United States
Environmental Protection
Agency

Atmospheric Research and
Exposure Assessment Laboratory
Research Triangle Park NC 27711

Research and Development

EPA/600/S3-90/051 Sept. 1990

EPA

Project Summary

The Across North America Tracer Experiment (ANATEX) Model Evaluation Study

Terry L. Clark and Richard D. Cohn

During the first three months of 1987, three perfluorocarbon tracer gases were released at 2.5-day or 5.0-day intervals from two sites in central North America (Glasgow, Montana and St. Cloud, Minnesota) and sampled for 24-h periods at 77 surface sites. The source-receptor distances ranged from less than 30 km to 3,000 km. These Across North America Tracer Experiment (ANATEX) data serve as a unique evaluation data set with which to evaluate the long-range transport and diffusion simulations of acid deposition models and to establish a range of uncertainty for various model genres. The performances of three single-layer Lagrangian, six multiple-layer Lagrangian, and two multiple-layer Eulerian models are assessed using quantifiable measures based on comparisons of ensemble mean concentrations and plume widths as well as trajectory errors expressed as a function of transport time.

This Project Summary was developed by EPA's Atmospheric Research and Exposure Assessment Laboratory, Research Triangle Park, NC, to announce key findings of the research project that is fully documented in a separate report of the same title (see Project Report ordering information at back).

Introduction

The U.S. Environmental Protection Agency, the National Oceanic and Atmospheric Administration, and the U.S. Air Force have completed an evaluation of 11 operational models to assess the performances of simple and state-of-the-science, long-range transport and diffusion models. The model calculations were compared to observations of surface concentration data compiled during the Across North America Tracer Experiment (ANATEX).

During the first three months of 1987, three perfluorocarbon tracer gases were released at 2.5-day or 5.0-day intervals from two sites in central North America (Glasgow, Montana and St. Cloud, Minnesota) and sampled for 24-h periods at 77 surface sites. The source-receptor distances ranged from less than 30 km to 3,000 km. These ANATEX data serve as a unique evaluation data set with which to evaluate the long- range transport and diffusion simulations of acid deposition models and to establish a range of uncertainty for various model genres. The performances of three single-layer Lagrangian models (SRL, TCAL, and VCAL), six multiple- layer Lagrangian models (ARL, BAT, GAMUT, HY-SPLIT, MLAM-FINE, and MLAM- COARSE), and two multiple-layer Eulerian models (ADPIC and ADOM) are assessed using quantifiable measures based on comparisons of ensemble mean concentrations and plume widths as well as trajectory errors expressed as a function of transport time.

Before the distribution of the ANATEX data, modelers applied their models in a "blind" applications mode using required meteorological input data and the actual periodic 3-hour (h) ANATEX tracer emission rates according to the prescribed schedule during the first 3 months of 1987. Some of these models were very similar to others in this study; the only differences were variations of the modeling assumptions or the selection of modeling options.

Model performance measures were developed on the basis of the features of the surface sampling network and the sampling protocol. These measures were quantified using either ensemble concentration means or relative distances of the centroids of tracer "footprints"--composite tracer plumes defined by the 24-h-mean measurements.

Several aspects of the evaluation study are discussed in this report. First, the performances of the three genres of models--(1) single-layer Lagrangian, (2) multiple-layer Lagrangian, and (3) multiple-layer Eulerian--are compared to each other to relate model performance to model approach. Secondly, the performances of the various model versions are related to the differences in the modeling codes to relate model performance to model assumptions/options. Thirdly, model performance is related to three meteorological scenarios to relate model performance to the degree of complexity of the air flow.

Objectives

The objectives of this model evaluation are fourfold:

(1) to assess the overall performance, as well as the model errors on temporal scales of 24 h, of prognostic long-range atmospheric models with respect to transport and diffusion as a function of transport time and distance,
(2) to intercompare the model performances and relate performance to fundamental modeling approaches,
(3) to identify the periods and associated meteorological conditions when each model performed best and worst, and
(4) to compare and contrast the AMES conclusions with those of similar studies using CAPTEX data.

Model Performance Summaries

The model performance assessment was based on seven performance measures and charts using either both halves of the entire data set or a subset of this data set relating concentrations to specific tracer releases. The performance measures summarized here are box plot distributions, frequency distributions, mean concentrations as a function of transport distance, mean lateral diffusion, footprint transport speed and location errors, and mean trajectory errors.

Single-Layer Lagrangian (SLL) Models

Box Plot Distributions

Each of the three models of this genre (SRL, TCAL, and VCAL) revealed a tendency to overestimate the frequencies of higher concentrations. During the first half-period, the medians and third quartiles of each model were 2-to-6 times greater than those of the measurements. For the second half-period, the same was true for SRL only; TCAL and VCAL values were much closer (i.e., within a factor of 2) to those of the measurements.

Frequency Distributions

During the first half-period the SLL models generally approximated the frequency of concentrations above the thresholds (5 dfL/L and 8 dfL/L for PTCH and PDCH, respectively). However, SRL overestimated the frequencies of concentrations exceeding 99 dfL/L by approximately a factor of 4. TCAL, and to a lesser degree, VCAL closely approximated the distributions. During the second half-period SRL greatly underestimated the frequency of PTCH concentrations above the threshold (<1% versus 18%), as well as the frequency of nonzero concentrations (1% versus 36%). Meanwhile, the percentage of sites with nonzero concentrations calculated by both TCAL and VCAL was much greater than that for the measurements of both tracers (55% to 80% versus approximately 40%).

Mean Concentrations as a Function of Transport Distance

For transport distances ranging from 300 to 2,300 km, the mean TCAL and VCAL concentrations during both half-periods and the mean SRL concentrations during the first half-period along several bands of sites tended to be higher than those of the other models, as well as the measurements. Deviations of factors of 2 and 3 from the measured means were common; some SRL deviations were as great as a factor of 6.

Mean Lateral Diffusion

The comparison of the model and measured mean plume widths showed an inconsistency between half-periods. For the first half-period, the model mean plume widths were generally underestimated, but within an average of 20% of the measured width for TCAL, 30% for VCAL, and 50% for TCAL. For the second half-period, TCAL and VCAL mean plume widths were greater than the measured mean plume widths especially for PTCH, by as much as 130%. SRL mean plume widths for this period (less than 250 km) were much less than the actual widths--in fact, nearly zero--indicating a serious problem with the diffusion.

Footprint Transport Speeds and Centroid Locations

Of these three models, VCAL performance was clearly best in calculating the transport speeds and centroid locations of tracer footprints. Although VCAL, as well as TCAL, demonstrated a tendency to overestimate the transport speeds (10% and 40 % for VCAL PTCH and PDCH, respectively; 10% and 180% for TCAL PTCH and PDCH, respectively), the VCAL overestimates exceeded a factor of 2 for only 6 footprint-days, compared to 24 footprint-days for TCAL. In addition, VCAL showed little bias in placing its PTCH and PDCH footprint centroids (mean ratios +20% D m). TCAL was less consistent, showing no bias for PTCH footprints, but a large positive bias (i.e., its centroids generally were to the south of the measured centroids) for PDCH footprints. TCAL also tended to place its centroids to the right of the actual centroids when the transport speeds were overestimated. SRL tended to both overestimate transport speeds by +40% and place the footprint centroids to the right of the actual centroids.

Mean Trajectory Errors

The mean centroid location errors of SRL and TCAL were greater than any other model. These errors increased linearly with transport time from approximately 350 km to 800 km after 13.5 h and 61.5 h, respectively. On the other hand, the mean centroid location errors for VCAL were among the least, half those of SRL and TCAL.

Multiple-Layer Lagrangian (MLL) Models

Box Plot Distributions

In general, with the exception of HY-SPLIT, the means, medians, and third quartiles of the MLL models more closely corresponded to those of the measurements than did those of the single-layer Lagrangian models. This was especially true for the PTCH concentrations, where the means, medians, and third quartiles were within ±50% of those for the measurements. Those for the PDCH concentrations were generally greater than those for the measurements by a factor of 2. HY-SPLIT means, medians, and third quartiles tended to be greater than those of the other MLL models, greater than those of the measurements by as much as factors of 3 and 4, and more closely resembled those of the SLL models. During the first half-period, HY-SPLIT medians were comparable to the third quartiles of the measurements.

Frequency Distributions

The frequency distributions of the MLL models generally corresponded more favorably to those of the measurements than did the SLL models. With the exceptions of BAT, MLAM-FINE, and MLAM-COARSE, the MLL model frequencies of concentrations above the thresholds approximated those of the measurements; MLAM-FINE and MLAM-COARSE frequencies tended to be higher by a factor of 2. HY-SPLIT concentrations and MLAM-FINE and MLAM- COARSE PDCH concentrations exceeding 99 dfL/L during the first half-period occurred at least twice as often as those measured; the opposite was true for ARL PTCH concentrations. ARL and GAMUT distributions of PDCH concentrations were virtually identical to those of the measurements.

Mean Concentrations as a Function of Transport Distance

Although the mean concentrations along the 300-m and 1,000-m bands for the MLL models tended to be lower than the actual means by factors of 2 to 4, the mean concentrations along the bands for the MLL models tended to more closely resemble those of the measurements than did the SLL models. This was especially true at distances farther downwind of the release sites, where the means of all MLL models but MLAM-COARSE were within ±2 dfL/L of the actual means; MLAM-COARSE means generally were high by a factor of 2 to 3 at all distances.

Mean Lateral Diffusion

During the first half-period the mean dispersions of MLAM-FINE and its sibling, MLAM-COARSE, were within ±30% of the actual dispersion. GAMUT and HY-SPLIT dispersions were the lowest of all models, factors of 2 to 3 lower than the actual dispersions. The dispersions for the other MLL models were 50% to 100% lower than the actual dispersion. During the second half-period both BAT and, once again, HY-SPLIT dispersions were the lowest of any model (lower than the actual dispersions by factors of 2 to 3). MLAM-COARSE dispersions were very high, generally double those of the measurements. The dispersions of the remaining MLL models were lower than the actual, but within 30%. For both half-periods, BAT and HY-SPLIT plume widths showed virtually no change with transport distance.

Footprint Transport Speeds and Centroid Locations

Each of the MLL models demonstrated skill in simulating the tracer transport. With two exceptions (i.e., BAT and HY-SPLIT PDCH footprints), mean transport speeds were within an average of 30% of actual mean speeds and mean relative location errors were within 30% of the actual transport distance. Furthermore, only VCAL--a SLL model--had a lower mean absolute location error than the MLL with the greatest error (GAMUT: 417 km).

The performances of half of the MLL models in simulating footprint transport speeds and centroid locations did not vary for each tracer. Performances for the three exceptions--ARL, BAT, and HY-SPLIT--were better for the PTCH data, partly because the PTCH data set tended to be dominated by simpler, northwest flows, while the PDCH data set included a wider range of flow patterns. For example, BAT showed the lowest speed and location errors and no significant biases for the PTCH footprints; however, for the PDCH footprints, BAT speeds strongly tended to be lower than actual speeds and its centroids tended to be to the right of the actual centroids. MLAM-FINE and MLAM-COARSE speeds tended to be lower while GAMUT speeds tended to be greater, but for each model more so for the PDCH footprints. Although MLAM-COARSE showed minimal location errors for both PTCH and PDCH sets, MLAM-FINE locations tended to be to the right of the actual locations for both tracers while GAMUT locations tended to be to the left of the actual PTCH footprints. Both ARL and HY-SPLIT showed the most scatter for speeds and location errors; speeds tended to be greater than the actual speeds. ARL centroid locations for the PDCH footprints tended to be to the left of the actual centroids.

Mean Trajectory Errors

The six MLL models were divided into two types of behavior. BAT, MLAM-FINE, and MLAM-COARSE mean errors peaked to 400 km after 3.5 days of transport and showed no significant additional increase beyond that. On the other hand, ARL, GAMUT, and HY-SPLIT errors increased more sharply with transport time, reaching 530 ±30 km after 2.5 days and 920 ±160 km after 3.5 days, or about 3 times greater than the other MLL models. Errors decreased to 750 ±100 km after 4.5 days.

Multiple-Layer Eulerian (MLE) Models

Box Plot Distributions

The correspondence between model and measurement distributions varied with tracer. The ADOM median and third quartile for PTCH concentrations were 30% to 60% less than those of the measurements, while the opposite was true for the PDCH concentrations. The PTCH box plot distribution for ADPIC was virtually identical to those of the measurements; the ADPIC median and third quartile for PDCH concentrations were within 60% of those for the measurements.

Frequency Distributions

The ADOM distributions for the first half-period corresponded very closely to those of the measurements; ADOM was not applied for the second half-period. The comparisons of the ADPIC with the actual distributions were inconsistent; while ADPIC distributions were quite similar to the actual distributions for PTCH during the first half-period and PDCH concentrations for the second half-period, the ADPIC frequencies for concentrations exceeding the thresholds deviated by 75% for the remaining half-periods for each tracer.

Mean Concentrations as a Function of Transport Distance

The mean PTCH concentrations for three bands of sites for both ADOM and ADPIC tended to be lower than those of the measurements. Ratios of calculated-to-predicted means for all three bands ranged from 0.4 to 0.9. However, the opposite was true for first-period PDCH concentrations; these ratios approximated 1.8 for both ADOM and ADPIC. Second-period ADPIC mean PDCH concentrations were lower by an average of 30%.

Mean Lateral Diffusion

During the first half-period the lateral dispersion of both models was generally greatest of all the models and greater than the actual dispersion an average of 10% for ADOM and 50% for ADPIC. During the second half-period, ADPIC dispersion for PTCH was low by factors of 2 to 3 for some bands and varied little with transport distance; however, its PDCH dispersion was within ±10% of the actual dispersion at all distances.

Footprint Transport Speeds and Centroid Locations

ADOM speeds for both tracers tended to be lower than actual speeds by 20% to 40%. For its largest location errors, ADOM speeds were greater than actual speeds by at least a factor of 2 and its centroids tended to be to the left of the actual centroids. However, in general, its centroids tended to be to the right. Similarly, ADPIC speeds tended to understate the actual speeds, sometimes by as much as 60% and 80%; ADPIC centroids for the PDCH footprints were to the right of the actual centroids in nearly every case and to the left for the PTCH footprints.

Mean Trajectory Errors

Both ADOM and ADPIC mean errors were among the greatest of all models. The ADOM mean error after 1.5 days was low, approximately 200 km, but quickly increased to 600 km after 2.5 days (greatest of all models), then decreased to 500 km after 3.5 days. Similarly, the ADPIC mean error increased sharply from 300 km at 1.5 days to 850 km after 2.5 days, among the greatest.

Conclusions

The limitations of the ANATEX data (e.g., virtually no vertical tracer distributions beyond 300 km of the release sites, the spacing between sites, and 24-h integrated sampling at surface sites), limited the scope of the model evaluation study. Firstly, evaluations based on point-to-point comparisons of simultaneous tracer concentrations were not practical. Secondly, model errors could not be related to specific model processes. Consequently, this model evaluation study focused on identifying model biases for whatever reason. When appropriate, possible problems with modeled processes were offered as explanations for the observed biases. However, a more resolved data base and additional model applications are required to reveal the actual causes of these errors.

Single-Layer Lagrangian (SLL) Models

The SRL transport vectors, based on surface pressure gradients, clearly are biased: speeds tend to be overestimated and directions tend to be to the right of the actual vectors. This bias, as well as its direction, is not surprising given the nature of geostrophic wind vectors. The high bias in the transport speed can explain SRL's tendency of overestimating the mean concentrations and the frequency of high concentrations, as well as underestimating the number of sites with nonzero concentrations and the lateral diffusion. That is, for a model that overestimates transport speed, the plumes will be narrower and the concentrations will tend to be greater for fixed distances and transport times.

The definition of the height of the mixed layer--the only difference between TCAL (fixed height at 1,500 m AGL) and VCAL (variable height based on potential temperature profiles)--has a large influence on the performance of a single-layer model. This underscores the need to choose carefully the layer through which wind vectors are to be calculated. The low mean centroid location errors for VCAL indicates that a single-layer model can perform as well as the models of the other genres.

The general tendency of TCAL, and to a lesser degree VCAL, to overestimate the transport speeds can explain their tendencies to overestimate the mean concentrations, as a consequence of slower tracer diffusion relative to transport distance. Since wind speeds generally increase with height, the higher TCAL transport speeds could have resulted from a mixed layer height that was too high; climatological data suggest that the 1,500-m height is a factor of 2 too high.

Multiple-Layer Lagrangian (MLL) Models

The MLL models clearly outperformed all others except VCAL in simulating the transport of tracer footprints. With the possible exceptions of ARL and GAMUT, none of the MLL models clearly outperformed any other in its genre. For most of these models, the majority of the performance measures indicated relatively good performance, but the remaining performance measures indicated biases in the model results. For instance, ARL showed little if any bias in its results as did GAMUT (with the exception of its high bias in the transport speeds), but the mean location errors were relatively great. This indicated that although their mean errors were rather substantial, their centroids, in general, were to the left of the actual centroids as often as they were to the right. The relatively good comparison of its distribution statistics with those of the measurements appears to indicate that ARL and GAMUT simulated rather well the lateral/vertical diffusion; however, they both appeared weak in simulating the transport.

BAT's underestimates of the mean lateral diffusion and the frequencies of occurrence of concentrations above the thresholds would appear to be related to each other. That is, a model that underestimated plume widths will show fewer cases of nonzero concentrations and concentrations above the thresholds. The tendency to overstate the PDCH footprint transport speeds and to place its PDCH footprint centroids to the right of the actual centroids indicated that BAT's vertical mixing could be overstated, effectively giving more influence to the higher-altitude winds, which tend to have greater speeds and directions to the right of the lower-altitude winds. The very low mean centroid location errors, however, indicate that BAT simulates well the transport.

HY-SPLIT's tendencies to understate the plume widths and overstate its third quartiles could be symptomatic of its algorithm for calculating atmospheric stability from the NGM results, as opposed to the algorithm of ARL (its sibling), which interpolates surface and rawinsonde data. Like ARL, HY-SPLIT mean centroid location errors were relatively great, yet no substantial biases were evident in its calculation of footprint speeds and locations. The relatively high turbulent K_z profiles used by HY-SPLIT may have exaggerated the vertical diffusion, thereby adversely affecting its performance.

MLAM-FINE's footprint transport speeds tended to be lower and to the right of the actual speeds and locations for the first half-period, the only period for which it was applied. These slower speeds could explain its other tendency to overestimate the frequency of concentrations exceeding the thresholds. That is, its footprint widths would be wider relative to transport distance. Furthermore, at any one site the concentrations from one release can be nonzero for two days rather than one day. For both half-periods, MLAM-COARSE showed the same tendencies as MLAM-

FINE. In addition, MLAM-COARSE tended to overstate the frequency of concentrations exceeding the thresholds as well as the mean concentrations at all distance downwind of the release sites. Especially during the second half-period, MLAM-COARSE plume widths tended to be greater than those of the actual widths. All but one of these biases could be explained by the slower MLAM-COARSE transport speeds; the high bias in the mean concentrations could be symptomatic of a low bias in the vertical mixing, causing concentrations near the surface to be biased high.

Multiple-Layer Eulerian (MLE) Models

In general, the MLE models performed quite similarly and better for the ensemble measures than they did for the footprint comparison measures. This implies that these two models performed relatively well for the average, but performed relatively poorly for individual cases. The only substantial bias observed in the ensemble measures was for lateral diffusion; both ADOM and ADPIC tended to overstate the footprint widths in the first half-period, the only period for which ADOM was applied. However, both models tended to understate the footprint speeds, which could by itself explain the high bias in lateral diffusion, as well as the large mean centroid location errors.

The strong relationship between the large ADOM centroid location errors and overestimated transport speed errors could indicate a problem with its vertical diffusion for several cases (PTCH-15, PDCH-4, -10, and -15), all of which were intercepted by cyclones or fronts. That is, the model could have overestimated vertical diffusion and, as a consequence, relied more on the faster wind speeds at higher levels.

The strong ADPIC tendency to understate the transport speeds and place footprint centroids to the right of the actual footprint centroids demonstrates its weakness in simulating the transport for the 27 footprints of this study. Additional data are needed to substantiate this conclusion. The reason for relatively few ADPIC footprints related to the fact that ADPIC concentrations very often did not return to zero days after actual tracer footprints were transported across regions.

Richard D. Cohn is with Analytical Sciences, Inc., Durham, NC 27713.
***Terry L. Clark**, the EPA author, is also the Project Officer, (see below).*

The complete report, entitled "The Across North America Tracer Experiment (ANATEX) Model Evaluation Study," (Order No. PB-90-261-454AS; Cost: $23.00 subject to change) will be available only from:

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Telephone: 703-487-4650

The EPA Project Officer can be contacted at:

Atmospheric Research and Exposure Assessment Laboratory
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

United States
Environmental Protection
Agency

Center for Environmental Research
Information
Cincinnati OH 45268

Official Business
Penalty for Private Use $300

EPA/600/S3-90/051