United States
Environmental Protection
Agency

Atmospheric Research and
Exposure Assessment Laboratory
Research Triangle Park NC 27711

Research and Development

EPA/600/S3-88/043 Mar. 1989

EPA

Project Summary

Meteorological Processor for Regulatory Models (MPRM-1.1) – User's Guide

John S. Irwin, James O. Paumier, and Roger W. Brode

Version 1.1 of MPRM provides a general purpose computer processor for organizing available meteorological data into a format suitable for use by air quality dispersion models. Specifically, the processor is designed to accommodate those dispersion models that have gained EPA approval for use in regulatory decision making. A unique feature of the processor is the ability to employ user collected meteorological measurements as well as those routinely collected by the National Weather Service (NWS).

MPRM-1.1 will support the following dispersion models listed in the Guideline on Air Quality Models (Revised) (EPA, 1986), as well as three screening models:

- **Those requiring RAMMET formatted data: BLP, RAM, ISCST, MPTER, CRSTER, and COMPLEX1.**
- **Those requiring STAR formatted data: CDM (with either 16 or 36 wind direction sectors), ISCLT, and VALLEY (long-term).**
- **Those requiring special formats: CALINE-3 and RTDM (default).**

MPRM can be envisioned as a three-stage system. The first stage retrieves the meteorological data from computer tape or disk files and processes the data through various quality assessment checks. The second stage collects all data available for a 24-hour period (upper air observations, hourly surface weather observations, and data collected as part of an on-site meteorological measurement program) and stores these data in a combined (merged) format. The third stage reads the merged meteorological data and performs the necessary processing to produce a meteorological data file suitable for use by the specified dispersion model.

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

EPA has recently issued guidance on the use of meteorological data, collected via an on-site measurement program, for regulatory modeling applications. The meteorological processors currently available from EPA do not have the capability of processing user collected on-site meteorological data as directed by the guidance. Therefore, MPRM-1.1 has been designed to construct meteorological data files of upper air, mixing height, surface observations, and on-site data for air pollution dispersion models that are routinely used in regulatory decision making by EPA. Specifically, the processor is designed to accommodate those dispersion models recommended for use in the Guideline on Air Quality Models (Revised in 1986).

Why This Design?

Recognizing that the list of approved dispersion models is apt to change, and that those changes are likely to call for use of various processing methods, the new processor is designed to have a highly modular structure. Additionally, the new processor is designed to avoid computer conflict problems. It was further decided to develop an input structure that would ultimately support use of a menu data entry system. This would allow construction of the input through a computer-controlled question and answer session, and possibly facilitate the usability of the processor by a variety of users.

What Does MPRM Do?

MPRM can be envisioned as a three-stage processing system, as depicted in Figure 1. During the first stage, the processor extracts upper air, mixing height, and surface data from the raw data files delivered from NCDC and on-site data from the raw data files developed from the on-site measurement program. The extracted data are processed through a series of quality assessment checks. As a result, reports of missing and suspect values are generated. During the second stage, the processor combines the available data for each midnight-to-midnight 24-hour period (twice-daily upper air soundings and mixing height data, hourly surface weather observations, and hourly on-site data) and stores these data in a combined (merged) format. During the third and final stage, the processor reads the merged data and develops a meteorological data file for the dispersion model selected by the user.

Extraction and Quality Assessment (Stage 1 Processing)

The goal of this first stage of processing is to:

- Read the on-site and NWS meteorological data files
- Find the data within the time period specified by the user
- Store these data in American Standard Code for Information Interchange (ASCII) data files
- Scan the stored values and report occurrences of missing or suspect values.

MPRM can currently process hourly surface observations in CD-144 format, upper air soundings in TD-5600 format, and mixing height data in TD-9689 format. Persons experienced with RAMMET have a working knowledge of hourly surface weather (CD-144) data and twice-daily mixing height (TD-9689) data. NCDC can provide CD-144 and TD-9689 data on computer magnetic tape or on 5 1/4-inch diskettes in a format suitable for use in IBM compatible personal computers (PCs). MPRM installed on a mainframe computer can process these data either from the computer tape or from mass storage data files. MPRM installed on a PC can process these data either from data files on the hard disk or from 5 1/4-inch diskettes. We would at least suggest an IBM-AT or equivalent with 640K random access memory (RAM) and a hard disk.

As of 1984, NCDC had converted to new data storage formats for surface observations and upper air soundings namely. elemental formats TD-3280 and TD-6201, respectively. NCDC now converts from these formats to service requests for CD-144 and TD-5600 formatted data. MPRM is designed internally for ready adaptation to these new formats. One, if not the first, upgrade will expand the data formats supported by MPRM to include TD-3290 and TD 6201 formats.

Because there is no standard format for storage of on-site meteorological data, MPRM is designed to process a variety of on-site data formats by having the user define the structure of the input data. The two major restrictions to consider in processing the on-site data are 1) the order in which the data values are presented and 2) the data file must be a standard ASCII data file. In principle MPRM will be able to process the on-site data as long as the data values for each observation are ordered correctly (date and time, then meteorological values) and the observation can be read using a FORTRAN FORMAT statement.

An additional capability of this first stage is assessing the quality of the data by checking for possible missing or suspect values. Any occurrences of missing or suspect data values are reported before the upper air soundings, mixing height data, surface observations and on-site data are combined.

The output files from this first stage of processing can be edited using standard text editors routinely available on computer systems. The only foreseeable problem in editing these data files is that some text editors have a limitation on the size of the file to be edited. This problem is of most concern when the editing is performed on a PC where the text editor is typically limited to the available RAM. For example, because of the RAM limitation, a file consisting of 700 Kb of year of hourly surface observations in CD-144 format could not be edited. A possible solution to this problem is to break the file into parts that can be edited with software designed for this purpose. The modified larger file is recreated by concatenating the smaller files.

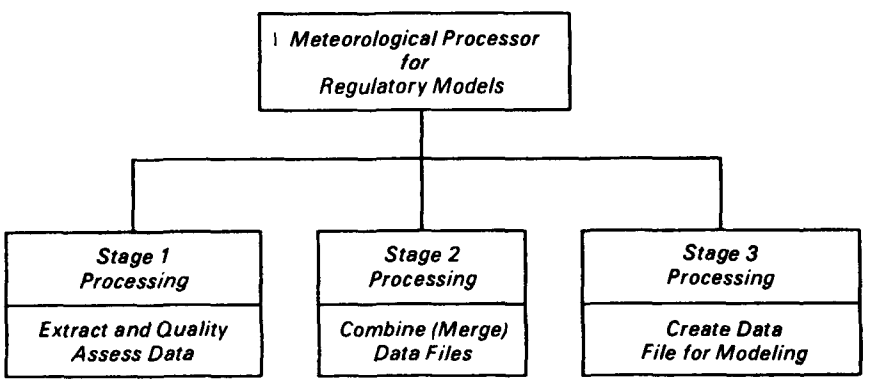

Figure 1. Overview of processing stages within MPRM.

Combining Data (Stage 2 Processing)

The goal of this second stage of processing is to:

- Combine into one file the available on-site and NWS meteorological data

files created during Stage 1 processing

- Store the data in a more compact format.

For the specification of the dispersive state of the atmosphere, it is most convenient to consider the physics of the atmosphere on a daily basis, i.e., a 24-hour period. Estimation of the depth of convective mixing is in reality the summation of effects starting with the heating of the surface shortly after sunrise. Thus, the merging of the available data for each 24-hour period is the next logical step in the processing before developing the characterization of the input meteorological data files for the dispersion models.

The merged data are stored in unformatted form because this format provides a more efficient usage of storage than the formatted ASCII data file storage used during the first stage of processing. The ASCII files are convenient for text editors but are no longer needed once the quality assessment and editing are completed.

Creating a Model Input File (Stage 3 Processing)

The goal of this third stage of processing is to:

- Create a meteorological data file for use with a regulatory dispersion model chosen by the user.

MPRM can generate any one of several output formats to meet the input requirements of the regulatory dispersion model chosen by the user. The RAMMET format has been selected as the default output with default methods for processing wind, temperature, stability category and mixing heights. These methods employ the NWS hourly surface weather observations and NCDC twice-daily mixing heights and duplicate the processing performed by the RAMMET meteorological processor.

Since no "modeling" has been performed prior to the third stage of processing, it is anticipated that future changes to the modeling guidelines will have the most impact on this stage of meteorological processing. Acceptance of new algorithms for mixing height estimation, or methods for characterizing the variation of wind speed and wind direction with height, would require that new computer algorithms be supplied within MPRM for use at this stage of processing. Acceptance of a new dispersion model might require changes to the output subroutine within MPRM in order to provide the meteorological data in the format required by the new accepted dispersion model. In consideration of these possibilities, MPRM has a highly modularized design. This allows upgrading of specific parts of the computer code without having to redesign the processor.

Relationship of MPRM to EPA Air Pollution Modeling Guidance

The data processing methods incorporated into MPRM are intended to implement the recommendations contained in the EPA on-site meteorological program guidance document. These recommendations include the determination of Pasquill stability categories from on-site measurements, based on the recommendations in the EPA Guideline on Air Quality Models (Revised). As data processing recommendations are modified, MPRM will be upgraded to reflect the latest guidance. Moreover, any discrepancies that might exist between MPRM and current regulatory guidance should not be construed as guidance, but as errors within the MPRM system.

It is not the purpose of this user's guide to provide a comprehensive summary of all relevant guidance on dispersion modeling for regulatory applications. Other recommendations from the guidance document for conduct of an on-site meteorological measurement program may be relevant to a particular application. For example, the on-site meteorological program guidance document contains recommendations on instrument siting and quality assurance. An important recommendation is that a minimum of 90 percent valid data recovery exist for each variable before a data set can be used for regulatory modeling. The issue of handling missing data values for dispersion modeling is discussed below in more detail. Please note that data substitution cannot be used to reach the 90 percent data recovery rate required by regulatory modeling guidance.

Missing Values, the Bane of All Data Sets

When the meteorological conditions are insufficiently declared, a dispersion model will not be able to produce concentration estimates. In the case of dispersion models currently accepted for regulatory applications, the situation is aggravated by the fact that none of the hourly dispersion models can continue processing if the meteorological record is not continuous. The only way to continue processing is to present the dispersion model with an unbroken meteorological record having no missing values.

Substitutions based on other on-site data, if available and deemed to be representative, may provide the best solution for providing an unbroken meteorological record. If the situation is such that only 1 hour's data is missing, it might be practical to use linear interpolation between adjacent hours to estimate the missing meteorological conditions. As the time period of missing values increases, the usefulness as well as the reasonableness of linear interpolation to fill in missing values becomes increasingly more dubious. Rationalizations involving use of monthly mean values from climatological records are sometimes employed. The fact remains that, given the right circumstances, any technique employed for filling in missing values can prove to be inadequate.

A clear consensus has yet to be reached on how best to resolve the dilemma created by a broken meteorological record. A possible solution may involve making substitutions for missing meteorological values for isolated 1-hour periods, and treating longer breaks in the meteorological record within the dispersion model. At the very least, the dispersion models might be modified to process available valid data and to skip (or output a missing value indicator for the concentration estimate) those hours when processing could not continue due to missing values in the meteorological record.

While the guidance document for conduct of on-site meteorological measurement programs contains a recommended hierarchy of data substitution strategies for regulatory applications, the implementation of these recommendations requires expert judgment of "representativeness" of the data substitutions. Yet to be developed is a system of numerical rules having sufficient expertise that we can confidently recommend their use in a universal sense for automatic processing of missing data values.

Since MPRM is constrained to include only processing methods that have been accepted for use in regulatory applications and since we have yet to develop a set of numerical techniques for universal use in handling missing values, MPRM has no automatic method for

correcting missing values. If and when techniques are accepted for handling missing values on an automatic basis, they will be incorporated into MPRM, unless of course the resolution is within the dispersion model algorithm.

The EPA author, ***John S. Irwin,*** *is with the Atmospheric Research and Exposure Assessment Laboratory, Research Triangle Park, NC 27711; James 0. Paumier is with the Computer Sciences Corporation, P.O. Box 12767, Research Triangle Park, NC 27709; and* ***Roger W. Brode*** *is also an EPA author with the Office of Air Quality Planning and Standards, Research Triangle Park, NC 27711*

D. Bruce Turner *is the EPA Project Officer (see below).*

The complete report, entitled, "Meteorological Processor for Regulatory Models (MPRM-1.1) User's Guide," (Order No. PB 89-127 526/AS; Cost: $28.95, 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 Research and Exposure Assessment Laboratory
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

EPA/600/S3-88/043