United States
Environmental Protection
Agency

Atmospheric Sciences Research
Laboratory
Research Triangle Park NC 27711

Research and Development

EPA/600/S3-85/067 Dec. 1985

EPA

Project Summary

EPA Regional Oxidant Model: Description and Evaluation Plan

K. L. Schere and A. J. Fabrick

The USEPA Regional Oxidant Model and NEROS data base are described. The model incorporates a comprehensive description of the physical and chemical processes thought to be important to tropospheric O_3 production on 1000 km scales. The data base employed for the first application of the ROM was collected during the summers of 1979 and 1980 in the Northeast U.S. It contains meteorological and air quality data from regular monitoring networks and from enhanced networks or special field project measurements made during that period.

The evaluation procedure that will be used to determine the ROM performance on this data base is outlined. A number of episodes will be simulated from the period July 23 through August 16, 1980, for which performance statistics will be developed. The evaluation of any given day within an episode will proceed in two distinct stages. The first state will focus on model performance for an individual model realization, irrespective of all other realizations. Model realizations for a given day are functions of the possible flow fields that existed for the day. The second state will attempt to evaluate model performance using the full probabilistic abilities of the ROM that consider all realizations concurrently. The focus of the evaluation will be on O_3. The exact pathway through the evaluation study will be determined by the resources available at the time.

This Project Summary was developed by EPA's Atmospheric Sciences Research 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 perception of air quality problems has increased in recent years from an urban scale to a larger regional scale as the effects of multi-day and long-range transport of air pollutants have been studied and understood. During the past five years the U.S. Environmental Protection Agency (EPA) has undertaken a model development effort to produce an air quality simulation model capable of treating all of the chemical and physical processes that are thought to affect the concentrations of air pollutants over several-day/1000 km scale domains. The EPA Regional Oxidant Model (ROM) is now operational. Among the processes it treats are horizontal transport, atmospheric chemistry, nighttime wind shear and turbulence episodes associated with the nocturnal jet, cumulus cloud effects on vertical mass transport, mesoscale vertical motions induced by terrain and the large scale flow, diffusion and deposition, subgrid scale chemistry processes, emissions of natural and anthropogenic precursors, and wet and dry removal. These processes are simulated in a three-dimensional (3-D) Eulerian framework with 3 1/2 vertical layers extending through the boundary layer and the capping inversion or cloud layer. In the present configuration of the ROM domain, horizontal resolution is approximately 20 km.

The data base that will be used in the evaluation study is that of the Northeast Corridor Regional Modeling Project (NECRMP). The NECRMP ambient data base consists of measurements made during four EPA field measurement pro-

grams in 1979 and 1980 in the northeastern U.S. These include the 1979 Northeast Regional Oxidant Study (NEROS I), the 1980 NEROS II, the Persistent Elevated Pollution Episodes Study, and the 1980 Urban Field Studies. Together, these studies have provided a variety of air quality and meteorological measurements on regional, urban, and site-specific scales. Also, a parallel effort within the NECRMP program assembled a complete source emissions inventory specifically addressing the ROM requirements. This emissions inventory will be supplemented by updated emissions for the U.S. and southern Canada contained in the National Acid Precipitation Assessment Program data base, and the revised inventory will be used in the ROM evaluation study.

The individual grid cells of the model domain for the NECRMP application are 15 minutes of longitude wide in the E-W direction and 10 minutes of latitude in the N-S direction, or about 18.5 km^2. There are 60 cells E-W and 42 cells N-S, giving a total horizontal extent of approximately 1100 km by 780 km. The vertical structure of the ROM consists of 3 1/2 layers. The bottom (1/2) layer is actually a diagnostic surface layer less than 100 m deep where surface deposition and subgrid scale chemical effects are modeled in a diagnostic manner. Layers 1 and 2 are prognostic model layers and extend through the depth of the well mixed layer during the day and the surface inversion and old mixed layer at night. The top prognostic model layer extends up to 1 km above the top of the mixed layer and includes any convective cloud elements.

The ROM System

The ROM described here is the 2nd generation version of the model. It is designed to simulate air pollution chemistry and dispersion over a 1000-km by 1000-km area for multiple day periods. The simulation of photochemical air pollution over such scales is a complex problem. The ROM reflects this complexity, consisting of some 25 programs accessing almost 100 files.

The input files contain the "raw" data accessed by the modeling system. Actually, these data files are produced by a combination of manual and computer manipulation of raw data. However, the input data files constitute a boundary between the ROM and the rest of the world, and therefore define the input data for the model. The initial preprocessors take the "raw" input data and transform them for use by other preprocessors. These data are transferred between preprocessor programs through the processor input files (PIF). The preprocessor programs are used to develop intermediate parameters that are then transformed into final model inputs. Examples of these intermediate parameters are the wind fields, emissions, and turbulent fluxes. The preprocessors transfer the processed data through the PIF files and the model input files (MIF).

The data contained in the MIF files are converted into the form required by the CORE model through the execution of the b-matrix compiler (BMC) program. The BMC translates the parameter fields in the MIF files (layer thicknesses, horizontal winds, interfacial fluxes, deposition velocities, etc.) into the matrix and vector elements necessary to operate the CORE model. The BMC transfers the information needed by the CORE model via the final data input files.

The CORE model is the computer language analogue of the differential equations that describe the processes involved in the chemistry and dispersion of regional photochemical air pollution. The CORE model is expressed in fundamental mathematical form. All inputs to the model are matrices and vectors whose elements are composites of meteorological parameters, chemical rate constants, etc.

The output of the CORE model consists of the layer-averaged pollutant concentrations for every grid cell every 30 minutes. However, these data are not the final model results. The ROM incorporates two new concepts in air pollution modeling. The first is a method for simulating the physical and chemical processes that occur within about the first 100 meters of the ground. The result of this scheme is that both the ground-level concentration and the root-mean-square concentration variation within each selected cell are produced. These data can be extracted using the results of the CORE model output (contained in the model output files) and parameters contained in the model input files.

The other concept is the incorporation of the uncertainty of the wind field representation. It is known that many different wind fields can be constructed that match the observed wind data and empirical and theoretical constraints. In the ROM, a family of possible wind fields is used in the model and leads to a result of an ensemble of concentration fields. Therefore, the model results will consist of a distribution of concentration values for each grid cell for each time step. The implementation of the multiple wind field concept requires that the model be run many (~10) times for each simulation. ROM postprocessor programs will be used to transform the results of the individual CORE outputs into ensemble concentration distributions. Other model postprocessors will transform the layer 1 concentration values into ground level (layer 0) concentration distributions. These data constitute the model results. Measures of model accuracy will be made by comparing the model predictions of concentration probability distributions with observed measurements, as well as examining individual model runs (realizations) on a single wind field.

Model Evaluation

The comprehensive evaluation of a 3-D gridded air quality simulation model is a complex task. At a basic level it is necessary to test whether the mathematical representations of the individual physical and chemical processes are correct. This is done for each component process in the ROM. The validity of the chemical kinetic mechanism, for example, may be tested with data from controlled smog chamber experiments. The vertical cloud flux algorithm can be validated with aircraft measurements of material below, within, and above a convective-type cloud from the NEROS field program. The wind field algorithm cannot be precisely tested in an independent fashion, although individual trajectories can be evaluated from tetroon data from the NEROS study, and the individual wind realizations can be analyzed for their consistency with measurements from fixed monitoring sites. The model evaluation outlined here takes the model as a whole and attempts to compare its predictions with ambient observations.

This view would suggest that the ROM could then be viewed as a "black box" model producing results that could be statistically analyzed against observations. Although it is tempting to take this view, it must be avoided. The blind application of statistical tests will not provide sufficient understanding of model performance to draw any meaningful conclusions from the study. The goal of a model evaluation study, such as the one proposed here, is to gain insight into the model predictions and the observed data. One wants to know if the model is producing the right answer for the right reason and whether the model prediction

is good enough for the user's purposes. The careful analysis of observational data will provide the direction for a meaningful path through the course of a model evaluation. The proper choice of statistical comparisons between observations and predictions can be established from an understanding of the phenomena shown by the data sets. This is the approach that we take in the ROM evaluation.

The approach will consist of three levels of analysis. The first level will explore the observed data set with diagnostic tests to bring out the important features, both on temporal and spatial scales, and to order the data according to the features that are found. Because the principal pollutant of interest included in the ROM simulations is O_3, the analysis of observed data will focus on it. The second level of analysis will be an evaluation of the ROM results for an individual realization of an episode. This approach is a deterministic analysis of each simulated realization without regard to the other members in the family of realizations. Finally, the third analysis level will consider concurrently the entire family of simulated realizations and the resulting probability distributions of concentration. This analysis structure is actually a hierarchy, with the results obtained at each lower level guiding the steps taken at the next higher level.

The ROM model evaluation must account for the stochastic nature of the predicted concentration field. This is a result of the multiple wind field realizations (interpolations) that the ROM wind field processor generates from a given set of wind observations. These multiple wind fields are each consistent with both the observations and physical laws governing atmospheric flow and are assigned probabilities of occurrence based on the inherent kinetic energy contained within the field. When all resulting wind fields are considered, the model generates a concentration probability distribution for a given receptor site within the domain instead of a single concentration value. When all realizations are considered concurrently, this aspect of the model prediction makes the ROM different than most other air quality simulation models and presents more of a challenge to the evaluation effort.

The first and second moments of the predicted concentration distribution can be used to determine the expected frequency with which the observed concentrations should fall within a given interval. The utility of the model for regulatory use is measured partly by the width of this interval. The second moment of the distribution (the concentration variance) is the parameter that defines this width. It is a measure of the inherent uncertainty of the model. If the width is very large the model may provide no more information than one would gather by guessing the expected concentration. This is true even if the model is shown to be accurate in other respects. Therefore part of the ROM evaluation should consist of an analysis of the predicted concentration variances. For regulatory use, the predicted concentration for maximum ozone at a given receptor location for a single realization is also a parameter of interest. This value may be compared to the corresponding observed value. Also, if groups of receptors can be identified having similar predicted concentration distributions, the observed distribution composed of the measurements from each receptor within the group can be compared to the predicted distribution from the group. Such groups might be stratified by receptor location with respect to downwind distance from major source emissions areas.

Base level evaluation studies with the ROM will be conducted for four or five regional smog episodes from the data base. The individual episodes range from 2 to 6 days in duration, and are mostly from the 1980 summer period. A full 2-week simulation will also be performed.

*The EPA author **K. L. Schere** (also the EPA Project Officer, see below) is with the Atmospheric Sciences Research Laboratory, Research Triangle Park, NC 27711, and A. J. Fabrick is with MEF Environmental Inc., Austin, TX 78758.*

The complete report, entitled "EPA Regional Oxidant Model: Description and Evaluation Plan," (Order No. PB 86-103 090/AS; Cost: $11.95, 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 Sciences Research Laboratory
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

United States
Environmental Protection
Agency

Center for Environmental Research
Information
Cincinnati OH 45268

EPA/600/S3-85/067