FINAL REPORT

Performance Evaluation of Atmospheric Transport Models

Part of Task 3: Independent Analysis of Exposure, Dose, and Health Risk to Offsite Individuals

Revision 1
August 1999

Submitted to the Colorado Department of Public Health and Environment, Disease Control and Environmental Epidemiology Division, Rocky Flats Health Studies in partial fulfillment of Contract No. 100APPRCODE 391

"Setting the standard in environmental health"

Radiological Assessments Corporation
417 Till Road Neeses, South Carolina 29107
phone 803.536.4883 fax 803.534.1995

FINAL REPORT

Performance Evaluation of Atmospheric Transport Models

Part of Task 3: Independent Analysis of Exposure, Dose, and Health Risk to Offsite Individuals

Revision 1
August 1999

Author
Arthur S. Rood, K-Spar, Inc.

Principal Investigator
John E. Till, Ph.D., *Radiological Assessments Corporation*

SUMMARY

This study evaluates five atmospheric transport models for use in the Rocky Flats Environmental Dose Reconstruction Project. These models are the Terrain-Responsive Atmospheric Code (TRAC) (Hodgin 1991); the Industrial Source Complex, Short Term Version 2 (ISC) (EPA 1992); Regional Atmospheric Transport Code for Hanford Emission Tracking (RATCHET) (Ramsdell et al. 1994); INPUFF2 (Petersen and Lavdas 1986); and TRIAD (Hicks et al. 1989). Simulations involving TRAC and ISC are documented in Haugen and Fotino (1993) and were used without modification. The RATCHET, INPUFF2, and TRIAD simulations were performed by the author using the 15-minute meteorological data at the 10-m level taken from the 61-m tower located at the Rocky Flats Plant (RFP). RATCHET and TRIAD simulations also incorporated data taken in 1-hour increments at Denver Stapleton International Airport located approximately 25 km southeast of the plant.

Evaluations were based on how well model predictions compared with measured tracer concentrations taken during the Winter Validation Tracer Study (Brown 1991) conducted in February 1991 at the RFP. The study consisted of 12 separate tests; 6 tests were conducted during nighttime hours, 4 during daytime hours, and 2 during day-night transition hours. For each test, an inert tracer (sulfur hexafluoride) was released at the RFP at a constant rate for 11 hours. Two sampling arcs, 8 and 16 km from the release point, measured tracer concentrations every hour for the last 9 hours of each test period. Seventy-two samplers were located on the 8-km arc and 68 samplers were located on the 16-km arc. Predicted concentrations were then compared to the observed tracer concentrations at each of the samplers.

Modeling objectives considered previous performance evaluations using the Winter Validation Tracer Study data set, the nature of the primary release events at the RFP, and the limited meteorological data available during those events. The primary release events were two fires in 1957 and 1969, which lasted about 12 hours each, and suspension of plutonium-contaminated soil from the 903 drum storage pad (1964 through 1969). Fifty percent of the releases from the 903 area were attributed to 25 high wind days that occurred in 1968 and 1969 (Weber et al. 1997). In addition to the major release events, estimates of exposure from routine releases are also to be calculated.

The endpoints of the dose reconstruction project are estimates of lifetime carcinogenic incidence risk to hypothetical receptors residing at fixed locations in the model domain during the operation of the RFP from 1953 to 1989. Therefore, the time-integrated concentration was the most appropriate quantity to evaluate. Other endpoints, such as the maximum hourly average concentration may be of interest to provide bounding estimates of risk and were included in the evaluation.

Modeling objectives were separated into paired and unpaired comparisons. Paired comparisons included the evaluating the time-averaged point concentration (9-hour average concentration at each sampler) and the arc-integrated concentration (9-hour average ground-level concentration integrated over either the 8- or 16-km sampling arc). Unpaired comparisons included the 25 highest time-averaged concentrations for each 9-hour test and the maximum hourly average concentration at any sampler in the 8- or 16-km arc during the 9-hour test. Results were presented separately for the 8- and 16-km sampling arcs.

Performance measures were based on the work of Fox (1981), EPA (1988), Hanna (1989), Cox and Tikvart (1990), and Weil et al. (1992). Initially, the fractional bias (*FB*), normalized

mean square error (*NMSE*), and correlation coefficient (*r*) were selected. The geometric mean bias (*MG*) and geometric mean variance (*VG*) were added based on the recommendations in Hanna et al. (1991). These measures are more appropriate when several of the predicted and observed concentration pairs in a sample differ by a factor of 10 or greater. Correlation coefficients based on a log-transformed regression were used for some of the data categories instead of the linear correlation coefficient. The *FB* and *MG* are measures of model bias; the *NMSE* and *VG* are measures of model variance. A perfect model would have an *FB* and *NMSE* value of 0 and an *MG* and *VG* value of 1.0. A negative *FB* value and a *MG* value less than 1.0 indicated model overprediction. On recommendation of the Health Advisory Panel, which oversees the project, we also included the geometric mean (*GM*) and geometric standard deviation (GSD) of the ratio of predicted to observed concentration (C_p/C_o).

Ninety-five percent confidence intervals on the *FB*, *NMSE*, *VG*, *MG*, and *r* performance measures were calculated. The confidence interval on the differences between the performance measures for any two models was also calculated to test whether the difference was significant (at the 95% level). Confidence limits were calculated using the bootstrap resampling method (Efron 1982) as implemented in the BOOT software (Hanna et al. 1991) and a FORTRAN program written by the author of this report. Bootstrapping assumes the sample members to be of equal variance and independent of one another. These assumptions are violated to some degree in some of the data sets. However, we included the bootstrap estimates of confidence intervals in our analysis because of its widespread use in the modeling community. Readers may chose to ignore these analyses and draw their own conclusions from the deterministic results. Similar conclusions about model performance have been arrived at regardless of the measures used to evaluate the models in this study.

Time-averaged paired comparisons were limited to only those points where the predicted and observed concentrations were greater than the time-averaged sampler minimum detectable concentration of 3.7 ng m^{-3}.

No one model consistently outperformed the others in all modeling objectives and performance measures. Differences between some model predictions and corresponding observations exceeded a factor of 10 for the time-averaged paired comparisons and were somewhat less for the other modeling objectives. Models tended to overpredict lower observed concentrations and underpredict higher observed concentrations. Models often overestimated the observed concentrations taken during daytime and transition period tests and underestimated the concentrations during nighttime tests.

For the maximum concentration modeling objective, 75–100% of the model predictions were within a factor of 5 of the observations. Measures of bias indicated that the ISC model overpredicted concentrations. Little bias was observed for the other models. TRIAD and INPUFF2 exhibited the highest correlation with observations at the 8-km ($r = 0.88$) and 16-km ($r = 0.60$) sampling arc and had least amount of variability

Compared to the other models, performance measures for RATCHET were closer to their optimum values for the time-averaged pair comparisons. Seventy-nine percent of the RATCHET predictions were within a factor of 5 of the observations at the 8-km arc and 72% at the 16-km arc. The other models had between 64 and 72% of their predictions within a factor of 5 of the observations. RATCHET exhibited the least amount of variability among the models and had the highest correlation with observations (r ≈ 0.6). Geometric standard deviations of C_p/C_o were around 4.5 for RATCHET, but were about 6 for the other models except TRIAD, which had *GSD* values near 5.

Differences between models were less distinct for the unpaired time-averaged comparisons. Models typically performed better relative to the paired comparisons. Over 90% of the model predictions were within a factor 5 of the observations for the 8-km arc and over 80% for the 16-km arc. Correlation coefficients ranged from 0.68 to 0.84 and were highest for ISC at the 8-km sampling arc. RATCHET and ISC exhibited the least amount of variability, but all models showed less variability compared to the paired comparisons.

Arc-integrated results showed that 66 to 91% of the model predictions were within a factor of 3 of the observations. Models tended to exhibit higher *Cp*/*Co* ratios at the 16-km arc than the 8-km arc. The TRAC model was biased low at the 8-km arc and the ISC model was biased high at the 16-km arc. RATCHET and TRIAD showed no significant bias at both sampling arcs. RATCHET and ISC exhibited the least amount of variability.

Paired time-averaged performance measures suggested that the overall performance of the RATCHET model was somewhat better than the other models. Inspection of the paired and unpaired comparisons suggests that models experienced difficulty defining plume trajectories. Much of this difficulty was attributed to the influence of multilayered flow initiated by terrain complexities and the diurnal flow patterns characteristic of the Colorado Front Range that was only partially accounted for in some of the model simulations. Recent surface observations indicated substantial differences in wind velocities within the model domain. These differences were not accounted for in the ISC and INPUFF2 simulations. The performance of these two models relative to the others suggested that plume trajectory was primarily influenced by surface observations at Rocky Flats. However, slightly better performance was observed for models that incorporated Denver Stapleton International Airport meteorological data. The ability of a puff trajectory model to incorporate spatially-varying wind fields leads us to favor such models for use in the dose reconstruction study.

CONTENTS

FIGURES

TABLES

ACRONYMS

ASCOT	Atmospheric Studies in Complex Terrain
cfd	cumulative density function
CDPHE	Colorado Department of Public Health and Environment
C_o	observed concentration
C_p	predicted concentration
CWL	cylinder weight loss
DOE	U.S. Department of Energy
EPA	U.S. Environmental Protection Agency
FB	fractional bias
GM	geometric mean of C_p/C_o ratios
GSD	geometric standard deviation of C_p/C_o ratios
ISC	Industrial Source Complex, Short Term Version 2
mdc	minimum detectable concentration
MFC	mass flow controllers
MG	geometric mean bias
NCAR	National Center for Atmospheric Research
NMSE	normalized mean square error
r	correlation coefficient
RAC	*Radiological Assessments Corporation*
RATCHET	Regional Atmospheric Transport Code for Hanford Emission Tracking
RFP	Rocky Flats Plant
SF_6	sulfur hexafluoride
TRAC	Terrain-Responsive Atmospheric Code
VG	geometric mean variance
WVTS	Winter Validation Tracer Study

INTRODUCTION

The Rocky Flats Environmental Technology Site is owned by the U.S. Department of Energy (DOE) and is currently contractor-operated by Kaiser-Hill Company. For most of its history, the site was called the Rocky Flats Plant (RFP) and was operated by Dow Chemical Company as a nuclear weapons research, development, and production complex (Figure 1). The RFP is located on approximately 2,650 ha (6,500 acres) of Federal property, about 8–10 km (13–16 mi) from the cities of Arvada, Westminster, and Broomfield, Colorado and 26 km (16 mi) northwest of downtown Denver, Colorado. The original 156-ha (385-acre) main production area is surrounded by a 2,490-ha (6,150-acre) buffer zone that now delineates the RFP boundary.

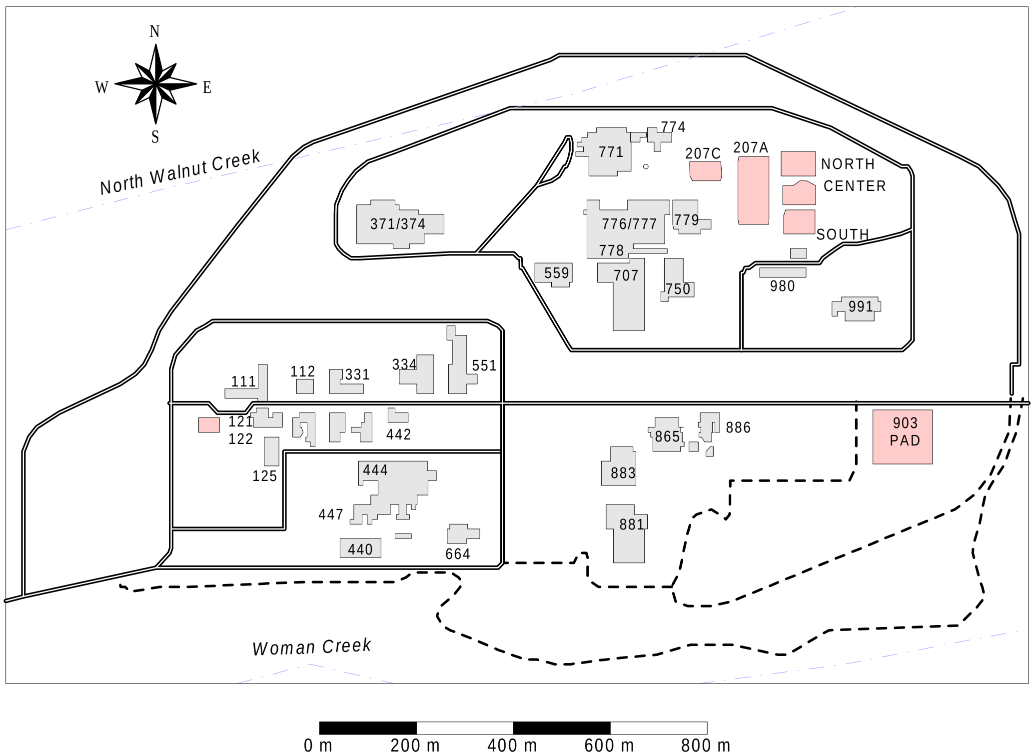

Figure 1. Main production area of the Rocky Flats Plant as it appeared in 1990. Originally, the buildings were identified with two-digit numbers. Later, a third digit was added. The production area, now sometimes called the industrial area, is surrounded by a security perimeter fence. The area between the perimeter fence and Indiana Street to the east is the buffer zone. The buffer zone was expanded to Indiana Street in the 1970s.

Through a 1989 Agreement in Principle between the DOE and the State of Colorado, DOE provided the State with funding and technical support for health-related studies. The purpose of the Historical Public Exposures Studies on Rocky Flats is to identify potential health effects in residents in nearby communities who may have been exposed to past toxic and radioactive releases. The Colorado Department of Public Health and Environment (CDPHE) first invited a national panel of experts to help design the health studies. Because of intense public concern about Rocky Flats contamination among Denver metropolitan area residents following a Federal Bureau of Investigation raid of Rocky Flats in June 1989, the panel decided to stress public involvement and to separate the research into two major phases conducted by two different contractors to enhance accountability and credibility.

Phase I of the study was performed by ChemRisk (a division of McLaren/Hart, Environmental Engineering). In Phase I, ChemRisk conducted an extensive investigation of past operations and releases from the RFP. The Phase I effort identified the primary materials of concern, release points and events, quantities released, transport pathways, and preliminary estimates of dose and risk to offsite individuals. The conclusions from Phase I were released in a public summary document (HAP 1993), a series of task reports by ChemRisk, and several articles in the journal *Health Physics*.

Radiological Assessments Corporation (*RAC*) was awarded the contract to conduct Phase II of the study, which is an in-depth investigation of the potential doses and risks to the public from historical releases from Rocky Flats. Recommendations for work to be performed in Phase II are outlined in the Phase I summary document HAP (1993).

This report compares atmospheric transport models using tracer experiment data taken in the winter of 1991 at Rocky Flats. Model performance is evaluated using six performance measures. These evaluations will assist in selecting a model or models that best meets the modeling objectives and endpoints of the Historical Public Exposures Studies on Rocky Flats.

WINTER VALIDATION TRACER STUDY

During February 1991, a tracer study was conducted near the RFP on the Front Range of the Colorado Rocky Mountains (Brown 1991). The study gathered data for validation of atmospheric transport models used in emergency response. The Winter Validation Tracer Study (WVTS), as it became known, consisted of 12 separate tests (Table 1). For each test, an inert tracer (sulfur hexafluoride [SF_6]) was emitted continuously for 11 hours from a 10-m high stack located on the east side of the main plant complex. One-hour average air concentrations were then measured for the last 9 hours of the release at 165 samplers located in radial arcs, 8 and 16 km from the release point (Figure 2). Six tests were performed under nighttime conditions, four under daytime conditions, one under day-night transition, and one under night-day transition. A total of 108 hours of data was recorded. Seventy-two samplers were distributed at or near an 8-km radial arc, and 68 samplers were distributed at or near a 16-km radial arc. These arcs encircled the release point at the RFP. Sampler elevations ranged from about 1,600 m to 2,600 m above sea level. The remaining samplers were located in clusters near high density population areas. For this study, we only considered the samplers located along the 8 and 16-km radial arcs and ignored the cluster samplers.

The WVTS was combined with the Atmospheric Studies in Complex Terrain (ASCOT) field study, which participated in a limited number of tests. During the tests in which ASCOT participated, additional meteorological measurements were taken at locations within the nocturnal canyon drainage flows.

Previous investigators (Haugen and Fotino 1993) have used this data set in a performance evaluation of the Terrain-Responsive Atmospheric Code (TRAC) and the Industrial Source Complex Code, Version 2 (ISC) models. An electronic copy of the data set was obtained from Dr. Duane Haugen of Colorado School of Mines. These data included the observed concentrations for all 12 tests, the sampler ID numbers, and the TRAC and ISC predicted concentrations. Azimuth location and radial arc distances for each of the samplers were obtained from the written documentation of the study (Haugen and Fotino 1993). Colorado School of Mines performed the ISC model simulations, and the meteorological data used in these simulations were also included in the data set. The ISC and TRAC results provided by Haugen were used in this study without modification. Haugen provided 17,820 predicted and observed concentration triplets (TRAC, ISC, and observed) (9 hours per test × 12 tests × 165 samplers per test). This number included the sample clusters located in high density areas. Of that total, 15,120 triplets were used in this analysis. As stated previously, the cluster samplers were omitted.

ATMOSPHERIC TRANSPORT MODELS

Five atmospheric transport models were evaluated in this study: TRAC (Hodgin 1991); ISC Version 2 (EPA 1992); the Regional Atmospheric Transport Code for Hanford Emission Tracking (RATCHET) (Ramsdell et al. 1994); TRIAD (Hicks et al. 1989); and INPUFF2 (Petersen and Lavdas 1986). The models were selected to cover a range of complexity, from the relatively simple straight-line Gaussian Plume model represented by ISC to a complex terrain model such as TRAC.

Table 1. Winter Validation Tracer Source Strength (Brown 1991)

Test	Start date	Start time (MST)[a]	End date	End time (MST)[a]	MFC[b] (kg hr $^{-1}$)	CWL[c] (kg hr $^{-1}$)	Average (kg hr $^{-1}$)
1	02/03/91	20:00:00	02/04/91	07:00:00	13.71	13.24	13.48
2	02/04/91	20:00:00	02/05/91	07:00:00	13.05	12.16	12.61
3	02/06/91	20:00:00	02/07/91	07:00:00	13.71	13.33	13.52
4	02/07/91	20:00:00	02/08/91	07:00:00	16.53	16.84	16.69
5	02/09/91	13:00:00	02/09/91	00:00:00	23.61	22.63	23.12
6	02/11/91	07:00:00	02/11/91	18:00:00	23.61	22.94	23.28
7	02/12/91	07:00:00	02/12/91	18:00:00	23.61	23.99	23.80
8	02/14/91	01:00:00	02/14/91	12:00:00	23.61	23.44	23.53
9	02/15/91	07:00:00	02/15/91	18:00:00	23.61	23.29	23.45
10	02/16/91	20:00:00	02/17/91	07:00:00	23.61	23.47	23.54
11	02/17/91	20:00:00	02/18/91	07:00:00	23.61	23.04	23.33
12	02/19/91	07:00:00	02/19/91	18:00:00	23.21	22.97	23.09

[a] Mountain standard time.
[b] Release rate calculated from mass flow controllers (MFC) which were calibrated at 760 mm Hg, 21.11 °C
[c] Release rate determined from cylinder weight loss (CWL).

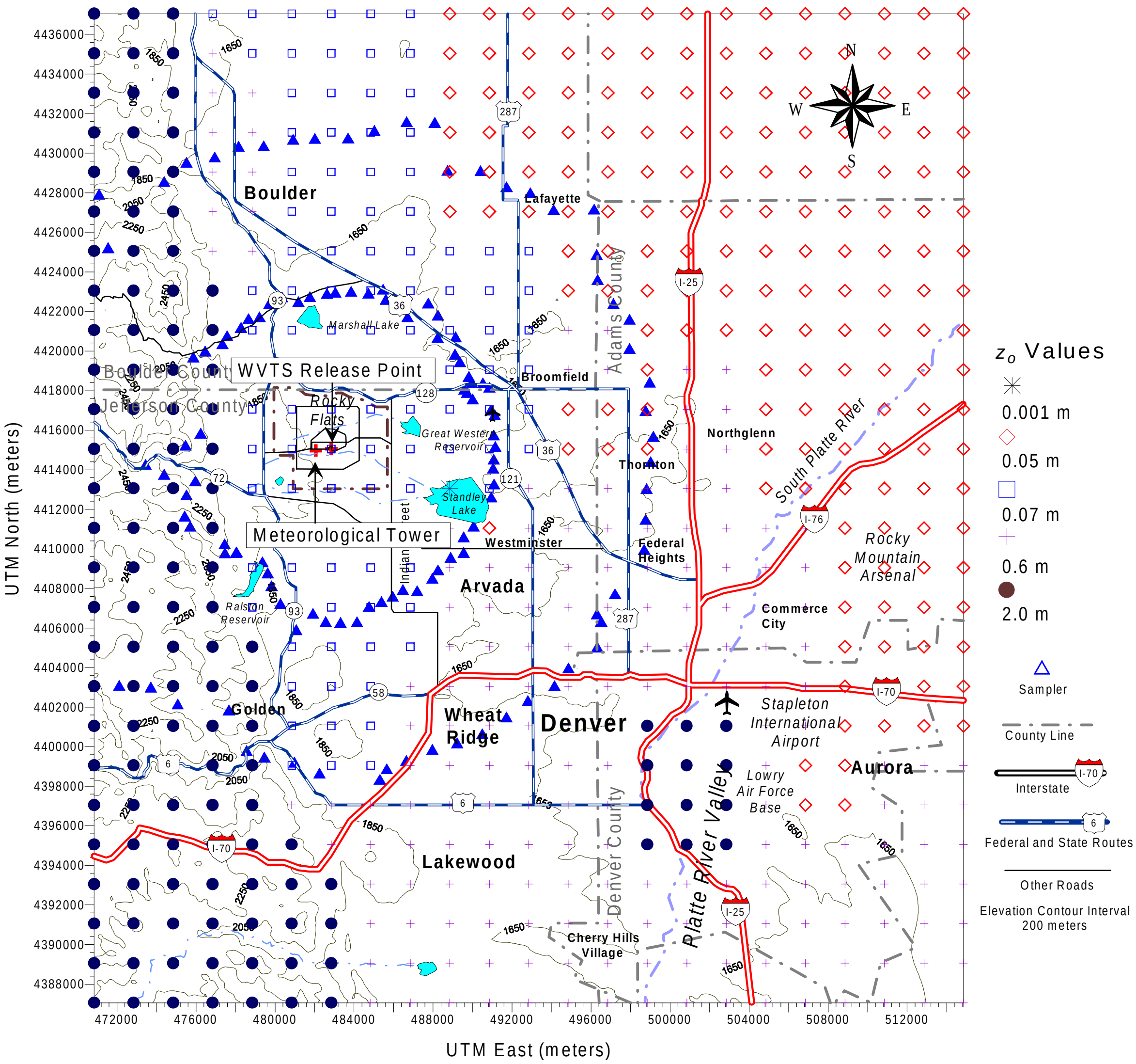

Figure 2. Sampler locations and topography for the 8- and 16-km sampling arcs for the Winter Validation Tracer Study. Five of the 16-km samplers in the western part of the model domain were outside digital elevation model provided by the U.S. Geological Survey. The roughness lengths (Z_o) illustrated were used in the RATCHET simulations.

The ISC model is a simple Gaussian plume model primarily used for regulatory compliance calculations. It is based on the work done by Pasquill (1961) and Gifford (1961) who provided a useful, practical set of guidelines for atmospheric diffusion downwind from a continuous point source under atmospheric conditions ranging from stable to very unstable. The short-term version of the model used in this study was designed to estimate hourly average concentrations from steady-state releases. The model is often called a straight-line model because it assumes that the mean average wind speed and direction do not change in time or space over the period being analyzed. It also assumes that diffusion takes place over a homogeneous flat plane. These assumptions definitely do not apply at the RFP and its environs. Terrain differences up to the release height may be accounted for in the model, but these options were not used in the simulations reported. Meteorological data taken at the 10-m level from the 61-m station located 780 m west and 80 m south of the release point were used in these simulations.

The ISC simulations were performed by researchers at Colorado School of Mines. The ISC simulations were performed unconventionally. The short-term version of ISC is designed to accept hourly meteorological data and provide 1-hour average concentrations at designated receptor locations. The meteorological data provided by Rocky Flats contains data taken every 15 minutes (15-minute average wind speed and direction). Rather than average the four 15-minute average observations (wind speed and direction) to yield hourly average conditions, Haugen performed four ISC runs, one for each 15-minute period. He then averaged the four concentrations to yield a 1-hour average concentration at the designated receptor location. Stability classes for these simulations were determined using the lateral turbulence and wind speed method described in EPA (1987).

The TRAC model represents a substantially more complex model that incorporates terrain complexities and a three-dimensional, temporally and spatially variable wind field over the modeling domain. Model input includes the terrain elevations, surface roughness features, and meteorological data taken at various locations and heights in the model domain. The TRAC model uses a similarity theory approach where diffusion coefficients are based on the Bulk Richardson Number and Monin-Obukhov stability scaling length. The model was designed for short-term releases that require emergency response actions and is an integral part of the emergency response program at Rocky Flats. Rocky Flats personnel performed the TRAC simulations under independent technical oversight at CDPHE facilities. These simulations used meteorological data at the 10- and 61-m level from the 61-m station at Rocky Flats and 16 additional stations located outside the 16-km sampling arc. The results of these simulations (and the ISC simulations) were transferred to Dr. Duane Haugen of Colorado School of Mines. Dr. Haugen generously provided this information to the author of this report in electronic format.

The RATCHET code was written for the Hanford Environmental Dose Reconstruction Project and is a major rework of the MESOIL2 model (Ramsdell and Burk 1991). Significant changes were made to the representation of atmospheric dispersion process and incorporation of Monte Carlo methods to represent uncertainty. RATCHET and was specifically designed for the Hanford Dose Reconstruction Project. The model uses puff dispersion algorithms to transport material in a temporally and spatially varying wind field. Terrain complexities are not explicitly treated within the model, but effects of topography can be incorporated by using meteorological data that reflect the effects of major topographical features and incorporation of spatially varying roughness length. Diffusion coefficients are estimated from statistics for atmospheric turbulence, which are in turn estimated from atmospheric conditions, e.g., wind speed, atmospheric stability, and surface roughness.

Simulations described in this study were performed by the author and only used the deterministic features of the code. The RATCHET source code was obtained from J. V. Ramsdell of Pacific Northwest Laboratories and required modification for compilation on a MS DOS♭-based microcomputer. Other modifications were needed to adapt the code output for use in this study. Code modifications are documented in Rood (1995). RATCHET allows users to input environmental parameters such as roughness length and precipitation, which may be spatially variable over the model domain. Roughness length varied considerably within the model domain. West of the release point, the foothills of the front range are encountered and roughness lengths are about 2 m (see Figure 2). East of the release point, flat farmland and grass fields predominate where roughness lengths of 0.07 to 0.05 m are appropriate. Southeast of the release point, residential communities are present and roughness lengths are closer to 0.6 m. With a few exceptions, most tracer plumes traveled east of the facility, so the roughness length estimates in this region of the model domain may have a greater influence on the results. A deposition velocity of zero was assumed for all simulations, which is appropriate for the inert gas (SF_6) that was released. Meteorological data for the RATCHET simulations included data at the 10-m level from the 61-m tower at the RFP and hourly observations from Denver Stapleton International Airport located about 25 km southeast of the plant.

The TRIAD model (Hicks et al. 1989) is basically an enhanced version of the INPUFF2 model (Petersen and Lavdas 1986). TRIAD is a Gaussian Integrated Puff model and is capable of simulating the accidental release of a substance over several minutes or modeling the more typical continuous release from a stack. Unlike INPUFF2, TRIAD includes a wind field interpolator so transport in a spatial and temporally varying wind field can be calculated. We included TRIAD in our evaluation because of this feature. Three dispersion algorithms are available in the TRIAD and INPUFF2 codes. These dispersion algorithms are the Pasquill-Gifford (P-G) scheme (Turner 1970), the onsite scheme (Irwin 1983) for short travel time dispersion, and a scheme for long travel times in which the growth of the puff becomes proportional to the square root of time. The long travel time scheme is used when the value of the horizontal diffusion coefficient (σ_y) exceeds a user-defined limit and can by disabled by setting this limit to a large value. For the simulations reported in this study, the long travel time scheme was disabled by setting the limiting horizontal diffusion coefficient value at 10,000 m. Meteorological data at the 10-m level from the 61-m tower located at Rocky Flats included measurements of the standard deviation of the horizontal angle and standard deviation of the vertical wind speed, which could have been used in the onsite dispersion scheme. However, these data are not available for the time frame of interest (1952–1988) and we would probably have to resort to Pasquill's scheme to reconstruct past release events. For this reason, Pasquill's scheme was used in these simulations. As mentioned previously, a deposition velocity of zero was assumed for all simulations. The TRIAD model also allows for terrain differences between source and receptor. This feature does not represent a true terrain following algorithm because plumes are allowed to pass through elevated features and not over or around them. For this reason, we ignored the terrain difference feature of TRIAD. The INPUFF2 simulations used meteorological data from the RFP 61-m meteorological tower at the 10-m level. The TRIAD simulations used the 10-m level RFP data in addition to hourly observations taken at Denver Stapleton International Airport.

Except for TRAC, the models evaluated in this study had limited capability to handle complex terrain. Complex terrain is one of the more prominent physical features of the Rocky Flats environs. The limited amount of meteorological data available during the important release events at the RFP may limit the usefulness of a complex terrain model because these models typically require a greater characterization of the wind field than could be obtained from only one or two recording stations. The TRAC model simulations were performed with multiple meteorological recording stations; however, only one station was within the 16-km sampling arc. The lack of meteorological stations limited the effectiveness of the model's ability to predict the plume location and density.

METEOROLOGICAL DATA

Meteorological data for 1991 at Rocky Flats and Denver Stapleton International Airport were obtained in electronic format from the Rocky Flats meteorologist and the National Climatic Data Center, respectively. The Rocky Flats data were measured at a height of 10-and 61-m from a 61-m tower located at the RFP. Only data from the 10-m level were used in the RATCHET, INPUFF2, and TRIAD simulations. The 61-m tower was located 786 m west and 87 m south of the release point (UTM coordinates 482064 E 4414963 N). Each record represented an average over a 15-minute recording period and included wind speed and direction, temperature, heat flux, and standard deviations of these parameters. The 10-m data were cross-checked with the meteorological data provided by Haugen, and good agreement was found between both sets. Haugen's data also included estimates of mixing height and atmospheric stability class, calculated using the lateral turbulence and wind speed method (standard deviation of the horizontal wind direction fluctuations) as described in EPA (1987). The mixing height estimates were derived from linear interpolation for each 15-minute period from the rawinsonde data furnished routinely every 12 hours by the National Weather Service for Denver Stapleton International Airport. Mixing height data were used in the TRIAD and INPUFF2 simulations. The RATCHET model estimates mixing height using a procedure derived in Zilitinkevich (1972). It compares this estimate to a default value provided by the user and ultimately uses the higher of the two estimates in the computations. Default mixing height values were obtained from Haugen's data. No precipitation was measured during any of the 12 tests, and there was not any snow cover during the month of February 1991.

The Denver Stapleton meteorological data were hourly observations taken from a 2.0-m (15-ft) tower located on top of the terminal building. These data were instantaneous observations of wind speed, direction, temperature, cloud cover, and ceiling height recorded on-the-hour. This method of recording weather observations was typical of most major airports before changing to the Automated Surface Observation Site recording system. Additional airport data were obtained from the Jefferson County Airport located about 8 km east of the RFP. These data were also on-the-hour observations taken between the hours of 06:00–23:00 mountain standard time and while the airport was open. These data were used primarily to obtain cloud cover and ceiling height estimates for Rocky Flats. Wind speed and direction were not used because these data were incomplete for half of the tests conducted.

The TRAC, RATCHET, TRIAD, models are capable of incorporating a spatially and temporally varying wind field based on interpolation from several meteorological recording stations in the model domain. TRAC simulations used 16 additional stations located in an 80-km (50-mi) radius from the RFP. Haugen and Fotino (1993) showed these stations had little impact on plume trajectories within 16 km of the plant and suggest they were too far away to have any significant impact on the predicted concentrations at the samplers. For these reasons, no attempt was made to locate these data and incorporate them in the RATCHET and TRIAD simulations.

Denver Stapleton airport data are the only complete meteorological record we have for the period the RFP operated. Meteorological data at Rocky Flats before 1984 is sporadic and of questionable integrity. However, conditions at Denver Stapleton airport are known to differ from those at Rocky Flats. Wind roses constructed using RFP data from 1989–1993 indicate the predominant surface wind direction to be from the west-northwest. Wind roses for the Denver Stapleton International Airport during the same period show the predominate surface wind direction to be from the south. Surface observations from Denver Stapleton are strongly influenced by air movement within the Platte River Valley, which flows to the northeast from the city center. By contrast, Rocky Flats is more strongly influenced by its proximity to the foothills. Both locations are influenced by the diurnal pattern of upslope-downslope conditions that characterize the general air movement on the Colorado Front Range environs. Downslope conditions typically occur during the evening hours and are characterized by drainage flow of cooler surface air from the foothills and upper reaches of the Platte River Valley to the plains. Airflow at Rocky Flats is typically from the west-northwest and converges with the flow from the south within the Platte River Valley in a broad zone 25 to 30 km northwest of the RFP. During daylight hours and after surface heating has eliminated the cooler surface layer, the downslope conditions cease. This is followed by a brief period of relatively calm winds, which in turn is followed by return of air up the valley or upslope conditions. Surface airflow at Rocky Flats is typically from the east during upslope conditions and from the northeast at Denver Stapleton International Airport.

Ideally, we would like to have several meteorological stations east of Rocky Flats to capture this air movement, but they were not in existence during the WVTS. The inclusion of the Denver Stapleton data was intended to at least capture the salient features of air flow in the model domain. Discrepancies between model predictions and observation for this tracer study are likely to be similar to those encountered while making air concentration predictions from actual releases, because the same constraints on the meteorological data are applied to both situations.

Data Transformations

Hourly average meteorological conditions needed to be calculated from the 15-minute data for RATCHET. Hourly average wind speed and direction were calculated from the RFP meteorological data using the protocol described in EPA (1987) and Turner (1964). An arithmetic average of the wind direction was computed first and then segregated into 1 of 36, 10-degree increments. Atmospheric stability was also calculated from these data using a general classification scheme discussed in Pasquill (1961), Gifford (1961), and Turner (1964). This typing scheme employs seven stability classes ranging from A (extremely unstable) to G (extremely stable) and requires estimates of sky cover and ceiling height. Cloud cover and ceiling height data were not available at the RFP and were incomplete at Jefferson County Airport. Cloud cover data from Denver Stapleton International Airport were used to supplement the missing Jefferson County Airport data.

The INPUFF2 and TRIAD models were capable of incorporating meteorological data every 15 minutes. Therefore, the unmodified 15-minute RFP data were used in these simulations. The stability classes estimated by Haugen were also used but were modified according to the procedure outlined in the INPUFF2 and TRIAD user's manual. INPUFF2 uses the P-G curves as described in Turner (1970). However, for neutral atmospheric conditions, two dispersion curves, as suggested by Pasquill (1961) are incorporated into the model. According to the procedure, the neutral stability class (D) is separated into a day (D-day) and night (D-night) curve resulting in seven stability class categories. Hourly average wind speed, wind direction, and stability class are summarized in Appendix A.

MODELING OBJECTIVES

Modeling objectives should support the endpoints of the overall study. For example, the modeling objectives would be quite different if the endpoints of the study were to determine a bounding estimate of the maximum exposed individual as opposed to the endpoint of estimating annual-average exposure to residents downwind of a facility.

The overall endpoint of the Rocky Flats Dose Reconstruction Study are estimates of lifetime carcinogenic incidence risk to hypothetical receptors residing at fixed locations in the model domain during all or part of the operation time of the RFP (1952–1989). The primary pathway of exposure is inhalation of contaminated airborne particulates. Lifetime cancer incidence risk requires estimates of the integrated exposure one would receive while present in the model domain. Therefore, the 9-hour averaged concentration was a more appropriate quantity to evaluate rather than hourly-averages. The study domain extends 28 km south, 12 km west, 22 km north, and 32 km east from the RFP. Most of the Denver metropolitan area and the city of Boulder are included in the domain. Other endpoints, such as the maximum hourly average concentration a receptor was exposed to may be interest in terms of providing bounding estimates of exposure and risk.

Major release events at the RFP include glove box fires in 1957 and 1969 and suspension of plutonium-contaminated soil from the 903 pad area, which occurred primarily in 1969. Duration of these events ranged from about 12 hours for the fires to several days for significant releases from 903 pad. In addition to the major release events, exposure from routine releases of plutonium and other contaminants are also to be evaluated.

Four modeling objectives are defined for this study:

1. Maximum hourly average concentration (unpaired in space and time)
2. Paired time-averaged concentration
3. Unpaired time-averaged concentration
4. Arc-integrated concentration.

For each modeling objective, results are presented separately for the 8- and 16-km sampling arcs. Descriptions of each modeling objective follow.

Maximum Hourly Average Concentration

This modeling objective compares the predicted and observed maximum, 1-hour average concentration measured at a sampler during the 9-hour test period at either an 8- or 16-km distance from the release point. It evaluates how well a model predicts the maximum hourly average concentration within a 9-hour sampling interval. The predicted maximum concentration was not paired in space and time with the maximum observed concentration. Each test contributed five predicted-observed data pairs to the evaluation for each sampling arc (one predicted-observed data pair per model $\times$ five models).

Paired Time-Averaged Concentration

This modeling objective compares the predicted and observed time-averaged (9-hour) concentration at a specific sampler. It evaluates how well a model predicts both plume density and location over the 9-hour sampling interval. The 9-hour average concentration was determined by a simple arithmetic average of the nine, 1-hour average concentrations. Sampler data that were missing were not included when computing the average. For example, if 2 of the 9 hours of observed data for a sampler were missing, then the time-averaged concentration for that sampler would be the sum of the concentrations from the 7 valid hours divided by 7. Each individual test contributed 360 predicted-observed data pairs to the evaluation for the 8-km arc (72 predicted-observed data pairs per model simulation × 5 model simulations) and 340 predicted-observed data pairs for the 16-km arc (68 predicted-observed data pairs per model simulation × 5 model simulations). Of the 700 data pairs, only a subset was selected for evaluation based on the criteria stated in the performance measures section of this report.

Unpaired Time-Averaged Concentration

This modeling objective compares the 25 highest predicted and observed time-averaged (9-hour) concentrations. It evaluates how well a model predicts plume density over the averaging interval regardless of its location. The value of 25 was selected based on U.S. Environmental Protection Agency (EPA) guidance for comparing regulatory models. The 9-hour average concentration was determined by a simple arithmetic average of the nine, 1-hour average concentrations as described above. Each individual test contributed 125 predicted-observed data pairs to the evaluation for the 8- and 16-km arc (25 predicted-observed data pairs per model simulation × 5 model simulations).

Arc-Integrated Concentration

This modeling objective compares the predicted and observed time-averaged concentration, integrated along an 8- or 16-km arc surrounding the release point. It evaluates how well a model predicts the ground-level tracer mass in the model domain at a given distance from the release point. This value is related to the vertical dispersion coefficient and the mixing height used in the simulation. The arc integrated concentration is estimated by

$$\Psi = \int_0^x C(x)\ dx \qquad (1)$$

where

Ψ = arc integrated concentration (ng m^{-2})

$C(x)$ = nine-hour averaged concentration as a function of arc length (ng m^{-3})

x = circumference of either the 8- or 16-km arc.

Each test contributed five predicted-observed data pairs to the evaluation for each sampling arc (one predicted-observed data pair per model simulation × five model simulations).

PERFORMANCE MEASURES

Comparisons of field tracer measurements to predictions made by models (Hanna 1989; Tangirala et al. 1992; Rao and Hosker 1993; Hanna et al. 1993) have focused on using several simplified measures to evaluate model performance (Fox 1981; EPA 1988; Hanna 1989; Cox and

Tikvart 1990; Weil et al. 1992). These measures are the fractional bias (*FB*) and normalized mean square error (*NMSE*). Fractional bias is given by

$$FB = \frac{2(\overline{C_o} - \overline{C_p})}{\overline{C_o} + \overline{C_p}} \tag{2}$$

where C_p and C_o are the predicted and observed concentrations, respectively. Overbars indicate averages over the sample. The *NMSE* given by

$$NMSE = \frac{\overline{(C_o - C_p)^2}}{\overline{C_o}\ \overline{C_p}} \tag{3}$$

where C_p and C_o are the predicted and observed concentration, respectively. Overbars indicate averages over the sample. The *FB* is a measure of mean bias. A *FB* of 0.6 is equivalent to model underprediction by about a factor of 2. A negative value indicates model overprediction. The *NMSE* is a measure of variance, and a value of 1.0 indicates that a typical difference between predictions and observations is approximately equal to the mean. The *NMSE* and *FB* are appropriate when the typical difference between the predictions and observations are approximately a factor of 2 (Hanna et al. 1991). When several of the predictions differ from the observations by a factor of 10 or 100, then a log-transformed measure of model bias and variance is more appropriate because it provides a more balanced approach (Hanna et al. 1991). The log-transformed measures are the geometric mean bias (*MG*) and the geometric mean variance (*VG*) and are defined by

$$MG = \exp\left(\overline{\ln C_o} - \overline{\ln C_p}\right) \tag{4}$$

$$VG = \exp\left[\overline{\left(\ln C_o - \ln C_p\right)^2}\right] \tag{5}$$

where the overbars indicate averages over the sample. Geometric mean bias values of 0.5 and 2.0 indicate a factor of two overprediction and underprediction, respectively. A *VG* value of 1.6 indicates a typical factor of about 2 between the predicted and observed data pairs. A perfect model would have *FB* and *NMSE* values of 0 and *MG* and *VG* values of 1.0.

In response to a recommendation from the Rocky Flats Health Advisory Panel, we have also include the geometric mean (*GM*) and geometric standard deviation (*GSD*) of the predicted-to-observed ratio (C_p/C_o) as a performance measure. The *GM* and *GSD* are given by

$$GM = \exp\left(\overline{\ln \frac{C_p}{C_o}}\right) \tag{6}$$

$$GSD = \exp\left[\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\ln \frac{C_{p_i}}{C_{oi}} - \overline{\ln \frac{C_p}{C_o}}\right)^2}\right] \tag{7}$$

where n = the sample size and overbars indicate averages over the sample. Note that the *MG* is simply the geometric mean of the observed to predicted ratio. Similar conclusions concerning bias and variance may be drawn from either the *FB*, *NMSE*, *MG*, *VG*, *GM*, or *GSD*. Some readers may be more familiar with a given measure; for this reason, we have included all six measures in our evaluation.

In addition to the above measures, the correlation coefficient between predicted and observed values and the number of predictions within a factor of 5 of the observations were also reported. The correlation coefficient was determined using simple linear regression techniques and is a measure of the fraction of the total variation in the predicted concentration that is accounted for by the regression. Scatter plots were also included as qualitative measure of performance for the paired and unpaired time-averaged concentration modeling objectives. The scatter plots are shown on log-log scale and include the ideal correlation line (i.e., r = 1.0) and a log-transformed fit to the data given by

$$\log\left(C_o\right) = a \log\left(C_p\right) + b . \tag{8}$$

The log-transformed regression was used for the time-averaged comparisons because differences between predictions and observations were often quite large, exceeding a factor of 10 in some cases. Log-transformed measures provide a more meaningful comparison when differences between several of the prediction-observed pairs in a data set exceed a factor of 10 (Hanna et al. 1991).

Selection Criteria

The observed data set only reported nonzero hourly average concentrations greater than the minimum detectable sampler concentration (*mdc*) of 33 ng m^{-3}. Measured concentrations below this value were reported as zero. A sampler that had only 1 hour of data (in the 9-hour measurement period) greater than the *mdc* would have a 9-hour average concentration of 33 ng m^{-3} / 9 = 3.7 ng m^{-3}. This value represents the time-averaged mdc for a sampler. We applied this cutoff to the data set so that only predicted *and* observed time-averaged concentration pairs greater than 3.7 ng m^{-3} were considered. If a sampler had missing data for some of the hours, then the average concentration was the average of the valid hourly observations. Samplers missing 9 hours of observation were eliminated from the data set.

Confidence Intervals

Confidence intervals on the previously discussed performance measures have often been estimated to test whether a given measure is significantly different from its optimum or ideal value (Hanna 1989; Hanna et al. 1993; Tangirala et al. 1992; Rao and Hosker 1993). For example, the optimum *FB* value is 0.0. The analysis is then used to see if the confidence interval includes 0. Confidence intervals have also been used for testing the significance of model differences, for example, whether the *NMSE* for model "A" is significantly different than model "B." Bootstrap resampling (Efron 1982) is a popular means of estimating the statistics used to determine confidence intervals. Bootstrapping is a statistical resampling technique used to make nonparametric estimates of statistics of the cumulative density function (*cdf*). This method does not depend on an assumed sample distribution (e.g., normal or lognormal). Rather, it uses the sampled or observed distribution to obtain the estimates. A key assumption in bootstrapping is that the sampled observations are independently and identically distributed. The primary problem with applying this technique to field validation tracer studies is that these assumptions are violated to some degree. Blocking data into sets comprising similar characteristics (i.e., meteorology and source conditions) is an attempt to assure the sampled observations come from identical distributions. However, correlation may still exist between concentrations observed close together in time and space, violating the requirement for sample independence. Tangirala et al. (1992) recognizes this problem and admits the difficulty in estimating these correlations in unequally spaced data. Not accounting for these correlations generally results in underestimating the confidence interval. In our case, time correlations are not an issue because we are only comparing time-averaged concentrations from independent tests and not consecutive hourly observations. However, spatial correlations remain a problem.

To overcome this problem, the sample would have to consist of time-averaged concentrations from a single sampler. The sample can then be blocked according to similar meteorological conditions (i.e., nighttime, daytime, or day-night) (see the Sample Blocking section in this report). For example, a sample might consist of the time-averaged concentration at a given sampler, blocked for nighttime, daytime, and day-night conditions. We are then confronted with the problem of a small sample size. For the previous example, the sample size would be 12 if all concentrations were above the *mdc*. Predicted or observed concentration pairs that are 0 would have to be deleted from the sample (at least for the *MG* and *VG* performance objectives), further reducing the sample size. The maximum concentration and arc-integrated concentration modeling objectives do not have this problem; however, the sample size is limited to 12.

Despite these shortcomings, the modeling community has used these techniques to quantify differences between the performance of models for about the last 15 years. Because of this precedent, estimates of confidence intervals were included in this study. Readers may choose to ignore these analyses and draw their own conclusions from the other information presented in the results section. Similar conclusions about model performance have been arrived at regardless of the measures used to evaluate model performance in this study.

Confidence intervals were estimated for all modeling objectives. Ninety-five percent confidence limits were calculated for the *FB, NMSE, MG, VG,* and correlation coefficient (*r*) performance measures using the blocked bootstrap resampling method (Efron 1987; Hanna 1989). Confidence intervals for the maximum concentration and arc-integrated concentration were estimated using the BOOT software (Hanna et al. 1991). The BOOT program calculates the 95% confidence intervals (2.5% and 97.5% points on the *cdf*) on the previously mentioned performance measures and tests whether the performance measures for two models are significantly different at the 95% level. Two methods are used to estimate confidence intervals; the so called seductive method and the robust moment method (Hanna 1989). The seductive method estimates confidence limits based on the *cdf* generated from bootstrapping. The robust-moment method uses the Student's-*t* procedure to estimate confidence limits based on the mean and variance of the resampled distribution. The robust method is the preferred method for estimating points on the tails of the distribution (Hanna 1989); therefore, results are presented for the robust method only.

Sample Blocking

The predicted-observed sample pairs were blocked into those performed at night (Tests 1, 2, 3, 4, 10, and 11); those performed during the day (Tests 6, 7, 9, and 12); and those performed during transition periods (Tests 5 and 8). Blocks were selected based on the persistent meteorological conditions present during the tests. Nighttime tests were performed under downslope conditions where valley drainage flow (winds *from* the west) predominated. Typically, upslope conditions (winds *from* the east) occur during the daylight hours with transition from upslope to downslope conditions occurring during the evening and transition from downslope to upslope occurring during the morning. Results of the paired and unpaired time-averaged concentration are also presented separately for the nighttime, daytime, and day-night transition blocks. However, confidence intervals are only presented for the entire data set. Separate reporting of the different blocks allows for a qualitative comparison of model performance between different meteorological conditions that persisted during the tests.

RESULTS

Appendix B contains the time-averaged concentration plots for each test and sampling arc. A cursory review of these plots reveals that no one model consistently out performs the rest. However, some general observations can be drawn from these graphs. Observed concentrations were substantially lower for the daytime and transition period tests compared to nighttime tests. Most tests showed multiple observed peaks, suggesting several shifts in the mean wind direction during the sampling period. Model-predicted plume widths for the nighttime tests were generally smaller than the corresponding observed plume width. The results that follow are presented by modeling objective.

Maximum One-Hour Concentration

Measures of bias [*FB*, *MG*, and GM (Table 2)] indicated that the ISC model overpredicted concentrations. Confidence intervals on the *FB* and *MG* performance measures included 1.0 for most models except ISC. Confidence intervals on the differences between model performance measures indicated that the bias measures for ISC were significantly different from other models (Figure 3). For TRAC, RATCHET, TRIAD, and INPUFF2, no significant differences between the bias performance measures were noted, and predicted-observed ratios were near 1.0. Correlation coefficients were highest for TRIAD and INPUFF2; in some cases, the differences between these values were significantly different from the other models at the 8-km distance but not at the 16-km distance. Variability measures (*NMSE*, *VG*, and *GSD*) showed TRIAD and INPUFF2 to have the least amount of variability at the 8- and 16-km distance.

Paired Time-Averaged Concentration

Paired time-averaged concentration results are presented in Tables 3, 4, and 5 for night, day, and transition period (day-night) tests, respectively. For the night and day tests, performance measures for the RATCHET model were generally closest to their optimum value (i.e., *FB* and *NMSE* = 0, *MG* and *VG* = 1.0) compared to the other models. Most models had C_p/C_o ratios greater than 1.0 for day tests indicating model overprediction. For transition period tests, all models showed greater variability as indicated by the *NMSE*, *VG*, and *GSD* performance measure. Correlation coefficients were generally lower for transition period tests compared to those for the night and day blocks, especially at the 16-km distance. The ISC model's performance measures were generally closest to their optimum value, and the RATCHET model consistently underpredicted for transition period tests.

Table 2. Performance Measure Results for the Maximum Concentration Modeling Objective

Performance Measure	ISC	TRAC	RATCHET	TRIAD	INPUFF2
8-km Results					
Geometric Mean C_p/C_o	1.9	1.0	0.91	1.2	1.0
Geometric Std C_p/C_o	2.2	2.3	2.7	2.0	1.9
FB	–0.54	0.21	0.14	–0.02	0.19
FB Confidence Interval	–0.95 — –0.17	–0.20 — 0.61	–0.30 — 0.55	–0.27 — 0.20	–0.07 — 0.47
NMSE	0.93	0.66	0.59	0.14	0.32
NMSE Confidence Interval	–0.28 — 2.2	0.044 — 1.2	0.036 — 1.0	0.08 — 0.20	0.072 — 0.55
MG	0.53	1.0	1.1	0.84	0.93
log(MG) Confidence Interval	-1.1 — -0.12	–0.42 — 0.42	–0.54 — 0.67	–0.60 — 0.23	–0.47 — 0.28
VG	2.7	1.9	2.6	1.6	1.5
log(VG) Confidence Interval	0.31 — 1.6	0.20 — 1.1	0.45 — 1.5	0.17 — 0.66	0.17 — 0.66
r	0.62	0.56	0.57	0.89	0.88
r Confidence Interval	0.42 — 0.91	0.28 — 0.90	0.29 — 0.94	0.83 — 0.98	0.81 — 0.97
% within a factor of 5	92	100	83	100	100
16-km Results					
Geometric Mean C_p/C_o	2.7	1.9	0.93	1.6	1.7
Geometric Std C_p/C_o	2.2	3.9	2.5	2.5	2.2
FB	–0.53	0.05	0.54	0.024	–0.003
FB Confidence Interval	–0.87 — –0.23	–0.47 — 0.53	0.07 — 0.95	–0.48 — 0.39	–0.41 — 0.35
NMSE	0.57	0.82	1.4	0.45	0.36
NMSE Confidence Interval	–0.05 — 1.3	0.39 — 1.2	0.29 — 2.3	0.24 — 0.63	0.14 — 0.57
MG	0.38	0.53	1.3	0.65	0.60
log(MG) Confidence Interval	-1.4 — -0.52	-1.3 — 0.001	-0.48 — 0.62	-1.0 — 0.047	-0.94 — -0.072
VG	4.7	8.5	2.3	2.4	2.4
log(VG) Confidence Interval	0.51 — 2.6	0.56 — 3.7	0.49 — 1.1	0.25 — 1.8	0.94 — 0.072
r	0.82	0.48	0.68	0.72	0.80
r Confidence Interval	0.71 — 0.97	0.006 — 0.98	0.51 — 0.91	0.57 — 0.96	0.64 — 1.0
% within a factor of 5	75	75	92	83	92

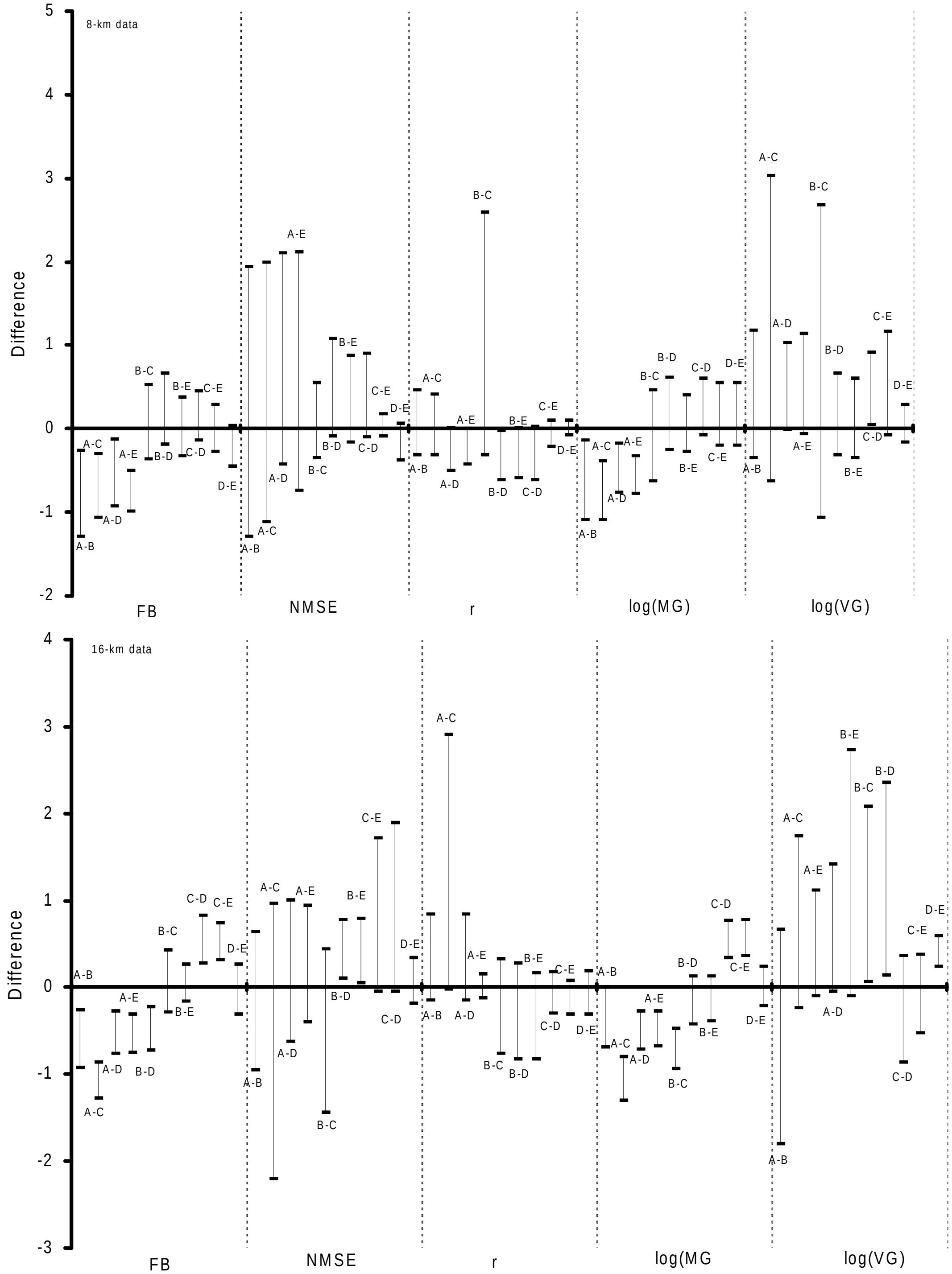

Figure 3. Ninety-five percent confidence intervals on differences in the fractional bias (*FB*), normalized mean square error (*NMSE*), correlation coefficient (*r*), geometric mean bias (*MG*), and geometric mean variance (*VG*) for the maximum concentration modeling objective. Models are identified as follows: ISC = A, TRAC = B, RATCHET = C, TRIAD = D, and INPUFF2 = E.

Table 3. Performance Measure Results for Paired Time-Averaged Concentration, Nighttime Tests

	ISC	TRAC	RATCHET	TRIAD	INPUFF2
8-km Sampling Arc					
Geometric Mean C_p/C_o	0.88	0.43	1.0	0.90	1.1
Standard Deviation C_p/C_o	6.2	5.3	3.6	4.1	4.0
Number of Data Pairs	177	173	258	180	180
Fractional Bias	0.19	0.85	0.13	0.36	0.35
Normalized Mean Square Error	2.9	5.6	1.9	2.7	2.5
Correlation Coefficient[a]	0.16	0.26	0.75	0.53	0.49
Geometric Mean Bias	1.1	2.3	0.97	1.0	0.92
Geometric Mean Variance	28	33	5.0	7.2	6.8
% within a factor of 5	72	74	83	76	77
16-km Sampling Arc					
Geometric Mean C_p/C_o	1.4	0.8	1.5	1.2	1.5
Standard Deviation C_p/C_o	6.9	5.6	3.6	5.9	5.7
Number of Data Pairs	137	132	174	121	128
Fractional Bias	–0.0058	0.47	–0.0032	0.27	0.16
Normalized Mean Square Error	3.4	4.9	2.7	3.0	3.7
Correlation Coefficient[a]	0.15	0.26	0.67	0.0098	0.15
Geometric Mean Bias	0.70	0.12	0.69	0.83	0.67
Geometric Mean Variance	47	20	6.0	23	24
% within a factor of 5	71	73	78	69	71

[a]. A log-transformed regression was performed.

Table 4. Performance Measure Results for Paired Time-Averaged Concentration, Daytime Tests

	ISC	TRAC	RATCHET	TRIAD	INPUFF2
8-km Sampling Arc					
Geometric Mean C_p/C_o	1.8	1.5	1.2	0.93	1.1
Standard Deviation C_p/C_o	5.3	5.3	3.5	4.3	5.0
Number of Data Pairs	94	97	104	97	103
Fractional Bias	-0.58	0.56	0.10	-0.21	-0.31
Normalized Mean Square Error	4.2	2.5	2.5	3.6	3.1
Correlation Coefficient	0.36	0.37	0.49	0.49	0.46
Geometric Mean Bias	0.57	0.69	0.81	1.1	0.94
Geometric Mean Variance	22	18	4.8	8.5	13
% within a factor of 5	59	71	83	64	71
16-km Sampling Arc					
Geometric Mean C_p/C_o	2.0	2.4	1.4	0.96	0.99
Standard Deviation C_p/C_o	5.1	6.4	3.0	3.5	4.8
Number of Data Pairs	65	55	65	53	54
Fractional Bias	–0.89	–1.2	–0.35	–0.21	–0.33
Normalized Mean Square Error	3.6	13	1.2	1.4	1.6
Correlation Coefficient	0.15	0.013	0.53	0.35	0.16
Geometric Mean Bias	0.49	0.42	0.72	1.0	1.0
Geometric Mean Variance	23	64	3.7	4.9	11
% within a factor of 5	63	42	83	75	65

[a] A log-transformed regression was performed.

Table 5. Performance Measure Results for Paired Time-Averaged Concentration, Transition Period Tests

	ISC	TRAC	RATCHET	TRIAD	INPUFF2
8-km Sampling Arc					
Geometric Mean C_p/C_o	1.1	0.52	0.34	0.59	0.39
Standard Deviation C_p/C_o	7.4	10	6.2	7.7	9.5
Number of Data Pairs	43	47	93	57	59
Fractional Bias	0.79	1.5	1.3	1.2	1.3
Normalized Mean Square Error	7.2	20	28	15	18
Correlation Coefficient	0.56	0.44	0.53	0.53	0.44
Geometric Mean Bias	0.87	1.92	2.9	1.7	2.6
Geometric Mean Variance	51	294	95	80	359
% within a factor of 5	65	49	63	72	58
16-km Sampling Arc					
Geometric Mean C_p/C_o	0.96	1.4	0.38	0.94	0.84
Standard Deviation C_p/C_o	7.0	7.6	6.2	7.0	8.9
Number of Data Pairs	32	38	57	25	27
Fractional Bias	0.33	0.67	0.81	0.48	0.63
Normalized Mean Square Error	3.5	7.7	7.1	4.8	4.6
Correlation Coefficient	0.32	0.14	0.40	0.49	0.23
Geometric Mean Bias	1.0	0.70	2.6	1.1	1.2
Geometric Mean Variance	39	64	65	24	103
% within a factor of 5	66	68	51	70	56

[a] A log-transformed regression was performed.

A qualitative observation of scatter plots of all tests combined (Figures 4 and 5) suggests models tend to underpredict higher observed concentrations and overpredict lower observed concentrations. RATCHET predictions appear to most closely matched observations, and the log-transformed performance measures appear to support this observation. Box-and-whisker plots illustrate how the C_p/C_o ratios are distributed (Figure 6). All models had predictions that exceeded the corresponding observed value by a factor of 10, indicating the log-transformed performance measures are a more appropriate measure of performance. Ninety percent of the RATCHET predictions were within a factor of 10 of the observations. Bias measures at the 8-km distance (Table 6) suggest TRAC and RATCHET were biased slightly low. Ninety-five percent confidence intervals on the *MG* performance measure indicated no significant bias for ISC, TRIAD, and INPUFF2 at the 8-km distance. At the 16-km distance, the 95% confidence interval on the *MG* performance measure included 1.0 for all models except ISC. The ISC model overpredicted at that distance. The log-transformed measure of variability (*VG*) indicated RATCHET had the least amount of variability among the models at both sampling distances. Also, the *VG* value for RATCHET was significantly different (at the 95% level) from the *VG* value for the others models (Figure 7). Correlation coefficients for RATCHET were also the highest among the models and were significantly different from the other models at both sampling arcs.

Table 6. Performance Measure Results for Paired Time-Averaged Concentration, All Tests

Performance Measure	ISC	TRAC	RATCHET	TRIAD	INPUFF2
8-km Results					
Geometric Mean C_p/C_o	1.1	0.64	0.86	0.88	0.91
Geometric Std C_p/C_o	6.2	6.4	4.4	4.7	5.3
FB	0.26	0.89	0.46	0.57	0.58
FB Confidence Interval	0.023 — 0.46	0.72 — 1.1	0.27 — 0.65	0.36 — 0.78	0.39 — 0.77
NMSE	5.0	12	7.8	7.7	7.8
NMSE Confidence Interval	3.3 — 6.6	8.3 — 15	4.7 — 11	5.1 — 11	4.8 — 11
MG	0.89	1.6	1.2	1.1	1.1
log(MG) Confidence Interval	–0.29 — 0.097	0.29 — 0.67	0.041 — 0.29	–0.016 — 0.32	–0.052 — 0.29
VG	28	38	9.1	11	16
log(VG) Confidence Interval	2.7 — 3.9	3.0 — 4.2	1.8 — 2.6	1.9 — 2.9	2.2 — 3.3
r[a]	0.41	0.36	0.67	0.61	0.55
r Confidence Interval	0.31 — 0.51	0.26 — 0.45	0.60 — 0.72	0.53 — 0.67	0.47 — 0.63
% within a factor of 5	67	69	79	72	72
16-km Results					
Geometric Mean C_p/C_o	1.5	1.2	1.1	1.1	1.3
Geometric Std C_p/C_o	6.5	6.4	4.3	5.3	5.9
FB	–0.02	0.35	0.12	0.29	0.22
FB Confidence Interval	–0.32 — 0.20	0.11 — 0.63	–0.085 — 0.31	0.027 — 0.50	-0.04 — 0.48
NMSE	4.0	6.1	3.7	4.0	4.6
NMSE Confidence Interval	1.2 — 6.6	2.2 — 10	1.2 — 6.1	1.6 — 6.2	1.3 — 7.7
MG	0.67	0.87	0.90	0.91	0.80
log(MG) Confidence Interval	–0.66 — –0.19	–0.31 — 0.16	–0.25 — 0.062	–0.32 — 1.2	–0.45 — 0.018
VG	37	32	8.5	16	24
log(VG) Confidence Interval	2.9 — 4.4	2.8 — 4.2	1.8 — 2.4	2.2 — 3.3	2.5 — 3.8
r[a]	0.30	0.28	0.58	0.41	0.31
r Confidence Interval	0.16 — 0.41	0.14 — 0.39	0.51 — 65	0.29 — 0.52	0.21 — 0.43
% within a factor of 5	68	64	74	71	67

[a] A log-transformed regression was performed.

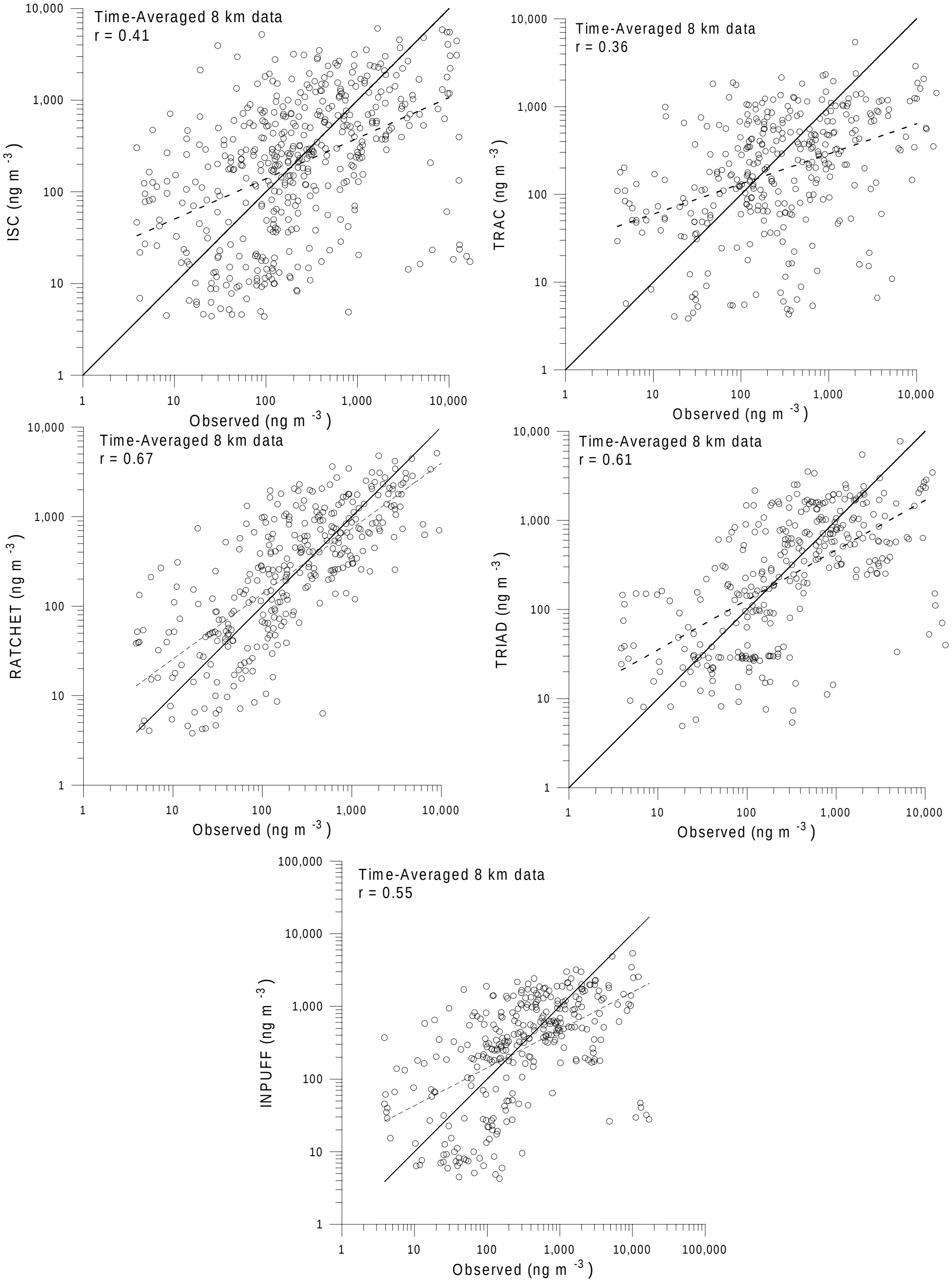

Figure 4. Scatter plots of predicted and observed time-averaged concentration at the 8-km sampling arc. The solid line represents an ideal fit (i.e., Y=X) to the observations. The dashed line represents the regression fit using a power function (Equation 8).

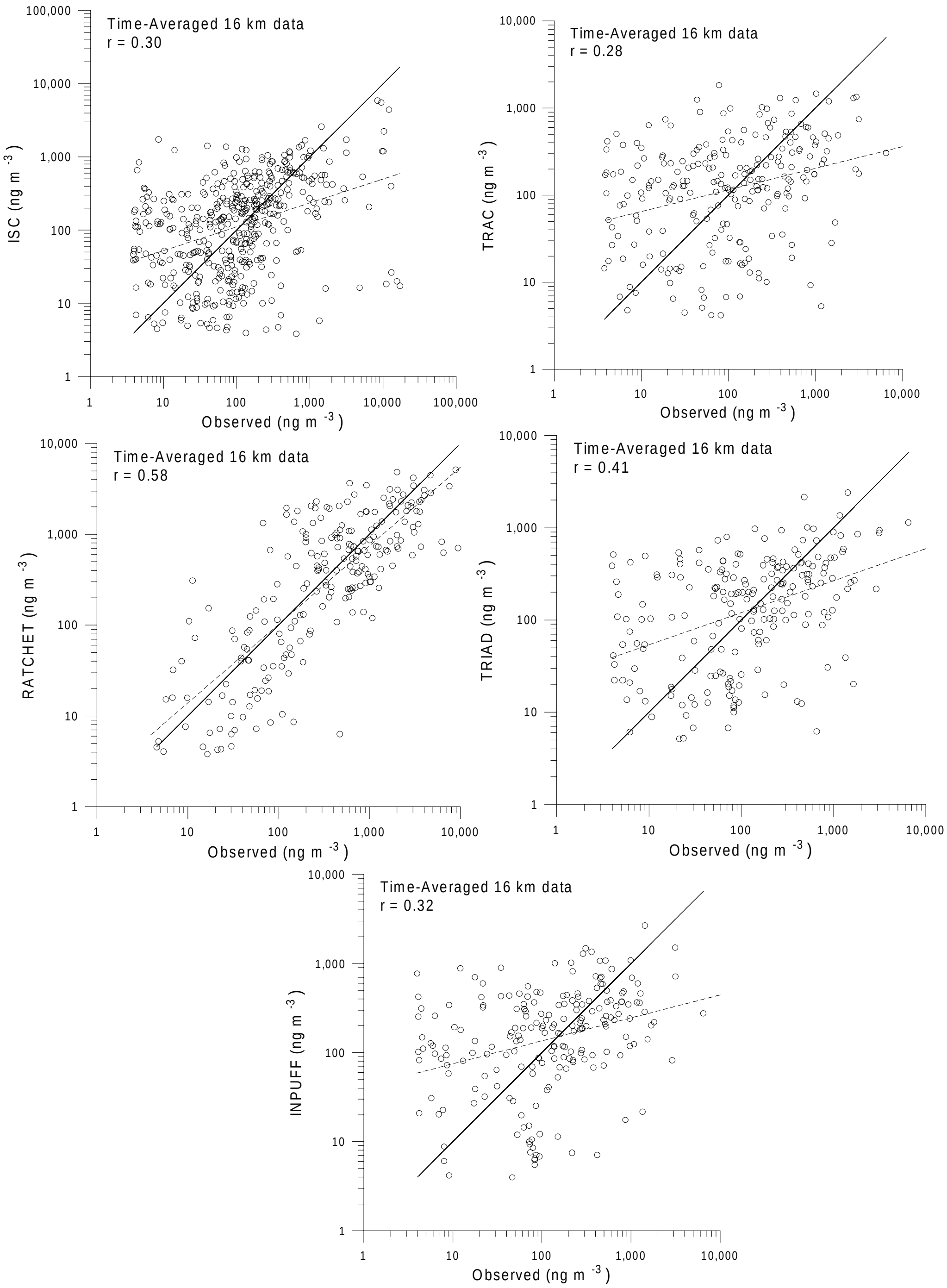

Figure 5. Scatter plots of predicted and observed time-averaged concentration at the 16-km sampling arc. The solid line represents an ideal fit (i.e., Y=X) to the observations. The dashed line represents the regression fit using a power function (Equation 8).

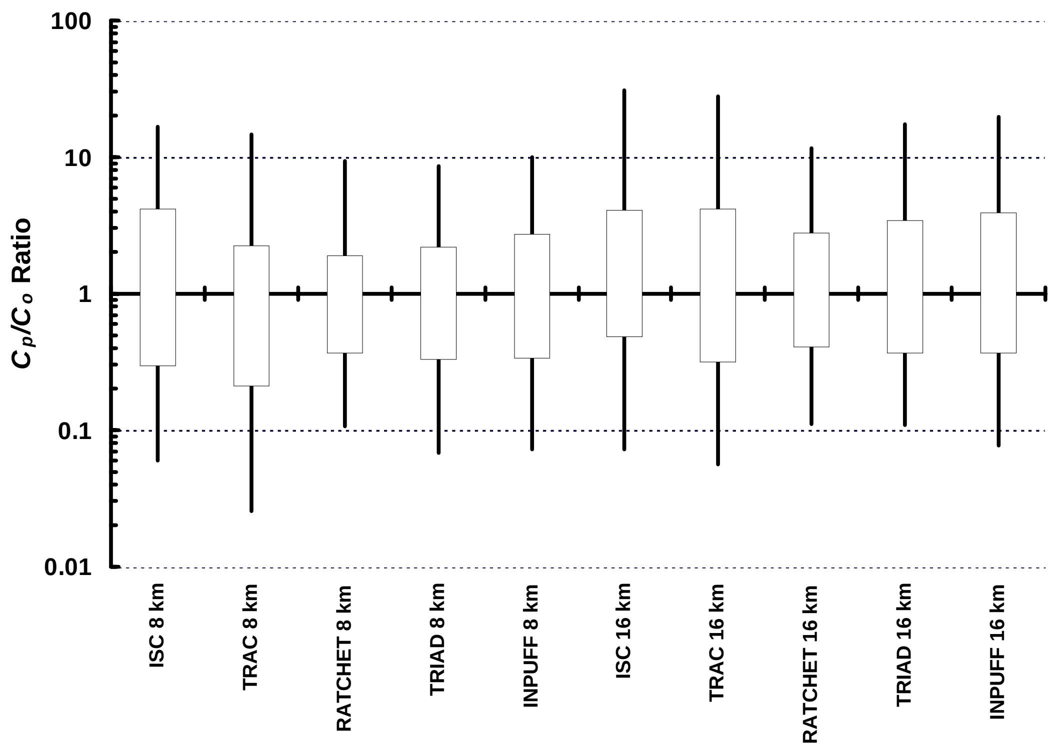

Figure 6. Time-averaged paired C_p/C_o ratios for all tests combined. The shaded rectangles represent the lower and upper quartiles of the distribution of C_p/C_o. The span of the vertical lines represent the 5th and 95th percentile values of the distribution.

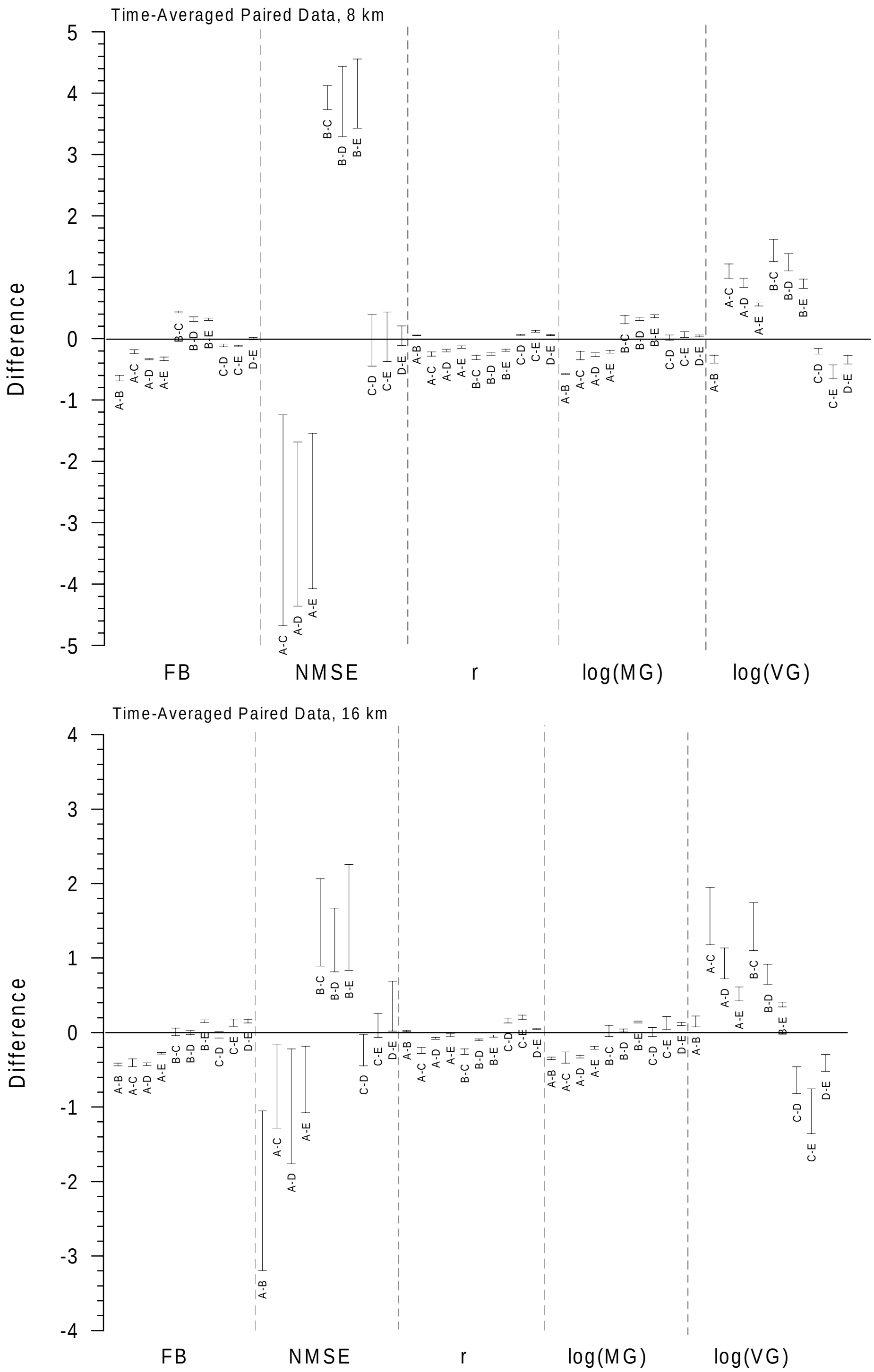

Figure 7. Ninety-five percent confidence intervals on differences in the fractional bias (*FB*), normalized mean square error (*NMSE*), correlation coefficient (*r*), geometric mean bias (*MG*), and geometric mean variance (*VG*) for paired time-averaged concentrations. Models are identified as follows: ISC = A, TRAC = B, RATCHET = C, TRIAD = D, and INPUFF2 = E.

Unpaired Time-Averaged Concentration

The most striking difference between the paired and unpaired comparisons is the improvement in the variability performance measures (*GSD*, *VG*, and *NMSE*) and correlation coefficients (*r*). Approximately 90% of the model predictions were within a factor of 5 of the observations Correlation coefficients typically ranged from about 0.6 to 0.8 with a few exceptions. For the nighttime tests (Table 7), bias performance measures (*FB*, *MG*, and *GM*) indicated overprediction by the ISC and RATCHET model and underprediction by the TRAC model. Bias measures were near their optimum value for the INPUFF2 and TRIAD models.

For daytime tests (Table 8), bias measures indicated overprediction by the ISC and TRAC models and slightly smaller positive bias for RATCHET, TRIAD and INPUFF2. RATCHET generally showed the least amount of variability. The transition period tests again showed erratic results (Table 9). For example, RATCHET has a correlation coefficient of 0.91 for the 8-km arc and 0.29 for the 16-km arc – the smallest and largest correlation coefficient of the data set. Bias measures indicated model underprediction at both sampling arcs.

Scatter plots (Figures 8 and 9) of all the tests combined suggest models tend to underpredict the higher observed concentrations and overpredict lower observed concentrations. A similar conclusion was reached from the paired comparison plots. The ISC predictions appear to best match the observations at the 8-km arc and the correlation coefficient supports this observation. Box-and-whisker plots illustrate how the C_p/C_o ratios are distributed (Figure 10). Ninety-percent of all model predictions were within a factor of 10 of the observations, but some models had differences that exceeded a factor of 10 at the 16-km arc. Therefore, it was not clear whether the arithmetic (*FB* and *NMSE*) or log-transformed (*MG* and *VG*) performance measures were more appropriate.

Correlation coefficients for ISC were significantly different (at the 95% level) from the other models (Figure 11). Model performance was degraded somewhat at the 16-km arc as indicated by lower correlation coefficients. The *FB* performance measure for all tests (Table 10) at the 8-km distance suggest all models underpredicted (most notably, TRAC), and the 95% confidence interval on *FB* excluded 0 for all models. However, the log-transformed performance measures indicated only TRAC and TRIAD underpredicted concentrations. At the 16-km distance, bias measures indicated overprediction by the RATCHET and ISC models. The box-and-whisker plots confirm this observation. However, the 95% confidence interval on the *FB* performance measure included 1.0 for these models indicating no significant bias. Bias measures, for the most part, were significantly different among the models. Smaller variability relative to the paired comparisons was noted. The ISC model exhibited the least amount of variability as indicated by the *NMSE* performance measure at both sampling arcs. However, the log-transformed performance measures suggested RATCHET performed similarly. The *NMSE* value for ISC was significantly different (at the 95% level) from the other models.

Table 7. Performance Measure Results for Unpaired Time-Averaged Concentration, Nighttime Tests

	ISC	TRAC	RATCHET	TRIAD	INPUFF2
8-km Sampling Arc					
Geometric Mean C_p/C_o	1.2	0.51	1.5	0.92	1.0
Standard Deviation C_p/C_o	1.9	2.5	2.1	2.6	2.3
Number of Data Pairs	150	149	150	145	147
Fractional Bias	0.081	0.71	–0.11	0.23	0.21
Normalized Mean Square Error	0.68	1.9	0.53	0.86	0.88
Correlation Coefficient[a]	0.78	0.59	0.68	0.52	0.59
Geometric Mean Bias	0.84	2.0	0.69	1.1	0.98
Geometric Mean Variance	1.5	3.7	2.0	2.4	2.0
% within a factor of 5	97	91	93	95	95
16-km Sampling Arc					
Geometric Mean C_p/C_o	1.7	0.84	2.0	1.2	1.4
Standard Deviation C_p/C_o	2.4	2.5	2.3	3.0	2.7
Number of Data Pairs	147	137	150	126	128
Fractional Bias	–0.16	0.40	–0.14	0.18	0.096
Normalized Mean Square Error	1.3	2.4	2.2	1.4	1.5
Correlation Coefficient[a]	0.75	0.70	0.86	0.53	0.60
Geometric Mean Bias	0.58	1.2	0.49	0.85	0.73
Geometric Mean Variance	2.8	2.3	3.2	3.4	3.0
% within a factor of 5	87	92	86	83	87

[a]. A log-transformed regression was performed.

Table 8. Performance Measure Results for Unpaired Time-Averaged Concentration, Daytime Tests

	ISC	TRAC	RATCHET	TRIAD	INPUFF2
8-km Sampling Arc					
Geometric Mean C_p/C_o	1.7	1.8	1.0	1.0	1.7
Standard Deviation C_p/C_o	2.5	1.8	1.7	2.6	2.0
Number of Data Pairs	90	88	84	86	91
Fractional Bias	–0.53	–0.60	0.099	–0.23	–0.39
Normalized Mean Square Error	3.0	0.96	1.4	2.2	2.0
Correlation Coefficient	0.73	0.87	0.85	0.68	0.81
Geometric Mean Bias	0.60	0.55	1.0	0.97	0.59
Geometric Mean Variance	3.0	2.0	1.4	2.5	2.1
% within a factor of 5	93	100	100	94	99
16-km Sampling Arc					
Geometric Mean C_p/C_o	2.7	2.1	1.3	0.98	1.1
Standard Deviation C_p/C_o	2.2	3.0	2.0	1.9	2.1
Number of Data Pairs	79	80	69	65	63
Fractional Bias	–1.0	–1.1	–0.44	–0.37	–0.44
Normalized Mean Square Error	4.3	9.8	0.53	1.3	1.0
Correlation Coefficient	0.72	0.73	0.81	0.84	0.86
Geometric Mean Bias	0.37	0.48	0.75	1.0	0.94
Geometric Mean Variance	5.0	5.6	1.7	1.5	1.7
% within a factor of 5	72	79	97	100	100

[a] A log-transformed regression was performed.

Table 9. Performance Measure Results for Unpaired Time-Averaged Concentration, Transition Period Tests

	ISC	TRAC	RATCHET	TRIAD	INPUFF2
8-km Sampling Arc					
Geometric Mean C_p/C_o	0.75	0.68	0.70	0.64	0.67
Standard Deviation C_p/C_o	2.7	3.1	3.3	3.1	3.6
Number of Data Pairs	50	47	50	50	50
Fractional Bias	1.0	1.4	1.3	1.1	1.2
Normalized Mean Square Error	5.9	15	11	8.1	10
Correlation Coefficient	0.89	0.90	0.91	0.85	0.81
Geometric Mean Bias	1.3	1.5	1.4	1.6	1.5
Geometric Mean Variance	2.8	4.1	4.4	4.4	6.0
% within a factor of 5	98	77	84	92	86
16-km Sampling Arc					
Geometric Mean C_p/C_o	0.52	0.80	0.89	0.45	0.50
Standard Deviation C_p/C_o	3.7	2.6	5.4	4.7	4.6
Number of Data Pairs	43	40	50	46	46
Fractional Bias	0.48	0.71	0.70	0.62	0.71
Normalized Mean Square Error	1.5	4.6	5.6	3.2	3.3
Correlation Coefficient	0.67	0.72	0.29	0.53	0.54
Geometric Mean Bias	1.9	1.2	1.1	2.2	2.0
Geometric Mean Variance	8.1	2.7	17	18	15
% within a factor of 5	84	93	56	54	62

[a] A log-transformed regression was performed.

Table 10. Performance Measure Results for Unpaired Time-Averaged Concentration, All Tests

Performance Measure	ISC	TRAC	RATCHET	TRIAD	INPUFF2
8-km Results					
Geometric Mean C_p/C_o	1.2	0.79	1.1	0.89	1.1
Geometric Std C_p/C_o	2.3	2.8	2.3	2.7	2.6
FB	0.30	0.78	0.27	0.47	0.45
FB Confidence Interval	0.17 — 0.44	0.67 — 0.92	0.14 — 0.44	0.33 — 0.60	0.31 — 0.60
NMSE	2.8	6.7	3.2	3.7	4.0
NMSE Confidence Interval	2.3 — 3.5	5.2 — 8.8	2.4 — 4.0	2.8 — 4.4	3.2 — 4.9
MG	0.82	1.3	0.88	1.1	0.90
log(MG) Confidence Interval	–0.29 — –0.12	0.12 — 0.31	–0.22 — 0.041	–0.008 — 0.23	–0.22 — 0.008
VG	2.1	3.1	2.1	2.7	2.4
log(VG) Confidence Interval	0.65 — 0.85	0.89 — 1.3	0.55 — 0.89	0.81 — 1.2	0.75 — 1.0
r	0.84	0.73	0.83	0.78	0.80
r Confidence Interval	0.81 — 0.87	0.68 — 0.80	0.79 — 0.87	0.73 — 0.82	0.76 — 0.84
% within a factor of 5	96	91	94	94	95
16-km Results					
Geometric Mean C_p/C_o	1.6	1.1	1.6	0.93	1.1
Geometric Std C_p/C_o	2.9	2.9	2.9	3.2	3.0
FB	–0.14	0.25	–0.21	0.24	0.17
FB Confidence Interval	–0.29 — 0.015	0.089 — 0.43	–0.18 — 0.18	0.097 — 0.37	0.011 — 0.33
NMSE	1.6	3.3	2.6	2.1	2.1
NMSE Confidence Interval	-0.14 — 3.0	1.1 — 5.3	0.53— 4.3	2.7 — 3.5	0.057 — 3.8
MG	0.62	0.91	0.64	1.1	0.93
log(MG) Confidence Interval	–0.59 — –0.38	–0.24 — 0.008	–0.54 — –0.31	–0.061 — 0.21	–0.21 — 0.057
VG	3.9	3.1	3.6	3.8	3.3
log(VG) Confidence Interval	1.1 — 1.6	0.92 — 1.2	1.1 — 1.4	1.0 — 1.6	0.88 — 1.4
r	0.71	0.71	0.71	0.68	0.69
r Confidence Interval	0.65 — 0.78	0.67 — 0.78	0.66 — 0.76	0.60 — 0.76	0.61 — 0.77
% within a factor of 5	82	88	83	82	85

[a] A log-transformed regression was performed.

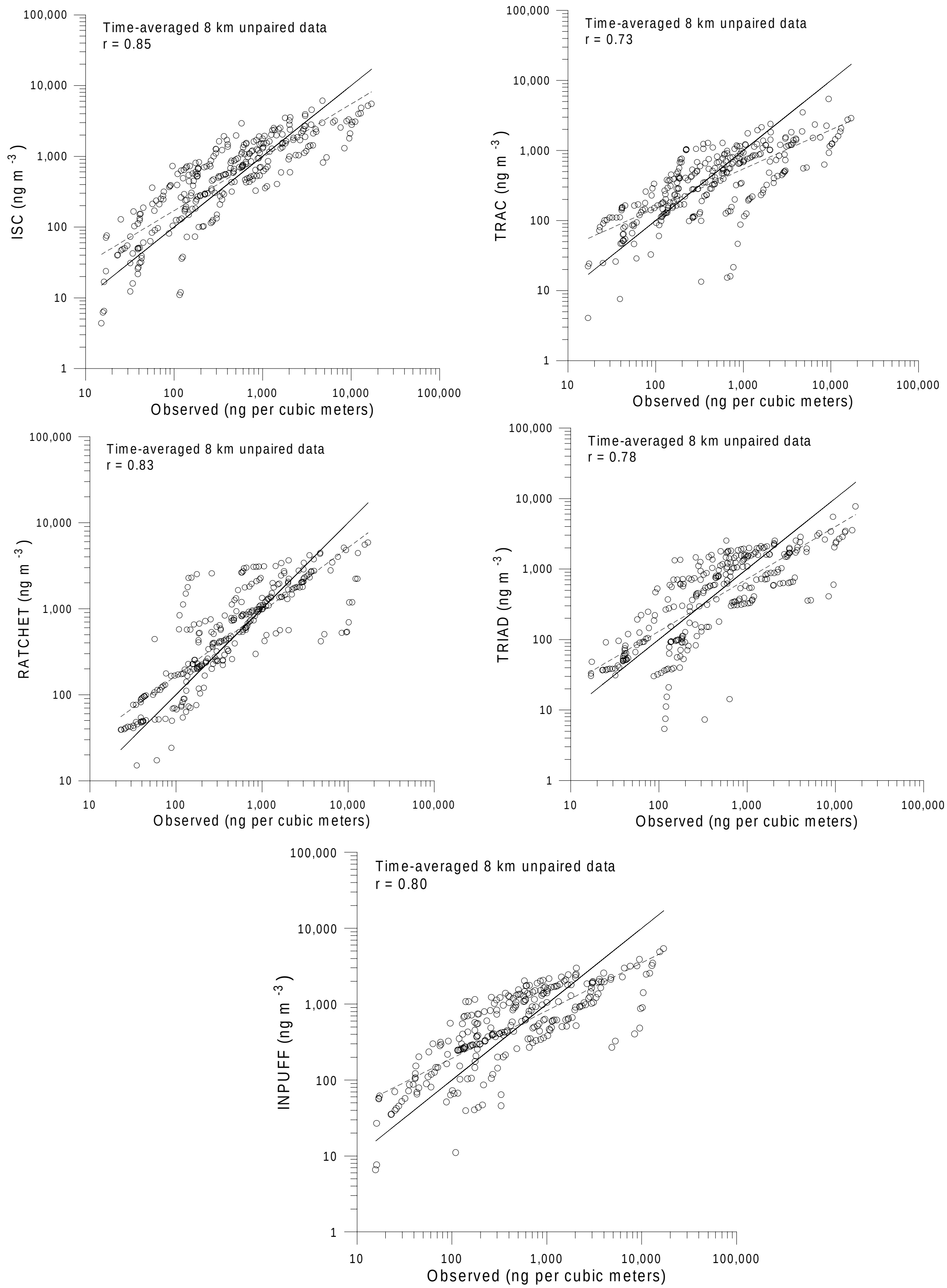

Figure 8 Scatter plots of predicted and observed unpaired time-averaged concentration at the 8-km sampling arc. The solid line represents an ideal fit (i.e., Y = X) to the observations. The dashed line represents the regression fit using a power function (Equation 8).

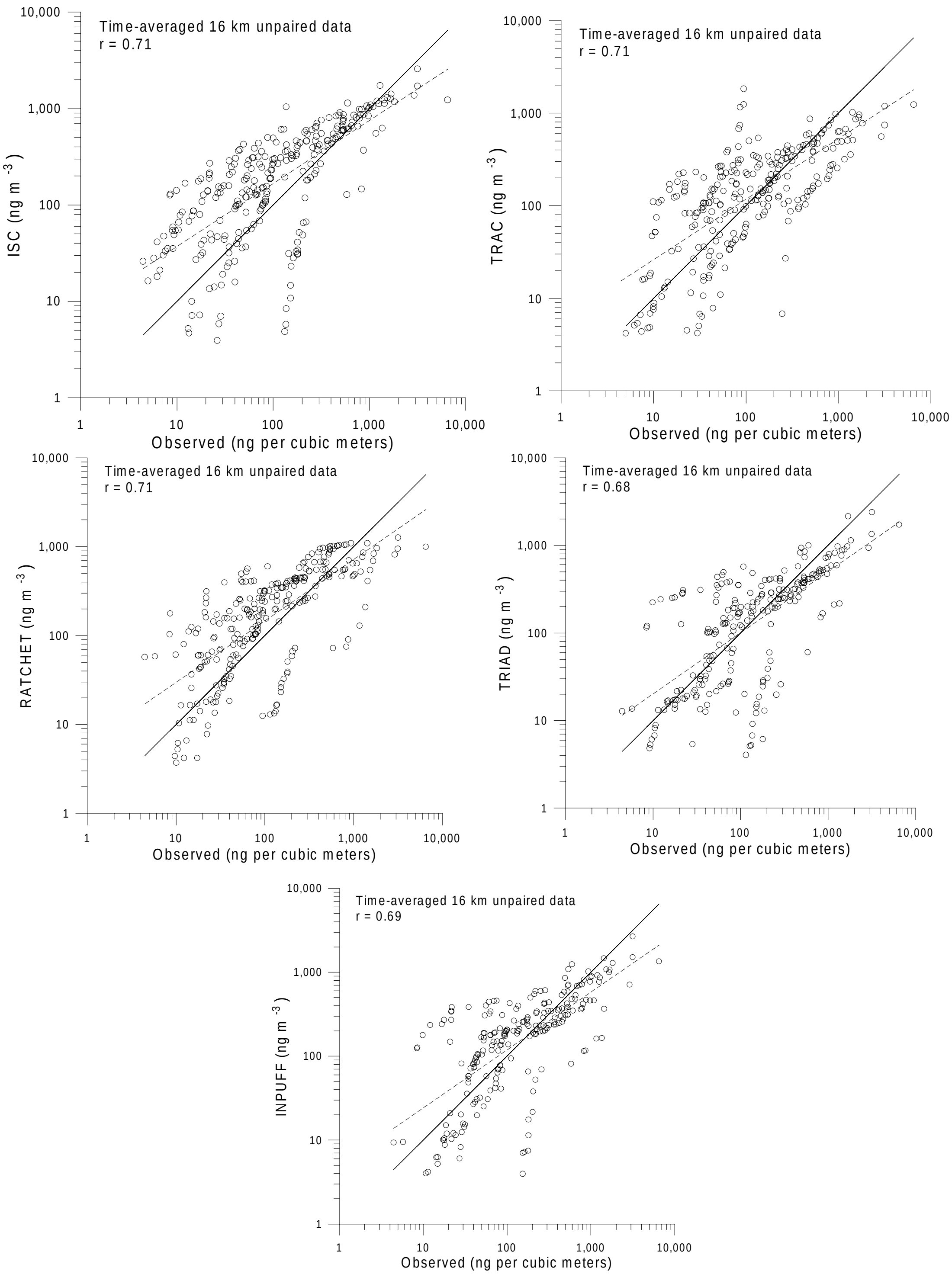

Figure 9 Scatter plots of predicted and observed unpaired time-averaged concentration at the 16-km sampling arc. The solid line represents an ideal fit (i.e., Y = X) to the observations. The dashed line represents the regression fit using a power function (Equation 8).

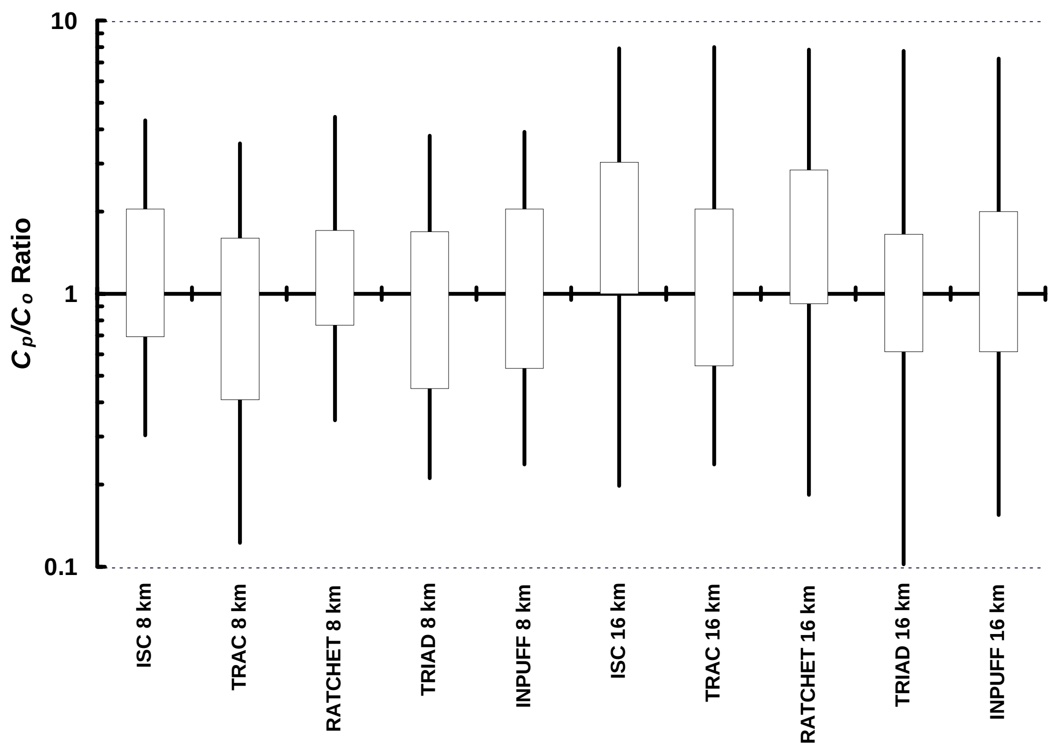

Figure 10. Time-averaged unpaired C_p/C_o ratios for all tests combined. The shaded rectangles represent the lower and upper quartiles of the distribution of C_p/C_o. The span of the vertical lines represent the 5th and 95th percentile values of the distribution.

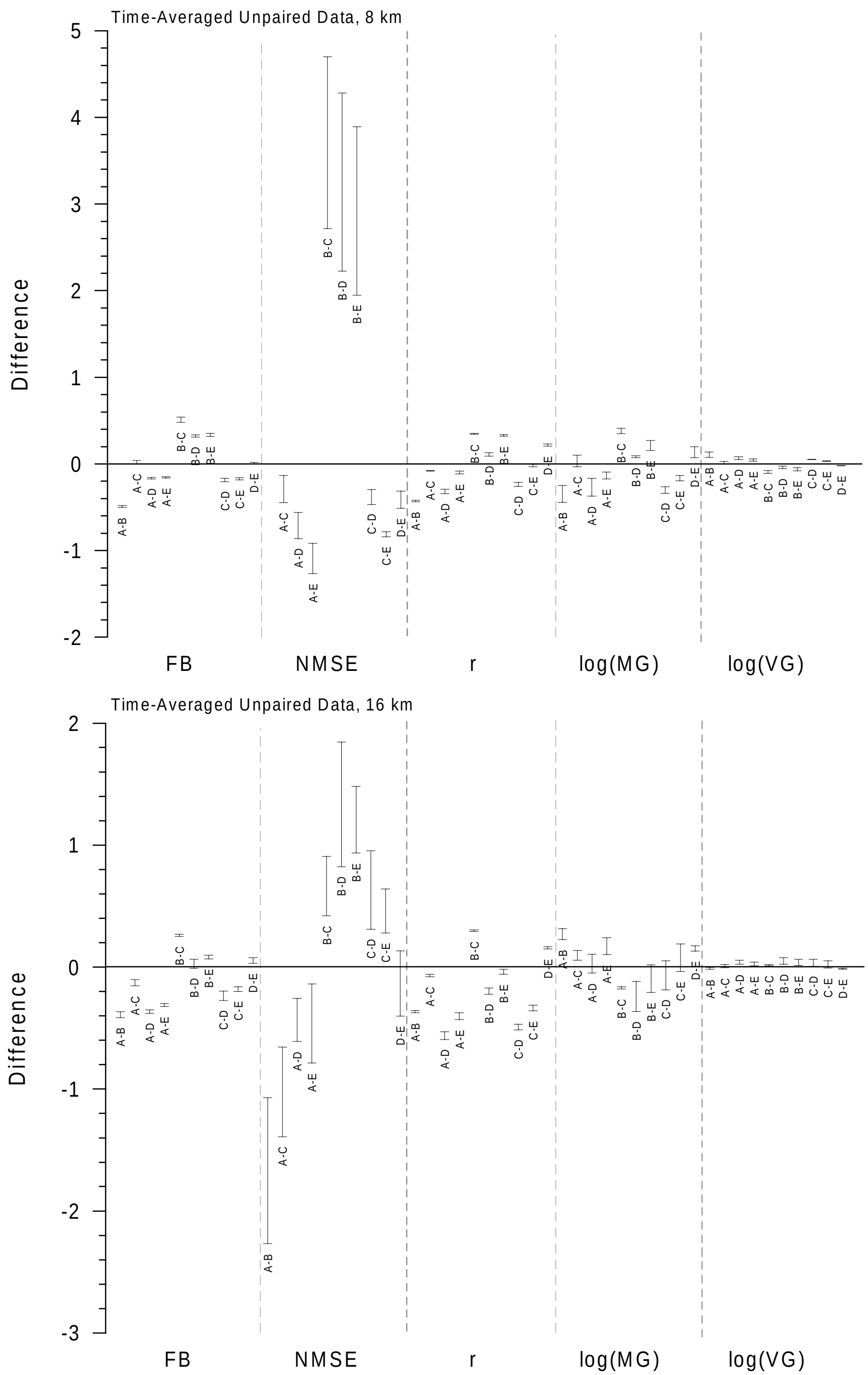

Figure 11. Ninety-five percent confidence intervals on differences in the fractional bias (*FB*), normalized mean square error (*NMSE*), correlation coefficient (*r*), geometric mean bias (*MG*), and geometric mean variance (*VG*) for unpaired time-averaged concentrations. Models are identified as follows: ISC = A, TRAC = B, RATCHET = C, TRIAD = D, and INPUFF2 = E.

Arc-Integrated Concentration

This performance objective provides a measure of a model's ability to estimate the integrated ground-level mass at a given distance from the release point. This measure also indirectly evaluates the vertical diffusion component of each model. A C_p/C_o ratio less than 1.0 would suggest either a mass balance problem or a vertical diffusion coefficient that is too high, resulting in a larger than expected portion of tracer mass above the receptor height (about a meter above ground level). A C_p/C_o ratio greater than 1.0 would suggest a vertical diffusion coefficient that is too low, resulting in a smaller than expected portion of tracer mass above the receptor height. This performance measure removes the plume trajectory from the evaluation and only compares the ground-level tracer mass at a given distance from the release point and not the location of the mass. Models that perform well in this modeling objective but poorly in paired time-average comparisons suggest a problem with the modeled plume trajectory and not with the diffusion coefficients.

Sixty-six to ninety-one percent of the model predictions were within a factor of 3 of the observations. All models tended to have higher C_p/C_o ratios at the 16-km arc compared to the 8-km arc (Table 11). Predicted-observed ratios ranged from 0.66 (TRAC) to 1.1 (ISC) at the 8-km distance and 0.94 (TRIAD) to 1.5 (ISC) at the 16-km distance. The TRAC model exhibited the greatest amount of variability at both sampling arcs as indicated by the *NMSE* and *VG* performance measures. Confidence intervals on the *NMSE* and *VG* performance measures for RATCHET indicated no significant difference (at the 95% level) from their optimum value at the 8-km arc. Bias measures indicated model underprediction at the 8-km arc for all models except ISC. However, the *FB* and *MG* confidence interval included 0 and 1.0 respectively for the RATCHET and ISC models at that distance. At the 16-km distance, the *FB* and *MG* confidence interval included 0 and 1.0, respectively, for all models except ISC. The ISC model showed a positive bias at that distance. The ISC model had the highest linear correlation coefficients at both sampling arcs, but these values were not significantly different from the other models (Figure 12) except for TRIAD at the 16-km distance.

Table 11. Performance Measure Results for the Arc Integrated Concentration Modeling Objective

Performance Measure	ISC	TRAC	RATCHET	TRIAD	INPUFF2
8-km Results					
Geometric Mean C_p/C_o	1.1	0.66	0.80	0.90	0.94
Geometric Std C_p/C_o	2.3	2.7	1.8	2.3	2.5
FB	0.33	0.85	0.35	0.53	0.52
FB Confidence Interval	–0.09 — 0.70	0.44 — 1.2	–0.086 — 0.75	0.089 — 0.92	0.04 — 0.95
NMSE	0.90	2.9	1.1	1.4	1.5
NMSE Confidence Interval	0.077 — 1.6	0.62 — 4.9	–0.39 — 2.3	0.098 — 2.5	–0.069 — 2.8
MG	0.88	1.5	1.2	1.1	1.1
log(MG) Confidence Interval	–0.58 — 0.32	0.002 — 0.83	–0.12 — 0.54	–0.37 — 0.54	–0.43 — 0.54
VG	1.9	3.0	1.4	1.9	2.2
log(VG) Confidence Interval	0.45 — 0.86	0.32 — 1.8	–0.12 — 0.82	0.32 — 0.93	0.36 — 1.2
r	0.68	0.38	0.54	0.58	0.47
r Confidence Interval	0.49 — 0.95	–0.041 — 0.90	0.27 — 0.99	0.27 — 1.0	0.074 — 0.99
% within a factor of 5	100	83	92	92	92
16-km Results					
Geometric Mean C_p/C_o	1.5	1.0	1.2	0.94	0.97
Geometric Std C_p/C_o	2.0	2.8	2.1	1.9	1.9
FB	0.003	0.41	0.10	0.38	0.34
FB Confidence Interval	–0.37 — –0.34	–0.026 — 0.81	–0.355 — 0.52	–0.034 — 0.75	–0.11 — 0.74
NMSE	0.37	1.1	0.55	0.82	0.83
NMSE Confidence Interval	0.129 — 0.57	0.19 — 2.0	0.079 — 0.98	0.12 — 1.4	0.16 — 1.4
MG	0.68	0.97	0.86	1.1	1.0
log(MG) Confidence Interval	-0.73 — -0.043	–0.43 — 0.37	–0.62 — 0.31	–0.34 — 0.46	–0.36 — 0.42
VG	1.8	2.6	1.7	1.5	1.5
log(VG) Confidence Interval	0.24 — 0.93	0.41 — 1.5	0.41 — 0.62	0.24 — 0.60	0.22 — 0.60
r	0.74	0.47	0.63	0.69	0.60
r Confidence Interval	0.56 — 0.97	0.074 — 0.92	0.31 — 1.0	0.50 — 0.94	0.33 — 0.97
% within a factor of 5	100	92	92	100	100

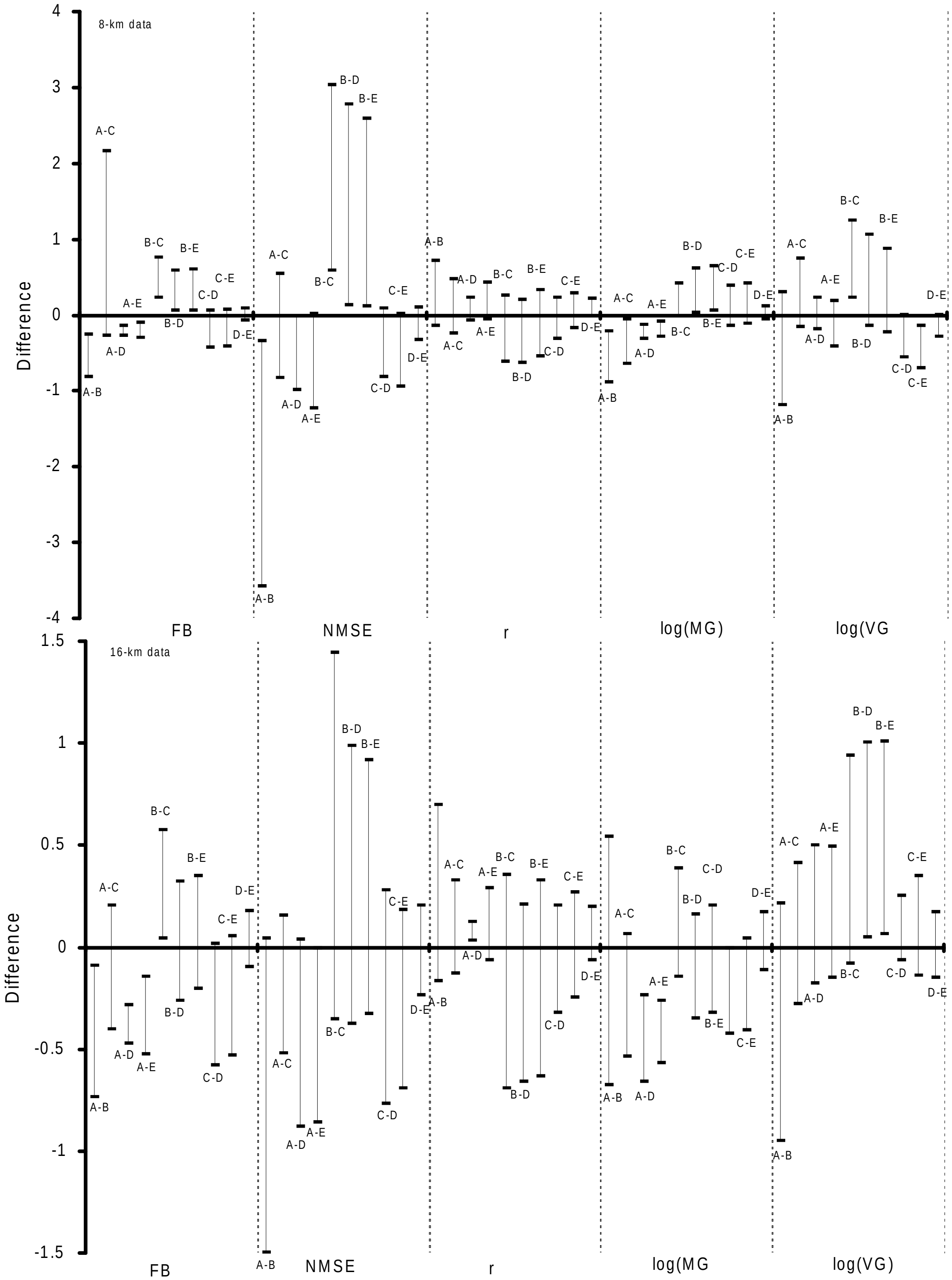

Figure 12. Ninety-five percent confidence intervals on differences in the fractional bias (*FB*), normalized mean square error (*NMSE*), correlation coefficient (*r*), geometric mean bias (*MG*), and geometric mean variance (*VG*) for the arc integrated concentration. Models are identified as follows: ISC = A, TRAC = B, RATCHET = C, TRIAD = D, and INPUFF2 = E.

DISCUSSION

Comparison of the time-averaged paired and unpaired results suggests that models experienced a greater difficulty defining plume trajectory than plume density. Terrain complexities and interactions between local, regional, and synoptic flows along the Front Range certainly contributed to this difficulty. Analysis of hourly concentrations by Lange (1992) indicated terrain interactions in the horizontal, strong vertical motions, and speed- and direction wind shears often created rapid shifts and multiple peaks in the observed plume pattern. Nighttime experiments were influenced by local and regional flows, and daytime and transition period experiments were influenced by local, regional, and synoptic flows. During nighttime hours, flow was dominated by nocturnal drainage flow from the valleys and canyons west of the release point. The plume was confined to a shallow layer no greater than 150 m thick, resulting in high surface concentrations measured at the samplers. These westerly flows converged with southeast flow in the Platte River Valley northeast of the Denver metropolitan area. During morning transition periods, Lange noted strong synoptic westerly winds with a weak easterly return flow near the surface, resulting in a complex dispersion pattern. Low surface concentrations and strong vertical mixing to 800 m were characteristic of these conditions. During daytime hours, similar conditions existed. Low surface concentrations were noted, especially at the 16-km distance. To some extent, these observations explain the patterns seen in the model results. That is, C_p/C_o ratios were generally lower for the nighttime tests and higher for the daytime and transition period tests.

Elderkin and Gudiksen (1993) studied several of the nighttime tests in which additional instrumentation was installed and monitored as part of the ASCOT program. They found dispersion was controlled by multiple scales of motion that created interacting layers which varied in three dimensions and on an hourly basis. Tracer plumes were mostly confined to a stable drainage layer that followed regional flow features, intermittently interrupted by evolving mountain-canyon flows. Interactions between the surface layer and the mountain-canyon flow layer caused unexpected tracer trajectories that were not anticipated based on conventional surface observations.

With the exception of TRAC, the models evaluated in this study are not capable of incorporating three dimensional varying wind field into model simulations. Furthermore, meteorological data required to characterize horizontal shifts due to local terrain and elevation differences within the sampling arcs were not available for every test conducted and are also lacking for the assessment period (1952–1989). At best, we are limited to surface observations at Rocky Flats and Denver Stapleton International Airport. Data from Denver Stapleton International Airport are the only complete meteorological record for the period. Reliable meteorological data from Rocky Flats during the assessment period exists only after 1984. Before that, data are sporadic, incomplete, and of questionable integrity. Limited meteorological data are available from Jefferson County Airport located about 8-km east of the RFP, but this station only operated from 06:00 to 23:00 mountain standard time.

Denver Stapleton International Airport is about 25-km southeast from the RFP, well outside the WVTS sampling domain, and is strongly influenced by flow up and down the Platte River Valley. Modeled plume trajectories within the WVTS sampling arcs were strongly influenced by the data from the 61-m tower at RFP. The ISC and INPUFF2 model simulations used RFP data exclusively. Comparisons by Haugen and Fotino (1993) of the hourly average plume center-of-mass with the mean hourly wind direction indicated TRAC and ISC predictions were strongly influenced by the RFP meteorological station, but the observed center-of-mass showed little correlation to the mean hourly average wind direction. TRAC simulations employed 16 additional stations within a 50-km radius from the release point, but these stations were believed to be too far away to have any significant impact on plume trajectories in the sampling domain.

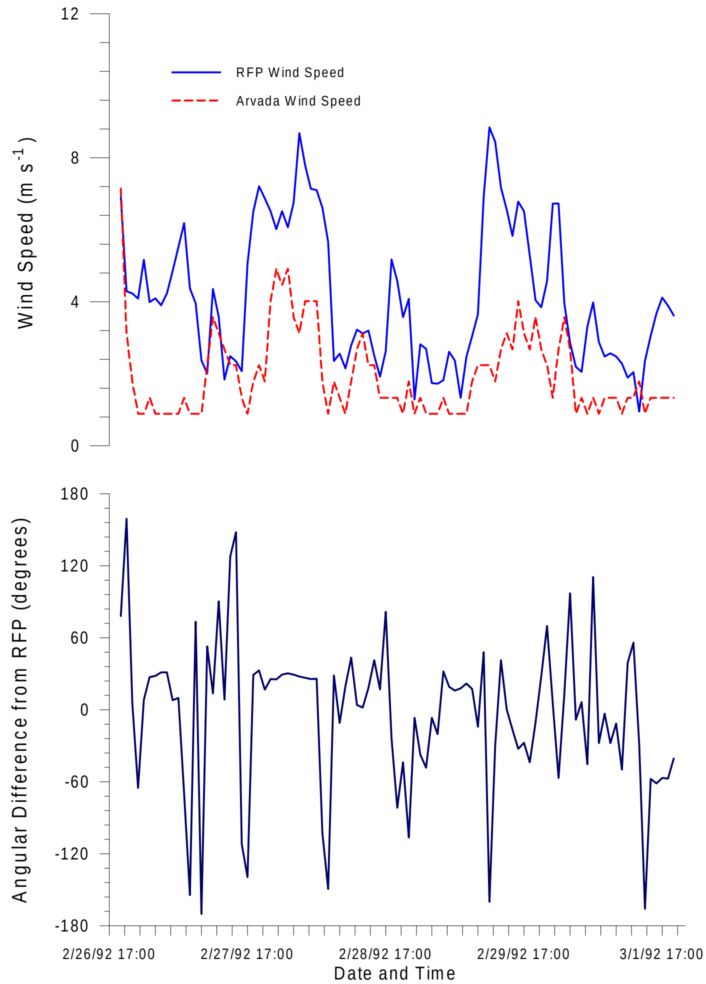

Figure 13. Comparison of wind speed and direction measured in February 1992 at Rocky Flats and Arvada. Negative angular differences indicate the Arvada wind direction was clockwise relative to the RFP. Positive angular differences indicate the Arvada wind direction is counterclockwise from the RFP.

In 1992, additional monitoring stations were installed in the vicinity of Rocky Flats. These stations are maintained by the CDPHE. Comparing hourly surface conditions (wind speed and direction) in the city of Arvada with those at the RFP for several days in February of 1992 (Figure 13) indicates that wind direction and wind speed are correlated to some degree, but conditions are definitely different at the two locations. Wind directions differed by as much as 170 degrees. Wind speeds were typically lower in Arvada by about a factor of 2, but they differed in some cases by a factor of 4. To further complicate the situation, the plume may be entrained in an elevated layer that has flow characteristics showing little resemblance to surface conditions. These observations reinforce Haugen and Fotino's conclusion that without more detailed surface and elevated observations in the model domain, we cannot expect complex models to perform any better than simpler models.

Comparison of the TRIAD and INPUFF2 simulations indicates the relative importance of the Denver Stapleton International Airport meteorological station in defining plume trajectory. Paired time-averaged performance measures for TRIAD indicated less variability and better correlation between predicted and observed pairs compared to those of INPUFF2. Paired time-averaged performance measures for the RATCHET model, which also included Denver Stapleton data, were generally closest to their optimum values compared to other models. From these results, there appears to be some benefit to using the Denver Stapleton data in model simulations, especially for estimating concentrations at distant receptors.

A cursory review of the concentration plots in Appendix B indicates RATCHET concentrations were smoother and showed less defined peaks than the other models, which may result from the use of hourly-average meteorological conditions instead of 15-min updates used in the other models. Also, RATCHET generally had the greatest number of valid predicted and observed data pairs compared to the other models. One might conclude that RATCHET spreads the plume over a larger area, thereby providing a greater number of points to compare. However, if this were the case, then we would expect underprediction in the unpaired analysis. This was not the case, and in fact, bias performance measures for this modeling objective suggested little bias in the RATCHET results at the 8-km arc and positive bias at the 16-km arc. In addition, arc-integrated results indicated that the predicted ground-level tracer mass in the sampling domain was consistent with the observations for RATCHET at both sampling arcs.

Conditions of historical release differed from the conditions in which the models were tested. For the 1957 fire, the majority of the release occurred as a buoyant plume containing plutonium particulates from the Building 771 stack (44 m high) at the plant. The same can be said about releases from the 1969 fire, but these releases occurred from vent sites at the top of buildings. Releases in 1969 from the 903 pad were at ground level and may have involved particles up to 100 μm in diameter. Releases from normal operations occurred from effluent stacks, ground-level evaporation ponds, and building vents. Other factors, such as plume rise and plume depletion, were not evaluated in this experiment. Certainly, the release conditions under which the WVTS took place differed from the conditions under which most release events occurred. Therefore, the results of this study are important in terms of atmospheric diffusion only. Impacts of release height and plume depletion on model performance will be considered in the specific reports for each release incident.

SUMMARY AND CONCLUSIONS

The suite of atmospheric transport models tested ranged from relatively simplistic (ISC) to moderately complex (TRAC). No one model outperformed the others in all modeling objectives and data categories. However, paired time-averaged performance measures suggest that the overall performance of RATCHET was somewhat better than the other models.

One might have expected TRAC to outperform the other models because it is a site-specific model that accounts for terrain complexities and three-dimensional wind fields. Time-averaged paired performance measures indicated that the overall performance of TRAC was no better than the other models, and in some cases, worse. Variability performance measures (*NMSE*, *VG*, and

GSD) indicated TRAC exhibited the highest degree of variability among the models compared. Performance in the time-averaged unpaired comparisons was somewhat better; but overall there was no substantial improvement in performance over the other models. We attribute this outcome not so much to the TRAC model itself, but to the inadequacies of meteorological data in the sampling arcs required to drive the model. Without substantial improvement in the quality and quantity of meteorological data, we cannot expect models such as TRAC to performed any better than simpler models.

The ISC model performed surprisingly well considering its simplicity relative to the other models. The model tended to overpredict concentrations, but in many cases, its variability performance measures (*GSD*, *NMSE*, and *VG*) were close to those of the other models. A drawback of this model it its inability to accept a temporally and spatially varying wind field in the model domain. This apparently made little difference given the lack of meteorological data in the sampling domain and distance to the samplers (≤16 km). However, the overall study encompasses a larger area, and influence of the Platte River Valley on plume trajectories cannot be accounted for in this model. The performance of the ISC model relative to the other models highlights the need for additional meteorological data when using puff trajectory models. The use of this model for Phase I of the dose reconstruction project probably resulted in bounding estimates of atmospheric diffusion.

The INPUFF2 and TRIAD simulations used the Pasquill-Gifford dispersion parameters as did ISC, but these models incorporated temporally and in the case of TRIAD, spatially varying wind fields. Time-average paired comparisons showed overall, slightly better performance by TRIAD and INPUFF2 compared to ISC. Log-transformed variability measures indicated slightly less variability and higher correlation with observed values for the INPUFF2 and TRIAD simulations compared to those of ISC. Such was not the case in the time-averaged unpaired comparisons suggesting a straight-line model performs as well as a puff trajectory model in terms of predicting plume density at distances ≤ 16 km.

There were only two samplers located outside the 16-km sampling arc (about 20 km from the release point), so it would be difficult to evaluate how these models performed at distances greater than 16 km. As shown previously, wind velocities vary spatially during the same time increment over the distances we intend to model, which will include an area extending 40 km southeast of the plant. We would expect a puff trajectory model to perform better at these distances relative to the straight-line model because puff models are capable of incorporating a spatially varying wind field within the model domain. Meteorological data from the Denver metropolitan area are available from Denver Stapleton International Airport for the entire assessment period. For these reasons, we favor using one of the puff-trajectory models over a straight-line plume model in this study.

The complexity of airflow patterns on the Colorado Front Range certainly contributed to the relatively large variability seen between model predictions and observations. Complex modeling does not appear to be feasible because detailed meteorological data in the model domain are lacking. Consequently, uncertainty in model predictions are expected to be large and probably dominate the overall uncertainty in risk estimates.

REFERENCES

Briggs, G.A. 1973. *Diffusion Estimation for Small Emissions*. Rep. ATDL Contribution File No. 79. Atmospheric Turbulence and Diffusion Laboratory, Oak Ridge, Tennessee.

Brown, K.J. 1991. *Rocky Flats 1990-91 Winter Validation Tracer Study*. Report AG91-19. North American Weather Consultants, Salt Lake City, Utah.

CDPHE (Colorado Department of Public Health and Environment). 1994. *Project Task 6 Report: Exposure Pathway Identification and Transport Modeling.* Health Studies on Rocky Flats Phase I: Historical Public Exposures.

Cox, W.M. and J.A. Tikvart. 1990. "A Statistical Procedure for Determining the Best Performing Air Quality Simulation Model." *Atmospheric Environment* 24A: 2387–2395.

Efron, B. 1982. *The Jackknife, the Bootstrap and Other Resampling Plans.* Society for Industrial and Applied Mathematics.

Efron, B. 1987. "Better Bootstrap Confidence Intervals." *J. Amer. Statist. Assoc.* 82: 171–185.

Elderkin, C.E. and P.H. Gudiksen. 1993. "Transport and Dispersion in Complex Terrain." *Radiation Protection Dosimetry* 50: 265–271.

EPA (U.S. Environmental Protection Agency). 1987. *On-Site Meteorological Program Guidance for Regulatory Modeling Applications*. EPA-450/4–87–013. Research Triangle Park, North Carolina.

EPA. 1988. *Procedures for Determining the Best Performing Model*. EPA Office of Air Quality Planning and Standards, Research Triangle Park, North Carolina.

EPA. 1992. *User's Guide for the Industrial Source Complex (ISC) Dispersion Models Vol. 1, User's Instructions*. EPA-450/4–92–008a. Research Triangle Park, North Carolina.

Fox, D.G. 1981. "Judging Air Quality Model Performance–a Review of the Woods Hole Workshop." *Bull. Am. Meteorol. Soc*. 62: 599–602.

Gifford, F.A. 1961. "Use of Routine Meteorological Observations for Estimating Atmospheric Dispersion." *Nuclear Safety* 2 (4): 47–51.

Hanna, S.R. 1989. "Confidence Limits for Air Quality Models Evaluations as Estimated by the Bootstrap and Jackknife Resampling Methods." *Atmospheric Environment* 23: 1385–1398.

Hanna, S.R., D.G. Strimaitis, and J.C. Chang. 1991. *Hazard Response Modeling Uncertainty (A Quantitative Method), Volume 1: User's Guide for Software for Evaluating Hazardous Gas Dispersion Models*. Air Force Engineering and Service Center, Tyndall Air Force Base, Florida.

Hanna, S.R., J.C. Chang, and D.G. Strimaitis. 1993. "Hazardous Gas Model Evaluation with Field Observations." *Atmospheric Environment* 27A: 2265–2285.

HAP (Health Advisory Panel). 1993. *Health Advisory Panel's Report to Colorado Citizens on the Phase I Study of the State of Colorado's Health Studies on Rocky Flats*. Colorado Department of Public Health and Environment. Denver, Colorado.

Haugen, D.A. and I.P. Fotino. 1993. *Performance Evaluation of the Terrain-Responsive Atmospheric Code (TRAC) Model.* Colorado School of Mines, Golden Colorado.

Hicks, B.B., K.S. Rao, R.J. Dobosy, R.P. Hosker, J.A. Herwehe, and W.R. Pendergrass. 1989. *TRIAD: A Puff-Trajectory Model for Reactive Gas Dispersion with Application to UF_6 Releases to the Atmosphere.* ERL ARL-168. National Oceanic and Atmospheric Administration, Air Resources Laboratory, Silver Springs, Maryland.

Hodgin, C.R. 1991. *Terrain-Responsive Atmospheric Code (TRAC) Transport and Diffusion: Features and Software Overview*. Report RFP-4516. EG&G Rocky Flats, Golden, Colorado.

Irwin, J.S. 1983. "Estimating Plume Dispersion–A Comparison of Several Sigma Schemes." *J Climate Applied Meteorology* 22: 92–114.

Lange, R. 1992. "Modeling the Dispersion of Tracer Plumes in the Colorado Front Range Boundary Layer During Night-and Day-Time Conditions." American Meteorological Society Tenth Symposium on Turbulence and Diffusion, Portland, Oregon, September 29 to October 2, 1992.

Pasquill, F. 1961. "The Estimation of the Dispersion of Windborne Material." *The Meteorological Magazine* 90: 33–49.

Petersen, W.B. and L. Lavdas, 1986. *INPUFF 2.0: A Multiple Source Gaussian Puff Dispersion Algorithm*. EPA-600/8–86/024. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Ramsdell, J.V., Jr. and K.W. Burk. 1991. *MESOIL2, a Lagrangian Trajectory Climatological Dispersion Model.* PNL–7430 HEDR, Battelle Pacific Northwest Laboratories, Hanford, Washington.

Ramsdell, J.V., Jr., C.A. Simonen, and K.W. Burk. 1994. *Regional Atmospheric Transport Code for Hanford Emission Tracking (RATCHET)*. PNWD–2224 HEDR UC–000, Battelle Pacific Northwest Laboratories, Hanford, Washington.

Rao, K.S. and R.P. Hosker. 1993. "Uncertainty in the Assessment of Atmospheric Concentrations of Toxic Contaminants from an Accidental Release." *Radiation Protection Dosimetry* 50: 281–288.

Rood, A.S. 1995. *Configuration of the RATCHET Code for Compilation on a Personal Computer and Use in the Rocky Flats Dose Reconstruction Project*. *Radiological Assessments Corporation*, Draft Technical Memorandum. March.

Tangirala, R.S., K.S. Rao, and R.P. Hosker. 1992. "A Puff Model Simulation of Tracer Concentrations in the Nocturnal Drainage Flow in a Deep Valley." *Atmospheric Environment* 26A: 299–309.

Turner, D.B. 1964. "A Diffusion model for an Urban Area." *Journal of Applied Meteorology* 3 (1): 83–91.

Turner, D.B. 1970. *Workbook of Atmospheric Dispersion Estimates*. Office of Air Programs Publication No. AP-26 (NTIS PB 191 482). U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Weber, J.M., A.S. Rood, J. Binder, H.R. Meyer, and J.E. Till. 1997. *Development of the Rocky Flats Plant 903 Area Plutonium Source Term*. 3 CDPHE-RFP-1997-DRAFT *Radiological Assessments Corporation*, Neeses, South Carolina. September.

Weil, J.C., R.I. Sykes, and A. Venkatram. 1992. "Evaluating Air Quality Models: Review and Outlook." *Journal of Applied Meteorology* 31: 1121–1145.

Zilitinkevich, S.S. 1972. "On the Determination of the Height of the Ekman Boundary Layer." *Boundary Layer Meteorology* 3 (2): 141–145.