Evaluation of an Air Pollution Analysis System for Complex Terrain

D. G. ROSS

Centre for Applied Mathematical Modelling, Monash University, Melbourne, Australia

D. G. FOX

Rocky Mountain Forest and Range Experiment Station, USDA, Fort Collins, Colorado

(Manuscript received 17 July 1990, in final form 4 December 1990)

ABSTRACT

This paper describes results from a study to evaluate components of an operational air quality modeling system for complex terrain. In particular, the Cinder Cone Butte (CCB) "modeler's dataset" is used to evaluate the current technique for incorporating terrain influences and atmospheric stability into the system's 3D diagnostic wind-field model.

The wind-field model is used in conjunction with a Gaussian puff model to compare predicted and observed tracer concentrations for different configurations, chosen to highlight the influence of the model's technique for incorporating terrain and atmospheric stability in the final flow field. A quantitative statistical basis, including the use of a bootstrap resampling procedure to estimate confidence limits for the performance measures, is used for the evaluation. The results show that the model's technique for incorporating terrain and atmospheric stability yields a significant improvement in predictive performance. Even when only routinely available input data are used, the performance is shown to be as good as that of models based directly on the CCB dataset itself.

1. Introduction

Models to estimate air quality in complex terrain face a number of challenges in their development and application. There are many unanswered scientific questions about both the mean and turbulent characteristics of airflow in complex terrain. As a result, a number of major research efforts have been undertaken, which include the Complex Terrain Model Development (CTMD) program, with results reported as a series of "milestone" reports (e.g., DiCristofaro et al. 1985); and the Atmospheric Studies in Complex Terrain (ASCOT) program, with results published as a set of papers in the June and July 1989 volumes of the *Journal of Applied Meteorology*. The CTMD program focused on plume impaction—the dynamics of plume behavior in the vicinity of an isolated topographic feature, while the ASCOT program focused on an understanding of plume behavior in nocturnal drainage flows. Both programs provide not only detailed data required to understand the physics of each flow type, but also databases for the development and evaluation of models.

The topographic air pollution analysis (TAPAS) is a flexible computer modeling system being developed jointly by the Centre for Applied Mathematical Modelling (CAMM) at Monash University and the Rocky Mountain Forest and Range Experiment Station, U.S. Department of Agriculture—Forest Service. TAPAS contains simulation models of varying complexity and ranges of applicability, input data management routines, and graphical display procedures designed to assist air resource managers. The details of the overall system are described elsewhere (e.g., Fox et al. 1987). Predicting the concentration of atmospheric pollutants in complex terrain typically requires an air quality modeling system that is capable of simulating both spatial and temporal variations in meteorology and turbulent diffusion. The TAPAS system configured to use the NUATMOS and CITPUFF models represents a cost-effective system capable of simulating these variations in complex terrain, while maintaining much of the simplicity of the basic Gaussian approach.

CITPUFF (Ross et al. 1987) is a multiple-source puff model capable of taking as its transport field a fully three-dimensional (3D) wind field. The current version is designed to accept wind fields from the complex terrain diagnostic wind field model NUATMOS (Ross et al. 1988a,b; Smith and Ross 1988); however, modification to accept the output of an alternative wind field model is straightforward. The models have undergone an extensive testing and evaluation program, both as separate models and in combination (e.g., Lorimer 1989). The program has included comparison of model predictions with exact solutions, data from lab-

Corresponding author address: Dr. D. Graeme Ross, Centre for Applied Mathematical Modelling, Monash University, P.O. Box 197, Caulfield East, Melbourne, Victoria 3145, Australia.

oratory experiments, and observations from field studies.

The testing and evaluation program is ongoing and is intimately coupled with further development of the models and their components. This paper describes results of a study aimed at evaluating the NUATMOS/CITPUFF configuration of TAPAS using experimental results from the CTMD program. In particular, results for Cinder Cone Butte (CCB) (a roughly axisymmetric isolated hill in Idaho) are used to evaluate the performance of the current technique for incorporating terrain and atmospheric stability effects into NUATMOS. A quantitative statistical comparison between predictions and measured tracer concentrations is used, with the results presented in this paper focusing on the case study hours in the modeler's dataset examined by Strimaitis et al. (1983).

2. The NUATMOS and CITPUFF models

a. *NUATMOS*

The model generates a 3D mass-consistent wind field based on arbitrarily located observations. This is achieved by interpolating throughout the domain of interest and then making minimal adjustments in order to eliminate divergence (for mass conservation).

The divergence elimination phase minimizes the functional

$$E(u, v, w) = \iiint \Big[(u - u_0)^2 + (v - v_0)^2 + \alpha^{-2}(w - w_0)^2\Big] dV$$

subject to the constraint

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0,$$

where x, y are the horizontal coordinates; z is the vertical coordinate; $\mathbf{V}_0 = (u_0, v_0, w_0)$ is the corresponding initial (interpolated) velocity; $\mathbf{V} = (u, v, w)$ is the final velocity; and the parameter α allows horizontal and vertical winds to be adjusted differentially.

Mathematically, the functional is minimized using variational calculus methods with the resulting Poisson equation for the Lagrange multiplier solved in terrain-following coordinates using multiple grids and a tridiagonal solver to speed up convergence. NUATMOS (version 5) has been optimized to the degree that it can be run comfortably on current-generation personal computers.

The NUATMOS model has been extensively tested using comparisons with exact solutions, data from laboratory experiments, and data from field flow measurements (e.g., Ross et al. 1988a,b; Connell 1988). The tests have revealed that for an initially uniform wind the divergence reduction stage with $\alpha = 1$ will yield the correct potential flow solution for an isolated terrain shape. The adjustment to the initial wind field can be interpreted as the imposition of a "terrain effect" upon a uniform background wind. This leads to speedup, retardation, and channeling as the wind encounters changes in the shape of the underlying terrain surface. The potential flow solution can also be regarded as an approximation to flow past obstacles in neutral atmospheric stability conditions.

Ross et al. (1988a) use a hybrid approach to develop and test a framework for a Froude number dependent expression for α based on conservation of energy arguments and laboratory data. Subsequent extension and evaluation using additional laboratory data (Ross et al. 1988b) suggest that the relationship

$$\alpha^{-2} = 1 + \frac{3}{(S^2 - 1)\mathrm{Fr}^2}, \qquad (1)$$

where S is the "speedup" over the terrain and Fr is a characteristic Froude number of the flow, provides a robust relationship that incorporates the influence of terrain shape indirectly through the calculated speedup. Testing of this relationship using "direct" flow field data from the CTMD program was inconclusive, largely because of the considerable uncertainties in the limited data available.

The testing of the α–Fr relationship has been restricted to isolated topography embedded in a uniform approach flow of speed U_∞ at an upstream reference height $z = H_s$, and a linearly thermally stratified environment characterized by a buoyancy frequency $N^2 = (g/\theta_R)(d\theta_e/dz)$, where θ_e is the environmental potential temperature with θ_R a representative value of the air between heights $z = H_s$ and the topography height $z = H$. The hill Froude number $\mathrm{Fr} = U_\infty/NH$, and the speedup $S = U_{\max}/U_\infty$ is for the potential flow solution (obtained from NUATMOS with $\alpha = 1$). In the present analysis we follow Spangler (1983, 1986) and use a "release height" Froude number where U_∞ is taken as the release height wind speed and H is the characteristic height of the CCB terrain (taken to be 95 m).

b. *CITPUFF*

CITPUFF (Ross et al. 1987) is a multisource Gaussian puff model that uses the 3D wind fields generated by NUATMOS, interpolated in both time and location. Plume rise, stability class, and other hourly model inputs are also interpolated in time. Other features include a range of optional dispersion formulas, partial penetration of an elevated stable layer and introduction of an elevated layer of arbitrary stability, and dynamic determination of the puff advection step length.

TAPAS, configured to use NUATMOS and CITPUFF, has been extensively evaluated using data from the Latrobe Valley Airshed Study conducted in southeastern Australia (Lorimer 1989; Ross et al. 1989).

3. The Cinder Cone Butte database

Cinder Cone Butte is a two-peaked, roughly axisymmetric hill about 100 m high. It's nearly circular base is about 1 km in diameter. Figure 1 illustrates the terrain (with a vertical exaggeration of a factor of approximately five) with a view looking south, with a section from the topographic file used in the modeling. The experimental program included SF_6 tracer releases upwind of CCB with tracer sampling at receptors located on the hill. A meteorological monitoring network was used to obtain direct and derived meteorological parameters corresponding to each tracer release experiment. In addition, lidar and photographic data provided information on plume trajectories and plume spread.

Strimaitis et al. (1983) present the results of case study analyses from a total of 45 hourly experiments conducted. For each case study period the meteorological and tracer gas databases were analyzed in detail to provide a comprehensive modeler's dataset for evaluating the performance of air quality models for stable flows in complex terrain. Each case study dataset includes hourly average values of the observed SF_6 concentration at each receptor, source release information, and meteorological characteristics. In the case of the meteorological information, data from a number of locations were used to derive:

- an hourly averaged wind vector (at the source release height h);
- atmospheric stability parameters, including the Brunt–Väisälä frequency N, the critical streamline height H_c, a characteristic Froude number, and the Pasquill stability category;
- turbulence data, including the crosswind and vertical turbulent intensities.

The case study hours are subdivided according to whether the release height was above or below the critical streamline height. The critical or dividing streamline height (e.g., Sheppard 1956) represents a division between two regions of flow that occur in stable stratified flow around a terrain obstacle. Below the critical streamline, the flow has insufficient kinetic energy to surmount the terrain feature and consequently passes around it in an approximately horizontal plane. Above the critical streamline, the flow passes up and over the terrain feature. It is important to note that the critical streamline concept is only a simplistic approximation. In fact, its role in airflow blocking and splitting has been questioned by Smith (1988).

4. TAPAS evaluation

a. TAPAS configurations

Ground-level predictions at the tracer receptor locations have been made for each of the CCB case study hours using the following configurations of the NUATMOS wind field model:

(i) a uniform wind field throughout the computational domain. The value of the wind vector is taken as the source height wind vector (see Table 1). Operationally this is achieved by completing only the initial interpolation phase of NUATMOS.

(ii) Adjusting the wind field produced by configuration (i) to be divergence-free by applying the divergence reduction phase of NUATMOS with the parameter $\alpha = 1$. Ross et al. (1988a) show that this approach will generate the potential flow solution for any terrain shape.

(iii) As for configuration (ii), but with the value of α specified using Eq. (1).

(iv) As for configuration (ii), but with the value of α determined empirically by matching the critical streamline height predicted by NUATMOS with that reported in the modeler's dataset for each particular case study hour.

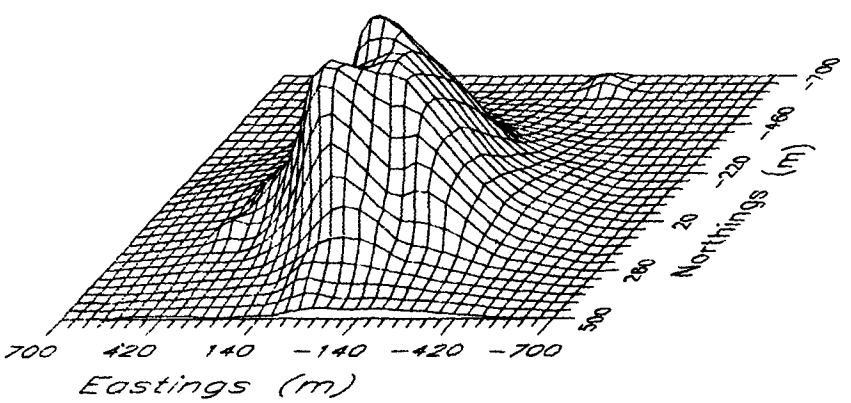

FIG. 1. Ground relief of CCB (exaggerated by a factor of ~5) within the computational grid used by NUATMOS and CITPUFF with a view looking south. The complete terrain file used consisted of 61 × 61 points or 60 × 60 cells with a horizontal resolution of 40 m.

The CITPUFF model was configured to move puffs along the terrain-following surface passing through their point of release for configuration (i). For the other configurations the CITPUFF model used the fully 3D wind field (including the vertical velocities) generated by NUATMOS to transport puffs.

The TAPAS configurations have been chosen to highlight the influence of the technique for incorporating terrain and atmospheric stability effects in the NUATMOS wind-field model. We are interested in, first, the addition of a 3D wind field and, second, the quality (mathematical and physical correctness) of the field on the concentration predictions in complex terrain.

Predictions from the CCB models (Strimaitis et al. 1983) for $h > H_c$: NEUTRAL and LIFT, and for $h \leq H_c$: IMPINGEMENT and WRAP, are also included in the analysis for comparison with the TAPAS configurations. It is relevant to note that the LIFT and WRAP models were developed as improvements to the NEUTRAL and IMPINGEMENT models, respectively. The CCB models are an earlier form of the CTDM (Strimaitis et al. 1987) and CTDMPLUS (Perry et al. 1989) modeling systems.

In what follows, the models or system configurations will be referred to by the following code:

CCB–N/I:	NEUTRAL ($h > H_c$) and/or IMPINGEMENT ($h \leq H_c$)
CCB–L/W:	LIFT ($h > H_c$) and/or WRAP ($h \leq H_c$)
TAP–UNI:	TAPAS [NUATMOS configuration (i), uniform wind field]
TAP–POT:	TAPAS [NUATMOS configuration (ii), potential flow field]
TAP–αF:	TAPAS [NUATMOS configuration (iii), α determined using equation (1) formulation]
TAP–αE:	TAPAS [NUATMOS configuration (iv), α determined empirically].

The basic output from a deterministic air quality model is an estimate of the ensemble average concentration at a receptor point. The uncertainty associated with this prediction is made up of a reducible component due to inadequacies of the model formulation and input data, and an irreducible component due to the random nature of turbulence in the atmosphere (Fox 1984). The basic input requirements for the CITPUFF model are the emission rate, release height, dispersion relations, gridded terrain heights, and the flow field provided by NUATMOS. The CCB database enables us to minimize the uncertainties associated with the first four of these inputs and to focus attention on the influence of the flow field, and in particular the effects of terrain and atmospheric stability on the flow. The uncertainties associated with emission rate, release height, and terrain heights are clearly small. We adopt the dispersion relations used in the CCB models

$$\sigma_y = i_y x,$$

$$\sigma_z = \frac{i_z x}{(1 + Nx/U_\infty p)^{1/2}},$$

where σ_y and σ_z are the crosswind and vertical standard deviations of the assumed Gaussian distribution, respectively; i_y and i_z are the crosswind and vertical turbulence intensity (at source height), respectively; U_∞ is the wind speed at source height; x is travel distance from the source; and p is a parameter determined by calibrating against the lidar measurement ($p = 1.5$). Although the dispersion relations for CCB are based on the experimental data, they are still subject to considerable uncertainty (Strimaitis et al. 1983).

We also examine the effect of using the Pasquill–Gifford dispersion scheme, but with the standard scheme adjusted for roughness height and averaging time in the manner recommended by Hanna et al. (1977).

Table 1 summarizes the basic input data used for the model simulations for each of the case study hours examined in this analysis.

b. *Evaluation procedure*

We adopt the approach of Hanna (1989) and use two basic statistical performance measures to assess the performance of the alternative models or system configurations. The first measure is the normalized or fractional bias of the mean concentration:

$$\text{FB} = 2(\bar{c}_0 - \bar{c}_p)/(\bar{c}_0 + \bar{c}_p),$$

where $\bar{c}_0$ and $\bar{c}_p$ are the mean observed and predicted concentrations, respectively. The fractional bias FB has an ideal value of zero and can vary between -2 and $+2$. The second performance measure is the normalized mean-square error NMSE:

$$\text{NMSE} = \overline{(c_p - c_0)^2}/\bar{c}_0\bar{c}_p,$$

which provides a global error estimate of the scatter between observations and predictions.

Means and confidence intervals of the two performance measures are calculated using the resampling and model evaluation software described by Hanna (1989). The so-called bootstrapping resampling procedure has been used to estimate the confidence limits or statistical uncertainty of the performance measure results. The estimates of 95% confidence limits have been obtained using both the "seductive" and "moment" bootstrap procedures, although only the results of the latter are presented here. Hanna suggests that the latter procedure is preferable as it overcomes problems that can arise with the former at the tails of the distribution. For the present case the differences in estimates obtained from the alternative procedures are negligible and do not alter our conclusions.

TABLE 1. Observed and derived input data for the model simulations.

	Meteorological data[a]							Source data[a]				Derived model inputs		
Experiment	U_∞ (h) (m s^{-1})	θ (deg)	H_c	i_z	i_y	N (s^{-1})	Stability class	h (m)	Q (g s^{-1})	Distance (m)	Direction (deg)	Fr[b]	α[c]	α[d]
202H4	7.3	327	0	0.041	0.054	0.037	E	20	0.082	1014.6	319	2.08	0.60	0.26
202H5	7.8	326	7	0.037	0.060	0.047	E	30	0.086	1014.6	319	1.75	0.53	0.25
204H1	2.1	359	58	0.049	0.341	0.051	F	30	0.094	1035.7	5	0.43	0.15	0.10
205H4	5.9	121	10	0.031	0.069	0.042	F	40	0.083	1155.1	120	1.48	0.47	0.24
205H5	6.5	120	24	0.049	0.104	0.048	F	50	0.087	1155.1	120	1.43	0.46	0.19
206H4	4.5	128	20	0.031	0.078	0.044	F	35	0.039	595.9	124	1.08	0.36	0.21
206H6	2.1	127	35	0.086	0.212	0.035	F	35	0.062	595.9	124	0.63	0.22	0.17
206H8	2.0	127	37	0.097	0.180	0.040	F	35	0.062	595.9	124	0.53	0.19	0.16
209H7	2.5	105	58	0.047	0.121	0.094	F	40	0.160	999.2	101	0.28	0.10	0.10
210H3	6.3	118	26	0.018	0.105	0.044	F	57	0.169	1084.3	114	1.51	0.48	0.19
210H7	7.3	118	14	0.057	0.108	0.035	F	58	0.178	1086.2	122	2.20	0.62	0.22
211H1	1.9	107	45	0.084	0.326	0.058	F	30	0.179	1001.2	101	0.34	0.12	0.14
211H5	2.4	120	26	0.068	0.277	0.092	F	20	0.175	1155.1	120	0.27	0.10	0.19

[a] Meteorological and source data from Strimatis et al. (1983). Hourly average meteorological data—U_∞ (h): wind speed at release height (m s^{-1}); θ: wind direction (deg) ($\theta = 0$, northerly wind); H_c: dividing streamline height (m); i_y, i_z: crosswind and vertical turbulence intensities; N: Brunt–Väisälä frequency (1 s^{-1}); Pasquill stability class (Lavery et al. 1982; Table 42). Source Data—h: release height (relative to terrain height datum) (m); Q: hourly average emission rate of SF_6 (g s^{-1}); Source position relative to "center" of CCB grid (see Fig. 1).

[b] Release height Froude number Fr = U_∞/NH, where H = 95 m is taken as the characteristic terrain height of CCB.

[c] α values determined from Eq. (1) using the release height Froude number and a "neutral" flow ($\alpha = 1$) speedup of $S = 1.18$.

[d] α values determined by matching the experimental and simulated value of H_c.

Scatterplots of observed versus predicted concentrations are also used as a simple depiction of the model's results.

The evaluation procedure is applied to an ensemble of 13 case study hours with the data combined and blocked; that is, the observation/prediction pairs at each receptor for all hours are combined into a single dataset (455 observation/prediction pairs), which is divided into blocks containing data with similar characteristics. The justification for combining and examining all the case study hours as a single dataset is based on the fact that all 13 hours represent stable atmospheric conditions (Pasquill stability categories E or F). The obvious blocking is then to divide the data based on whether $h > H_c$ or $h \leq H_c$.

With the blocked bootstrap resampling procedure, each resample of 455 observation/prediction pairs is forced to contain 233 pairs (equivalent to the 7 case study hours where $h > H_c$) from the first block and 222 pairs (equivalent to the 6 case study hours where $h \leq H_c$) from the second block. Consequently, it is impossible to pick all 455 pairs from the same block in a given sample.

The evaluation procedure is also applied separately to the two component ensembles where $h > H_c$ and $h \leq H_c$.

5. Results

A simple but instructive way of illustrating the results is through a scatterplot of predictions versus observations, paired in location and time. Figures 2a and b show such plots for each model configuration, but with the results broken into ensembles of hours for which $h > H_c$ and $h \leq H_c$, respectively. For the ensemble with $h > H_c$, the improvement in predictive performance resulting from including each of the wind field adjustments associated with moving from TAP-UNI through to TAP-αE is clear, with the TAP-αE configuration being the best performer. The ensemble with $h \leq H_c$ also indicates some progressive improvement; however, the predictive performance of all model configurations is relatively poor. In particular, each configuration badly underpredicts the highest observed concentrations, as well as tending to overpredict the low end of the range. In fact, there is evidence that the inclusion of each wind-field adjustment associated with moving through TAP-UNI to TAP-αE leads to a deterioration in predictive performance for the lower observed concentrations.

It is useful to be able to quantify these observations, and to have estimates of the statistical confidence of the findings.

Table 2 contains the calculated values of FB and NMSE for the ensemble of all hours and the two component ensembles, for each model or system configuration.

Figures 3 and 4 illustrate FB and NMSE and their 95% confidence intervals for the total and each of the component ensembles. If the confidence limits overlap the zero of a statistical measure, then the calculated value of the measure is not significantly different from

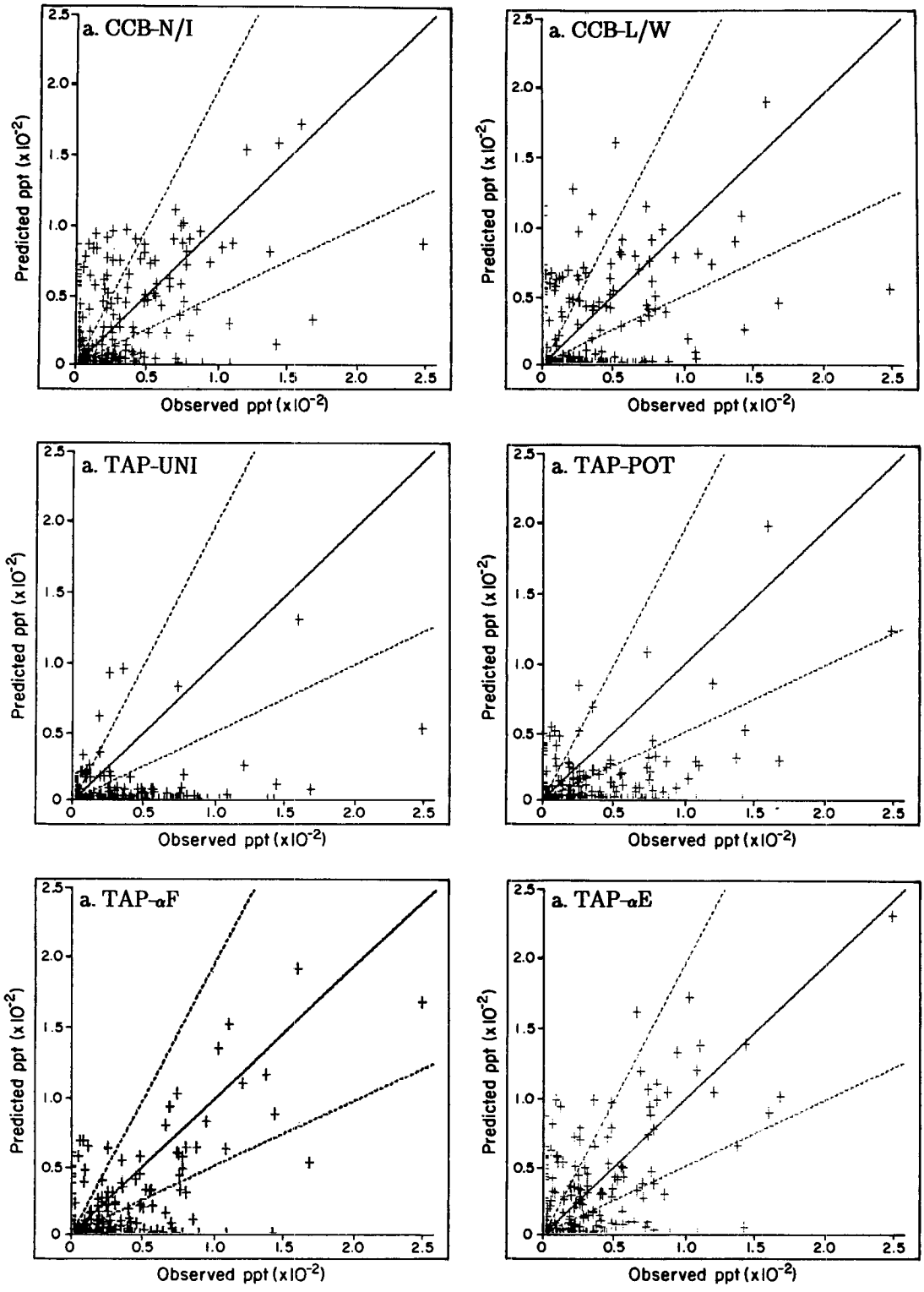

FIG. 2. Scatterplots of predicted versus observed concentrations (ppt), paired in location and time, for each model or system configuration (a) $h > H_c$ and (b) $h \leq H_c$. The area between the dashed lines contains points that are within a factor of 2 of the solid (perfect fit) line.

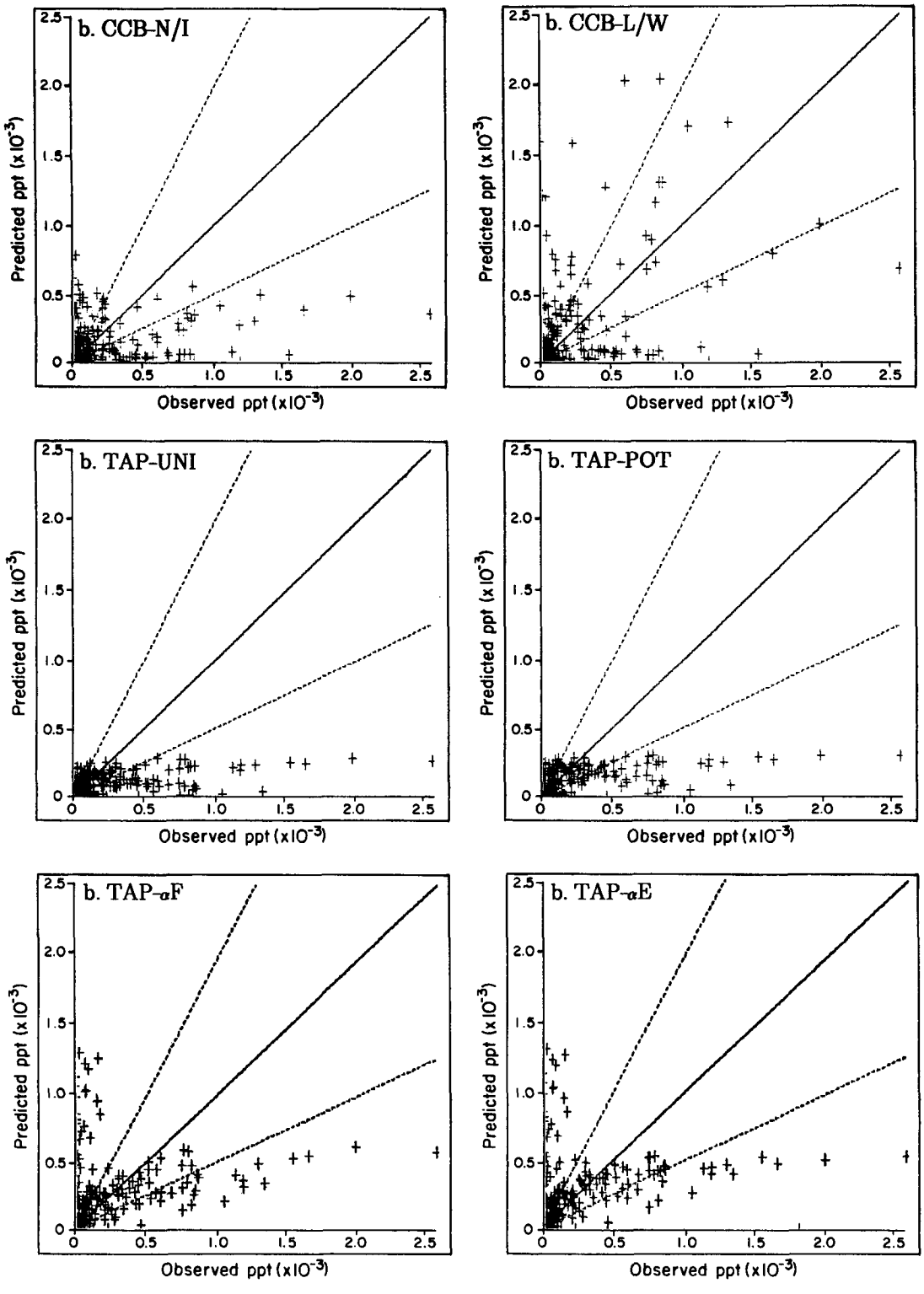

FIG. 2. (*Continued*)

TABLE 2. Statistical performance measures.

Fractional bias (FB)						
	Model					
Ensemble	CCB–N/I	CCB–L/W	TAP–UNI	TAP–POT	TAP–αF	TAP–αE
Total	0.26	−0.003	0.97	0.72	−0.04	−0.16
$h > H_c$	−0.06	0.17	1.27	0.80	0.37	−0.004
$h \leqslant H_c$	0.37	−0.05	0.89	0.69	−0.14	−0.2
Normalized mean-square error (NMSE)						
	Model					
Ensemble	CCB–N/I	CCB–L/W	TAP–UNI	TAP–POT	TAP–αF	TAP–αE
Total	4.6	3.9	9.3	6.3	3.7	3.3
$h > H_c$	1.6	2.4	10.8	4.3	1.9	1.4
$h \leqslant H_c$	4.0	2.9	6.3	4.6	2.6	2.4

0.0 at the 95% confidence level. Figure 4 shows that the NMSE for all models is significantly different from zero at the 95% confidence level. For the total ensemble we see from Fig. 3 that the FBs for CCB–N/I, TAP–UNI, and TAP–POT are also significantly different from zero, with each model underpredicting. The FBs for CCB–L/W, TAP–αF, and TAP–αE are not significantly different from zero at the 95% confidence level. The same finding also holds for each component ensemble for TAP–UNI, TAP–POT, and TAP–αE and for the component ensemble with $h \leqslant H_c$ for TAP–αF, while TAP–αF underpredicts for the $h > H_c$ ensemble.

When examining the CCB models, it is important to remember that each total ensemble comprises predictions from distinct models whose design and range of applicability are restricted to either $h > H_c$ or $h \leqslant H_c$. The total ensemble has some meaning and usefulness when comparing with the TAPAS configurations, but should not be examined in comparison to the results

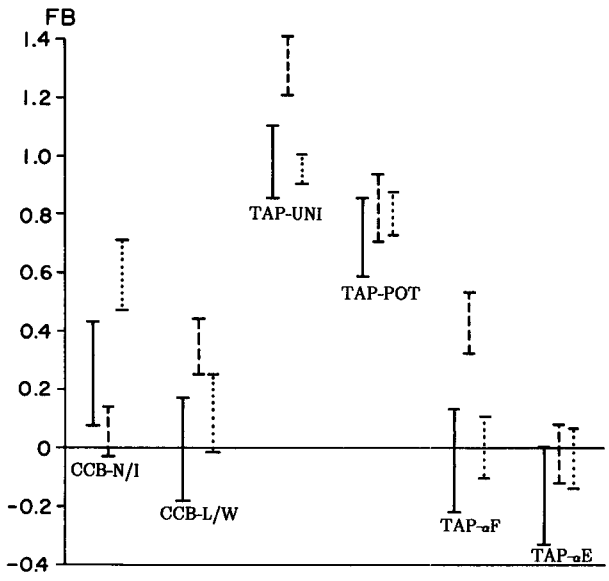

FIG. 3. Ninety-five percent confidence intervals on FB for each model or system configuration as estimated by Hanna's (1989) moment bootstrapping resampling procedure:

- total ensemble (solid line)
- $h > H_c$ ensemble (dashed line)
- $h \leqslant H_c$ ensemble (dotted line).

If the confidence interval does not overlap 0.0, then we have 95% confidence that the FB is different from 0.0.

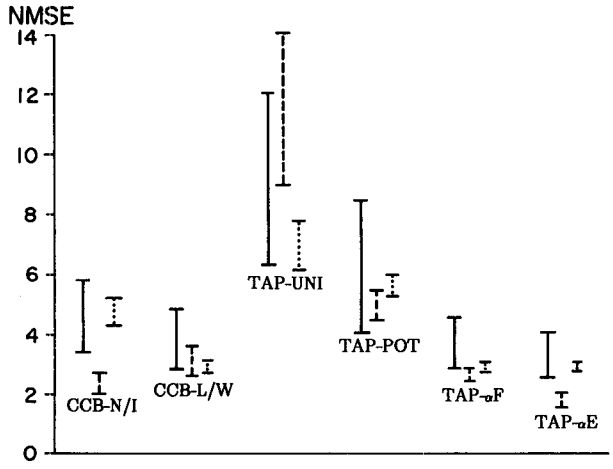

FIG. 4. Ninety-five percent confidence intervals on NMSE for each model or system configuration as estimated by Hanna's (1989) moment bootstrapping resampling procedure:

- total ensemble (solid line)
- $h > H_c$ ensemble (dashed line)
- $h \leqslant H_c$ ensemble (dotted line).

If the confidence interval does not overlap 0.0, then we have 95% confidence that the NMSE is different from 0.0.

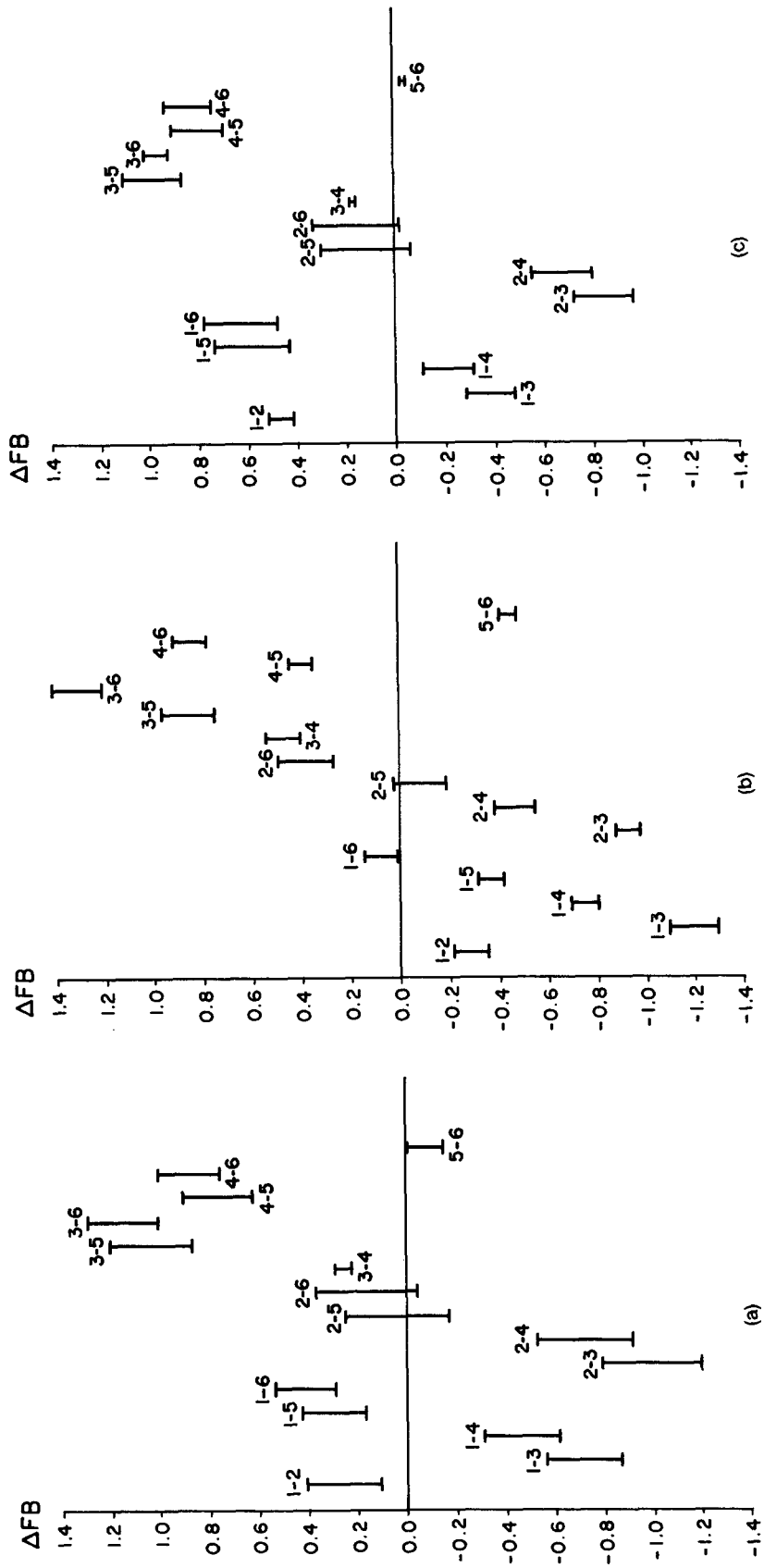

FIG. 5. Ninety-five percent confidence intervals on differences in FB (ΔFB) between models or system configurations (1 = CCB–N/I, 2 = CCB–L/W, 3 = TAP–UNI, 4 = TAP–POT, 5 = TAP–αF, and 6 = TAP–αE) (a) total ensemble, (b) $h > H_c$ ensemble, and (c) $h \leq H_c$ ensemble. If the confidence interval does not overlap 0.0, then we have 95% confidence that the FBs for the two models are different.

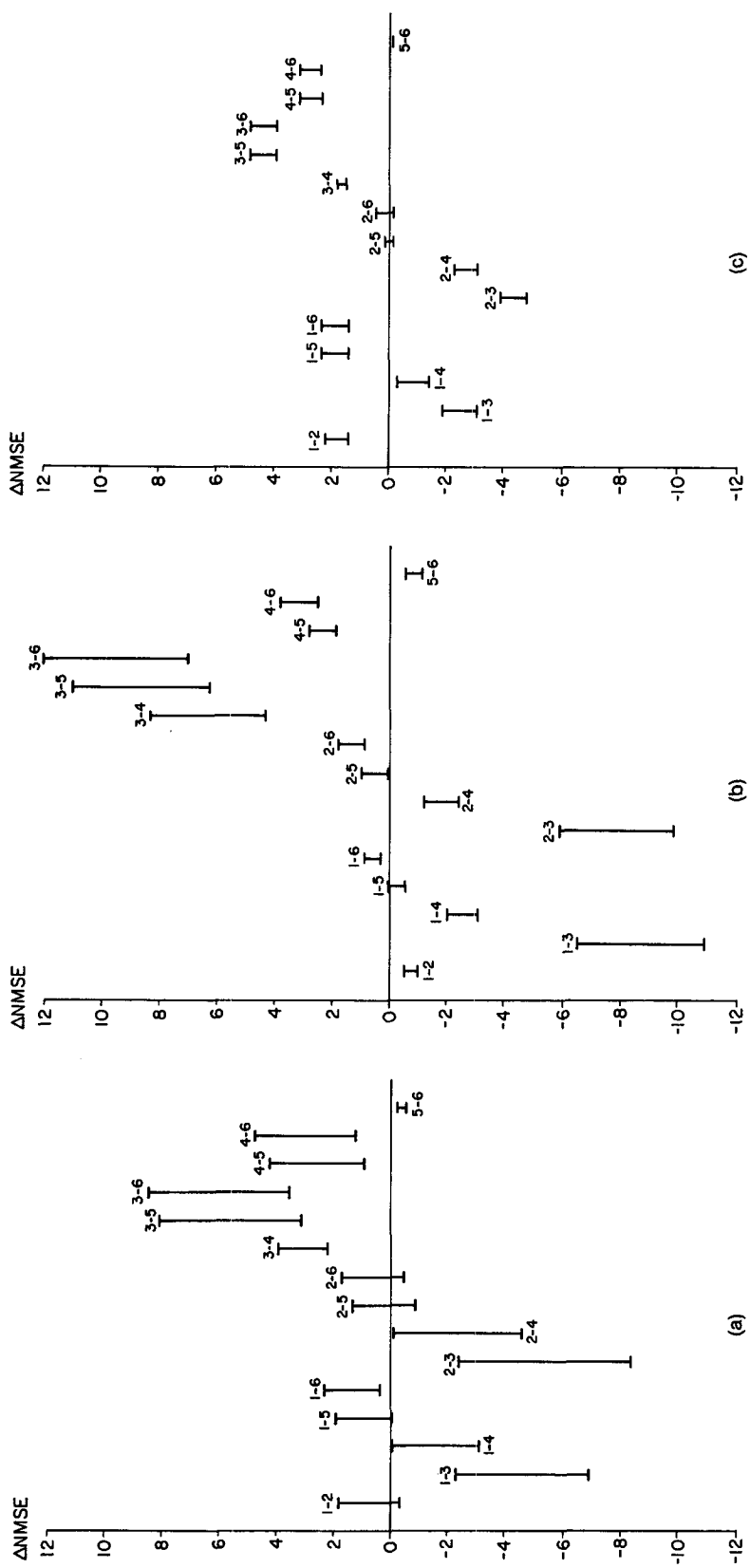

FIG. 6. Ninety-five percent confidence intervals on differences in NMSE (ΔNMSE) between models or system configurations (1 = CCB–N/I, 2 = CCB–L/W, 3 = TAP–UNI, 4 = TAP–POT, 5 = TAP–αF, and 6 = TAP–αE) (a) total ensemble, (b) $h > H_c$ ensemble, and (c) $h \leq H_c$ ensemble. If the confidence interval does not overlap 0.0, then we have 95% confidence that the NMSEs for the two models are different.

for its component ensembles. For the ensemble with $h > H_c$, the NEUTRAL model (CCB–N) yields a zero FB while its supposed improvement, the LIFT model (CCB–L), underpredicts. For the case $h \leq H_c$, however, the newer WRAP model (CCB–W) has a zero FB, while the older IMPINGEMENT model (CCB–I) underpredicts.

We now examine the influence of the relative performance of the model or system configurations by using the resampling procedure to estimate confidence limits on differences in FB (ΔFB) and NMSE (ΔNMSE) between models. The results of all possible pairings of the models are illustrated in Figs. 5 and 6 for each of the three ensembles.

We consider first the influence of the flow field adjustments contained in the alternative NUATMOS configurations. The confidence intervals on both ΔFB and ΔNMSE confirm our conclusions from examining the scatterplots and show that there is a significant improvement in predictive performance as we move from TAP–UNI through to TAP–αE. This conclusion is valid also for the cases with $h > H_c$ and $h \leq H_c$. However, it should be noted that the differences between TAP–αF and TAP–αE are at the margin of significance for both the total ensemble and the component ensemble with $h \leq H_c$. The improvement in predictive performance in moving from TAP–αF to TAP–αE is clear for the cases where $h > H_c$.

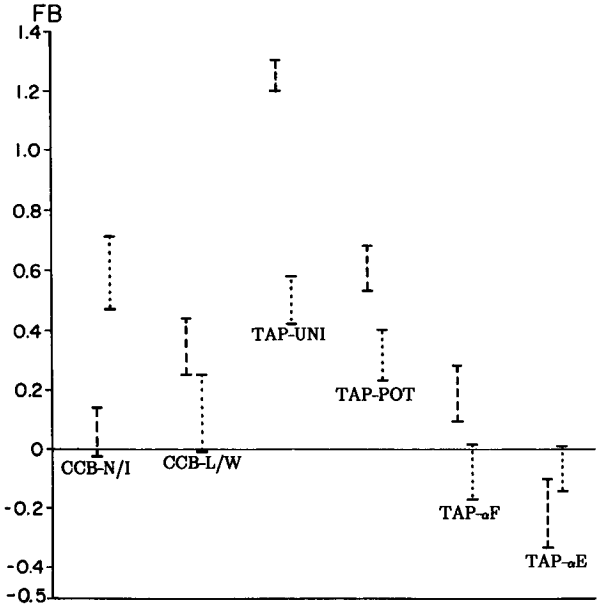

FIG. 7. Ninety-five percent confidence intervals on FB for each model or system configuration when CCB dispersion scheme is replaced by Pasquill–Gifford scheme in the TAPAS configurations:

- $h > H_c$ ensemble (dashed line)
- $h \leq H_c$ ensemble (dotted line).

If the confidence interval does not overlap 0.0, then we have 95% confidence that the FB is different from 0.0.

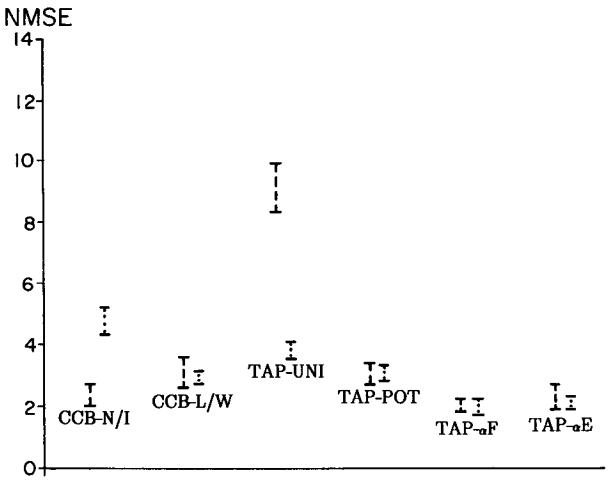

FIG. 8. Ninety-five percent confidence intervals on NMSE for each model or system configuration when CCB dispersion scheme is replaced by Pasquill–Gifford scheme in the TAPAS configurations:

- $h > H_c$ ensemble (dashed line)
- $h \leq H_c$ ensemble (dotted line).

If the confidence interval does not overlap 0.0, then we have 95% confidence that the NMSE is different from 0.0.

For the ensemble $h > H_c$, the NEUTRAL (CCB–N) and LIFT (CCB–L) models are significantly different with regard to both NMSE and FB, with the older NEUTRAL model performing significantly better than its "improvement." The NMSE and FB for the IMPINGEMENT (CCB–I) and WRAP (CCB–W) models are also significantly different for the case $h \leq H_c$; however, now the improved model WRAP is a significantly better performer than its predecessor IMPINGEMENT with respect to both statistical measures.

Let us now compare the best of the TAPAS configurations with the best of the CCB configurations. For the total ensemble, the performances of CCB–L/W and TAP–αE are not significantly different with respect to either NMSE or FB. This is also true for the case where $h \leq H_c$. However, for the ensemble with $h > H_c$, the TAPAS configuration TAP–αE has a significantly smaller NMSE than the NEUTRAL model (CCB–N), with both models having no significant difference in their FB. Indeed, we saw earlier that the FB for both models is not significantly different from zero (Fig. 3).

Figures 7–10 illustrate the results of repeating the analysis when the CCB dispersion scheme is replaced by the Pasquill–Gifford scheme in the TAPAS configurations. Only the results for the ensembles with $h > H_c$ and $h \leq H_c$ are presented.

The results confirm our conclusion that the inclusion of the flow field adjustments in the alternative NUATMOS configurations do yield a significant improvement in predictive performance, although the choice of the best TAPAS configuration is less clear than for the CCB dispersion scheme.

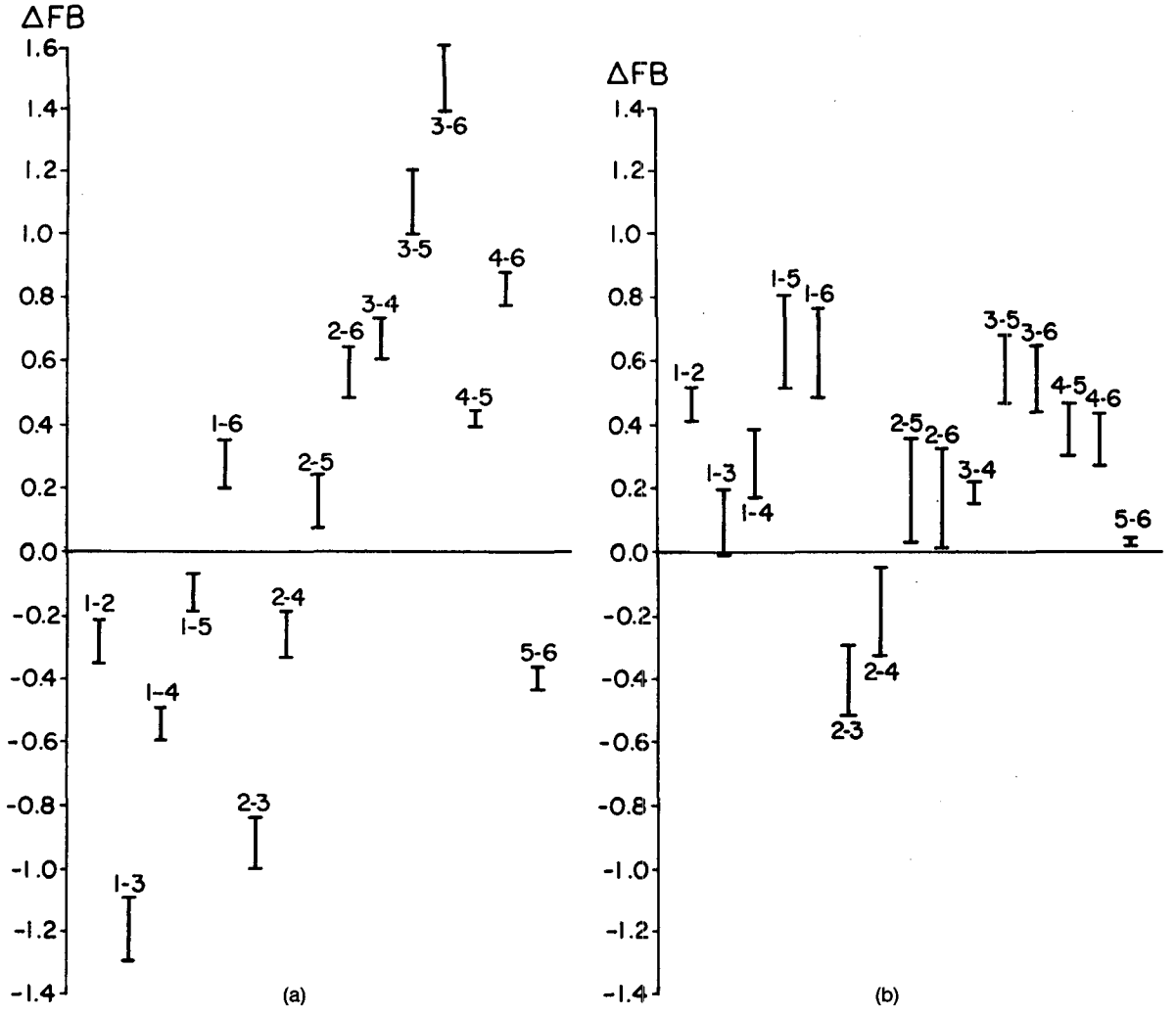

FIG. 9. Ninety-five percent confidence intervals on differences in FB (ΔFB) between models or system configurations when CCB dispersion scheme is replaced by Pasquill–Gifford scheme in the TAPAS configurations (1 = CCB–N/I, 2 = CCB–L/W, 3 = TAP–UNI, 4 = TAP–POT, 5 = TAP–αF, and 6 = TAP–αE) (a) $h > H_c$ ensemble and (b) $h \leq H_c$ ensemble. If the confidence interval does not overlap 0.0, then we have 95% confidence that the FBs for the two models are different.

For the ensemble with $h > H_c$, both TAP–αF and TAP–αE are not significantly different with respect to NMSE; however, TAP–αF slightly underpredicts while TAP–αE tends to slightly overpredict. As a consequence, both models have a slightly inferior predictive performance compared to the best of the CCB models (CCB–N), which have a FB that is not significantly different from zero. The NMSE values of all three models (TAP–αF, TAP–αE, and CCB–N) are not significantly different.

For the ensemble with $h \leq H_c$, TAP–αF is the best of the TAPAS configurations with a significantly smaller NMSE than TAP–αE, although both models have a FB that is not significantly different from zero. Comparison with the best of the CCB models (CCB–W) indicates that the TAPAS configuration may be superior on the basis of its significantly smaller NMSE, although both models (TAP–αF and CCB–W) have a FB that is not significantly different from zero.

Figures 11a and b contain scatterplots of predictions versus observations for the best of the TAPAS configurations based on the Pasquill–Gifford scheme for the ensemble of hours for which $h > H_c$ and $h \leq H_c$, respectively. Comparison with the scatterplots for the appropriate CCB models in Fig. 2 provides a simple illustration of the results discussed previously.

6. Discussion and conclusions

A principal objective of this work is the evaluation of the current technique for incorporating terrain and atmospheric stability effects into the 3D diagnostic wind-field model NUATMOS when only a background wind vector is available. Although based on the use of

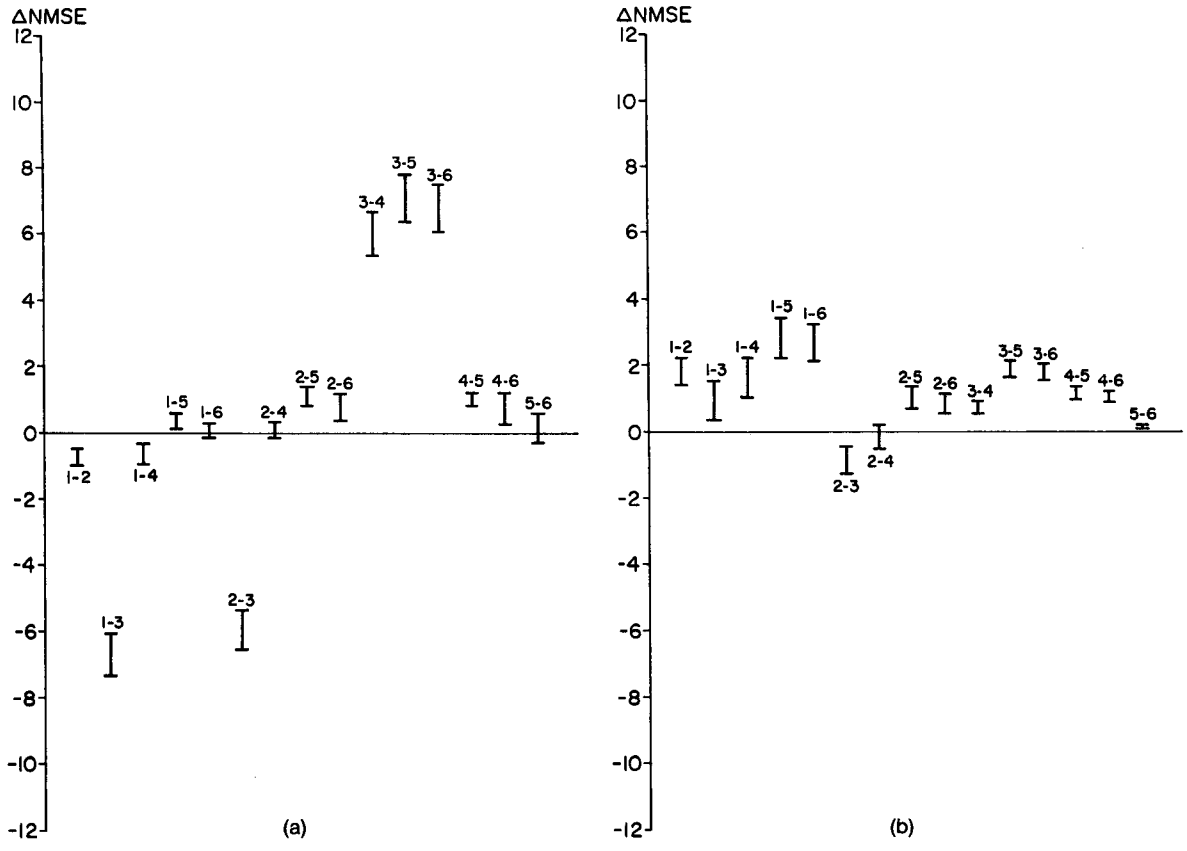

FIG. 10. Ninety-five percent confidence intervals on differences in NMSE (ΔNMSE) between models or system configurations when CCB dispersion scheme is replaced by Pasquill–Gifford scheme in the TAPAS configurations (1 = CCB–N/I, 2 = CCB–L/W, 3 = TAP–UNI, 4 = TAP–POT, 5 = TAP–αF, and 6 = TAP–αE) (a) $h > H_c$ ensemble and (b) $h \leq H_c$ ensemble. If the confidence interval does not overlap 0.0, then we have 95% confidence that the FBs for the two models are different.

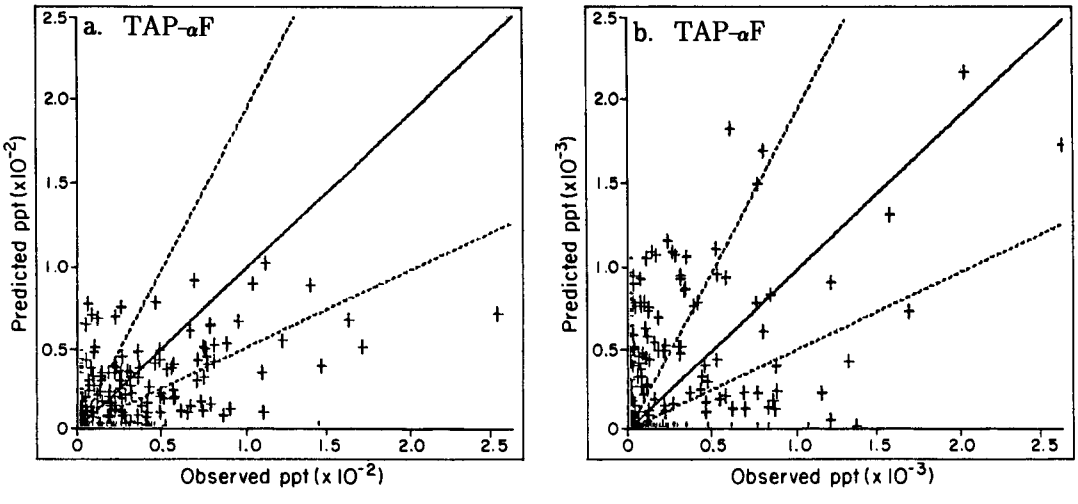

FIG. 11. Scatterplots of predicted versus observed concentrations (in ppt), paired in location and time, for the "best" of the TAPAS configurations TAP–αF based on the Pasquill–Gifford dispersion scheme (a) $h > H_c$ and (b) $h \leq H_c$. The area between the dashed lines contains points that are within a factor of 2 of the solid (perfect fit) line.

"indirect" field measurements, the statistical analysis adopted, together with the form of the CCB data, enables us to focus on these key flow field adjustments.

The results presented for the ensemble of case study hours and for two component ensembles; one with source release heights above the critical streamline height and the others below this height, demonstrate that significant improvement in the predictive ability of TAPAS results when the initially uniform flow field (based on interpolation of the background wind) is made divergence-free (with $\alpha = 1$) in response to the terrain shape. The improvement in performance is considerably larger for the case where $h > H_c$ than for that where $h \leq H_c$.

The predictive performance of TAPAS is seen to be further improved when the parameter α is used to incorporate atmospheric stability effects during the divergence reduction stage of NUATMOS. Determining α on the basis of a characteristic Froude number appears to be slightly inferior to using a value obtained by matching the experimental and predicted critical streamline heights, although this finding is at the margin of significance (at the 95% confidence level) and is reversed for the cases where $h \leq H_c$ when the CCB dispersion scheme is replaced by the Pasquill–Gifford scheme. The ensemble of cases where $h > H_c$ is inconclusive.

A secondary objective was to evaluate the overall performance of TAPAS and compare its performance with that of models specifically developed from the CCB dataset.

Comparison between the best of the TAPAS and CCB configurations indicates that TAPAS performs as well but not better than the CCB models, except for the ensemble of cases where $h > H_c$ when TAPAS yields a significantly lower NMSE. This result is encouraging given that the CCB models have been developed and calibrated using the CCB dataset itself, whereas, apart from the CCB dispersion scheme, TAPAS uses basic input information and subsequently derived flow field adjustments.

The results obtained by replacing the CCB dispersion scheme with the Pasquill–Gifford scheme are even more encouraging given that the TAPAS results are now almost totally based on inputs that are either routinely available or readily derived from routinely available information.

We have not attempted to compare the relative performance of the two dispersion schemes used in TAPAS or to conduct a "scientific evaluation" of them. The dispersion scheme and its interaction with the flow field represents a weak link in most models and, apart from the inherent uncertainties associated with the stochastic atmosphere, is a major cause of the relatively large NMSE resulting from even the best of the models examined here.

As air flows past CCB, the terrain-induced flow field will produce kinematic effects on the spread and growth of a plume. For example, the speedup over the crest will result in streamline compression in the vertical, with the streamlines also approaching the terrain surface more closely. In contrast, the streamlines in the crosswind direction will "fan out" over the crest. These kinematic influences will act to decrease σ_z and increase σ_y over the crest (e.g., Egan 1975). However, it is important to realize that final plume or puff spread depends also on the turbulent diffusion rates that result from these kinematic effects. Hunt and Mulhearn (1973), in examining the theory of turbulent plumes imbedded within potential flow fields, point out that turbulent diffusion across streamlines is enhanced by contraction in the distance between streamlines in the vertical and retarded by the expansion in the distance between streamlines in the lateral.

The TAPAS system, via NUATMOS, simulates the terrain-induced effects on the streamlines in terms of their relative separation and closeness to the terrain, with the parameters α playing a major role in controlling the magnitude of these influences. However, the CITPUFF dispersion algorithm and its coupling with the flow field does not directly include the full kinematic and turbulent diffusion rate influences on the spread of puffs. In particular, each puff is transported by only the velocity at its centroid, and its dispersion about the centroid is effectively independent of the flow field. There is, of course, an indirect influence of the flow field on the "effective" puff dispersion used in calculating the ground-level concentration field via the puff size at a particular location and the position of its centroid relative to the terrain.

The results presented are part of an ongoing study to evaluate and further develop TAPAS and its component models using experimental results from the CTMD program. The Hogback Ridge dataset will enable us to focus on an isolated, approximately 2D ridge, while the Tracy power plant dataset will enable us to evaluate the findings from the simple isolated 2D and 3D terrain in the context of more complex terrain.

One advantage of the analysis presented in this paper is that it provides confidence in the relative applicability of models for use in regulatory settings. Although the dataset is not "realistic" from a regulatory perspective, it does allow specific presentation of over- and underprediction. Inasmuch as responsible regulatory acceptance of models requires knowledge of and display of these characteristics, the TAPAS configurations presented here have regulatory utility.

Acknowledgments. The support provided under Cooperative Agreement 28-C8-478 between Monash University and the Rocky Mountain Forest and Range Experiment Station, USDA, is gratefully acknowledged.

The assistance of Mr. Andrew Lewis in performing the model runs and developing many of the graphics and analysis routines is also gratefully acknowledged.

REFERENCES

Connell, B. H., 1988: Evaluation of a three-dimensional diagnostic wind model: NUATMOS. Unpublished M.S. thesis, Department of Natural Resources. Colorado State University, Fort Collins, 135 pp.

DiCristofaro, D. C., D. G. Strimaitis, B. R. Greene, R. J. Yamartino, A. Venkatram, D. A. Godden, T. F. Lavery and B. S. Egan, 1985: Environmental Protection Agency complex terrain model development program: Fifth Milestone Rep.–1985. EPA-600/3-85-069, United States Environmental Protection Agency, Research Triangle Park, NC.

Egan, B. A., 1975: Turbulent diffusion in complex terrain. *Lectures in Air Pollution and Environment Impact Analysis,* D. A. Hauger, Ed. Amer. Meteor. Soc.

Fox, D. G., 1984: Uncertainty in air quality modeling. *Bull. Amer. Meteor. Soc.,* **65,** 27–36.

——, D. G. Ross, D. L. Deitrich and D. E. Mussard, 1987: An update on TAPAS and its model components. Preprint Volume, *Ninth Conference on Fire and Forest Meteorology,* San Diego, Amer. Meteor. Soc., 135–137.

Hanna, S. R., 1989: Confidence limits for air quality model evaluations, as estimated by bootstrap and jackknife resampling methods. *Atmos. Environ.,* **23,** 1385–1398.

——, G. A. Briggs, J. Deardorff, B. A. Egan, F. A. Gifford and F. Pasquill, 1977: Summary of recommendations made by the AMS workshop on stability classification schemes and sigma curves. *Bull. Amer. Meteor. Soc.,* **58,** 1305–1309.

Hunt, J. C. R., and P. J. Mulhearn, 1973: Turbulent dispersion from sources near two-dimensional obstacles. *J. Fluid Mech.,* **61,** 245–274.

Lorimer, G. S., 1989: Validation of air pollution dispersion models. *Clean Air (Aust.),* **23,** 82–88.

Perry, S. G., D. J. Burns, L. A. Adams, R. J. Paine, M. G. Dennis, M. T. Mills, D. G. Strimaitis, R. J. Yamartino and E. M. Insley, 1989: User's guide to the complex terrain dispersion model plus algorithms for unstable situations (CTDMPLUS) Vol. 1: Model description and user instructions. Environmental Protection Agency Rep. EPA/600/8-89/041, United States Environmental Protection Agency, RTP, NC, 196 pp.

Ross, D. G., G. S. Lorimer, L. Li and I. N. Smith, 1987: "CITPUFF": A Gaussian Puff Model for estimating pollutant concentration in complex terrain CAMM Rep. No. 21/87, 72 pp. [Available from Chisholm Institute of Technology: P.O. Box 197, Caulfield, Victoria, Australia.]

——, I. N. Smith, P. C. Manins and D. G. Fox, 1988a: Diagnostic wind field modeling for complex terrain: Model development and testing. *J. Appl. Meteor.,* **27,** 785–796.

——, M. Krautschneider, I. N. Smith and G. S. Lorimer, 1988b: Diagnostic wind field modeling: Development and validation. End of Grant Rep. to Department of Resources and Energy, National Energy Research Development, and Demonstration Program. End of Grant Rep. No. NERDDP EG89/776, 108 pp.

——, G. S. Lorimer, D. G. Fox and I. N. Smith, 1989: Results of applying wind and dispersion models in complex topography. Preprint Volume, *82nd Annual Meeting and Exhibition. Air and Waste Management Association,* Anaheim, 1–16.

Sheppard, P. A., 1956: Airflow over mountains. *Quart. J. Roy. Meteor. Soc.,* **82,** 528–529.

Smith, R. B., 1988: Linear theory of stratified flow past an isolated mountain in isosteric coordinates. *J. Atmos. Sci.,* **45,** 3889–3896.

Smith, I. N., and D. G. Ross, 1988: Diagnostic wind field studies. *Clean Air (Aust.),* **22,** 141–143.

Spangle2, T. C., 1983: Stagnation zone impaction on a simple terrain feature in stable flow. Ph.D. dissertation, Utah State University, Logan, Utah, 175 pp.

——, 1986: The role of near terrain turbulence in the prediction of ground level pollutant concentrations in complex terrain. *Atmos. Environ.,* **20,** 861–865.

Strimaitis, D. G., A. Venkatram, B. R. Greene, S. Hanna, S. Heisler, T. F. Lavery, A. Bass and B. A. Egan, 1983: Environmental Protection Agency complex terrain model development program: Second milestone rep.—1982. EPA-600/3-83-015, United States, 375 pp.

——, R. J. Paine, B. A. Egan and R. J. Yamartino, 1987: Environmental Protection Agency complex terrain model development: final rep. Environmental Protection Agency report EPA/600/3-88/006, Environmental Protection Agency, RTP, NC, 486 pp.