

H 96-56
V1-B-09(22)

CORRECTING RADM'S SULFATE UNDERPREDICTION: DISCOVERY AND CORRECTION OF MODEL ERRORS AND TESTING THE CORRECTIONS THROUGH COMPARISONS AGAINST FIELD DATA

ROBIN L. DENNIS,*† JOHN N. MCHENRY,‡ W. RICHARD BARCHET,§
FRANCIS S. BINKOWSKI† and DAEWON W. BYUN†

*Atmospheric Sciences Modeling Division, Air Resources Laboratory, National Oceanic and Atmospheric Administration, Research Triangle Park, NC 27711, U.S.A., ‡Computer Sciences Corporation, Applied Technology Division, Research Triangle Park, NC 27709, U.S.A. and §Pacific Northwest Laboratory, Atmospheric Sciences Department, Richland, WA 99352, U.S.A.

(First received 2 December 1991 and in final form 21 August 1992)

Abstract—A serious underprediction of ambient sulfate (SO_4^{2-}) by two comprehensive, Eulerian models of acid deposition, the Regional Acid Deposition Model (RADM) and the Acid Deposition and Oxidant Model (ADOM), was found in the National Acid Precipitation Assessment Program phase of the Eulerian Model Evaluation Field Study (EMEFS) model evaluation. Two hypotheses were proposed to explain the cause of the underprediction in RADM: insufficient SO_4^{2-} production by nonprecipitating convective clouds and insufficient primary SO_4^{2-} emissions. Modifications of the RADM cloud and scavenging module to simulate nonprecipitating cumulus clouds better are described in detail. Three contrasting pairs of tests using data from the EMEFS were applied to these hypotheses: source vs downwind regions, mid-summer vs late summer seasons and sunny-dry vs cloudy-wet synoptic types. The SO_4^{2-} emissions hypothesis, tested by artificially boosting SO_4^{2-} emissions, fared better than expected but was rejected because of its poor performance on the regional and seasonal contrast tests. The RADM nonprecipitating cumulus modification successfully captured the seasonal and the late summer synoptic contrasts but improvement is still needed for the regional and mid-summer synoptic contrasts.

Key word index: Regional acid deposition modeling, RADM, regional ambient sulfate, nonprecipitating clouds and regional sulfate, regional model evaluation, RADM evaluation, Eulerian Model Evaluation Field Study.

INTRODUCTION

The point of model evaluation is to establish the credibility of a model for use in decision-making. Most model applications require that the model extrapolate well beyond current precursor and primary emission conditions that could exist in any model evaluation data set. This is particularly true for issues that span the urban to global scales, such as oxidants, acidic deposition and visibility. Thus, a model evaluation needs to test the science in the models. Testing the science means looking for the "right" kind of answer (right answer for right reason and wrong answer for right reason), rather than simply looking for good comparisons of the final outcome.

In this paper we use model evaluation field data to discover and correct a major error in the Eulerian, Regional Acidic Deposition Model, RADM (Chang *et al.*, 1987, 1990), and then test the modified model to see if we obtained the right correction. Parallel corrections to the Acid Deposition and Oxidant Model, ADOM, the Canadian Eulerian acidic deposition

model (Venkatram *et al.*, 1988), are also mentioned. Secondly, we note the diagnostic capability of field data to test the correctness of modifications to models that improve their operational performance. Because the modification to RADM highlights the importance of cloud processes to SO_4^{2-} production, it is described in some detail.

The design objective of RADM is to incorporate all known major atmospheric physical and chemical processes related to acidic deposition so that we have a scientific basis for estimating the change in deposition due to major changes in precursor emissions. The present version covers the geographic domain east of central Texas and south of James Bay, Canada (2800 × 3040 km). It uses grid cells of 80 × 80 km and has 6 or 15 nonuniformly separated layers in the vertical, covering the distance from the ground to approximately 16 km in altitude. The horizontal and vertical transport time step is 300 s; the gas-phase chemistry time step is frequently considerably shorter than 300 s; changes due to cloud effects are computed using a time step limited by aqueous chemistry or scavenging (approximately 1 s); and integrated changes due to clouds are imposed at hourly intervals. The RADM predictions are output at hourly intervals. The high temporal resolution of the calculations

† On assignment to the Atmospheric Research and Exposure Assessment Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, U.S.A.

is required to describe the nonlinear chemical dynamics.

Original sulfate underprediction

RADM and ADOM are being evaluated in a two-phase, binational program called the Eulerian Model Field Study (EMEFS), described by Hansen *et al.* (1989) and Dennis *et al.* (1990). Data at surface sites were collected from the beginning of July 1988 to the end of May 1990, a period of 2 years. Special measurements by aircraft and at designated surface sites were taken during two periods: 15 July–27 September 1988 and 15 March–30 May 1990. The first stage of phase 1 of the model evaluation was performed for the National Acid Precipitation Assessment Program (NAPAP) using EMEFS data from 25 August to 27 September 1988. Results from this stage of the evaluation were reported in Dennis *et al.* (1990) for RADM. The second stage of phase 1 of the evaluation examines model performance after an interim round of improvements were incorporated into RADM and ADOM. The results of the improvements to RADM

are reported here. For clarity, we will designate RADM prior to the improvements as RADM2.1 and the improved version of RADM as RADM2.6. Phase 1 of the EMEFS evaluation, when completed in spring 1992, will have used data from the first 3 months of the EMEFS. The EMEFS evaluation of ADOM and RADM is expected to continue through 1993.

The NAPAP evaluation examined model performance with respect to air concentrations of SO_2 , SO_4^{2-} , HNO_3 , O_3 , H_2O_2 , and wet deposition and rainwater concentrations of SO_4^{2-} and NO_3^- (Dennis *et al.*, 1990). Although none of the comparisons for these species yielded perfect agreement, one underprediction in particular was of concern. Both ADOM and RADM underpredicted ambient SO_4^{2-} at the surface by approximately a factor of 0.6 (see top of Fig. 1). Yet, total sulfur was not underpredicted (see bottom of Fig. 1). Thus, there appeared to be a problem with how the models were converting SO_2 to SO_4^{2-} . The discrepancy was severe enough to consider that something must be missing or very wrong in the models.

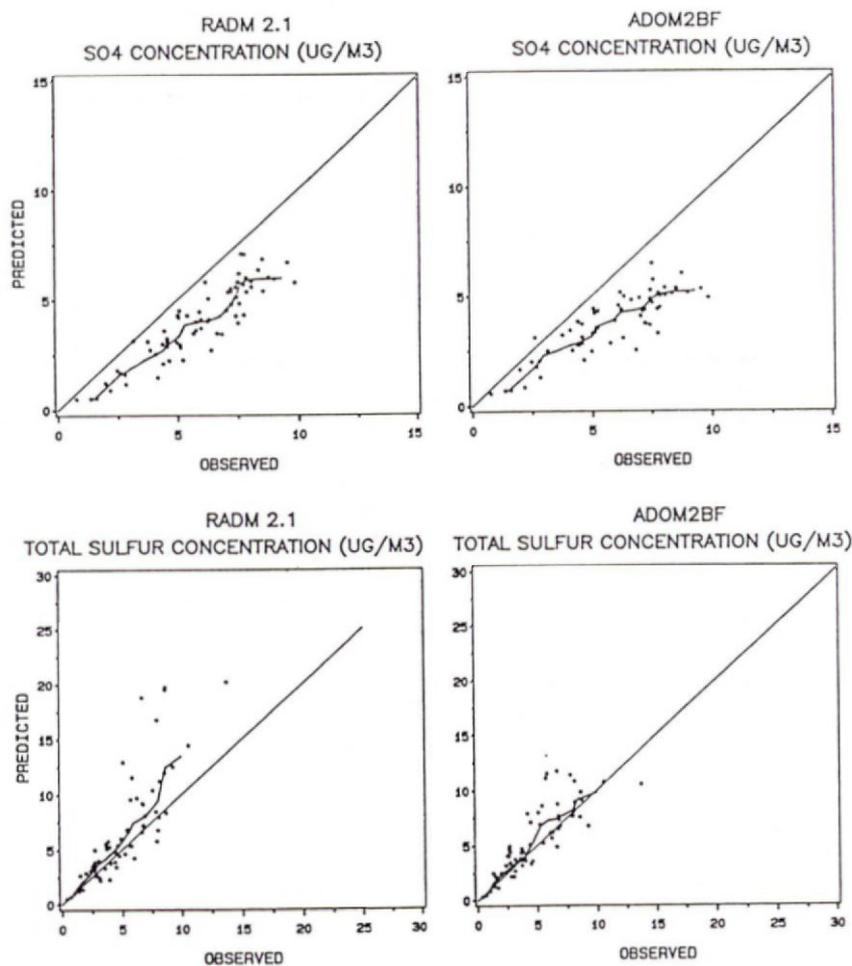


Fig. 1. The relationship between model-predicted and EMEFS-observed sulfate aerosol and total ambient sulfur for the 25 August–27 September 1988 period: (top) sulfate aerosol, (bottom) total sulfur; (left) RADM2.1, (right) ADOM2BF.

CORRECTION OF MODEL ERRORS

Search for a cause

An extensive examination of comparisons between model predictions and measurements yielded little in the way of clues for the underprediction. The aircraft flights were generally in clear air conditions and the underprediction aloft seemed to be much less than at the surface. However, the scatter was very large, more than a factor of 4. The degree of underprediction at the surface for the days in which the aircraft flew was basically the same as for "non-aircraft" days. Comparison to ambient SO_4^{2-} measurements from the National Dry Deposition Network showed that the degree of underprediction was similar in all seasons.

An analysis of how RADM2.1 simulated the production of SO_4^{2-} showed that a substantial amount of SO_4^{2-} from evaporating precipitating clouds was later scavenged by cloud processes. Total cloud-based, aqueous-phase production of SO_4^{2-} was predicted to account for 40–60% of the total sulfur deposition (McHenry and Dennis, 1991). Because clouds efficiently convert SO_2 to SO_4^{2-} , a major fraction of the predicted air concentration of SO_4^{2-} in RADM2.1 came from precipitating cloud systems. This is consistent with evidence (Altshuler, 1987; Gillani *et al.*, 1981; Gillani and Wilson, 1983) and theoretical studies (Seigneur and Saxena, 1988) that clouds produce a substantial amount of SO_4^{2-} . In order to remove the observed underprediction of SO_4^{2-} , an unrealistic doubling of the SO_4^{2-} production from clear air, gas-phase oxidation would be required. Thus, a leading candidate for a production process not adequately simulated was production of SO_4^{2-} by nonprecipitating clouds.

A detailed examination of how RADM2.1 handled nonprecipitating clouds was undertaken, and a set of scientific shortcomings that could potentially explain a significant portion of the SO_4^{2-} underprediction was enumerated. These are summarized below.

First, the average fractional cloud coverage by nonprecipitating clouds predicted by the model appeared to be unrealistically low, possibly by a factor of 5 or more. Fractional coverages for individual cloud fields simulated in one grid column for 1 h (the cloud lifetime) were limited to a maximum of 20%. By aggregating simulations for a series of meteorological cases into an annual average coverage (Samson *et al.*, 1990; Dennis *et al.*, 1990), we found that the annual average coverage predicted by RADM2.1 is less than 5% over most of the continental U.S.A. (Fig. 2a).

Additionally, the fractional coverages for nonprecipitating clouds were determined without any physical connection to the coverage determined for precipitating clouds. At times, this led to fractional coverages exceeding 100%. Computation of the annual averages for total cloud coverage (precipitating plus nonprecipitating) shown in Fig. 2b, demonstrates that the total coverage predicted by RADM2.1 is of the order of 5–15%.

Annual estimates derived from 2 years of satellite data show that total cloud coverage is rarely below 30% over the continental U.S.A. (Wylie and Menzel, 1989). This is depicted in Fig. 2c. Other studies of cloud climatologies, notably by Angell (1990), Rossow (1989) and Ardanuy *et al.* (1989) support Wylie and Menzel's results. Statistics from Angell (1990), compiled over a 38-year period from ground-based observations, suggest that the annual average fractional sky coverage is around 60%. These comparisons suggest that RADM2.1's total cloud cover may have been low by a factor of 2–3.

It is total cloud volume that, in reality, determines the amount of SO_4^{2-} production by in-cloud aqueous chemistry and subsequent wet deposition. For example, a partly cloudy sky of towering cumuli may occupy a larger volume in a grid column than a thin cloud that covers most of the grid. In this initial analysis of RADM2.1 cloud climatology, we used the fractional cloud coverage as a surrogate for the overall cloud volume.

Second, the vertical extent of nonprecipitating clouds appeared to be too low, possibly by a factor of 2 in many circumstances. RADM2.1 required that all nonprecipitating clouds have their cloud base (defined by the lifting condensation level) and cloud top (determined by relative humidity) be less than 1500 m above ground level. This constrained the frequency with which nonprecipitating clouds were simulated. For example, if no precipitation was associated with the grid cell and the temperature and humidity sounding within the grid column supported a cloud base below 1500 m AGL and a cloud top above 1500 m AGL, the presence of this "taller" nonprecipitating cloud, perhaps extending well into the free troposphere, was ignored. In a similar fashion, the simulation ignored nonprecipitating clouds whose bases were predicted to occur above 1500 m AGL. Hence, the limitation on vertical extent constrained both the overall number of clouds predicted as well as their resultant volume.

Third, RADM2.1's simulation of cloud-induced mixing of pollutants resulted in a mixing of pollutant mass throughout the depth of the model. In effect, the mixing algorithm allowed even the smallest, low-level clouds to influence concentrations of pollutants up to the top of the model (100 mb). It was clear that this unrealistic mixing needed correcting.

Sensitivity studies indicated that correction of these shortcomings had the potential to increase RADM production of SO_4^{2-} sufficiently to eliminate a major portion of the underprediction by the model. Therefore, revising the RADM2.1 nonprecipitating cloud module became the top priority for refinements related to eliminating the SO_4^{2-} underprediction.

Other, more general problems exist with RADM's simplified convective-cloud approach to modeling clouds. Ideally, RADM would have access to data that could, at least, distinguish between convective and stratiform types of clouds under both precipita-

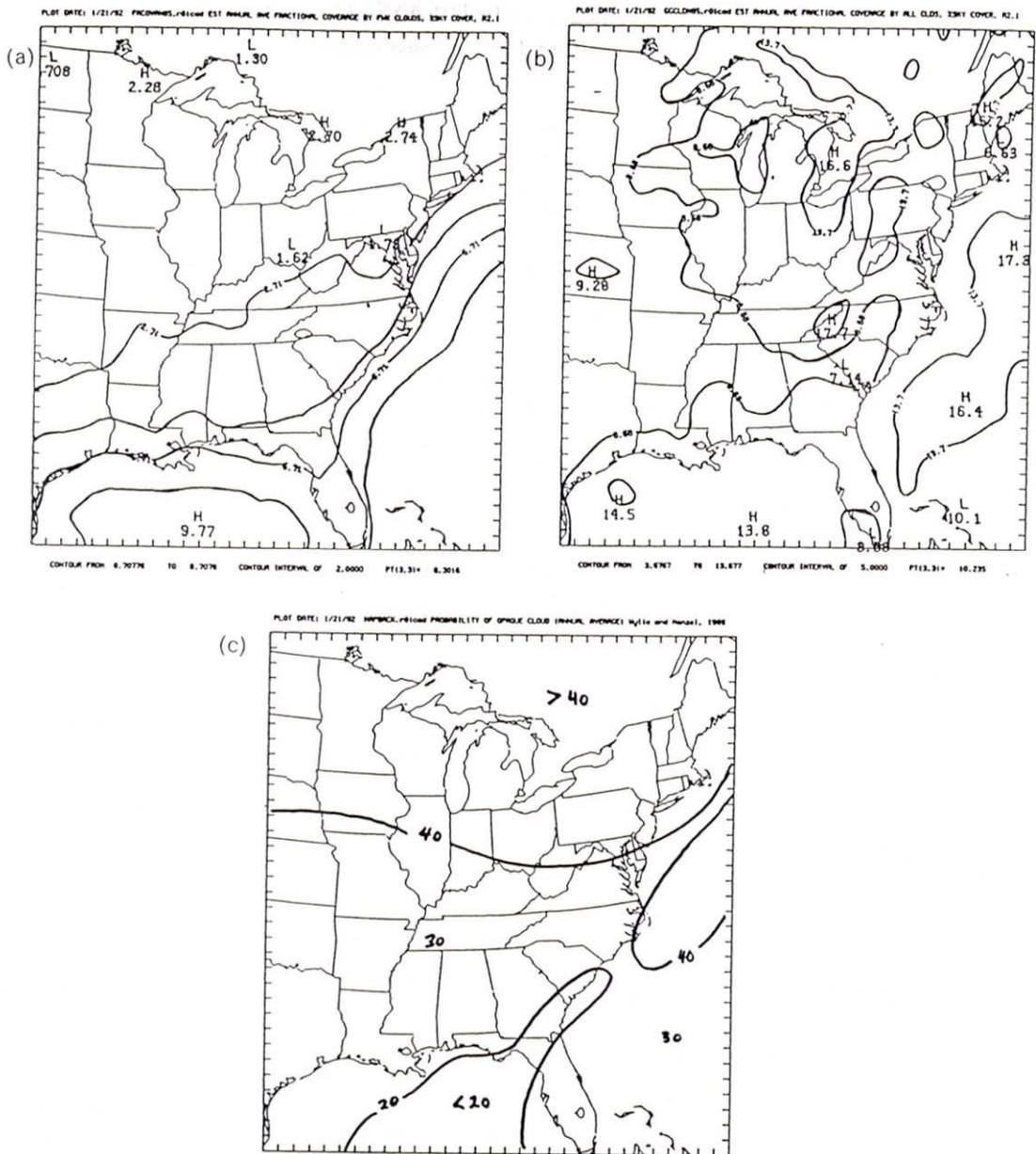


Fig. 2. Average annual fractional cloud cover: (a) RADM2.1 nonprecipitating clouds, (b) RADM2.1 all clouds, (c) observed.

ting and non-precipitating conditions. Dynamic consistency between the cloud model(s) in RADM's meteorological driver (MM4; see Anthes and Warner, 1978), and in RADM itself is felt to be a critical component to a "second generation" regional air pollution cloud model. Many desired features are not currently part of the MM4 system, (e.g. advection of cloud water; cloud parameters such as cloud base, top, fractional coverage; and liquid water content). Thus, our corrections are clearly within the limits of the simplified convective model originally conceived for RADM; more complex corrections await the development of a comprehensive cloud model within its me-

teorological driver. Our comparisons of the corrected (RADM2.6) model against the EMEFS data must also be viewed in this light.

With this in mind, a brief description of the functionality of the RADM cloud model will be given, followed by a more detailed explanation of the approach taken to resolve each of the three aforementioned shortcomings of the RADM2.1 cloud model.

Functionality of the cloud module in RADM

The RADM2.1 cloud module simulates both precipitating and nonprecipitating cumulus clouds.

Other cloud types such as low stratus, middle and high clouds are ignored. This shortcoming will be remedied in future versions of RADM when better cloud parameterizations are included in the meteorological driver, MM4. The purpose of the present work is to examine the hypothesis that there are not enough nonprecipitating cumulus clouds being simulated by RADM. Further, most of the comparisons with field data are for cases with cumulus clouds. Therefore, the neglect of the other cloud types has little or no influence on the results. Because of this limited purpose, aqueous chemistry and precipitation processes remain the same as RADM2.1. The modifications that are reported here are meant to be illustrative of the importance of nonprecipitating cumulus clouds. The modifications are not meant to be definitive, nor are they meant to be final.

All clouds present within each grid column are simulated by modeling a representative cloud that is assumed to exist for 1 h. The cloud simulation is a parameterization and not a modeling of basic physical processes. Chemical reactions within and scavenging of pollutants by the cloud changes the grid-averaged trace gas concentrations. The RADM2.1 cloud module has been described by Walcek and Taylor (1986) and, more recently, by Chang *et al.* (1990). In this section we will briefly summarize the key functional components of the RADM2.1 version of the model.

First, at each simulated hour, the module determines whether a cloud exists in the grid column. It does so by testing independently for the presence of precipitating and nonprecipitating clouds. A precipitating cloud is assumed to exist if the grid-averaged hourly precipitation predicted by the MM4 mesoscale meteorological model (Seaman and Stauffer, 1989) exceeds 0.01 cm. Nonprecipitating cumulus clouds are predicted to occur if the vertical moisture and temperature soundings for the grid column support a cloud base and top that are both less than 1500 m AGL and the relative humidity exceeds 70% at the cloud source level, which is defined as the layer with the highest equivalent potential temperature between the middle of the first model layer and 650 mb. If the cloud source level is not below 1500 m as well, no cloud is predicted to occur.

The cloud base and top are established next. Cloud base is defined as the lifting condensation level (LCL), which is computed by lifting air from the cloud source level. At a given hour, the LCL defines the cloud base for both precipitating and nonprecipitating clouds. Calculation of cloud top depends on cloud type.

The module then determines the fractional cloud coverage. A parameterization similar to Kuo's (1974) is used for precipitating clouds whereas a parameterization based on relative humidity is used for nonprecipitating clouds. Because the fractional coverages of the two cloud types are independent there is no limit to the total fractional coverage for all clouds which co-exist at the same hour.

Profiles of entrainment of environmental air, based on a top-down entrainment model, and profiles of liquid water content and cloud temperature are then determined. Following this, the cloud mixing algorithm is applied to compute the in-cloud pollutant concentrations. A time-dependent set of mass conservative chemical kinetics and scavenging equations for all chemical species in cloud and rain water is then solved in a box submodel. These calculations provide the post-reaction in-cloud concentrations from which the wet deposition of pollutants is determined and with which new gas-phase trace gas concentrations are computed.

Modifications to increase cloud volume

In the revised module, referred to here as the RADM2.6 cloud module, an increase in cloud volume is achieved by changing the parameterization for the fractional coverage of precipitating and nonprecipitating clouds and increasing the allowable depth of nonprecipitating clouds.

Mass flux constraints are used to prevent the sum total of convective cloud mass from exceeding a reasonable, although arbitrarily assigned, fraction of the mass potentially available for lifting into the convective portion of the cloud. The mass potentially available for cloud formation is the mass that exists below cloud base prior to cloud formation:

$$M_{bc} = A_t \sum_1^{k_b-1} \rho(k) \quad (1)$$

where A_t is the area of the grid cell, in m^2 , $\rho(k)$ is the air density of the k th model layer, in mol m^{-2} , and k_b is the model layer in which cloud base is located. The mass of air lifted to form the convective cloud is given by

$$M_{cv} = A_c \sum_{k_b}^{k_t} \rho(k) [1 - F(k)], \quad (2)$$

where A_c is the area of the convective cloud, k_t is the model layer in which cloud top is located, and $F(k)$ is the entrainment fraction. The method used to compute $F(k)$ is described by Walcek and Taylor (1986).

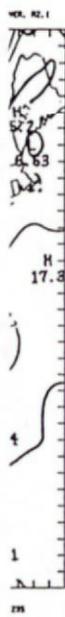
The mass flux constraint requires that the total convective cloud mass be limited by

$$M_{cv, \text{total}} \leq \alpha M_{bc}, \quad (3)$$

where $\alpha = 0.95$ when precipitating and coexisting but nonprecipitating (hereafter designated CNP) clouds occur and $\alpha = 0.5$ when only fair weather (FW) nonprecipitating clouds exist. For precipitating or FW clouds, the maximum allowable fractional coverage is given by

$$\text{maxfrac} = \frac{A_c}{A_t} = \frac{0.5 \sum_1^{k_b-1} \rho(k)}{\sum_{k_b}^{k_t} \rho(k) [1 - F(k)]} \quad (4)$$

The final fractional coverage by precipitating and



served.
 rrected
 i must
 e func-
 given,
 he ap-
 remen-
 model.
 th pre-
 clouds.

nonprecipitating cloud types is determined by the following procedures.

If the meteorological driver for RADM indicates that precipitation occurs in the grid cell, cloud base and cloud top are calculated and an initial guess is made for the fractional coverage (as in RADM 2.1):

$$\text{FRPR}_i = \min \left[\frac{P_r \tau_{\text{CLD}}}{\epsilon_s Q_{xs}}, 0.999 \right], \quad (5)$$

where P_r is the grid cell hourly average precipitation rate ($\text{kg m}^{-2} \text{h}^{-1}$), τ_{CLD} is the cloud lifetime of 1 h, ϵ_s is a storm efficiency roughly derived from Fritsch and Chappell's (1980) precipitation efficiency, and Q_{xs} is the excess water (kg m^{-2}) available for precipitating from the cloud.

The maximum allowable fractional cloud coverage is calculated from Equation (4) using the cloud base and top values for the precipitating cloud. The actual precipitating cloud fraction is then determined by

$$\text{FRPR}_f = \min[\text{FRPR}_i, \text{maxfrac}]. \quad (6)$$

If FRPR_f equals either maxfrac or 0.999, ϵ_s is calculated from

$$\epsilon_s = \frac{P_r \tau_{\text{CLD}}}{\text{FRPR}_f Q_{xs}} \quad (7)$$

so that wet deposition can be calculated correctly.

Next, a determination is made as to whether or not a CNP cloud exists. Increases in the frequency of occurrence of this type of nonprecipitating cloud are achieved by relaxing the RADM2.1 existence criteria. A CNP cloud exists if the LCL is located at or below 3000 m and the cloud source level relative humidity is 70% or higher. Additional cloud depth is gained by allowing the top of a CNP cloud to range up to 500 mb, provided the environmental relative humidity is above 65%. In essence, these new criteria establish areas of nonprecipitating clouds that are contiguous with the major precipitation regions and convective systems and whose tops may extend well into the free troposphere.

If a CNP cloud exists, an initial guess of its fractional coverage, in relation to FRPR_f , is made from

$$\text{FRNP}_i = \begin{cases} \min[1 - \text{FRPR}_f, 0.9] \left(\frac{\text{RH}_s - 0.7}{0.9 - 0.7} \right), & 0.7 < \text{RH}_s < 0.9 \\ \min[1 - \text{FRPR}_f, 0.9], & \text{RH}_s \geq 0.9, \end{cases} \quad (8)$$

where RH_s is the relative humidity at the cloud source level. Equations (6) and (8) together limit the initial guess for the total fractional coverage of both types of clouds to 1. If the precipitating cloud fraction is small, the CNP cloud cover fraction may be as large as 90%. This is a significant increase over the RADM2.1 limit of 20% on the fractional coverage of all nonprecipitating clouds. However, the maximum fractional coverage by a CNP cloud is constrained by allowing no

more than 90% of the amount of below-cloud air that remains after forming the precipitating cloud to form the CNP cloud. Algebraic manipulation leads to the following:

$$\text{maxfrac}_{\text{NP}} = 0.9 \left[\frac{M_{bc} - M_{cv,PR}}{A_i \sum_{k_b} \rho(k) [1 - F(k)]} \right], \quad (9)$$

where M_{bc} is the same for both the precipitating and nonprecipitating cloud, $M_{cv,PR}$ is the convective mass in the precipitating cloud, and quantities within the sum refer to the CNP cloud. Finally, the actual CNP cloud cover fraction is determined from

$$\text{FRNP}_f = \min[\text{FRNP}_i, \text{maxfrac}_{\text{NP}}]. \quad (10)$$

For hours in which no precipitation occurs in the cell, nonprecipitating, fair weather (FW) clouds may still exist if the temperature and humidity soundings support convection. As in RADM 2.1, FW clouds are not allowed unless the LCL lies at or below 1500 m AGL. Additional FW cloud frequency is gained by permitting the existence of FW clouds whose tops exceed 1500 m AGL. However, the tops of such "taller" FW clouds are then artificially reduced to about 1500 m AGL. As compared to RADM 2.1, this provides an increase in FW cloud frequency, but no depth, in an attempt to parameterize the net effect of large numbers of relatively low, thin FW clouds.

As with the other two cloud types, an initial guess is made of the fractional coverage of a FW cloud:

$$\text{FRFW}_i = \begin{cases} 0.9 \left(\frac{\text{RH}_s - 0.7}{0.9 - 0.7} \right), & 0.7 < \text{RH}_s < 0.9 \\ 0.9, & \text{RH}_s \geq 0.9 \end{cases} \quad (11)$$

While this initial guess allows the maximum permissible fractional coverage to be 90%, as compared to 20% in RADM 2.1, the final fractional coverage for FW clouds is determined by

$$\text{FRFW}_f = \min[\text{FRFW}_i, \text{maxfrac}_{\text{FW}}], \quad (12)$$

where $\text{maxfrac}_{\text{FW}}$ is evaluated according to Equation (4) but using the cloud base and top specific to the FW cloud.

Note that these equations assume that, in the lifting process, only one dynamic cycle of convection and subsidence occurs over the lifespan of the cloud, and that, within the grid column, the fundamental cloud formation process is identical for clouds which coexist. Additionally, the possibility of below-cloud convergence of air through the grid cell walls is ignored.

Modification of the mixing algorithms

In RADM 2.6, the RADM2.1 entrainment scheme is modified to allow a combination of top-down and sidewall entrainment. For precipitating clouds, the sidewall entrainment function decreases linearly from cloud base to cloud top. Entrainment in all nonprecipitating clouds is limited to the sidewall only.

For precipitating clouds, the RADM2.1 mixing equations have been modified to account for the portion of air that has been entrained through the cloud sidewall. Additionally, entrainment of air from above the top of the cloud, and also the adjustment of concentrations above cloud top, has been restricted to the layer immediately above the top of the cloud. This removes the problem associated with the RADM2.1 version that mixed mass up to the top of the model.

For nonprecipitating clouds of both types (CNP and FW), a simple direct exchange mixing mechanism has been implemented. Mixing is accounted for by the direct exchange of mass between the layer(s) below cloud base, the convective portion of the cloud, and the surrounding environment of the cloud. The mixing algorithm we use does not simulate the dynamics of air flow within the cloud field but parameterizes the net effect of the displacement of environmental air by the convective plume and the subsidence of that air back into layers below the cloud.

The concentration of pollutants in the cloud following mixing, $C_f(k)$, is a linear combination of the initial average concentration below cloud base, \bar{C}_{bi} , and the initial concentrations outside of the cloud, $C_{ei}(k)$:

$$C_f(k) = \bar{C}_{bi}[1 - F(k)] + C_{ei}(k)F(k). \quad (13)$$

The final average concentration of pollutants below cloud base, \bar{C}_{bf} , is the result of a loss of air (due to convection) at the initial mixing ratio and the replacement of an equivalent mass of cloud air with the variety of mixing ratios and densities characteristic of the cloud layer:

$$\bar{C}_{bf} = \bar{C}_{bi} - \frac{A_c}{A_i} \frac{\sum_{k_b}^{k_t} \rho(k) [1 - F(k)] [\bar{C}_{bi} - C(k)]}{\sum_1^{k_b-1} \rho(k)}. \quad (14)$$

In this equation, A_c/A_i is the fractional coverage of the cloud. It can be easily shown that the mass flux constraint prevents any one cloud from venting more than 50% of the pollutant mass available in the layer(s) below cloud base, or 95% when precipitating and CNP clouds co-exist. In this latter case, when the fractional coverages approach their maxima, the scheme essentially produces a "swap" of concentrations between the cloudy volume and the below-cloud layers.

For cases in which there is more than one model layer below cloud base, the final concentrations in the layers below cloud base are determined according to the ratio of the final to initial average concentrations in those layers:

$$C_{bf}(k) = \frac{\bar{C}_{bf}}{\bar{C}_{bi}} C_{bi}(k). \quad (15)$$

CHECKING THE CORRECTION

The corrections to the RADM2.1 cloud module were checked during their development against a data set that was totally independent of the EMEFS data period, the aggregation data set of annual average SO_2 and SO_4^{2-} concentrations. This procedure was followed to ensure that the retest of the revised model against the EMEFS evaluation period would be a truly quasi-independent retest and *not* simply a verification of a modification made using those data. The aggregation data set is derived from a stratified, random sample of 30 3-day meteorological periods from 1982 to 1985 used to represent average acidic deposition for that 4-year period. Cluster analysis was used to classify the wind-flow patterns for 1982-1985 into 19 sampling groups, or strata (Samson *et al.*, 1990). Observed and predicted concentrations for each meteorological episode are weighted according to the strata sampling frequencies to form annual averages (see Dennis *et al.*, 1990b, for a more complete description). Aggregation of predictions to annual averages gave the predictions a temporal resolution equivalent to the annual climatology of the observed cloud coverage.

We specifically chose the aggregation data set not only because it could be used to produce an annual average but also because it covers a period outside the EMEFS model evaluation period. Thus, we could return to the EMEFS data for a somewhat independent recheck of the modified model, since we had not used those data in developing the modifications. The EMEFS data also represent a different time period and a specific season rather than an annual average.

Aggregated annual cloud climatology

After implementing the improvements noted above, we calculated the aggregated annual average to determine whether an improvement in the overall cloud climatology of the model had indeed occurred. Figure 3a shows that the RADM2.6 estimated annual average fractional cloud coverage for nonprecipitating clouds has increased to around 15% (from about 2%) in the Ohio Valley source region, and for the domain, the average has increased to between 20 and 25%. Figure 3b shows that the estimated annual average fractional coverage for all cloud types, over most of the domain, is about double that predicted by the RADM2.1 cloud module (Fig. 2b). In addition, comparison of Fig. 3b with Fig. 2c, which shows Wylie and Menzel's (1989) satellite-derived cloud climatology, indicates that, though still low, RADM 2.6 is now doing a much better job at reproducing the amount of fractional cloud coverage on an annual average basis.

RADM-predicted cloud cover for the EMEFS evaluation period

Increasing the predicted cloud volume was a major objective of the modifications made to the RADM2.1

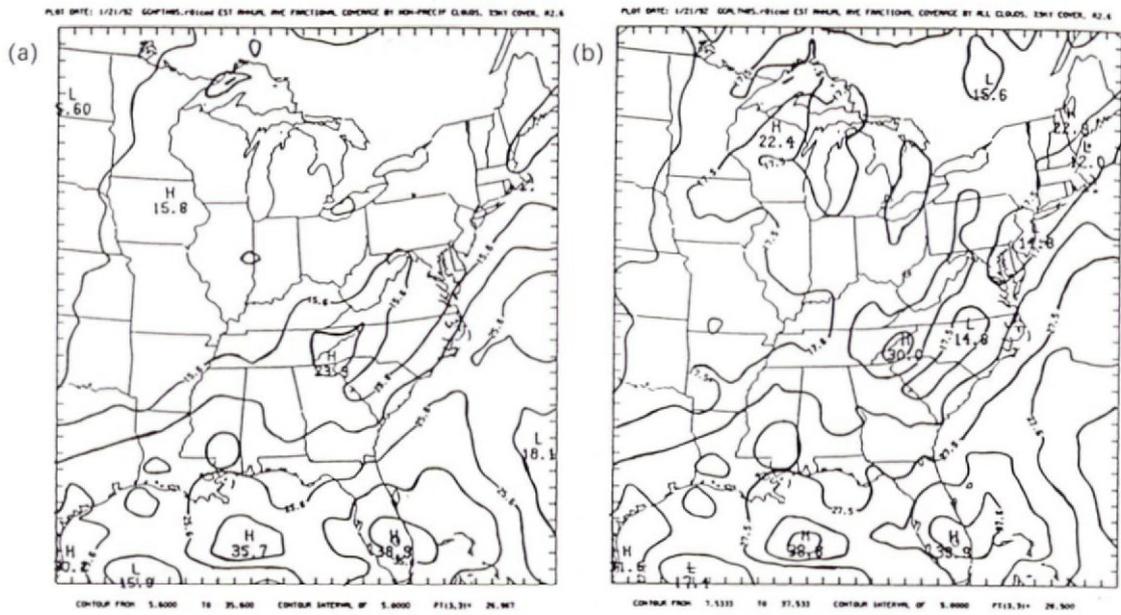


Fig. 3. Average annual fractional cloud cover for RADM2.6: (a) nonprecipitating clouds, (b) all clouds.

cloud module. Figure 3 indicates that this objective was achieved for the aggregated annual average. However, further testing was needed to show that the larger predictions of cloud coverage were temporally consistent with the observed cloud cover.

The Local Climatic Data (LCD) monthly summaries for selected weather observing stations in the U.S.A. were a convenient source of sky cover information for the EMEFS model evaluation period. The limitations of surface observed sky cover as surrogate for true cloud cover are recognized. Ideally, we should compare the observed and predicted cloud volumes. But because observational data on cloud volume for the EMEFS evaluation period are unavailable, we chose to compare the observed sky cover to the RADM-predicted cloud coverage and accept that neither represents the true cloud coverage. The comparison is intended to illustrate only that the temporal variation of the RADM2.6-predicted cloud coverage is consistent with observed cloudiness.

A set of 38 LCD stations was chosen to represent the daily variation of observed sky cover. The LCD stations were grouped according to the regional divisions found in the EMEFS data (the section below on spatial contrast provides the rationale for selecting the regions). A regional daily sky cover was calculated by averaging the reported midnight to midnight sky cover for all stations in the region. A comparable RADM-predicted regional daily cloud coverage was developed by averaging the predicted cloud coverage for the grid cells containing the LCD stations.

Figure 4 shows for two regions the time series of the regional observed sky cover and RADM-predicted cloud coverage for the period 25 August–27 September

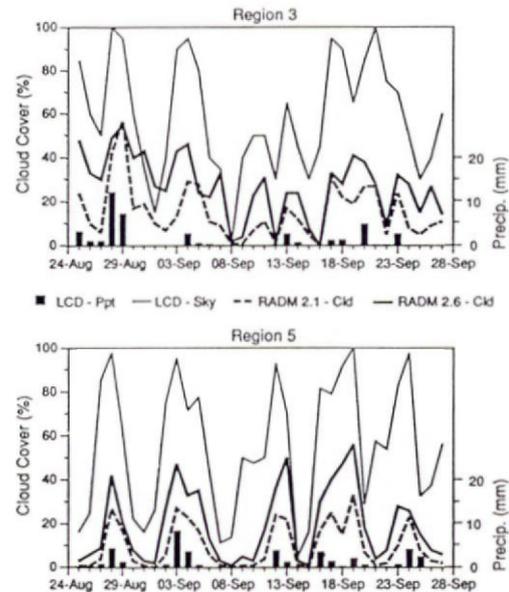


Fig. 4. Time series of RADM2.1- and RADM2.6-predicted regional average cloud coverage and regional average observed sky cover and precipitation amount from local climatic data for the 25 August–27 September 1988 period.

ber 1988. Also shown on the figure is the regional average daily precipitation amount based only on data from the LCD stations. The variations in daily sky cover indicate the passage of frontal systems through the region. The RADM2.6 total cloud amount reproduces these day-to-day changes fairly well. In a linear correlation, the regional average RADM2.6 total cloud cover explains from 14 to 64%

of the daily variation in sky cover. RADM2.6 total cloud amounts are clearly larger than those predicted by RADM2.1. In RADM2.1 most of the total cloud coverage is from precipitating clouds whereas in RADM2.6 nonprecipitating clouds dominate. The results shown in Fig. 4 and those not shown for other regions lead us to conclude that the daily variation in cloud amount and the timing of major changes in cloud coverage are being reasonably well predicted by RADM2.6.

SO₄²⁻ prediction for the aggregation data set

RADM predictions with the new nonprecipitating cloud module aggregated to annual averages were extremely encouraging. Comparisons of the SO₄²⁻ concentrations for the old and the new versions, RADM2.1 and 2.6, respectively, are shown in Fig. 5. The improved representation of the nonprecipitating cloud processes reduced the bias in the annual average SO₄²⁻ concentration from a factor of 0.6 with RADM2.1 to approximately a factor of 0.9 with RADM2.6. As expected, the overpredictions in annual SO₂ air concentrations was also reduced, especially for the lower concentrations, as shown in Fig. 6. Figure 7 shows that the RADM2.6-predicted annual

wet sulfur deposition is only very slightly increased due to the increase in SO₄²⁻ aerosol; the good agreement with observations is preserved. The bias in the RADM2.1-predicted annual average dry sulfur deposition that was noted in the NAPAP evaluation was also reduced by the RADM2.6 modifications to the nonprecipitating cloud module.

REVISED SO₄²⁻ PREDICTION FOR THE EMEFS EVALUATION PERIOD

Only after satisfactory performance on the aggregation data was achieved was the new version, RADM2.6, considered ready for further evaluation. The first step was a retest against the EMEFS period in which the problem was originally identified. An independent test of RADM2.6 against data from the EMEFS for the 25 August–27 September 1988 evaluation period was then carried out to check whether the underprediction was also eliminated for this period.

The ADOM modeling group independently had come to a similar conclusion that nonprecipitating clouds were missing in that model and needed to be

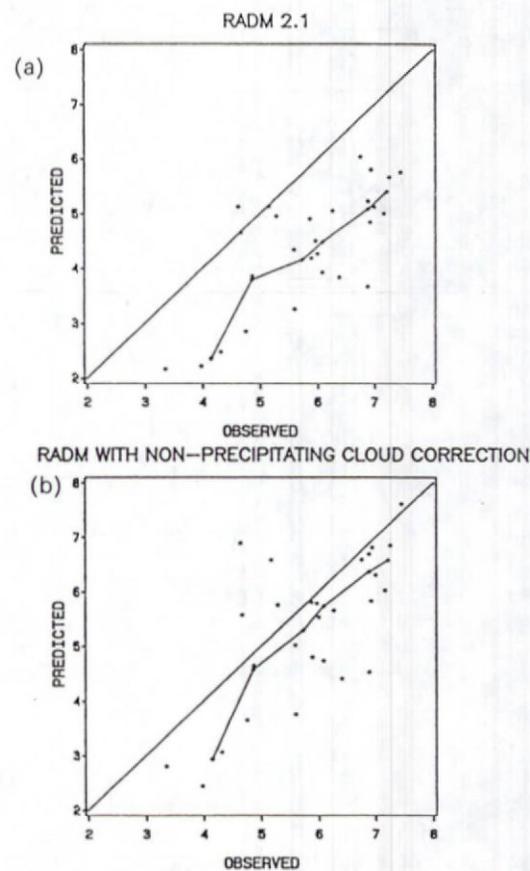


Fig. 5. Aggregated annual SO₄²⁻ concentrations (µg m⁻³): (a) RADM2.1 and (b) RADM2.6.

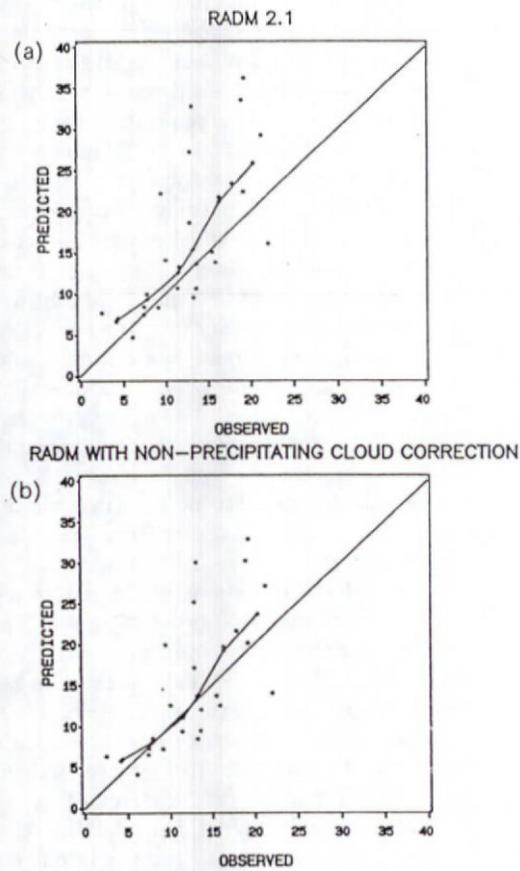


Fig. 6. Aggregated annual SO₂ air concentrations (µg m⁻³): (a) RADM2.1 and (b) RADM2.6.

COVER, 82.4
27.8
32.0
8

20
10
0
Precip. (mm)
Cd
20
0
Precip. (mm)
pre-
onal
unt
'em-
ional
y on
daily
stems
cloud
fairly
erage
64%

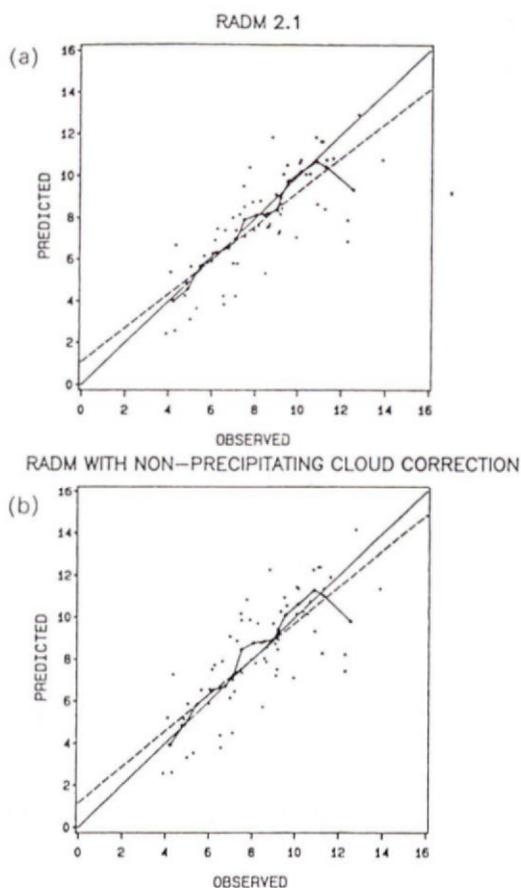


Fig. 7. Aggregated annual SO_4^{2-} wet deposition (kg ha^{-1}): (a) RADM2.1 and (b) RADM2.6.

included. They, however, followed a somewhat different path than the RADM group. The ADOM modelers focused on the SO_4^{2-} production from non-precipitating stratiform clouds (Karamchandani and Venkatram, 1991). Using EMEFS data from the 25 August–6 September 1988 subperiod, Venkatram and Karamchandani (1991) found that incorporation of nonprecipitating stratus cloud processes could eliminate the SO_4^{2-} aerosol underprediction.

When predictions from both of the new models, RADM2.6 and ADOM2Bg, were compared to measurements averaged for the EMEFS evaluation period from 25 August to 27 September 1988, the results looked quite good. As shown in Fig. 8, the original underprediction for the average over the 34-day period was eliminated for both models. In fact, for RADM the agreement was better than expected. Because our approach does not account for non-precipitating clouds associated with stratiform systems, but uses a surrogate, convective cloud, we did not expect the bias to be completely removed by these corrections. Our expectation was that a test of the new model version against the EMEFS data set would show a small bias of the order of 10%. However, there was essentially no bias in the comparison.

DIAGNOSTIC ANALYSIS OF THE RADM CORRECTION

A model evaluation needs to test the science in the models. As noted above, this means establishing through diagnostic tests that the model is producing the "right" kind of answer (the right answer for right reason and/or the wrong answer for the right reason). A model evaluation should not simply look for good comparisons of the final outcome. Therefore, we needed to be skeptical about the success of our corrections and take the next step of diagnostically testing the models to ascertain whether the non-precipitating cloud correction was the right one or was fortuitous.

To be skeptical regarding the advanced models requires more than testing the correction on an independent data set, because the corrections deal with incorporation of a new process, one that had been absent or significantly underrepresented in earlier model versions. We need to establish that replacement of our correction with an alternative process would result in comparisons that were not as good as those resulting from our proposed correction. If the comparisons turned out to be equivalently good for all corrections across all of the tests, then we could not say, with any confidence, that we had necessarily incorporated the right scientific process to correct the model error.

One alternative hypotheses was tested along with the RADM2.6 corrections to the RADM2.1 cloud module. We also focused on expanding the original set of tests to create pairs of contrasting conditions that should give the tests better discriminating power between the various hypotheses.

Development of alternative hypotheses

The RADM group focused on the role of non-precipitating, fair weather cumulus and cumulus clouds accompanying precipitating clouds. As discussed above, the testing of this correction to the model indicates it could explain the SO_4^{2-} underprediction. A second hypothesis was suggested by the regional acid deposition modeling group at the University of Cologne, Germany. They hypothesized that some rapid physical or chemical process, such as heterogeneous conversion, might be causing the ratio of SO_4^{2-} to SO_2 to increase just after emissions. Sensitivity studies by the University of Cologne group on a winter case indicated that doubling the SO_4^{2-} emissions ratio from their original 1.7 to 3% significantly reduced the SO_4^{2-} underprediction they had observed (Ebel and Hass, pers. commun., 1991). The two hypotheses examined are:

- SO_4^{2-} production by nonprecipitating cumulus clouds (RADM's approach).
- Significantly increasing primary SO_4^{2-} emissions to represent rapid, near-source, in-plume conversion (German hypothesis).

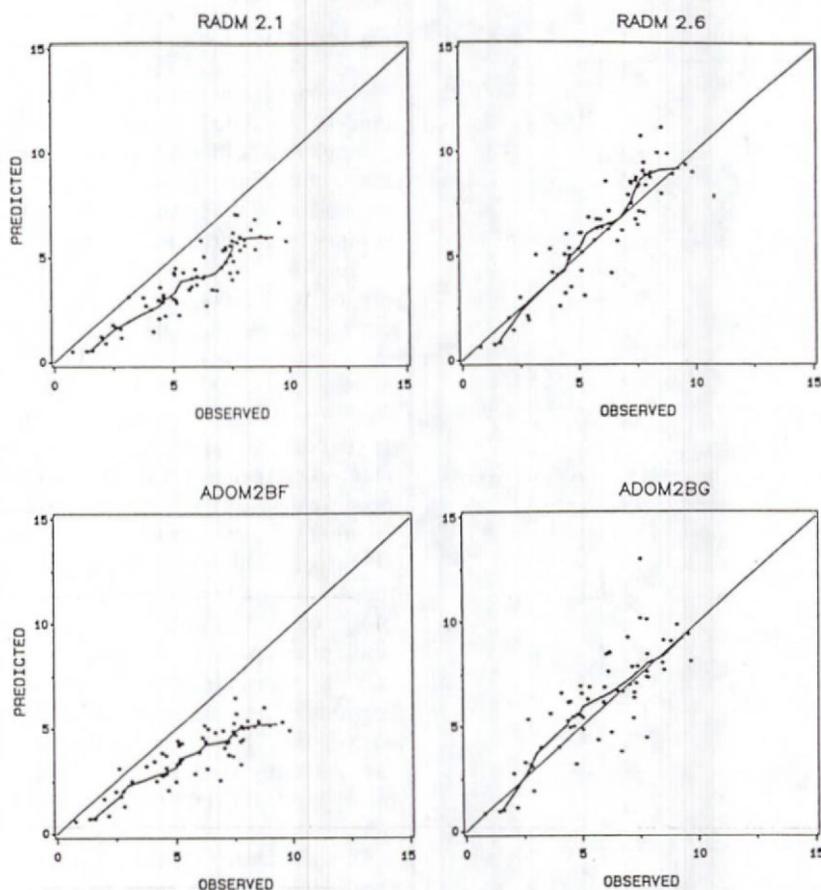


Fig. 8. Model-predicted and EMEFS-observed SO_4^{2-} air concentrations for the 25 August–27 September 1988 period: (top) RADM2.1 and RADM2.6 and (bottom) ADOM2BF and ADOM2Bg.

The NAPAP emissions inventory uses emission factors for calculating SO_4^{2-} emissions. Application of the emission factors for the eastern U.S.A. result in a range of percentage of emissions that are in the form of SO_4^{2-} from 0.7 to 4.3% with a median of 1%. Top emitting point sources, accounting for two-thirds of the total SO_2 emissions, have calculated SO_4^{2-} percentages ranging from 0.46 to 2.16% with a median of 0.64%. The emission-weighted average percentage of SO_4^{2-} is 0.7%. Analysis of aircraft experiments in the plume downwind of a single power plant show SO_4^{2-} percentages close to the stack (10–20 km) that are consistent with the NAPAP emissions factors (Sverdrup and Spicer, 1987). The range of SO_4^{2-} percentage is 0.16–1.7% with a median of 0.9%. The upper end of this range is not as high as the 3% used by Ebel and Hass. Older data from 1977 show results that include higher percentages, up to 8% (Easter *et al.*, 1980). Based on discussions with Spicer, the use of a primary SO_4^{2-} emissions ratio greater than 5% appears unjustifiable.

To establish the magnitude of a significant increase in primary SO_4^{2-} emissions for the second hypothesis we performed a budget analysis for the 25 August–27

September 1988 period. We found that directly emitted SO_4^{2-} constitutes 3–6% of the predicted surface SO_4^{2-} concentrations for RADM2.1. Elimination of the factor of 0.6 underprediction would require more than a 50% increase in ambient sulfate. Such an increase could be achieved by a factor of 10 increase in primary emitted sulfate. Independent data suggest that such an increase is unjustifiable, whereas a factor of 3 increase, bringing the percentage of primary sulfate emitted close to that used by the Germans, would be within the range of acceptability, albeit at its high end.

We chose, therefore, to split the second hypothesis into two cases: a low case with a sulfate emissions increase of a factor of 3 ($3\times$), and a high case with a sulfate emissions increase of a factor of 10 ($10\times$). We were particularly interested in seeing how the factor of 10 would fare in the diagnostic tests. This case represents a good "litmus test" regarding the power of the EMEFS data and the tests we have chosen to discriminate among hypotheses. In essence we are putting a "sample" of known character into the evaluation "system," much like a laboratory blank or spike. We know from other considerations that the

$10 \times$ case of the second hypothesis should be wrong. The question is whether comparisons between model predictions and EMEFS data would be able to show that the $10 \times$ case of the second hypothesis does not explain all aspects of the underprediction as well as the nonprecipitating cloud hypothesis or even the $3 \times$ case. If we cannot discriminate between the three cases, then our tests are very weak indeed and we cannot be sure we have the right correction without additional corroborative evidence outside the EMEFS database.

Establishing comparisons against field data

Our experience has been that single tests, by themselves, are relatively weak. Several different tests are needed to develop effective judgments regarding model skill. This is because at the spatial and temporal scales of the measurements many degrees of freedom, so to speak, are involved in production of the final outcome when several, inter-connected processes are involved. To gain the discriminating power, therefore, we needed to use several data sets that involve as much contrast as possible. We selected three data sets that emphasized the following contrasts:

- seasonal contrast
- spatial contrast
- synoptic contrast.

Seasonal contrast. Processes involved in the production of ambient SO_4^{2-} are expected to vary with the seasons. Photochemistry is at a maximum in summer and virtually ceases in winter. The relative importance of convective cloud processes also varies with season. Thus, inclusion of a seasonal contrast was deemed important to the model evaluation in general and to the testing of the different hypotheses

in particular. Selection of different 30-day time periods to provide comparison of model predictions with measurements across different seasons was already designed into the model evaluation process. The first three seasonal periods selected, in order of priority, were: late summer/early fall—end of August through September 1988 (covering the U.S. aircraft field intensive program); summer—mid-July to mid-August 1988 (covering the Canadian aircraft field intensive); and late fall/early winter—November through much of December 1988. Model predictions for only the first two periods were available for this study. The problem of SO_4^{2-} underprediction was initially noticed in comparisons with data from the first seasonal period. The second seasonal period, summer, contributes not only to the seasonal contrast test for this study, but also represents an independent data set against which to test the models.

Spatial contrast. One would expect the spatial distribution of SO_4^{2-} produced by cloud processes to differ from the distribution produced by primary emission of a majority of the SO_4^{2-} . Therefore, a spatial contrast was included in the study. The NAPAP evaluation work found that the EMEFS data were quite "noisy" across space and time. As a result, comparisons against the models would not be very stringent. Rather than impose an empirically driven model to smooth the spatially noisy EMEFS data, the model evaluation team elected to reduce the noise, to provide for more stringent testing, by forming regional groups of sites and averaging the EMEFS data over the sites in each region. These regions are shown in Fig. 9. Sites were grouped together to maintain geographical contiguity and because they had similar daily patterns of variations in trace gas concentrations associated with the passage of frontal systems (Dennis *et al.*, 1990a). The NAPAP evaluation found



Fig. 9. Centers of RADM grid cells containing EMEFS sites active during the evaluation period and the regional groupings of those cells.

that region 5 represents a source-dominated domain whereas region 3 represents a downwind domain with no major emissions. Also, the NAPAP evaluation found that transport errors appeared to be minimal for these regions (Dennis *et al.*, 1990a). These two regions were selected for tests that required spatial contrasts.

Synoptic contrast. In order to smooth out the temporal noise in the EMEFS data, the model evaluation team examined model performance for approximately 30-day averages. Much synoptic detail is averaged out in the approximately 30-day averages examined thus far for the 25 August–27 September 1988 evaluation period. Gross model error (the absolute difference between predicted and observed) decreases rapidly for successively longer averaging periods up to 5–7-day averages. Then gross error decreases only very little, if at all. This means the models are being stressed most when called on to replicate the synoptic cycles on a day-by-day basis. As noted above, the temporal "noise" in the comparisons is too great to examine each day individually. Therefore, days that were as synoptically similar as possible needed to be collected to form contrasting sets of synoptic conditions for the tests. Because the emphasis was on the influence of clouds, we defined subdivisions characteristic of sunny conditions with few to no clouds, cloudy conditions with no rain, and raining conditions.

Per cent possible sunshine, sky cover and precipitation amount observed at first-order, U.S. weather stations, as reported in the LCD monthly summary,

were used to define four classes of days. The definition of the classes is given in Table 1. The daily occurrence of these four classes of conditions for the 25 August–27 September 1988 and 19 July–5 August 1988 periods is provided in Table 2. The arithmetic average over all sites in a region is calculated for each parameter for each day. Because of the low number of occurrences in the cloudy, dry category, we did not use it in this study. We will re-examine this decision when the anticipated longer evaluation sequence, from 15 July to 30 September 1988, is available.

Results of testing the hypotheses for correcting the SO_4^{2-} underprediction

Figure 10 shows the comparisons of model predictions with EMEFS measurements for SO_4^{2-} aerosol concentrations averaged for the two seasonal periods, 19 July–6 August 1988 and 25 August–27 September 1988. The two different approaches do well for the 25 August–27 September 1988 period. The RADM approach is essentially unbiased and the 3 × and 10 × sulfate boosts span the one-to-one line. The RADM approach and the high and low sulfate boosts would be given credit for their ability to explain the earlier SO_4^{2-} underprediction. For the summer period, the RADM approach is nearly unbiased whereas the 3 × SO_4^{2-} primary emissions boost underpredicts SO_4^{2-} slightly more than the original underprediction identified for the 25 August–27 September 1988 period. The 10 × SO_4^{2-} boost results fall between the RADM

Table 1. Classification based on synoptic stratification

Attributes	Sunny, clear, dry	Dull, cloudy, dry	Dull, cloudy, wet	All other days
Sunshine (%)	≥ 70	≤ 40	≤ 40	> 40 and ≤ 70
Cloud cover	≤ 3/10	≥ 8/10	≥ 8/10	> 3/10 and ≤ 8/10
Precipitation	0 or T or ≤ 0.05 in.	0 or T or ≤ 0.05 in.	> 0.05 in.	

Table 2. Number of days by classification for the evaluation periods

Region	Sunny, clear, dry	Dull, cloudy, dry	Dull, cloudy, wet	All other days
19 July–5 August 1988				
3	2	0	3	13
5	4	0	2	12
6	1	0	5	12
7	1	0	0	17
25 August–27 September 1988				
3	5	2	6	21
5	9	1	5	19
6	7	5	5	17
7	7	2	5	20

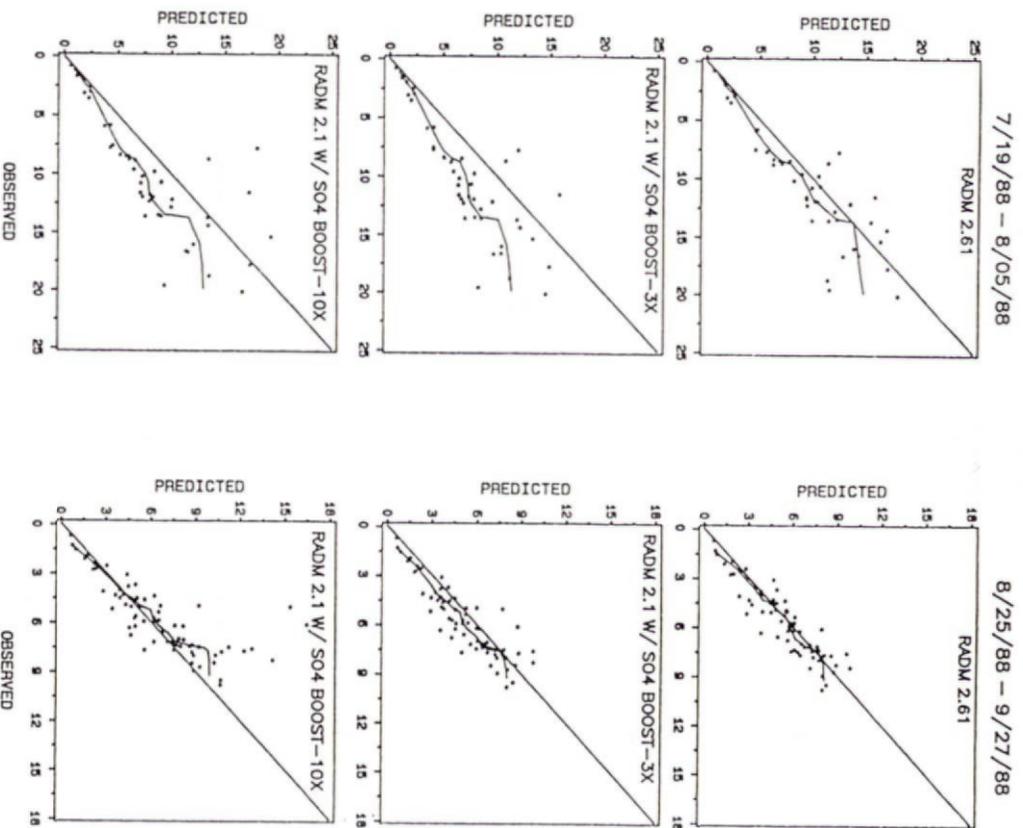


Fig. 10. Performance of the two hypotheses in simulating seasonal contrasts: (top) RADM2.6, (middle) 3 \times sulfate boost, (bottom) 10 \times sulfate boost; (left) 19 July-6 August 1988, (right) 25 August-27 September 1988.

approach and the 3 \times boost for high sulfate concentrations and equals the 3 \times boost for low sulfate concentrations. The degree of agreement is quantified in Table 3, using per cent deviation as the measure. Per cent deviation here is defined as prediction minus observed divided by observed. We conclude that the RADM approach just passes the seasonal contrast test and that the 3 \times boost does not. The 10 \times boost case gives pause, but is sufficiently biased in the mid-range concentrations to judge that it, too, does not pass.

Figure 11 shows the comparisons of model predictions with EMEFS measurements for SO_4^- aerosol concentrations averaged across the days comprising each of the two synoptic stratifications, sunny-clear-dry and dull-cloudy-wet, within the 25 August-27 September 1988 period. The comparisons for the RADM approach are reasonably good. The compar-

isons for the high and low sulfate boosts are fairly good; the 3 \times boost underpredicts the dull-cloudy-wet and the 10 \times boost overpredicts the sunny-clear-dry stratification. Although the RADM approach appears to be slightly better for the dull-cloudy-wet stratification, one would probably not reject the SO_4^- emission increase hypothesis based on its performance on this test. We conclude that the RADM approach passes the synoptic contrast test for the 25 August-27 September 1988 period and that the 3 \times and 10 \times primary sulfate emissions increases receive good marks, though less than for the RADM approach.

Both approaches have difficulty on the synoptic contrast test for the 19 July-6 August 1988 period, as shown in Fig. 12. Only one comparison out of six, the RADM2.6 approach on the dull-cloudy-wet stratification, is relatively unbiased. None of the approaches

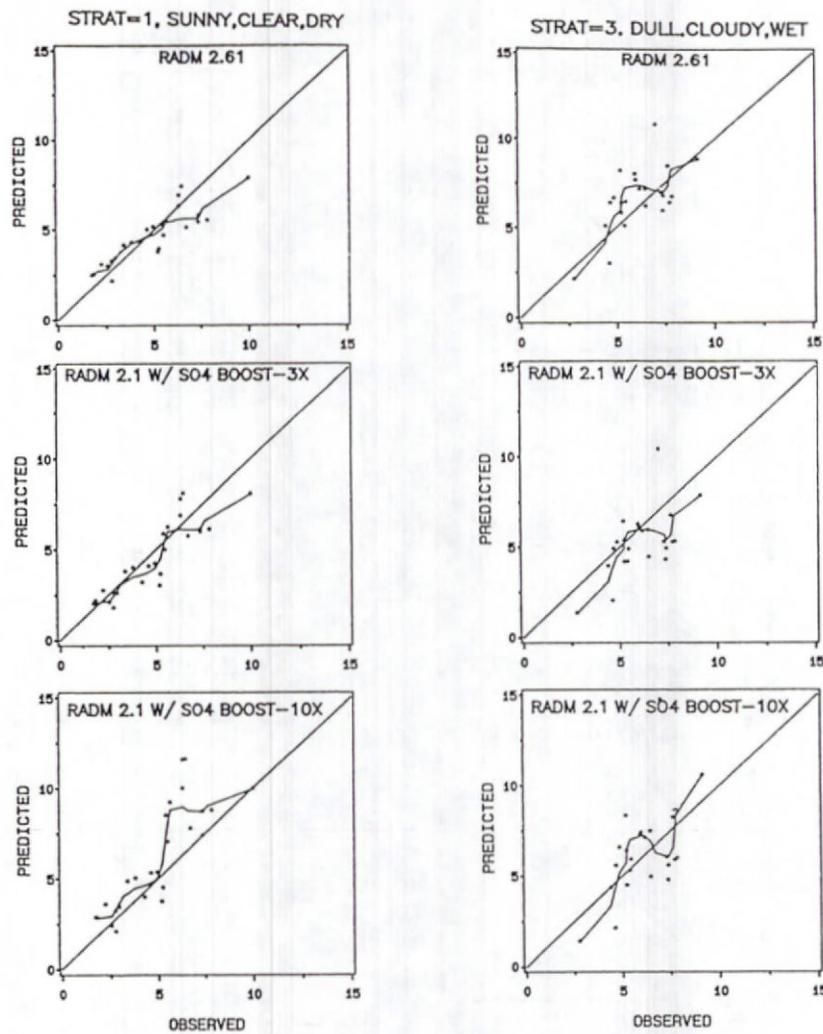


Fig. 11. Performance of the two hypotheses in simulating synoptic contrasts during the 25 August–27 September 1988 period: (top) RADM2.6, (middle) 3 × sulfate boost, (bottom) 10 × sulfate boost; (left) sunny-clear-dry, (right) dull-cloudy-wet.

Table 3. Measure of agreement for seasonal and synoptic contrast tests (per cent deviation = $100 \times [\text{pred.} - \text{obs.}] / \text{obs.}$)

Seasonal contrast	August–September	July–August
	RADM2.61	–5
RADM2.1 with 3 × boost	–11	–32
RADM2.1 with 10 × boost	10	1
Synoptic contrast	Sunny, clear, dry	Dull, cloudy, wet
	August–September	
RADM2.61	–6	9
RADM2.1 with 3 × boost	–8	–14
RADM2.1 with 10 × boost	25	4
July–August		
RADM2.61	–40	7
RADM2.1 with 3 × boost	–45	–38
RADM2.1 with 10 × boost	–34	–32

airly
dry-
ear-
ap-
-wet
the
per-
DM
e 25
3 ×
eive
ap-
ptic
l, as
the
itifi-
ches

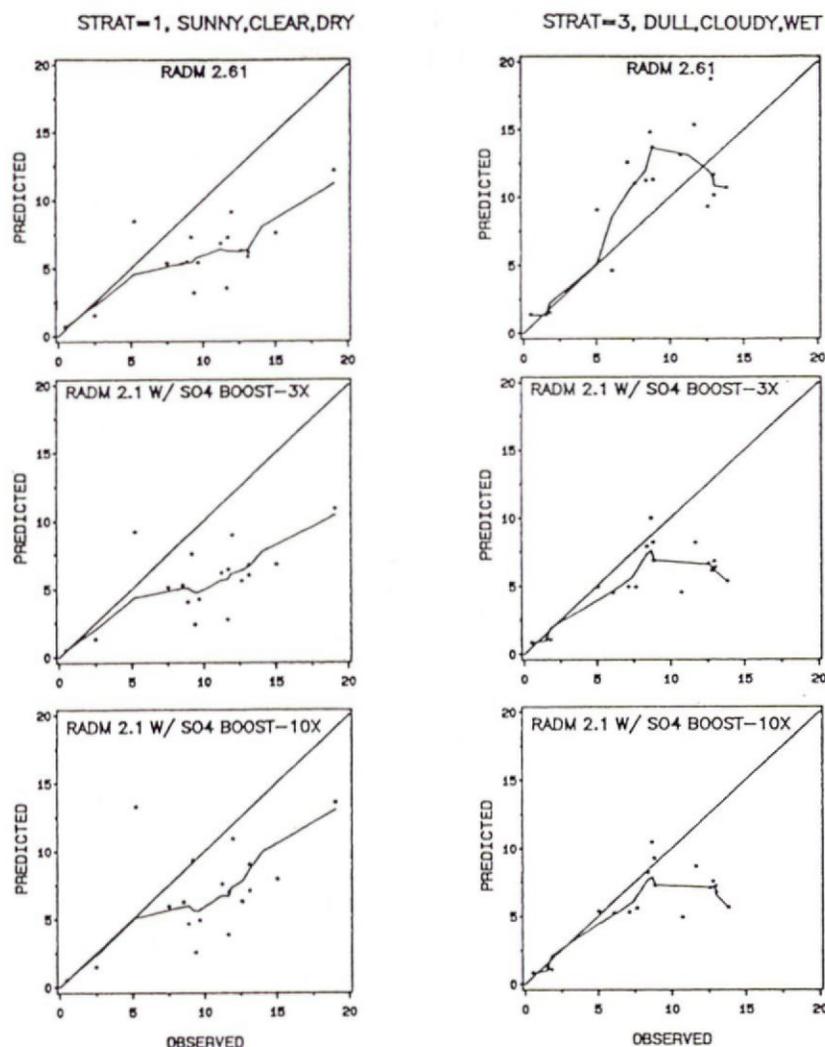


Fig. 12. Performance of the two hypotheses in simulating synoptic contrasts during the 19 July-6 August 1988 period: (top) RADM2.6, (middle) 3 \times sulfate boost, (bottom) 10 \times sulfate boost; (left) sunny-clear-dry, (right) dull-cloudy-wet.

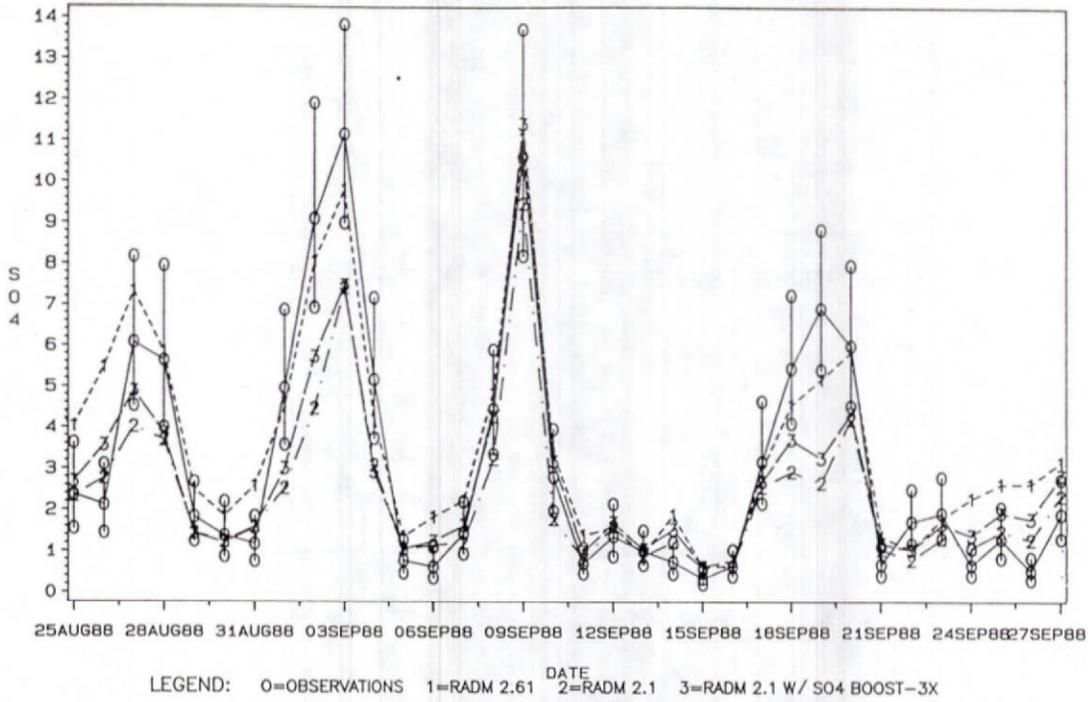
do well for the sunny-clear-dry stratification. We judge that all approaches fail this test, although the RADM2.6 approach did produce a relatively unbiased comparison for one of the meteorological stratifications.

Time series of daily SO_4^{2-} for two regions were used for the spatial contrast tests. The contrast was between source regions, represented here by region 5, and downwind regions, represented here by region 3. Figures 13 and 14 show the time series for the 25 August-25 September 1988 period for RADM and the 3 \times primary sulfate emissions increase and RADM and the 10 \times sulfate boost, respectively. As shown in Fig. 13, the RADM2.6 approach and trebling of the SO_4^{2-} emissions both significantly improve the agreement in the source regions. On two of the days RADM2.6 performs better and on two different days the 3 \times boost performs better. In Fig. 14 the 10 \times boost performs better than the RADM2.6 ap-

proach on some days and then overdoes it (overprediction) on other days for the source region. For the downwind region the RADM2.6 approach significantly improves the agreement for the cycle of four peaks in sulfate concentrations. The 3 \times boost shows little improvement and the 10 \times boost shows fair improvement over RADM2.1 predictions, the original model, in the downwind region. The degree of agreement is more difficult to quantify for the time series than for the paired comparisons. We have chosen to use gross error as a measure of agreement and present it in Table 4 for the time-series comparisons. Gross error is the mean for the period of the absolute value of the daily predictions minus observations.

Figure 15 shows the time series for the 19 July-6 August 1988 period for RADM and the 10 \times sulfate primary emissions boost. The 3 \times boost is hardly different from the RADM2.1 predictions shown in Fig. 15. The RADM2.6 approach gives some im-

S04 CONCENTRATION (UG/M3)
 MODEL COMPARISON FOR 8/25/88-9/27/88
 AVERAGE OF PREDICTIONS AND OBSERVATIONS
 REGION-3



S04 CONCENTRATION (UG/M3)
 MODEL COMPARISON FOR 8/25/88-9/27/88
 AVERAGE OF PREDICTIONS AND OBSERVATIONS
 REGION-5

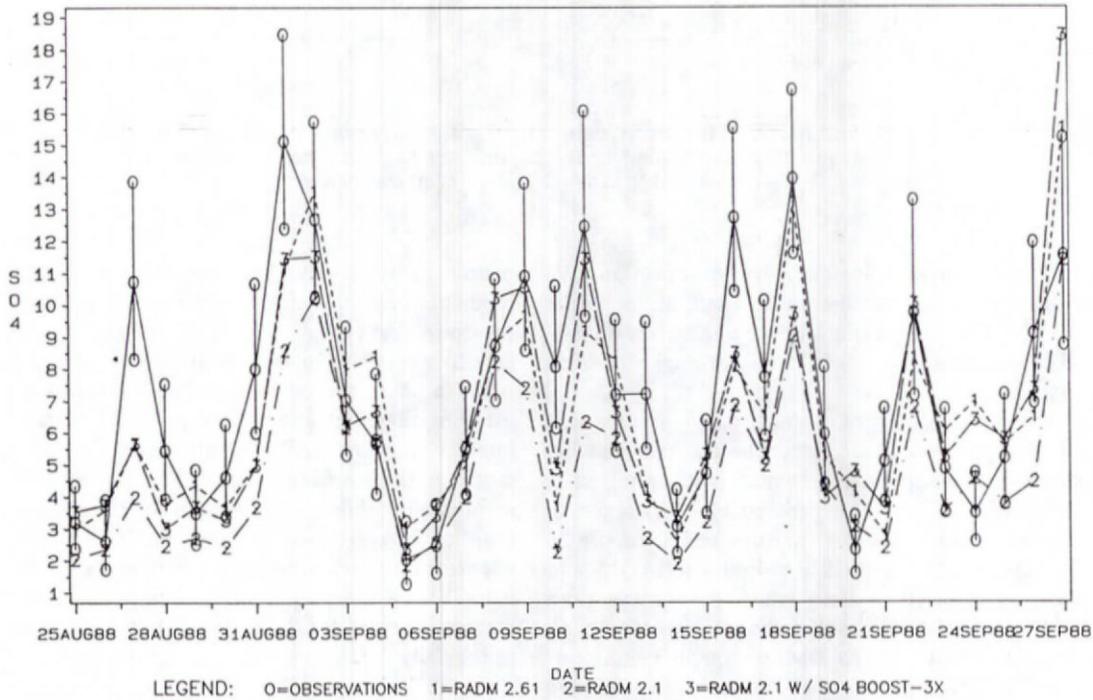


Fig. 13. Time series of RADM2.6-, RADM2.1- and RADM2.1 with 3 × sulfate boost-predicted and EMEFS-observed daily SO₄²⁻ aerosol concentrations for the 25 August-27 September 1988 period: (top) region 3, downwind; (bottom) region 5, source.