GENERATING SHAPEFILES FROM CMAQ, SMOKE, AND CAMX FILES

Neil J. M. Wheeler*, Kenneth J. Craig, and Bryan M. Penfold
Sonoma Technology, Inc., Petaluma, CA, USA

1. INTRODUCTION

Air quality modeling systems, such as the Community Multiscale Air Quality model (CMAQ) (National Exposure Research Laboratory, 1999), the Comprehensive Air Quality Model with extensions (CAMx) (ENVIRON International Corporation, 2004), the Urban Airshed Model (UAM), require gridded inputs of meteorology, emissions, air quality, and geophysical data, and produce two- and three-dimensional gridded outputs for multiple species. Because of the large volume of data used and produced by these systems, visualization tools such as the Package for Analysis and Visualization of Environmental data (PAVE) (Thorpe et al., 1996) are routinely used to quality assure, investigate, and analyze their input and output files. In addition, various post-processing tools have been developed for use in model performance evaluation and other analyses. To allow these tools to be used with model input and output files for which they were not designed, Sonoma Technology, Inc. (STI) has developed a series of file conversion programs.

Using subroutines from these file conversion programs and a publicly available C-library, a new utility program, GSHAPES, was developed to extract data from common model input and output files and write Environmental Systems Research Institute, Inc. (ESRI) shapefiles. Shapefiles can be viewed and analyzed with ESRI software such as ARC/INFO and ArcView. Further, by converting model input and output data to shapefile format, high-quality maps of the data can be prepared and the model data can be combined with standard geophysical and socioeconomic data to perform complex analyses.

2. APPROACH

To build the GSHAPES program, a simple user interface, a file reader, and a file writer were programmed in FORTRAN. The File Reader program was developed from file readers previously developed for use in format conversion programs. Currently the File Reader can read four model file formats:

1. the I/O API netCDF format used by CMAQ;

2. the original UAM format that includes headers;

3. the newer UAM/CAMx format without headers that requires additional metadata; and

4. the CALPUFF output format.

The File Writer program creates an ESRI projection (.prj) file based on the metadata provided to it and passes data that was read from a model input or output file to the Shapefile Interface program, which calls routines in the Shapefile C Library (Warmerdam, 2006) to create the shapefile. The components and data flows for this shapefile creation process are summarized in **Figure 1**.
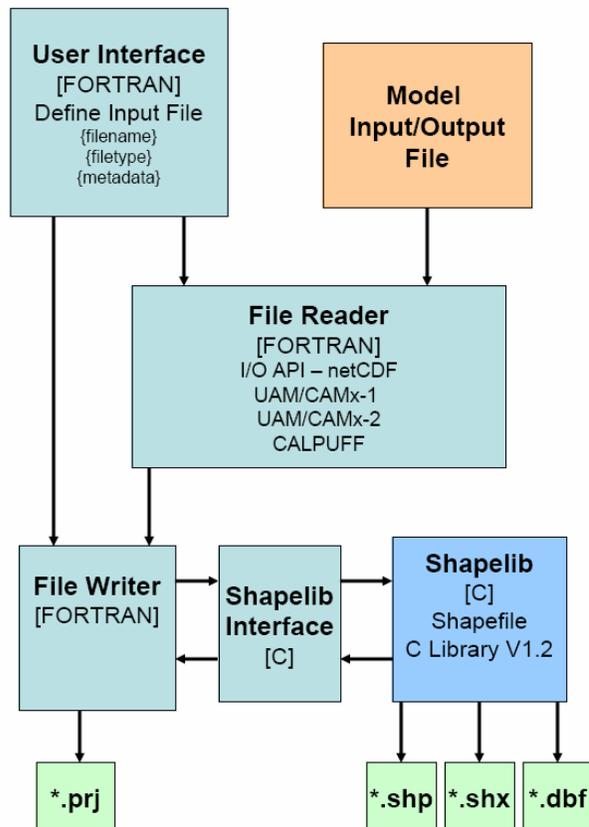
Fig. 1. GSHAPES components and data flows.

Corresponding author: Neil Wheeler, Sonoma Technology, Inc., 1360 Redwood Way, Suite C, Petaluma, CA 94954-1169; e-mail: neil@sonomatech.com

The shapefile format (Environmental Systems Research Institute, 1998) is an interchange format promulgated by ESRI for simple vector data with attributes. The file format actually consists of three files:

1. the *.shp file that holds the actual shape vertices,
2. the *.shx file that holds index data pointing to the structures in the .shp file, and
3. the *.dbf file that holds the attributes (e.g., concentrations).

In addition, the projection file (*.prj) provides information about the map projection used in the model files.

A shapefile stores nontopological geometry and attribute information for the spatial features in a data set. The geometry for a feature is stored as a shape comprising a set of vector coordinates. Because shapefiles do not have the processing overhead of a topological data structure, they have advantages over other data sources such as faster drawing speed and edit ability. Shapefiles handle single features that overlap or are noncontiguous. They also typically require less disk space and are easier to read and write. Shapefiles can support point, line, and area features. Area features are represented as closed loop, double-digitized polygons. Attributes are held in a dBASE® format file. Each attribute record has a one-to-one relationship with the associated shape record.

3. RESULTS

A CMAQ simulation for the southeastern United States was processed with a utility program that calculated peak 8-hr ozone concentrations and output those concentrations to a NetCDF-based I/O API file. The peak 8-hr ozone concentrations in this file for July 18, 2002, as visualized with PAVE, are shown in **Figure 2**. Concentrations are shown in units of ppm and state boundaries were drawn using data sets provided with PAVE.

The I/O API file was then processed with GSHAPES to produce the four files used in its implementation of the shapefile format. The data in those files were then visualized with ArcView GIS software. The results of this visualization are shown in **Figure 3**.

In the ArcView visualization, concentrations were converted to units of ppb and displayed as color-filled contours similar to those produced by PAVE. The resulting contours were overlayed on a map covering a slightly larger area than the modeling domain and annotated with a legend and distance scale. The patterns of peak 8-hr ozone concentrations seen in ArcView visualization match those in the PAVE visualization, indicating the conversion was performed correctly.
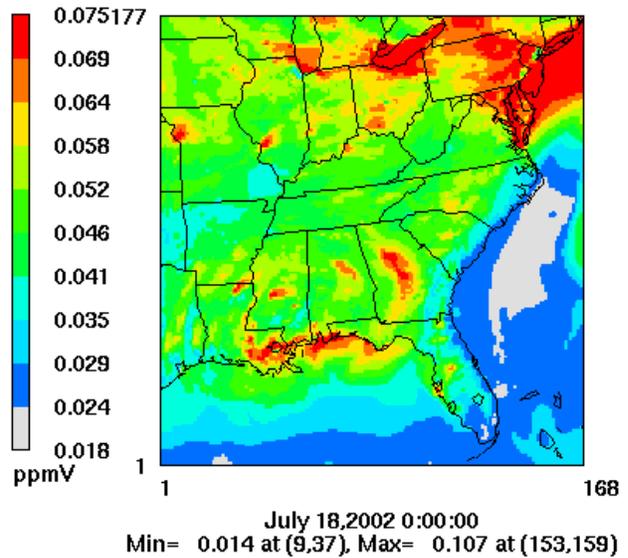
Fig 2. Peak 8-hr ozone concentrations from CMAQ visualized with PAVE.
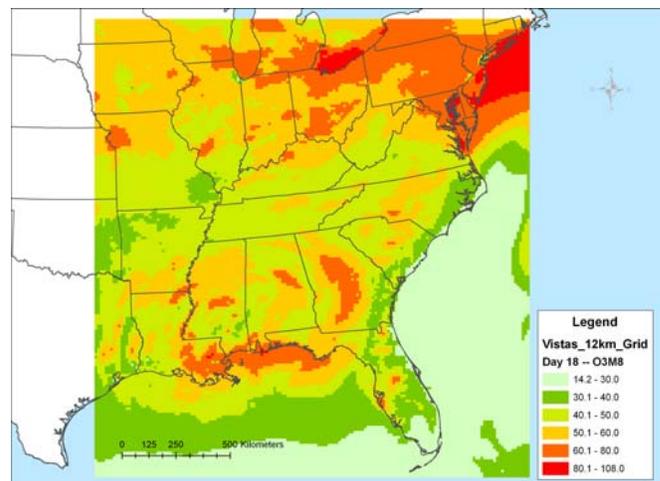
Fig 3. Peak 8-hr ozone concentrations from CMAQ visualized with ArcView.

4. STATUS AND AVAILABILITY

GSHAPES is currently undergoing alpha testing and is expected to be made publicly available in early 2007. At that time, source code, Linux executables, and documentation will made available on the STI web site. GSHAPES will be licensed under the terms of the GNU General Public License published by the Free Software Foundation.

The source for the Shapefile C Library is copyrighted (1998) by Frank Warmerdam. As of Shapelib Version 1.2.6, the core portions of the library are made available under either of two licenses: the Library GNU Public License (LGPL) or an MIT style license.

5. REFERENCES

ENVIRON International Corporation, 2004: User's guide to the Comprehensive Air Quality Model with Extensions (CAMx). Version 4.00, Prepared by ENVIRON International Corporation, Novato, CA.

Environmental Systems Research Institute, I., 1998: ESRI Shapefile Technical Description. White Paper, Environmental Systems Research Institute, Inc., Redlands, CA 34 pp.

National Exposure Research Laboratory, 1999: Science algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) modeling system. Report, EPA/600/R-99/030 (peer reviewed) prepared by the National Exposure Research Laboratory, Research Triangle Park, NC.

Thorpe, S., J. Ambrosiano, C. Coats, A. Eyth, S. Fine, D. Hils, T. Smith, A. Trayanov, R. Balay, M. Vouk, and T. Turner, 1996: The package for analysis and visualization of environmental data. *Proceedings, Computing in Environmental Resource Management, 2-4 December*, Research Triangle Park, NC, 241-249.

Warmerdam, F., 2006: Shapefile C Library V1.2. Available on the Internet at http://shapelib.maptools.org August 15, 2006.