

# **The MIMS Spatial Allocator: A Tool for Generating Emission Surrogates without a Geographic Information System**

Alison Eyth, Kimberly Hanisak  
Carolina Environmental Program  
University of North Carolina at Chapel Hill  
CB# 105, 100 Miller Hall  
Chapel Hill, NC 27599-1105  
<http://www.cep.unc.edu>  
[eyth@unc.edu](mailto:eyth@unc.edu)

## **ABSTRACT**

The Multimedia Integrated Modeling System (MIMS) Spatial Allocator is a free tool for generating spatial surrogates for emissions, changing the map projection of shapefiles, and performing other types of spatial allocation without requiring users to have a Geographic Information System (GIS). The tool outputs spatial surrogates for regularly spaced air quality model grids on one of several classes of map projections, including Universal Transverse Mercator (UTM), Lambert Conformal, and Latitude-Longitude. It supports a variety of standard and user-specified ellipsoids to approximate the shape of the earth. In addition to creating emission surrogates, other types of spatial computations can be performed by the tool, such as copying shapefiles to new map projections and aggregation. For example, county polygons and their attributes can be aggregated into state polygons with corresponding attributes. The tool can run on Unix or Windows computers, and scripts are provided to assist with running the tool on both Unix and Windows. This new tool bridges the gap between emission modelers and GIS analysts by providing a zero cost, easy-to-use, powerful, and portable system for spatial surrogate generation, changing the map projection of shapefiles, and other spatial analyses.

## **INTRODUCTION**

The Multimedia Integrated Modeling System (MIMS) Spatial Allocator was developed by members of the Carolina Environmental Program at the University of North Carolina at Chapel Hill (<http://www.unc.cep.edu>). The project was sponsored by the EPA Office of Research and Development to provide a tool for generating spatial surrogates for emissions and performing other types of spatial allocation without requiring a Geographic Information System (GIS). By making this tool available to the public, the task of generating surrogates can be moved from the EPA and a few contractors into the hands of the emission modelers. The tool can generate surrogates for point, line, or polygon-based data sets such as ports, airports, housing, population, agriculture, water area, and railroads. It can also perform some other useful operations.

The drawbacks of using a GIS to generate surrogates are that GISs can cost thousands of dollars and they are complex software systems that can be difficult to use. In contrast to a GIS, the tool is provided for free from the EPA and is a comparatively smaller piece of software tailored specifically for creating spatial surrogates for emissions. Thus, the learning curve is substantially smaller than that of a GIS. The tool is written in C, and can run on Windows or Unix systems. It can be scripted and run as a background job on both Unix and Windows. It uses ESRI Shapefiles – a standard in the GIS industry – for input data bases (see <http://www.esri.com>). The ellipsoid used to approximate the shape of the earth in the input shapefiles can be specified. New input data sets can easily be used as they become available in the form of shapefiles.

The tool outputs surrogates for regularly spaced air quality model grids in a format that can be input to the Sparse Matrix Operator Kernel Emissions (SMOKE - see <http://www.cmascenter.org/doc.html>) modeling system. This format is similar to the one used by EMS-2001 (see [http://www.ladco.org/emis/ems\\_2001/index.html](http://www.ladco.org/emis/ems_2001/index.html)). The user may choose one of several classes of map projections on which to output surrogates. The available map projection classes include: Universal Transverse Mercator (UTM), Lambert Conformal, Latitude-Longitude, Mercator, and Stereographic. In addition to creating spatial surrogates for emissions, other types of spatial computations can be performed by the tool. For example, the attributes of county-level polygons can be aggregated into attributes for state-level polygons. The attributes can be either averaged or summed, depending on which method is appropriate for each attribute. The tool can also convert shapefiles from one map projection and earth ellipsoid to another.

## **BODY**

### **Background on Shapefiles**

A basic understanding of shapefiles is important for users of the Spatial Allocator because these files serve as both inputs to and outputs from the program. Shapefiles contain geographic data objects in the form of points, lines, or polygons with associated informational attributes. Shapefiles are a GIS industry standard format developed by ESRI. Each shapefile stores data for a single type of object (i.e. points, lines or polygons). Shapefiles have a non-topological vector data format in which each object has its own set of coordinates specified. Point shapes are single-coordinate features such as stacks, schools, churches or monuments. Line shapes can be continuous lines such as fault lines on a map, or they can be polylines like branches of a river. Since there are no topological relationships between lines in a shapefile, they are not associated with any points or polygons. Polygon shapes can be simple areas or multi-part areas like states and counties. Polygon shapes can overlap, but the topological relationship of this overlap is not stored in the shapefile.

Each shapefile consists of a set of at least three component files. The three mandatory component files that comprise "a shapefile" contain the coordinates of the shapes, the attributes associated with each shape (e.g., population, number of households, area), and an index that maps the attribute entries to the coordinate entries. Each component of the shapefile has the with the same base file name but different file extensions. The attributes are stored in a dBASE format (.dbf) file, coordinates are stored in a .shp file, and the index is stored in a .shx file. There are some additional optional components of shapefiles - such as the .prj file, which contains information about the map projection of the coordinates in the .shp file.

### **Available Operating Modes**

The tool currently has four modes of operation, which are specified by setting the value of the MIMS\_PROCESSING environment variable. The available modes are:

- SURROGATE - create spatial surrogates for emission modeling
- CONVERT\_SHAPE - convert a shape file to a different map projection and/or ellipsoid
- AGGREGATE - aggregate data from one geospatial unit to another by summing attributes
- AVERAGE - average data from one geospatial unit to another by averaging attributes

The primary output from the SURROGATE mode is an ASCII surrogate file that can be input directly into SMOKE. The output from the CONVERT\_SHAPE mode is a shapefile with the same attributes as

the input shapefile, but with the coordinates of the geospatial objects modified to be with respect to the user-specified output map projection and/or ellipsoid. The outputs from the AGGREGATE and AVERAGE modes are shapefiles. The .shp file output from these modes is the same, but the method by which the attributes are computed for the output shapes differs. AGGREGATE should be used for attributes that are additive (e.g. population), whereas AVERAGE should be used for attributes that are per unit area (e.g. population density). The list of attributes to include in the output shapefile can be specified.

## How Surrogates are Computed

Spatial surrogates for emissions are most commonly used to map county-level emissions data onto the rectangular grid cells used by an Eulerian air quality model such as the Community Multiscale Air Quality model (CMAQ - see <http://www.cmascenter.org/doc.html>). A spatial surrogate is a value greater than zero and less than or equal to one that specifies the fraction of the emissions in the spatial region of interest (e.g., county) that should be allocated to a particular grid cell. Three types of surrogates can be created with the MIMS Spatial Allocator: polygon-based, line-based, and point-based. The difference between these is the type of attribute that is used to "weight" the county level data into grid cells. This weighting process allows the emissions to be allocated in a manner more specific than a simple uniform spread over the county. Polygon-based surrogates use attribute information that is based on area (e.g., population in a census tract). Some other examples of attributes that are used as weights are points that represent ports or airports, the area of or land use in a county polygon, and lines representing the location of roads or railroads.

The surrogate value is calculated as the ratio of the weight attribute's value in the spatial intersection of the county and the grid cell to the total value of the weight attribute in the county. Note that a surrogate does not have to be for a county, but could be for some other geographic region such as state, nation, or census tract. The polygons that represent these geographic regions are known by the spatial allocator as the "data polygons", whereas the polygons, lines, or points used for the weight are known as the "weight polygons". The spatial surrogate value for county *Cty*, and grid cell *GC* is computed according to the equation:

$$srg(Cty, GC) = \frac{Wt(Cty \cap GC)}{Wt(Cty)} \quad (1)$$

where  $Wt(Cty \cap GC)$  is the value of the weight attribute in the region of intersection between the county *Cty* and the grid cell *GC*, and  $Wt(Cty)$  is the value of the weight attribute in the entire county. The emissions modeling system uses this value to compute the emissions in a particular grid cell as follows:

$$emis(GC) = srg(Cty, GC) * emis(Cty) \quad (2)$$

where  $srg(Cty, GC)$  is the surrogate value for the county *Cty* and grid cell *GC*, and  $emis(Cty)$  are the emissions in the county *Cty*.

The numerator of the spatial surrogate ratio in Equation (1), which is the value of the weight attribute in the region of intersection between county *Cty* and grid cell *GC*, is calculated according to the equation:

$$Wt(Cty \cap GC) = \sum_{wp} Wt(wp) * \frac{area(wp \cap Cty \cap GC)}{area(wp)} \quad (3)$$

where  $Wt(wp)$  is the value of the weight attribute in the weight polygon  $wp$ ;  $area(wp \cap Cty \cap GC)$  is the area of the region of intersection between the weight polygon  $wp$ , county  $Cty$ , and grid cell  $GC$ ; and  $area(wp)$  is the area of the weight polygon.

The denominator for the surrogate ratio (i.e. the total weight attribute value in the county or other data polygon) is calculated according to the equation:

$$Wt(cty) = \sum_{wp} Wt(wp) * \frac{area(wp \cap Cty)}{area(wp)} \quad (4)$$

where  $Wt(wp)$  is the value of the weight attribute in the weight polygon  $wp$ ;  $area(wp \cap Cty)$  is the area of the region of intersection between the weight polygon  $wp$  and county  $Cty$ ; and  $area(wp)$  is the area of the weight polygon  $wp$ .

For line-based surrogates the length of the linear weight feature (e.g., railroad, river, road) replaces area in the above equations. For point-based surrogates, instead of area, allocate a value of 1 if the weight point falls within the region of interest and a zero if it doesn't. In some cases, no special weight attribute is desired, and instead the surrogate is based purely on the area of the polygon, length of polyline, or count of the points. This mode is supported by the tool. In this case, the above equations simplify to the following:

$$Wt(Cty \cap GC) = \sum_{wp} \frac{area(wp \cap Cty \cap GC)}{area(wp)} = area(Cty \cap GC) \quad (5)$$

$$Wt(cty) = \sum_{wp} \frac{area(wp \cap Cty)}{area(wp)} = area(Cty) \quad (6)$$

As discussed above, if the weight polygons are lines or points (e.g. railroads or ports), replace area with length or count, respectively.

### Using the Spatial Allocator to Generate Surrogates

Scripts for Unix and Windows that generate a set of about six different spatial surrogates for emissions are provided with the spatial allocator. On both Unix and Windows, the tool is driven by setting environment variables. The required environment variables for generating surrogates are the following:

- MIMS\_PROCESSING - the operating mode of the program (e.g. SURROGATE)
- GRIDDESC - Directory and filename for the GRIDDESC file
- GRID\_NAME - The name of the output grid (this must exist as a grid in the GRIDDESC file)
- POLY\_DATA\_TYPE - Type of file containing data polygons (use ShapeFile for surrogates).
- POLY\_DATA - Directory and base file name for file containing data polygons
- ATTR\_DATA\_ID - Name of attribute in data polygon file that specifies a unique entry (e.g. FIPS)
- POLY\_WEIGHT\_TYPE - Type of file containing weight polygons (should be ShapeFile)
- POLY\_OUT\_TYPE - Type of file for output (use RegularGrid for surrogates)
- POLY\_WEIGHT - Directory and base file name for file containing weight polygons; or NONE.
- ATTR\_WEIGHT - Attribute used as the weight in the surrogate calculation, or NONE.
- CATEGORY\_WEIGHT - The integer used to designate a particular surrogate

- SURROGATE\_FILE - Directory and filename of output surrogate file (including .txt extension)

Below is an example of a simple Windows .bat file that sets the required environment variables for generating a single surrogate for ports.

**Example 1. A Basic Windows Batch File for Generating a Single Spatial Surrogate**

```
set MIMS_PROCESSING=SURROGATE
set DATA=C:\surrogates\inputs

set GRIDDESC=%DATA%\GRIDDESC.txt
set GRID_NAME=M_08_99NASH

set POLY_DATA_TYPE=ShapeFile
set POLY_DATA=%DATA%\cnty_tn
set ATTR_DATA_ID=FIPS_CODE

set POLY_WEIGHT_TYPE=ShapeFile
set POLY_OUT_TYPE=RegularGrid

echo Generating ports surrogate
set POLY_WEIGHT=%DATA%\tn_ports
set ATTR_WEIGHT=BERTHS
set CATEGORY_WEIGHT=4

set SURROGATE_FILE=C:\surrogates\output\srg_ports.%GRID_NAME%.txt
C:\surrogates\bin\mims_spatial.exe
```

Table 1 illustrates the impact of the weight attribute. In this case there are four grid cells that overlap the county with FIPS code 45001. The leftmost column is the grid cell number (note that these normally consist of a row and column). The second column shows the number of port objects in the grid cell. The third column shows the surrogate values for the county if ATTR\_WEIGHT is set to NONE, and the fourth column shows the surrogate values for the county if ATTR\_WEIGHT is set to BERTHS. Note that in the third column, the number of points in each grid cell is used to compute the surrogate, but in the fourth column to number of berths in each grid cell is used. The total of both surrogates is 1 because the county lies total within the grid. If the county did not totally lie within the grid, the total of the surrogate fractions might be less than 1.

**Table 1. The Impact of the Weight Attribute on a Point Surrogate**

Grid Cell	Number of Ports	Number of Berths	Srg (wt=NONE)	Srg (wt=BERTHS)
1	0	0	0	0
2	2	6	0.5 = 2/4	0.5 = 6/12
3	1	6	0.25 = 1/4	0.5 = 6/12
4	1	0	0.25 = 1/4	0
Total	4	12	1	1

## Specifying Map Projections and Ellipsoids

The default map projection used by the spatial allocator for weight and data polygon files is Latitude-Longitude, and the default ellipsoid is a sphere with a radius of 6370977m. Non-default values for the map projection and ellipsoid can be specified using the conventions of the PROJ.4 library (see <http://www.remotesensing.org/proj>). Information on specifying map projection parameters for PROJ.4 can be found at the bottom of the web pages linked from <http://www.remotesensing.org/proj>. The projection parameters should start with "+proj=" and the parameters should be in a comma separated list without white space. The environment variables to set are DATA\_POLY\_MAP\_PRJN, WEIGHT\_POLY\_MAP\_PRJN, and OUTPUT\_POLY\_MAP\_PRJN.

To use a Lambert conformal conic projection with standard parallels of 33 and 45 degrees, a central latitude of 40 degrees, and a central longitude of -97 degrees as the projection for data polygons, use the following syntax:

```
setenv DATA_POLY_MAP_PRJN "+proj=lcc,+lat_1=33,+lat_2=45,+lat_0=40,+lon_0=-97"
```

To use a Latitude-Longitude projection, leave the variable unset or set it to LATLON. To specify a UTM projection for zone 17 for the weight polygons, use the following syntax:

```
setenv WEIGHT_POLY_MAP_PRJN "+proj=utm,+zone=17"
```

The shape of the Earth's ellipsoid can be set independently for data, weight, and output polygons using the DATA\_POLY\_ELLIPSOID, WEIGHT\_POLY\_ELLIPSOID, and OUTPUT\_POLY\_ELLIPSOID environment variables. If these variables are not specified, or they are specified as SPHERE, the default spherical ellipsoid is used. To specify a non-default ellipsoid, you must specify either the name of a standard Ellipsoid, or the values that traditionally describe the shape of the Earth including a (Ellipsoidal Semi-Major Axis), b (Ellipsoidal Semi-Minor Axis), R (radius), and rf. A plus sign must precede either the name of the Earth Shape or each ellipsoidal parameter. When supplying multiple parameters, the list must be comma separated without any white space between the parameters. For example, you may specify the GRS80 spheroid using the following syntax:

```
setenv WEIGHT_POLY_ELLIPSOID=+GRS80
```

Or you may specify the ellipsoid parameters as follows:

```
setenv WEIGHT_POLY_ELLIPSOID=+a=6378137.0,+rf=298.257222101
```

The OUTPUT\_POLY\_MAP\_PRJN and the OUPUT\_POLY\_ELLIPSOID are not used in the SURROGATE mode. However, they are used for the other modes such as CONVERT\_SHAPE and AGGREGATE.

## Converting the Map Projection of Shapefiles

A copy of a shapefile with a different map projection can be made by using the CONVERT\_SHAPE mode of the spatial allocator. The method for accomplishing this uses several of the concepts that were discussed above. Unix and Windows scripts are provided to aid with this operation. Below is a Unix C-shell script that illustrates how to use this mode.

```

#!/bin/csh -f
setenv MIMS_PROCESSING CONVERT_SHAPE
setenv POLY_OUT_TYPE      RegularGrid      # Type of grid
setenv POLY_DATA_TYPE     ShapeFile       # Type of input data file
setenv POLY_DATA          $argv[1]        # shape file name - no extension
setenv POLY_OUT_NAME      $argv[2]        # shape file name - no extension

setenv DATA_POLY_MAP_PRJN "+proj=lcc,+lat_1=33,+lat_2=45,+lat_0=40,+lon_0=-97"
setenv DATA_POLY_ELLIPSOID +WGS84
setenv OUTPUT_POLY_MAP_PRJN  LATLON      # map projection for data poly file
setenv OUTPUT_POLY_ELLIPSOID SPHERE
/apps/mims_spatial/bin/mims_spatial.exe

```

## Specifying an Output Grid for Surrogates

Output grids for spatial surrogates are selected using a “GRIDDESC file”. The GRIDDESC file is part of the EDSS / Models-3 I/O API (see <http://www.cmascenter.org/doc.html>). This file is also used by SMOKE and CMAQ to specify modeling grids. The GRIDDESC file contains both map projection and grid description information. Note that the I/O API assumes that the earth's ellipsoid is a sphere. Map projections are listed at the top of the GRIDDESC file. Each map projection has a name, type, and five descriptive parameters: PROJ\_ALPHA, PROJ\_BETA, PROJ\_GAMMA, X-CENT, and Y-CENT. The meaning of the five parameters is dependent upon the projection type. The supported types of map projections are: Latitude-Longitude, Lambert, Mercator, Stereographic, UTM, Polar, Equatorial Mercator, Transverse Mercator, and Custom - each type is represented by a different integer.

Below the map projection section of the GRIDDESC file is a section for grids. Each grid has a name, references one of the map projections at the top of the file, and the following parameters: x origin, y origin, the size of the grid cells in the x and y directions, the number of columns and rows in the grid, and the thickness of the external boundary around the grid. The names of the grids and map projections are given on the line before the remaining parameters. An example of a GRIDDESC file is given below.

### Example 2. A GRIDDESC File

```

! coords --line:  name; type, P-alpha, P-beta, P-gamma, xcent, ycent
'LAT_LON'
1, 0.0D0, 0.0D0, 0.0D0, 0.0D0, 0.0D0
'UTM_10'
5, 10.0D0, 0.0D0, 0.0D0, 0.0D0, 0.0D0
'LAM_40N90W'
2, 30.0D0, 60.D0, -90.D0, -90.D0, 40.D0
'LAM_40N105W'
2, 30.0D0, 60.D0, -105.D0, -105.D0, 40.D0
' ' ! end coords.  grids:  name; xorig,yorig,xcell,ycell,ncols,nrows,nthik
'EPAW36_56X78'
'LAT_LON' , -127.0D0, 26.0D0, 0.5000D0, 0.33333D0, 56, 78, 1
'NEW_YORK'
'UTM_18', 480.0D3, 4440.0D3, 5.0D3, 5.0D3, 58, 46, 1'
'SMRAQ36_72X74'
'LAM_40N90W', -972.D3 , -1728.D3, 36.D3, 36.D3, 72, 74, 1
'DENVER8_34X45'
'LAM_40N105W', -116.D3, -188.D3, 8.D3, 8.D3, 34, 45, 1

```

The output grid for a surrogates file is specified by setting POLY\_OUT\_TYPE to RegularGrid, setting the variable GRIDDESC to the full path to the GRIDDESC file, and setting GRID\_NAME to the name of one of the grids in the GRIDDESC file.

## Options for Quality Assurance of Surrogates

The spatial allocator has several options that are useful for performing quality assurance on the surrogates it generates. The environment variables that activate these options are:

- **OUTPUT\_SRG\_NUMERATOR:** If set to YES, output the computed surrogate numerator from Equation (1) as a comment that follows the actual surrogate values in the surrogate file. The surrogate numerator is the value of the weight attribute in the intersection of the grid cell with the county.
- **OUTPUT\_SRG\_DENOMINATOR:** If set to YES, output the computed surrogate denominator from Equation (1) as a comment that follows the actual surrogate values in the surrogate file. The surrogate denominator is the value of the weight attribute in the county.
- **MIMS\_QASUM:** If set to YES, output the sum of all the surrogate values computed so far for the county as a comment that follows the actual surrogate values in the surrogate file. For counties entirely within the output grid, the final QA sum should equal 1.
- **POLY\_OUT\_NAME:** When a file name is specified for this variable in SURROGATE mode, output the gridded weight attribute as a shapefile. The gridded values are also output as a comma separated value (.csv) file that can be loaded into a spreadsheet. The shapefile can be viewed using a GIS to check the consistency of the surrogates with the input data bases. The map projection of the shapefile is the map projection of the output grid using a SPHERE for the ellipsoid.

An excerpt from a surrogate file created with the above options turned on follows. On the second through last lines of the file, the first number is the surrogate category (from the value of CATEGORY\_WEIGHT). Note that multiple surrogate categories may exist in the same file. Typically multiple categories result from multiple runs of the tool. The category is followed by the county's FIPS code, the column and row of the grid cell (where 1,1 is the lower left) and the surrogate value. The "!" denotes that the rest of the line is a comment. The order of the values following the comment delimiter ("!") is: numerator, denominator, and QA sum. This file can be input directly into SMOKE because it ignores any characters after the "!".

### Example 3. An Excerpt from a Surrogate File with Quality Assurance Options Turned On

Cat	County	Col	Row	Surrogate val		Srg Numerator	Srg Denominator	QA Sum
3	47001	45	32	0.0083578	!	7.14390e+06	8.5475e+08	0.0083579
3	47001	46	31	0.0159004	!	1.35909e+07	8.5475e+08	0.024258
3	47001	46	32	0.0720642	!	6.15969e+07	8.5475e+08	0.096322
3	47001	46	33	0.0268572	!	2.29562e+07	8.5475e+08	0.12318
4	47011	44	20	1	!	8	8	1
4	47037	19	28	0.2	!	2	10	0.2
4	47037	20	27	0.3	!	3	10	0.5
4	47037	20	28	0.3	!	3	10	0.8
4	47037	21	29	0.2	!	2	10	1
4	47039	6	18	1	!	3	3	1
4	47071	6	10	1	!	1	1	1
4	47079	3	27	1	!	2	2	1
4	47085	7	24	0.9	!	18	20	0.9
4	47085	9	25	0.1	!	2	20	1

The first line of the surrogate file is a comment with information about the output grid. An example is given below. This example wraps over two lines as shown here, but would be a single line in an actual surrogate file. The information on this line from left to right is: the grid name, x and y coordinates of the lower left corner of the grid, the size of the grid cells in x and y, the number of grid cells in x and y, the external boundary layer thickness, the type of map projection and units thereof, and the map projection parameters (in this case the standard parallel latitudes and the longitude and latitude of the origin).

#### **Example 4. The First Line of a Surrogate File**

```
#GRID M_08_99NASH 1000000.0 -536000.0 8000.0 8000.0 46 42 1  
LAMBERT meters 30.0 60.0 -100.0 -100.0 40.0
```

When the POLY\_OUT\_NAME environment variable is used to specify a file name for an output shapefile that shows the total of the surrogate numerator weights for each grid cell, a corresponding .csv file is created that contains the same information in a format that can be loaded into a spreadsheet. An excerpt from one of these files for a population surrogate is shown below. In its entirety, the file contains the column, row, and sum of the numerator weights for each grid cell. Essentially it is a gridded version of the weight attribute.

#### **Example 5. An Excerpt from a Numerator Sum Comma Separated Value (.csv) File**

```
COL,ROW,NUMERSUM  
1,1,0.000000  
1,2,0.000000  
1,3,0.000000  
1,4,0.000000  
1,5,0.000000  
1,6,0.000000  
1,7,0.000000  
1,8,10.621275  
1,9,425.006375  
1,10,387.477077  
1,11,614.337478  
1,12,827.052993  
1,13,859.772380  
1,14,1201.650099  
1,15,5486.228720
```

## **System Architecture**

The MIMS Spatial Allocator is written in C and makes use of several public domain software packages. The Shapelib package (see <http://gdal.velocet.ca/projects/shapelib/>) is used to read and write shapefiles. When shapefiles are read in, the coordinates of the objects are converted from the input map projection to the output map projection for the surrogates or output Shapefile using PROJ.4. The intersection of weight, data, and grid polygons is performed using the Generic Polygon Clipper (see <http://www.cs.man.ac.uk/aig/staff/alan/software/>). Wrapper functions serve as an interface between each external package from the rest of the spatial allocator.

The main program of the spatial allocator first determines the mode of operation and then follows a sequence of steps appropriate to that mode. To compute surrogates, the tool performs the following operations:

- Creates polygon objects representing the grid cells after obtaining the grid information from the GRIDDESC file;
- Reads in the data polygons while converting the coordinates to the map projection of the output grid and excluding the coordinate data for any polygons that do not intersect the output grid;
- Reads the weight polygons while converting the coordinates to the map projection of the output grid and excluding the coordinate data for any polygons that do not intersect the output grid;
- Computes the intersection of the weight polygons with the data polygons;
- Computes the intersection of the new weight/data polygons with the grid cells;
- Computes the surrogates by first computing the values for all the possible surrogate numerators and denominators, and then computing the value of each surrogate fraction and writing them out to the output surrogate file with any accompanying quality assurance information requested by the user.

The function used to compute the actual surrogate fractions calls special functions to compute the numerator and the denominator of the surrogate fraction. The function used for the numerator computes the value of the weight attribute in the intersection of each data polygon with each grid cell. The function used for the denominator computes the value of the weight attribute in each data polygon. The functions used during the process of computing the surrogates are combined in different ways to support the other operating modes of the tool.

## **Limitations and Future Directions**

The current version of the spatial allocator has several limitations. First, the only output format for the surrogates is the SMOKE-ready input format, on a grid that assumes a spherical earth. Also, only the I/O API supported map projections can be used. At some point, it may be desirable to make the tool output surrogates in different formats and for additional map projections and spheroids. The tool could be updated to read and write other additional file formats, such as netCDF and the new ArcGIS format

Another limitation of the current version is that for input files that are lines, the lines must be “dissolved” at the county boundaries (see ESRI’s ArcGIS documentation for more information). If the lines are not dissolved in this way, the denominator of the surrogate is not computed properly. In the future it may be possible to remove this limitation.

The spatial allocator currently contains code to read a version of Biogenic Emissions Land Use Data (BELD) and output a type of surrogate. However, the output format cannot be read into SMOKE. Thus, the tools could be updated to read the BELD3 1-km land use data and output a file that can be used to drive SMOKE biogenic processing.

The initial concept for the spatial allocator included its use for regridding of data files and other more sophisticated spatial allocation features. This was difficult to integrate with the spatial surrogate production. The tool could be updated to include some of these additional features.

The process of computing surrogates can use a lot of memory on a computer. The original version of the spatial allocator declared enough memory to store for each data polygon an attribute value for every grid cell. When the program was applied to a national scale modeling scenario, this caused the memory usage to become very large. It was not necessary to store the data this way because each data polygon is likely to intersect only a small subset of all the grid cells. Thus, the most recent version includes an optimization that stores for each data polygon a list of attribute values for only the grid cells that actually intersect the bounding box of the data polygon. Another optimization was added to read in only the portions of line and polygon shapefiles that intersect the bounding box of the output grid / set of

output polygons.

These optimizations greatly cut down on memory usage. However, the tool still allocates memory for each data polygon in two places, in addition to using memory for each weight polygon and each grid cell. This means that memory usage is still an issue when large national scale shapefiles are used as inputs to process a national scale grid. In this case, not many of the data polygons will be thrown out because some portion output grid will intersect most of the data polygons. Thus, the memory usage will be greater than the size of the sum of the size of the .shp files for the weight and data polygons. This is a problem when the size of the input .shp files is close to or more than the size of the available memory on the computer running the program. This could cause the tool to run slowly or not to complete. In a future version of the tool, it may be possible to process portions of the input data files such that not all the data needs to be read into memory at the same time.

## CONCLUSIONS

This new tool bridges the gap between emission modelers and GIS analysts by providing a zero cost, easy-to-use, powerful, and portable system for spatial surrogate generation, changing the map projection of shapefiles, and other spatial analyses. The released version of the software is available from the Web page [http://www.epa.gov/asmdnerl/mims/software/spatial\\_allocator.html](http://www.epa.gov/asmdnerl/mims/software/spatial_allocator.html). Until this page is updated with the Spring 2003 version, the latest version is available at <http://www.emc.mcnc.org/projects/mims/spatial>. Note that the latter Web page may not be available after June 30, 2003. Future updates may be found at <http://www.cep.unc.edu/EMPD/>.

## BIBLIOGRAPHY

Houyoux, M. R.; Vukovich, J. M.; Coats, C. J. Jr.; Wheeler, N. M.; Kasibhatla, P. S. "Emission inventory development and processing for the Seasonal Model for Regional Air Quality (SMRAQ) project", J. Geophys. Res., 2000, 105, D7, pp 9079-9090.

Documentation index for SMOKE, CMAQ, and the EDSS/Models-3 I/O API;  
<http://www.cmascenter.org/doc.html>

EMS 2001 Information: [http://www.ladco.org/emis/ems\\_2001/index.html](http://www.ladco.org/emis/ems_2001/index.html)

ESRI Shapefile Technical Description: <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>, July, 1998.

Generic Polygon Clipper home page: <http://www.cs.man.ac.uk/aig/staff/alan/software/>.

MIMS home page: [http://www.epa.gov/asmdnerl/mims/software/spatial\\_allocator.html](http://www.epa.gov/asmdnerl/mims/software/spatial_allocator.html).

PROJ.4 Information: <http://www.remotesensing.org/proj>

Shapelib home page: <http://gdal.velocet.ca/projects/shapelib/>.

## KEY WORDS

Spatial Surrogates  
Emission Model  
Polygon Overlay

Shapefile  
Geographic Information System

## **ACKNOWLEDGMENTS**

This work was funded by the US Environmental Protection Agency's Office of Research and Development via a GSA task order.

## **DISCLAIMER**

The software described herein is provided as-is with no warranties. Users should quality assure any outputs from it using the provided quality assurance features. Any issues found may be reported to the authors.