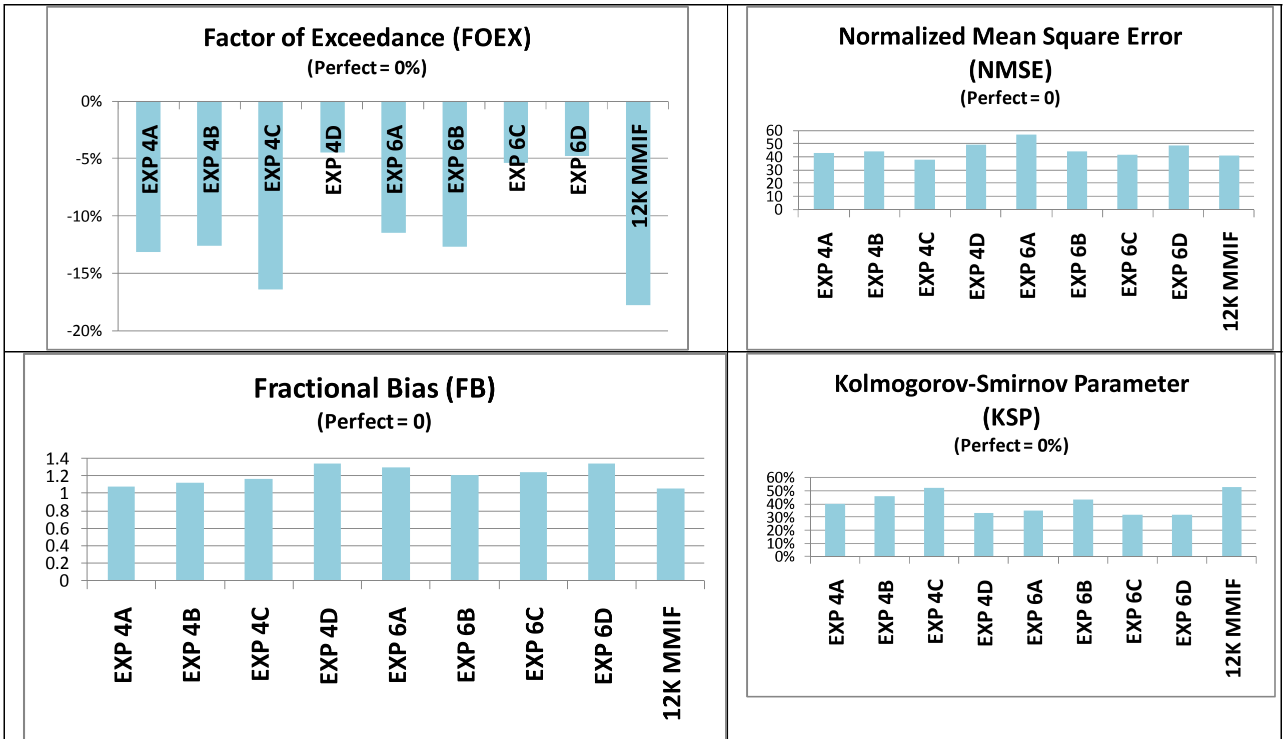


Figure 5-7a. Global model performance statistics for the CTEX3 CALPUFF sensitivity tests using the 12 km MM5 data.

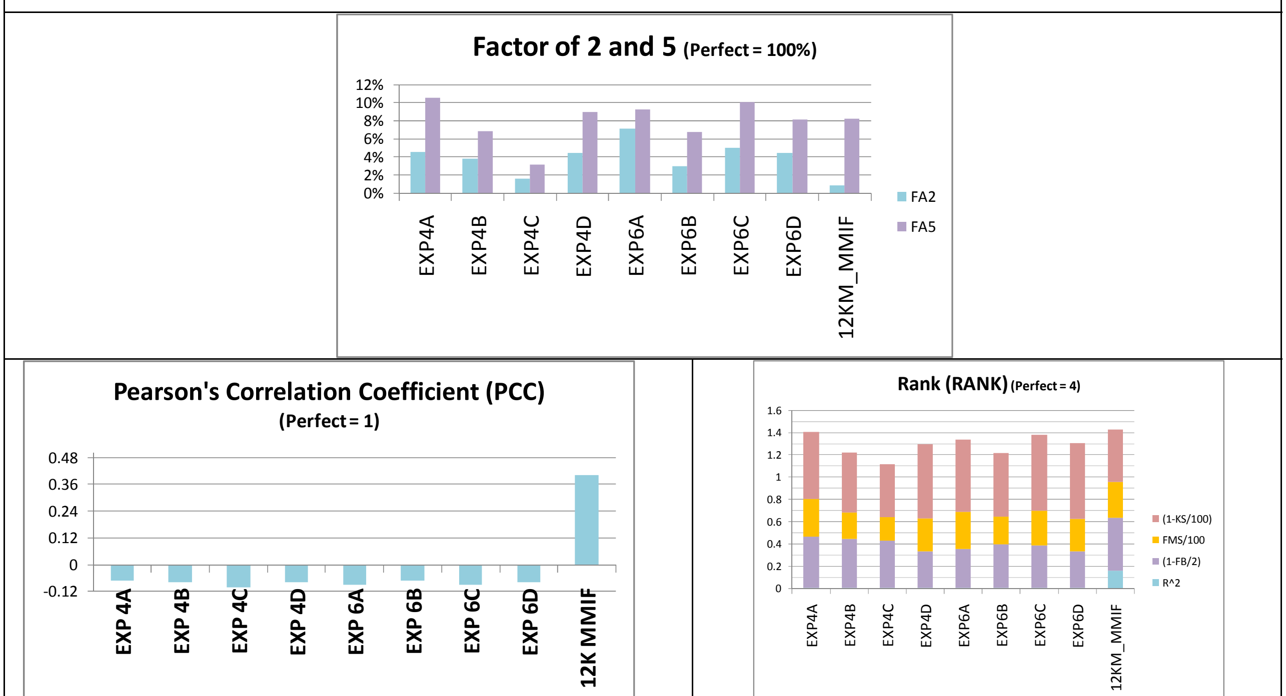


Figure 5-7b. Global model performance statistics for the CTEX3 CALPUFF sensitivity tests using the 12 km MM5 data (high scores indicate better model performance).

5.4.1.4 Comparison of CALPUFF CTEX3 Model Evaluation using Different MM4/MM5 Grid Resolutions

In the final series of CTEX3 CALPUFF sensitivity tests we grouped the "B" and "D" series of CALPUFF/CALMET sensitivity tests that use the EPA-FLM recommended RMAX1/RMAX2 settings (100/200) and no met observations, respectively, using the various MM5 data and grid resolutions in CALMET with the 12KM_MMIF and 36KM_MMIF CALPUFF sensitivity tests. The spatial model performance statistics are shown in Figure 5-8. The 36KM_MMIF and 12KM_MMIF have the best and second best FMS statistics (36% and 32%) followed by EXP3D and EXP6D (29%). The worst performing FMS statistics are given by the "B" series of CALPUFF/CALMET sensitivity tests with values ranging from 23% to 25%. The 36KM_MMIF has by far the lowest (best) FAR value (68%) followed by 12KM _MMIF (74%) with the "B" series of CALPUFF/CALMET sensitivity tests having the worst (highest) FAR values that approach 80%. A clear pattern is seen in the POD statistic for the CALPUFF/CALMET sensitivity tests with the "D" series using no met observations clearly performing better (33%) than the "B" series (19% to 25%). However, the best performing CALPUFF sensitivity test using the POD statistics is 36KM_MMIF (36%). Oddly, the 12KM_MMIF is one of the worst performing configurations with POD value the same as many of the "B" series (25%). 36KM_MMIF (20%) is also the best performing CALPUFF sensitivity test according to the TS statistic with the no met observations ("D" series) CALPUFF/CALMET sensitivity tests (15% to 16%) and 12KM_MMIF (15%) having better TS values than when met observations are used with CALMET (10% to 14%).

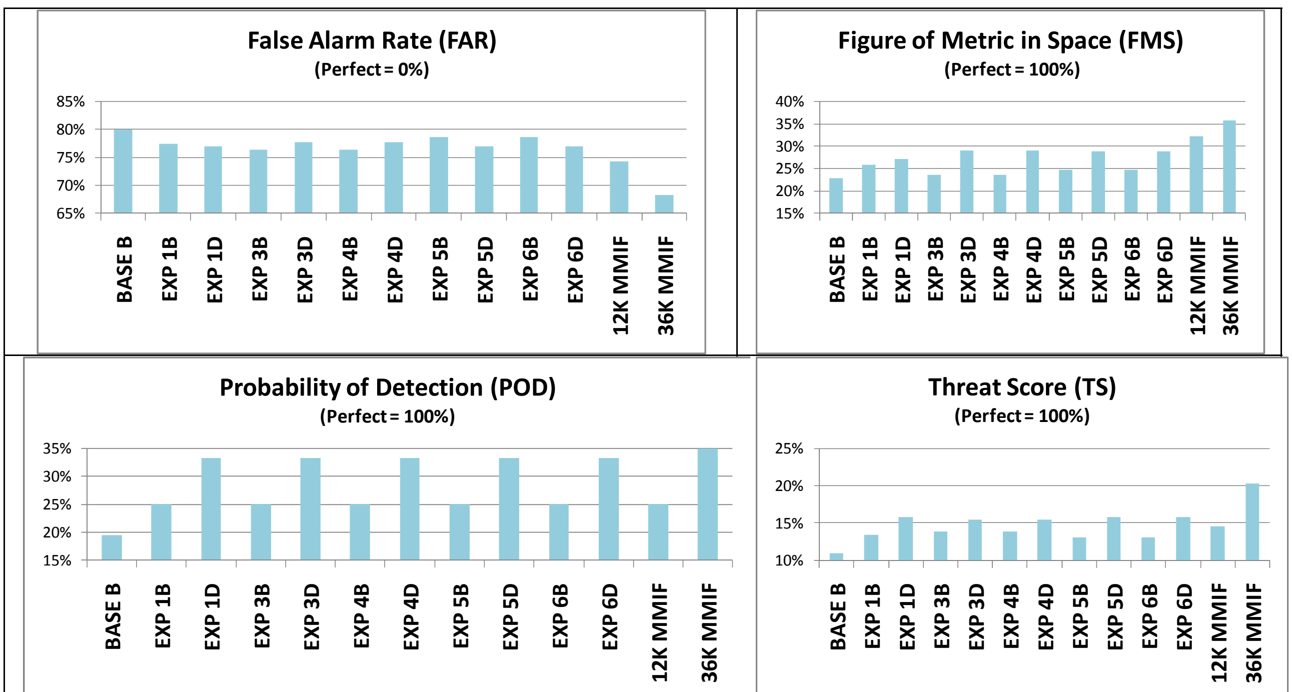


Figure 5-8. Spatial model performance statistics for the CTEX3 CALPUF sensitivity tests using different MM4/MM5 grid resolutions.

For the FOEX and KSP global statistical metrics, the "D" series of CALPUFF/CALPUFF sensitivity tests is clearly performing better than the "B" series with 12KM_MMIF one of the worst performing model configurations for these two statistics (Figure 5-9a). However, the "B" series of CALPUFF/CALMET sensitivity tests is exhibiting lower bias (FB) and error (NMSE) than the “D” series of CALPUFF/CALMET sensitivity tests with the 36KM_MMIF exhibiting the lowest bias and error statistics; 12KM_MMIF has the second lowest FB and third lowest NMSE. For the within a factor of 2 and 5 statistics the “D” series performs better than the “B” series of CALPUFF/CALMET sensitivity tests. The 12KM_MMIF has by far the lowest FA2 metric but has a FA5 metric that is comparable to the “D” series of CALPUFF/CALMET sensitivity tests. By far the best performing model configuration for the FA5 metric is 36KM_MMIF whose value (15%) is almost double the next best performing CALPUFF model configurations (7% to 9%). The 36KM_MMIF (0.43) followed closely by the 12KM_MMIF (0.40) are by far the best performing sensitivity tests according to the correlation coefficient statistical metric with the CALPUFF/CALMET tracer estimates showing a small negative correlation with the observations (-0.07 to -0.08). According to the composite RANK statistic, 36KM_MMIF (1.61) is the best performing CALPUFF sensitivity test of this group followed by 12KM_MMIF (1.43). The CALPUFF/CALMET RANK statistics range from 1.16 to 1.32 with the “D” series typically performing better (~1.3) than the “B” series (~1.2) with the exception of EXP1B (1.3).

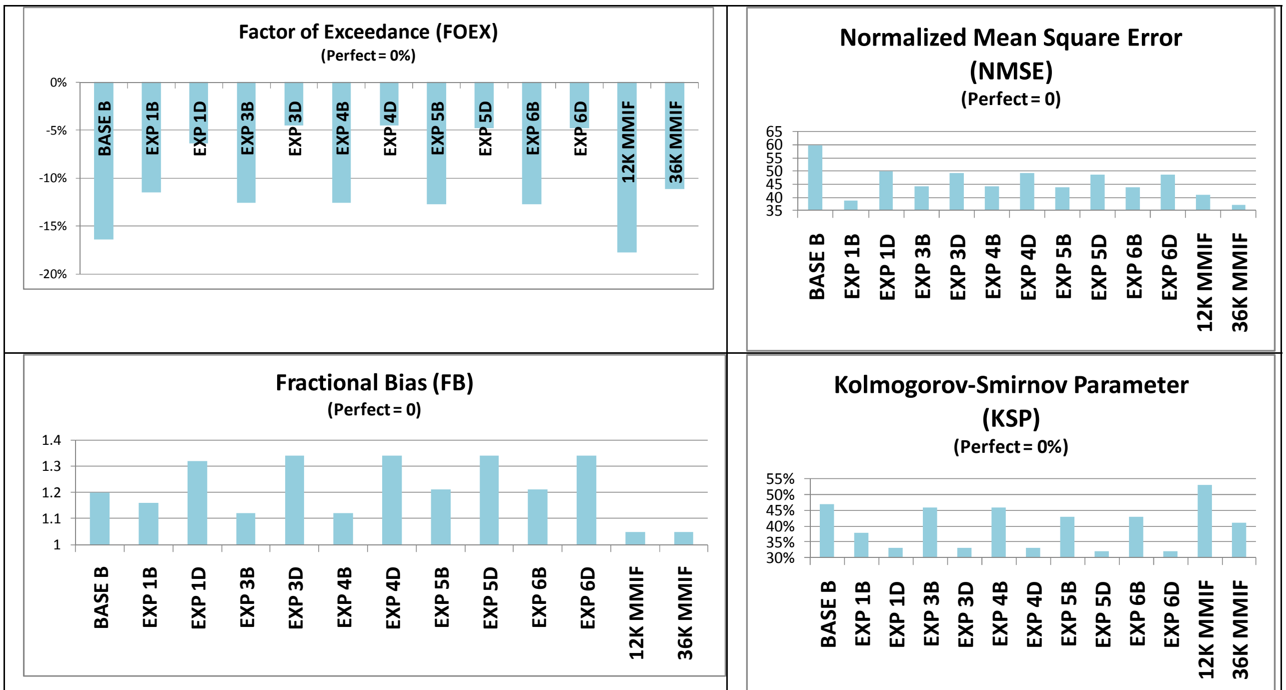


Figure 5-9a. Global model performance statistics for the CTEX3 CALPUFF sensitivity tests using different MM4/MM5 grid resolutions.

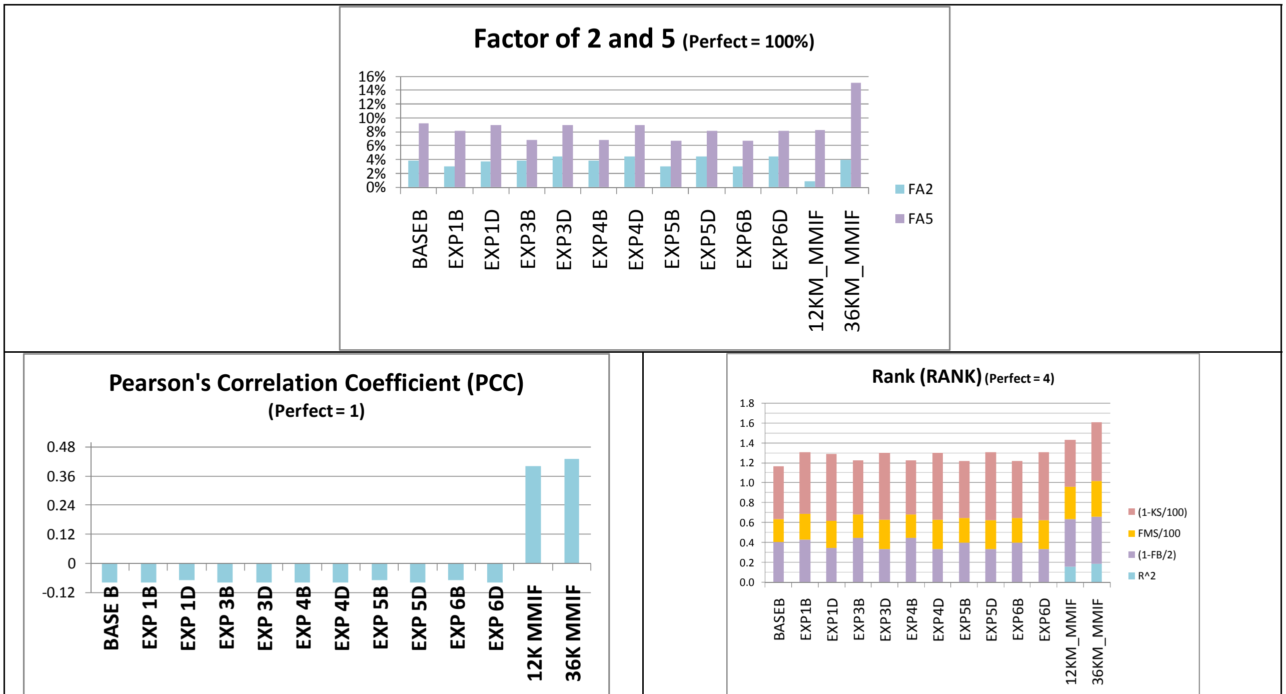


Figure 5-9b. Global model performance statistics for the CTEX3 CALPUFF sensitivity tests using different MM4/MM5 grid resolutions (high scores indicate better model performance).

5.4.1.5 Rankings of CTEX3 CALPUFF Sensitivity Tests using the RANK Statistic

The ranking of all of the CTEX3 CALPUFF sensitivity tests using the composite RANK model performance statistics is given in Table 5-9. The 36KM_MMIF (1.61) is the highest ranked CALPUFF sensitivity test using RANK followed by 12KM_MMIF (1.43) which is very close to EXP3A and EXP4A that are tied for third with a RANK value of 1.40. It is interesting to note that the EXP3A and EXP4A CALPUFF/CALMET sensitivity test that uses the, respectively, 36 km and 12 km MM5 data with 12 km CALMET grid resolution and RMAX1/RMAX2 values of 500/1000 is tied for third best performing CALPUFF/CALMET configuration using the RANK statistic, but the same model configuration with alternative RMAX1/RMAX2 values of 10/100 (EXP3C and EXP4C) degrades the model performance to the worst performing CALPUFF configuration according to the RANK statistics with a RANK value of 1.12.

Based on the RANK statistic and the CALPUFF sensitivity test rankings in Table 5-9 we conclude the following for the CTEX3 CALPUFF sensitivity tests:

- The CALPUFF MMIF sensitivity tests are the best performing configuration for the CTEX3 experiments.
- The CALPUFF/CALMET “B” series (RMAX1/RMAX2 = 100/200) appears to be the worst performing configuration for RMAX1/RMAX2.
- The CALMET/CALPUFF “A” series seems to be the best performing RMAX1/RMAX2 setting (500/1000) followed by the “C” series (10/100) then “D” series (no met observations).
- Ignoring the “B” series of sensitivity tests, the CALPUFF/CALMET sensitivity tests that use higher MM5 grid resolution (36 and 12 km) tend to produce better model performance than those that used the 80 km MM4 data.

- When using the "A" series model configuration, the use of higher CALMET resolution does not produce better CALPUFF model performance, however for the "C" and "D" series of CALMET runs use of higher CALMET grid resolution does produce better CALPUFF model performance.
- Note that the finding that CALPUFF/CALMET model performance using CALMET wind fields based on setting RMAX1/RMAX2 = 100/200 (i.e., the "B" series) produces worse CALPUFF model performance for simulating the observed atmospheric tracer concentrations is in contrast to the CALMET evaluation that found the "B" series produced winds closest to observations (see Appendices A and B). Since the CALPUFF tracer evaluation is an independent evaluation of the CALMET/CALPUFF modeling system, whereas the CALMET surface wind evaluation is not, the CALPUFF tracer evaluation may be a better indication of the best performing CALMET configuration. The CALMET "B" series approach for blending the wind observations in the wind fields may just be the best approach for getting the CALMET winds to match the observations at the monitoring sites, but at the expense of degrading the wind fields.

Table 5-9. Final Rankings of CALPUFF CTEX3 Sensitivity Tests.

Ranking	Sensitivity Test	RANK Statistics	MM5 (km)	CALGRID (km)	RMAX1/RMAX2	Met Obs
1	36KM_MMIF	1.610	36	--	--	--
2	12KM_MMIF	1.430	12	--	--	--
3	EXP3A	1.400	36	12	500/1000	Yes
4	EXP4A	1.400	12	12	500/1000	Yes
5	EXP5C	1.380	36	4	10/100	Yes
6	EXP6C	1.380	12	4	10/100	Yes
7	EXP1C	1.340	36	18	10/100	Yes
8	EXP5A	1.340	36	4	500/1000	Yes
9	EXP6A	1.340	12	4	500/1000	Yes
10	EXP5D	1.310	36	4	--	No
11	EXP6D	1.310	12	4	--	No
12	EXP1B	1.300	36	18	100/200	Yes
13	EXP3D	1.300	36	12	--	No
14	EXP4D	1.300	12	12	--	No
15	BASEA	1.290	80	18	500/1000	Yes
16	EXP1D	1.290	36	18	--	No
17	EXP1A	1.280	36	18	500/1000	Yes
18	EXP3B	1.220	36	12	100/200	Yes
19	EXP5B	1.220	36	4	100/200	Yes
20	EXP4B	1.220	12	12	100/200	Yes
21	EXP6B	1.220	12	4	100/200	Yes
22	BASEC	1.170	80	18	10/100	Yes
23	BASEB	1.160	80	18	100/200	Yes
24	EXP3C	1.120	36	12	10/100	Yes
25	EXP4C	1.120	12	12	10/200	Yes

5.4.2 CALPUFF CTEX5 Model Performance Evaluation

The model performance of the CALPUFF sensitivity tests for the CTEX5 (October 25, 1983) field experiment are presented below grouped by MM5 grid resolution. The MM5 output were used as input to the CALMET or MMIF meteorological drivers for CALPUFF, as was done for the

CTEX3 discussed in Section 5.4.1. As noted in Table 5-6, CTEX5 CALPUFF sensitivity tests were not performed for the EXP1 and EXP2 series of experiments.

5.4.1.1 CALPUFF CTEX5 Model Evaluation using 80 km MM5 Data

The spatial model performance statistics for the CTEX5 CALPUFF sensitivity tests using the 80 km MM5 data are shown in Figure 5-10. The BASEA and BASEB sensitivity tests are performing the best followed by BASEC, then BASED with 80KM_MMIF coming in last.

Figure 5-10. Spatial model performance statistics for the CTEX5 CALPUF sensitivity tests that used the 80 km MM5 data.

Although 80KM_MMIF has the lowest FOEX statistic, for all the other global statistic it is the worst or almost worst performing CALPUFF sensitivity test using 80 km MM5 data. BASEA has the best bias, error, FA2 and FA5 statistics of this group with either BASEB or BASEC coming in second and then BASED next to last and 80KM_MMIF last. The RANK composite statistics ranks BASEA (2.06) and BASEC (2.05) the highest followed by BASEB (1.82) and BASED (1.79) next and 80KM_MMIF (1.42) in last.

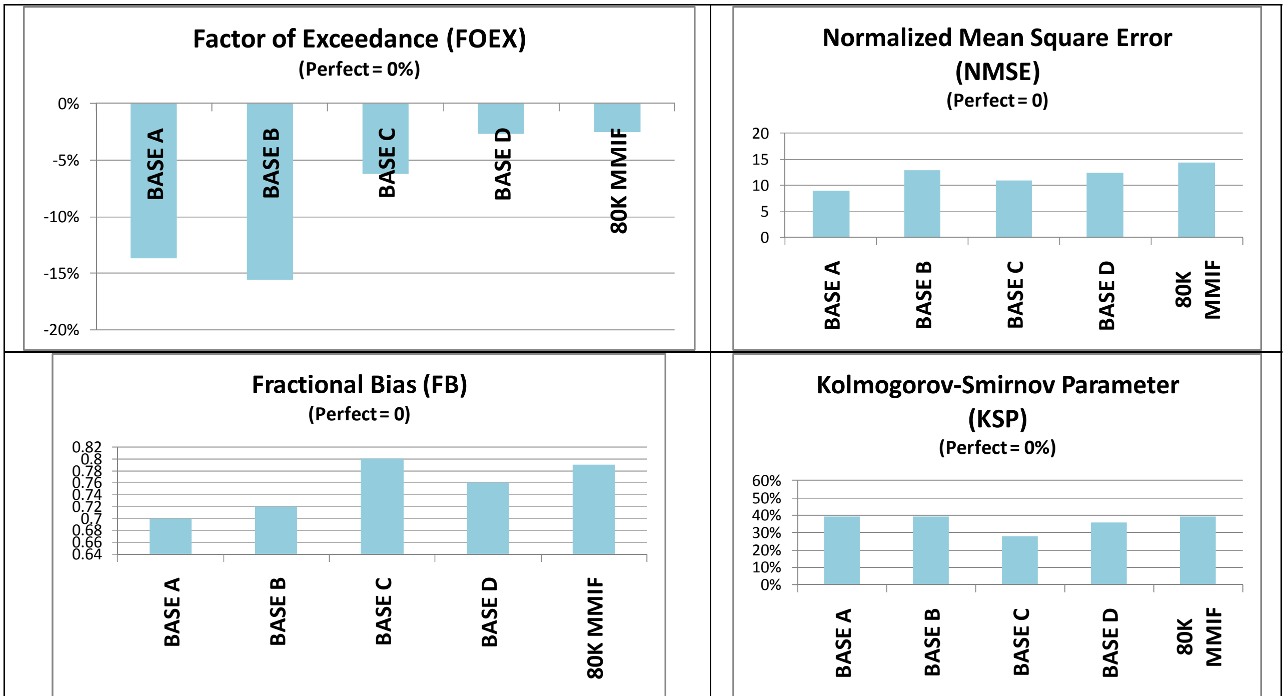


Figure 5-11a. Global model performance statistics for the CTEX5 CALPUFF sensitivity tests using the 80 km MM5data (lower values indicate better performance).

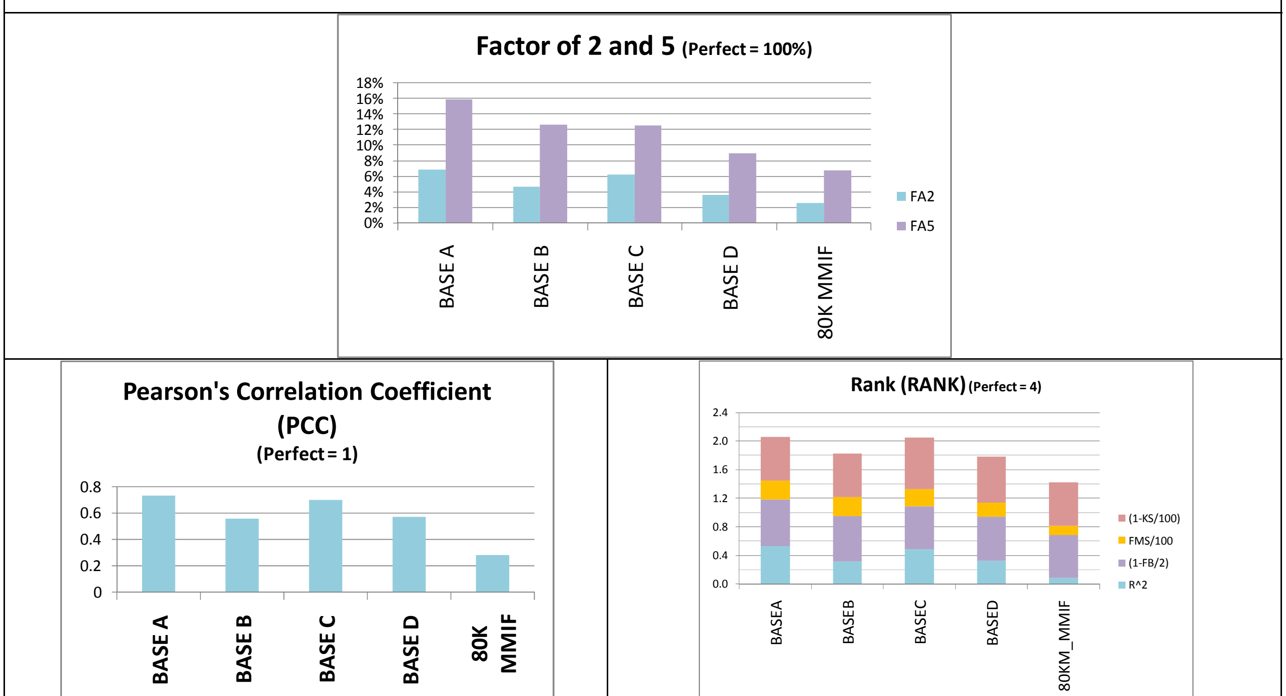


Figure 5-11b. Global model performance statistics for the CTEX5 CALPUFF sensitivity tests using the 80 km MM5 data (higher values indicate better performance).

5.4.2.2 CALPUFF CTEX5 Model Evaluation using 36 km MM5 Data

Figure 5-12 displays the spatial statistical metrics for the CTEX5 CALPUFF sensitivity tests using the 36 km MM5 data. Note that the CALMET simulation for EXP3D encountered an error in HTOLD so no CALPUFF sensitivity modeling results are available. The “A” and “B” series of

CALPUFF sensitivity simulations are performing best for the spatial performance statistics with the 36KM_MMIF performing worst.

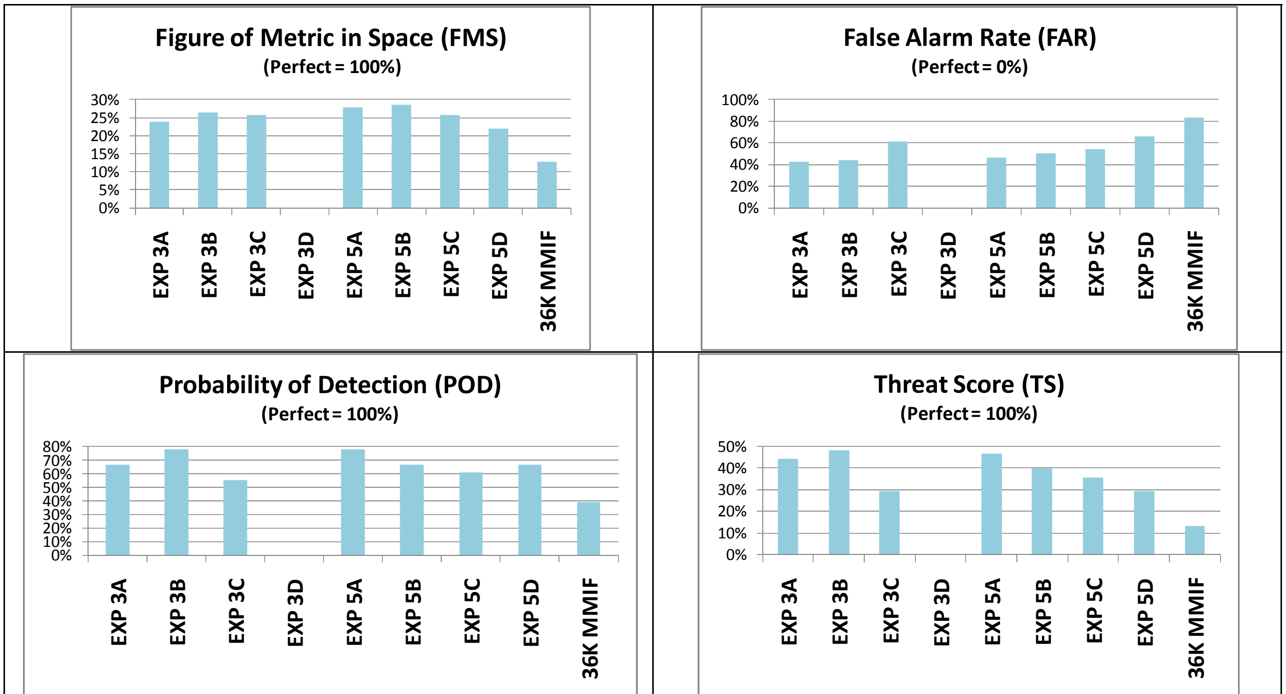


Figure 5-12. Spatial model performance statistics for the CTEX5 CALPUF sensitivity tests that used the 36 km MM5 data.

The global statistics for the CALPUFF sensitivity tests using the 36 km MM5 data are shown in Figure 5-13. EXP5D and 36KM_MMIF have the FOEX that is closest to zero. The EXP5B and EXP5D sensitivity simulations have the lowest bias and error followed by EXP3C with 36KM-MMIF having the worst bias and error metrics. The lowest (best) KSP statistics is given by EXP5D followed by 36KM_MMIF and EXP3C. EXP3B and EXP5A have the best FA2 and FA5 values, with 36KM_MMIF having the worst ones. EXP3A, EXP3C and EXP5A all have correlation coefficients above 0.7 with 36km_MMIF having the lowest correlation coefficient that is below 0.3. Using the overall composite RANK statistics, EXP3C and EXP5D (2.1) are ranked first followed by EXP3A and EXP5A (2.0) with 36KM_MMIF (1.4) having the lowest RANK statistic.

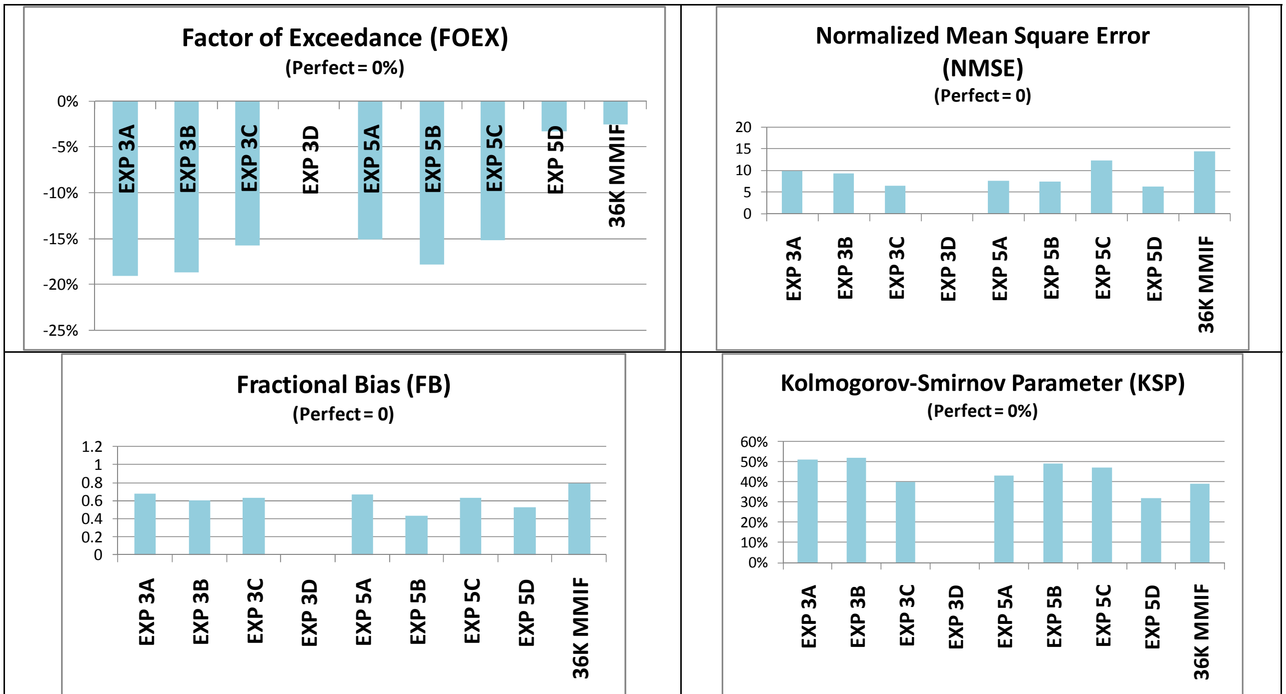


Figure 5-13a. Global model performance statistics for the CTEX5 CALPUFF sensitivity tests using the 36 km MM5 data.

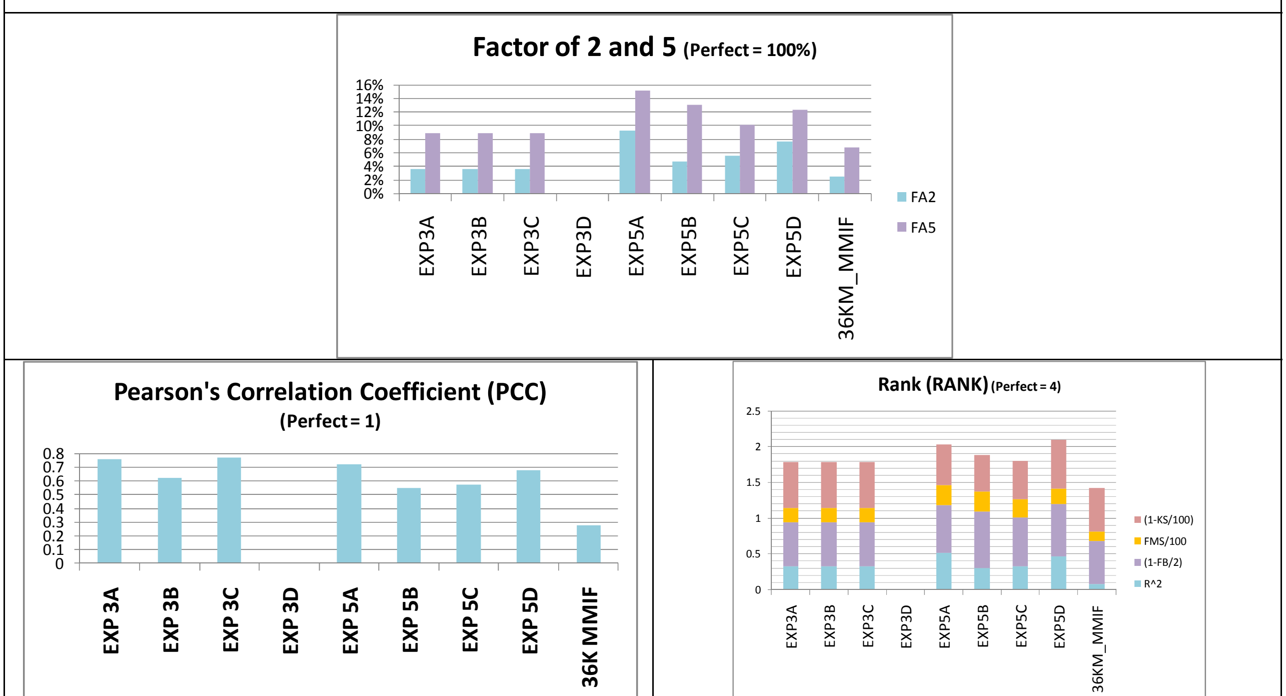


Figure 5-13b. Global model performance statistics for the CTEX5 CALPUFF sensitivity tests using the 36 km MM5 data (higher values indicate better performance).

5.4.2.3 CALPUFF CTEX5 Model Evaluation using 12 and 4 km MM5 Data

The spatial statistics for the CALPUFF/CALMET and CALPUFF/MMIF sensitivity tests using the 12 km MM5 data along with the 4 km MM5 CALPUFF/MMIF sensitivity test are given in Figure 5-

14. Across all the spatial statistics, EXP6A performs the best with EXP4A, EXP4B, EXP6B and 4KM_MMIF next best and 12KM_MMIF being worst.

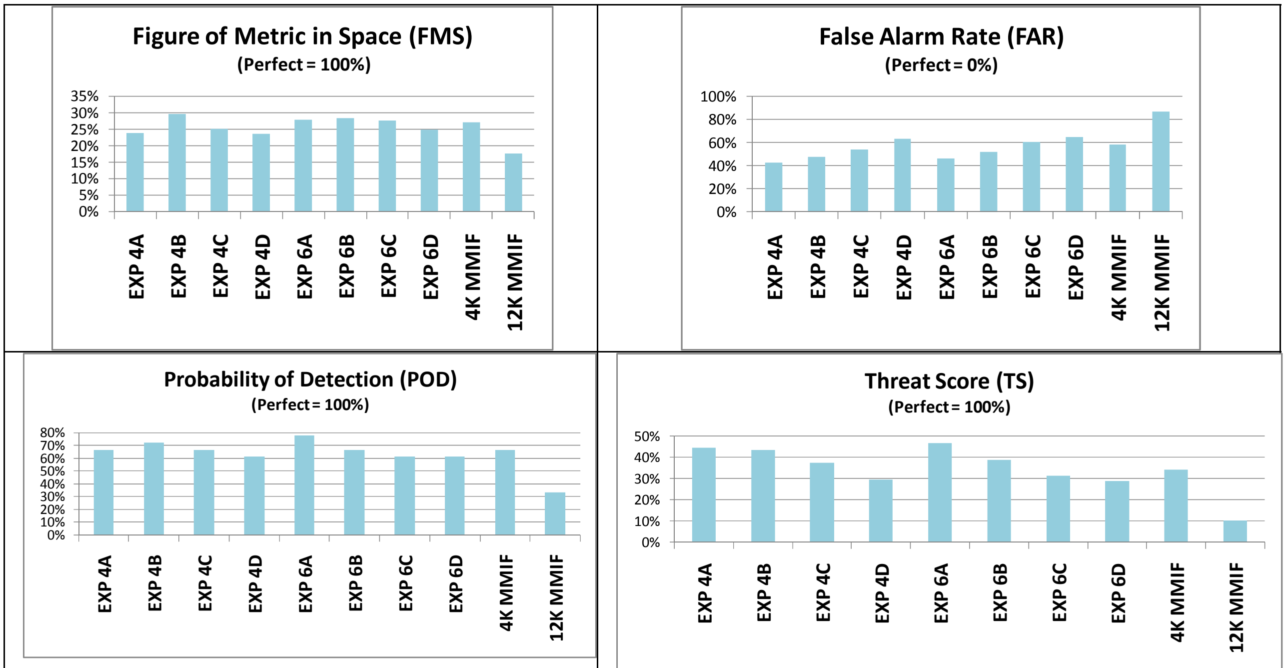


Figure 5-14. Spatial model performance statistics for the CTEX5 CALPUF sensitivity tests that used the 12 and 4 km MM5 data.

The lowest error of the 12 km MM5 CALPUFF sensitivity tests is given by EXP4B and EXP6A-C, with 12KM_MMIF having the highest error (Figure 5-15). EXP6B has the lowest bias follows by EXP6C and EXP4B, with 12KM_MMIF having the largest bias. EXP6A and EXP6B have the most model predictions within a factor of 2 of the observations and EXP4C has the most within a factor of 5. The CALMET/CALPUFF correlation coefficients range from 0.57 to 0.76 with EXP4A (0.76) and EXP6C (0.75) have the highest values and EXP6B (0.57) having the lowest value. The 4KM_MMIF has an even lower correlation coefficient (0.48) with the 12KM_MMIF having no to slight anti-correlation with the observed values (-0.07). According to the RANK composite statistics the best performing 12 km CALPUFF sensitivity test in EXP6C (2.19) followed by EXP6A (2.02) and EXP4A (1.98) with 12KM_MMIF (1.28) performing worst.

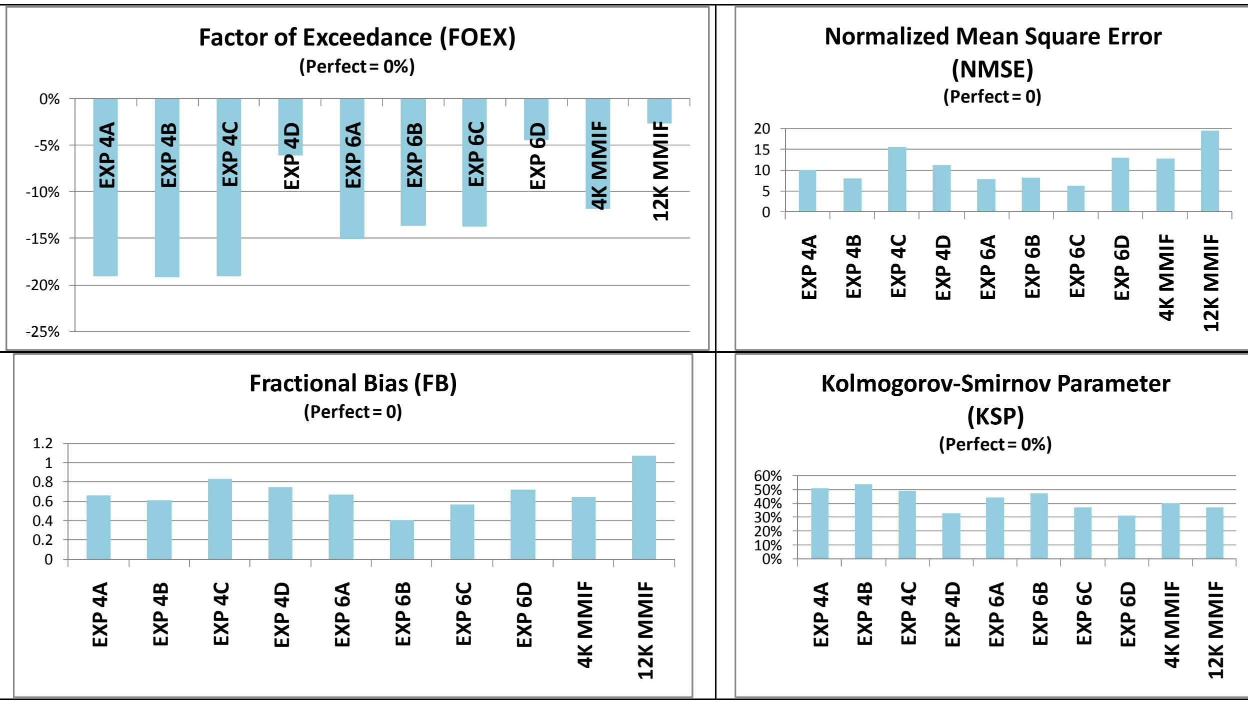


Figure 5-15a. Global model performance statistics for the CTEX5 CALPUFF sensitivity tests using the 12 km MM5 data.

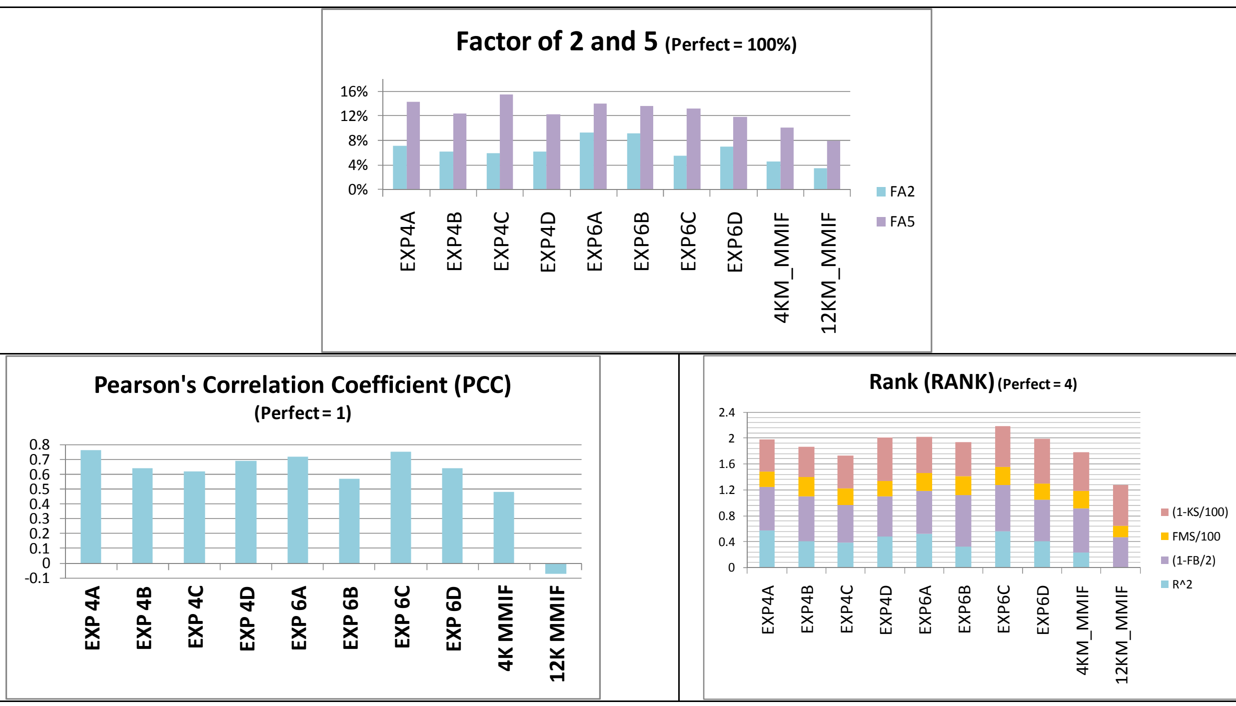


Figure 5-15b. Global model performance statistics for the CTEX5 CALPUFF sensitivity tests using the 12 km MM5 data (higher values indicate better performance).

<u>5.4.2.4 Comparison of CALPUFF CTEX5 Model Evaluation using Different MM5 Grid Resolutions</u>

This section analyzes the CALPUFF model performance across different MM5 and CALMET grid resolutions using the "B" and "D" series of CALPUFF/CALMET and the CALPUFF/MMIF sensitivity tests. The "B" series (EXP4B, EXP5B, EXP6B and EXP3B) and 4KM_MMIF have the highest FMS values between 25% and 30% with 36KM_MMIF and 80KM_MMIF have the lowest FMS scores between 10% and 15%. Again the "B" series of CALPUFF/MMIF sensitivity tests have the best FAR scores with the worst scores given by the 12KM, 36KM and 80KM MMIF sensitivity tests. The "D" series using no met observations has the worst (highest) FAR scores of the CALPUFF/CALMET sensitivity tests. EXP3B has the best POD value follows by EXP4B with EXP5B, EXP5D, EXP6B and 4KM_MMIF ties for third best; the 12KM, 36KM and 80KM MMIF CALPUFFs have the worst FAR scores. Similar results are seen with the TS statistics with the four top performing sensitivity tests ordered by EXP4B, EXP5B, EXP6B and 4KM_MMIF,

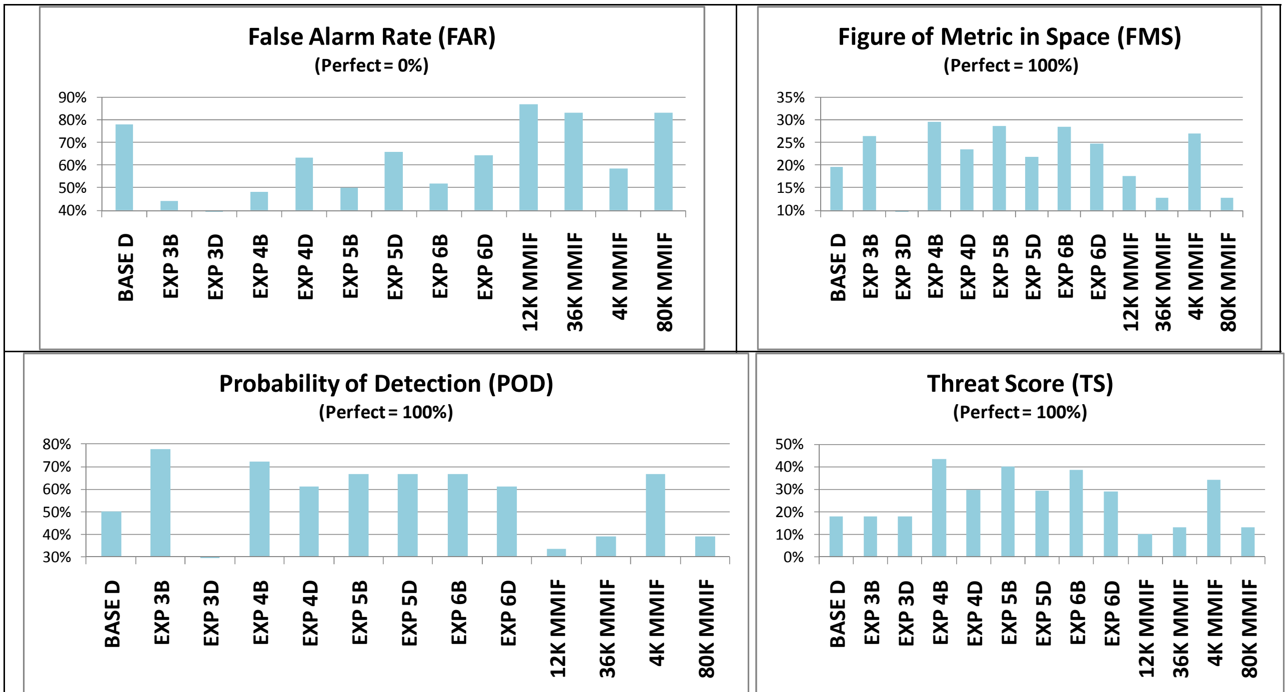


Figure 5-16. Spatial model performance statistics for the CTEX5 CALPUF sensitivity tests using different MM4/MM5 grid resolutions.

Although EXP5D has the lowest error, the "B" series of sensitivity tests consistency have the lowest bias and error (Figure 5-17). The 12KM_MMIF has the highest bias and error followed by the 80KM_MMIF sensitivity test. The "D" series of sensitivity tests have the best (lowest) KS parameter at ~30% with the "B" series of tests have the worst (highest) KSP at ~50% with the other sensitivity test in between.

The best sensitivity test for predicting the observed tracer within a factor of 2 and 5 is EXP3B followed by EXP6B with the 12KM, 36KM and 80KM MMIF runs being the worst. The correlation coefficients for the CALPUFF/CALMET CTEX sensitivity tests in this group range from 0.57 to 0.69 with 4km_MMIF being the best performing MMIF configuration with a 0.48 PCC with the other MMIF runs being much worse. The composite RANK statistics scores the CALPUFF/CALMET and 4KM_MMIF sensitivity tests in the 1.8 to 2.1 range with EXP5D (2.1) scoring the highest followed by EXP4D (1.99), EXP6D (1.99), EXP6B (1.94), EXP5B (1.89) and EXP4B (1.86). The 12KM, 36KM and 80KM CALPUFF/MMIF sensitivity tests have the lowest RANK scores (1.28 to 1.42).

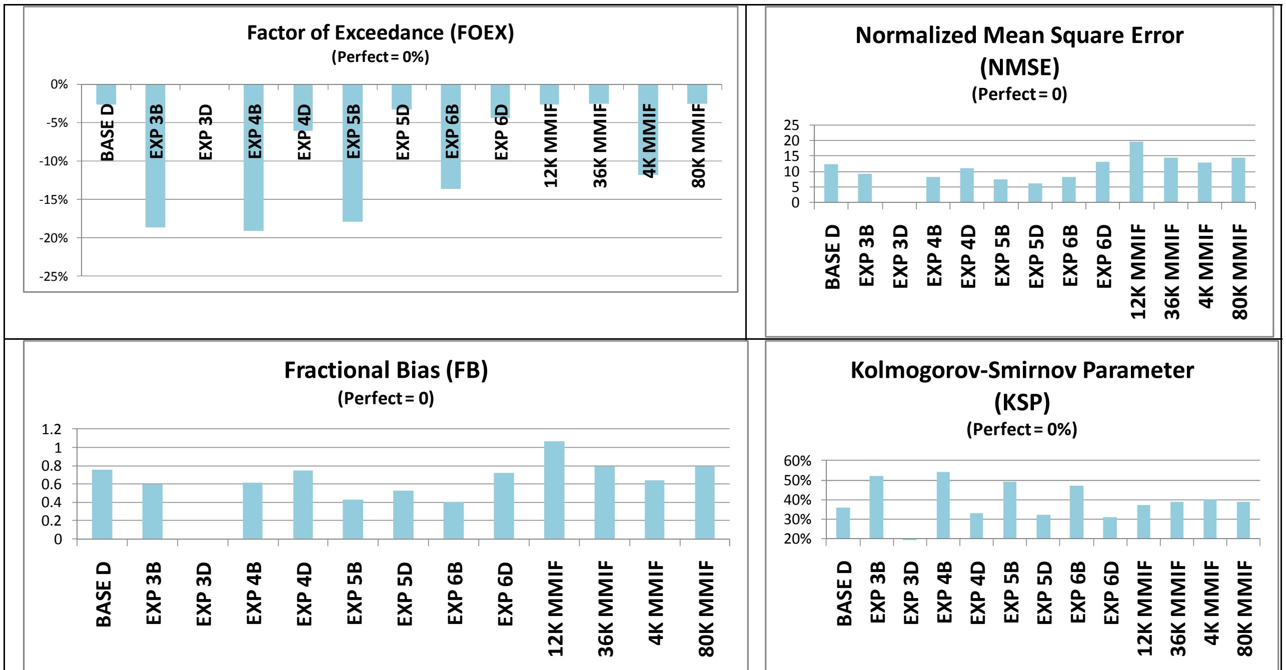


Figure 5-17a. Global model performance statistics for the CTEX5 CALPUFF sensitivity tests using different MM4/MM5 grid resolutions.

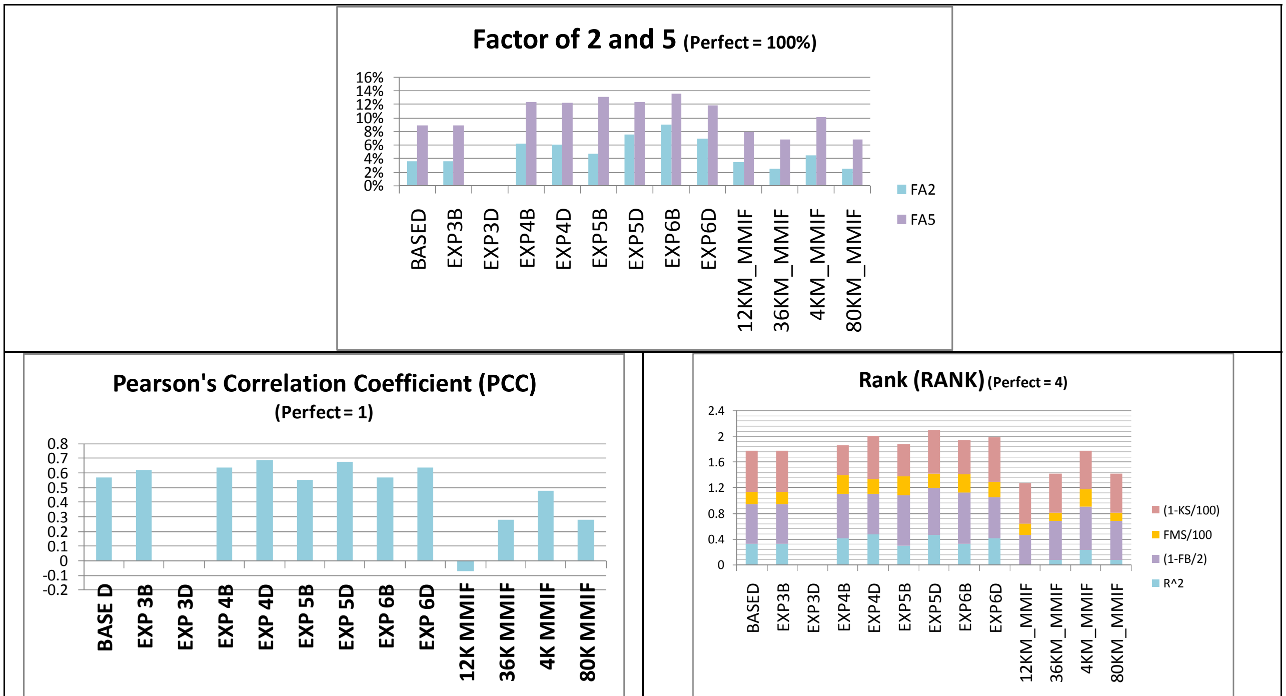


Figure 5-17b. Global model performance statistics for the CTEX5 CALPUFF sensitivity tests using different MM4/MM5 grid resolutions (higher values indicate better performance).

5.4.2.5 Rankings of CTEX5 CALPUFF Sensitivity Tests using the RANK Statistic

Table 5-10 ranks the model performance of the CTEX5 CALPUFF sensitivity tests using the RANK composite statistic. Outside of the 12KM, 36KM and 80KM MMIF CALPUFF sensitivity tests being by far the worst performing configurations with RANK values in the 1.28 to 1.42 range, the remaining sensitivity tests have RANK values in the 1.7 to 2.2 range, with the 4KM_MMIF run being in the lower end of this range. Examining trends in the CALPUFF sensitivity tests, the EXP6 series that uses the highest MM5 (12 km) and CALMET (4 km) grid resolution tends to have better model performance, whereas the "B" series of sensitivity tests tends to have worst model performance. Although the BASEA scenario is ranked 4th, the other BASE series using the 80 km MM5 and 18 km CALMET grid resolution have RANK scores on the lower end of the distribution. Based on these results we conclude the following for the CTEX5 sensitivity tests:

- Use of higher MM5 grid resolution (12 km) produces better CALPUFF model performance using both CALMET and MMIF.

5.5 CONCLUSIONS OF THE CAPTEX TRACER SENSITIVITY TESTS

There are some differences and similarities in CALPUFF's ability to simulate the observed tracer concentrations in the CTEX3 and CTEX5 field experiments. The overall conclusions of the evaluation of the CALPUFF model using the CAPTEX tracer test field experiment data can be summarized as follows:

- Regarding use of CALMET versus MMIF as a meteorological driver for CALPUFF, no definitive conclusion can be made since the CALPUFF/MMIF was the best performing model configuration for CTEX3 and the worst performing configuration for CTEX5.

Table 5-10. Final Rankings of CALPUFF CTEX5 Sensitivity Tests using the RANK statistic.

Ranking	Sensitivity Test	RANK Statistics	MM5 (km)	CALGRID (km)	RMAX1/RMAX2	Met Obs
1	EXP6C	2.19	12	4	10/100	Yes
2	EXP5D	2.10	36	4	--	No
3	BASEA	2.06	80	18	500/1000	Yes
4	BASEC	2.05	80	18	10/100	Yes
5	EXP5A	2.03	36	4	500/1000	Yes
6	EXP6A	2.02	12	4	500/1000	Yes
7	EXP4D	2.00	12	12	--	No
8	EXP6D	1.99	12	4	--	No
9	EXP4A	1.98	12	12	500/1000	Yes
10	EXP6B	1.94	12	4	100/200	Yes
11	EXP5B	1.89	36	4	100/200	Yes
12	EXP4B	1.86	12	12	100/200	Yes
13	BASEB	1.82	80	18	100/200	Yes
14	EXP5C	1.80	36	4	10/100	Yes
15	BASED	1.79	80	18	--	No
16	EXP3A	1.79	36	12	10/100	Yes
17	EXP3B	1.79	36	12	100/200	Yes
18	EXP3C	1.79	36	12	500/1000	Yes
19	EXP3D	1.79	36	12	--	No
20	4KM_MMIF	1.78	4	--	--	No
21	EXP4C	1.72	12	12	10/100	Yes
22	36KM_MMIF	1.42	36	--	--	No
23	80KM_MMIF	1.42	80	--	--	No
24	12KM_MMIF	1.28	12	--	--	No

- The use of 12 to 36 km resolution MM5 data tends to produce better CALPUFF model performance than using coarse grid data (e.g., 80 km).
- Regarding the effects of the RMAX1/RMAX2 parameters on CALPUFF/CALMET model performance, the “A” series (500/1000) is performing best for CTEX3 but the “C” series (10/100) is performing best for CTEX5 with both CTEX3 and CTEX5 agreeing that the “B” series (100/200) is the worst performing setting for RMAX1/RMAX2.
    - This is in contrast to the CALMET surface wind model evaluation that found the EPA-FLM Clarification Memorandum recommended settings used in the “B” series of CALMET experiments produced the wind fields that most closely matched observations (see Appendices A and B).
    - However, the CALMET surface wind evaluation was not a valid independent evaluation since surface wind observations are also used as input to CALMET for some of the experiments.

6.0 1994 EUROPEAN TRACER EXPERIMENT

6.1 DESCRIPTION OF THE 1994 EUROPEAN TRACER EXPERIMENT

The European Tracer Experiment (ETEX) was initiated in 1992 by the European Commission (EC), International Atomic Energy Agency (IAEA), and World Meteorological Organization (WMO) to address many of the questions that arose from the 1986 Chernobyl accident regarding the capabilities of LRT models and the ability to properly handle and disseminate large volumes of data. ETEX was designed to validate long-range transport models used for emergency response situations and to develop a database which could be used for model evaluation and development purposes.

6.1.1 ETEX Field Study

Two releases of a perflurocarbon tracer called perfluromonomethylcyclohexane (PMCH) were made in October and November 1994 from France. For this evaluation, model simulations are focused upon the first PMCH release. The first ETEX release has been used extensively to evaluate operational LRT models for numerous countries so was also used in this study. In many ways, it represents an ideal database for LRT evaluation because of the volume and high frequency of observations taken.

The PMCH was released at a constant rate of approximately 8 g/s (340 kg total) for 12 hours beginning at 1600 UTC on 23 October 1994 from Monterfil, France. The release of PMCH was a dynamic release, with an outlet temperature of 84°C and velocity of 47.6 m/s (JRC, 2008). Air concentrations were sampled at 168 monitoring sites in 17 European countries with a sampling frequency of every three hours for approximately 90 hours. Figure 6-1 displays the location of the PMCH release point in northwestern France and the array of sampling receptors.

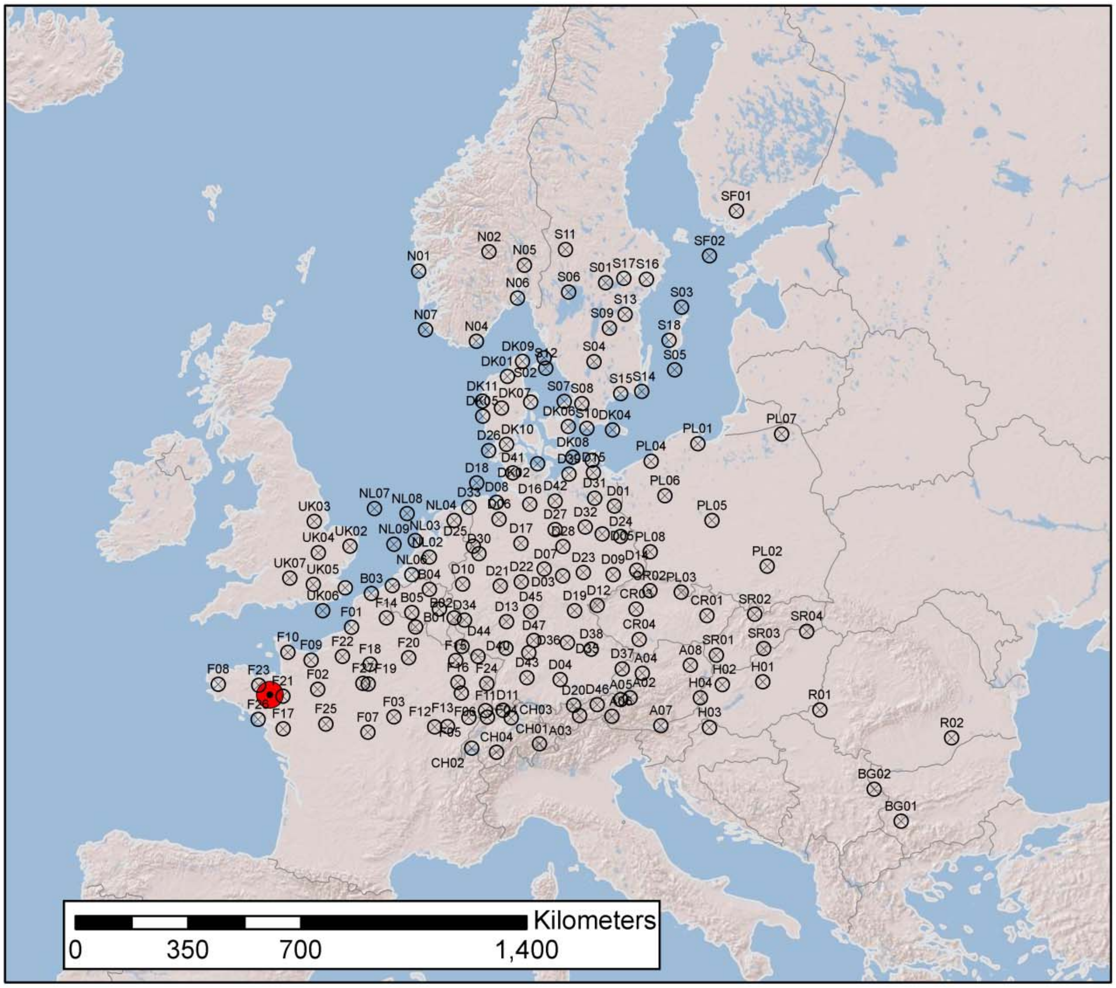


Figure 6-1. Locations of the PMCH tracer release point in Monterfil , France and sampling receptors for the 1994 European Tracer Experiment (ETEX).

6.1.2 Synoptic Conditions

Numerous synoptic surface and upper air observations were made by participating meteorological agencies as part of this experiment. Two separate extratropical cyclonic systems were present over the European continent at the time of the release. A strong extratropical cyclone was located over the North Sea with a central pressure of 980 mb. A second, significantly weaker, extratropical cyclone was located near the Balkan Peninsula over the Black Sea, having a central pressure of 1010 mb. These cyclonic systems were important to the transport of the PMCH tracer cloud. Figure 6-2 depicts the locations of the extratropical cyclonic systems during the ETEX field study.

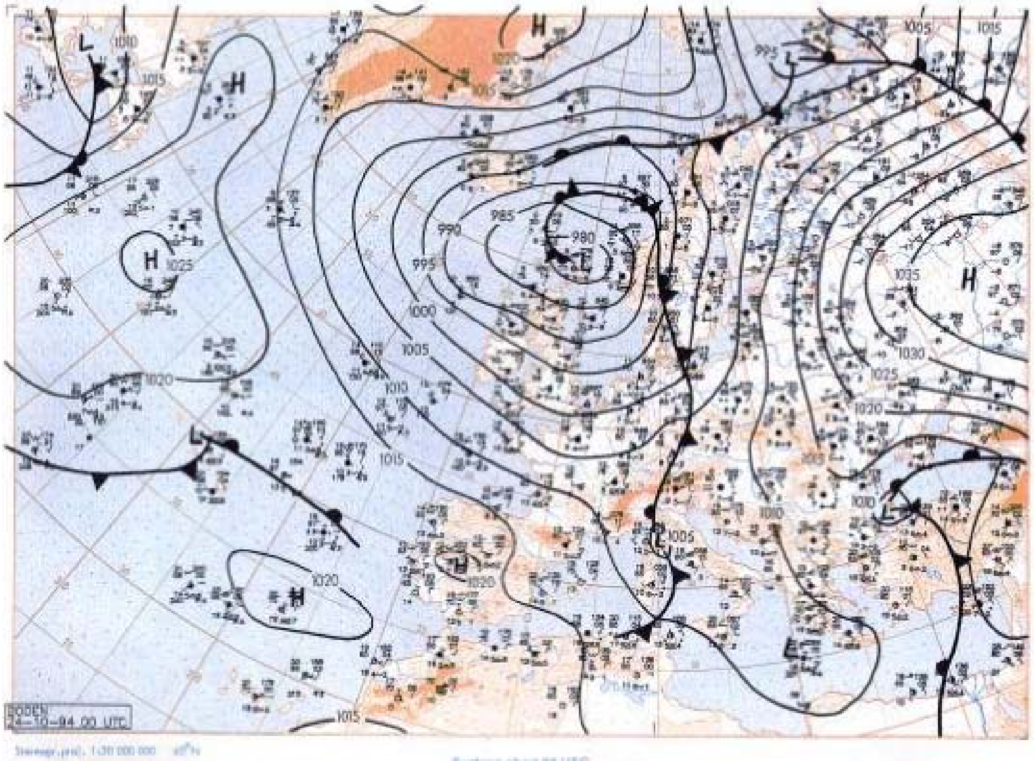


Figure 6-2a. Surface synoptic meteorological conditions for Europe at 0000 UTC on October 24, 1994 eight hours after the release of the PMCH tracer in ETEX (Source: http://rem.jrc.ec.europa.eu/etex/).

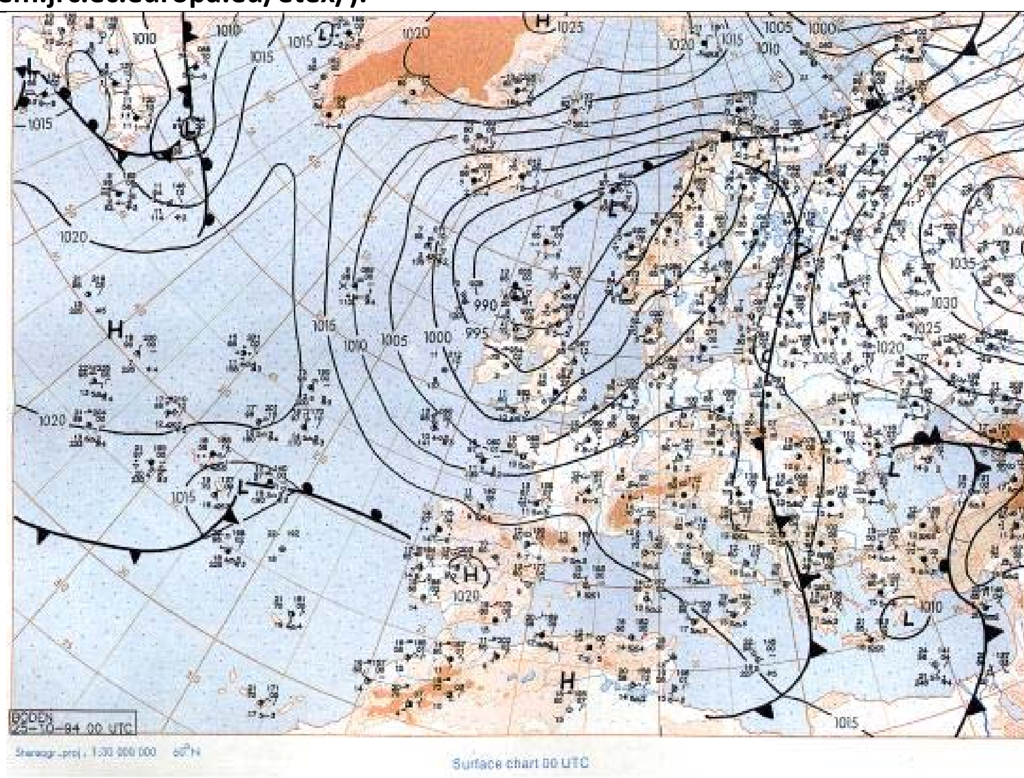


Figure 6-2b. Surface synoptic meteorological conditions for Europe at 0000 UTC on October 25, 1994 32 hours after the release of the PMCH tracer in ETEX (Source: http://rem.jrc.ec.europa.eu/etex/).

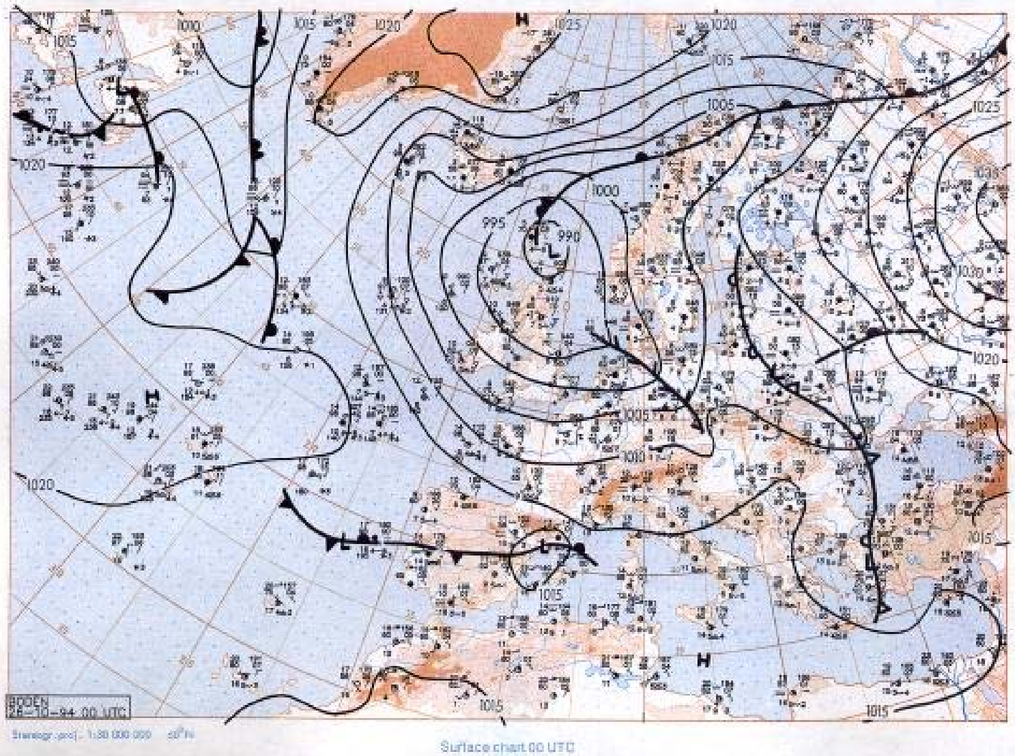


Figure 6-2c. Surface synoptic meteorological conditions for Europe at 0000 UTC on October 26, 1994 56 hours after the release of the PMCH tracer in ETEX (Source: http://rem.jrc.ec.europa.eu/etex/).

Figure 6-3 displays the spatial distribution of the observed PMCH tracer concentration in pico-grams per cubic meter (pgm^{-3}) 24, 36, 48 and 60 hours after the release in Monterfil, France. During the first 24 hours after the release of the PMCH, the tracer cloud was advected generally east-northeast from the release point in northwestern France into the Netherlands and Luxembourg and into western Germany. By 36 hours after the initial release, the tracer cloud had advected well into Germany (Figure 6-3, top right). In this region, the wind flow split between the two cyclonic systems northwest and southeast of Germany (see Figure 6-2), causing the tracer cloud to essentially bifurcate, with one portion advecting around the core of the cyclonic system over the North Sea, and the other portion advecting southeast towards the cyclonic system in the Balkan Peninsula region. 48 and 60 hours after the tracer release (Figure 6-3, lower panels), the tracer cloud stretches from Norway to the Black Sea in a narrow northwest to southeast orientation.

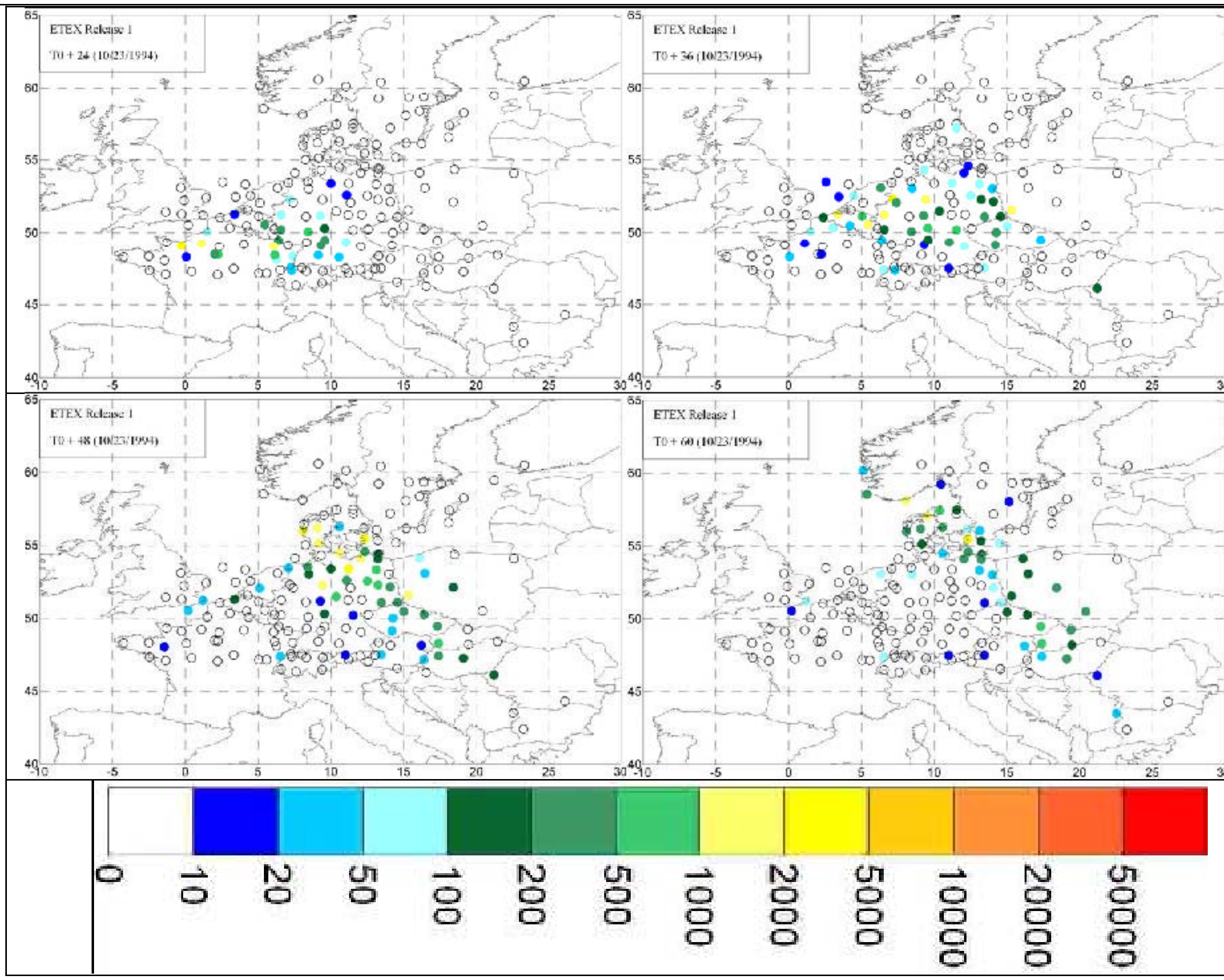


Figure 6-3a. Distribution of the observed PMCH tracer concentrations (pgm^{-3}) 24 (top right), 36 (top right), 48 (bottom left) and 60 (bottom right) hours after the release.

6.2 MODEL CONFIGURATION AND APPLICATION

6.2.1 Experimental Design

The objectives of the LRT model evaluation using the ETEX field study database was somewhat different than the other three tracer test evaluations. In the GP80, SRL75 and CAPTEX tracer test LRT model evaluations, one major objective was an evaluation of the CALPUFF LRT dispersion model using two different sets of meteorological inputs, one based on the CALMET diagnostic wind model and the other using the MMIF WRF/MM5 pass-through tool. However, in the ETEX LRT model tracer test evaluation an objective was to use the same meteorological inputs in all of the LRT dispersion models. This approach is similar to the one taken by Chang and co-workers (2003) who conducted an evaluation of three Lagrangian puff models (HPAC/SCIPUFF, VLSTRACK, and CALMET/CALPUFF). While all three puff models are based on a Gaussian puff formulation, these models varied significantly in terms of the level of sophistication of their technical formulation. Chang and co-workers (2003) proposed a framework to perform an objective and meaningful evaluation when such models vary significantly in their formulation. A primary focus of their model evaluation framework centered upon the use of the same observed meteorological data and similar modeling domains. To the extent practical, default model options were selected for all models in their evaluation. Reflecting that evaluation paradigm, a major focus of the LRT model evaluation using the ETEX database in this study was to provide a common source of meteorological fields to each of the dispersion models evaluated.

Five different LRT dispersion models were evaluated using the ETEX database. Each of the LRT models in this exercise requires three-dimensional meteorological fields as input to the model. For the majority of these models, meteorological fields from prognostic meteorological models are the primary source of the meteorological inputs. However, CALPUFF and SCIPUFF typically rely upon their own diagnostic meteorological models to provide three-dimensional meteorological fields to the dispersion model. In cases where prognostic meteorological model data are ingested to set the initial conditions within the diagnostic meteorological model, much of the original prognostic meteorological data is not preserved, and key parameters are rediagnosed. This compromises a key component of the evaluation paradigm of Chang et al. (2003) that we have adopted for the ETEX evaluation, namely a common meteorological database. The Mesoscale Model Interface (MMIF) software program (Emery and Brashers, 2009) was developed to facilitate direct ingestion of prognostic meteorological model data by the LRT dispersion model, bypassing the diagnostic meteorological model component and rediagnosing algorithms effectively overcoming the challenge to this evaluation paradigm.

6.2.2 Meteorological Inputs

During the original ATMES-II project, participating agencies during ETEX were required to calculate concentration fields for their respective models using analysis fields from the European Center for Medium-Range Weather Forecasts (ECMWF). ECMWF analysis fields were available at 6-hour intervals and a horizontal resolution of 0.5° (~50 km) latitude-longitude (D'Amours, 1998). Participating agencies could also submit results obtained using different meteorological analyses. Van Dop et al. (1998) and Nasstrom et al. (1998) found that increasing the resolution of the input meteorological fields enhanced the performance of the dispersion models evaluated in the ATMES-II study. Similarly, Deng et al. (2004) found that SCIPUFF model performance for the Cross-Appalachian Tracer Experiment (CAPTEX) improved by increasing meteorological model horizontal and vertical resolution, use of four dimensional data assimilation (FDDA), and more advanced meteorological model physics. However, they also noted that use of the more advanced physics options were responsible for more improvement in model performance than merely increasing horizontal grid resolution.

For the LRT model evaluation exercise using the ETEX database presented in this report, meteorological inputs were generated using a limited-area mesoscale meteorological model to produce higher temporally and spatially resolved meteorological data than used in the ATMES-II project. By producing more accurate meteorological fields, it should be possible to maximize performance of the LRT models under evaluation in this study. Furthermore, by using a common source of meteorological data between each of the five modeling systems, it reduces the potential contribution of differences in meteorological data on dispersion model performance and facilitates a more direct intercomparison of dispersion model results.

Hourly meteorological fields were derived from the PSU/NCAR Mesoscale Meteorological Model (MM5) Version 3.74 (Grell et al., 1995). MM5 was initialized with National Center for Environmental Prediction (NCEP) reanalysis data (NCAR, 2008). NCEP reanalysis fields are available every 6 hours on a 2.5° x 2.5° (~275 km) grid. The MM5 horizontal grid resolution was

36 kilometers and the vertical structure contained 43 vertical layers. Physics options were not optimized for northern European operations, but were based upon more advanced physics options available in MM5, reflecting the findings of Deng et al. (2004). Key MM5 options included:

- ETA Planetary Boundary Layer (PBL) scheme;
- Kain-Fritsch II cumulus parameterization (Kain, 2004);
- Rapid Radiative Transfer Model (RRTM) radiation scheme (Mlawer et al. 1997);
- NOAH land surface model (LSM) (Chen et al. 2001); and
- Dudhia Simple Ice microphysics scheme (Dudhia, 1989).

Four dimensional data assimilation (FDDA) (Stauffer et al. 1990, 1991) was employed for this study. “Analysis nudging” based upon the NCEP reanalysis fields were used with default values for nudging strengths.

6.2.3 LRT Model Configuration and Inputs

Three distinct classes of LRT dispersion models were included as part of the ETEX tracer evaluation including four Lagrangian models and one Eulerian model. CALPUFF Version 5.8 (Scire et al. 2000b) and SCIPUFF Version 2.303 (Sykes et al., 1998) are Lagrangian Gaussian puff models. HYSPLIT Version 4.8 (Draxler 1997) and FLEXPART Version 6.2 (Siebert 2006) are Lagrangian particle models. CAMx Version 5.2 (ENVIRON, 2010) is an Eulerian grid model. The respective user’s guides provide a complete description of the technical formulations of each of these models.

Both CALPUFF and SCIPUFF are based upon Gaussian puff formulation. The two puff models have the advantage of more robust capabilities for source characterization, having the ability to treat dispersion for point, area, or line sources. Furthermore, these models can more accurately characterize dynamic releases of pollutants by accounting for initial plume rise of the pollutant. Conversely, the two particle models are very limited in their capability to characterize sources, having no direct ability to account for variations in source configurations or consider plume rise. The CAMx grid model is limited in its ability to simulate “plumes” by the grid resolution specified. CAMx includes a subgrid-scale Plume-in-Grid (PiG) module to treat the early evolution, transport and dispersion of point source plumes whose effect on model performance was investigated using sensitivity tests.

Since plume rise varies from hour-to-hour as a function of ambient temperature, wind speed and stability it is not possible to define a release height which would reflect this variation. Therefore, a constant release height of 10 meters was assigned for the two particle models in this study. This limitation of the particle models is problematic when comparing against models such as CALPUFF, SCIPUFF and CAMx that can simulate dynamic releases of emissions and calculate hour-specific plume rise using hourly meteorological data. Iwasaki et al. (1998) found that the initial release height assigned to the Japan Meteorological Agency (JMA) particle model had a large impact on the predicted ground level concentrations. Investigation of initial release height sensitivity of the two particle models was beyond the scope of this evaluation. However, this limitation should be noted when considering the uncertainty of concentration estimates from the two particle models.

Each of the four models requires gridded meteorological fields for dispersion calculations. CALPUFF normally uses output from the CALMET diagnostic wind field model (Scire et al., 2000a). SCIPUFF also has its own simplified mass-consistent wind field processor referred to as MC-SCIPUFF (Sykes et al., 1998). Gridded meteorological fields are normally supplied to HYSPLIT and FLEXPART using software that converts prognostic meteorological data into formats that are directly ingested into the respective dispersion models. The CAMx model also uses software to reformat output from a prognostic meteorological model into the variables and formats used by CAMx.

Use of a diagnostic wind field model (DWM) as the primary method to supply meteorological data to the dispersion models under review creates additional uncertainty in the intercomparison of the five dispersion models. DWM's, such as CALMET, have the ability to ingest prognostic data from models such as the PSU/NCAR MM5 (Grell et al.,1995) or the Advanced Research Weather Research and Forecasting (WRF-ARW) (Skamarock et al. 2008) as its first guess wind field. However, this method of using the prognostic meteorological data as the first guess field for the DWM does not preserve the integrity of the original meteorological field. For example, the CALMET DWM adjusts the wind fields for kinematic and thermodynamic effects of terrain and also rediagnoses key meteorological parameters such as planetary boundary layer heights. Thus, to conduct a proper evaluation of the dispersion models on the same basis, each of the models should be operated with the same meteorological dataset. In order to maintain consistency with this study objective, it would not have been appropriate to use either MC-SCIPUFF or CALMET to produce three-dimensional meteorological fields for their respective dispersion model.

In order to facilitate direct intercomparison of models using a common prognostic meteorological dataset, it is necessary to supply meteorological fields to CALPUFF and SCIPUFF in the same manner as the particle models and grid model included in this study. SCIPUFF has the ability to ingest prognostic data sets directly in either MEDOC (Multiscale Environmental Dispersion Over Complex terrain) (Sykes et al., 1998) or HPAC formats. The Pennsylvania State University developed the MM5SCIPUFF utility program (A. Deng, pers. comm.) to convert MM5 fields into the MEDOC format which is directly ingested into the SCIPUFF. Similarly, the US EPA developed the Mesoscale Model Interface (MMIF) software to convert MM5 fields into the CALPUFF meteorological input format (Emery and Brashers, 2009). With these two utility programs, it was now possible to evaluate the five LRT models using a consistent set of meteorological inputs.

Due to the inherent differences that exist between each of the five LRT models, it was not possible to standardize dispersion model options. Rather, options selected for each class of models were similar to the extent possible. For example, more advanced model features (turbulence dispersion, puff splitting) were used for CALPUFF simulations as these represent the state-of-the-practice for puff dispersion models and are most consistent with the capabilities of the SCIPUFF modeling system, helping to facilitate greater inter-model consistency for this evaluation.

CALPUFF is typically only recommended to distances of about 300 km or less (EPA, 2003). This would effectively limit the useful range of CALPUFF to the first 24-36 hours of ETEX simulation. However, recent enhancements to the CALPUFF modeling system include both horizontal and vertical puff splitting, incorporating the effects of wind shear on puff growth, potentially allowing for use of CALPUFF at distances greater than the nominal recommended limit of about 300 km, and allowing for more direct intercomparison with the two particle models and one grid model used in this study which are free of this restriction. The default method for CALPUFF vertical puff splitting is to allow for splitting to occur once per day by turning on the puff splitting flag near sunset (hour 17), artificially limiting the number of split puffs that are generated by the model. However, for the ETEX evaluation puff-splitting was enabled for each simulation hour instead of the default option of once per day in order to allow for full treatment of wind shear. The puff splitting feature of the CALPUFF modeling system does not have a complementary puff "merging" feature which aggregates puffs according to specified rules when they occupy the same space. Without the complementary puff merging capability, the number of puffs generated by puff-splitting can rapidly increase, resulting in extensive computational requirements of the model and eventual simulation termination once the maximum number of puffs allowed by the model is exceeded. Since the ETEX CALPUFF application was of short duration, the number of puffs allowed was increased so no termination occurred. However, the use of all hour puff splitting with CALPUFF in an annual simulation could be problematic. The SCIPUFF Lagrangian puff model also performs puff splitting when a sheared environment is encountered, however it can perform puff merging when two puffs occupy the "same" space so does not suffer from the extensive computer time of CALPUFF when aggressive puff splitting is desired.

The horizontal and vertical grid structures of CALPUFF were similar to the parent MM5 data. Twenty-seven (27) vertical levels were used in CALPUFF with each of the first 27 MM5 layers matched explicitly to the CALPUFF vertical structure, through the lowest 4,900 m vertical depth of the atmosphere. Additionally, 168 discrete receptors were included in the modeling analysis, with the location of each corresponding to the location and elevation of the ETEX monitors. AERMOD (EPA, 2004) turbulence coefficients, no complex terrain adjustment, and puff-splitting were selected for this analysis. A constant emission rate of 7.95 g/s was assigned for twelve hours of release of the PMCH tracer. Plume rise and momentum were also simulated in CALPUFF according to the release characteristics detailed on the ETEX website. CALPUFF results were integrated for 90 hours, and model results were post-processed in order to generate 30 three (3) hour averages for each of the 168 discrete receptors.

For SCIPUFF simulations, the horizontal and vertical grid structures of the extracted MM5 data were similar to the original MM5 data. Twenty-eight (28) vertical levels were extracted, encompassing a depth of approximately 5,000 m, similar to the CALPUFF simulations. Plume rise and momentum were also simulated in SCIPUFF in the same manner as the CALPUFF simulations. SCIPUFF results were also integrated for 90 hours, and model results were post-processed in order to generate 30 three (3) hour averages for each of the same 168 discrete receptors.

FLEXPART simulations used a 375 x 175 horizontal grid at a resolution 0.16° (~18 km) latitude/longitude. All MM5 vertical layers were extracted for the transport simulation. The FLEXPART concentration grid consisted of 15 vertical levels from the surface to 1,500 m with 9 layers below the first 500 m. Emissions were released at 10 meters. Concentrations were bi-linearly interpolated to grid cells corresponding to the 168 ETEX monitoring locations that were used.

HYSPLIT simulations used a 60 x 60 concentration grid with a horizontal resolution 0.25° (~28 km) latitude/longitude, consistent with NOAA's model configuration for ETEX described on the DATEM website. All MM5 vertical layers to 5000 meters were extracted for the transport simulation. Emissions were released at 10 meters. The gridded concentration output was linearly interpolated to the sampling locations utilizing software from NOAA's Data Archive of Tracer Experiments and Meteorology (DATEM) project. HYSPLIT was configured as a puff-particle hybrid (same used by the NOAA ARL for their ETEX evaluation) was used for the model intercomparison (i.e., INITD = 104)

Note that the FLEXPART and HYSPLIT meteorological inputs were based on the 36 km MM5 meteorological model output, so they used the same transport conditions and resolution as the other LRT models. The FLEXPART (~18 km) and HYSPLIT (~28 km) horizontal grid resolution is used to convert the particles (mass) to concentrations (mass divided by volume).

CAMx was operated on a 148 x 112 horizontal grid with 36 km grid resolution with 25 vertical layers up to a 50 mb pressure level (~15 km). CAMx is a photochemical grid model that includes state-of-science gas, aerosol and aqueous phase chemistry modules and dry and wet deposition algorithms. However, for the ETEX tracer modeling CAMx was operated with no chemistry and no wet or dry removal mechanisms. The MM5CAMx processor was used to process the MM5 output to the variables and formats required by CAMx. CAMx has several options for vertical mixing (from MM5CAMx), horizontal advection as well as a subgrid-scale Plume-in-Grid (PiG) module. Several alternative configurations of CAMx were investigated using sensitivity tests. When comparing with the other LRT models, we used a CAMx configuration with the following attributes, which are fairly typical for many CAMx simulations:

- CMAQ-like vertical diffusion coefficients from MM5CAMx;
- Piecewise Parabolic Method (PPM) horizontal advection solver; and
- No PiG module.

6.3 QUALITY ASSURANCE

Quality assurance (QA) of the LRT dispersion runs was conducted by evaluating the MM5 meteorological model output against surface meteorological observations and by examining of the LRT model inputs and outputs, as available, to assure that the intended options and configurations were used.

6.3.1 Quality Assurance of the Meteorological Inputs

A limited statistical evaluation of the MM5 simulation for the ETEX period was conducted as part of this evaluation. The meteorological observations collected at the 168 sampling stations during the ETEX exercise were not used as part of the MM5 data assimilation strategy; therefore, these observations could reliably be used to provide an independent evaluation of the MM5 simulation.

MM5 model performance evaluation results are presented in Figure 6-4. The MM5 performance statistics presented in Figure 6-4 are compared to performance criteria typically recommended for meteorological model applications for regional air quality studies in the United States (Emery et al. 2001) that were presented previously in Table 5-7. In general, MM5 verification scores indicate a persistent negative bias and higher error for both wind speed (-1.67 m/s and 4.73 m/s, respectively) and temperature (-1.1 °K and 2.36 °K, respectively) averaged across all 168 sites that are outside of target performance benchmark values for each of these meteorological parameters. Wind direction bias and error were within the performance benchmarks. Typically, these performance statistics would likely cause the modeler to consider experimenting with additional physics configurations and/or altering the data assimilation strategy to enhance meteorological model verification statistics. However, the MM5 simulation was not optimized for this project for several reasons:

- First, from an operational perspective, the meteorological model errors are likely consistent with the magnitude of model prediction errors that would have been experienced during the original ETEX exercise if forecast fields rather ECMWF analysis fields had been employed. Additionally, the MM5 simulation has the added advantage of data assimilation to constrain the growth of forecast error as a function of time.

- Second, since each of the five LRT model platforms evaluated in this project are presented with the same meteorological database; a systemic degradation of performance due to advection error would have been observed if the meteorology was a primary source of model error. However, since poor model performance was only noted in one of the five models, meteorological error was not considered the primary cause of poor performance.
- Finally, since wind direction is likely one of the key meteorological parameters for LRT simulations, the operational decision to use the existing MM5 forecasts was made because the MM5 wind direction forecasts were within acceptable statistical limits.

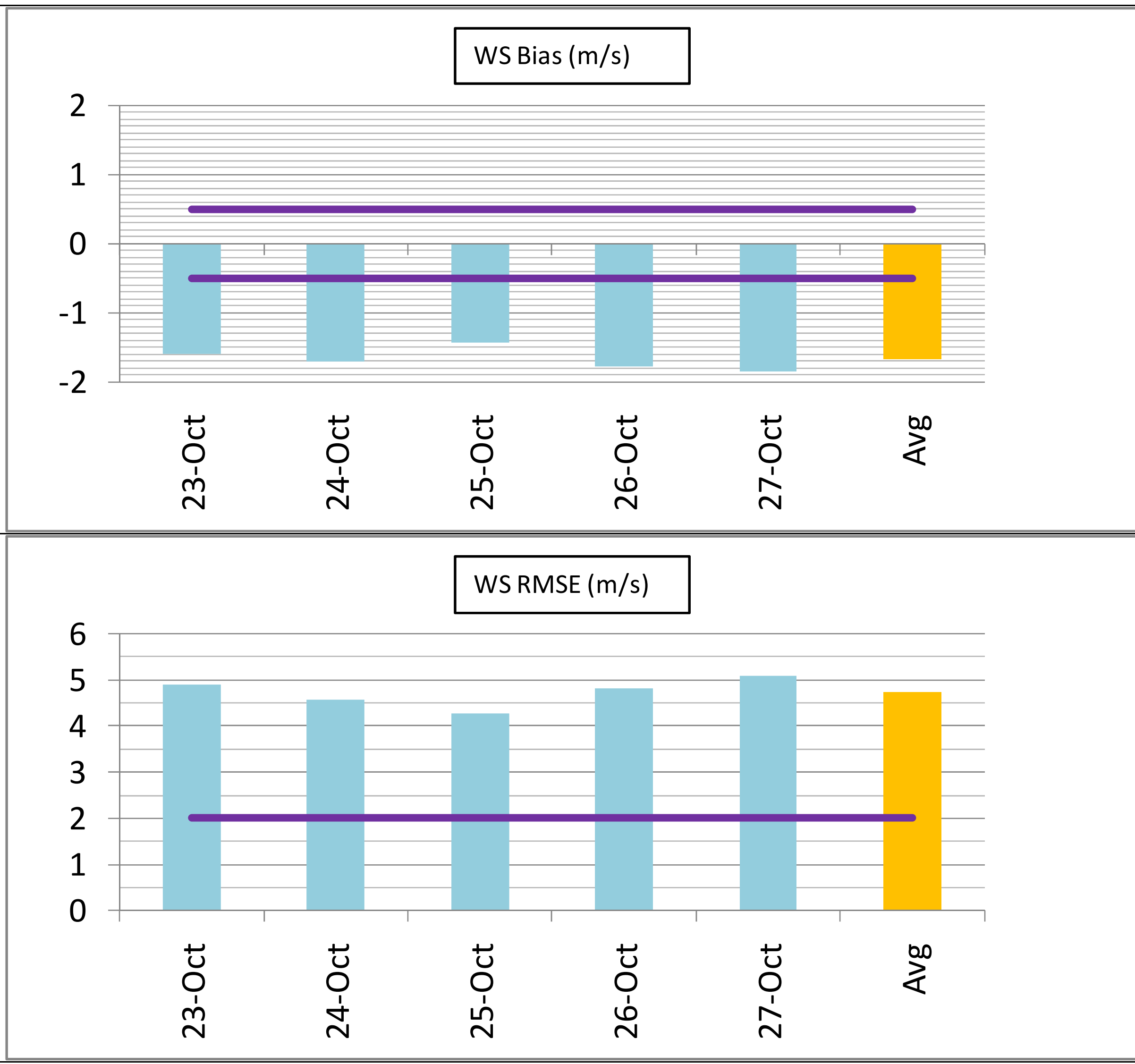


Figure 6-4a. ETEX MM5 model performance statistics of Bias (top) and RMSE (bottom) for wind speed and comparison with benchmarks (purple lines).

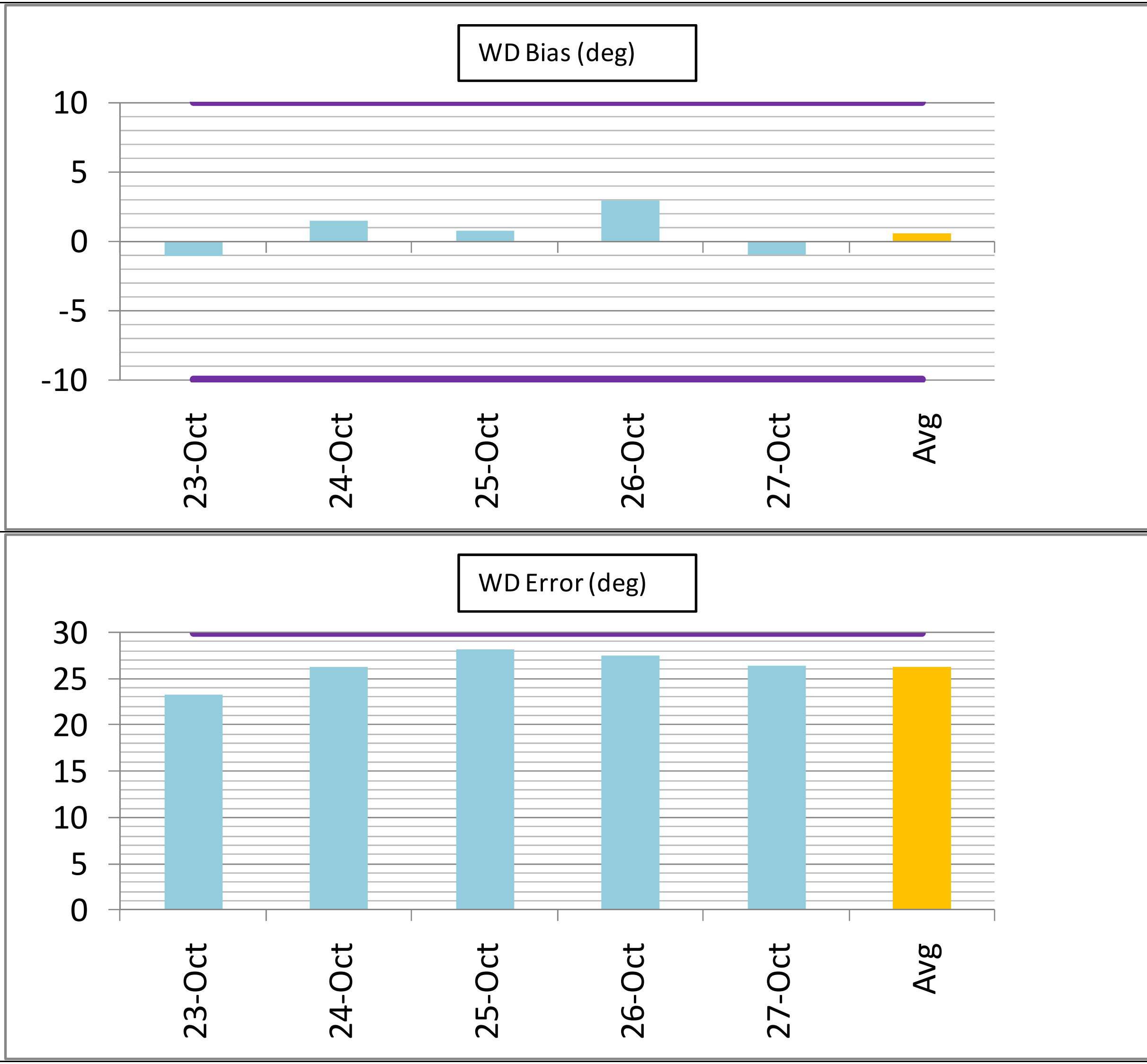


Figure 6-4b. ETEX MM5 model performance statistics of Bias (top) and Error (bottom) for wind direction and comparison with benchmarks (purple lines).

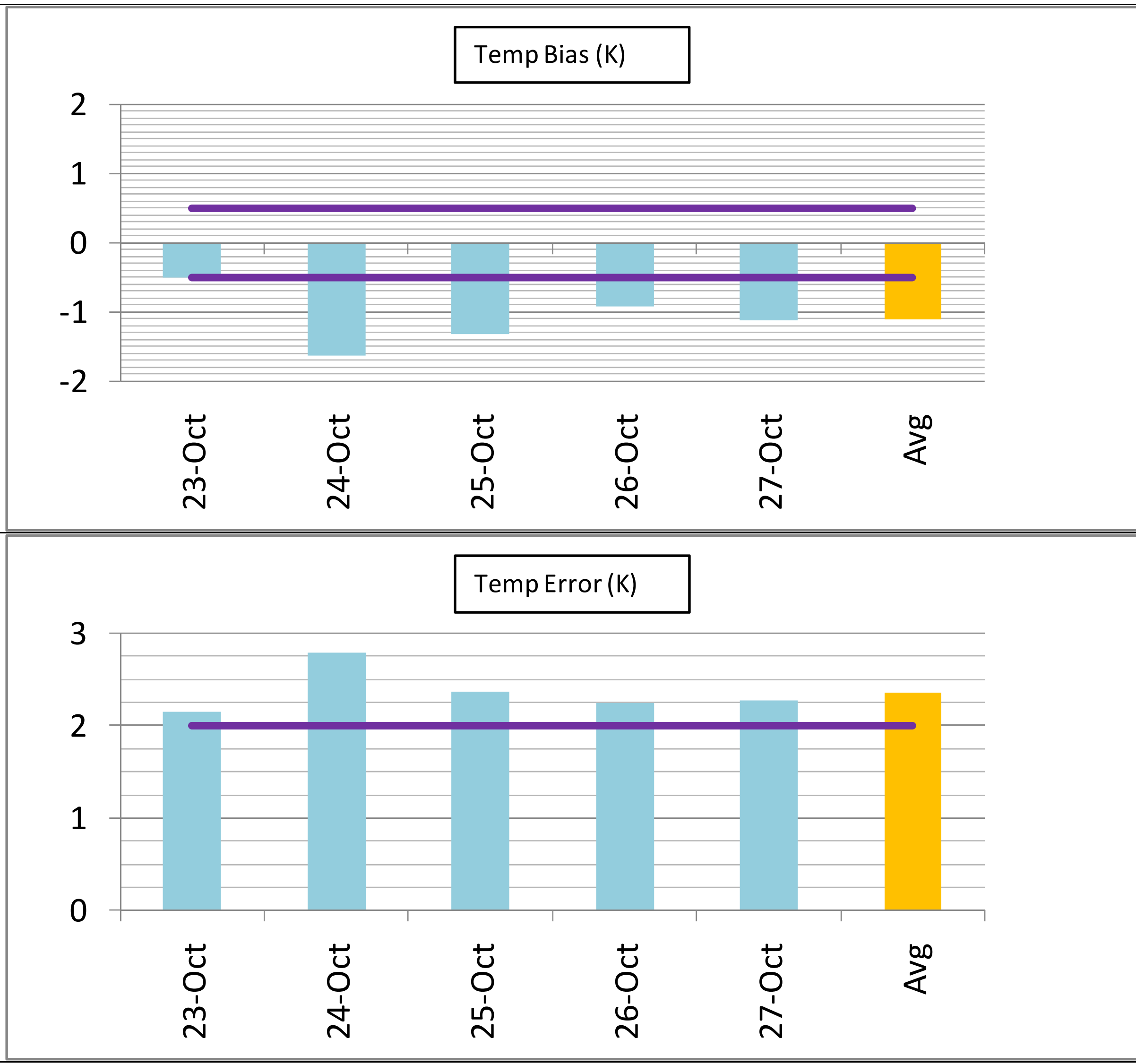


Figure 6-4c. ETEX MM5 model performance statistics of Bias (top) and Error (bottom) for temperature and comparison with benchmarks (purple lines).

6.3.2 Quality Assurance of the LRT Model Inputs

The input control files for the five LRT dispersion models were examined to assure that the intended model options were used in each of the simulations.

6.4 MODEL PERFORMANCE EVALUATION

The model performance of the five LRT dispersion models are evaluated using statistical measures as used in the ATMES-II study (Mosca et al., 1998) and recommended by DATEM (Draxler, Heffter and Rolph, 2002). Graphical comparisons are generated of the predicted and observed tracer spatial distributions.

6.4.1 Statistical Model Performance Evaluation

The spatial, temporal and global model performance of the five LRT models is evaluated using the statistical model performance metrics described in Section 2.4.

6.4.1.1 Spatial Analysis of Model Performance

Four spatial analysis model performance statics have been identified and are discussed in this section: FMS, FAR, POD and TS. Figure 6-5 displays the FMS spatial analysis performance metrics for the five LRT models and the ETEX tracer study field experiment. Recall that the FMS statistic is define as the overlap divided by the union of the predicted and observed tracer clouds with a perfect model receiving an FMS score of 100%.

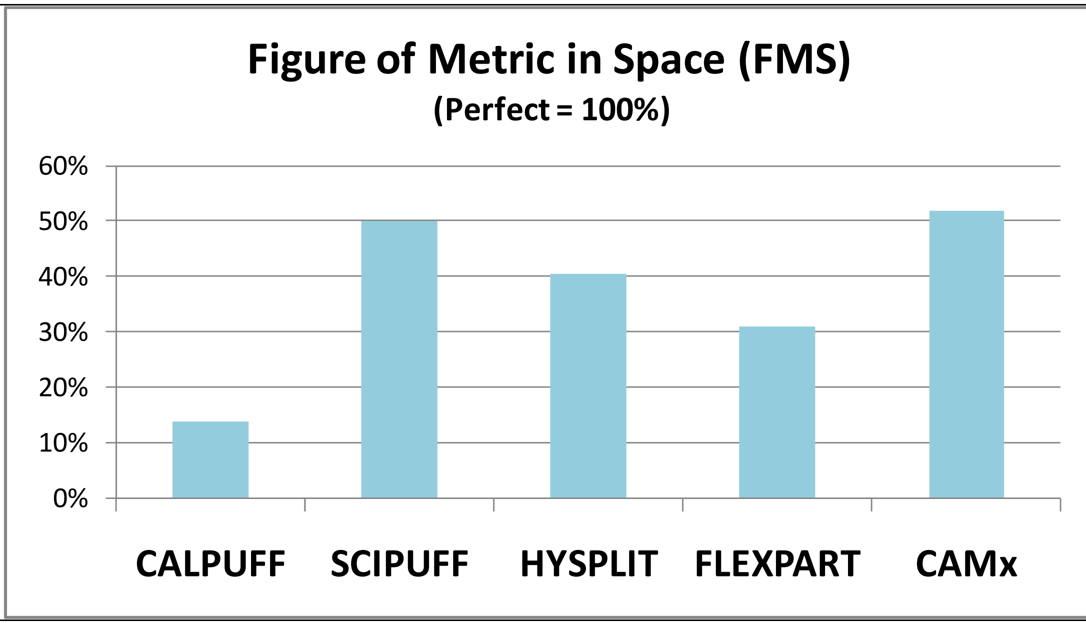


Figure 6-5. Figure of Merit in Space (FMS) statistical performance metric for the five LRT models and the ETEX tracer field experiment.

Figure 6-6 displays the False Alarm Rate (FAR) performance metrics. The FAR metric is defined by the number of times that a tracer concentration was predicted to occur at a monitor-time when no tracer was observed (i.e., a miss) divided by the number of times a tracer was predicted to occur at a monitor-time (i.e., sum of misses and hits); a perfect model (i.e., one that had no misses) would have a FAR score of 0%.

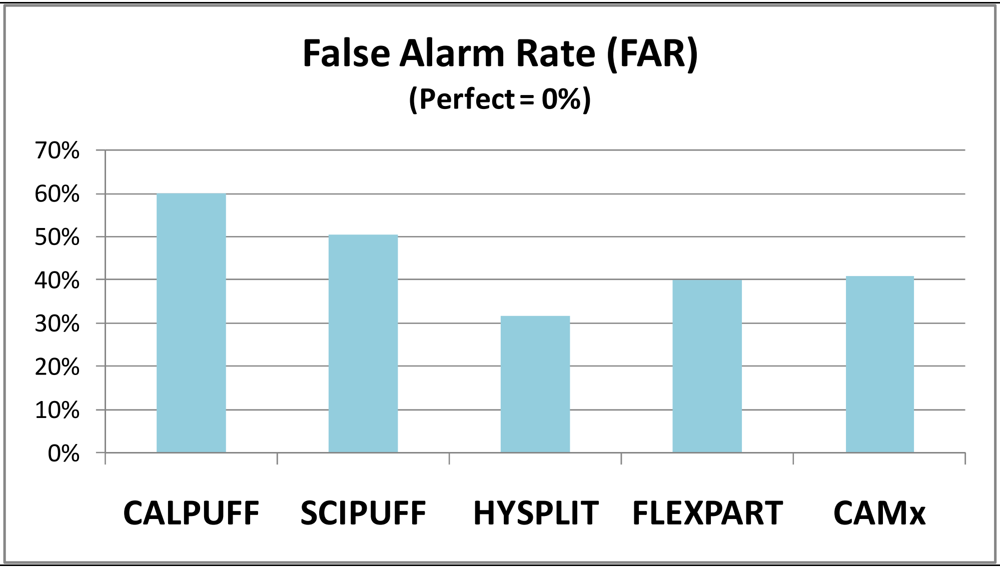


Figure 6-6. False Alarm Rate (FAR) statistical performance metric for the five LRT models and the ETEX tracer field experiment.

The Probability of Detection (POD) performance statistic is defined as the percent of the time the predicted and observed tracer both occurred at a monitor-time (i.e., a hit of tracer concentrations greater than 1 ngm^{-3}) divided by the number of times that the tracer was observed at any monitor-time (i.e., sum of hits and misses); a perfect model POD score would be 100% (i.e., anytime there was observed tracer at a monitor there was also predicted tracer at the monitor).

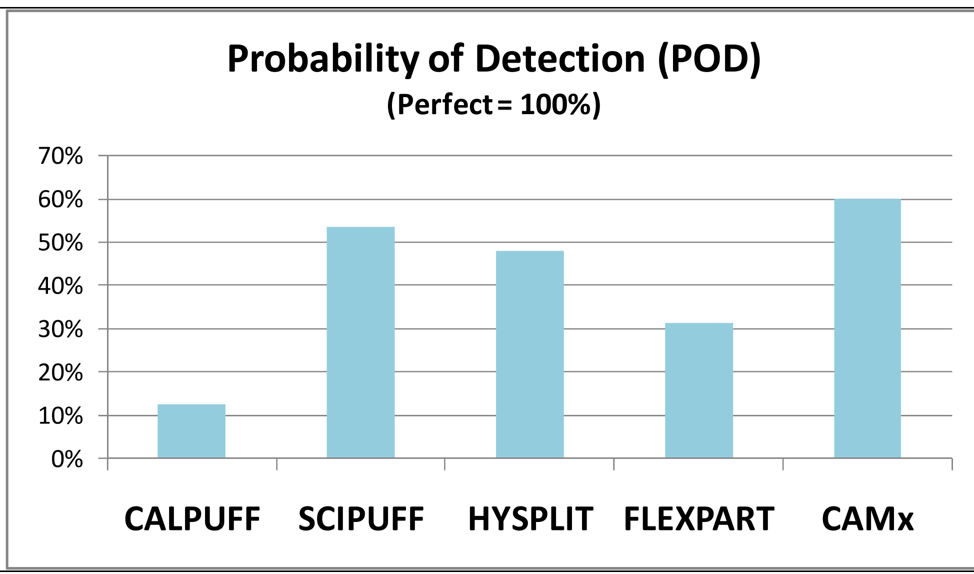


Figure 6-7. Probability of Detection (POD) statistical performance metric for the five LRT models and the ETEX tracer field experiment.

The Threat Score (TS) is the ratio of the number of times that a tracer is both predicted and observed at a monitor-time at the same time (i.e., common hits among the predictions and observations) divided by the number of monitor-time events that either a prediction or observed tracer occurred at a monitor (i.e., either a predicted or observed hits), with a perfect score of 100% (which means there were no occurrences when there was a predicted hit but an observed miss and vice versa).

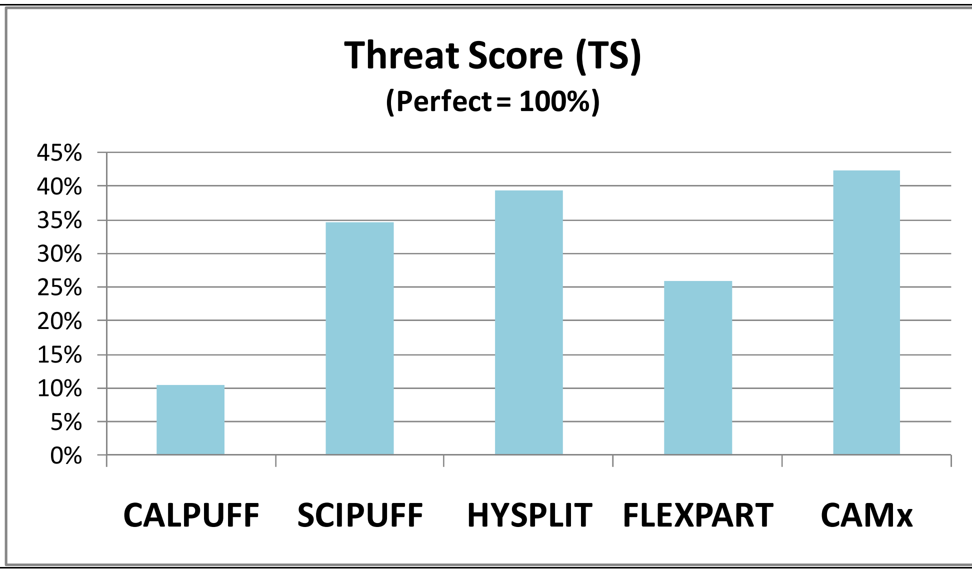


Figure 6-8. Threat Score (TS) statistical performance metric for the five LRT models and the ETEX tracer field experiment.

6.4.1.2 Global Analysis of Model Performance

Eight global statistical analysis metrics are used to evaluate the five LRT model performance using the ETEX data base that are described in Section 2.4 and consist of the FOEX, FA2, FA5, NMSE, PCC, FB, KS and RANK statistical metrics.

The Factor of Exceedance (FOEX) gives a measure of the scatter of the modeled predicted and observed and a level of underestimation versus overestimation of the model. FOEX is bounded by -50% to +50%. The within a Factor of α (FAα), where we used within a Factor of 2 (FA2) and 5 (FA5), also gives an indication of the amount of scatter in the predicted and observed tracer pairs, but no information on whether the model is over- or under-predicting. A perfect model would have an FAα score of 100%. A good performing model would have a FOEX score near zero and high FAα values. A model with a large negative FOEX and low FAα values would indicate an under-prediction tendency. Whereas a model with a large positive FOEX and low FAα would suggest a model that over-predicts.

Figure 6-9 displays the FOEX performance metrics for the five LRT models and the ETEX modeling period.

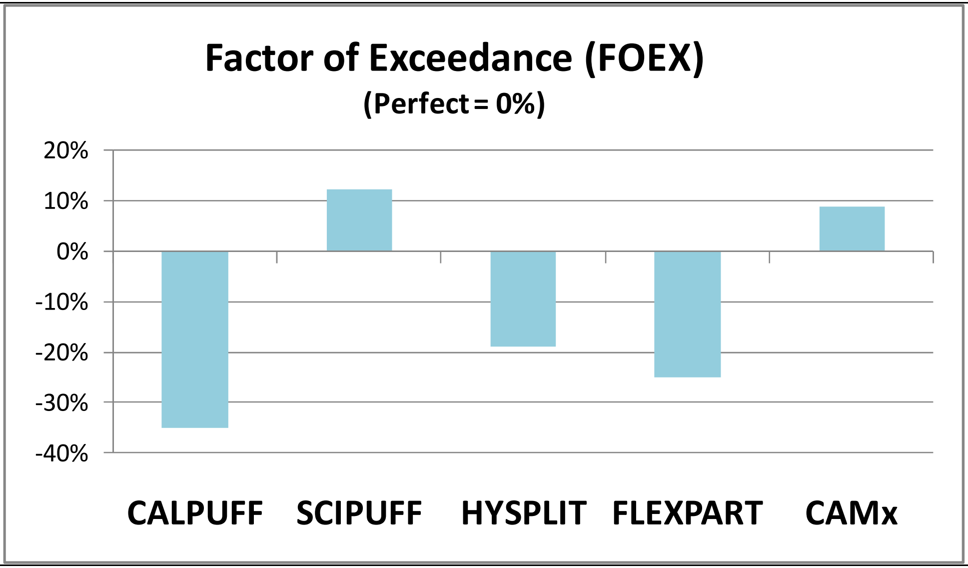


Figure 6-9. Factor of Exceedance (FOEX) statistical performance metric for the five LRT models and the ETEX tracer field experiment.

The rankings of the five LRT models are the same whether using the FA2 or FA5 performance metric.

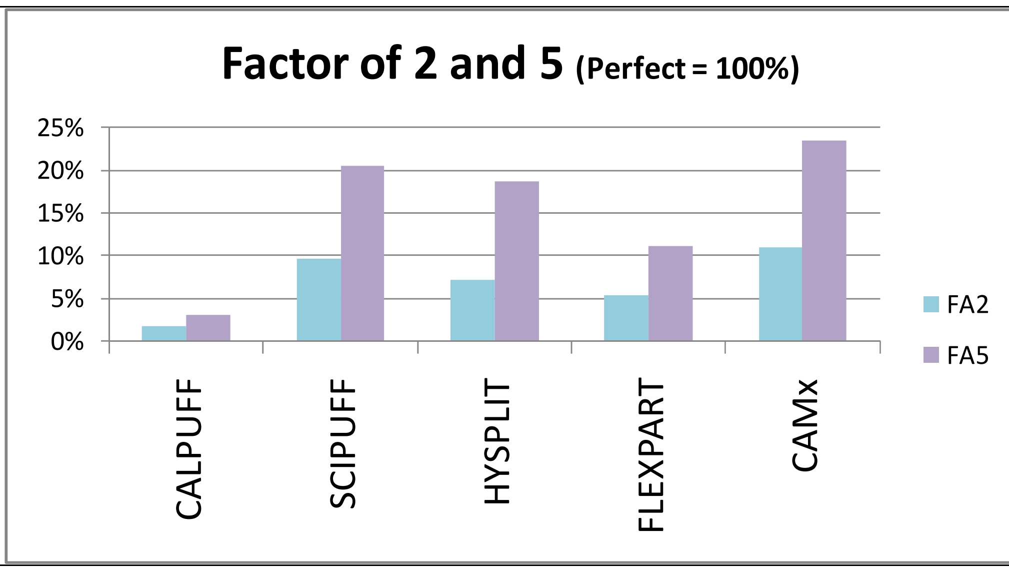


Figure 6-10. Factor of 2 (FA2, top) and Factor of 5 (FA5, bottom) statistical performance metric for the five LRT models and the ETEX tracer field experiment.

The scores for the Normalized Mean Squared Error (NMSE) statistical metrics for the five LRT models are given in Figure 6-11. The NMSE provides an indication of the deviations between the predicted and observed tracer concentrations paired by time and location with a perfect model receiving a 0.0 score.

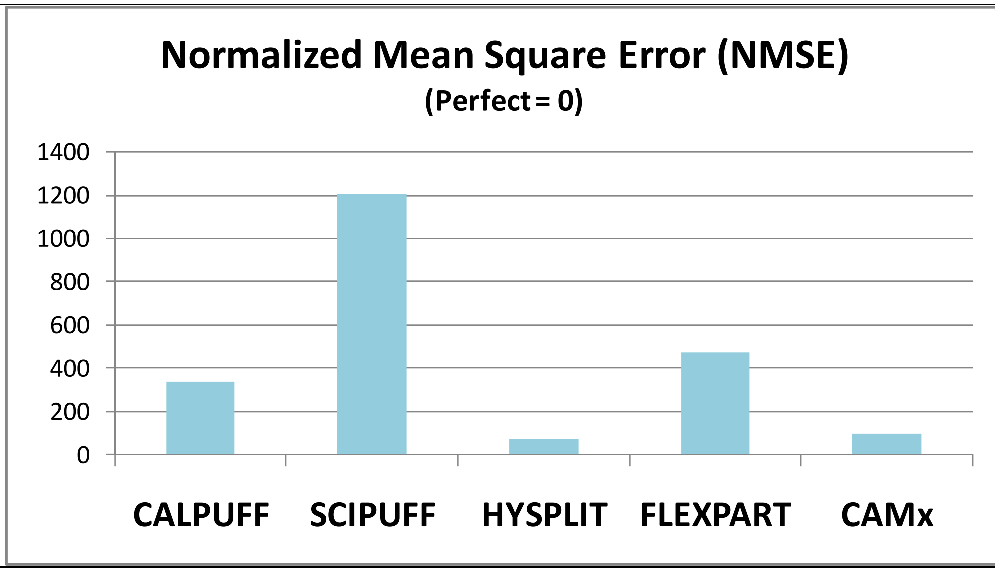


Figure 6-11. Normalized Mean Square Error (NMSE) statistical performance metric for the five LRT models and the ETEX tracer field experiment (pgm^{-3}).

The Pearson's Correlation Coefficient (PCC or R) ranges between -1.0 and +1.0, a model that has a perfect correlation with the observations would have a PCC value of 1.0. The PCC values for the five LRT models are shown in Figure 6-12. All of the models have positive PCCs so none are negatively correlated with the observe data.

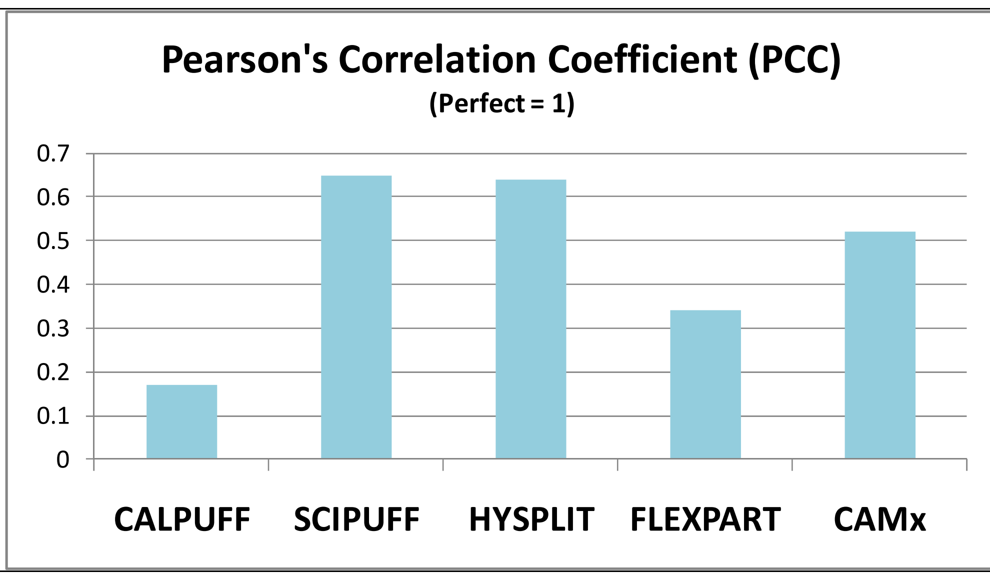


Figure 6-12. Pearson's Correlation Coefficient (PCC) statistical performance metric for the five LRT models and the ETEX tracer field experiment.

The Fractional Bias (FB) is a measure of bias in the deviations between the predicted and observed paired tracer concentrations and ranges from -2.0 to +2.0 with a perfect model receiving a 0.0 score. Figure 6-13 displays the FB parameter for the five LRT models. All five models exhibit a positive FB, which suggests an overestimation tendency.

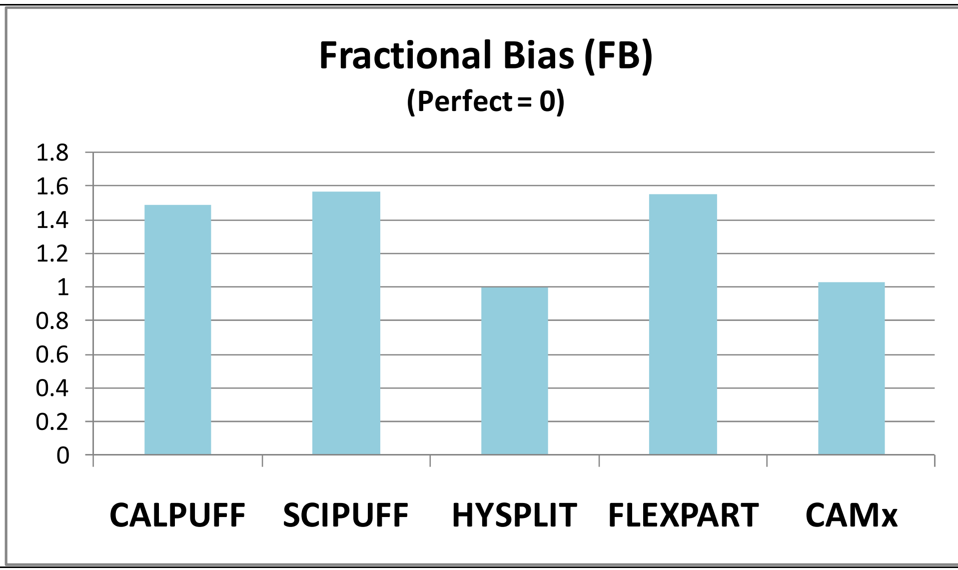


Figure 6-13. Fractional Bias(FB) statistical performance metric for the five LRT models and the ETEX tracer field experiment.

The Kolmogorov-Smirnoff (KS) parameter compares the frequency distributions of the predicted and observed tracer concentrations unmatched by time and location. It is the only unpaired statistical metric in the global statistics. The KS parameter ranges from 0% to 100% with a perfect model receiving a score of 0%. The KS parameters for the five LRT models and the ETEX modeling are shown in Figure 6-14.

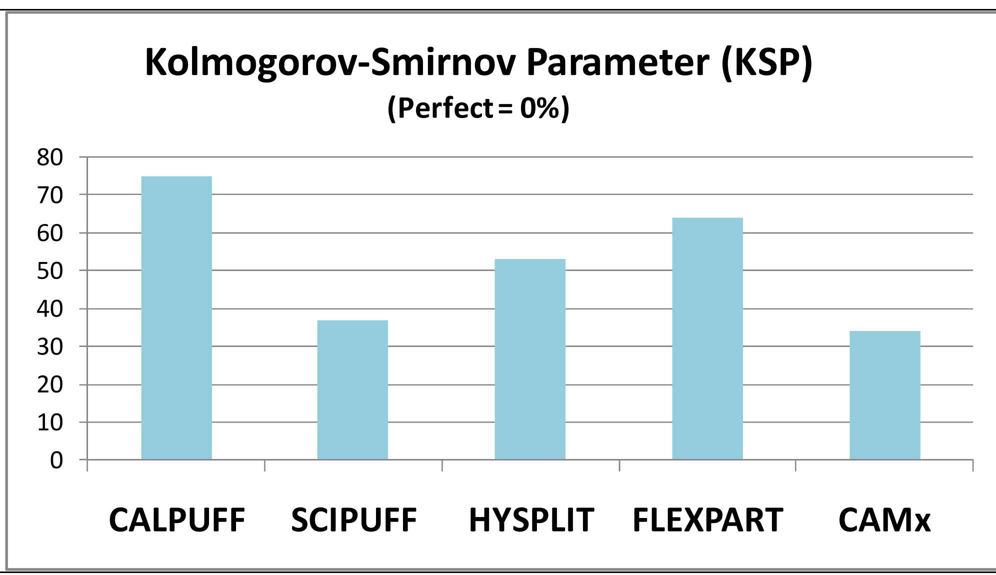


Figure 6-14. Kolmogorov – Smirnov Parameter (KSP) statistical performance metrics for the five LRT models and the ETEX tracer field experiment.

The RANK statistical performance metric was proposed by Draxler (2001) as a single model performance metric that equally ranks the combination of performance metrics for correlation (PCC or R), bias (FB), spatial analysis (FMS) and unpaired distribution comparisons (KS). The RANK metrics ranges from 0.0 to 4.0 with a perfect model receiving a score of 4.0. Figure 6-15 lists the RANK model performance statistics for the five LRT models. CAMx is the highest ranked model using the RANK metric with a value of 1.9. Note that CAMx scores high in all four areas of model performance (correlation, bias, spatial and cumulative distribution). The next highest ranking models according to the RANK metric are SCIPUFF and HYSPLIT with a score of 1.8.

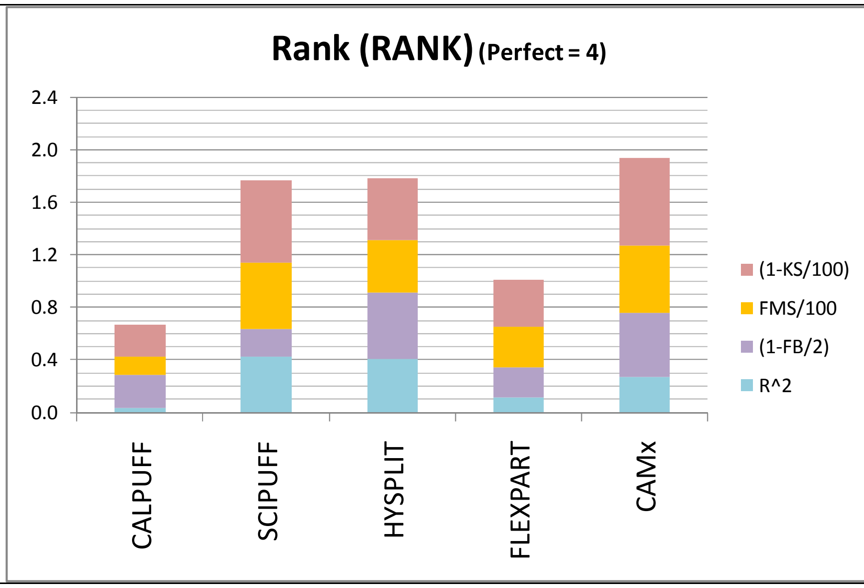


Figure 6-15. RANK statistical performance metric for the five LRT models and the ETEX tracer field experiment.

6.4.1.3 Summary of Model Ranking using Statistical Performance Measures

Table 6-1 summarizes the rankings between the five LRT models for the 11 performance statistics analyzed. Depending on the statistical metric, three different models were ranked first for a particular statistic with CAMx being ranked first most of the time (64%) and HYSPLIT ranked first second most (27%). In order to come up with an overall rank across all eleven statistics we average the modeled ranking order in order to come up with an average ranking that listed CAMx first, HYSPLIT second, SCIPUFF third, FLEXPART fourth and CALPUFF the fifth. This is the same ranking as produced by the RANK integrated statistics that combines the four statistics for correlation (PCC), bias (FB), spatial (FMS) and cumulative distribution (KS) giving credence that the RANK statistic is a potentially useful performance statistic for indicating over all model performance of a LRT dispersion model.

Table 6-1. Summary of model ranking using the statistical performance metrics.

Statistic	1st	2nd	3rd	4th	5th
FMS	CAMx	SCIPUFF	HYSPLIT	FLEXPART	CALPUFF
FAR	HYSPLIT	FLEXPART	CAMx	SCIPUFF	CALPUFF
POD	CAMx	SCIPUFF	HYSPLIT	FLEXPART	CALPUFF
TS	CAMx	HYSPLIT	SCIPUFF	FLEXPART	CALPUFF
FOEX	CAMx	SCIPUFF	HYSPLIT	FLEXPART	CALPUFF
FA2	CAMx	SCIPUFF	HYSPLIT	FLEXPART	CALPUFF
FA5	CAMx	SCIPUFF	HYSPLIT	FLEXPART	CALPUFF
NMSE	HYSPLIT	CAMx	CALPUFF	FLEXPART	SCIPUFF
PCC or R	SCIPUFF	HYSPLIT	CAMx	FLEXPART	CALPUFF
FB	HYSPLIT	CAMx	CALPUFF	FLEXPART	SCIPUFF
KS	CAMx	SCIPUFF	HYSPLIT	FLEXPART	CALPUFF
Avg. Ranking	CAMx	HYSPLIT	SCIPUFF	FLEXPART	CALPUFF
Avg. Score	1.55	2.27	2.73	3.82	4.64
RANK Ranking	CAMx	HYSPLIT	SCIPUFF	FLEXPART	CALPUFF

6.4.2 Spatial Displays of Model Performance

Figure 6-16 displays the observed tracer distribution 24, 36, 48 and 60 hours after the beginning of the tracer release as well as the predicted tracer distribution by CALPUFF, SCIPUFF, FLEXPART and CAMx. Note that the observed tracer spatial distribution plots in Figure 6-16 are color coded at the monitoring sites. Previously the spatial distribution of the observed tracer distribution was also presented using spatial interpolation from the monitoring sites in Figure 6-3b. However, such an interpolation is in itself a model and may not be correct, so in Figure 6-16 the observed tracer concentrations at the monitoring sites is presented for comparison with the five LRT models.

24 hours after the tracer release, the observed tracer was advected to the east-northeast and was present across northern France and Germany (Figure 6-16a, top left). CALPUFF advected the tracer with a more northeasterly direction than observed and underestimated the plume spread thereby missing the observed tracer concentrations in southern Germany (Figure 6-16a, top right). SCIPUFF (Figure 6-16a, middle left) also appeared to advect the tracer with more of a northeast direction than observed, but had more plume spread so was better able to capture the occurrence of observed tracer concentrations in southern Germany. FLEXPART (Figure 6-16a, middle right) and HYSPLIT (Figure 6-16a, bottom left) both correctly advect the tracer initially in the east-northeast direction, but FLEXPART greatly underestimates the observed plume spread on the ground with HYSPLIT also underestimating the plume spread but not as much as FLEXPART. CAMx also appears to initially transport the tracer with more of a northeasterly than east-northeast direction as seen with SCIPUFF. Like SCIPUFF, the CAMx tracer plume has a southerly bulge that begins to capture the occurrence of the observed tracer concentrations in southern Germany that the other three LRT dispersion models miss completely. All of the models fail to reproduce the leading edge of the observed tracer cloud in northeastern Germany, with SCIPUFF and CAMx best able to simulate the observed front of the tracer cloud. The LRT dispersion models underestimation of the location of the leading edge of the observed tracer cloud is likely related to the MM5 model wind speed underestimation bias (see Figure 6-4a). SCIPUFF tends to have an overestimation bias of both concentrations and spatial extend of the observed tracer 24 hours after its release.

The predicted and observed tracer distribution 36 hours after its release is shown in Figure 6-16b. The observed tracer plume moved eastward and traverses Germany 36 hours after the start of the release and is stretched from the west coast of Sweden in the north to Hungary in the South. CALPUFF is displacing the tracer too far to the northeast with the centerline over the North Sea stretching from the northern tip of France to southern tip of Sweden and missing most of the observed tracer concentrations in France, Germany and Czechoslovakia. SCIPUFF covers the spatial extent of the observed tracer cloud, and then some, correctly estimating the coverage across Germany and Czechoslovakia. FLEXPART reproduces the easterly transport of the observed tracer clouds 36 hours after the start of the release, but greatly underestimates the ground level plume spread. HYSPLIT also reproduces the easterly transport of the observed tracer plume but also understates the plume spread missing the observed tracer concentrations in southern Germany and Czechoslovakia. CAMx has a similar distribution as SCIPUFF with less of an overestimation bias locating the tracer center of mass slightly too far north. After 36 hours from the start of the tracer release the leading edge of the observed tracer is just entering Poland from Germany, which is reproduced well by SCIPUFF, HYSPLIT and CAMx with FLEXPART having a lag and CALPUFF locating the leading edge of the tracer too far north.

By 48 hours after the beginning of the tracer release, the observed tracer cloud is exhibiting a northwest to southeast orientation stretching from Denmark in the northwest to Hungary in the southeast (Figure 6-16c). The CALPUFF tracer plume, however, is advected too far north into the North Sea and southern Finland with a circular Gaussian puff distribution. SCIPUFF correctly reproduces the northwest to southeast orientation of the observed tracer cloud and almost completely covers the observed tracer cloud but appears to overestimate the spatial extent and concentrations of the observed tracer. HYSPLIT and FLEXPART also are exhibiting a northwest to southeast orientation of the observed tracer cloud but both models, and especially FLEXPART, understate the spatial spread of the observed ground level tracer concentrations. CAMx reproduces the northwest to southeast orientation of the observed tracer distribution and appears to better match the observed tracer plume spread than SCIPUFF (overstated) and FLEXPART and HYSPLIT (understated).

After 60 hours after the beginning of the tracer release, the observed tracer cloud still has the northwest to southeast orientation that stretches from southern Finland in the northwest to the most western point of Romania. The CALPUFF model has advected its circular puffs to the north with the center over the North Sea just west of southern Finland almost completely missing the spatial extent of the observed tracer. The other four LRT dispersion models are correctly estimating the northwest to southeast orientation of the observed tracer pattern 60 hours after the beginning of the tracer release. However, the remaining four LRT models (less CALPUFF) estimate different amounts of plume spread with FLEXPART estimating a very narrow predicted tracer cloud that understates the observed spread of the tracer footprint. SCIPUFF estimated the largest spatial extent of the tracer cloud that is much larger than observed. HYSPLIT and CAMx estimated tracer spread that is closer to what was observed.

The comparison of the spatial distribution of the predicted and observed tracer concentrations from the ETEX1 experiment helps explain the statistical model performance presented earlier. The poor performance of the CALPUFF model is because it keeps the tracer in a circular Gaussian plume distribution that is advected too far north and fails to reproduce the elongation and stretching of the observed tracer cloud in the northwest to southeast orientation. The other four LRT dispersion models do allow the predicted tracer cloud to take on the northwest to southeast distribution matching the basic features of the observed tracer footprint well, but with different amounts of plume spread. FLEXPART greatly understates the amount of tracer plume spread and observed surface concentrations, whereas SCIPUFF overstates the amount of plume spread as well as the surface concentrations.

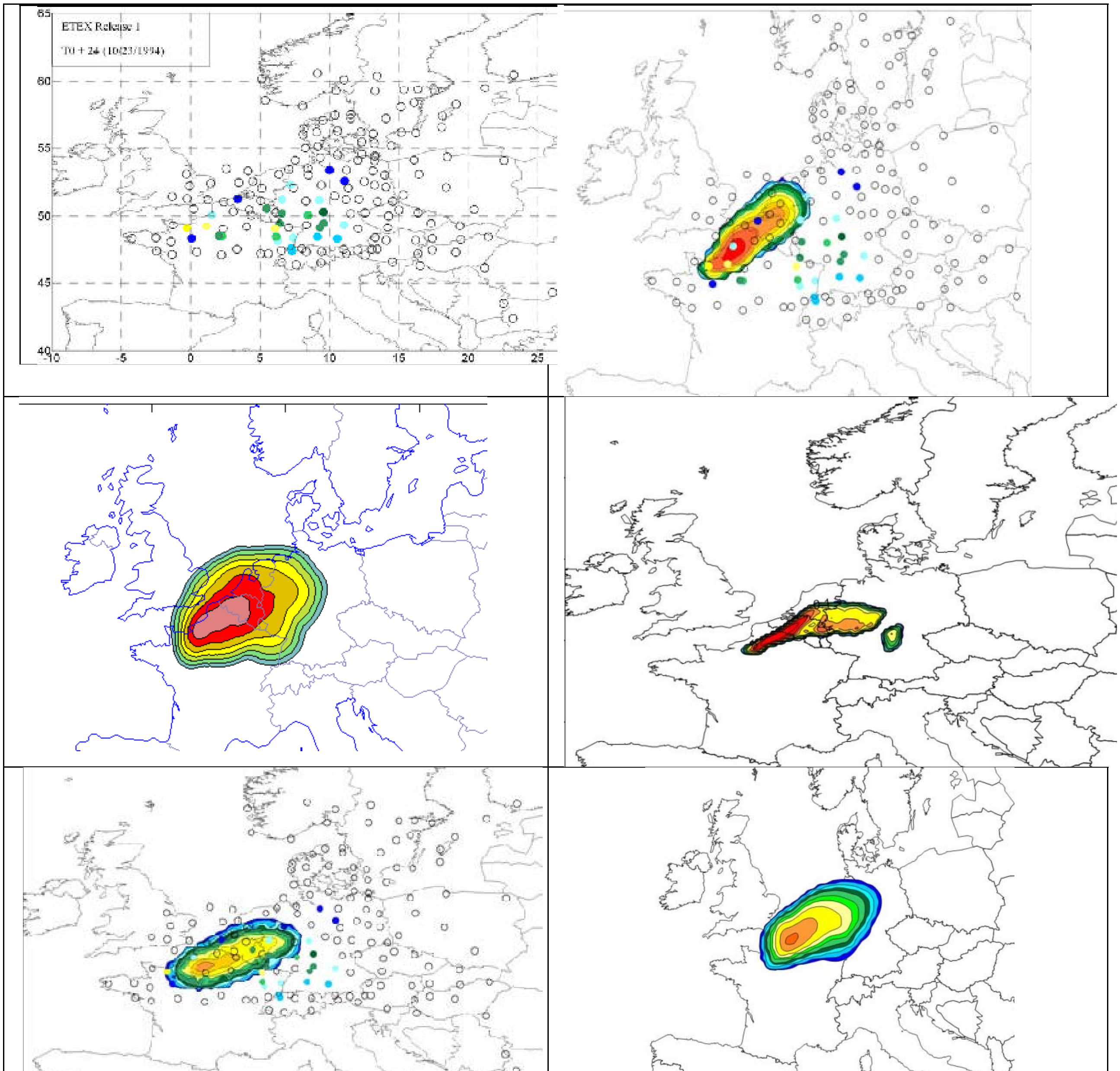


Figure 6-16a. Comparison of spatial distribution of the ETEX tracer concentrations 24 hours after release for the observed (top left), CALPUFF (top right), SCIPUFF (middle left), FLEXPART (middle right), HYSPLIT (bottom left) and CAMx (bottom right).

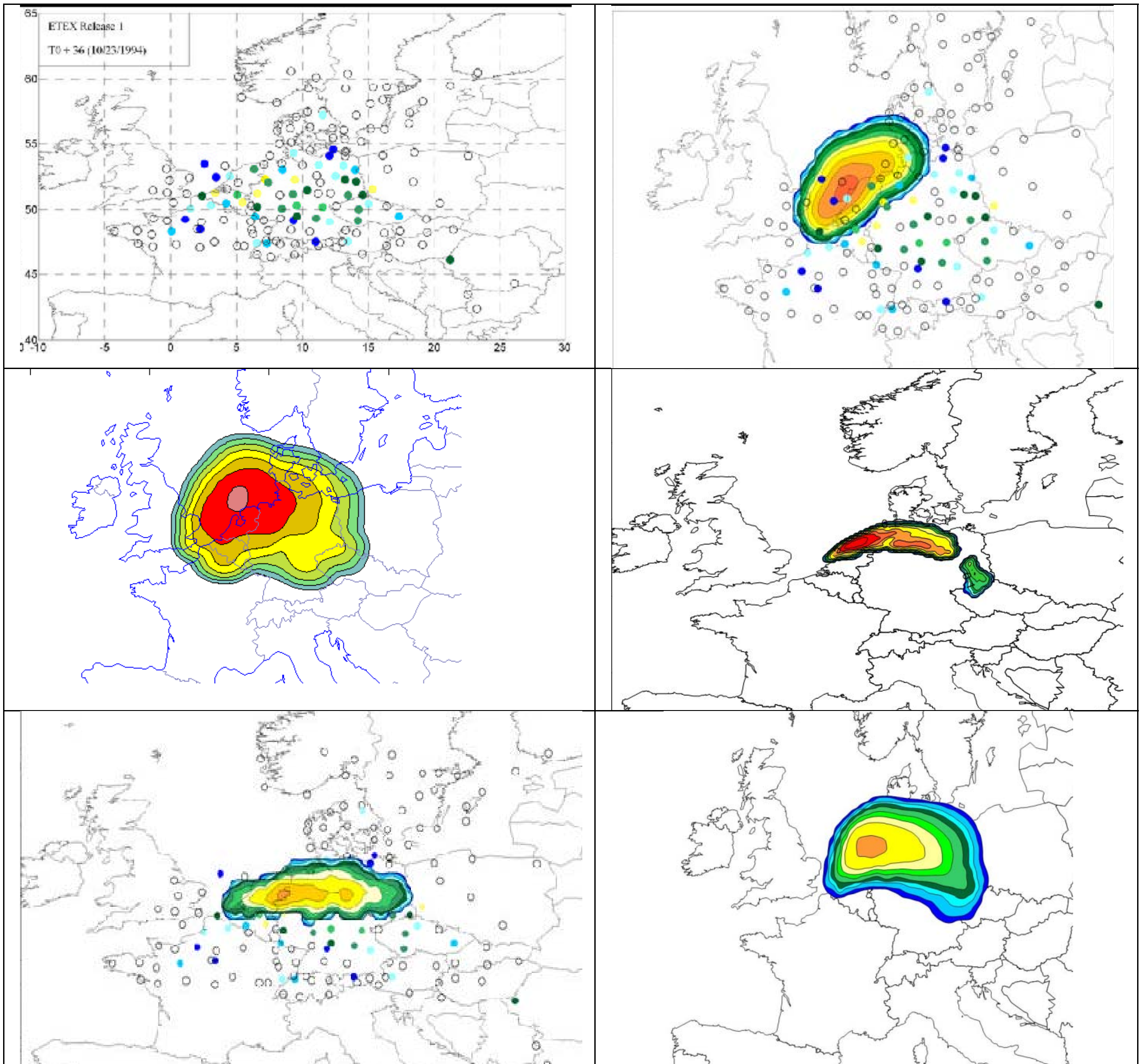


Figure 6-16b. Comparison of spatial distribution of the ETEX tracer concentrations 36 hours after release for the observed (top left), CALPUFF (top right), SCIPUFF (middle left), FLEXPART (middle right), HYSPLIT (bottom left) and CAMx (bottom right).

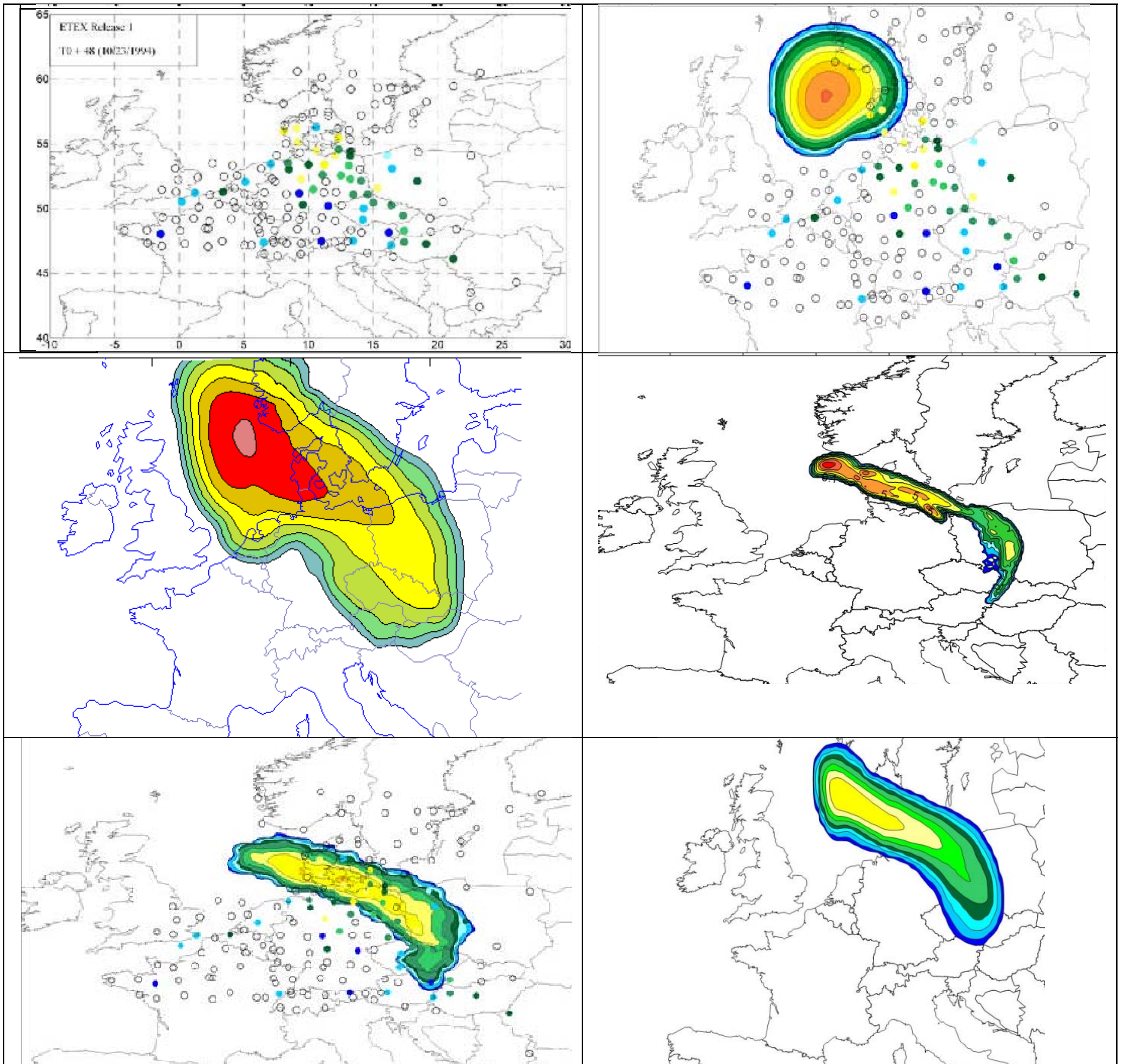


Figure 6-16c. Comparison of spatial distribution of the ETEX tracer concentrations 48 hours after release for the observed (top left), CALPUFF (top right), SCIPUFF (middle left), FLEXPART (middle right), HYSPLIT (bottom left) and CAMx (bottom right).

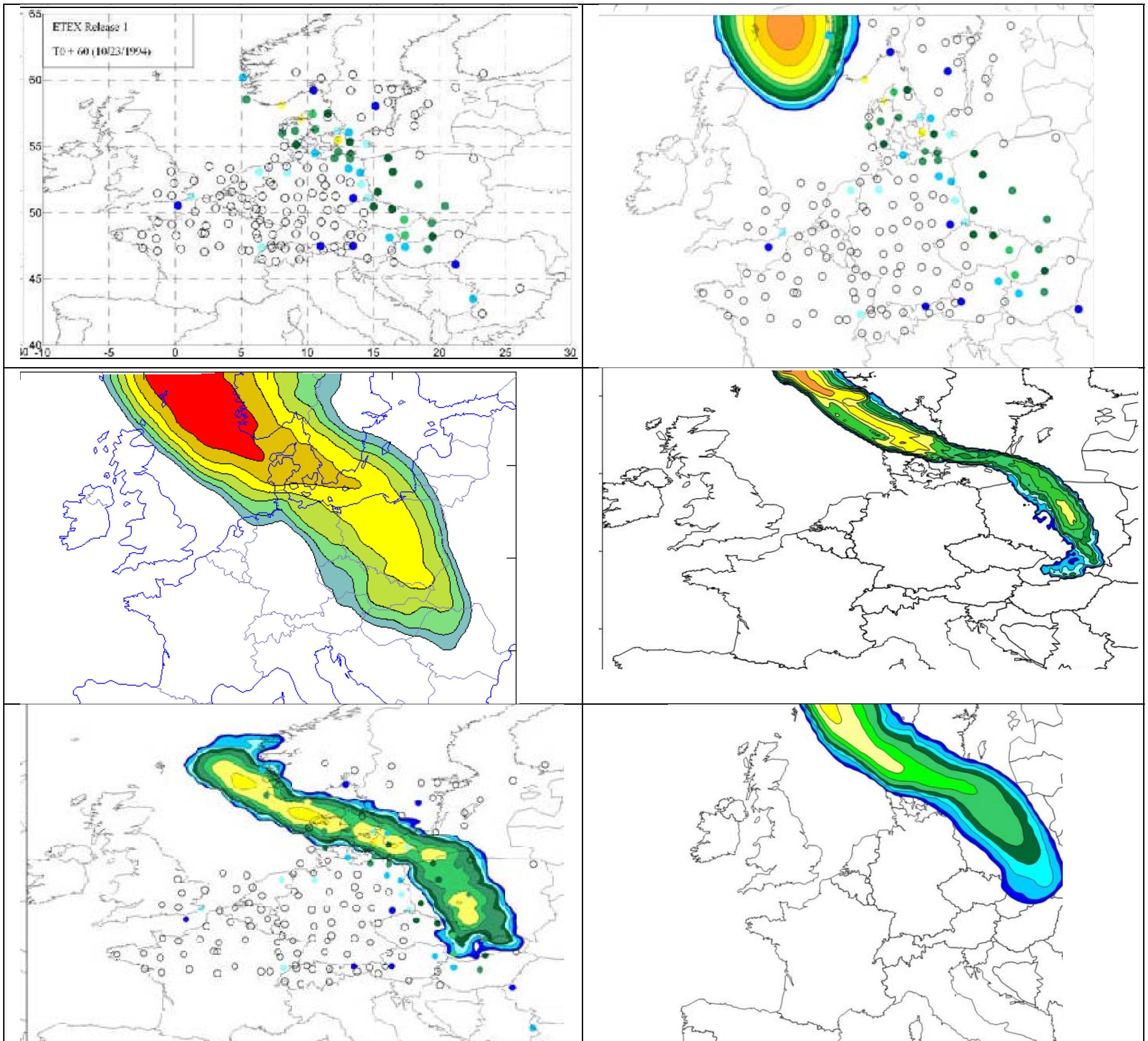


Figure 6-16d. Comparison of spatial distribution of the ETEX tracer concentrations 60 hours after release for the observed (top left), CALPUFF (top right), SCIPUFF (middle left), FLEXPART (middle right), HYSPLIT (bottom left) and CAMx (bottom right).

6.4.3 CAMx Sensitivity Tests

Sixteen CAMx sensitivity tests were conducted to investigate the effects of vertical diffusion, horizontal advection solvers and use of the sub-grid scale Plume-in-Grid (PiG) module on the model performance for the ETEX tracer experiment.

<u>Plume-in-Grid (PiG) Module</u>: The PiG module treats the near-source plume dispersion (and chemistry if applicable) of a point source plume using a subgrid-scale Lagrangian puff module. The mass from the PiG puff module is transferred to the grid model when the plume size is commensurate with the grid cell size used in the CAMx simulation. Two types of PiG sensitivity tests were conducted in this study to investigate the effects of the PiG module on model performance:

- NoPiG: The tracer emissions were released directly into the CAMx 36 km grid cell containing the tracer release location calculating plume rise using the local meteorological conditions to inject the emissions into the appropriate vertical layer.
- PiG: Calculate plume rise using local meteorological conditions and simulate the early evolution of plume dispersion using the PiG module.

Vertical Diffusion Coefficients (Kz): The Kz coefficients define the rate of vertical mixing in a column of grid cells in CAMx. MM5 meteorological model does not directly output Kz, thus the MM5CAMx pre-processor has several different algorithms for diagnosing the Kz coefficients. Four different Kz algorithms were evaluated in the CAMx sensitivity tests in this study:

- OB70: O'Brien 1970 algorithm for calculation Kz values by diagnosing them from the MM5 output.
- TKE: The Eta planetary boundary layer (PBL) scheme used in the ETEX MM5 meteorological modeling has a Turbulent Kinetic Energy (TKE) formulation. When using a TKE PBL scheme, MM5CAMx can calculate the Kz coefficients directly from the TKE values, rather than diagnosing them from the other meteorological variables in the MM5 output.
- ACM2: The Asymmetric Convective Mixing (ACM2) algorithm has two components: a standard Kz scheme that calculates diffusion between two adjacent grid cells in a column; and a non-local diffusion scheme that can calculate diffusion between grid cells in a column that are not adjacent. In CAMx, the ACM2 scheme will deduce when convective activity is present in a column of grid cells and add the non-local diffusion to the standard local diffusion based on the Kz coefficients.
- CMAQ: Use the algorithm for calculating Kz from the CMAQ modeling system (Byun and Ching, 1999).

Horizontal Advection Solver: Horizontal advection (transport) is solved in CAMx using finite difference algorithms that were explicitly developed for simulating transport and limit numerical diffusion that can artificially reduce concentration peaks. Two horizontal transport algorithms are implemented in CAMx and their effect on model performance for the ETEX experiment was evaluated:

Bott: The Bott (1989) scheme is a positive definite transport scheme that limits numerical diffusion.

PPM: The Piecewise Parabolic Method (PPM; Colella and Woodward, 1984) is a higher order positive definite transport scheme that is also designed to limit numerical diffusion.

The configuration of CAMx presented in the previous sections comparing model performance against the other four LRT models was a standard configuration used in many regional model applications:

- Don't use PiG subgrid-scale puff module (NoPiG)
- Use of CMAQ-like Kz vertical diffusion coefficients (CMAQ)
- Use of PPM horizontal advection solver (PPM)

6.4.3.1 NoPiG CAMx Sensitivity Tests

Figure 6-17 displays the CAMx spatial model performance statistics for the sensitivity tests that were run without using the PiG subgrid-scale puff module. For the FMS statistic, the CMAQ Kz and PPM horizontal transport sensitivity test (CMAQ/PPM) is performing the best with a FMS value of 51.8% followed by CMAQ/Bott (50.9%) and ACM2/PPM (50.8%). Vertical diffusion has the biggest effect with the ranking of the algorithms from best to worst using the FMS statistic being CMAQ, ACM2, TKE and OB70. Whereas, for the horizontal advection solver the PPM algorithm performs slightly better than Bott using the FMS statistic.

For the FAR statistic, CMAQ/Bott has the best score (39.0%) followed by CMAQ/PPM (41.0%). Overall CMAQ is the best performing vertical diffusion formulation and Bott performs better than PPM for horizontal advection using the FAR statistic.

For the POD and TS spatial statistics, the CMAQ and TKE vertical diffusion algorithms perform substantially better than the OB70 and ACM2 approaches. There are much smaller differences in the model performance using the two advection solvers for the POD and TS statistics.

In summary, based on the spatial statistics, the CMAQ Kz algorithm appears to be the best performing approach for vertical mixing followed by TKE. And with the exception of the FAR statistic, PPM produces slightly better spatial model performance statistics than the Bott horizontal advection solver. The differences in vertical diffusion algorithms has a greater effect on CAMx model performance than the differences in horizontal advection solvers.

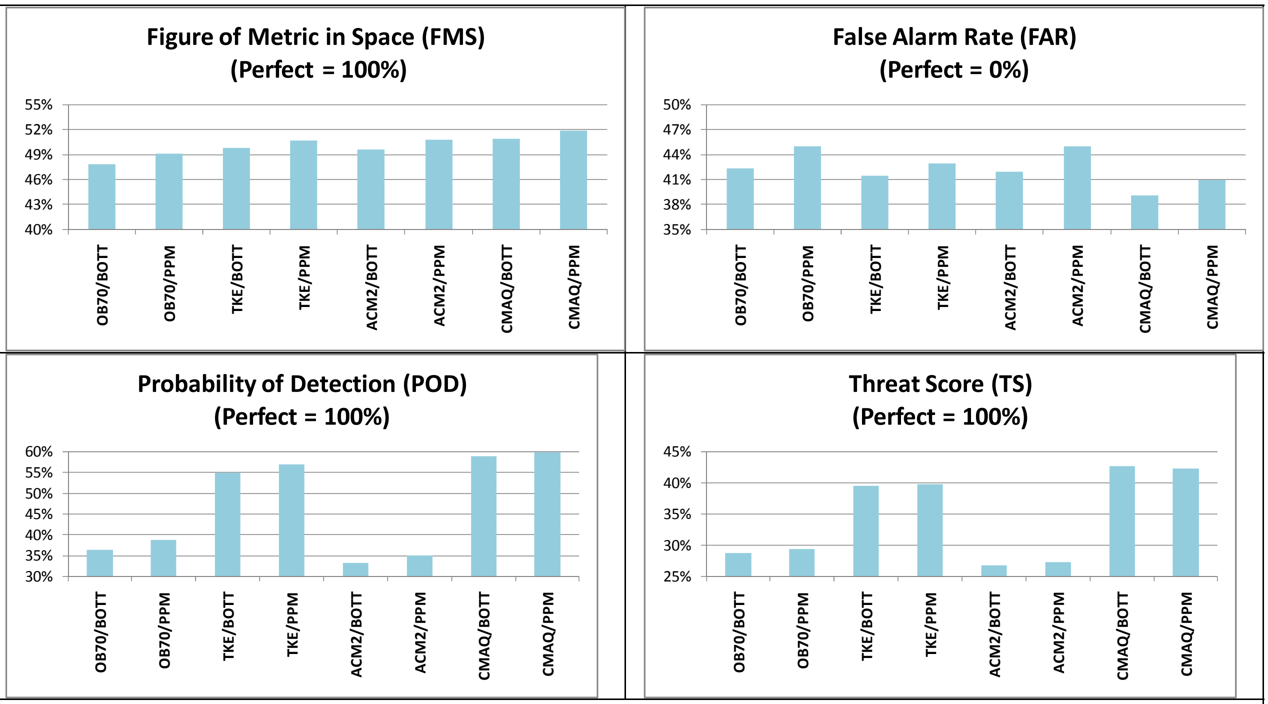


Figure 6-17. Spatial model performance statistics for the CAMx sensitivity tests without using the PiG subgrid-scale puff module (NoPiG).

Figure 6-18 displays the global statistics for the CAMx NoPiG sensitivity tests with Figures 6-18a and 6-18b containing the statistical metrics where the best performing model has the, respectively, lowest and highest score. For the FOEX metric, the vertical diffusion algorithm has the biggest effect with the ACM2 scoring the best with an essentially zero FOEX score followed by OB70 with values -2.4% (OB70/Bott) and -3.4% (OB70/PPM). The TKE (8.5%) and PPM (8.7% and 9.1%) have the highest (worst) FOEX scores. The FOEX metrics using the two alternative horizontal advection algorithms are essentially the same.

Using the NMSE statistical performance metric, the CMAQ vertical diffusion scheme performs best with OB70 and TKE producing very similar results next, with the ACM2 exhibiting the worst NMSE performance results (Figure 6-18a, top right). The PPM horizontal advection scheme is performing slightly better than the Bott algorithm based on the NMSE metric.

The CMAQ vertical diffusion scheme is also the best performing method according to the FB metrics followed by the TKE then ACM2 and then OB70 in last. According to the FB metrics, PPM performs slightly better than Bott.

For the KS parameter, the OB70 is the best vertical mixing method with CMAQ barely beating out ACM2 in second and TKE slightly worse. The PPM horizontal advection solver is performing slightly better than Bott for the KS parameter.

For the within a factor of 2 and 5 metrics (FA2 and FA5, Figure 6-18b, top), the CMAQ and TKE vertical mixing approaches are clearly performing better than the OB70 and ACM2 methods and the PPM horizontal advection solver is clearly performing better than Bott. For the FA2, the TKE/PPM is the best performing configuration (11.4%) followed by CMAQ/PPM (10.9%), Whereas for the FA5 the reverse is true with CMAQ/PPM being the best performing configuration (23.4%) followed by TKE/PPM (22.2%).

There is essentially no difference in the PCC statistic using the two horizontal advection solvers (Figure 6-18b, bottom right). According to the PCC metric, CMAQ is the best performing vertical diffusion approach (0.52) followed by TKE (0.37 and 0.38), OB70 (0.35) and ACM2 (0.26 and 0.27).

The final panel in Figure 6-18b (bottom right) displays the overall RANK statistic. The RANK statistics orders the model performance of the CAMx configurations without PiG as follows:

1. CMAQ/PPM (1.94)
2. CMAQ/Bott (1.90)
3. TKE/PPM (1.70)
4. OB70/PPM (1.66)
5. TKE/Bott (1.65)
6. ACM2/PPM (1.60) (tied)
7. OB70/Bott (1.60) (tied)
8. ACM2/Bott (1.54)

Based on this analysis the CMAQ Kz coefficients is the best performing vertical diffusion approach followed by TKE and the PPM horizontal advection algorithm is performing slightly better than Bott. The vertical diffusion algorithm has a greater effect on CAMx model performance compared to the choice of horizontal advection solvers.

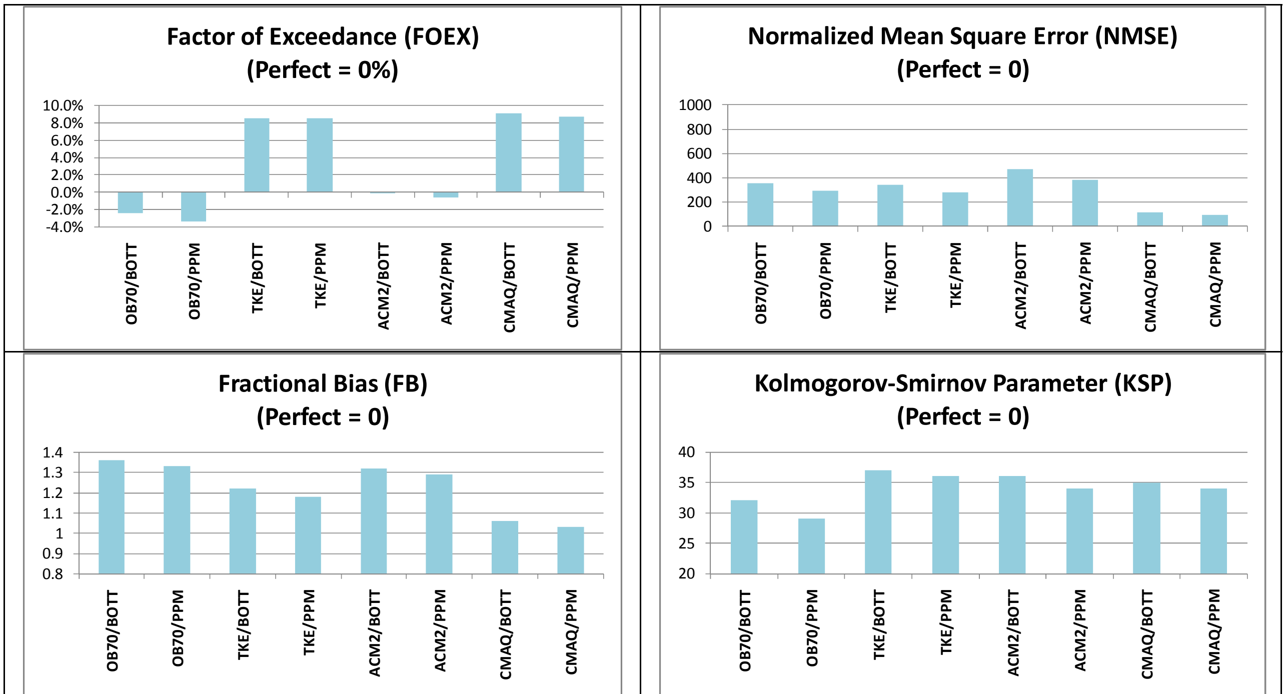


Figure 6-18a. Global model performance statistics for the CAMx sensitivity tests without using the PiG subgrid-scale puff module.

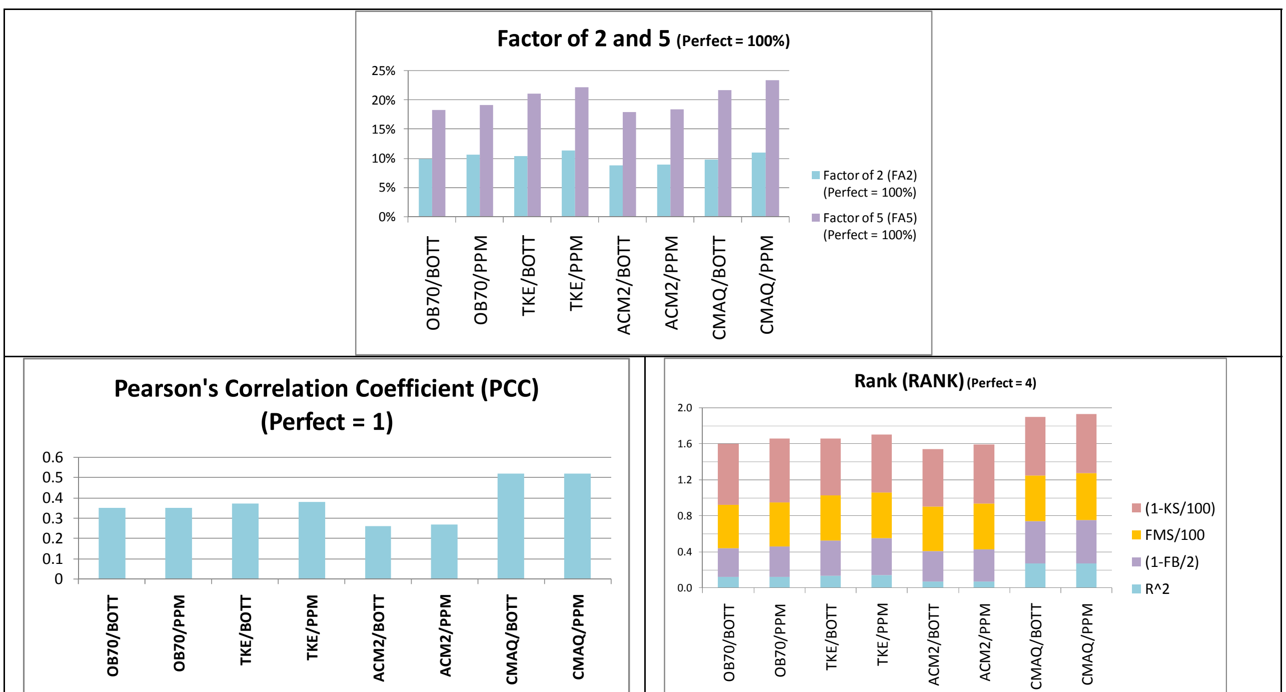


Figure 6-18b. Global model performance statistics for the CAMx sensitivity tests without using the PiG subgrid-scale puff module.

6.4.3.2 Effect of PiG on Model Performance

Whether better model performance is obtained using the PiG module or not frequently depends on the statistical metric being analyzed and the CAMx model configuration (vertical diffusion algorithm and horizontal advection solver). However, whether the PiG is used or not has very little difference on the rankings of the CAMx model performance using the alternative vertical mixing and horizontal advection approaches. In general, it appears that the CAMx model performance without the PiG is performing slightly better than its performance using the PiG.

The spatial performance statistics are sometimes improved and sometimes degraded when the PiG module is invoked. For the global statistics, the PCC performance statistic is degraded by -11% to -37% (-0.03 to -0.13 points) when the PiG module is invoked. Similarly, use of the PiG versus NoPiG module increases (degrades) the FB metric by 5 to 18 percent and also increases (degrades) the NMSE metrics for all model configurations.

Table 6-2 summarized the RANK model performance statistic for the different CAMx model configurations with and without the PiG module. For each model vertical diffusion/horizontal advection configuration, using the PiG module always results in slightly lower RANK statistics that are from -3.9% to -8.5% lower than when the PiG module is not used. The ranking of the top four CAMx vertical diffusion/horizontal advection configurations remains unchanged whether the PiG module is used or not. And by far the most important parameter examined in regards to the RANK model performance statistics for the ETEX experiment in the CAMx sensitivity tests is the vertical mixing algorithm, with the CMAQ Kz parameterization producing the best four RANK model performance statistics out of the 16 sensitivity tests: (1) NoPiG/CMAQ/PPM; (2) NoPiG/CMAQ/Bott; (3) PiG/CMAQ/PPM; and (4) PiG/CMAQ/Bott.

Table 6-2. CAMx RANK model performance statistic and model rankings for different model configurations with and without using the PiG subgrid-scale puff model.

Model Configuration	Without PiG Module		With PiG Module		PiG-NoPiG	
	RANK	Model Ranking	RANK	Model Ranking	ΔRANK	Percent
OB70/BOTT	1.60	7[a]	1.53	6[a]	-0.07	-4.4%
OB70/PPM	1.66	4	1.55	4	-0.11	-6.6%
TKE/BOTT	1.65	5	1.51	7	-0.14	-8.5%
TKE/PPM	1.70	3	1.56	3	-0.14	-8.2%
ACM2/BOTT	1.54	8	1.48	8	-0.06	-3.9%
ACM2/PPM	1.60	6[a]	1.53	5[a]	-0.07	-4.4%
CMAQ/BOTT	1.90	2	1.76	2	-0.14	-7.4%
CMAQ/PPM	1.94	1	1.80	1	-0.14	-7.2%

[a] tied

6.4.4 CALPUFF Sensitivity Tests

Most CALPUFF applications have limited the distance downwind that the model is applied for to less than 300 km from the source. However, the evaluation of CALPUFF in the ETEX study has applied the model to much farther downwind distances. The issue of the downwind applicability of the CALPUFF model was raised in the FLAG (2000) report and EPA's June 26-27, 2000 7th Conference on Air Quality Modeling[19] that proposed to list CALPUFF as an EPA recommended model for far-field applications. However, when CALPUFF was designated an EPA recommended far-field model in a 2003 Federal Register (FR) notice, EPA noted that "*…since the 7th Modeling Conference, enhancements were made to CALPUFF that allow puffs to be split both horizontally (to address wind direction shear) and vertically (to address spatial variation in meteorological conditions). These enhancements likely will extend the system's ability to treat transport and dispersion beyond 300 km*" (68 FR 18441). EPA goes on to further state that "*…Future performance comparisons for transport beyond 300 km are likely to extend the applicability and use of the modeling system, and we intend to watch for such evaluations very diligently. In an effort to keep the public abreast with the latest findings, EPA requests that evaluation results of the CALPUFF modeling system be sent to us (SCRAM webmaster) in an electronic format suitable for distribution, or that citations be provided for copyrighted material. EPA will post this information on its website for review and assessment*" (EPA, 2003).

Despite the passage of eight years since EPA's request for CALPUFF evaluation regarding its suitability for application beyond 300 km, no such documentation has been submitted. Thus, the ETEX CALPUFF evaluation serves as an important source of information on the downwind applicability of CALPUFF. In this section we present two types of performance analysis:

- Analyze the CALPUFF model performance as a function of distance from the source to determine whether the poor performance of CALPUFF relative to the other LRT models is related to applying the model beyond its downwind distance of applicability; and
- Perform CALPUFF puff splitting sensitivity tests to determine whether puff splitting can increase the downwind distance applicability of CALPUFF, as suggested in the 2003 Federal Register notice.

6.4.4.1 Time Dependent Model Performance

Figure 6-19 displays the FMS model performance statistic for the five LRT models as a function of time from the beginning of the tracer release in the ETEX experiment. Although the CALPUFF model performance does degrade with time (distance), even close to the source it is performing worse than the other LRT models. This was also seen in the spatial maps of the model performance presented previous in Figure 6-16 where the CALPUFF model had spatial alignment problems compared with the observed tracer 24 hours after the tracer was released. Thus, CALPUFF does not perform comparably to the other evaluated LRT models even within 300 km of the source.

19 http://www.epa.gov/ttn/scram/7thmodconf.htm

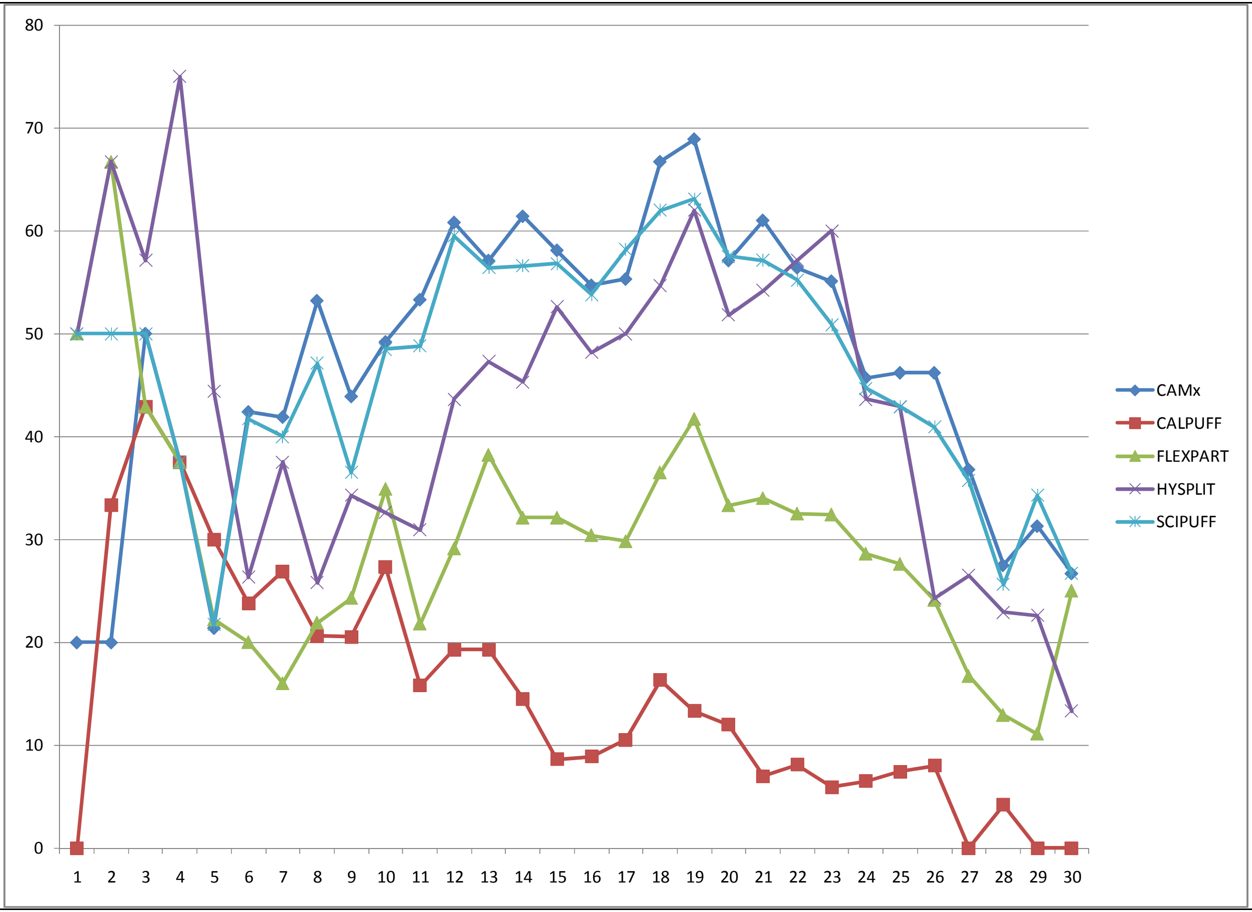


Figure 6-19. Figure of Merit (FMS) spatial model performance statistics as a function of time since the beginning of the tracer release.

6.4.4.2 CALPUFF Puff Splitting Sensitivity Tests

The CALPUFF puff splitting algorithm is controlled by several model options that are defined in the CALPUFF control input file. Two types of puff splitting may be invoked in CALPUFF: (1) vertical puff splitting when vertical wind shear is present across vertical layers in a well-mixed puff; and (2) horizontal puff splitting when there is sufficient horizontal wind shear across the horizontal extent of the puff.

The MSPLIT control option turns on puff splitting when set to 1, when MSPLIT is 0 no vertical or horizontal puff splitting is allowed to occur.

Four criteria must occur in order for vertical puff splitting to occur in CALPUFF:

1. The puff must be in contact with the ground.
2. The puff splitting flag must be turned on (i.e., IRESPLIT = 1).
3. The previous hours mixing height must be above a certain height (mixing height > ZISPLIT).
4. The ratio of the last hours mixing height to the maximum mixing height encountered by the puff is less than a maximum value (current mixing height/maximum mixing height > ROLDMAX).

The puff splitting flag (item 2) is turned on using the IRESPLIT input option. IRESPLIT consists of 24 values corresponding to the hour of the day with values that are either 0 or 1, where 1 turns on the puff splitting flag for puffs. Once the puff splitting flag is turned on, it remains on until the puff splitting occurs at which point the puff splitting flagged is turned off until it is turned back on again by IRESPLIT. The default setting for IRESPLIT is to have all hours zero except for setting hour 17 to 1. The reasoning behind this is to invoke puff splitting in the evening when a nocturnal inversion occurs and there is a decoupling of the winds between the nocturnal inversion layer and above the nocturnal inversion (i.e., the residual “mixed layer”). Setting IRESPLIT to all zeros will result in the puffs never performing vertical puff splitting and setting IRESPLIT to all ones will result in the puff splitting flag always turned on and puffs will always split when the other three criteria for vertical puff splitting are met.

The default value for the previous hours minimum mixing height value (item 3) is ZISPLIT = 100 m. This minimum value is used to assure that the current mixing height is not negligible.

The ratio of the previous hours mixing height to maximum mixing height encountered by the puff (item 4) is controlled by the ROLDMAX parameter with a default value of 0.25.

When vertical puff splitting occurs in CALPUFF, the number of puffs that the puff is split into is controlled by the NSPLIT parameter that has a default value of 3.

Horizontal puff splitting occurs when the puff concentrations are above a minimum value (CNSPLITH), the puff has a minimum width that is defined by its sigma-y in grid cell units (SYSPLITH) and the minim puff elongation rate (SYSPLITH per hour) is above a SHSPLITH factor. The default minimum concentration is CNSPLITH = 10^{-7} g/m^3 (0.1 μg/m^3). Default SYSPLITH value is 1.0 and default SHSPLITH factor is 2.0. When horizontal puff splitting occurs in CALPUFF the number of puffs the puff is split into is controlled by the NSPLITH parameter that has a default of 5.

Eight CALPUFF puff splitting sensitivity tests were conducted, which are defined in Table 6-3. When vertical and horizontal puff splitting occurs in CALPUFF, the default number of puffs to split into was used in the CALPUFF sensitivity tests (i.e., NSPLIT = 3 and NSPLITH = 5). The NOSPLIT sensitivity test set MSPLIT = 0 so no vertical or horizontal puff splitting was allowed to occur. The DEFAULT puff splitting turned on puff splitting (MSPLIT = 1) but only turned on the vertical puff splitting flag at hour 17 every day. Whereas, the ALLHRS sensitivity test made sure that the vertical puff splitting flag was turned on all the time (i.e., IRESPLIT = 24*1) removing criteria 2 from the vertical puff splitting requirement. The ZISPLIT sensitivity test set ZISPLIT to zero thereby removing criteria 3 in the vertical puff splitting, as well as requirement 2 (like ALLHRS). ROLD relaxed the minimum ratio of the previous hours to maximum mixing height for vertical puff splitting from 0.25 to 0.50. The SYS sensitivity test allows horizontal puff splitting to occur more frequently by allowing puff splitting to occur with a puff sigma-y value is greater than SYSPLITH values of 0.1 (2.6 km) versus the default 1.0 (36 km) value. The last sensitivity test combines the ROLD and SYS sensitivity tests.

Table 6-3. Summary of CALPUFF puff splitting sensitivity tests performed using the ETEX database.

Sensitivity Test	MSPLIT	NSPLIT	IRESPLT	ZISPLIT	ROLDMAX	NSPLITH	SYSPLITH	CNSPLITH
NOSPLIT	0	NA	NA	NA	NA	NA	NA	NA
DEFAULT	1	3	Hr 17=1	100	0.25	5	1.0	10^{-7}
ALLHRS	1	3	24*1	100	0.25	5	1.0	10^{-7}
CNSMIN	1	3	24*1	100	0.25	5	1.0	10^{-20}
ZISPLIT	1	3	24*1	0	0.25	5	1.0	10^{-20}
ROLD	1	3	24*1	0	0.50	5	1.0	10^{-20}
SYS	1	3	24*1	0	0.25	5	0.1	10^{-20}
SYSROLD	1	3	24*1	0	0.50	5	0.1	10^{-20}

Figure 6-20 displays the spatial model performance statistics for the CALPUFF puff splitting sensitivity tests. The DEFAULT, ALLHRS and CNSMIN CALPUFF sensitivity tests obtained the exactly same model performance statistics indicating that CALPUFF model performance was not affected by the IRESPLT and CNSMIN puff splitting parameters. There are some small difference in the spatial model performance statistics for the other CALPUFF puff splitting sensitivity tests with the ROLD parameter having the biggest effect when changed from 0.25 to 0.50 that improved model performance a couple of percentage points for the FMS, POD and TS spatial statistics but degraded the FAR spatial statistic by several percentage points.

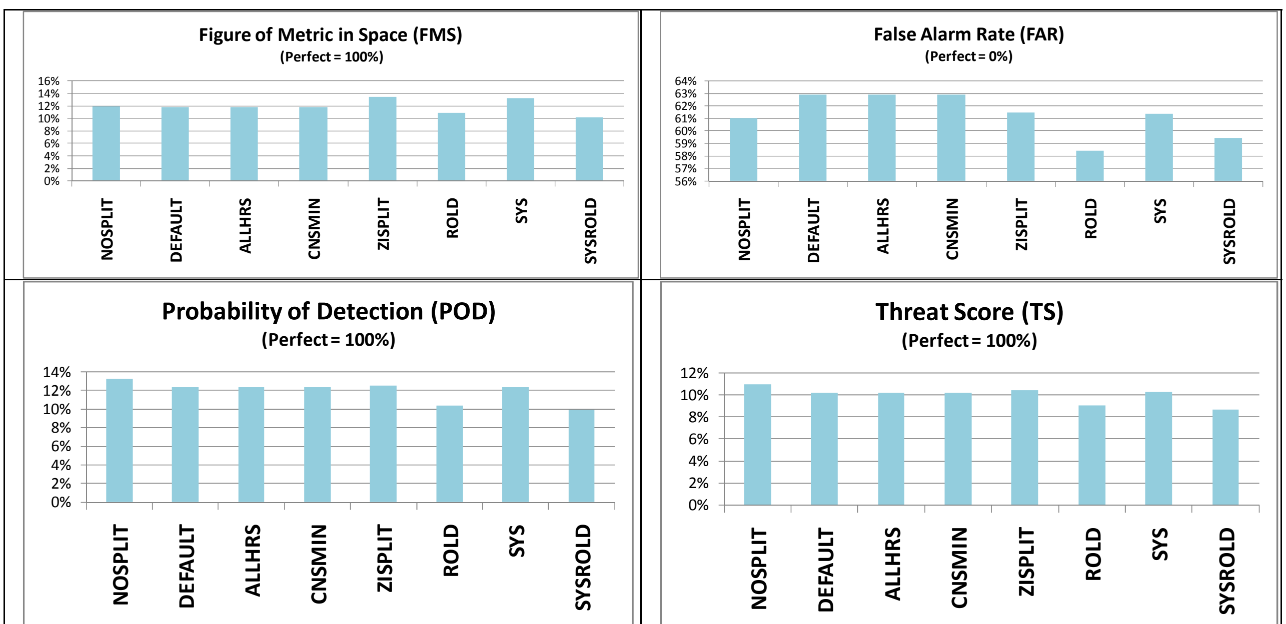


Figure 6-20. Spatial model performance statistics for the CALPUFF puff splitting sensitivity tests.

The global model statistics for the CALPUFF puff splitting sensitivity tests are shown in Figure 6-20, with Figures 6-21a and 6-21b displays statistics where the best performing model configuration has the lowest and highest score, respectively. The puff splitting sensitivity tests have a very small effect on the CALPUFF model performance. Again, the biggest effect on CALPUFF performance of all the puff splitting parameters comes from changing ROLD from 0.25 to 0.50, which appears to slightly degrade most CALPUFF model performance metrics with the exception of bias and error that are improved. Again, in terms of the CALPUFF global model performance versus other four LRT dispersion models (Figures 6-9 through 6-15), the CALPUFF puff splitting sensitivity tests are exhibiting by far the worst model performance. For example, the RANK model performance statistic varies from 0.6 to 0.7 across the CALPUFF puff splitting sensitivity tests as compared to much higher values for CAMx (1.9), SCIPUFF (1.8), HYPLIT (1.8) and FLEXPART (1.0).

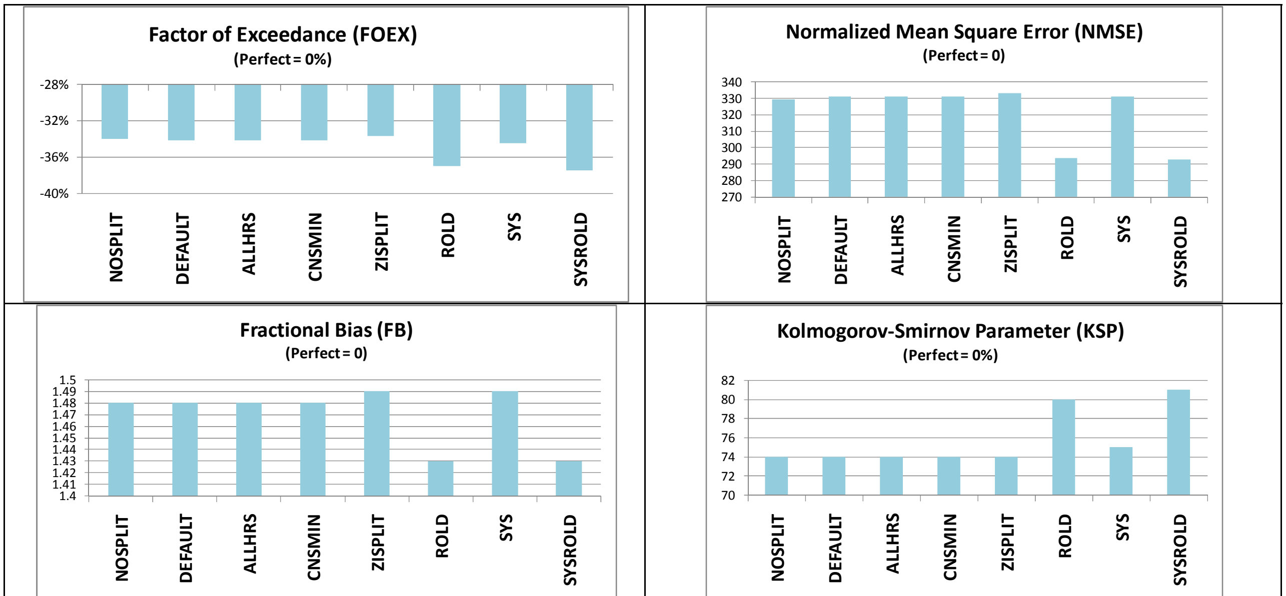


Figure 6-21a. Global model performance statistics for the CALPUFF puff splitting sensitivity tests.

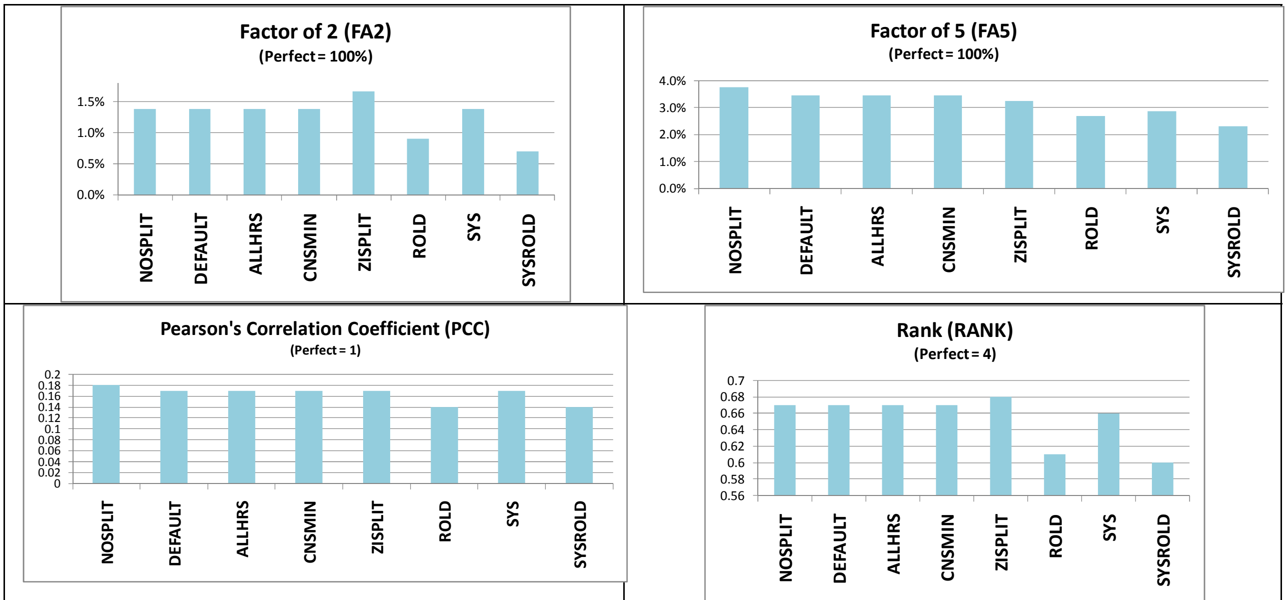


Figure 6-21b. Global model performance statistics for the CALPUFF puff splitting sensitivity tests.

In conclusion, the CALPUFF puff splitting sensitivity tests did not have any significant effect on CALPUFF model performance. Whether puff splitting was used or not produced essentially identical model performance for the ETEX experiment and certainly did not improve the CALPUFF model performance.

6.4.5 HYSPLIT Sensitivity Tests

HYSPLIT is unique among the models analyzed in this project in that its configuration is highly flexible, allowing for treatment of atmospheric dispersion purely as a Lagrangian particle model (default configuration), puff-particle hybrid model, or purely as a puff model. Nine sensitivity analyses were conducted against the ETEX database to provide information about the various configurations of HYSPLIT, but more importantly to provide additional information regarding the two distinct classes (puff and particle) of Lagrangian models evaluated as part of this project. Model configuration (puff, particle, puff-particle hybrid) are governed through the HYSPLIT parameter INITD. A description of the INITD variable options is provided in Table 6-4.

Model configuration options for the nine sensitivity runs are detailed in Table 6-5. In general, model control options were held to default values with two notable exceptions, the INITD and NUMPAR variables. HYSPLIT performance is highly sensitive to the number of particles released in the simulation. The HYSPLIT parameter NUMPAR controls the number of particles released over the duration of the emissions release. The default value for NUMPAR is set to 2500, but the user must take caution to insure that a sufficient number of particles are released to provide a "smooth temporal change" in concentration fields (NOAA, 2009). The original NOAA configuration for HYSPLIT was for INITD = 104 that is a particle/puff hybrid configuration (3D part – THh-Pv) with NUMPAR set to 1500. Original sensitivity runs found that the concentration fields were spotty; therefore, NUMPAR was set to 10000 to provide for smoother temporal evolution of the concentration fields.

Table 6-4. HYSPLIT INITD options and descriptions.

INITD Value	Description
0 (Default)	3D Particle Horizontal and Vertical
1	Gaussian horizontal and top-hat vertical puff (Gh-THv)
2	Top-hat horizontal and vertical puff (THh-THv)
3	Gaussian horizontal puff and vertical particle distribution (Gh-Pv)
4	Top-hat horizontal puff and vertical particle distribution (THh-Pv)
103	3D particle (#0) converts to Gh-Pv (#3)
104	3D particle (#0) converts to THh-Pv (#4)
130	Gh-Pv (#3) converts to 3D particle (#0)
140	THh-Pv (#4) converts to 3D particle (#0)

Table 6-5. HYSPLIT sensitivity runs and relevant configuration parameters.

Sensitivity Test	INITD	NUMPAR	ISOT	KSPL	FRHS	FRVS	FRTS	FRME
INITD0	0	10000	1	NA	NA	NA	NA	NA
INITD1	1	10000	1	1	1.0	0.01	0.10	0.10
INITD2	2	10000	1	1	1.0	0.01	0.10	0.10
INITD3	3	10000	1	1	1.0	0.01	0.10	0.10
INITD4	4	10000	1	1	1.0	0.01	0.10	0.10
INITD103	103	10000	1	1	1.0	0.01	0.10	0.10
INITD104	104	10000	1	1	1.0	0.01	0.10	0.10
INITD130	130	10000	1	1	1.0	0.01	0.10	0.10
INITD140	140	10000	1	1	1.0	0.01	0.10	0.10

Figure 6-22 displays the spatial model performance statistics for the HYSPLIT INITD sensitivity tests. Wide variation in spatial performance is noted across the nine runs and requires closer examination. For example, the two puff based configurations (INITD1 and INITD2) showed the poorest spatial performance of all of the runs with low POD and TS values and much higher FAR values compared to all other configurations. The 3D particle based configuration (INITD0) had higher POD, TS, and lower FAR in comparison, yet it had a comparably low FMS to INITD1 and INITD2. Since the FMS score examines all model/observed values greater than 0 and the additional spatial metrics use a contingency level of 100 pg m^{-3}, it can be interpreted that the 3D particle configuration performed significantly better at concentration levels above 100 pg m^{-3}, but its spatial performance degraded with concentration ranges below the contingency level. The puff-particle hybrid configurations (INITD3, INTID4, INITD103, INITD104) performed consistently better overall across all four spatial metrics.

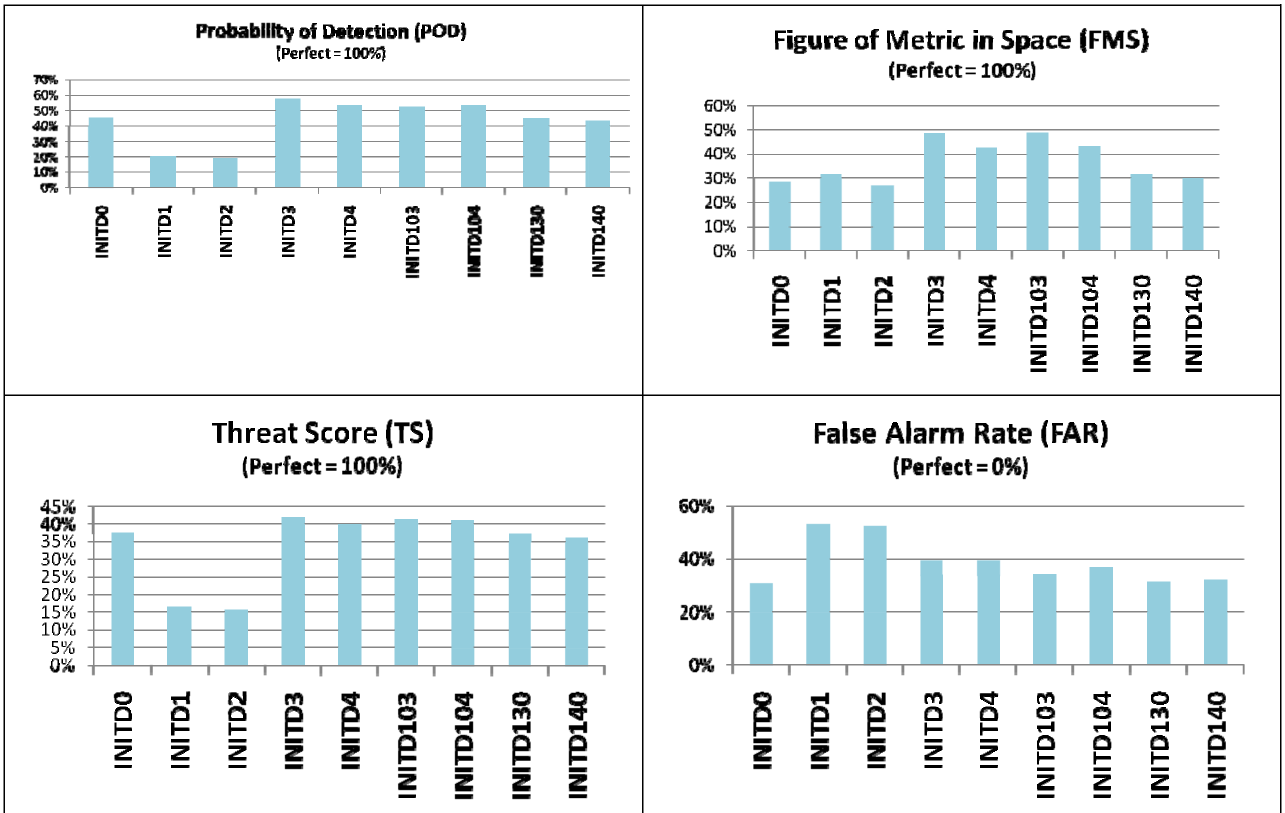


Figure 6-22. Spatial model performance statistics for nine HYSPLIT INITD sensitiviity experiments.

Figure 6-23 displays the global statistics for the HYSPLIT sensitivity tests with Figures 6-23a and 6-23b containing the statistical metrics where the best performing model has the, respectively, lowest and highest score. For the FOEX metrics, the INITD3 scores the best with a 3.4% FOEX score followed by INITD4 (-7%), INITD104 (-8.6%), and finally INITD103 (-9.5%). INITD2 scored worst with a -28.9%. INITD0, 1, 130, and 140 performed nearly as poorly with scores ranging between -21.9% to -24%. Using the NMSE statistical performance metric, the best performing configuration was INITD130, 140, and 3 with values of 17, 18, 19 pg m^{-3} respectively. The model configurations with the highest predicted error were INITD1 and INITD2 with values of approximately 325 and 333 pg m^{-3}. For the KS parameter, the four puff-particle model configuration options (INITD3,4,103,104) again showed the best scores.

For the within a factor of 2 and 5 metric (FA2 and FA5, Figure 6-23b, top), the hybrid puff-particle configurations INITD3 and INITD4 and their counterpart particle-puff configurations INITD103 and INITD104 are clearly performing better than pure particle (INITD0) or puff (INITD1 and INITD2) configurations. For the PCC metric, INITD140 had the highest (0.69) followed by INITD104 (0.64) and INITD0 and 103 (0.63). Interestingly, it appears that the higher PCC score for INITD103 is the main reason for the highest overall model RANK as both INITD3 and 103 had nearly identical spatial performance while INITD3 had slightly better KS scores.

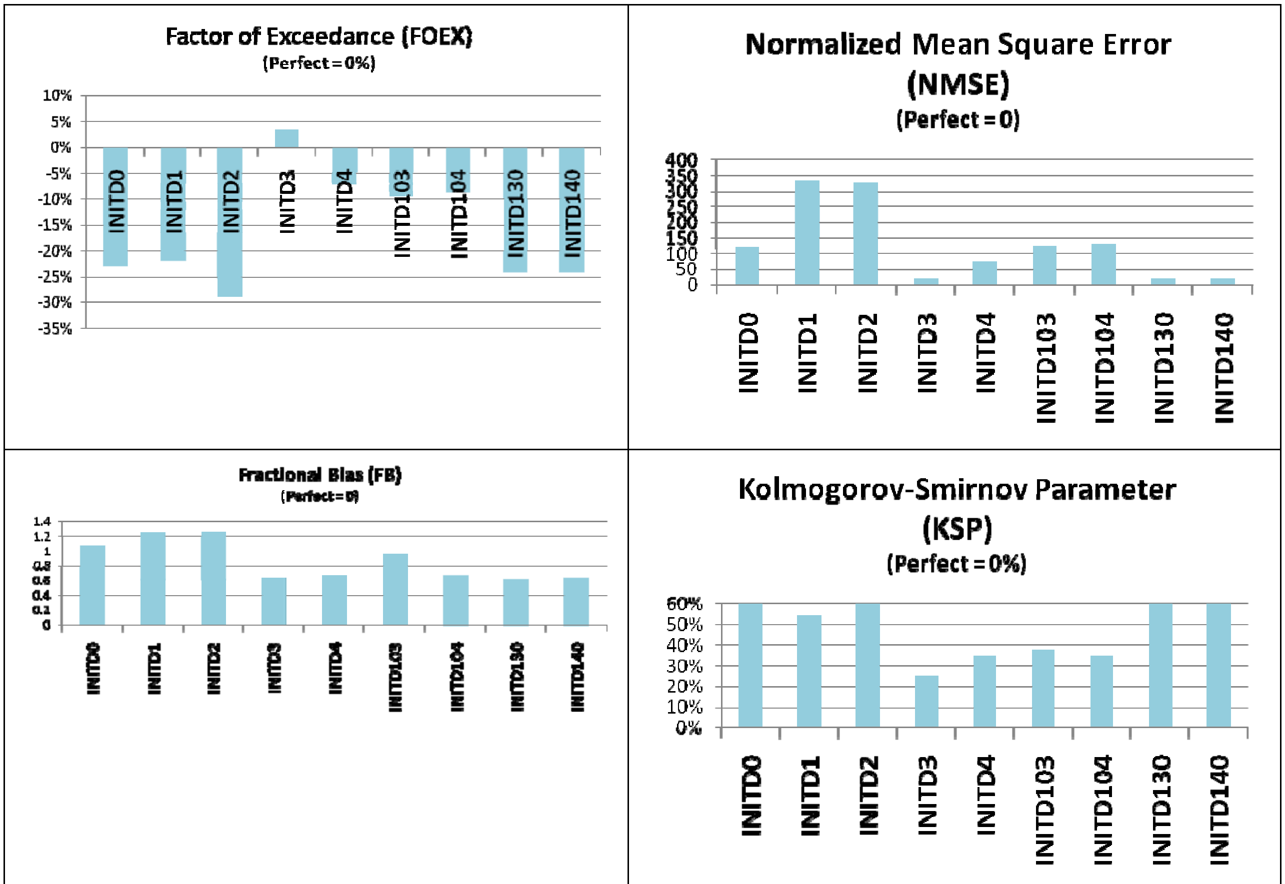


Figure 6-23a. Global model performance statistics for nine HYSPLIT INITD sensitivity tests.

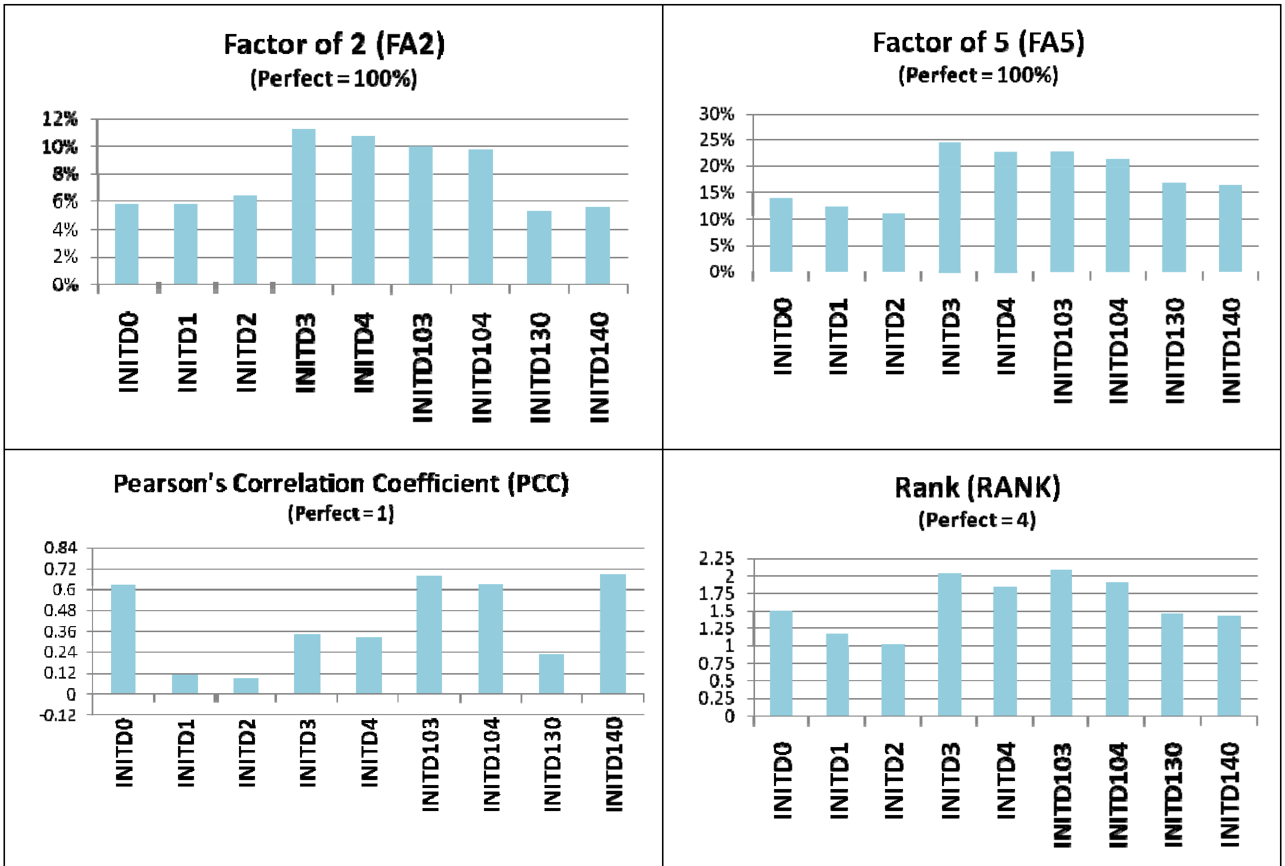


Figure 6-23b. Global model performance statistics for nine HYSPLIT INITD sensitivity tests.

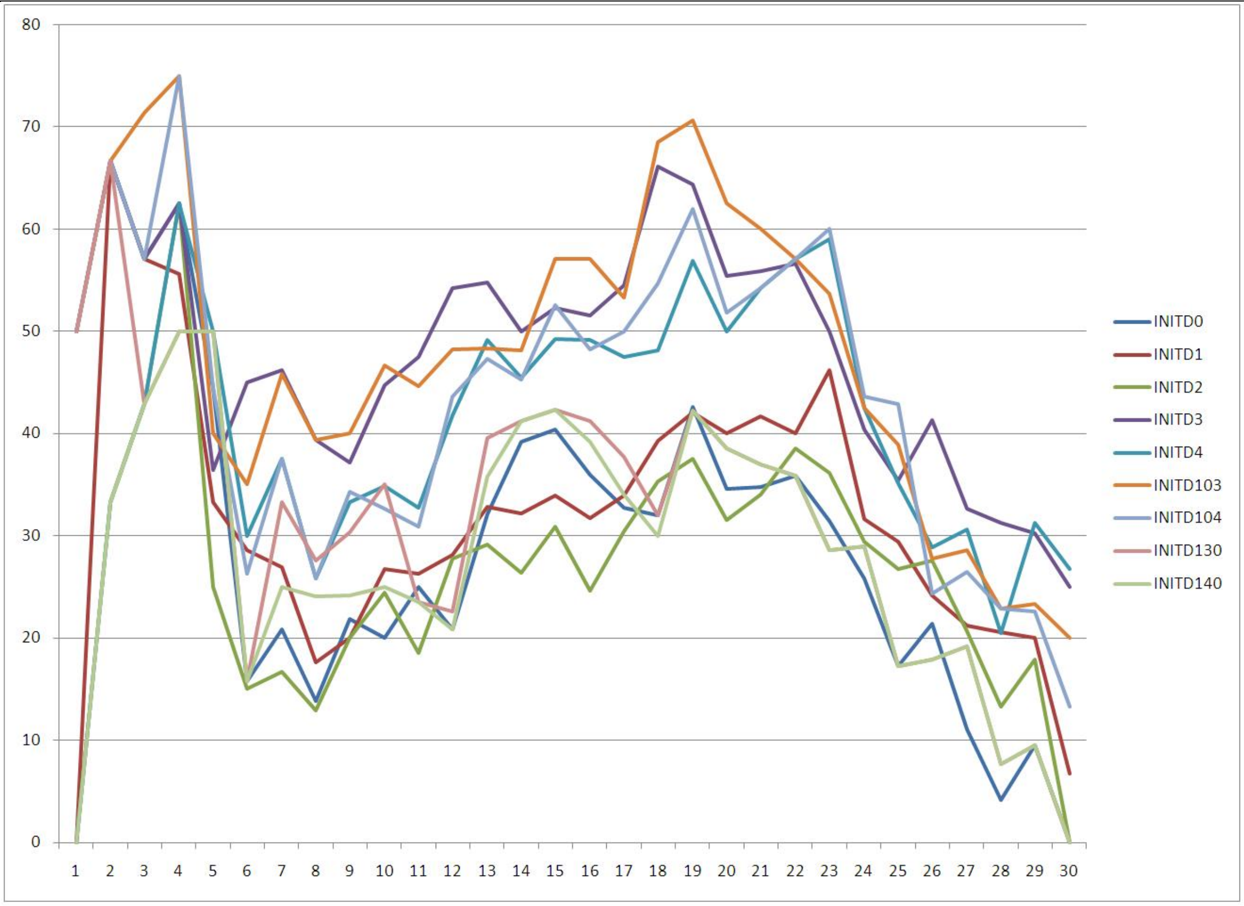


Figure 6-24. Figure of Merit (FMS) spatial model performance statistics as a function of time since the beginning of the tracer release for HYSPLIT INITD sensitivity analyses.

The final panel in Figure 6-23b (bottom right) displays the overall RANK statistic. The RANK statistics orders the model performance of the HYSPLIT INITD configurations are as follows:

1. INITD103 (2.09)
2. INITD3 (2.03)
3. INITD104 (1.91)
4. INITD4 (1.85)
5. INITD0 (1.50)
6. INITD130 (1,47)
7. INITD140 (1.44)
8. INITD1 (1.16)
9. INITD2 (1.01)

Based on this analysis the puff-particle and particle-puff hybrid configurations of the HYSPLIT system are clearly the best performing, indicating a distinct operational advantage over pure puff or particle configurations.

6.5 CONCLUSIONS OF THE MODEL PERFORMANCE EVALUATION OF THE LRT DISPERSION MODELS USING THE ETEX TRACER EXPERIMENT FIELD STUDY DATA

The evaluation of the five LRT dispersion models using a common MM5 dataset and the ETEX database has provided interesting results about the current capability of LRT models to reproduced observed tracer concentrations. Four of the five LRT models were able to reproduce the observed tracer bifurcation at the farther downwind distances. The CALPUFF model was unable reproduce the observed bifurcation of the tracer cloud and kept the estimated tracer cloud in a circular Gaussian distribution that was advected too far north. CALPUFF puff splitting sensitivity tests were performed to determine whether it would help simulate the bifurcation of the tracer cloud but puff splitting had little effect on the CALPUFF predictions.

CAMx sensitivity tests were conducted to examine vertical mixing and horizontal advection solvers and the best performing CAMx model configuration was the one that is most frequently used in applications, which includes using the CMAQ-like vertical diffusion coefficients in MM5CAMx and the PPM advection solver. The vertical diffusion algorithm had a much bigger effect on CAMx model performance than the choice of horizontal advection solver.

The HYSPLIT sensitivity tests with different particle-puff variations resulted in a wide range of model performance with RANK scores that varied from 1.01 to 2.09.

7.0 REFERENCES

Anderson, B. 2008. The USEPA MM5CALPUFF Software Project. 12th Annual George Mason University Conference on Atmospheric Transport and Dispersion Modeling, Fairfax, VA, July 8-10, 2008.

Anthes, R.A. and T.T. Warner. 1978. Development of Hydrodynamic Models Suitable for Air Pollution and other Mesometeorological Studies. Mon. Wea. Rev., 106, 1045-1078. 1078. (http://journals.ametsoc.org/doi/abs/10.1175/1520-0493(1978)106%3C1045%3ADOHMSF%3E2.0.CO%3B2).

Bott, A. 1989. A Positive Definite Advection Scheme Obtained by Nonlinear Renormalization of the Advective Fluxes. Mon. Wea. Rev., **117**, 1006-1015.

Boybeyi, Z., N. Ahmad, D. Bacon, T. Dunn, M. Hall, P. Lee, R. Sarma, and T. Wait. 2001. Evaluation of the Operational Multiscale Environment Model with Grid Adaptivity against the European Tracer Experiment. J. App. Meteor., **40**, 1541-1558.

Brashers, B. and C. Emery. 2011. The Mesoscale Model Interface Program (MMIF) Version 2.0 – Draft User's Manual. Prepared by ENVIRON International Corporation Novato California and Lynnwood Washington. EPA Contract No. EP-D-07-102, Work Assignments 2-06 and 4-06. September 30.

Brashers, B. and C. Emery. 2012. The Mesoscale Model Interface Program (MMIF) Version 2.1, 2012-01-31 – Draft User's Manual. Prepared by ENVIRON International Corporation Novato California and Lynnwood Washington. EPA Contract No. EP-D-07-102, Work Assignments 2-06, 4-06 and 5-08. January 31.

Byun, D.W., and J.K.S. Ching. 1999. "Science Algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) Modeling System", EPA/600/R-99/030.

Carhart, R.A., A. Policastro, M. Watag and L. Coke. 1989. Evaluation of Eight Short-Term Long-Range Transport Models using Field Data. Atmos. Env., 23, 85-105.

Chang, J.C., K. Chayantrakom, and S.R. Hanna. 2003. Evaluation of CALPUFF, HPAC, and VLSTRACK with Two Mesoscale Field Datasets. J. App. Meteor., **42**, 453-466.

Chang, J. C., S.R. Hanna, Z. Boybeyi, and P. Franzese. 2005. Use of Salt Lake City URBAN 2000 Field Data to Evaluate the Urban Hazard Prediction Assessment Capability (HPAC) Dispersion Model. J. App. Meteor., **44**, 485 – 501.

Chen, F. and J. Dudhia. 2001. Coupling an advanced land-surface/hydrology model with the Penn State/NCAR MM5 modeling system. Part I: Model implementation and sensitivity. Mon. Wea. Rev., **129**, 569 – 585.

Colella, P., and P.R. Woodward. 1984. The Piecewise Parabolic Method (PPM) for Gas-dynamical Simulations. J. Comp. Phys., **54**, 174-201.

D'Amours, R. 1998. Modeling the ETEX Plume Dispersion with the Canadian Emergency Response Model. Atmos. Environ., **32**, 4335 – 4341. 34

Deng, A. N.L. Seaman, G.K. Hunter, and D.R. Stauffer, 2004: Evaluation of Interregional Transport Using the MM5-SCIPUFF System. J. App. Meteor., **43**, 1864-1885.

Deng, A. N. and D.R. Stauffer. 2006. On Improving 4-km Mesoscale Model Simulations. J. App. Meteor., **45**, 361-381. Director General Joint Research Centre, cited 2008: European Tracer Experiment: Overview of the ETEX Project. (http://rem.jrc.europa.eu/etex).

DOE. 1978. Heavy Mathane-SF6 Tracer test Conducted at the Savannah River Laboratory, December 10, 1975. DP-1469. U.S. Department of Energy. Prepared by E.I. du Pont de Nemours and Company, Savannah River Laboratory, Aiken, South Carolina.

Draxler, R.R., and B.J.B. Stunder. 1988. Modeling the CAPTEX vertical tracer concentration profiles. J. Appl. Meteorol., 27:617-625.

Draxler, R.R. and J.L. Heffter. 1989. Across North America Tracer Experiment (ANATEX). Volume I: Description, Ground-Level Sampling at Primary Sites, and Meteorology. January. NOAA Tech Memo ERL ARL-142.

Draxler, R.R. and G.D. Hess. 1997. Description of the HYSPLIT_4 Modeling System. NOAA Technical Memorandum ERL ARL-224. National Oceanic and Atmospheric Administration, Air Resources Laboratory, Silver Springs, MD. Revised August 1998, September 2002, October 2003 and January 2004. (http://www.arl.noaa.gov/documents/reports/arl-224.pdf).

Draxler, R.R., J.L. Heffter, and G.D. Rolph. 2002. DATEM: Data Archive of Tracer Experiments and Meteorology. National Oceanic and Atmospheric Administration, Silver Springs, MD. Last Revised 23 July 2002. (http://www.arl.noaa.gov/documents/datem/datem.pdf).

DTRA, 2001: Hazard prediction and assessment capability (HPAC), User Guide Version 4.0.1. Prepared for Defense Threat Reduction Agency, Contract DSWA01-98-C-0110, by Science Applications International Corporation, Rep. HPAC-UG-01-U-R0C0, 631 pp.

Dudhia, J. 1989. Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. J. Atmos. Sci., **46**, 3077-3107. Draxler, R.R., and G.D. Hess, 1997: Description of the Hysplit_4 modeling system, Tech. Rep. NOAA Tech Memo ERL ARL-224, National Oceanic and Atmospheric Administration, Silver Springs, MD, 24 pp.

Dudhia, J. 1993. A Non-hydrostatic Version of the Penn State/NCAR Mesoscale Model: Validation Tests and Simulation of an Atlantic Cyclone and Cold Front. Mon. Wea. Rev., Vol. 121. pp. 1493-1513.

Earth Tech. 2005. CALGRID (Version 2) User Instructions. Final report prepared for the Ozone Transport Commission, Washington D.C. Earth tech, Inc., Concord, MA. May.

Emery, C.A., E. Tai, and G. Yarwood. 2001. Enhanced Meteorological Modeling and Performance Evaluation for Two Texas Ozone Episodes. Prepared for the Texas Natural Resource Conservation Commission, by ENVIRON International Corp., Novato, CA.

Emery, C. and B. Brashers. 2009. The Mesoscale Model Interface Program (MMIF) – Version 1.0. Draft User's Manual. ENVIRON International Corporation, Novato, CA. Prepared for U.S. Environmental Protection Agency. June 11.

ENVIRON. 2010. User's Guide – Comprehensive Air Quality Model with Extensions – Version 5.30. ENVIRON International Corporation, Novato, CA http://www.camx.com. May.

EPA, 1984: Interim Procedures for Evaluating Air Quality Models (Revised). EPA-450/4-84-023, Research Triangle Park, NC.

EPA, 1986a: Evaluation of Short-Term Long-Range Transport Models, Vol. I – Analysis Procedures and Results. EPA-450/4-86-016a, Research Triangle Park, NC.

EPA, 1986b: Evaluation of Short-Term Long-Range Transport Models, Vol. II – Appendices A through E. EPA-450/4-86-016b, Research Triangle Park, NC..

EPA, 1990.: Rocky Mountain Acid Deposition Model Assessment: ARM3 Model Performance Evaluation. EPA/600/S3-90/024, Research Triangle Park, NC. June.

EPA. 1989. Evaluation and Sensitivity Analysis Results of the MESOPUFF II Model with CAPTEX Measurements. United States Environmental Protection Agency, Atmospheric Research and Exposure Assessment Laboratory, Research Triangle Park, NC. July. (EPA/600/S3-89/058).

EPA. 1991. Guidance for Regulatory Application of the Urban Airshed Model (UAM). U.S., Environmental Protection Agency, Office of Air Quality, Planning and Standards, Research Triangle Park, NC. July. (http://www.epa.gov/ttn/scram/guidance/guide/uamreg.pdf).

EPA. 1995. User's Guide for the Industrial Source Complex (ISC) Dispersion Model – Volume I User Instructions. U.S. Environmental Protection Agency, Office of Air Quality, Planning and Standards, Research Triangle Park, NC. September. (http://www.epa.gov/ttn/scram/userg/regmod/isc3v1.pdf).

EPA. 1998a. A Comparison of CALPUFF Modeling Results to Two Tracer Field Experiments. Tech. Rep., EPA-454/R-98-009, Research Triangle Park, NC. (http://www.epa.gov/ttn/scram/7thconf/calpuff/tracer.pdf).

EPA. 1998b. Interagency Workgroup on Air Quality Modeling (IWAQM) Phase 2 Summary Report and Recommendations for Modeling Long Range Transport Impacts. Tech Rep., EPA-454/R-98-009, Research Triangle Park, NC, 160 pp. (http://www.epa.gov/scram001/7thconf/calpuff/phase2.pdf).

EPA. 2003. Revisions to the Guideline on Air Quality Models: Adoption of a Preferred Long Range Transport Model and Other Revisions, Final Rule. 40 CFR Part 51. Federal Register, Vol. 68, 72, Tuesday April 15, 2003. (http://frwebgate3.access.gpo.gov/cgi-bin/PDFgate.cgi?WAISdocID=xoakAq/0/2/0&WAISaction=retrieve).

EPA. 2004: AERMOD: Description of Model Formulation. Tech Rep., EPA-454/R- 03-004, Research Triangle Park, NC, 91 pp. (http://www.epa.gov/scram001/7thconf/aermod/aermod_mfd.pdf).

EPA. 2005. Revisions to the Guideline on Air Quality Models: Adoption of a Preferred General Purpose (Flat and Complex Terrain) Dispersion Model and Other Revisions; Final Rule. Federal Register, Vol. 70, No 216, Wednesday, November 9, 2005. (http://www.epa.gov/ttn/scram/guidance/guide/appw_05.pdf).

EPA. 2007. "Guidance on the Use of Models and Other Analyses for Demonstrating Attainment of Air Quality Goals for Ozone, PM2.5 and Regional Haze". U.S. Environmental Protection Agency, Research Triangle Park, NC. EPA-454/B-07-002. April. (http://www.epa.gov/ttn/scram/guidance/guide/final-03-pm-rh-guidance.pdf).

EPA. 2009a. Reassessment of the Interagency Workgroup on Air Quality Modeling (IWAQM) Phase 2 Summary Report: Revisions to Phase 2 Recommendations. Draft. U.S.

Environmental Protection Agency, Office of Air Quality, Planning and Standards, Air Quality Analysis Division, Air Quality Modeling Group, Research Triangle Park, NC. May 27. (http://www.epa.gov/scram001/guidance/reports/Draft_IWAQM_Reassessment_052709.pdf).

EPA. 2009b. Clarification on EPA-FLM Recommended Settings for CALMET. Memorandum from Tyler J. Fox, Group Leader, Air Quality Modeling Group, Office of Air Quality, Planning and Standards, U.S. Environmental Protection Agency to Regional Modeling Contacts. August 31. (http://www.epa.gov/ttn/scram/CALMET%20CLARIFICATION.pdf).

EPA. 2009c. AERMOD Implementation Guide. AERMOD Implementation Workgroup, U.S. Environmental Protection Agency, Office of Air Quality, Planning and Standards, Air Quality Assessment Division. Last Revised March 19, 2009. (http://www.epa.gov/scram001/7thconf/aermod/aermod_implmtn_guide_19March2009.pdf).

Ferber G.J., K. Telegadas, J.L. Heffter, C.R. Dickson, R.N. Dietz and P.W. Krey. 1981. Demonstration of a Long-Range Atmospheric Tracer System using Perfluorocarbons - Final Report. EPA-600/7-81-006, U.S EPA Office of Research and Development, Washington, D.C., January, 55 p.

Ferber G.J., J.L. Heffter, R.R. Draxler, R.J. Lagomarsino, F.L. Thomas, R.N. Dietz, and C.M. Benkovitz. 1986. Cross-Appalachian tracer experiment (CAPTEX-83). Final Report. NOAA Tech. Memo ERL ARL-142, Air Resources Laboratory, Silver Spring, MD, 60 p.

Fox, D. 1981. Judging Air Quality Model Performance. Bull. Amer. Meteor. Soc., 62, 599-609.

Graziani, G., W. Klug, and S. Mosca. 1998. Real-Time Long-Range Dispersion Model Evaluation of the ETEX First Release, EUR 17754 EN. ISBN 92-828-3657-6.

Grell, G.A., J. Dudhia, and D.R. Stauffer. 1995. A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5), NCAR Tech. Note NCAR/TN-398+STR, National Center for Atmospheric Research, Boulder, CO, 117 pp. (http://www.mmm.ucar.edu/mm5/documents/mm5-desc-pdf).

Heffter, J.L., J.F. Schubert and G.A. Meade. 1984. Atlantic Coast Unique Regional Atmospheric Tracer Experiment. ACCURATE. NOAA Tech Memo ERL ARL-130.

Irwin, J. 1997. A Comparison of CALPUFF Modeling Results with 1977 INEL Field Data Results. Presented at 22nd NATO/CCMS International technical Meeting on Air Pollution Modeling and its Application. Clermont-Ferrand, France. June 2-6, 1997.

Iwasaki, T., T. Maki, and K. Katayama. 1998: Tracer Transport Model at Japan Meteorological Agency and Its Application to the ETEX Data. Atmos. Environ., **32**, 4285 – 4295.

Kain, J.S.. 2004. The Kain-Fritsch convective parameterization: An update. J. App. Meteor., **43**, 170-181. 36

Kang, D., R. Mathur, K. Schere, S. Yu, and B. Eder. 2007. New Categorical Metrics for Air Quality Model Evaluation. J. App. Meteor., **46**, 549-555.

McNally, D. 2010. MMIFstat Evaluation Tool. Alpine Geophysics, LLC, Arvada, Colorado.

Mlawer, E.J., S.J. Taubman, P.D. Brown, M.J. Iacono, and S.A. Clough. 1997. Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the long-wave. J. Geophys. Res., **102** (D14), 16663-16682.

Mosca, S., G. Graziani, W. Klug, R. Ballasio, and R. Bianconi. 1998. A Statistical Methodology for the Evaluation of Long-Range Dispersion Models: An Application to the ETEX Exercise. Atmos. Environ., **32**, 4307- 4324.

National Center for Atmospheric Research, cited 2008: NCEP/NCAR Reanalysis – Data Product Description. (http://dss.ucar.edu/pub/reanalysis/prod_des.html).

Nasstrom, J.S., and J.C. Pace. 1998. Evaluation of the Effect of Meteorological Data Resolution on Lagrangian Particle Dispersion Simulations Using the ETEX Experiment. Atmos. Environ., **32**, 4187-4194.

Policastro, A.J., M. Wastag, L. Coke, R.A. Carhart and W.E. Dunn. 1986. Evaluation of Short-Term Long-Range Transport Models – Volume I. Analysis Procedures and Results. Argonne National Laboratory. Prepared for U.S. Environmental Protection Agency, Office of Air Quality, Planning and Standards. EPA-450/4-86-016a. October.

Scire, J.S., R. Yamartino, G.R. Carmichael, Y.S. Chang. 1989. CALGRID: A Mesoscale Photochemical Grid Model. Tech. Rep., Sigma Research Corporation, Concord, MA, 222 pp. (http://gate1.baaqmd.gov/pdf/2516_CALGRID_Mesoscale_Photochemical_Grid_Model_Volume_II_Users_Guide_Part1_1989.pdf)

Scire, J.S., F.R. Robe, M.E. Fernau, and R.J. Yamartino. 2000a. A User's Guide for the CALMET Meteorological Model (Version 5). Tech. Rep., Earth Tech, Inc., Concord, MA 332 pp. (http://www.src.com/calpuff/download/CALMET_UsersGuide.pdf).

Scire, J.S., D.G. Strimaitis, and R.J. Yamartino. 2000b. A User's Guide for the CALPUFF Dispersion Model (Version 5), Tech. Rep., Earth Tech, Inc., Concord, MA, 521 pp. (http://www.src.com/calpuff/download/CALPUFF_UsersGuide.pdf).

Seaman, N.L. 2000. Meteorological modeling for air quality assessments. Atmos. Environ., Vol. 34, No. 12-14, 2231-2260.

Seaman, N.L., D.R. Stauffer, and L.M. Lario. 1995. A multiscale four-dimensional data assimilation system applied to the San Joaquin Valley during SARMAP. Part I: Modeling design and basic performance characteristics. J. Appl. Meteo., Vol. 34, pp. 1739-1761.

Skamarock, W.C., J.B. Klemp, J. Dudhia, D.O. Gill, D.M. Barker, M.G. Duda, X. Huang, W. Wang, and J.G. Powers. 2008. A Description of the Advanced Research WRF Version 3, NCAR Tech. Note NCAR/TN-475+STR, National Center for Atmospheric Research, Boulder, CO, 125 pp. (http://www.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf).

Slade, D.H. 1968. Meteorology and Atomic Energy. Air Resources Laboratories, Research Laboratories, Environmental Sciences Services Administration, United States Department of Commerce. For the Division of Reactor Development and technology, United States Atomic Energy Commission, Office of Information Services. Silver Spring, Maryland (TID-24190). July.

Stauffer, D.R., and N.L. Seaman. 1990. Use of four-dimensional data assimilation in a limited-area mesoscale model. Part I: Experiments with synoptic-scale data. Mon. Wea. Rev., **118**, 1250 – 1277.

Stauffer, D.R., N.L. Seaman, and F.S. Binkowski. 1991. Use of four-dimensional assimilation in a limited-area mesoscale model. Part II: Effects of data assimilation within the planetary boundary layer. Mon. Wea. Rev., **119**, 734 – 754.

Stohl, A., M. Hittenberger, and G. Wotawa. 1998. Validation of the Lagrangian Particle Dispersion Model FLEXPART Against Large-Scale Tracer Experiment Data. Atmos. Environ., **32**, 4245 – 4264. 39.

Stohl, A., H. Sodemann, S. Eckhardt, A. Frank, P. Siebert and G. Wotawa. 2010. The Lagrangian Particle Dispersion Model FLEXPART Version 8.2. Norwegian Institute of Air Research, Kjeller, Norway. (http://zardoz.nilu.no/~flexpart/flexpart/flexpart82.pdf).

Stohl, A., C. Forster, A. Frank, P. Seibert, and G. Wotawa. 2005. Technical Note: The Lagrangian particle dispersion model FLEXPART version 6.2. Atmos. Chem. Phys. **5**, 2461 – 2474.

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 Research and Technology Division, Titan Corp., Princeton, NJ, 172 pp. (http://www.titan.com/appliedtech/Pages/TRT/pages/scipuff/scipuff_files.htm).

Telegadas, K., G.J. Ferber, R.R. Draxler, M.M Penderdast, A.L. Boni, J.P. Hughes and J. Gray. 1980. Measured Weekly and Twice-Daily Krypton-85 Surface Air Concentrations within 150 km of the Savannah River Plant (March 1975 through September 1977) – Final Report. NOAA Technical Memorandum ERL ARL-80. United States Department of Commerce, National Oceanic and Atmospheric Administration, Environmental Research Laboratories, Air Resources Laboratories, Silver Spring, Maryland. January.

Wendum, D., 1998: Three long-range transport models compared the ETEX experiment: A performance study. Atmos. Environ., **32**, 4297-4305.

Van Dop, H., R. Addis, G. Fraser, F. Girardi, G. Graziani, Y. Inoue, N. Kelly, W. Klug, A. Kulmala, K. Nodop, and J. Pretel. 1998. ETEX: A European Tracer Experiment: Observations, Dispersion Modeling and Emergency Response. Atmos. Environ., **32**, 4089 – 4094.

Xiu, A. and J.E. Pleim. 2000. Development of a land surface model. Part I: application in a mesoscale meteorology model. J. App. Met., 40, pp. 192-209.

Yamartino, R., J.S. Scire, S.R. Hanna, G.R. Carmichael, and Y.S. Chang. 1989. CALGRID: A Mesoscale Photochemical Grid Model. Volume I: Model Formulation Document, Tech. Rep., Sigma Research Corporation, Concord, MA, 81 pp.

Appendix A

Evaluation of the MM5 and CALMET Meteorological Models Using the CAPTEX CTEX5 Field Experiment Data

EVALUATION OF THE MM5 AND CALMET METEOROLOGICAL MODELS USING THE CAPTEX CTEX5 FIELD EXPERIMENT DATA

Statistical evaluation of the prognostic (MM5) and diagnostic (CALMET) meteorological model applications for the CTEX5 CAPTEX release was conducted using surface meteorological measurements. For the MM5 datasets, performance for meteorological parameters of wind (speed and direction), temperature, and humidity (mixing ratio) was examined. For the CALMET experiments, CALMET estimated winds (speed and direction) were examined because the two-dimensional temperature and relative humidity fields output are simple interpolated fields of the observations. Therefore, the evaluation for CALMET was restricted to winds where the majority of change can be induced by both diagnostic terrain adjustments and varying the OA strategy. Note that except for the NOOBS = 2 CALMET sensitivity tests (i.e., the "D" series of CALMET sensitivity tests), surface meteorological observations are blended with the wind fields in the CALMET STEP2 objective analysis (OA) procedure. Thus, the evaluation of the CALMET wind fields is not a true independent evaluation as the surface meteorological observations used in the evaluation are also used as input into CALMET. So we expect the CALMET wind fields to compare better with observations than MM5, but that does not mean that CALMET is producing better meteorological fields. As clearly shown by EPA (2009a,b), the CALMET diagnostic (STEP1) and blending of observations using the STEP2 OA procedure can introduce discontinuities and artifacts in the wind fields generated by the MM5/WRF prognostic meteorological model that is used as input to CALMET, even though the CALMET winds may match the observed surface winds at the locations of the monitoring sites does not necessarily mean that CALMET is performing better than MM5/WRF.

The METSTAT software (Emery et al., 2001) was used to match MM5 output with observation data. The MMIFStat software (McNally, 2010) tool was used to match CALMET output with observation data. Emery and co-workers (2001) have developed a set of "benchmarks" for comparing prognostic meteorological model performance statistics metrics. These benchmarks were developed after examining the performance of the MM5 and RAMS prognostic meteorological models for over 30 applications. The purpose of the benchmarks is not to assign a passing or failing grade, rather it is to put the prognostic meteorological model performance in context. The surface meteorological model performance benchmarks from Emery et al., (2001) are displayed in Table A-1. Note that the wind speed RMSE benchmark was also used for wind speed MNGE given the similarity of the RMSE and MNGE performance statistics. These benchmarks are not applicable for diagnostic model evaluations.

Table A-1. Wind speed and wind direction benchmarks used to help judge the performance of prognostic meteorological models (Source: Emery et al., 2001).

Wind Speed	Root Mean Squared Error (RMSE) Mean Normalized Bias (NMB) Index of Agreement (IOA)	≤ 2.0 m/s ≤ ±0.5 m/s ≥ 0.6
Wind Direction	Mean Normalized Gross Error (MNGE) Mean Normalized Bias (MNB)	≤ 30° ≤ ±10°
Temperature	Mean Normalized Gross Error (MNGE) Mean Normalized Bias (NMB) Index of Agreement (IOA)	≤ 2.0 K ≤ ±0.5 m/s ≥ 0.8
Humidity	Mean Normalized Gross Error (MNGE) Mean Normalized Bias (NMB) Index of Agreement (IOA)	≤ 2.0 g/kg ≤ ±1.0 g/kg ≥ 0.6

Table A-2 lists the CTEX5 MM5 sensitivity tests that are evaluated in this section. For the first set of MM5 experiments (EXP1) MM5 was configured as it would be run during the late 1980s and early 1990s using only 16 vertical layers, a single 80 km grid resolution and older (Blackadar) planetary boundary layer (PBL) and land soil module (LSM). There were several four dimensional data assimilation (FDDA) experiments using this first MM5 configuration from none (EXP1A) to analysis nudging above the PBL and at the surface (EXP1C).

The second set of MM5 experiments (EXP2A-C) used a more recent MRF PBL scheme and 33 vertical layers with three levels of grid nesting (108/26/12 km) and was meant to represent the way MM5 was run in the late 1990s/early 2000s. Three different levels of FDDA were used with this MM5 configuration: none (EXP2A), analysis nudging above the PBL (EXP2B) and analysis nudging above the PBL as well as at the surface (EXP2C). Note that additional sensitivity experiments were planned using this second MM5 configuration (e.g., EXP2D and EXP2E), but the MM5 model performance using the MRF PBL scheme was so poor that this MM5 configuration was abandoned.

The third set of MM5 experiments (EXP2F-J) used a MM5 configuration similar to the second set of MM5 experiments only with more vertical layers (43) and going back to the Blackadar PBL scheme due to the poor performance of MRF. Additional FDDA sensitivity tests were performed that increased the FDDA nudging strength by a factor of 2 and then added in observation nudging. The final MM5 configuration (EXP3) was exactly the same as the third configuration MM5 experiment EXP2H, only using the Pleim-Xiu PBL/LSM scheme.

The CALMET sensitivity tests are listed in Table A-3. The MM5 output from either MM5 EXP1C (80 km) or MM5 EXP2H (36 and 12 km) were used as initial guess winds in the CALMET experiments. The CALMET sensitivity tests varied by the CALMET grid resolution, the source and grid resolution of the MM5 output data used and how the surface and upper-air meteorological data were blended into the STEP1 wind fields in the STEP2 OA procedure. There were seven basic CALMET configurations:

BASE Use 80 km MM5 data from EXP1C and 18 km CALMET grid resolution.
1. Use 80 km MM5 data from EXP1C and 12 km CALMET grid resolution.
2. Use 80 km MM5 data from EXP1C and 4 km CALMET grid resolution.
3. Use 36 km MM5 data from EXP2H and 12 km CALMET grid resolution.
4. Use 12 km MM5 data from EXP2H and 12 km CALMET grid resolution.
5. Use 36 km MM5 data from EXP2H and 4 km CALMET grid resolution.
6. Use 12 km MM5 data from EXP2H and 4 km CALMET grid resolution.

The variations in the CALMET STEP2 OA procedures in the CALMET sensitivity test were as follows:

A. Use meteorological observations with RMAX1/RMAX2 = 500/1000.
B. Use meteorological observations with RMAX1/RMAX2 = 100/200.
C. Use meteorological observations with RMAX1/RMAX2 = 10/100.
D. Don't use any meteorological observations (NOOBS = 2).

Table A-2. Summary of CTEX5 MM5 sensitivity tests.

Sensitivity Test	Horizontal Grid	Vertical Layers	PBL	LSM	FDDA Used
1A_80km	80 km	16	BLKDR	5LAY	No FDDA
1B_80km	80 km	16	BLKDR	5LAY	Analysis Nudging
1C_80km	80 km	16	BLKDR	5LAY	Analysis Nudging Surface Analysis Nudging
2A_36km 2A_12km	108/36/12km	33	MRF	5LAY	No FDDA
2B_36km 2B_12km	108/36/12km	33	MRF	5LAY	Analysis Nudging
2C_36km	108/36/12km	33	MRF	5LAY	Analysis Nudging Surface Analysis Nudging
2F_36km 2G_12km	108/36/12km	43	BLKDR	5LAY	No FDDA
2G_36km 2G_12km	108/36/12km	43	BLKDR	5LAY	Analysis Nudging
2H_36km 2H_12km	108/36/12km	43	BLKDR	5LAY	Analysis Nudging Surface Analysis Nudging
2I_36km 2I_12km	108/36/12km	43	BLKDR	5LAY	Analysis Nudging Surface Analysis Nudging FDDA x 2 strength
2J_36km 2J_12km	108/36/12km	43	BLKDR	5LAY	Analysis Nudging Surface Analysis Nudging FDDA x 2 strength Observational Nudging
4_36km 4_12km	108/36/12km	43	PX	PX	Analysis Nudging Surface Analysis Nudging

Table A-3. Definition of the CTEX5 CALMET sensitivity tests and data sources.

Sensitivity Test	MM5 Experiment and Resolution	CALMET Resolution	RMAX1/RMAX2	NOOBS[A]
BASEA	EXP1C – 80 km	18 km	500/1000	0
BASEB	EXP1C – 80 km	18 km	100/200	0
BASEC	EXP1C – 80 km	18 km	10/100	0
BASED	EXP1C – 80 km	18 km	NA	2
1A	EXP1C – 80 km	12 km	500/1000	0
1B	EXP1C – 80 km	12 km	100/200	0
1C	EXP1C – 80 km	12 km	10/100	0
1D	EXP1C – 80 km	12 km	NA	2
2A	EXP1C – 80 km	4 km	500/1000	0
2B	EXP1C – 80 km	4 km	100/200	0
2C	EXP1C – 80 km	4 km	10/100	0
2D	EXP1C – 80 km	4 km	NA	2
3A	EXP2H – 36 km	12 km	500/1000	0
3B	EXP2H – 36 km	12 km	100/200	0
3C	EXP2H – 36 km	12 km	10/100	0
3D	EXP2H – 36 km	12 km	NA	2
4A	EXP2H – 12 km	12 km	500/1000	0
4B	EXP2H – 12 km	12 km	100/200	0
4C	EXP2H – 12 km	12 km	10/100	0
4D	EXP2H – 12 km	12 km	NA	2
5A	EXP2H – 36 km	4 km	500/1000	0
5B	EXP2H – 36 km	4 km	100/200	0
5C	EXP2H – 36 km	4 km	10/100	0
5D	EXP2H – 36 km	4 km	0/0	2
6A	EXP2H – 12 km	4 km	500/1000	0
6B	EXP2H – 12 km	4 km	100/200	0
6C	EXP2H – 12 km	4 km	10/100	0
6D	EXP2H – 12 km	4 km	NA	2
6K	EXP2H – 12 km	4 km	NA	2

A. NOOBS = 0 use surface and upper-air meteorological observations
NOOBS = 2 do not use surface and upper-air meteorological observations
NOOBS = 1 use surface but not upper-air meteorological observations

Figure A-1 compares the MM5 model estimated wind fields. Figures A-2 and A-3 display the temperature and humidity model performance for the MM5 simulations. As shown in Figure A-2, the temperature performance for the three MM5 sensitivity tests using the MRF PBL scheme (2A, 2B and 2C) is extremely poor using either the 36 or 12 km grid resolution having an underestimation bias greater than -4 degrees that does not meet the temperature bias performance goal (≤±0.5 degrees).

The wind speed and, especially, the wind direction performance of the MM5 simulations with no FDDA (1A, 2A and 2F) is noticeably worse than when FDDA is used with the wind direction bias and error exceeding the performance benchmarks when no FDDA is used. With the exception of the EXP2H temperature underestimation tendency that barely exceeds the performance benchmark, the MM5 EXP1C and EXP2H MM5 sensitivity tests that were used in the CALMET sensitivity tests achieve the model performance benchmarks for wind speed, wind direction, temperature and humidity.

Tables A-4 and A-5 show CALMET estimated winds compared to observations. The “A” series of CALMET sensitivity tests (RMAX1/RMAX2 =500/1000) tends to have a wind speed underestimation bias compared to the other RMAX1/RMAX2 settings for most of the base CALMET settings (Figure A-1). The “A” and “B” series of CALMET runs tend to have the winds that closest match observations compared to the “C” (RMAX1/RMAX2 = 10/100) and “D” (no observations) series of CALMET runs. The use of 12 km CALMET grid resolution appears to improve the CALMET model performance slightly compared to 80 and 36 km. The CALMET runs using the MM5 EXP2H 36/12 km data appear to perform better than the ones that used the MM5 EXP1C 80 km data. CALMET tends to slow down the MM5 wind speeds with the slowdown increasing going from the “D” to “C” to “B” to “A” series of CALMET configurations such that the “A” series has a significant wind speed underestimation tendency.

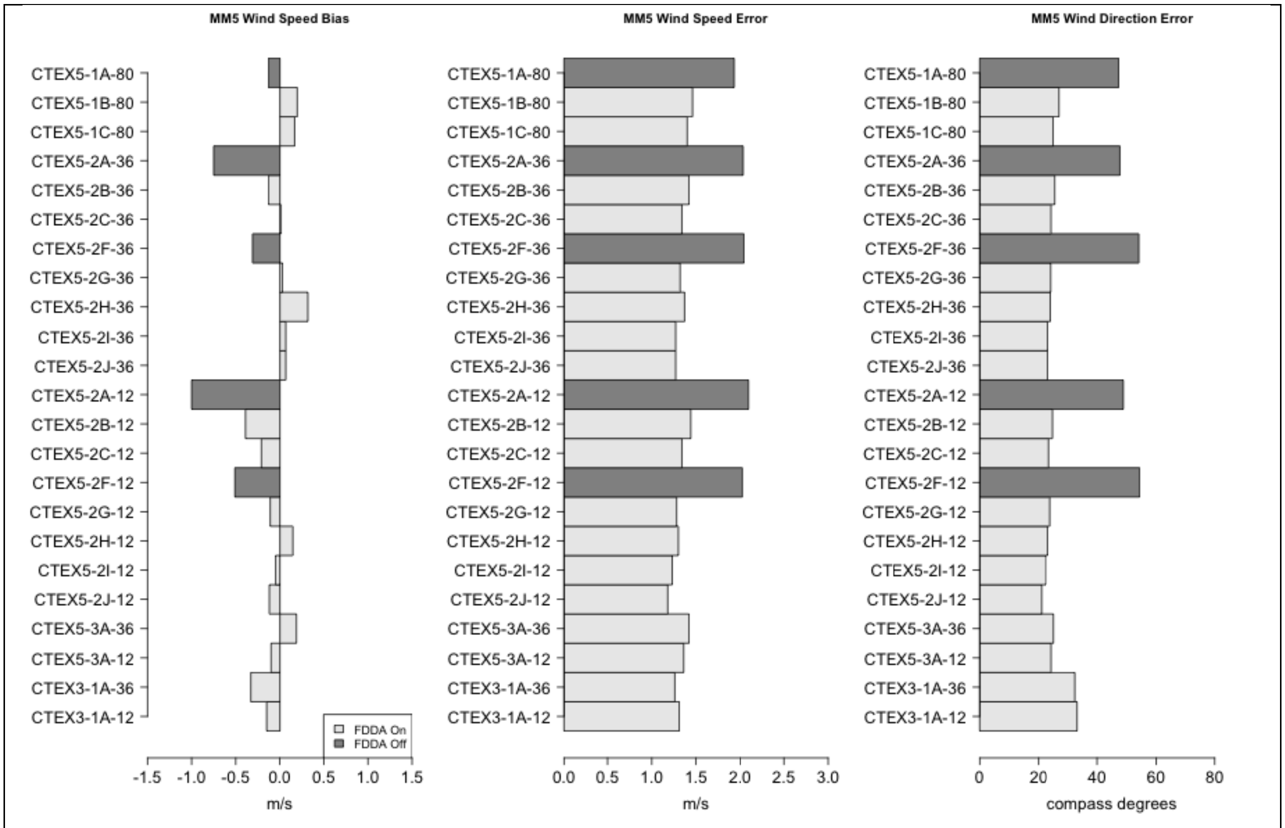


Figure A-1. Wind speed bias (m/s), wind speed error (m/s) and wind direction error (degrees) for MM5 runs.

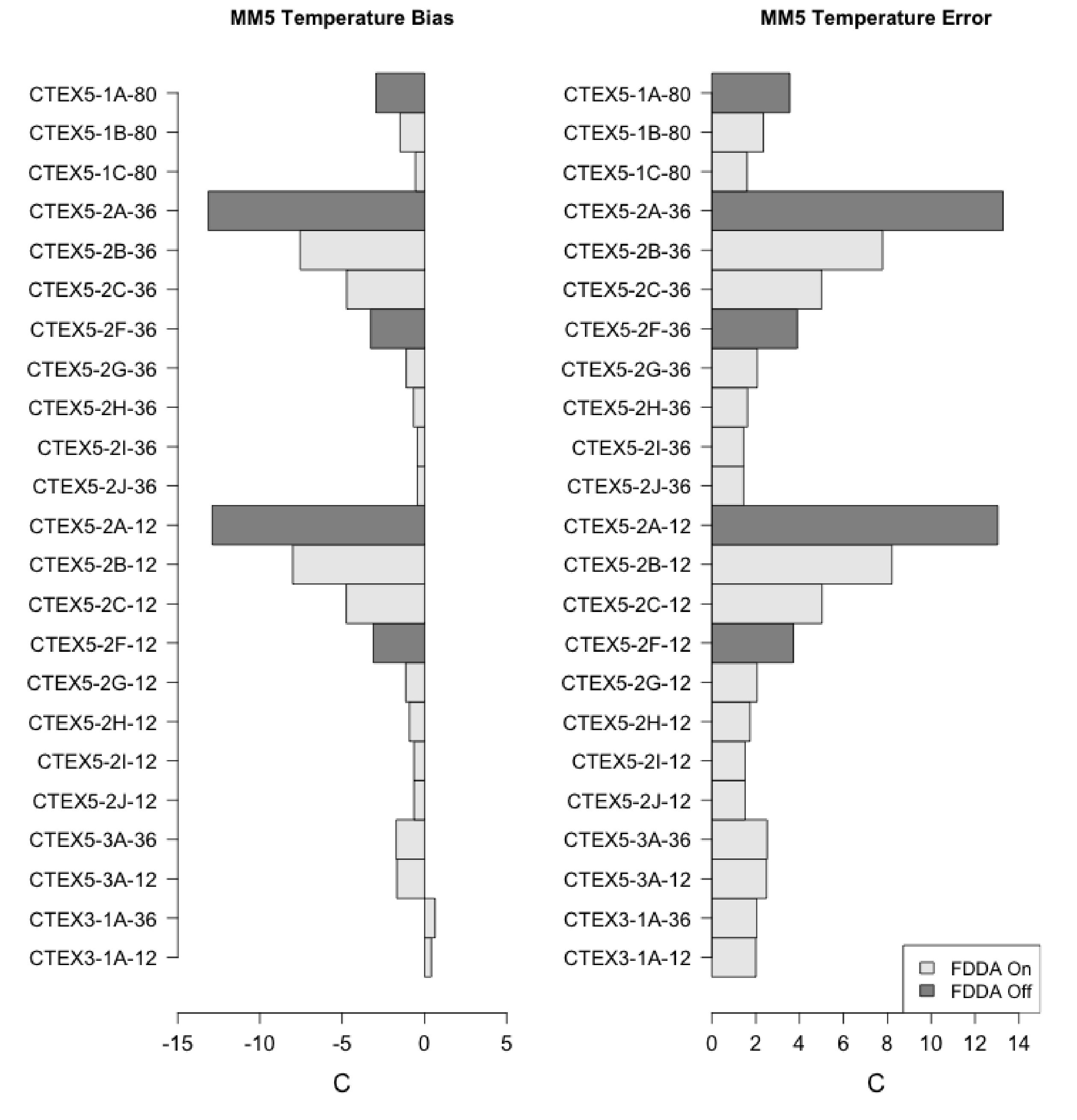


Figure A-2. Temperature bias and error (degrees K) of the CTEX5 MM5 meteorological modeling.

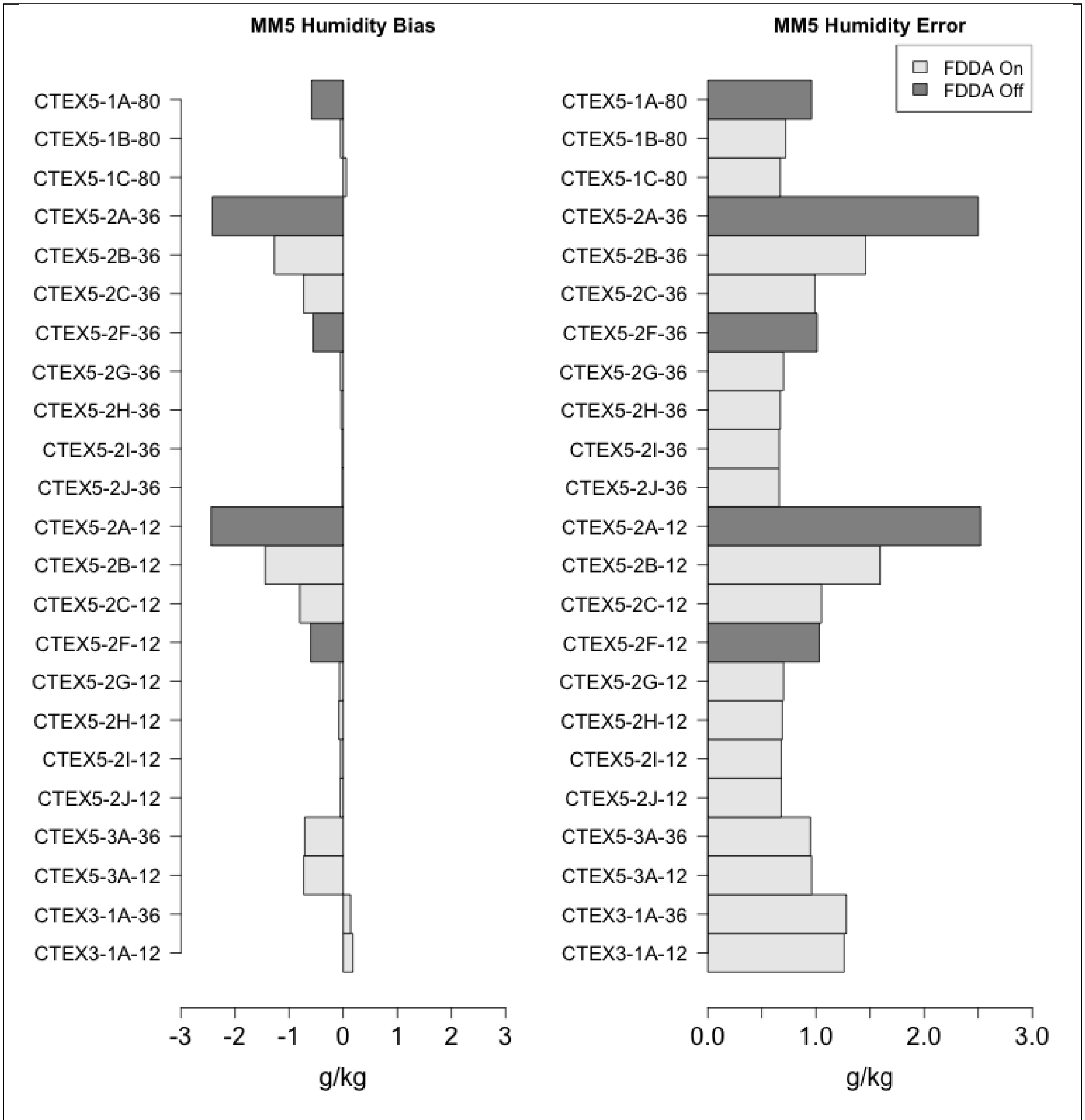


Figure A-3. Humidity bias and error (g/kg) of the CTEX5 MM5 meteorological modeling.

Table A-4. Comparison of CTEX5 MM5 meteorological simulation EXP1C and CALMET simulations using EXP1C MM5 80 km data as input.

	Wind Speed (m/s)			Wind Direction (°)	
	Bias	**Error**	**RMSE**	**Bias**	**Error**
Benchmark	≤±0.5	≤2.0	≤2.0	≤±10	≤30
MM5_EXP1C	0.17	1.40	1.83	4.52	25.1
CALMET					
BASEA	-0.35	0.89	1.38	-0.42	15.9
BASEB	-0.11	0.84	1.32	1.01	15.2
BASEC	-0.01	1.26	1.67	4.26	23.8
BASED	0.03	1.34	1.76	4.53	25.1
1A	-0.29	0.82	1.36	-0.56	14.9
1B	0.06	0.78	1.3	0.67	14.3
1C	-0.03	1.22	1.62	3.80	22.9
1D	0.02	1.34	1.77	4.45	25.1
2A	-0.21	0.71	1.33	-0.78	13.9
2B	-0.02	0.69	1.29	0.31	13.3
2C	-0.08	0.96	1.40	2.28	17.9
2D	0.00	1.34	1.77	4.08	25.0

Table A-5. Comparison of CTEX5 MM5 meteorological simulation EXP2H and CALMET simulations using EXP2H MM5 36 and 12 km data as input.

	Wind Speed (m/s)			Wind Direction (°)	
	Bias	Error	RMSE	Bias	Error
Benchmark	≤±0.5	≤2.0	≤2.0	≤±10	≤30
MM5_EXP2H	0.32	1.37	1.78	5.07	24.2
CALMET					
3A	-0.29	0.82	1.35	-0.56	14.9
3B	-0.01	0.78	1.29	0.83	14.2
3C	0.17	1.20	1.59	4.50	22.0
3D	0.24	1.34	1.74	5.13	24.1
4A	-0.29	0.82	1.35	-0.56	14.9
4B	-0.03	0.76	1.25	0.36	14.0
4C	0.07	1.16	1.54	3.36	21.4
4D	0.13	1.28	1.67	3.83	23.5
5A	-0.21	0.71	1.33	-0.78	13.9
5B	0.03	0.69	1.28	0.50	13.2
5C	0.08	0.95	1.39	2.87	17.4
5D	0.21	1.33	1.75	4.79	24.1
5K	0.04	0.69	1.28	0.47	13.2
6A	-0.21	0.71	1.33	-0.79	13.9
6B	-0.05	0.67	1.24	0.04	13.0
6C	-0.02	0.92	1.33	1.84	16.7
6D	-0.26	1.33	1.77	4.67	24.5
6K	0.00	0.66	1.23	0.00	13.0

Appendix B

EVALUATION OF VARIOUS CONFIGURATIONS OF THE CALMET METEOROLOGICAL MODEL USING THE CAPTEX CTEX3 FIELD EXPERIMENT DATA

B.1 CALMET MODEL EVALUATION TO IDENTIFY RECOMMENDED CONFIGURATION

The CAPTEX Release #3 (CTEX3) meteorological database was used to evaluate different configurations of the CALMET meteorological model for the purposes of helping to identify a recommended configuration for regulatory far-field CALMET/CALPUFF modeling. The results from these CALMET CTEX3 sensitivity tests were used in part to define the recommended CALMET model options in the August 31, 2009 Memorandum from the EPA/OAQPS Air Quality Modeling Group "Clarifications on EPA-FLM Recommended Settings for CALMET (i.e., the 2009 Clarification Memorandum). The EPA Clarification Memorandum on CALMET settings (EPA, 2009a) was a follow-up to a draft May 27, 2009 document: "Reassessment of the Interagency Workgroup on Air Quality Modeling (IWAQM) Phase 2 Summary Report: Revisions to Phase 2 Recommendations" (EPA, 2009a). The IWAQM Phase 2 Reassessment Report recommended settings for CALMET that were intended to facilitate the direct "pass through" of prognostic meteorological model (e.g., MM5 and WRF) output to CALPUFF as much as possible. However, in subsequent testing of the new recommended CALMET settings in the IWAQM Phase 2 Reassessment Report using the CTEX3 database, the performance of CALMET degraded compared to some other settings. This led to the August 31, 2009 Clarification Memorandum of recommended CALMET settings for regulatory far-field modeling.

EPA examined 31 different configurations of the CALMET diagnostic meteorological model using the CTEX3 database. The resultant CALMET wind fields were paired in space and time with observations using the CALMETSTAT tool. CALMETSTAT is an adaptation of the METSTAT program that is typically used to evaluate the MM5 and WRF prognostic meteorological models against surface meteorological observations.

Note that since CALMET uses some of the same meteorological observations as input as used in the evaluation database, this is not a true evaluation as by design CALMET's STEP2 objective analysis (OA) will modify the wind field to make the winds better match the observations at the locations of the monitoring sites. But as noted by EPA (2009a,b), this can be at the expense of degrading the wind fields.

Table B-1 lists the 31 CALMET sensitivity tests that were performed using the CTEX3 modeling database. These CALMET sensitivity tests differed in the following aspects:

- The resolution of the CALMET gridded fields (18, 12 and 4 km);
- The resolution of the MM5 prognostic meteorological model output used as input to CALMET (80, 36 and 12 km);
- How the MM5 data was used in CALMET (i.e., as a first guess field prior to the STEP 1 diagnostic effects, as the STEP 1 wind fields prior to STEP 2 blending (objective analysis or OA) of observations or the MM5 data are not used at all); and
- Whether the surface and upper-air meteorological observations were used (NOOBS=0) or not (NOOBS=2).

Table B-1. CTEX3 CALMET sensitivity simulations performed for the CTEX3 database.

RUN	CALMET Resolution	MM4/MM5 Resolution	NOOBS	RMAX1/RMAX2	IPROG
BASE A	18-km	80-km MM4	0	500/1000	STEP 1
BASE B	18-km	80-km MM4	0	500/1000	First Guess
BASE C	18-km	80-km MM4	0	10/100	First Guess
BASE D	18-km	80-km MM4	0	100/200	First Guess
BASE E	18-km	80-km MM4	0	10/100	STEP 1
BASE F	18-km	80-km MM4	0	10/100	First Guess
BASE G[A]	18-km	80-km MM4	2	NA	First Guess
BASE H	18-km	NA	0	500/1000	NA
BASE I	18-km	NA	0	100/200	NA
BASE J	18-km	NA	0	10/100	NA
BASE K	18-km	80-km MM4	0	100/200	First Guess[B]
EXP 1A	18-km	36-km MM5	0	500/1000	First Guess
EXP 1B	18-km	36-km MM5	0	100/200	First Guess
EXP 1C	18-km	36-km MM5	0	10/100	First Guess
EXP 1D	18-km	36-km MM5	2	NA	First Guess
EXP 3A	12-km	36-km MM5	0	500/1000	First Guess
EXP 3B	12-km	36-km MM5	0	100/200	First Guess
EXP 3C	12-km	36-km MM5	0	10/100	First Guess
EXP 3D	12-km	36-km MM5	2	NA	First Guess
EXP 4A	12-km	12-km MM5	0	500/1000	First Guess
EXP 4B	12-km	12-km MM5	0	100/200	First Guess
EXP 4C	12-km	12-km MM5	0	10/100	First Guess
EXP 4D	12-km	12-km MM5	2	NA	First Guess
EXP 5A	4-km	36-km MM5	0	500/1000	First Guess
EXP 5B	4-km	36-km MM5	0	100/200	First Guess
EXP 5C	4-km	36-km MM5	0	10/100	First Guess
EXP 5D	4-km	36-km MM5	2	NA	First Guess
EXP 6A	4-km	12-km MM5	0	500/1000	First Guess
EXP 6B	4-km	12-km MM5	0	100/200	First Guess
EXP 6C	4-km	12-km MM5	0	10/100	First Guess
EXP 6D	4-km	12-km MM5	2	NA	First Guess

A. Base G CALMET simulation obtained an Error in MIXDT2 – HTOLD so run not completed
B. Base K did not do any diagnostic adjustments to the wind fields

Figure B-1 displays the wind speed and direction model performance statistical metric for the Base A through Base K CALMET sensitivity test simulations that used either the 80 km MM4 or no prognostic meteorological model data as input. The dark gray bar represents the CALMET model configuration that is consistent with the recommendations in the August 31, 2009 Clarification Memorandum. The numerical values of the model performance statistics are provided in Table B-2. CALMET sensitivity simulations Base D, H, I and K are the best performing simulations for winds from this group. Base D is the current recommended CALMET settings, whereas Base H and I use no MM4 data and Base K is like Base D only CALMET does not perform any diagnostic wind field adjustments. The wind speed statistics for Base D and K are identical, whereas the ones for Base H and I are slightly worse than Base D and K. The wind direction statistics for Base D and K are almost identical and again the ones for Base H and I are slightly worse.

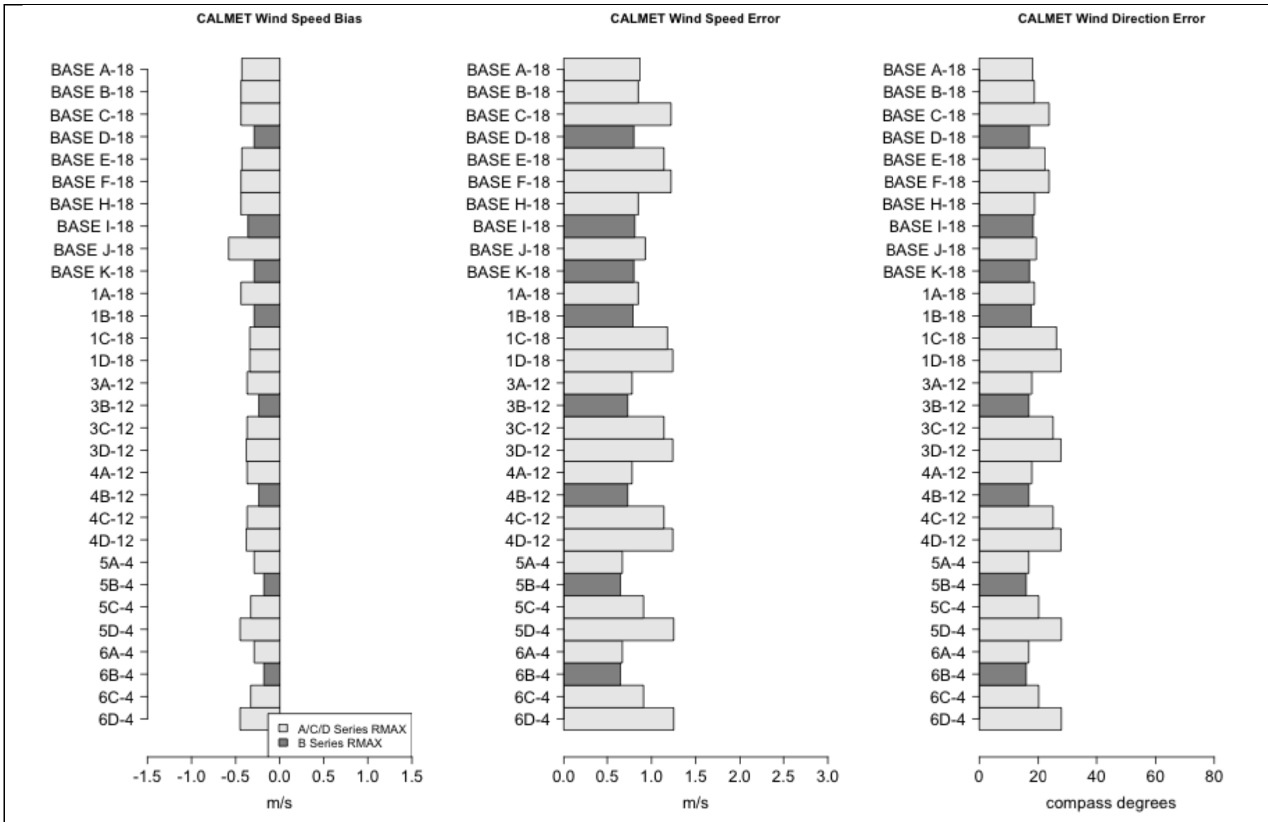


Figure B-1. Wind speed and wind direction comparisons with observations for CTEX3 CALMET sensitivity simulations.

Figure B-1 displays the wind speed and direction performance metrics for the second group of CALMET sensitivity tests that uses CALMET grid resolutions of 18 km (EXP1) and 12 km (EXP3 and EXP4) and uses 36 km MM5 (EXP1 and EXP3) and 12 km MM5 (EXP4) data as input to CALMET. EXP1B, EXP3B and EXP4B CALMET sensitivity tests all conform to the recommended settings in the Clarification Memorandum. The "B" series most closely matches observation data.

The CALMET model performance statistics for the final group of CTEX3 sensitivity tests corresponding to the EXP5 and EXP6 series of experiments are shown in Figure B-1. These experiments correspond to using a 4 km grid resolution in CALMET, which is the finest scale recommended in the Clarification Memorandum. They differ in the resolution of MM5 data used as input (36 or 12 km) and how observations are blended into the wind fields (different RMAX1/RMAX2 or no observations). When looking across all wind speed and direction statistics, the CALMET sensitivity simulations that conform to the CALMET settings in the August 2009 Clarification Memorandum (EXP5B and EXP6B) compare most closely to observations.

Figure B-1 displays the CALMET model performance statistics for all sensitivity tests that conform to the recommended CALMET settings in the Clarification Memorandum. The Clarification Memorandum specifies that prognostic meteorological model output data should be used as a first guess wind field in CALMET (IPROG = 14), but doesn't specify the resolution that the prognostic meteorological model should be run at. For the CALMET grid resolution, the Clarification Memorandum just specifies that it should be ≥ 4 km. Thus these CALMET sensitivity tests vary by grid resolution used in the prognostic meteorological model (80, 36 and 12 km) whose output is used as input to CALMET and the CALMET grid resolution (18, 12 and 4 km).

Table B-2a. Summary wind speed model performance statistics for the CALMET CTEX3 sensitivity tests.

RUN	WS Gross Error (ms^{-1})	WS Bias (ms^{-1})	WS RMSE (ms^{-1})	IOA
BASE A	0.87	-0.43		0.81
BASE B	0.85	-0.44	1.60	0.82
BASE C	1.22	-0.44	1.62	0.63
BASE D	0.80	-0.29	1.23	0.83
BASE E	1.14	-0.43	1.52	0.68
BASE F	1.22	-0.44	1.62	0.63
BASE G	NA	NA	NA	NA
BASE H	0.85	-0.44	1.30	0.82
BASE I	0.81	-0.36	1.30	0.82
BASE J	0.93	-0.58	1.36	0.80
BASE K	0.80	-0.29	1.23	0.83
EXP 1A	0.85	-0.44	1.29	0.82
EXP 1B	0.79	-0.29	1.22	0.83
EXP 1C	1.18	-0.34	1.57	0.68
EXP 1D	1.24	-0.34	1.64	0.65
EXP 3A	0.78	-0.37	1.26	0.83
EXP 3B	0.73	-0.24	1.20	0.85
EXP 3C	1.14	-0.37	1.52	0.70
EXP 3D	1.24	-0.38	1.64	0.65
EXP 4A	0.78	-0.37	1.26	0.83
EXP 4B	0.73	-0.24	1.20	0.85
EXP 4C	1.14	-0.37	1.52	0.70
EXP 4D	1.24	-0.38	1.64	0.65
EXP 5A	0.67	-0.29	1.24	0.84
EXP 5B	0.65	-0.18	1.19	0.85
EXP 5C	0.91	-0.33	1.31	0.80
EXP 5D	1.25	-0.45	1.25	0.65
EXP 6A	0.67	-0.29	1.24	0.84
EXP 6B	0.65	-0.18	1.19	0.85
EXP 6C	0.91	-0.33	1.31	0.80
EXP 6D	1.25	-0.45	1.66	0.65

Table B-2b. Summary wind direction model performance statistics for the CALMET CTEX3 sensitivity tests.

RUN	WD Gross Error (deg.)	WD Bias (deg.)
BASE A	18.06	0.73
BASE B	18.62	-0.74
BASE C	23.91	2.63
BASE D	16.92	0.40
BASE E	22.31	2.68
BASE F	23.91	2.63
BASE G	NA	NA
BASE H	18.70	-0.79
BASE I	18.22	-0.65
BASE J	19.30	-0.85
BASE K	16.97	0.38
EXP 1A	18.64	-0.72
EXP 1B	17.59	1.15
EXP 1C	26.43	2.99
EXP 1D	27.99	3.11
EXP 3A	17.80	-0.82
EXP 3B	16.75	0.98
EXP 3C	25.25	2.57
EXP 3D	27.93	2.94
EXP 4A	17.80	-0.82
EXP 4B	16.75	0.98
EXP 4C	25.25	2.57
EXP 4D	27.93	2.94
EXP 5A	16.73	-1.00
EXP 5B	15.85	0.72
EXP 5C	20.11	1.42
EXP 5D	28.05	2.43
EXP 6A	16.73	-1.00
EXP 6B	15.85	0.72
EXP 6C	20.11	1.42
EXP 6D	28.05	2.43

B.2 CONCLUSIONS OF CTEX3 CALMET SENSITIVITY TESTS

The evaluation of the CALMET modeling system using the CTEX3 field experiment database is not a true independent evaluation because some of the surface meteorological observations used as the evaluation database are also used as input into CALMET. Thus, care should be taken in the interpretation of the CALMET meteorological model evaluation. In fact, EPA has demonstrated that CALMET's blending of meteorological observations with MM5 prognostic meteorological model fields can actually produce unrealistic results in the wind fields (e.g., discontinuities around the wind observation sites) at the same time as improving the CALMET statistical model performance at the meteorological monitoring sites.

Given these caveats, when looking at the alternative CALMET settings for RMAX1/RMAX2 the CALMET configuration that best matches observed winds is with the 100/200 RMAX1/RMAX2 setting as recommended in the 2009 Clarification Memorandum. Other recommended settings in the 2009 Clarification Memorandum (e.g., use of prognostic meteorological data as the initial first guess wind field) are supported by the CALMET CTEX3 model evaluation. Note that better wind field comparisons using the 2009 Clarification Memorandum recommended settings for RMAX1/RMAX2 was also seen for the CTEX5 CALMET evaluation presented in Appendix A.

Although the CALMET meteorological model performance evaluation for alternative model settings support the recommended 100/200 CALMET settings for RMAX1/RMAX2 in the Clarification Memorandum, the evaluation of the CALPUFF/CALMET modeling system for the CTEX3 and CTEX5 field experiments against observed tracer data presented in Chapter 5 come to an alternative conclusion. The CALPUFF/CALMET evaluation against the observed tracer observations in the CTEX3 and CTEX5 experiments found that different RMAX1/RMAX2 configurations produced better CALPUFF/CALMET tracer model performance for the two CAPTEX experiments, but that the 100/200 recommended setting always produced the worst CALPUFF/CALMET model performance. Given the large differences in the in the rankings of the ability of the CALPUFF to reproduce the observed tracer concentrations across the different meteorological model configurations in the two CAPTEX field experiments, it is unclear whether a third experiment would produce another set of rankings.

Appendix C

INTERCOMPARISON OF SIX LRT MODELS AGAINST THE CAPTEX RELEASE 3 AND RELEASE 5 FIELD EXPERIMENT DATA

C.1 INTRODUCTION

In this section, the evaluation of six LRT dispersion models (CALPUFF, SCIPUFF, HYSPLIT, FLEXPART, CAMx, and CALGRID) against the Cross Appalachian Tracer Study (CAPTEX) (Section 5) is presented. The ATMES-II evaluation framework described in Section 2.4.3.1 and 2.4.3.3 are utilized to conduct this evaluation. The CAPTEX evaluations generally follow the ETEX evaluation paradigm, all models presented in this section use a common 36 km MM5 meteorological data source. Thus the results from the CALMET/CALPUFF sensitivities are not presented because they are not within the scope of this evaluation framework. However, we do wish to note that CALPUFF/CALMET performance for CAPTEX-5 (EXP6C) was quite good, and exceeded that of the other models involved in the model intercomparison portion of this section; however, due to a different source of meteorology, only the MMIF/CALPUFF results for the same MM5 run and grid resolution are included.

In addition to the six model intercomparison, sensitivities of the HYSPLIT INITD and CAMx vertical diffusion and horizontal advection solver (Kz/advection solver) combinations are also presented. The best performing INITD and Kz/advection solver combinations are presented for purposes of model intercomparison.

C.2 HYSPLIT SENSITIVITY TESTS

Consistent with the approach taken for evaluating HYSPLIT for the European Tracer Experiment discussed in Section 6.4.5, HYSPLIT was evaluated using each of the nine INITD model configurations. The HYSPLIT INITD option defines the technical formulation of the dispersion model from fully particle to fully Lagrangian puff with several hybrid particle/puff combinations. A description of the INITD variable options is provided in Table 6-4. The HYSPLIT configurations for each INITD option are presented in Table C-1.

Table C-1. HYSPLIT sensitivity runs and relevant configuration parameters.

Sensitivity Test	INITD	NUMPAR	ISOT	KSPL	FRHS	FRVS	FRTS	FRME
INITD0	0	10000	1	NA	NA	NA	NA	NA
INITD1	1	10000	1	1	1.0	0.01	0.10	0.10
INITD2	2	10000	1	1	1.0	0.01	0.10	0.10
INITD3	3	10000	1	1	1.0	0.01	0.10	0.10
INITD4	4	10000	1	1	1.0	0.01	0.10	0.10
INITD103	103	10000	1	1	1.0	0.01	0.10	0.10
INITD104	104	10000	1	1	1.0	0.01	0.10	0.10
INITD130	130	10000	1	1	1.0	0.01	0.10	0.10
INITD140	140	10000	1	1	1.0	0.01	0.10	0.10

C.2.1 HYSPLIT SPATIAL PERFORMANCE FOR CAPTEX RELEASE 3

Figure C-1 displays the spatial model performance statistics for the HYSPLIT INITD sensitivity tests. Unlike the results from the HYSPLIT sensitivities from ETEX, the variation in spatial performance is much smaller. While the puff based INITD configurations showed slightly lower scores for POD, their scores for all other spatial categories is nearly identical to the other INITD configurations. INITD3 has the highest FMS score (34%) with INITD1 nearly the same at 33%. Consistent with the ETEX results, the puff configuration INITD1 (Gh-Thv) yielded slightly better performance statistics across spatial categories than the INITD2 puff configuration (Thh-Thv). Overall for CAPTEX Release 3, there appears to be little advantage of one INITD configuration over another for the four spatial categories of model performance metrics.

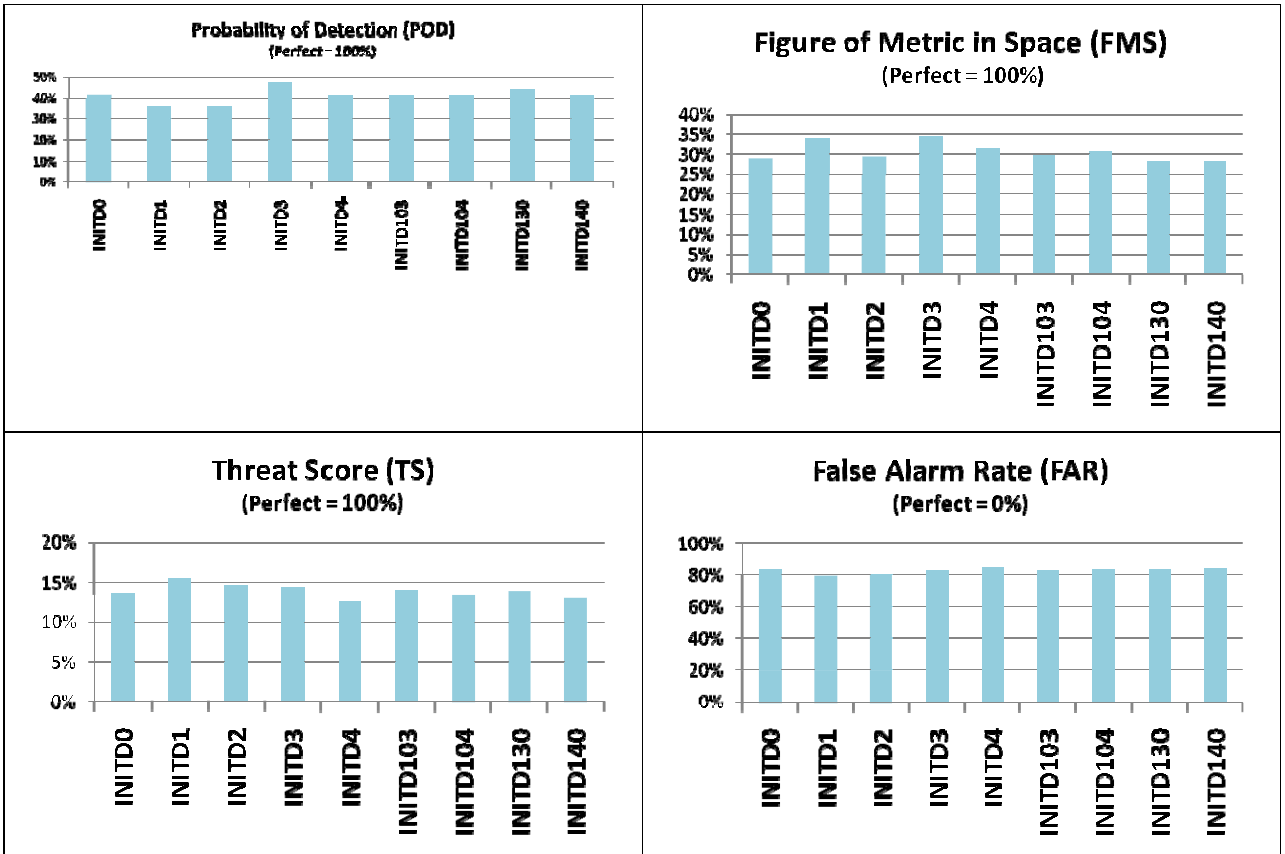


Figure C-1. Spatial model performance statistics for nine HYSPLIT INITD sensitiviity experiments for CAPTEX Release 3.

C.2.2 HYSPLIT GLOBAL STATISTICS FOR CAPTEX RELEASE 3

Figures C-2 and C-3 display the global statistics for the HYSPLIT sensitivity tests with Figures C-2 and C-3 containing the statistical metrics where the best performing model has the, respectively, lowest and highest score. For the FOEX metrics, INITD140 scores the best with nearly 0%, followed closely by INITD1 and INITD2. INITD3 scores the poorest with a 21% FOEX score followed by INITD4. The two puff configurations had the poorest NMSE and FB statistical performance metrics (with values of approximately 127 and 130 pg m^{-3} for error and 1.56 and 1.57 for FB). The four puff-particle model configuration options (INITD3,4,103,104) exhibited the best overall scores for both NMSE and FB. INITD1 and INITD2 exhibited the best overall KSP score with 30% and 31% respectively, with the poorest performing being INITD3 with 49%.

For the within a factor of 2 and 5 metric (FA2 and FA5, Figure C-3, top), the hybrid puff-particle configurations INITD3 and INITD4 and their counterpart particle-puff configurations INITD103 and INITD104 perform slightly better than the other configurations. For the PCC metric (PCC, Figure C.2.2-2, bottom left), all of the HYSPLIT configurations show a slight negative correlation ranging from -0.04 to -0.09.

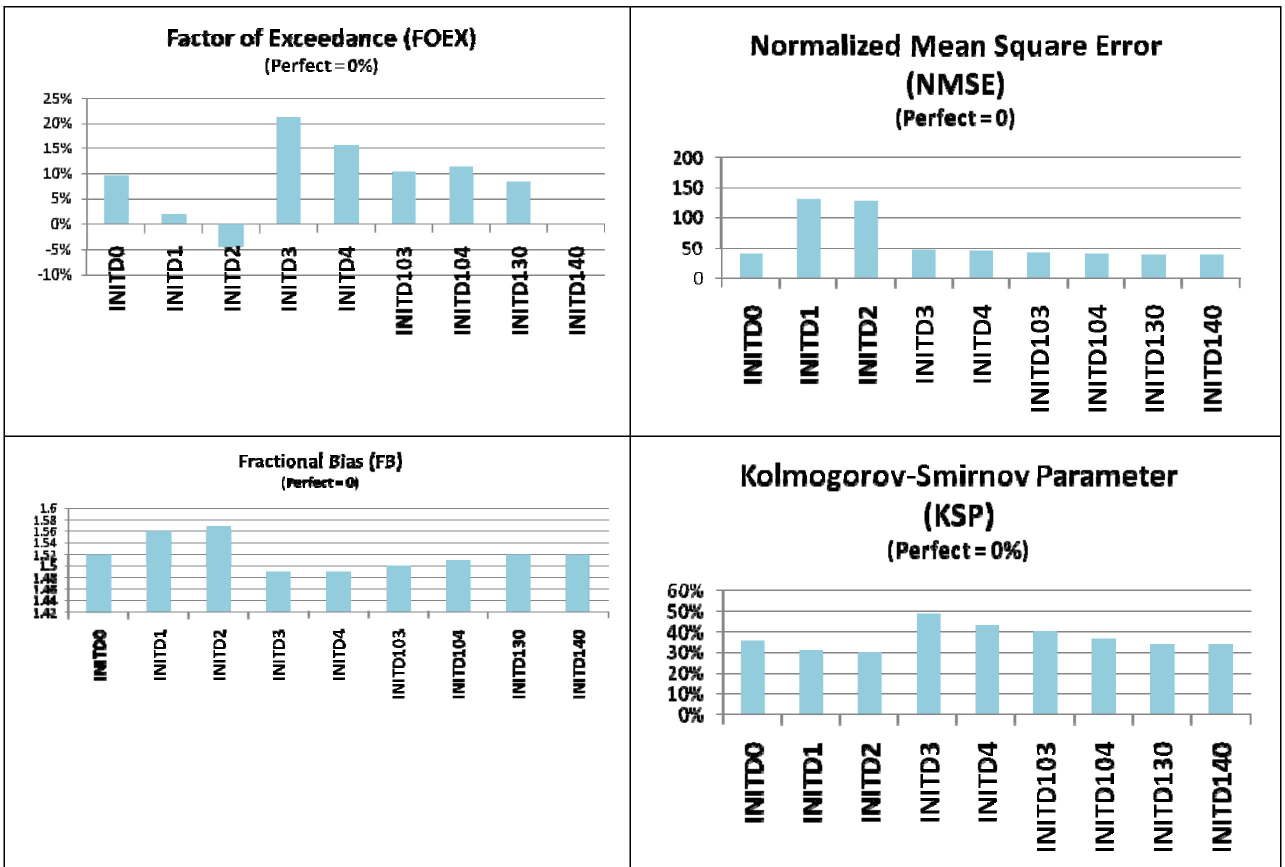


Figure C-2. Global model performance statistics for nine HYSPLIT INITD sensitivity tests for CAPTEX Release 3.

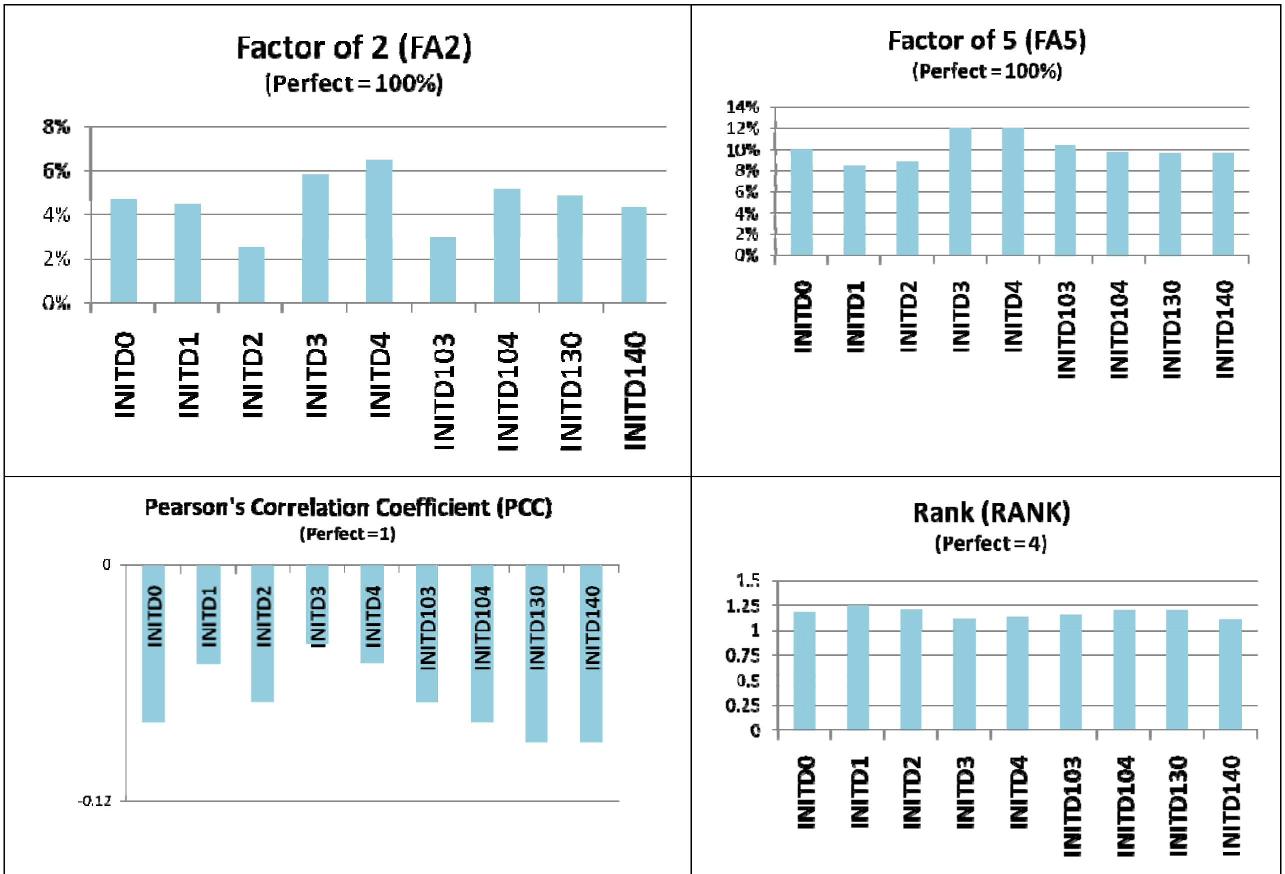


Figure C-3. Global model performance statistics for nine HYSPLIT INITD sensitivity tests for CAPTEX Release 3.

The final panel in Figure C-3 (bottom right) displays the overall RANK statistic. The RANK statistics orders the model performance of the HYSPLIT INITD configurations as follows:

1. INITD1 (1.25)
2. INITD2 (1.21)
3. INITD104 (1.19)
4. INITD130 (1.19)
5. INITD0 (1.18)
6. INITD103 (1.15)
7. INITD4 (1.14)
8. INITD3 (1.11)
9. INITD140 (1.10)

The RANK performance statistics results presented above raise some interesting questions about the RANK metric. The puff based configurations (INITD1 and INITD2) are the highest ranking with scores using the RANK metric with values of 1.25 and 1.21 respectively. However, each of these options had the worst (highest) NMSE and FB scores, while puff-particle configurations ranking slightly less using the RANK metric (1.1 to 1.19) have NMSE scores that are much better (only one-third) those for the puff configurations as well as slightly lower FB scores. On the basis of RANK scores, the INITD1 and INITD2 configurations are the best performing, but based upon other model performance statistics that are not included as the four statistical metrics that make up the RANK metric (i.e., PCC, FB, FMS and KSP), the puff-particle hybrid configurations are better performing. Thus care must be taken in interpreting model performance based solely on the RANK score and its use in performing model intercomparisons and we recommend examining the whole suite of statistical performance metrics, as well as graphical representation of model performance, to come to conclusions regarding model performance.

C.2.3 HYSPLIT SPATIAL STATISTICS FOR CAPTEX RELEASE 5

Figure C-4 displays the spatial model performance statistics for the HYSPLIT INITD sensitivity tests for CAPTEX Release 5. Overall, the spatial performance for this experiment is very similar to the results obtained from the ETEX INITD sensitivities for HYSPLIT. The puff configurations (INITD1 and INITD2) exhibited the poorest performance across all of the spatial statistics. INITD2 had the poorest FMS score with 5%, followed by INITD1 with 9.6%. INITD3 had the best FMS score of 19.66%, but less than 2% separated all of the remaining particle and puff-particle INITD configurations. The particle mode (INITD0) exhibited the best TS with 24.4% with less than 1.5% separating INITD103, 130, and 140 from INITD0. Consistently, the puff configurations exhibited the lowest TS among the nine configurations, both with 7.9%.

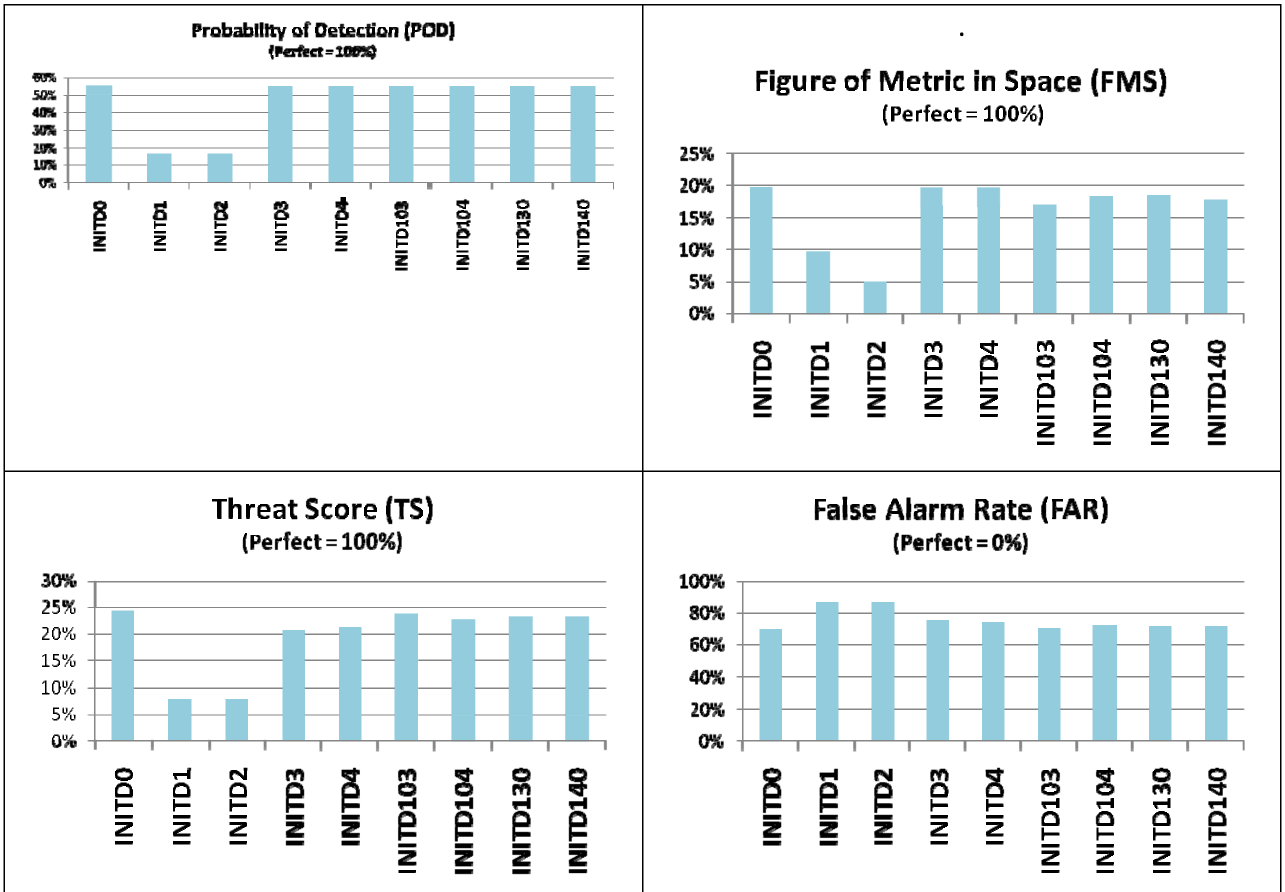


Figure C-4. Spatial model performance statistics for nine HYSPLIT INITD sensitiviity experiments for CAPTEX Release 5.

C.2.4 HYSPLIT GLOBAL STATISTICS FOR CAPTEX RELEASE 5

Figures C-5 and C-6 display the global statistics for the HYSPLIT sensitivity tests for CAPTEX Release 5 where the two figures containing the statistical metrics where the best performing model has the, respectively, lowest and highest score. For the FOEX metrics (Figure C-5, top left), INITD3 and INITD4 showed the best scores with -3% and -7.9% respectively. INITD2 scored the poorest with a -22% FOEX score followed by INITD1 with -18.3%. The two puff configurations had the poorest NMSE and FB statistical performance metrics (with values of approximately 72.7 and 63.6 pg m^{-3} for error and 1.45 and 1.41 for FB). INITD3 exhibited the best overall scores for both NMSE and FB (16.6 pg m^{-3} and 0.88 respectively. INITD1 and INITD2 exhibited the poorest KSP scores with 44% and 48% respectively. INITD104 had the best KSP score with 28%, followed by INITD4 (30%), INITD0 and INIT130 (31%), and INITD140 (32%).

For the within a factor of 2 and 5 metric (FA2 and FA5, Figure C-6, top), the puff INITD configurations performed the poorest with scores between 0% - 1% for FA2 and 1% - 4.8% for FA5. INITD0 showed the best FA2/FA5 scores with 6.9%/11.8%, followed by INITD130 and INITD140 for FA2 and INITD3 and INITD4 for FA5. Curiously, INITD3 and INITD4 had slightly lower FA2 scores (3.4%/2.8%) than the other puff-particle hybrid configurations, but higher FA5 scores. For the PCC metric (PCC, Figure C-6, bottom left), INITD3 had the highest score with 0.63, followed closely by the other puff-particle or particle configurations ranging from 0.51 (INITD0) to 0.62 (INITD2).

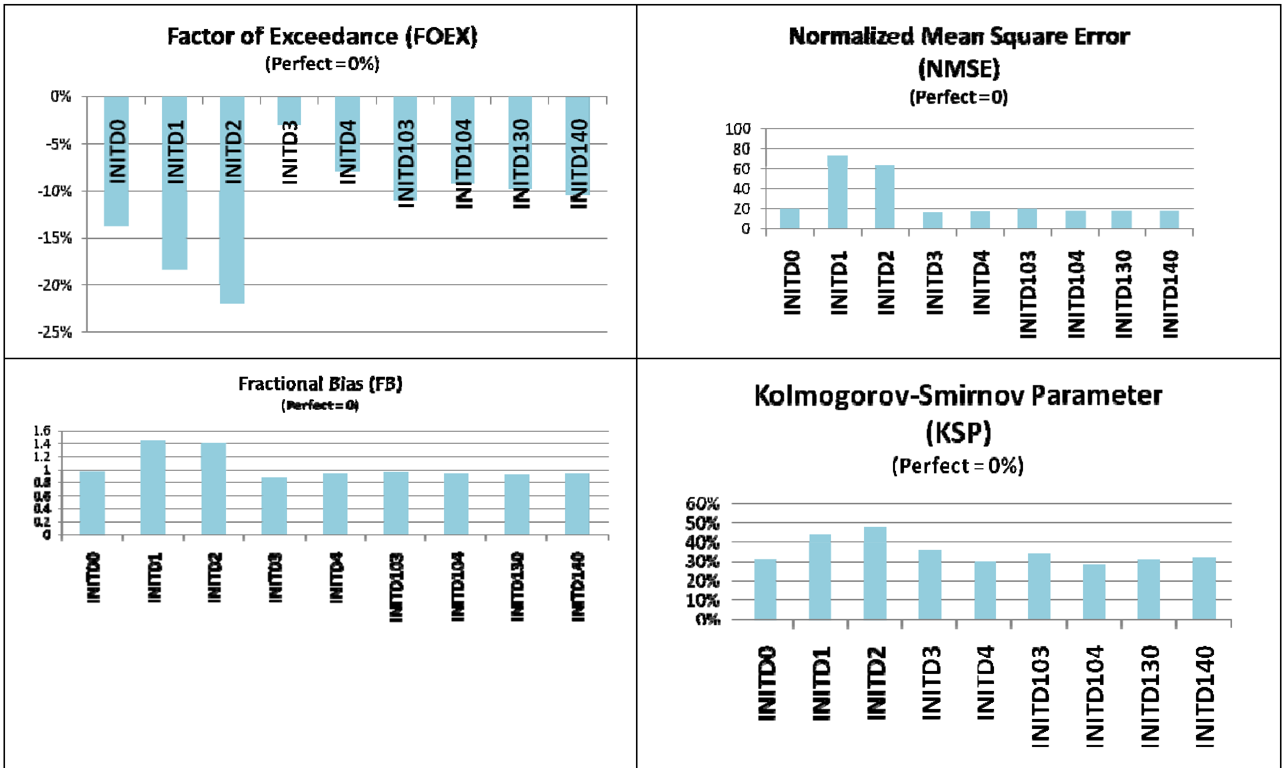


Figure C-5. Global model performance statistics for nine HYSPLIT INITD sensitivity tests for CAPTEX Release 5.

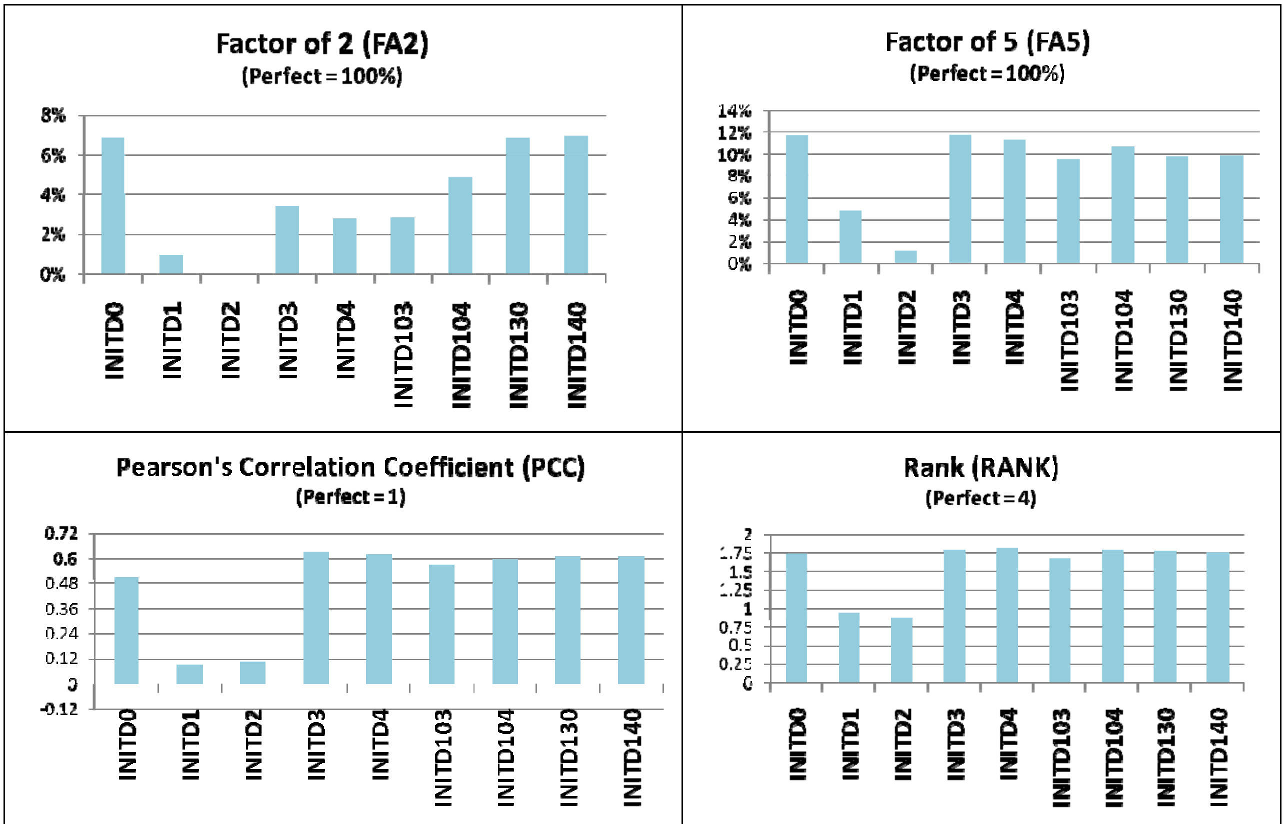


Figure C-6. Global model performance statistics for nine HYSPLIT INITD sensitivity tests for CAPTEX Release 5.

The final panel in Figure C-6 (bottom right) displays the overall RANK statistic. The RANK statistics orders the model performance of the HYSPLIT INITD configurations are as follows:

1. INITD4 (1.82)
2. INITD104 (1.80)
3. INITD3 (1.79)
4. INITD130 (1.78)
5. INITD140 (1.76)
6. INITD0 (1.75)
7. INITD103 (1.68)
8. INITD1 (0.94)
9. INITD2 (0.88)

C.3 CAMX SENSITIVITY TESTS

Following the general design of the study for the ETEX tracer database, CAMx sensitivity tests described in Section 6.4.3, thirty-two CAMx sensitivity tests were conducted to investigate the effects of vertical diffusion, horizontal advection solvers and use of the sub-grid scale Plume-in-Grid (PiG) module on the model performance for the CAPTEX tracer experiment releases 3 and 5. In addition to the sixteen sensitivities conducted for ETEX, a similar set of sensitivity analyses were conducted using the newer ACM2 vertical diffusion scheme (Pleim, 2007; ENVIRON, 2010) introduced into CAMx as of Version 5.20 as an alternative to the more traditional fully K-theory vertical diffusion schemes that were the only options available in previous versions of CAMx.

C.3.1 SPATIAL PERFORMANCE FOR CTEX3 NOPIG EXPERIMENTS

Figure C-7 displays the CAMx spatial model performance statistics for the sensitivity tests that were run without using the PiG subgrid-scale puff module. For the FMS statistic, the ACM2, TKE, and CMAQ Kz exhibit very similar performance (40.9%, 40.9%, and 39.4% respectively). OB70 exhibits the poorest performance with 33.5% for FMS.

For the FAR statistic, ACM2/Bott has the best score (55.6%) followed by TKE/PPM and ACM2/PPM (tied at 58.5%). Overall ACM2 is the best performing vertical diffusion formulation and PPM performs better than BOTT for horizontal advection using the FAR statistic.

For the POD and TS spatial statistics, the CMAQ, TKE, and ACM2 vertical diffusion algorithms perform similarly, and all are substantially better than the OB70 approach (15% lower than other vertical diffusion schemes). ACM2/BOTT has the best TS score with 28.6% followed by ACM2/PPM and TKE/PPM (tied at 28.33%). CMAQ/PPM exhibits the best POD score with 52.8% followed by CMAQ/BOTT, ACM2/PPM, and TKE/PPM (tied at 47.2%). Consistent with the ETEX spatial results, there are much smaller differences in the model performance using the two advection solvers for the POD and TS statistics compared to differences between Kz options.

In summary, based on the spatial statistics, the ACM2, CMAQ, and TKE Kz algorithms appear to be performing similarly, with the older OB70 option exhibiting much poorer overall performance. The differences in vertical diffusion algorithms have a greater effect on CAMx model performance than the differences in horizontal advection solvers.

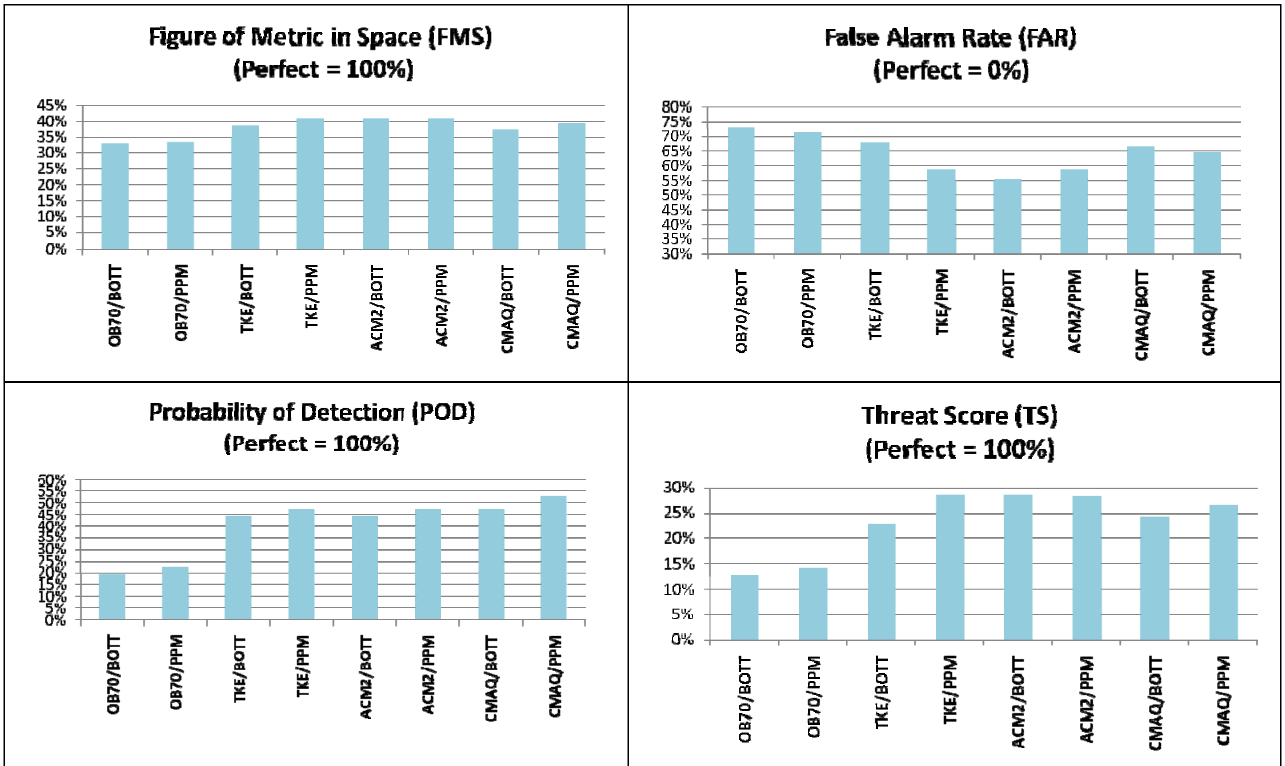


Figure C-7. Spatial model performance statistics for the CAMx sensitivity tests without using the PiG subgrid-scale puff module (NoPiG) for CAPTEX Release 3.

C.3.2 Global Statistical Performance for CTEX3 NoPiG Experiments

Figures C-8 and C-9 displays the global statistics for the CAMx NoPiG sensitivity and contain the statistical metrics where the best performing model has the, respectively, lowest and highest score. For the FOEX metric, the vertical diffusion algorithm that has the best score was the ACM2 algorithm with scores of 3.26% (ACM2/BOTT) and 3.6% (ACM2/PPM). The next best vertical diffusion algorithm/advection solver combination was OB70/BOTT with a -3.8% FOEX score. The CMAQ/BOTT (9.1%) and CMAQ/PPM (9.6%) have the highest (worst) FOEX scores.

For the NMSE statistical performance metric, the ACM2 and TKE vertical diffusion schemes perform best (28 to 30 pgm^{-3}) (Figure C-8, top right). Both OB70 scenarios yielded the poorest NMSE scores with error values more than twice that of the other Kz/advection solver configurations. Consistent with both FOEX and NMSE, the ACM2 vertical diffusion scheme is also the best performing method according to the FB and KSP metrics followed by the TKE scheme (Figure C-8, bottom left). The OB70 vertical diffusion algorithm performs the poorest for both of the FB and KSP metrics.

For the within a factor of 2 and 5 metrics (FA2/FA5, Figure C-9, top), the ACM2 combinations are the best performing with values of 8.8%/17.1% and 8.2%/16.3%. The TKE and CMAQ combinations perform similarly, with the TKE options having slightly higher FA2 percentages, but the CMAQ combinations exhibit higher FA5 percentages than the TKE.

For the PCC metric, the CMAQ and TKE combinations yield the best correlation performance with values ranging from 0.54 to 0.63 with CMAQ having slightly higher correlation values overall. Interestingly, for most other spatial and global statistical categories for the NoPiG tests

ACM2 Kz combinations rank as the best performing. However, the ACM2 Kz combinations have the lowest PCC correlation values of the four Kz combinations, with values of 0.22 and 0.30.

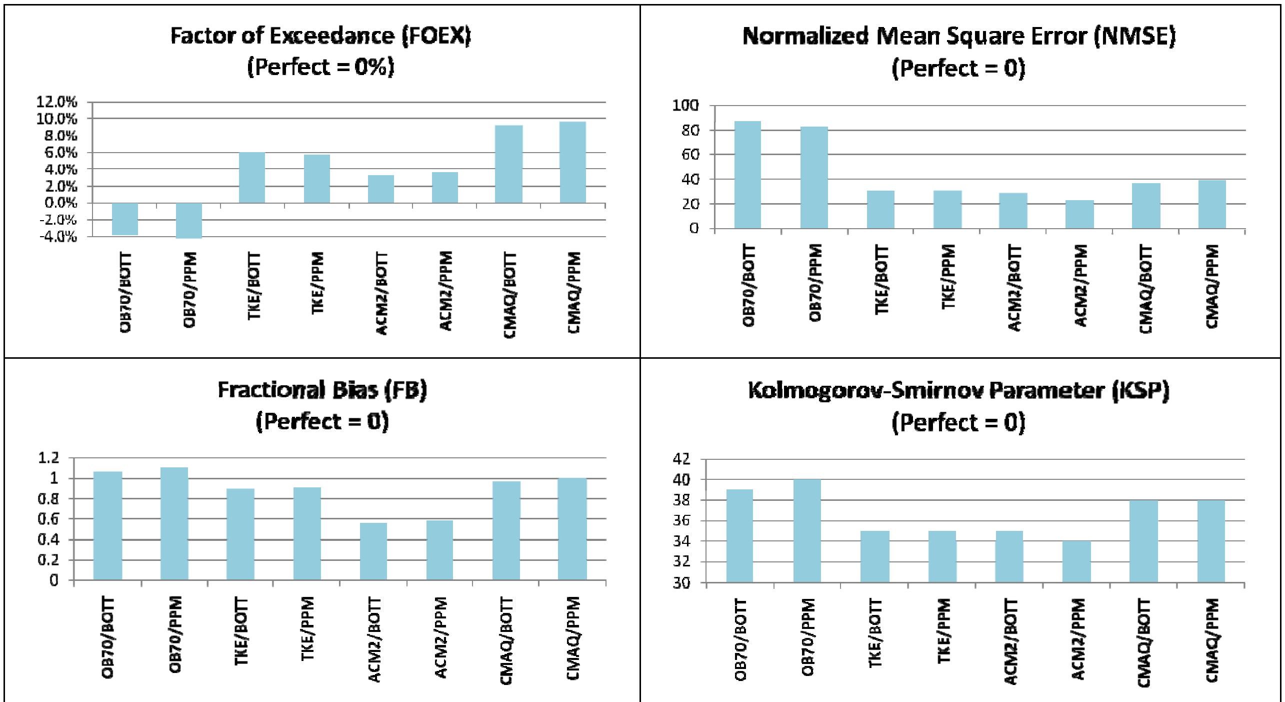


Figure C-8. Global model performance statistics for the CAMx sensitivity tests without using the PiG subgrid-scale puff module for CAPTEX Release 3.

The final panel in Figure C-9 (bottom right) displays the overall RANK statistic. The RANK statistics orders the model performance of the CAMx configurations without PiG as follows:

1. TKE/PPM (1.97)
2. CMAQ/PPM (1.91)
3. TKE/BOTT (1.89)
4. ACM2/PPM (1.87)
5. CMAQ/BOTT (1.83) (tied)
6. ACM2/BOTT (1.83) (tied)
7. OB70/PPM (1.67)
8. OB70/BOTT (1.56)

Based on this analysis, the TKE Kz coefficients is the best performing vertical diffusion approach followed closely by CMAQ. As noted previously, the vertical diffusion algorithm has a greater effect on CAMx model performance compared to the choice of horizontal advection solvers.

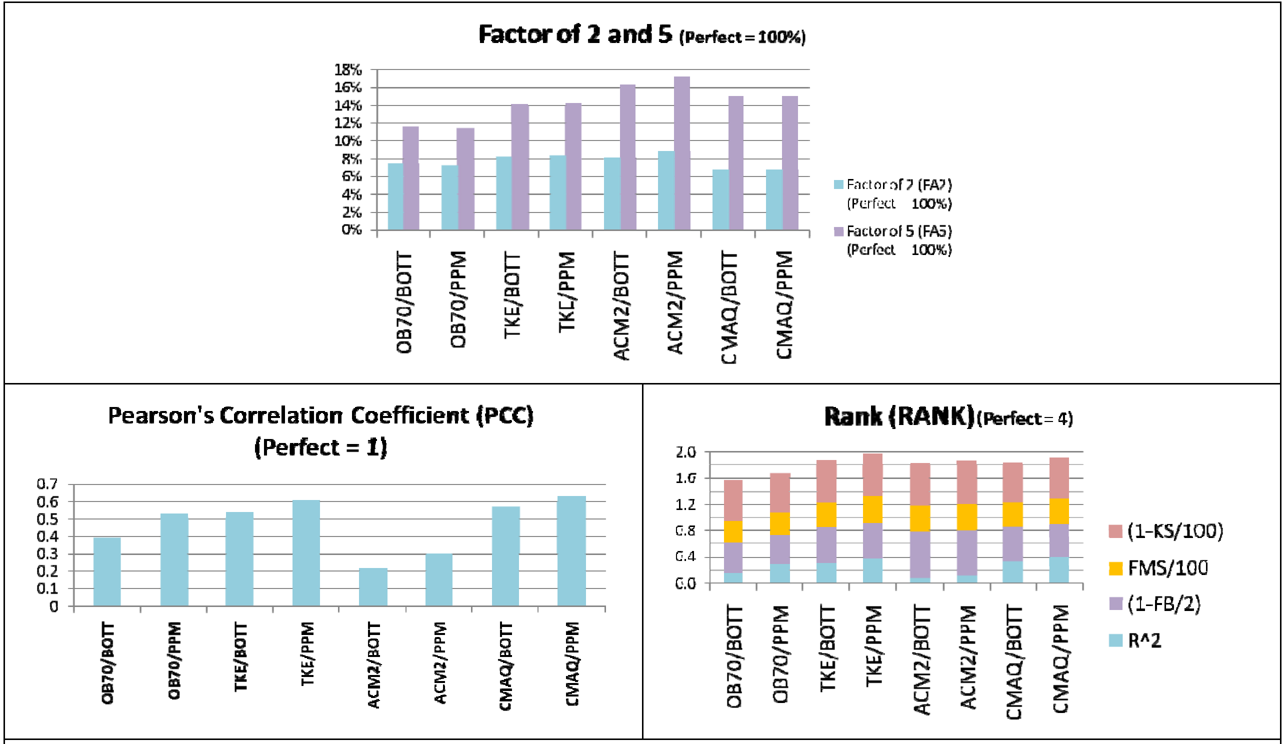


Figure C-9. Global model performance statistics for the CAMx sensitivity tests without using the PiG subgrid-scale puff module for CAPTEX Release 3.

C.3.3 SPATIAL PERFORMANCE FOR CTEX3 PIG EXPERIMENTS

Figure C-10 displays the CAMx spatial model performance statistics for the sensitivity tests that were run using the PiG subgrid-scale puff module. For the FMS statistic, the CMAQ Kz/BOTT combination has the best performance at 36.2%. TKE/BOTT and TKE/PPM followed closely with FMS scores of 36% and 35.6% respectively. OB70 exhibits the poorest performance with 23.5% for FMS.

For the FAR statistic, ACM2/Bott has the best score (61.5%) followed by TKE/BOTT and ACM2/PPM (62.8% and 63% respectively). OB70 exhibited the poorest performance with a 72% FAR. For the POD metric, the TKE/BOTT combination performs substantially better than the other Kz/advection solver combinations (44.4%). Both OB70 scenarios showed the poorest performance at 13.9%. For the TS spatial metric, the CMAQ/PPM exhibits the best score with 30.6% followed by TKE/BOTT and CMAQ/BOTT(25.4% and 23.3% respectively). OB70 again performs poorest with a TS value of 10.2% for both advection solver combinations. Consistent with the ETEX spatial results, there are much smaller differences in the model performance using the two advection solvers for the POD and TS statistics compared to differences between Kz options.

In summary, the effect of using the CAMx subgrid scale puff module appears to slightly degrade performance in comparison to the NoPiG experiments. A similar pattern was noted in the spatial statistics compared to the NoPiG experiments with the ACM2, CMAQ, and TKE Kz algorithms performing similarly, with the older OB70 option exhibiting much poorer overall performance.

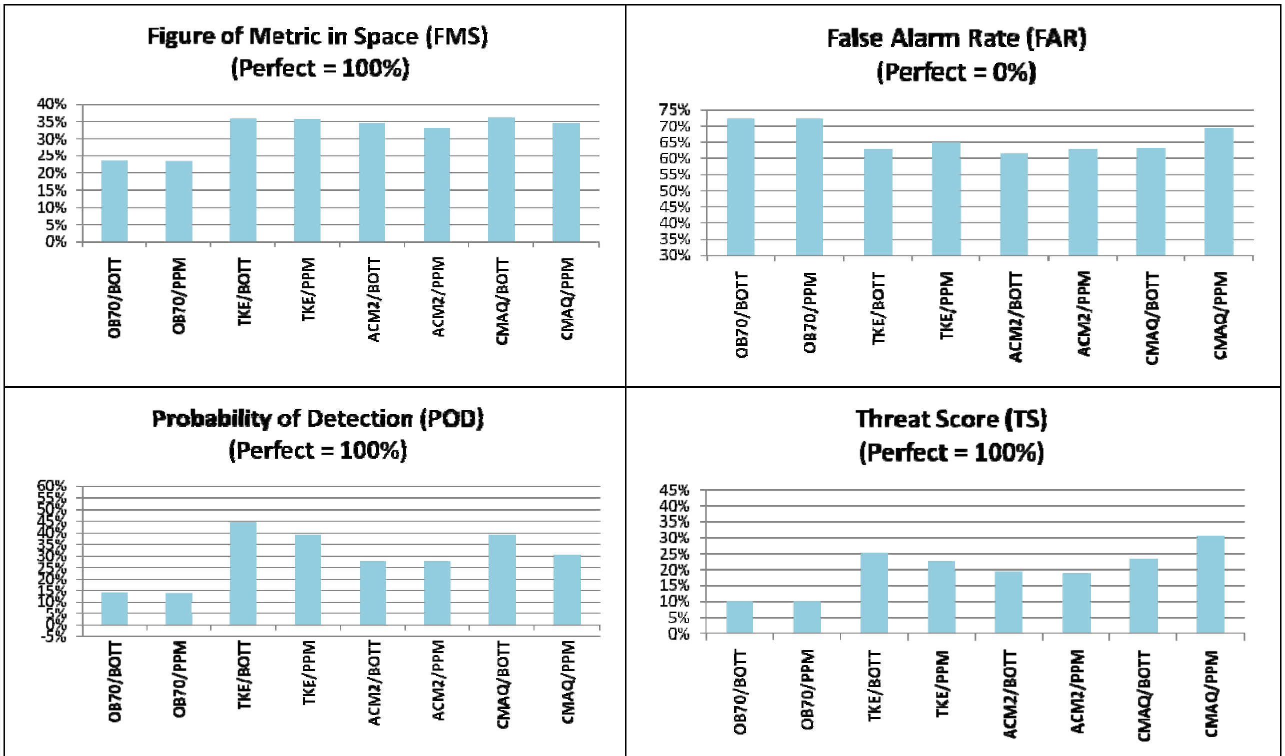


Figure C-10. Spatial model performance statistics for the CAMx CTEX3 sensitivity tests using the PiG subgrid-scale puff module (PiG).

C.3.4 GLOBAL STATISTICAL PERFORMANCE FOR CTEX3 CAMX PIG SENSITIVITY TESTS

Figures C-11 and C-12 displays the global statistics for the CAMx NoPiG sensitivity tests with the two figures containing the statistical metrics where the best performing model has the, respectively, lowest and highest score. For the FOEX metric, the vertical diffusion algorithm has the biggest effect was the TKE and CMAQ algorithms with scores near zero. The OB70 combinations exhibit significantly poorer FOEX performance with values of -13.6% (OB70/BOTT) and -16.9% (OB70/PPM).With the NMSE statistical performance metric, the ACM2 and TKE vertical diffusion schemes performs best (38 – 42 pgm^{-3}) (Figure C-11, top right). Both OB70 scenarios yielded the poorest scores with error values more than twice that of the other Kz/advection solver configurations (111 – 116 pgm^{-3}). For fractional bias (Figure C-11, bottom left), the ACM2 combinations have the best scores with 0.58/0.59 (BOTT/PPM). TKE Kz combinations follow with values 0.82/0.83 (BOTT/PPM). OB70 again has the poorest FB performance with values of 1.18/1.19 (BOTT/PPM).

For the within a factor of 2 and 5 metrics (FA2 and FA5, Figure C-12, top), ACM2, TKE, and CMAQ are very similar for FA2, but the TKE Kz option clearly performs best for the FA5 metric, followed by CMAQ. There is essentially no difference in the PCC statistic using the two horizontal advection solvers. According to the PCC metric (Figure C-11, bottom right), CMAQ is the best performing vertical diffusion approach (0.52/0.59 – BOTT/PPM) followed by TKE (0.33/0.44 – BOTT/PPM)ACM2 has the lowest PCC values with scores 0.17 – 0.23 (BOTT/PPM).

The final panel in Figure C-12 (bottom right) displays the overall RANK statistic. The RANK statistics orders the model performance of the CAMx configurations with PiG as follows:

1. CMAQ/PPM (1.86)

2. TKE/PPM (1.83)
3. CMAQ/BOTT (1.81)
4. TKE/BOTT (1.74)
5. ACM2/BOTT (1.69)
6. ACM2/PPM (1.68)
7. OB70/PPM (1.26)
8. OB70/BOTT (1.25)

Based on this analysis, the CMAQ Kz coefficients are the best performing vertical diffusion approach followed closely by TKE. Consistent with the NoPiG experiments, the vertical diffusion algorithm has a greater effect on CAMx model performance compared to the choice of horizontal advection solvers.

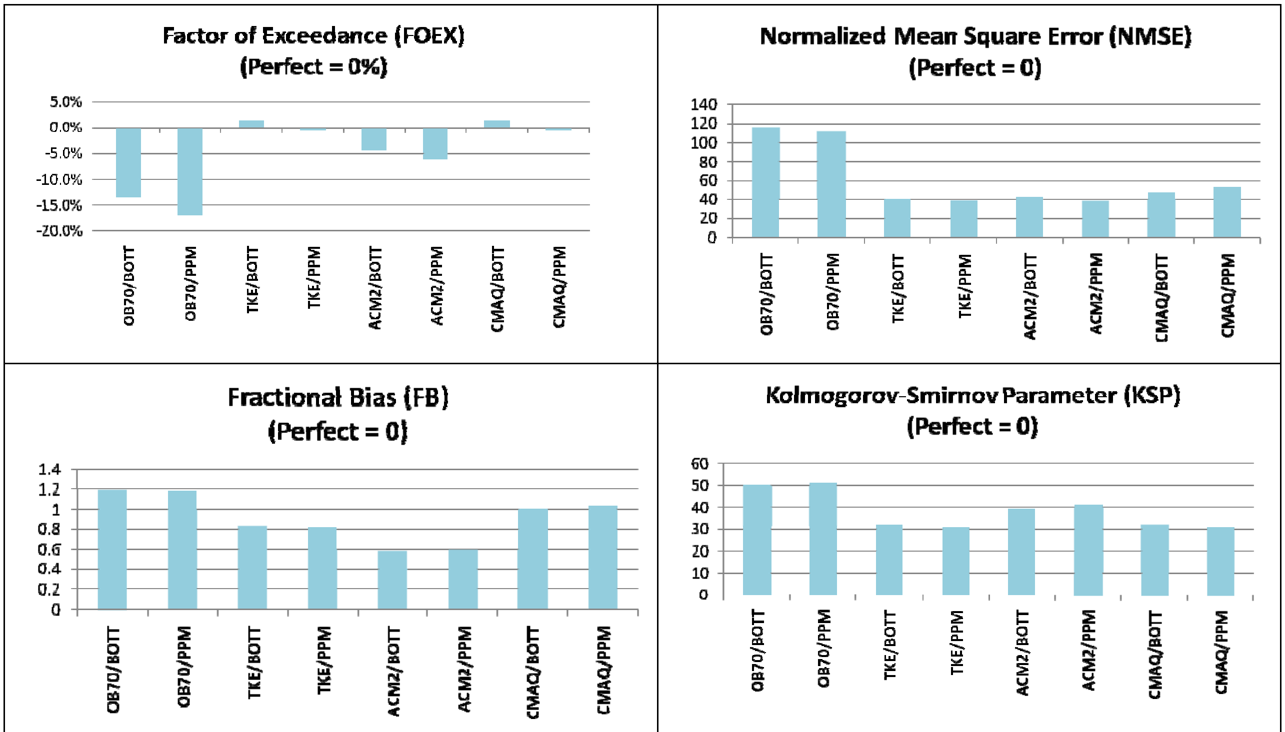


Figure C-11. Global model performance statistics for the CAMx CTEX3 sensitivity tests using the PiG subgrid-scale puff module.

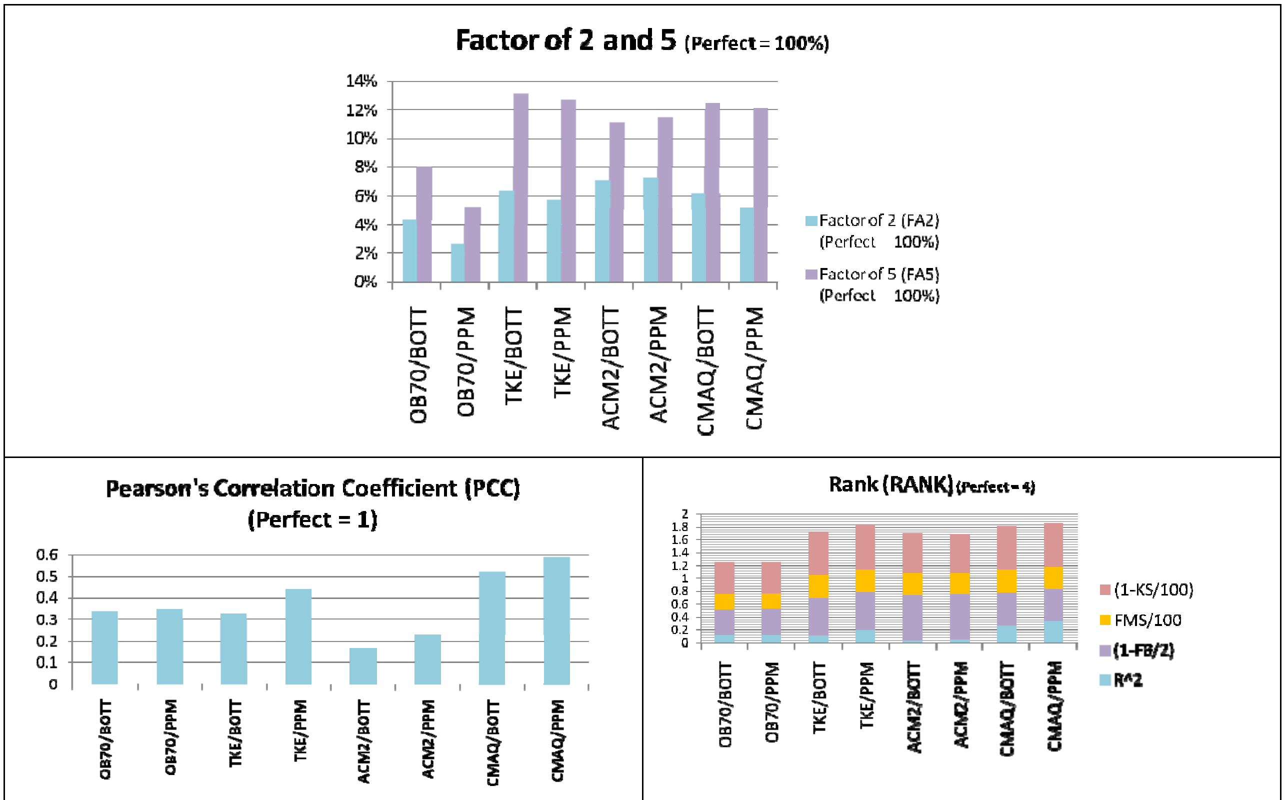


Figure C-12. Global model performance statistics for the CAMx CTEX3 sensitivity tests using the PiG subgrid-scale puff module.

C.3.4.1 EFFECT OF PIG ON MODEL PERFORMANCE

The effect of using the PiG module versus NoPiG results in similar results as seen in the ETEX experiments for CAMx; use of the subgrid-scale PiG module has very little effect on the CAMx model performance and the rankings of the CAMx model performance using the alternative vertical mixing and horizontal advection approaches. In general, it appears that the CAMx model performance without the PiG is performing slightly better than its performance using the PiG.

The spatial performance statistics are sometimes improved and sometimes degraded when the PiG module is invoked. For the global statistics, the PCC performance statistic is degraded by -11% to -37% (-0.03 to -0.13 points) when the PiG module is invoked. Similarly, use of the PiG versus NoPiG module increases (degrades) the FB metric by 5 to 18 percent and also increases (degrades) the NMSE metrics for all model configurations.

Table C-2 summarizes the RANK model performance statistic for the different CAMx model configurations with and without the PiG module. For each model vertical diffusion/horizontal advection configuration, using the PiG module always results in slightly lower RANK statistics that are from -3.9% to -8.5% lower than when the PiG module is not used.

Table C-2. CTEX3 CAMx RANK model performance statistic and model rankings for different model Kz/advection solver configurations with and without using the PiG subgrid-scale puff model.

Model Configuration	Without PiG Module		With PiG Module		PiG-NoPiG	
	RANK	Model Ranking	RANK	Model Ranking	ΔRANK	Percent
OB70/BOTT	1.56	8	1.25	8	-0.41	-26.4%
OB70/PPM	1.67	7	1.26	7	-0.41	-24.5%
TKE/BOTT	1.89	3	1.74	4	-0.15	-8.0%
TKE/PPM	1.97	1	1.83	2	-0.14	-7.1%
ACM2/BOTT	1.83	6[a]	1.69	5	-0.14	-7.6%
ACM2/PPM	1.87	4	1.68	6	-0.19	-10.1%
CMAQ/BOTT	1.83	5[a]	1.81	3	-0.02	-1.0%
CMAQ/PPM	1.91	2	1.86	1	-0.05	-2.6%

[a] tied

C.3.5 SPATIAL PERFORMANCE FOR CTEX5 NOPIG EXPERIMENTS

Spatial performance for CTEX5 and the NoPiG CAMx sensitivities posed a slightly more difficult challenge to interpret due to similarities amongst the Kz/advection solver options for the FMS metric (Figure C-13). The range of difference for the FMS between the minimum and maximum for all of the eight combinations was less than 2.1%, with all in the range of 22% to 24%. This would indicate that each of the model configurations performs similarly across all concentration ranges. However, in the extended spatial statistics of FAR, POD, and TS, greater differentiation in model spatial performance metrics are seen. For example, for POD and TS, the CMAQ Kz combinations perform best of all of the vertical diffusion options, and have POD/TS statistics that are nearly twice as good as the OB70 diffusion combinations with POD/TS values of ~60%/~22% for CMAQ versus ~33%/~10% for OB70 diffusion algorithm options. Since these statistics are valid for concentration ranges above the 100 pg m^{-3} concentration level, similarity in model performance for the FMS metric is likely due to better performance of OB70 and ACM2 at levels below the threshold concentration used for the FAR, POD and TS statistics. Above the concentration threshold spatial performance for OB70 and ACM2 lags behind that of the TKE and CMAQ, indicating that the TKE and CMAQ Kz options perform better across all concentration ranges compared to similar performance at the lower concentration levels below the threshold. Overall, it appears that the CMAQ Kz option yields the best performance of the diffusion options when examining the performance across all of the spatial metrics.

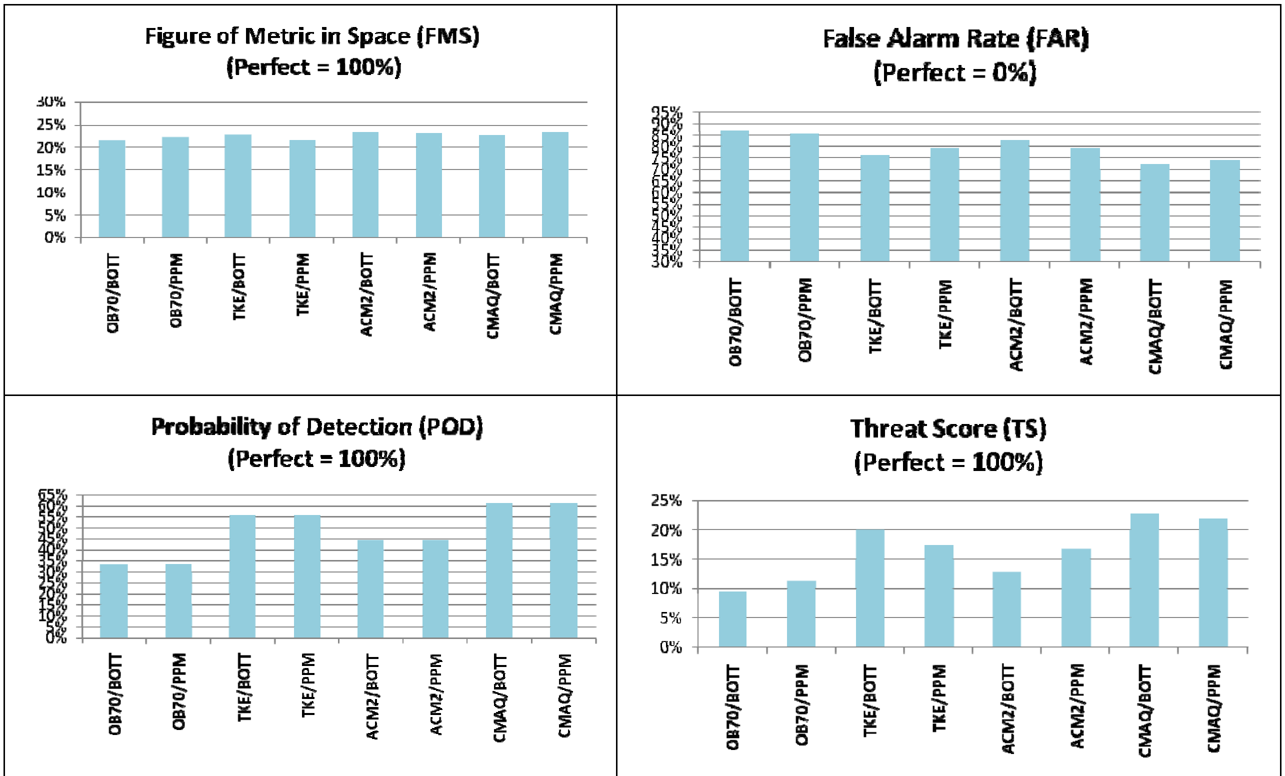


Figure C-13. Spatial model performance statistics for the CAMx CTEX5 sensitivity tests without using the PiG subgrid-scale puff module (NoPiG).

C.3.6 Global Statistical Performance for CTEX5 NoPiG Experiments

Figures C-14 and C-15 displays the global statistics for the CAMx NoPiG CTEX5 sensitivity tests for the statistical metrics with the best performing model has the, respectively, lowest and highest score. For the FOEX metric, all of the Kz/advection solver options are within 4% of each other (1% - 5%), with the best performance coming from OB70 and TKE options and degrading slightly across the ACM2 and CMAQ options.

The NMSE, FB, and PCC metrics provide clear differentiation in performance across the Kz options, with the TKE and CMAQ options yielding significantly better performance than either OB70 or ACM2. For NMSE, the CMAQ combinations have the best scores with 9.3 – 9.4 pg m^{-3}, followed by the TKE combinations with 12.6 – 13.6 pg m^{-3}. NMSE values are nearly double that for OB70 and ACM2. A similar relationship is found with the FB and PCC metrics, with the CMAQ and TKE performing significantly better than either ACM2 or OB70.

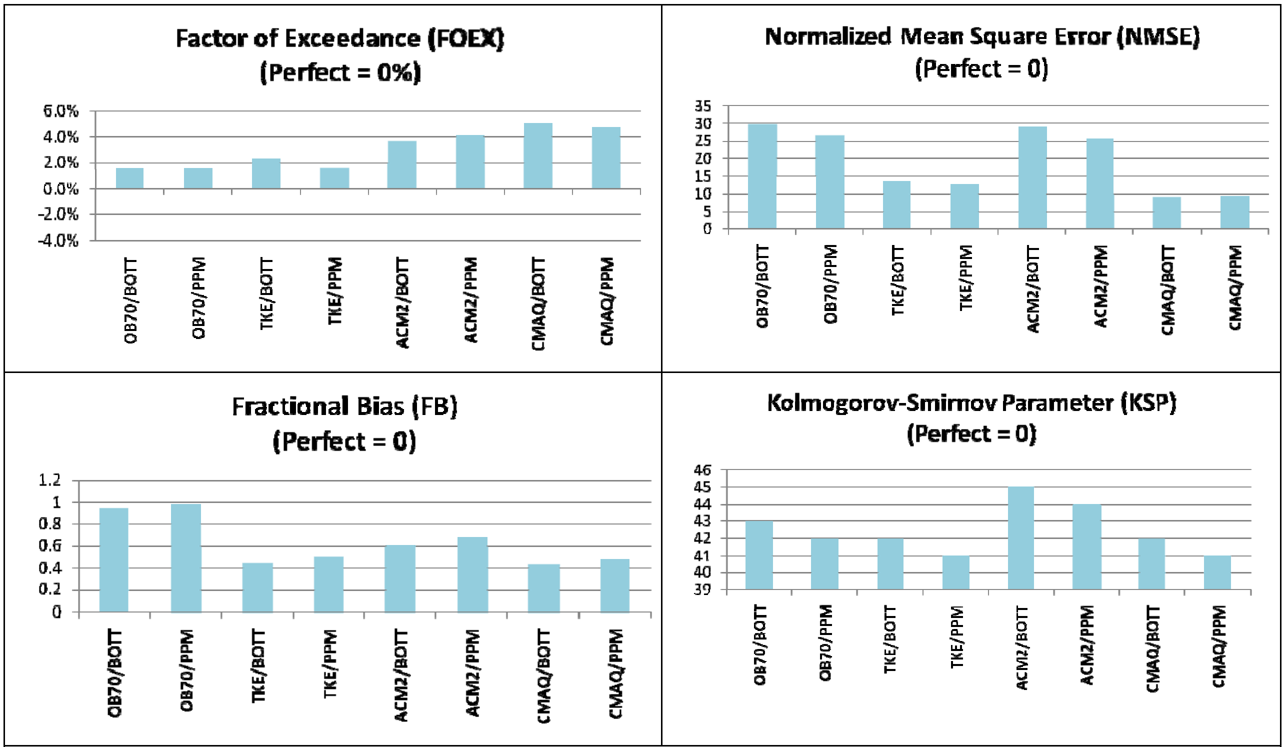


Figure C-14. Global model performance statistics for the CAMx CTEX5 sensitivity tests using the PiG subgrid-scale puff module.

The final panel in Figure C-15 (bottom right) displays the overall RANK statistic. The RANK statistics orders the model performance of the CTEX5 CAMx configurations without PiG as follows:

1. CMAQ/PPM (1.92) (tied)
2. CMAQ/BOTT (1.92) (tied)
3. TKE/PPM (1.73)
4. TKE/BOTT (1.71)
5. ACM2/BOTT (1.50)
6. ACM2/PPM (1.48)
7. OB70/PPM (1.34)
8. OB70/BOTT (1.33)

As with CTEX3, for the CTEX5 experiment the CMAQ Kz algorithm is the best performing vertical mixing approach in CAMx based on both the spatial and global statistical analyses. What differs in the CAMx CTEX3 and CTEX5 NoPiG sensitivity test performance is the composition of the RANK metric. Spatial performance for CTEX3 was significantly better than for CTEX5 (10% - 15% greater), thus FMS contributes less to the RANK statistical metric for CTEX5 compared to CTEX3. Similarly, the PCC metric is much more variable across the Kz options with CTEX5 experiment, with essentially no contribution of PCC to the RANK score for OB70 and ACM2 Kz options in CTEX5. Additionally, the KSP scores comprise a much greater portion of the RANK scores for CTEX5.

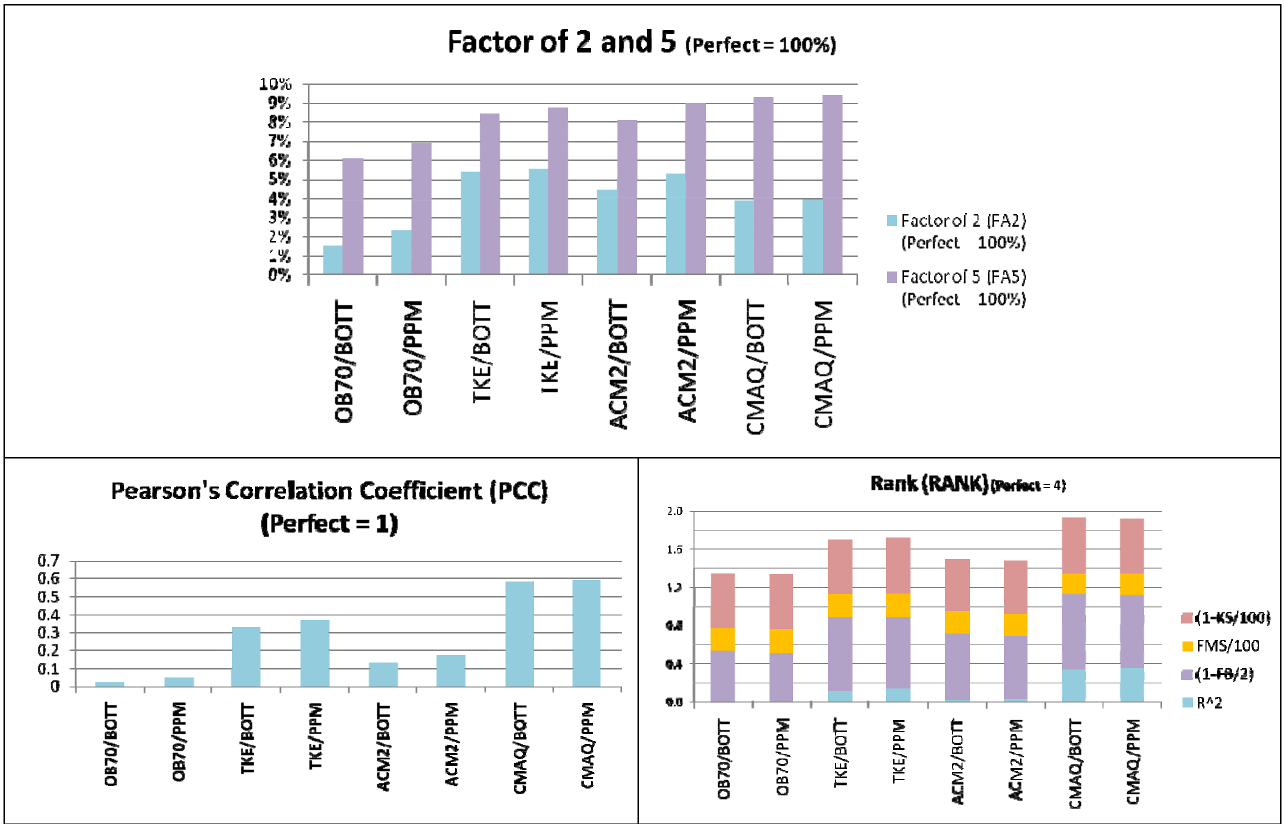


Figure C-15. Global model performance statistics for the CAMx sensitivity tests without using the PiG subgrid-scale puff module for CAPTEX Release 5 (NoPiG).

C.3.7 SPATIAL PERFORMANCE FOR CTEX5 PIG EXPERIMENTS

As with the CTEX5 NoPiG experiments, interpretation of the spatial performance for CTEX5 for the PiG CAMx sensitivities posed a slightly more difficult challenge to interpret due to similarities amongst the Kz/advection solver options for the FMS metric (Figure C-16). The range of difference for the FMS between the minimum and maximum for all of the eight combinations was less than 2%, with all in the range of 21.5% - 23.3%, noting a slight degradation across the board from the corresponding NoPiG experiments (0.2% - 3.1%).

Examination of the extended spatial statistics reveals a similar pattern in performance compared to the NoPiG equivalent tests. Greater differences in performance are observed across the various Kz/advection solver combinations, especially for the POD and TS metrics. For both of these metrics, the CMAQ Kz combinations clearly yield better spatial performance than the other Kz options (5% - 10% better for POD and 2% - 5% for TS than the second best Kz option (TKE)).

Consistent with results from the NoPiG scenarios, the CMAQ Kz option appears to perform best across the majority of the spatial metrics. The largest difference in spatial performance is determined by the user's selection of the Kz option than either the choice of advection solver or use of the subgrid scale PiG module in CAMx.

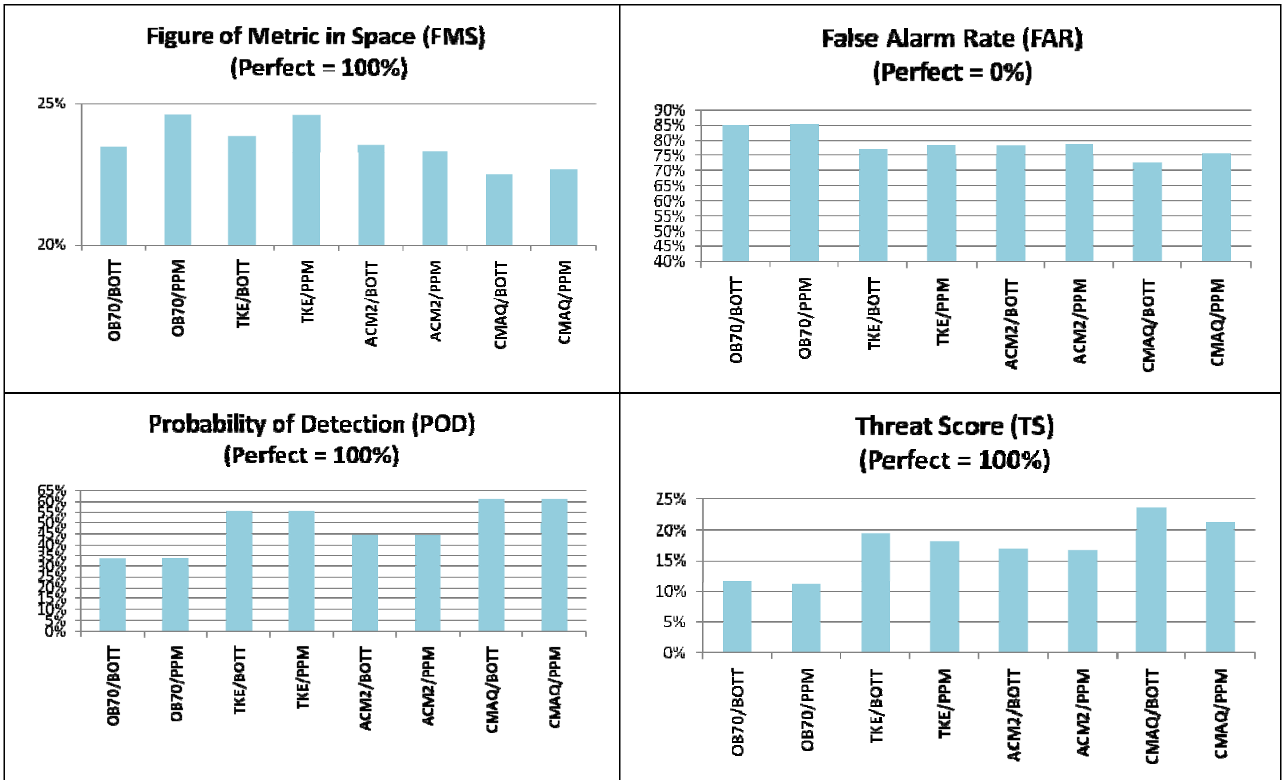


Figure C-16 Spatial model performance statistics for the CAMx CTEX5 sensitivity tests using the PiG subgrid-scale puff module (PiG).

C.3.8 Global Statistical Performance for CAMx CTEX5 PiG Experiments

Figures C-17 and C-18 displays the global statistics for the CAMx sensitivity tests using the PiG module with statistical metrics for the best performing model has the, respectively, lowest and highest score. For the FOEX metric, all of the Kz/advection solver options are within 5% - 6% of each other (-1.2% - 4.7%), with the best performance coming from OB70 and TKE options and degrading slightly across the ACM2 and CMAQ options, which is largely consistent with the equivalent NoPiG scenarios.

For NMSE and KSP, the CMAQ and TKE options perform better than either OB70 or ACM2. The CMAQ Kz option has the best NMSE values with 9.9 – 10.8 pg m^{-3} (PPM/BOTT), followed by TKE with values of 14.2 – 15.8 pg m^{-3} (PPM/BOTT). The TKE/PPM combination had the best scores for KSP, followed by CMAQ/PPM. All of the Kz options save OB70 had very similar FB and FA2/5 scores. OB70 consistently scored the poorest across all of the global statistical metrics.

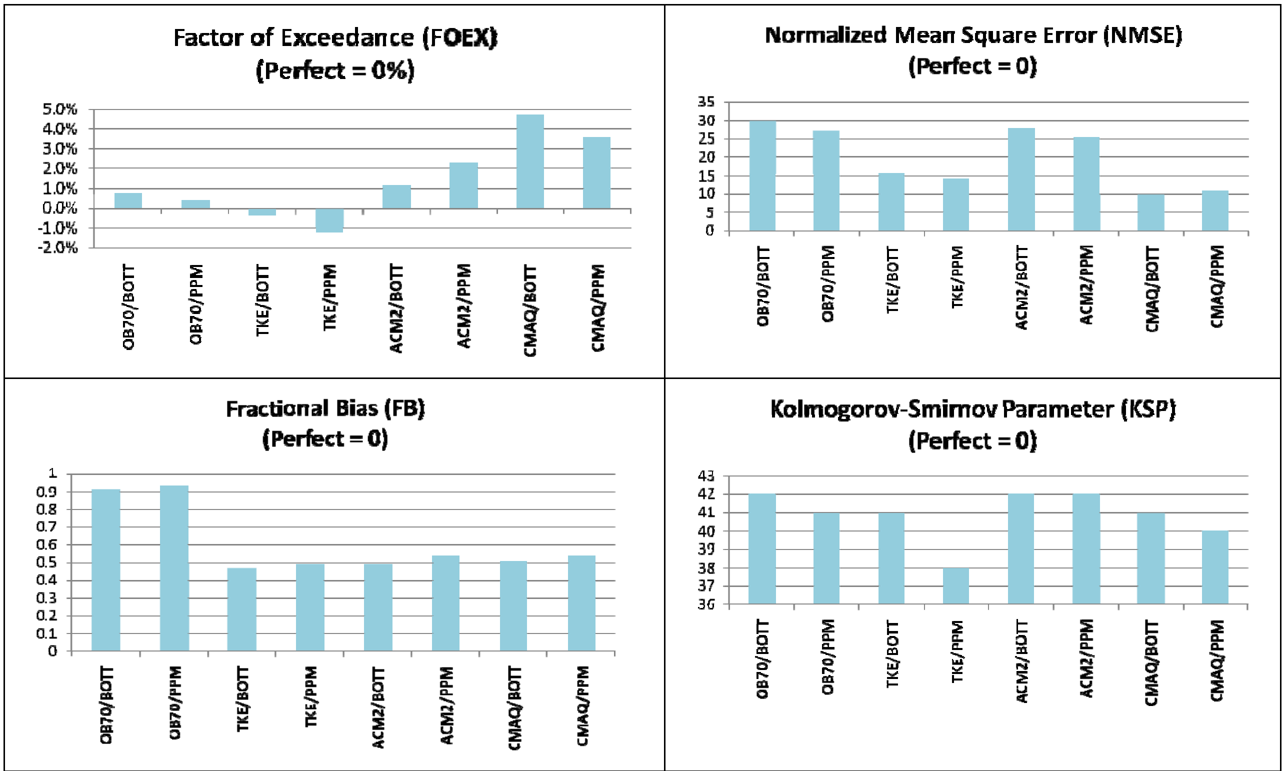


Figure C-17. Global model performance statistics for the CAMx sensitivity tests using the PiG subgrid-scale puff module for CAPTEX Release 5.

The final panel in Figure C-17 (bottom right) displays the overall RANK statistical metric. The RANK statistics orders the model performance of the CAMx configurations using the PiG module as follows:

1. CMAQ/BOTT (1.95)
2. CMAQ/PPM (1.92)
3. TKE/PPM (1.67)
4. TKE/BOTT (1.65)
5. ACM2/BOTT (1.58)
6. ACM2/PPM (1.56)
7. OB70/PPM (1.35)
8. OB70/BOTT (1.34)

Consistent with the NoPiG scenarios for CTEX5, CAMx performance using the CMAQ Kz option for vertical mixing is the best performing vertical diffusion algorithm overall for both the spatial and global statistical analyses and the choice of advection solver has a much smaller effect on model performance compared to vertical diffusion.

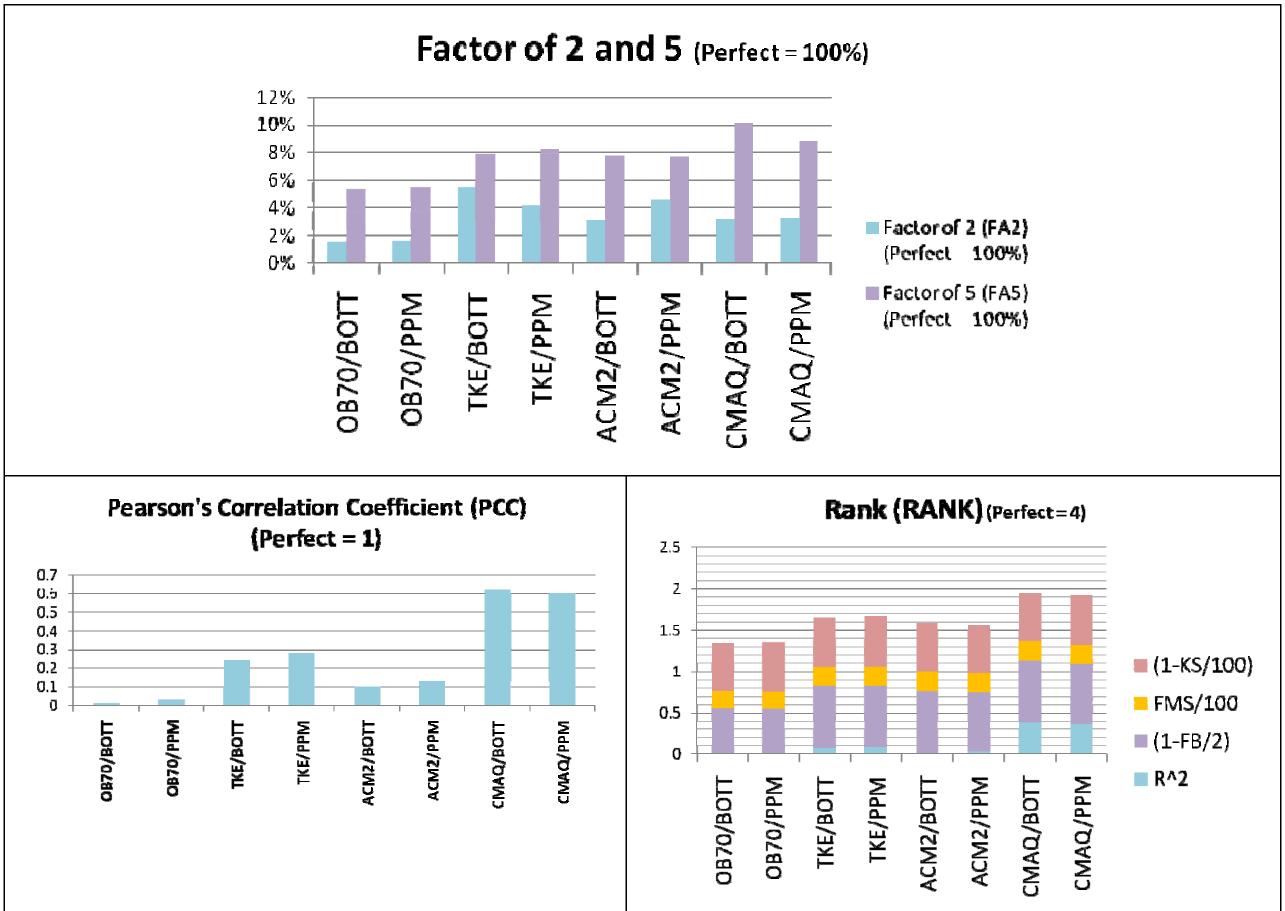


Figure C-18. Global model performance statistics for the CAMx sensitivity tests using the PiG subgrid-scale puff module for CAPTEX Release 5 (PiG).

C.3.8.1 EFFECT OF PIG ON MODEL PERFORMANCE FOR CTEX5

Similar to the results from the ETEX and CTEX3 experiments for CAMx, whether the PiG is used or not has very little effect on the CAMx model performance and the rankings of the CAMx model performance using the alternative vertical mixing and horizontal advection solver options. In general, it appears that the CAMx model performance without the PiG is performing slightly better than its performance using the PiG.

The spatial performance statistics are sometimes improved and sometimes degraded when the PiG module is invoked. Table C-3 examines the effect of PiG treatment of the tracer using two of the four spatial statistics, FMS and POD. Slight degradation of spatial performance when the PiG module is invoked is noted using the OB70, TKE, and ACM2 Kz diffusion combinations (from -3.1% to -0.2% for FMS and from -11.1% to 0% for POD). However, CAMX using the CMAQ Kz diffusion/advection solver combinations experienced a 0.2% to 0.5% improvement for FMS and no change for POD.

Table C-3. CAMx FMS and POD spatial performance statistic and model rankings for different model configurations with and without using the PiG subgrid-scale puff model for CAPTEX Release 5.

Model Configuration	Without PiG Module		With PiG Module		NoPiG-PiG	
	FMS	POD	FMS	POD	ΔFMS	ΔPOD
OB70/BOTT	23.48	33.33	21.54	27.78	-1.9%	-5.6%
OB70/PPM	24.62	33.33	22.05	33.33	-2.6%	0%
TKE/BOTT	23.85	55.56	22.83	55.56	-1.0%	0%
TKE/PPM	24.6	55.56	21.49	50	-3.1%	-5.6%
ACM2/BOTT	23.53	44.44	23.26	33.33	-0.3%	-11.1%
ACM2/PPM	23.31	44.44	23.08	44.44	-0.2%	0%
CMAQ/BOTT	22.48	61.11	22.66	61.11	0.2%	0%
CMAQ/PPM	22.66	61.11	23.2	61.11	0.5%	0%

Table C-4 summarizes the RANK model performance statistic for the different CAMx model configurations with and without using the PiG module. The results for the global statistics are somewhat varied across the Kz/advection solver configurations. OB70, ACM2, and CMAQ/BOTT showed slight improvements in their RANK score when using the PiG module (improvements ranged from 0.7% to 5.4%). However, the TKE combinations experienced performance degradations with changes ranging from -3.5% to 4.0%.

Table C-4. CAMx RANK model performance statistic and model rankings for different model configurations with and without using the PiG subgrid-scale puff model for CAPTEX Release 5.

Model Configuration	Without PiG Module		With PiG Module		PiG-NoPiG	
	RANK	Model Ranking	RANK	Model Ranking	ΔRANK	Percent
OB70/BOTT	8	1.33	1.34	8	+0.01	+0.7%
OB70/PPM	7	1.34	1.35	7	+0.01	+0.7%
TKE/BOTT	4	1.71	1.65	4	-0.06	-3.5%
TKE/PPM	3	1.73	1.67	3	-0.07	-4.0%
ACM2/BOTT	5	1.50	1.58	5	+0.08	+5.0%
ACM2/PPM	6	1.48	1.56	6	+0.08	+5.4%
CMAQ/BOTT	1[a]	1.92	1.95	1	+0.03	+1.5%
CMAQ/PPM	2[a]	1.92	1.92	2	0.0	0.0%

[a] tied

In general, it is difficult to discern a consistent pattern of performance across the Kz/advection solver combinations when using the CAMx subgrid scale PiG module or not. There appears to be only modest benefit in cases where performance improvement is detected and only modest degradation in model performance when the PiG module causes a worsening of model performance. The CAMx PiG module was originally developed primarily to treat the near-source chemistry of large point source plumes that can be quite different from its surrounding environment. The decision to employ the CAMx puff module relates not so much in improvement advection and diffusion performance, but rather whether or not it is appropriate to allow emissions of ozone and secondary $PM_{2.5}$ precursors from large point sources to be instantaneously mixed into the grid and what impact this would have on local chemical reactions.

C.4 COMPARISON OF SIX LRT DISPERSION MODELING USING CAPTEX RELEASE 3

The model performance of six LRT dispersion models (CALPUFF, SCIPUFF, HYSPLIT, FLEXPART, CAMx and CALGRID) are evaluated using common MM5 meteorological inputs and the CAPTEX Release 3 tracer experiment.

C.4.1 SPATIAL ANALYSIS OF MODEL PERFORMANCE

The performance of the six LRT dispersion models using the four spatial analysis model performance statics that were defined in Section 2.4 are discussed in this section. Figure C-19 displays the FMS spatial performance metrics for the six LRT models and the CTEX3 tracer study field experiment. The CAMx (39.4%) and SCIPUFF (35.2%) models are the two best performing models for the FMS statistic. They are followed by HYSPLIT (33.9%), CALPUFF (32.2%), and FLEXPART (32.1%). CALGRID has the poorest score for the FMS statistics with a value of only 24.1%.

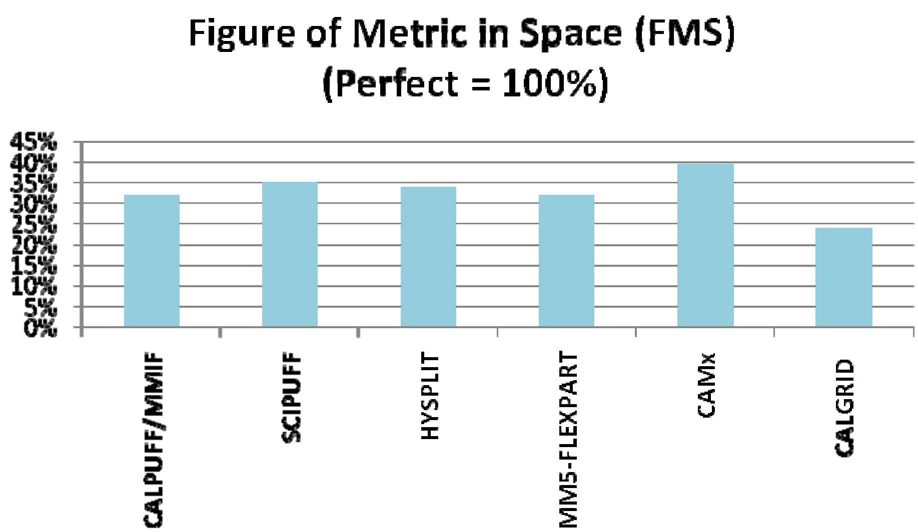


Figure C-19. Figure of Merit in Space (FMS) statistical performance metric for the six LRT models and CAPTEX Release 3.

Figure C-20 displays the FAR performance metric for the six LRT models. FLEXPART was the best performing model using the FAR statistics with a score of 54.1%. The next two best performing models using the FAR was CAMx (64.8%) and SCIPUFF (71.2%). CALPUFF and HYSPLIT exhibited similar performance for the FAR metric with values of 74.3% and 79.3% respectively. CALGRID had the worst FAR score with a value of 91.7%.

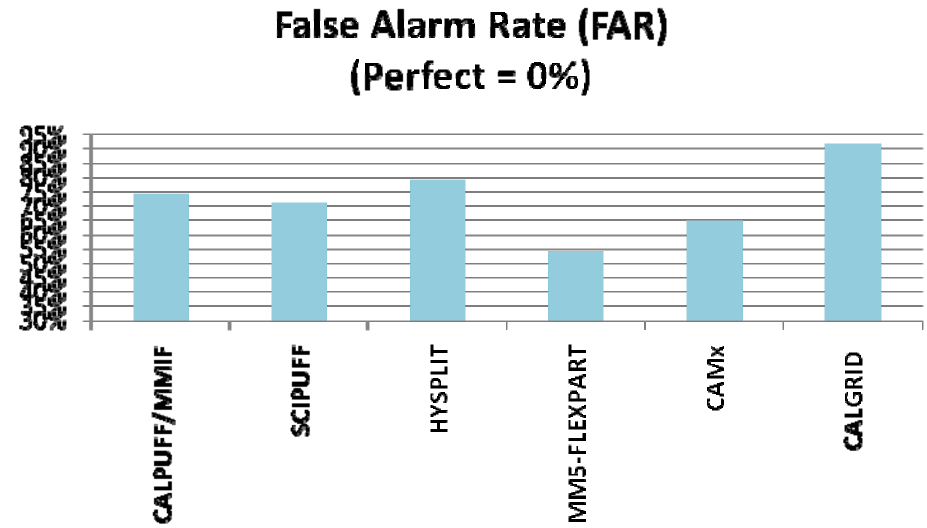


Figure C-20. False Alarm Rate (FAR) statistical performance metric for the six LRT models and CAPTEX Release 3.

Results for the Probability of Detection (POD) metric are presented in Figure C-21. CAMx was the best performing model using the POD performance statistic with a value of 52.6%. It is followed closely by FLEXPART with a score of 47.2%. SCIPUFF (41.7%) and HYSPLIT (36.1%) were in the middle, and CALPUFF and CALGRID had the worst POD score with a value of 25%.

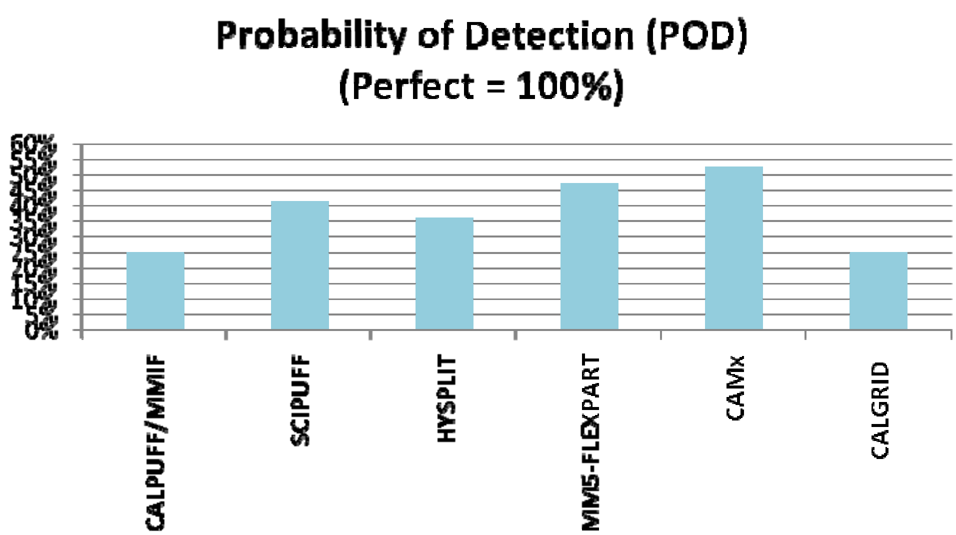


Figure C-21. Probability of Detection (POD) statistical performance metric for the six LRT models and CAPTEX Release 3.

Results for the TS metric and the six LRT models are presented in Figure C-22. FLEXPART had the highest TS statistics with a score with a value of 30.4% and was followed closely by CAMx

(26.8%). SCIPUFF (20.6%), HYSPLIT (15.1%), and CALPUFF (14.1%) were in the middle, and CALGRID exhibited the poorest TS performance with a score of 6.7%.

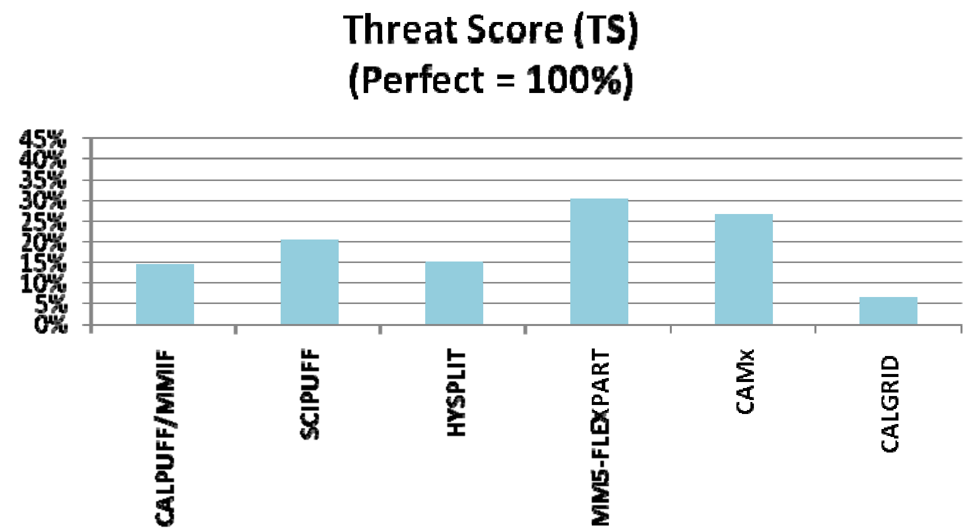


Figure C-22. Threat Score (TS) statistical performance metric for the six LRT models and CAPTEX Release 3.

Overall spatial performance was relatively equal between FLEXPART and CAMx, with CAMx having the best performance for the FMS and POD statistics and FLEXPART having better performance in the FAR and TS categories. CALPUFF, SCIPUFF, and HYSPLIT were comparable in their spatial performance for the CTEX3 experiment, with SCIPUFF showing marginally better scores in all of the four spatial performance metrics. CALGRID exhibited the poorest performance across all four spatial metrics.

C.4.2 GLOBAL ANALYSIS OF MODEL PERFORMANCE

Eight global statistical analysis metrics are used to evaluate the five LRT model performance using the ETEX data base that are described in Section 2.4 and consist of the FOEX, FA2, FA5, NMSE, PCC, FB, KS and RANK statistical metrics.

Figure C-23 displays the FOEX performance metrics for the six LRT models. HYSPLIT has the best FOEX score with a value of 2.0% that is closest to zero. The second best performing model using the FOEX metric is CAMx (9.6%) that is followed by SCIPUFF (11.5%). CALGRID has the poorest FOEX score with 20.6%.

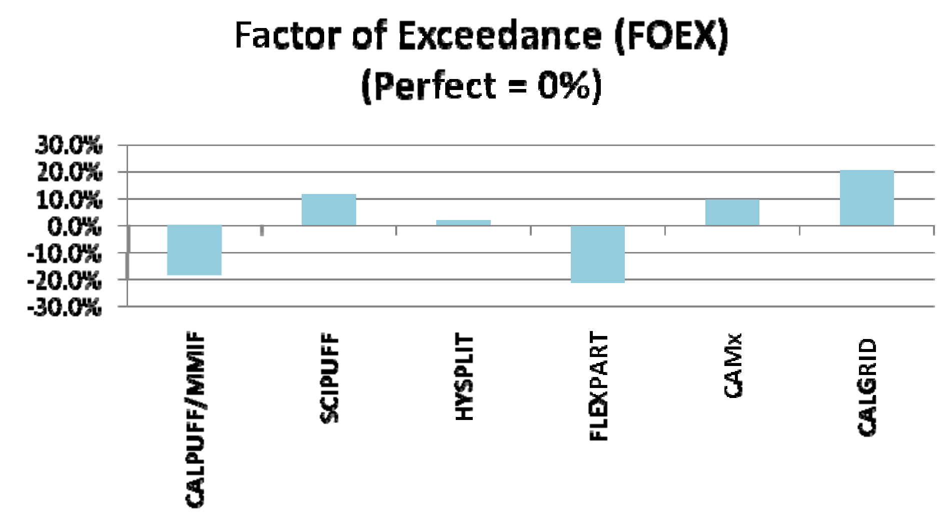


Figure C-23. Factor of Exceedance (FOEX) statistical performance metric for the six LRT models and CAPTEX Release 3.

FA2 and FA5 scores are presented in Figure C.-24. CAMx and SCIPUFF have nearly identical FA2 and FA5 scores with values of 6.6% - 6.7% (FA2) and 15% (FA5). The third best performing model for the FAα statistics is FLEXPART (5.3% and 12.2%) followed by HYSPLIT (4.5% and 8.5%). CALPUFF and CALGRID flip positions in FA2 and FA5 for the final position, with CALPUFF having a lower FA2 (0.8%) and a higher FA5 (8.2%) compared to CALGRID (FA2 – 1.05% and FA5 – 4.6%).

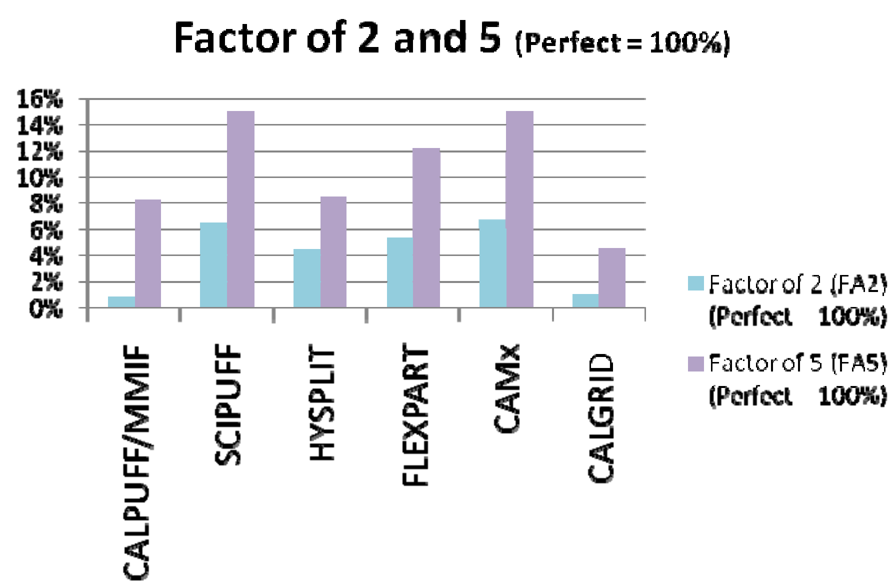


Figure C-24. Factor of 2 (FA2) and Factor of 5 (FA5) statistical performance metric for the six LRT models and CAPTEX Release 3.

The scores for the Normalized Mean Squared Error (NMSE) statistical metrics for the six LRT models are given in Figure C-25. The NMSE provides an indication of the deviations between the predicted and observed tracer concentrations paired by time and location with a perfect model receiving a 0.0 score. FLEXPART is the best performing model using the NMSE metric with a score of 21.4 pg m^{-3} followed closely by CAMx (38.9 pg m^{-3}) and CALPUFF (40.9 pg m-3). The worst performing LRT model according to the NMSE metric is HYSPLIT (130.5 pg m^{-3}).

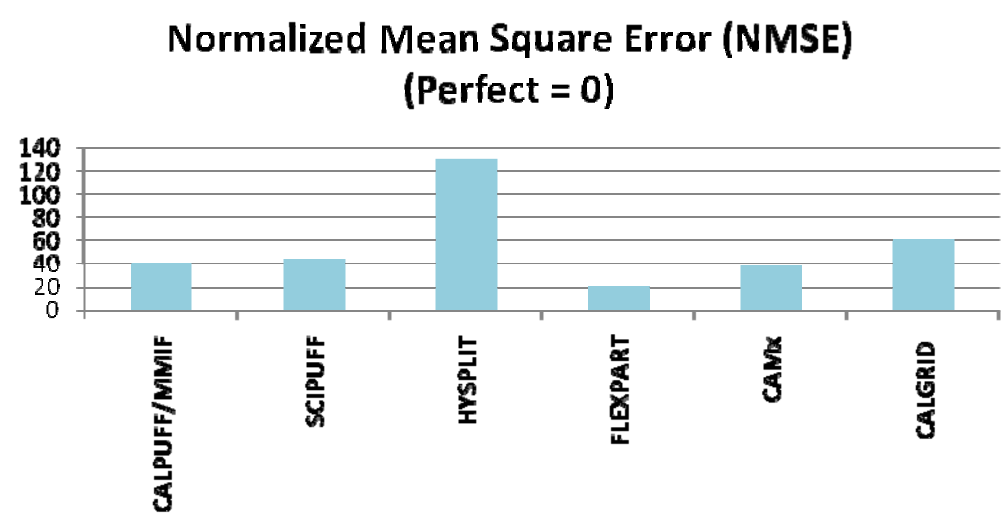


Figure C-25. Normalized Mean Square Error (NMSE) statistical performance metric for the six LRT models for CAPTEX Release 3. Values are expressed in pg m^{-3}.

The PCC values for the six LRT models are shown in Figure C-26. All models but HYSPLIT have positive correlation coefficients. The two best models according to the PCC statistical metric are CAMx (0.63) and SCIPUFF (0.56). The middle group of models consists of CALPUFF (0.4), CALGRID (0.23), and FLEXPART (0.19). The model with the least correlation with the observations is HYSPLTI (-0.1), indicating a weak negative correlation with observed data.

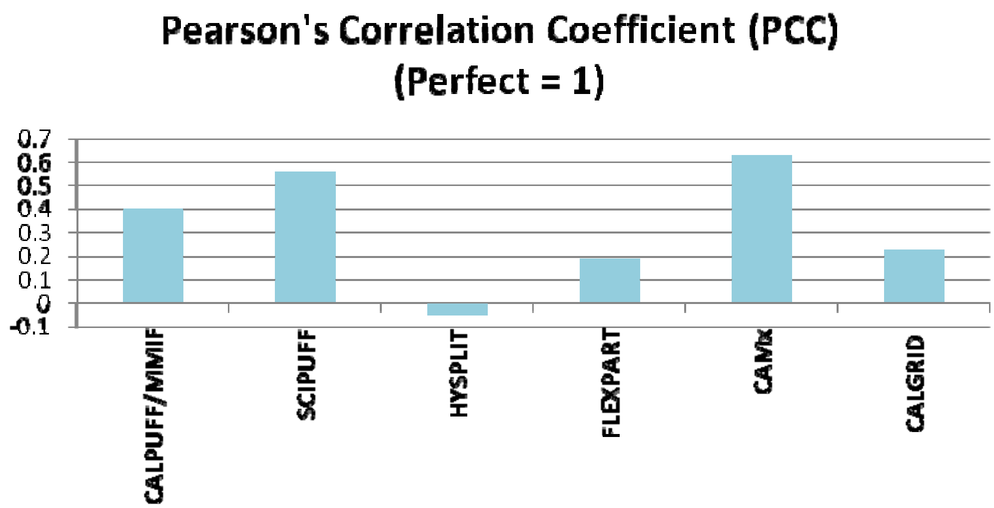


Figure C-26. Pearson's Correlation Coefficient (PCC) statistical performance metric for the six LRT models and CAPTEX Release 3.

Figure C-27 displays the FB parameter for the six LRT models. All six models exhibit a positive FB, which suggests an overestimation tendency. The best performing model with an FB value closest to zero are FLEXPART with a FB value of 0.68. CAMx (1.00), SCIPUFF (1.04) and CALPUFF (1.05) all have similar FB values are the second best performing group of models using the FB metric. CALGRID and HYSPLIT have the worst FB scores with values of 1.47 and 1.56 respectively.

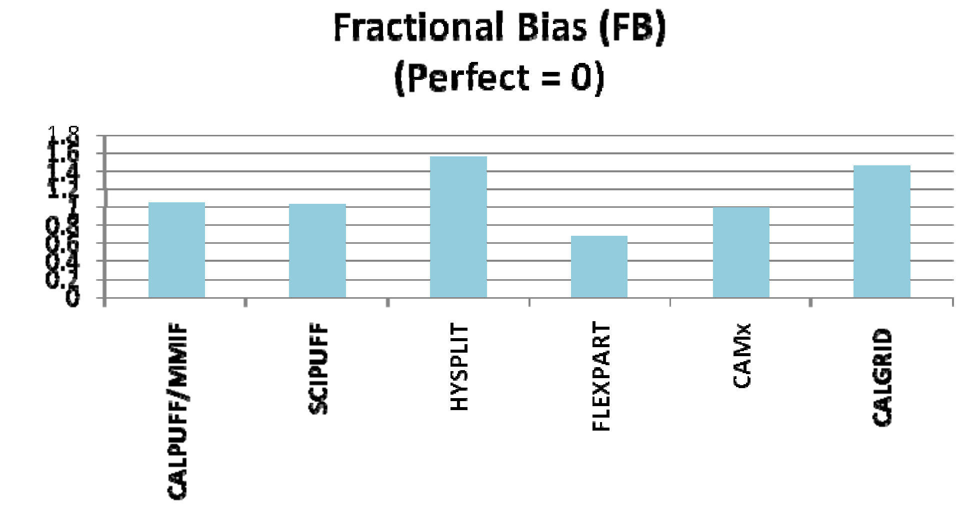


Figure C-27. Fractional Bias(FB) statistical performance metric for the six LRT models and CAPTEX Release 3.

The KS parameters for the six LRT models are shown in Figure C-28. HYSPLIT (31%) has the best KS parameter, which indicates the best match between the predicted and observed tracer concentration distributions, followed by CAMx (38%) and then SCIPUFF (43%). FLEXPART and CALGRID are essentially tied with the worst KS parameter with a value of 58%.

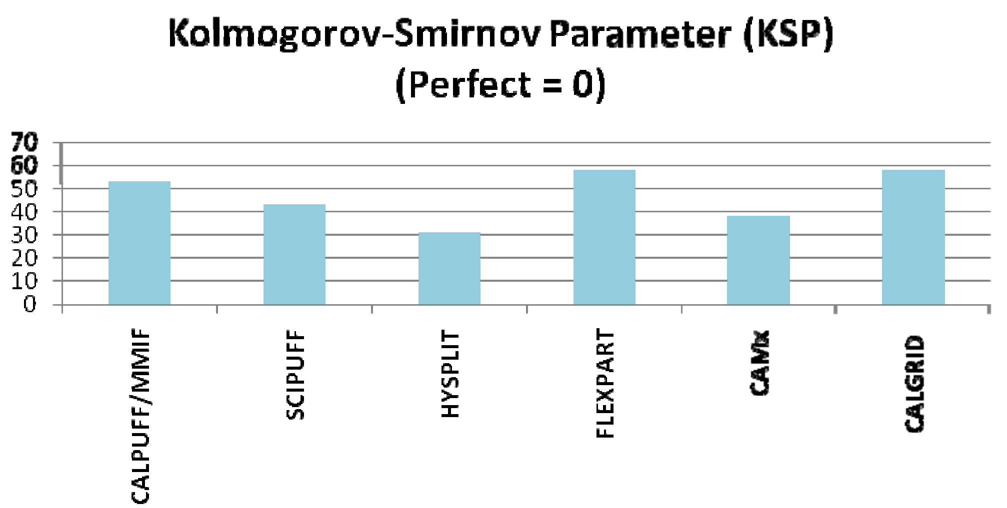


Figure C-28. Kolmogorov – Smirnov Parameter (KSP) statistical performance metrics for the six LRT models for CAPTEX Release 3.

The RANK statistical performance metric was proposed by Draxler (2001) as a single model performance metric that equally ranks the combination of performance metrics for correlation (PCC or R), bias (FB), spatial analysis (FMS) and unpaired distribution comparisons (KS). The RANK metrics ranges from 0.0 to 4.0 with a perfect model receiving a score of 4.0. Figure C-29 lists the RANK model performance statistics for the six LRT models. CAMx is the highest ranked model using the RANK metric with a value of 1.91. Note that CAMx scores high in all four areas of model performance (correlation, bias, spatial and cumulative distribution). The next best performing model according to the RANK metric is SCIPUFF with a score of 1.71. SCIPUFF scores relatively well across all of the four metrics, with slightly lower scores in cumulative distribution and correlation metrics compared to CAMx, contributing to its second rank. FLEXPART and CALPUFF are nearly even in terms of their performance with RANK values of 1.44 and 1.43 respectively. FLEXPART scores better than CALPUFF with the bias (FB) metric, whereas the reverse is true for the correlation (R^2) metric.

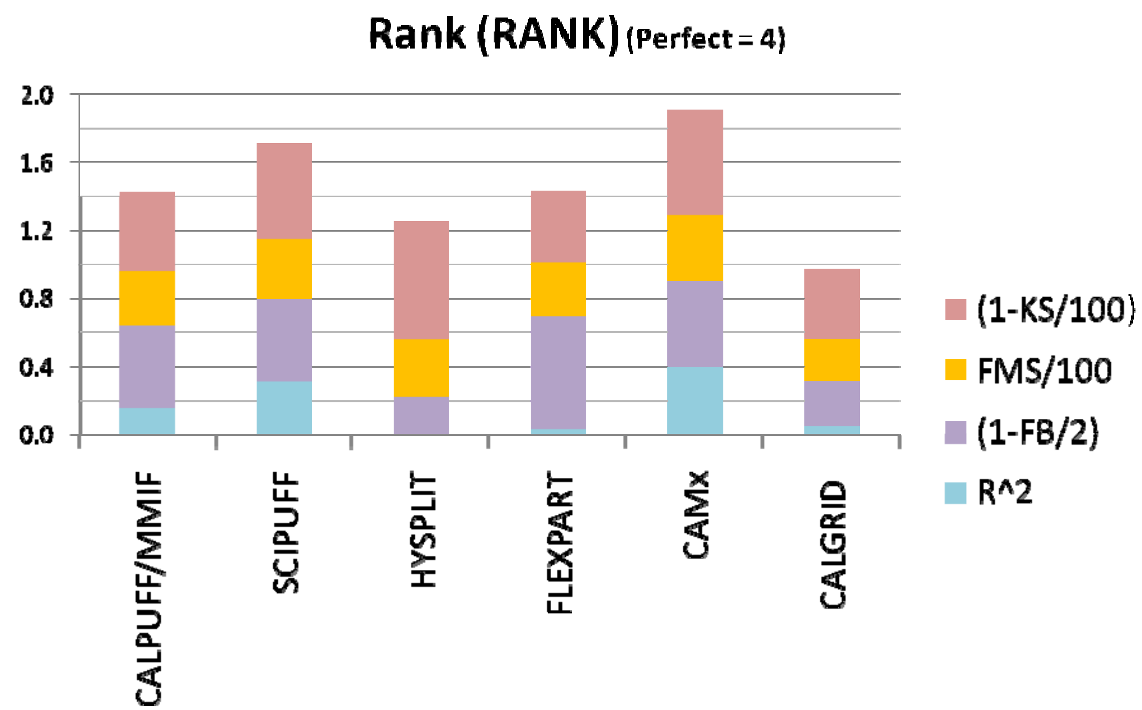


Figure C-29. RANK statistical performance metric for the six LRT models and CAPTEX Release 3.

C.4.2.1 SUMMARY OF LRT MODEL RANKINGS FOR CTEX3 USING STATISTICAL PERFORMANCE MEASURES

Table C-5 summarizes the rankings between the six LRT models for the 11 performance statistics analyzed. Depending on the statistical metric, three different models were ranked as the best performing model for a particular statistic with CAMx being ranked first more than the other models (46%) and FLEXPART ranked first second most (36%). CALGRID was consistently ranked the worst performing model being the poorest performing model for 6 of the 11 performance statistics.

In testing the efficacy of the RANK statistic for providing an overall ranking of model performance we the ranking of the six LRT models using the average rank of the 11 performance statistics versus the ranking from the RANK statistical metric (Table C-5). The average rank of model performance for the six LRT dispersion models and the CTEX3 experiment averaged across all 11 performance statistics and the comparison to the RANK rankings was as follows:

Ranking	Average of 11 Statistics	RANK
1.	CAMx	CAMx
2.	SCIPUFF	SCIPUFF
3.	FLEXPART	FLEXPART
4.	HYSPLIT	CALPUFF
5.	CALPUFF	HYSPLIT
6.	CALGRID	CALGRID

For the CTEX3 experiment, the average rankings across the 11 statistics is nearly identical to the rankings produced by the RANK integrated statistics that combines the four statistics for correlation (PCC), bias (FB), spatial (FMS) and cumulative distribution (KS) with only HYSPLIT and CALPUFF exchanging places as the 4th and 5th best performing models. CALPUFF performance was weighted down in the average statistic rankings due to lower scores in the FA2 and FA5 metrics compared to HYSPLIT. If not for this, the average rank across all 11 metrics would have been the same as Draxler's RANK score. Although this deviation did occur in the fourth and fifth ranked positions, the RANK statistic remains a valid performance statistic for indicating over all model performance of a LRT dispersion model. However, the analyst should use discretion in relying too heavily upon RANK score without consideration to which performance metrics are important measures for the particular evaluation goals. For example, if performance goals are not concerned with a model's ability to perform well in space and time, then reliance upon spatial statistics such as the FMS in the composite RANK value may not be appropriate. In the case of this evaluation, since space/time considerations are paramount for proper LRT model performance, the RANK metric is a valuable tool to rapidly assess model performance across a broad range of metrics being evaluated.

Table C-5. Summary of model ranking using the statistical performance metrics.

Statistic	1st	2nd	3rd	4th	5th	6th
FMS	CAMx	SCIPUFF	HYSPLIT	CALPUFF	FLEXPART	CALGRID
FAR	FLEXPART	CAMx	SCIPUFF	CALPUFF	HYSPLIT	CALGRID
POD	CAMx	FLEXPART	SCIPUFF	HYSPLIT	CALPUFF	CALGRID
TS	FLEXPART	CAMx	SCIPUFF	HYSPLIT	CALPUFF	CALGRID
FOEX	HYSPLIT	CAMx	SCIPUFF	CALPUFF	CALGRID	FLEXPART
FA2	CAMx	SCIPUFF	FLEXPART	HYSPLIT	CALGRID	CALPUFF
FA5	CAMx	SCIPUFF	FLEXPART	HYSPLIT	CALPUFF	CALGRID
NMSE	FLEXPART	CAMx	CALPUFF	SCIPUFF	CALGRID	HYSPLIT
PCC or R	CAMx	SCIPUFF	CALPUFF	CALGRID	FLEXPART	HYSPLIT
FB	FLEXPART	CAMx	SCIPUFF	CALPUFF	CALGRID	HYSPLIT
KS	HYSPLIT	CAMx	SCIPUFF	CALPUFF	FLEXPART	CALGRID
Avg. Ranking	CAMx	SCIPUFF	FLEXPART	HYSPLIT	CALPUFF	CALGRID
Avg. Score	1.55	2.72	3.0	4.0	4.27	5.55
RANK Ranking	CAMx	SCIPUFF	FLEXPART	CALPUFF	HYSPLIT	CALGRID
RANK	1.91	1.71	1.44	1.43	1.25	0.98

C.5 COMPARISON OF SIX LRT MODEL MODELING PERFORMANCE USING THE CAPTEX-5 EXPERIMENT

C.5.1 SPATIAL ANALYSIS OF MODEL PERFORMANCE

Figure C-30 displays the FMS spatial analysis performance metrics for the six LRT models and the CTEX5 tracer study field experiment. SCIPUFF (22.67%) and CAMx (22.66%) models are the two best performing models for the FMS statistic with nearly identical scores. They are followed by HYSPLIT (18.5%), CALPUFF (17.5%), FLEXPART (17.2), and CALGRID (16.1%).

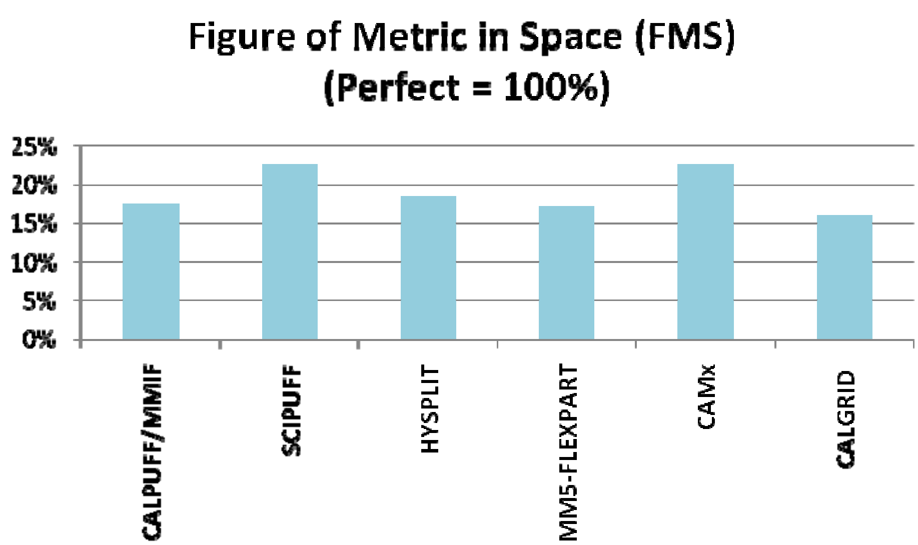


Figure C-30. Figure of Merit in Space (FMS) statistical performance metric for the six LRT models and CAPTEX Release 5.

Figure C-31 displays the FAR performance metrics. FLEXPART was the best (lowest) performing model using the FAR statistics with a score of 61.9%. The next two best performing models using the FAR metric were HYSPLIT (72.2%) and CAMx (75.6%). SCIPUFF, CALGRID, and CALPUFF had the worst (highest) FAR scores with values of 79.7%, 84.2%, and 87% respectively.

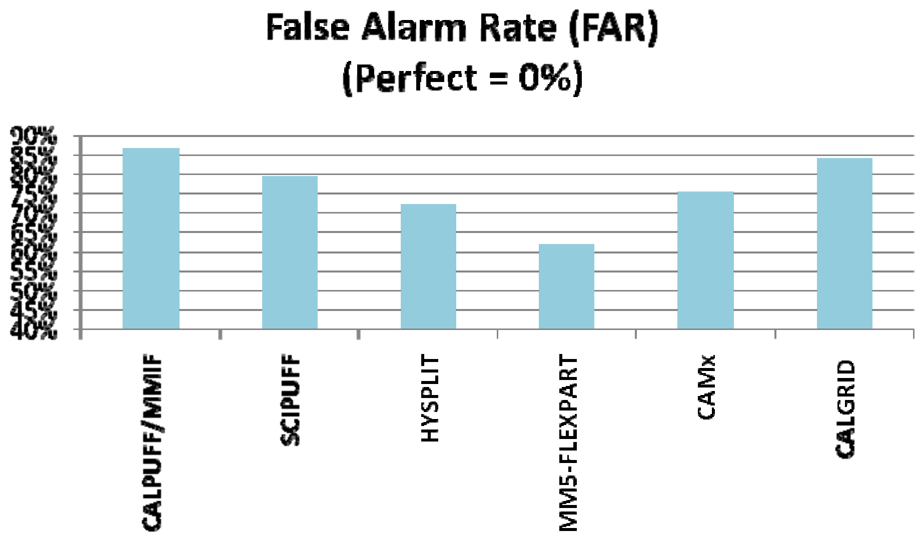


Figure C-31. False Alarm Rate (FAR) statistical performance metric for the six LRT models and CAPTEX Release 5.

Results for the Probability of Detection (POD) metric are presented in Figure C-32. SCIPUFF was the best performing model using the POD performance statistic with a value of 72.2%. CAMx was the second best performing model using POD followed by HYSPLIT, FLEXPART, CALPUFF and CALGRID.

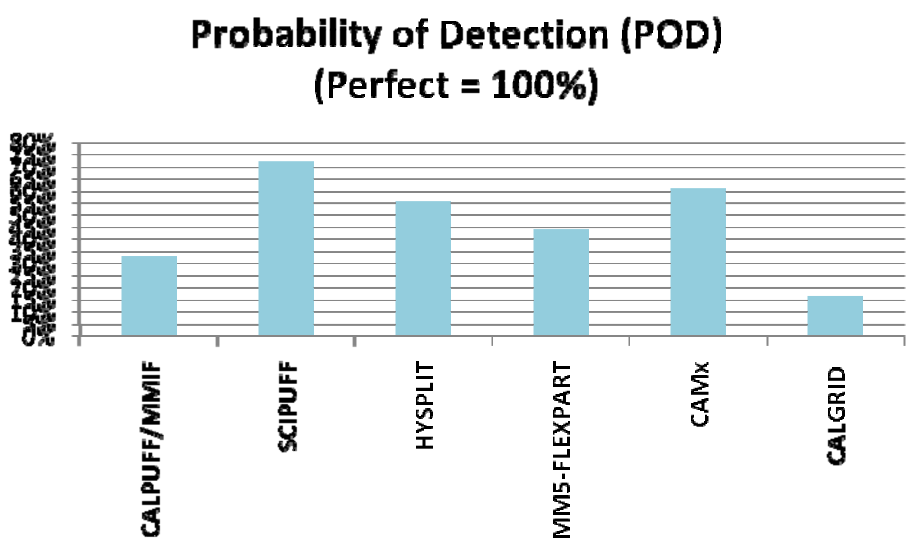


Figure C-32. Probability of Detection (POD) statistical performance metric for the six LRT models and CAPTEX Release 5.

Results for the TS metric are presented in Figure C-33. FLEXPART had the highest TS statistics with a score of 25.8%. HYSPLIT (22.7%), CAMx (21.2%), and SCIPUFF (18.8%) followed, and CALPUFF and CALGRID closed out with the worst (lowest) TS values of 10.3% and 8.8% respectively.

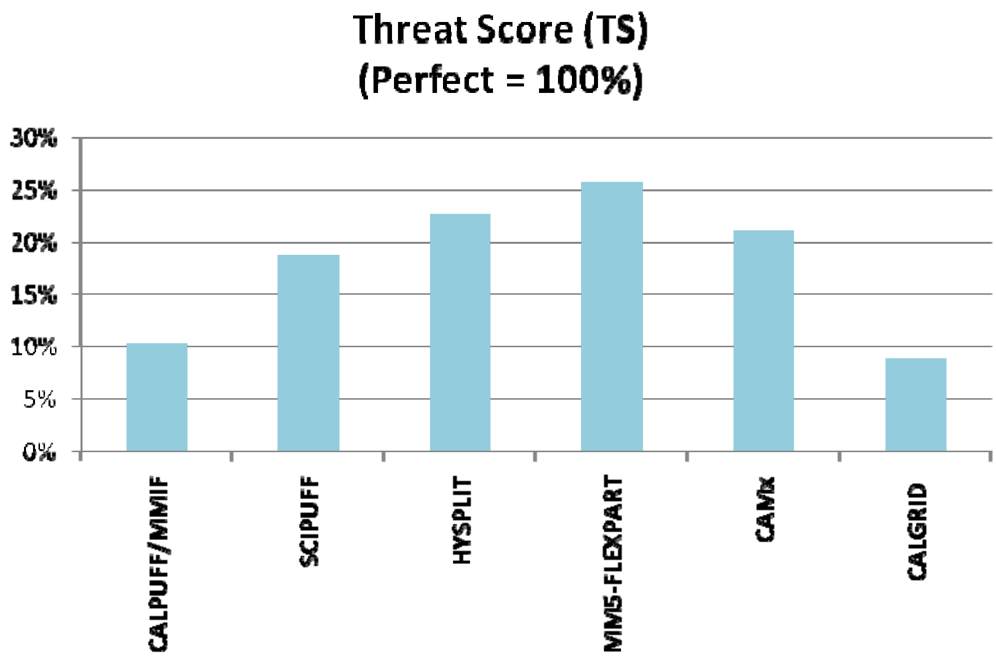


Figure C-33. Threat Score (TS) statistical performance metric for the six LRT models and CAPTEX Release 5.

Overall, the spatial performance for CTEX5 was relatively equal between FLEXPART and CAMx, with CAMx having the best performance for the FMS and POD statistics and FLEXPART having the best performance for the FAR and TS statistics. CALPUFF, SCIPUFF, and HYSPLIT were generally comparable in their spatial performance for CTEX5, with SCIPUFF showing marginally better scores in all four of the spatial metrics. CALGRID consistently exhibited the poorest performance across all four spatial metrics.

C.5.2 GLOBAL ANALYSIS OF MODEL PERFORMANCE

Figure C-34 displays the FOEX performance metrics for the six LRT models. CALPUFF had the best FOEX score (closest to zero) with a value of -2.6%. The second best performing model using the FOEX metric is CAMx (4.7%) followed by HYSPLIT (-9.2%) and CALGRID (-11.9%). SCIPUFF and FLEXPART had the poorest FOEX scores with values of 20.4% and -28.2% respectively.

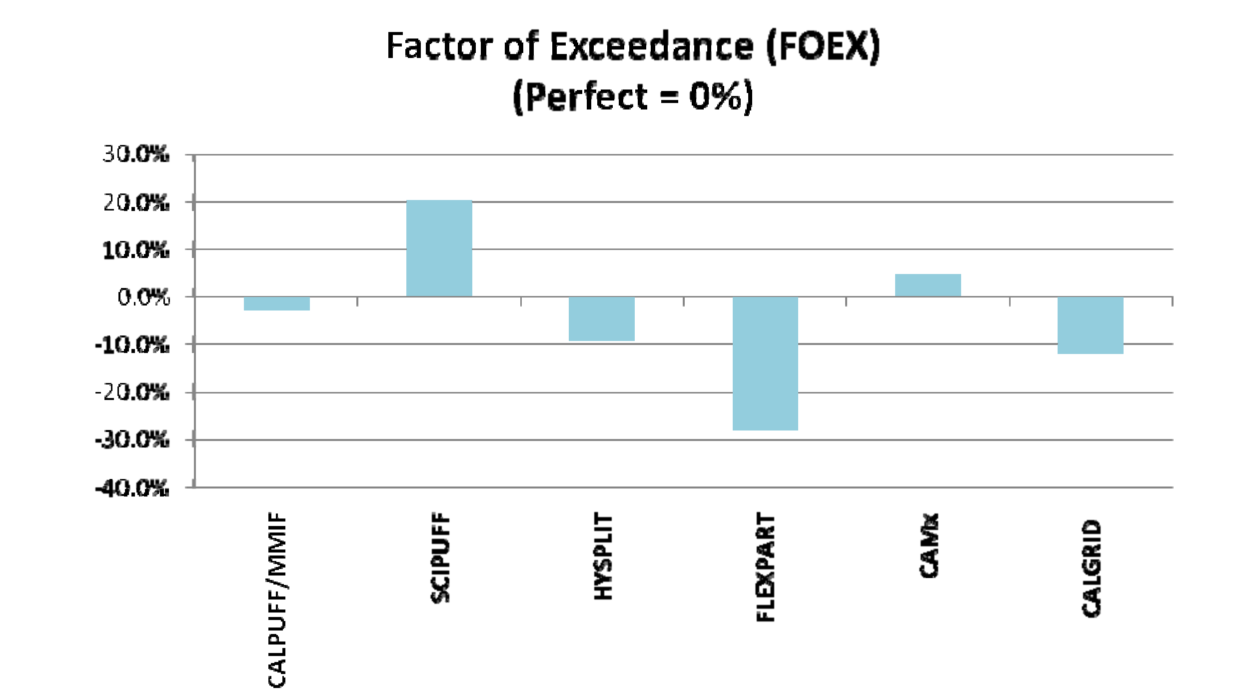


Figure C-34. Factor of Exceedance (FOEX) statistical performance metric for the six LRT models and CAPTEX Release 5.

The FA2 and FA5 scores are presented in Figure C-35. HYSPLIT has the best FAα scores with FA2 and FA5 values of 4.9% and 10.7%, respectively. CAMx and SCIPUFF have nearly identical FAα with values of 3.5% to 3.9% (FA2) and 9.3% to 9.4% (FA5). CALPUFF and FLEXPART follow with FA2/FA5 values of, respectively, 3.5%/7.9% and 2.3%/6.9%. CALGRID has the lowest FAα scores with FA2 and FA5 values of 0% and 0.9% respectively.

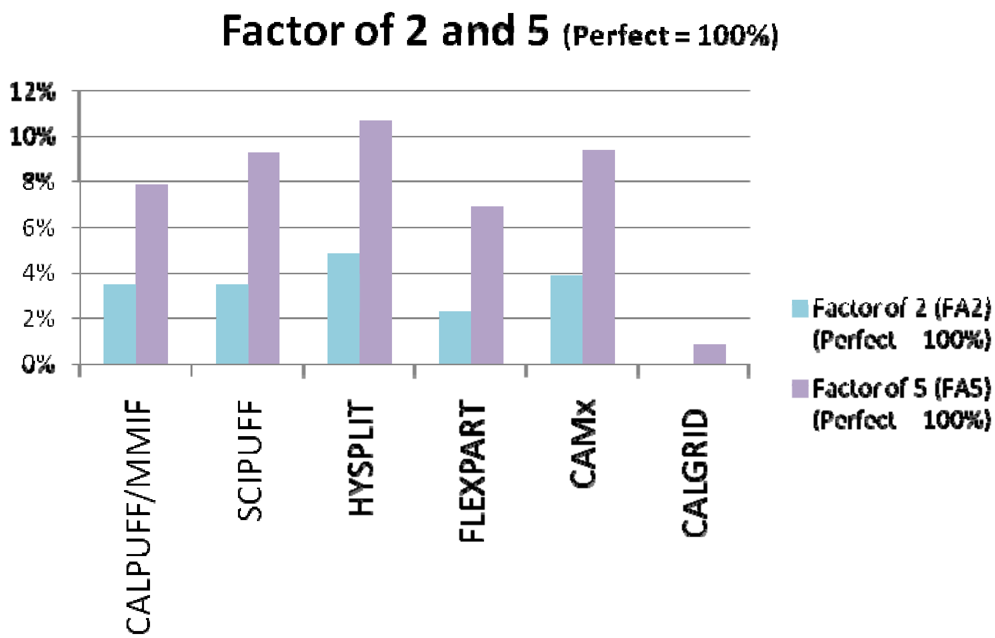


Figure C-35. Factor of 2 (FA2) and Factor of 5 (FA5) statistical performance metric for the six LRT models and CAPTEX Release 5.

The scores for the Normalized Mean Squared Error (NMSE) statistical metrics and the six LRT models are given in Figure C-36. CAMx is the best performing model using the NMSE metric with a score of 9.4 pg m^{-3} followed by SCIPUFF (14.8 pg m^{-3}). The middle tier of models are comprised of FLEXPART, HYSPLIT, and CALPUFF with values of 17.0, 17.2, and 19.5 pg m^{-3} respectively. CALGRID closes out with a NMSE value of 29.6 pg m^{-3}.

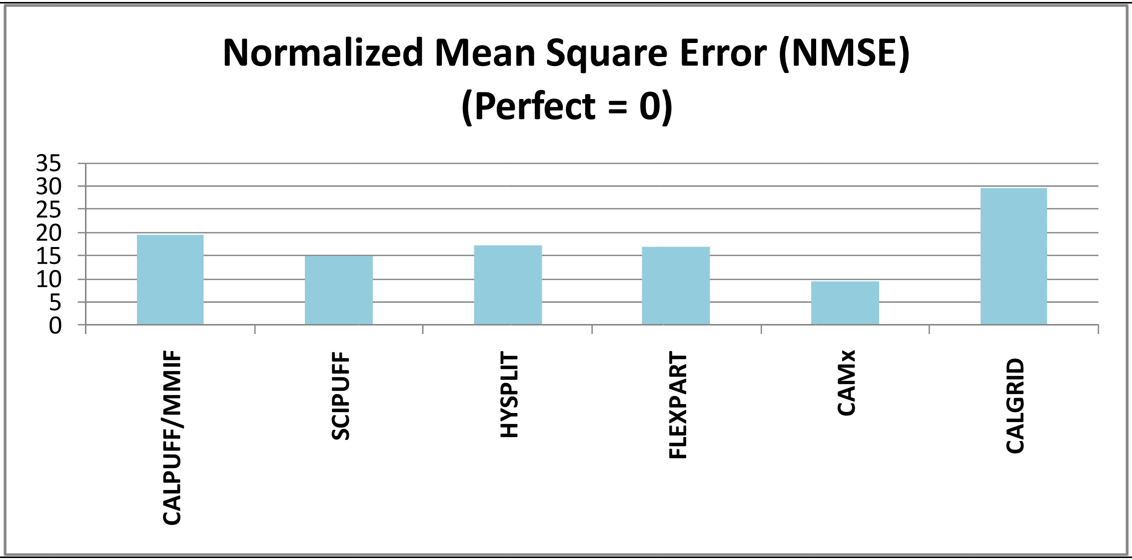


Figure C-36. Normalized Mean Square Error (NMSE) statistical performance metric for the six LRT models and CAPTEX Release 5. Values are expressed in pg m^{-3}.

The PCC values for the six LRT models are shown in Figure C-37. All models but CALGRID and CALPUFF have positive PCCs. The three best models according to the PCC statistical metric are HYSPLIT (0.60), CAMx (0.59) and SCIPUFF (0.56). FLEXPART has a PCC of 0.51. Both CALPUFF and CALGRID have negative PCC scores, indicating the modeled concentrations are anti-correlated with the observed data. CALGRID and CALPUFF have PCC values of -0.06 and -0.07, respectively.

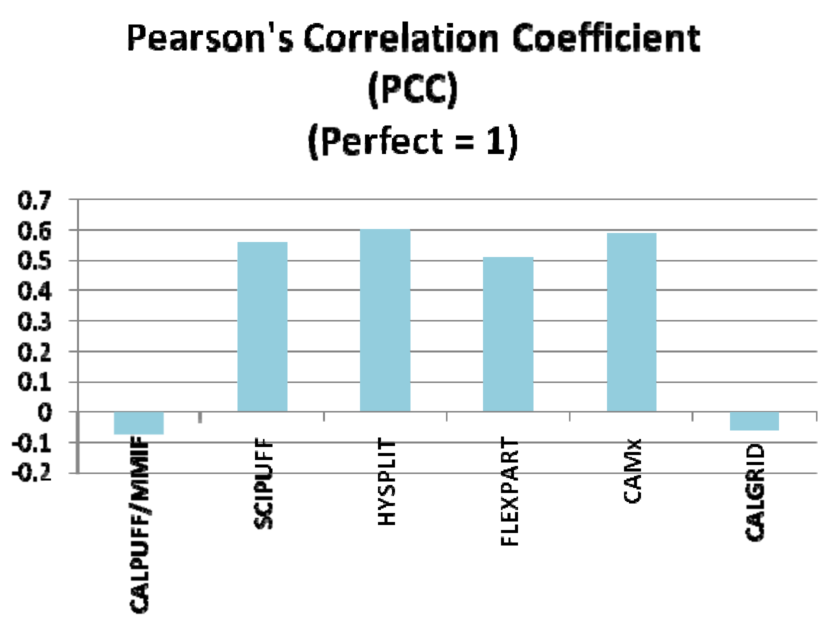


Figure C-37. Pearson's Correlation Coefficient (PCC) statistical performance metric for the six LRT models for CAPTEX Release 5.

Figure C-38 displays the FB parameter for the six LRT models and CTEX5. Four of the six models exhibit a positive FB, which suggests an overestimation tendency, whereas the other two have a negative FB. The best performing models with the FB parameter closest to zero are CAMx with a FB score of 0.49 indicating overestimation and CALPUFF with a FB value of -0.49 indicating underestimation. Next best is FLEXPART (-0.87) with an underestimation bias and SCIPUFF (0.89) with an overestimation bias followed by HYSPLIT (0.93) and CALPUFF (1.07).

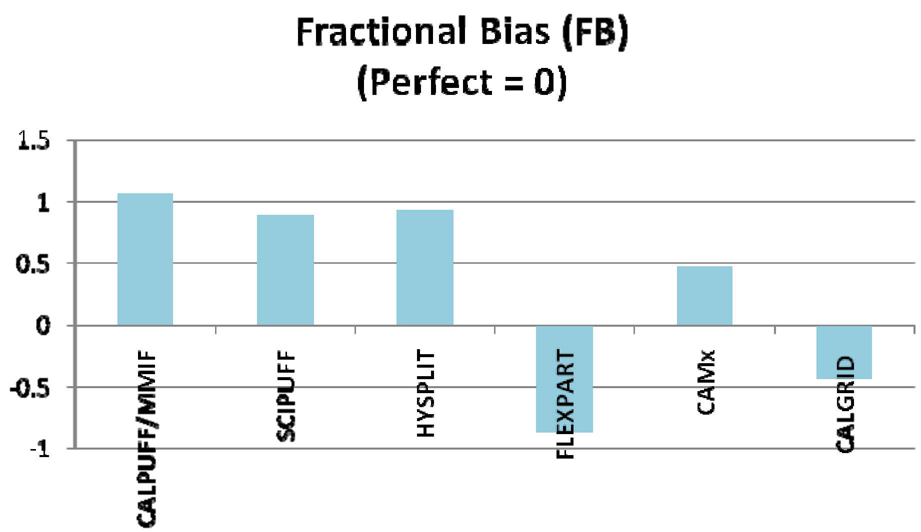


Figure C-38. Fractional Bias (FB) statistical performance metric for the six LRT models and CAPTEX Release 5.

The KS parameters for the six LRT models are shown in Figure C-39. HYSPLIT (28%) has the lowest KS parameter, which indicates the best match between the predicted and observed tracer concentration distributions according to the KS parameter, followed by CALPUFF (37%) and CALGRID (38%). CAMx follows with a score of 41% and FLEXPART and SCIPUFF close out with scores of 55% and 56% respectively.

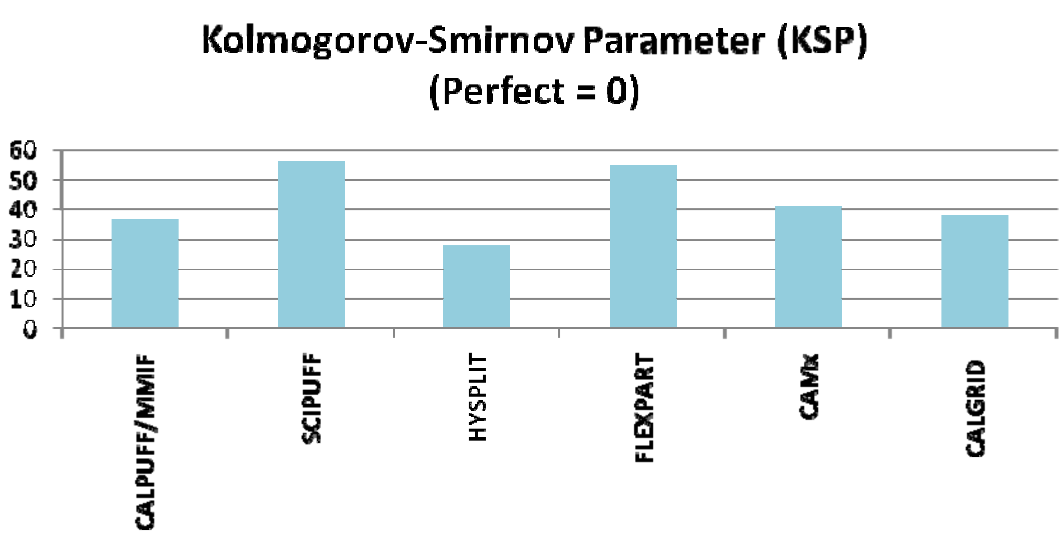


Figure C-39. Kolmogorov – Smirnov Parameter (KSP) statistical performance metrics for the six LRT models and CAPTEX Release 5.

Figure C-40 lists the RANK model performance statistics for the six LRT models. CAMx is the highest ranked model using the RANK metric with a value of 1.91 followed by HYSPLIT at 1.8. It is important to note, however, that both CAMx and HYSPLIT exhibit high scores in all four areas of model performance (correlation, bias, spatial and cumulative distribution). It is an important attribute of model performance to score high in all areas of model performance. The next best performing model according to the RANK metric is CALGRID with a score of 1.57, followed by SCIPUFF (1.53), FLEXPART (1.45), and finally CALPUFF (1.28). However, the CALGRID third best RANK metric comes at the expense of low spatial (FMS) and zero correlation (PCC or R2) performance skill.

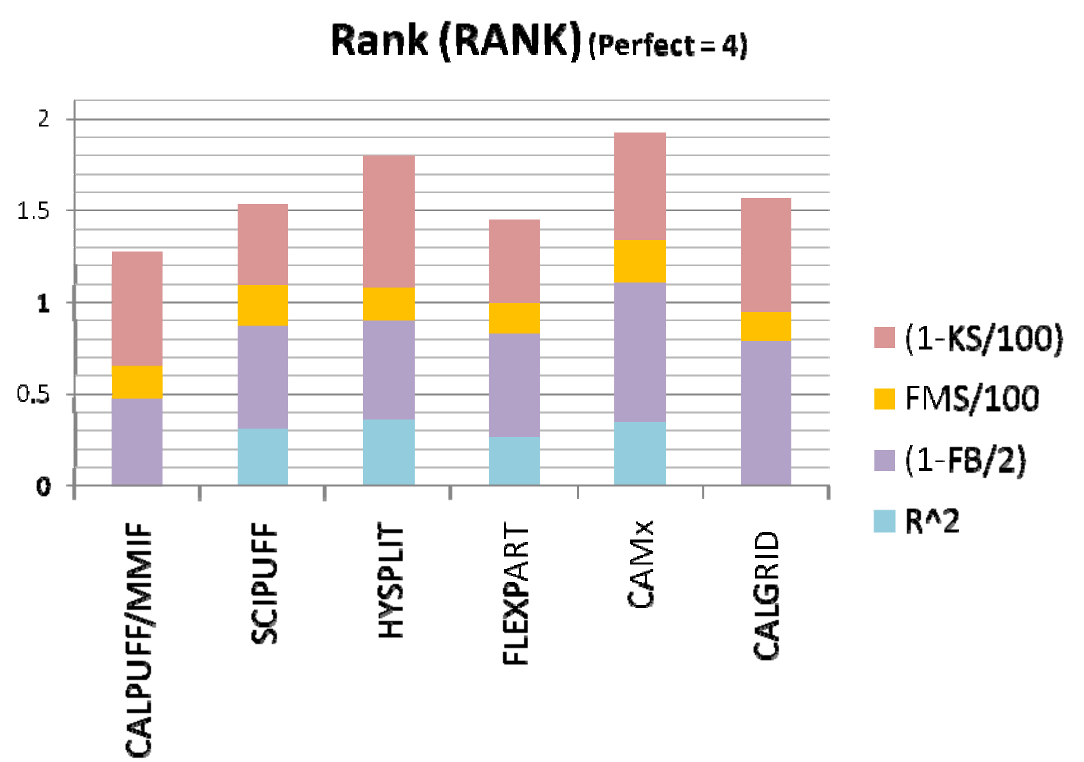


Figure C-40. RANK statistical performance metric for the six LRT models and CAPTEX Release 5.

C.5.2.1 SUMMARY OF SIX LRT MODEL RANKINGS FOR CTEX5 USING STATISTICAL PERFORMANCE MEASURES

Table C-6 summarizes the rankings of the six LRT models for the 11 performance statistics analyzed in CAPTEX Release 5 and compares the averaging ranking across the 11 statistics against the RANK metric. Depending on the statistical metric, five of the six models were ranked as the best performing model for a particular statistic. CALGRID was the only model not ranked first for any performance statistics, although it tied with CAMx for having the lowest (best) FB value (-0.49), but it was ranked behind CAMx (FB of +0.49) due to a desire for regulatory models to not have an underestimation bias. HYSPLIT was ranked the best performing model the most often scoring best in 4 of the 11 statistics (36% of the time). SCIPUFF, FLEXPART and CAMx all scored best with 2 of the 11 statistics (18%) with CALPUFF scoring best for just one statistical metric.

In testing the efficacy of the RANK statistic, overall rank across all eleven statistics was used to come up with an average modeled ranking to compare with the RANK statistic rankings. The average rank across all 11 performance statistics and the RANK rankings are as follows:

Ranking	Average of 11 Statistics	RANK
1.	CAMx	CAMx
2.	HYSPLIT	HYSPLIT
3.	SCIPUFF	CALGRID
4.	FLEXPART	SCIPUFF
5.	CALPUFF	FLEXPART
6.	CALGRID	CALPUFF

The results from CAPTEX Release 5 present an interesting case study on the use of the RANK metric to characterize overall model performance. As noted in Table C-6 and given above, the relative ranking of models using the average rankings across the 11 statistical metrics is considerably different than the RANK scores after the two highest ranked models. Both approaches rank CAMx as the best and HYSPLIT as the next best performing models for CTEX5, with rankings that are fairly close to each other. However, after that the two ranking techniques come to different conclusions regarding the ability of the models to simulate the observed tracer concentrations for the CTEX5 field experiment.

The most noticeable feature of the RANK metric for ranking models in CTEX5 is the third highest ranking model using RANK, CALGRID (1.57). CALGRID ranks as the worst or second worst performing model in 9 of the 11 performance statistics (82% of the time) and have an average ranking of 5.0, which means on average it is the 5th best performing model out of 6. In examining the contribution to the RANK metric for CALGRID, there is not a consistent contribution from all four broad categories to the composite score (Figure C-40). Recall from equation 2-12 in Section 2.4.3.2 that the RANK score is defined by the contribution of the four of the 11 statistics that represent measures of correlation/scatter (R2), bias (FB), spatial (FMS) and cumulative distribution:

$$RANK = \left|R^2\right| + \left(1 - \left|FB/2\right|\right) + FMS/100 + \left(1 - KS/100\right)$$

The majority of CALGRID's 1.57 RANK score comes from fractional bias and Kolmogorov-Smirnov parameter values. Recall from Figures C-36 and C-39 that the FOEX and FB metrics indicate that CALGRID consistently underestimates. The FB component to the composite score for CALGRID is one of the highest among the six models in this study, yet the underlying statistics indicate both marginal spatial skill and a degree of under-prediction (likely due to the spatial skill of the model).

The current form of the RANK score uses the absolute value of the fractional bias. This approach weights underestimation equally to overestimation. However, in a regulatory context, EPA is most concerned with models not being biased towards underestimation. When looking at all of the performance statistics, CALGRID is clearly one of the worst performing LRT models for CTEX5, and is arguably the worst performing model. Adaptation of RANK score for regulatory use will likely require refinement of the individual components to insure that this situation does not develop and to insure that the regulatory requirement of bias be accounted for when weighting the individual statistical measures to produce a composite score.

Table C-6. Summary of model rankings using the statistical performance metrics and comparison with the RANK metric.

Statistic	1st	2nd	3rd	4th	5th	6th
FMS	SCIPUFF	CAMx	HYSPLIT	CALPUFF	FLEXPART	CALGRID
FAR	FLEXPART	HYSPLIT	CAMx	SCIPUFF	CALGRID	CALPUFF
POD	SCIPUFF	CAMx	HYSPLIT	FLEXPART	CALPUFF	CALGRID
TS	FLEXPART	HYSPLIT	CAMx	SCIPUFF	CALPUFF	CALGRID
FOEX	CALPUFF	CAMx	HYSPLIT	CALGRID	SCIPUFF	FLEXPART
FA2	HYSPLIT	CAMx	CALPUFF	SCIPUFF	FLEXPART	CALGRID
FA5	HYSPLIT	CAMx	SCIPUFF	CALPUFF	FLEXPART	CALGRID
NMSE	CAMx	SCIPUFF	FLEXPART	HYSPLIT	CALPUFF	CALGRID
PCC or R	HYSPLIT	CAMx	SCIPUFF	FLEXPART	CALGRID	CALPUFF
FB	CAMx	CALGRID	FLEXPART	SCIPUFF	HYSPLIT	CALPUFF
KS	HYSPLIT	CALPUFF	CALGRID	CAMx	FLEXPART	SCIPUFF
Avg. Ranking	CAMx	HYSPLIT	SCIPUFF	FLEXPART	CALPUFF	CALGRID
Avg. Score	2.20	2.4	3.4	3.8	4.3	5.0
RANK Ranking	CAMx	HYSPLIT	CALGRID	SCIPUFF	FLEXPART	CALPUFF
RANK	1.91	1.80	1.57	1.53	1.45	1.28

C.5.3 SUMMARY AND CONCLUSIONS OF CAPTEX LRT MODEL EVALUATION

Following the ATMES-II evaluation paradigm described in Section 2.4.3.1 (spatial) and 2.4.3.3 (global), the performance of the six LRT dispersion models described in Section 2.2 have been evaluated for the Cross Appalachian Tracer Experiment (CAPTEX) Releases 3 and 5. Sensitivities of the INITD (particle/puff) configuration for HYSPLIT and Kz/advection solver combination for CAMx were examined for each CAPTEX release as well as in intercomparison of the model performance for the six models.

The model sensitivity results for HYSPLIT and CAMx are largely comparable to the conclusions from those of the ETEX experiment. For HYSPLIT, the puff-particle hybrid configurations appear to offer a distinct performance advantage over either HYSPLIT's pure particle or puff based formulations. For CAMx, the CMAQ Kz option typically performs the best, followed closely by TKE. The OB70 combination consistently performs the poorest for both CAPTEX releases. The evaluation of the use of the CAMx model's subgrid scale PiG module generally yields slightly degraded performance statistics over the NoPiG option.

United States Environmental Protection Agency
Office of Air Quality Planning and Standards
Air Quality Assessment Division
Research Triangle Park, NC
Publication No. EPA-454/R-12-003
May, 2012