Converting Gridded Inventories to ORL Format for Processing with SMOKE

Zachariah Adelman, Alison Eyth, Limei Ran
Center for Environmental Modeling for Policy Development
University of North Carolina Institute for the Environment
137 E. Franklin St. CB #6116
Chapel Hill, NC 27599-6116

Athanasios Sfetsos
Environmental Research Laboratory
Institute of Nuclear Technology and Radiation Protection
National Center for Scientific Research "Demokritos"
Patriarhou Grigoriou and Neapoleos Str
153 10, Agia Paraskevi, Greece

zac@unc.edu

ABSTRACT

To take advantage of the multi-source processing capabilities of SMOKE, we present an approach for converting gridded inventories to the ORL area source format using an open-source GIS. We demonstrate how to calculate administrative unit totals from a gridded inventory using the Spatial Allocator and then process the inventory through SMOKE. Advantages of converting gridded inventories to area source inventories include flexibility in how the data can be allocated to modeling grids and efficiency improvements in how the data are processed. We present graphical comparisons between gridded inventories on their native grids to the same data re-gridded with high-resolution population spatial surrogates. While we've observed some differences in the emissions totals using this approach, we present some ideas for future research to determine why these differences exist. This approach for extracting administrative unit inventories from gridded datasets is a way to make preparing gridded inventories for air quality modeling more accessible and efficient for areas of the world where only gridded emissions estimates are available.

INTRODUCTION

SMOKE (IE, 2008a) is optimized for processing source-based inventories, or datasets that contain information about specific classes of emissions processes at specific locations. Source-based inventories contain data about the locations where emissions occur, either by geographic coordinate or by administrative unit, such as a state or county. They also contain information about the type of emissions process using a numeric coding scheme like the source classification code (SCC). SMOKE uses the combination of location code and SCC to assign temporal variability, chemical speciation, spatial distributions, and projection information to an inventory. The use of unique location and source codes to represent different sources in an inventory allows SMOKE to process of inventories that contain multiple sources in one simulation. The implication is that different spatial distributions, temporal profiles, and chemical speciation profiles can be applied to many different types of sources in a single simulation. Figure 1 is a conceptual diagram of the SMOKE processing approach for area source inventories.

Figure 1. SMOKE area source inventory processing approach
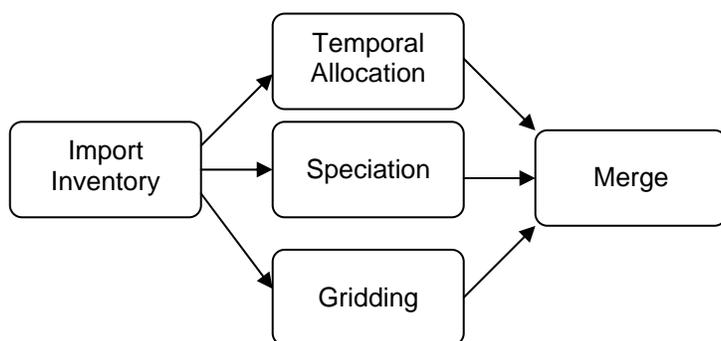
While SMOKE also supports the preparation of gridded emissions inventories for air quality modeling, there are several limitations to processing gridded emissions data. The efficiency of multi-source processing is not available with gridded inventories for a couple of reasons. First, gridded inventories typically contain emissions estimates for a single source. Second, if they do represent multiple sources, gridded inventories do not contain sufficient information to distinguish between the component sources, such as with SCCs and location codes/geographic coordinates. The result of this limitation is that a discrete SMOKE simulation must be performed for each gridded inventory in order to apply temporal profiles, chemical speciation profiles, and projection information to the inventory data. These profiles can only be applied uniformly across the entire model grid by pollutant (IE, 2008a). This approach is tedious when processing several gridded sources because it must be manually repeated for each source. Figure 2 illustrates the conceptual approach for processing gridded inventories with SMOKE. Notice how the gridding step is eliminated in this sequence and the gridded inventory that is input in the first processing step is a direct input to the merging process.

Figure 2. SMOKE pre-gridded inventory processing approach
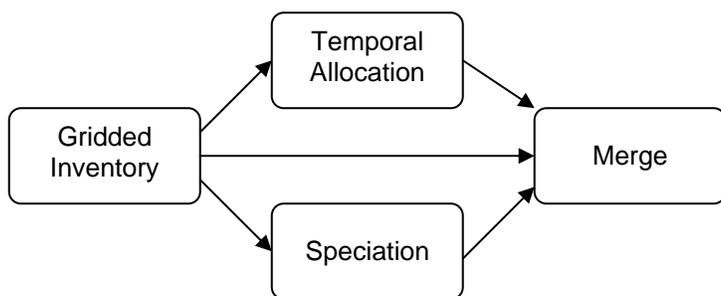
An additional limitation to using gridded inventory data is the constraint of the spatial resolution of the inventory. Common spatial resolutions of gridded inventories are on the order of 50 km – 200 km (Olivier et al., 2005; Vestreng et al., 2006). These resolutions are inconsistent with urban to regional scale air quality modeling applications, which are typically performed at 2 km – 50 km grid resolutions. High-resolution gridded inventories (<2-4 km) could be aggregated to urban/regional scales without sacrificing the spatial resolution of the data. For lower resolution gridded inventories, linear interpolation down to urban or regional scales is a common approach, but results in poorly resolved emissions locations (Streets et al., 2003). The incorporation of geographical data (e.g. land use / land

cover) data has proven an effective means of minimizing interpolation discrepancies (Winiwarter et al, 2003).

With source-based inventories, a Geographic Information System (GIS) is used to prepare spatial surrogates to allocate the emissions estimates to a variety of different grid resolutions, making them more suitable for urban to regional scale air quality modeling. This paper presents an approach for converting gridded emissions inventories to a source-based format in order to take advantage of the multi-source processing capabilities in SMOKE and to improve the spatial representation of the emissions data at regional and urban grid scales.

BODY

Methods

The conceptual approach for this work is to use a GIS to estimate emissions at the administrative unit (i.e. country, state, province, etc.) from a gridded inventory and convert the emissions values to a SMOKE-formatted area source inventory. The extraction of administrative unit inventories from gridded inventories has been documented by Streets et al. (2003) and Vestreng et al. (2006). The conversion of gridded data to area source inventories allows the emissions data from several gridded inventories to be combined to support multi-source and multi-pollutant processing with SMOKE. The inventory data can also be allocated to different modeling grids using spatial surrogates. Additionally, the proposed approach also has a prominent advantage of directly downscaling data from larger grids (e.g GEIA). A key objective of this work was to utilize only open-source software in order to provide free access to all of the tools needed to prepare gridded inventories for input to SMOKE.

The Spatial Allocator (SA) is an open-source Java-based GIS (IE, 2008b) that was developed to provide emissions modelers with a free tool to support common functions required in the preparation and analysis of emissions data for air quality modeling. One of the functions of the SA is the allocate mode, which allocates data from a gridded, polygon, or point file to an output grid or polygon file. In this application, a gridded netCDF inventory is input to the SA running in allocate mode and combined with a Shapefile of administrative unit boundaries to define the spatial attributes for allocating the emissions. The resulting Shapefile is a spatially weighted sum of the gridded emissions by administrative unit. As a final step, the output Shapefile is converted to a text file and post-processed to a SMOKE inventory format. Figure 3 shows the settings for the Spatial Allocator that we used to calculate administrative unit totals from the gridded inventories.

Figure 3. Spatial Allocator settings for allocating a gridded inventory to administrative unit totals

```
# Name and path of administrative unit Shapefile
OUTPUT_POLY_FILE      = $GISDIR/cntry02          # Shapefile name and location
OUTPUT_POLY_MAP_PRJN  = "+proj=latlong"          # Shapefile projection
OUTPUT_POLY_ELLIPSOID = "+ellps=WGS84"           # Shapefile ellipsoid
OUTPUT_POLY_ATTRS     = FIPS_CNTRY,CNTRY_NAME    # Attributes to allocate
OUTPUT_POLY_TYPE      = ShapeFile                # File type

# Gridded dataset being allocated
INPUT_FILE_NAME       = $NCFDIR/geiav1.NOx.ncf   # Input NCF name and location
INPUT_FILE_TYPE       = IoapiFile                # File type
INPUT_FILE_MAP_PRJN   = "+proj=latlong"          # NCF projection
INPUT_FILE_XCOL       = LONG                     # Spatial units in X
INPUT_FILE_YCOL       = LAT                      # Spatial units in Y
INPUT_FILE_ELLIPSOID  = "+ellps=GRS80"           # NCF ellipsoid
ALLOCATE_ATTRS        = NOX                      # Attributes to allocate
ALLOC_MODE_FILE       = ALL_AGGREGATE            # Allocation mode

# Name and path of resulting Shapefile
OUTPUT_FILE_NAME      = $NCFDIR/geia_global_nox  # Output file name and location
OUTPUT_FILE_TYPE      = Shapefile                # Output file type
OUTPUT_FILE_MAP_PRJN  = "+proj=latlong"          # Output projection
OUTPUT_FILE_ELLIPSOID = "+ellps=GRS80"           # Output spheroid
```

We developed methods to assign numeric codes to the different administrative units (country/state/county codes) and to assign SCCs to the inventory sources. We developed Fortran programs to convert the input ASCII gridded inventory data to a netCDF binary format (GEIA2NCF) for input to the SA and to convert the text file output by the SA to the ORL (CSV2ORL) area source inventory format. Figure 4 shows a schematic of the approach for converting gridded inventory data to ORL-formatted area source inventories.

Figure 4. Schematic of the gridded inventory preparation process
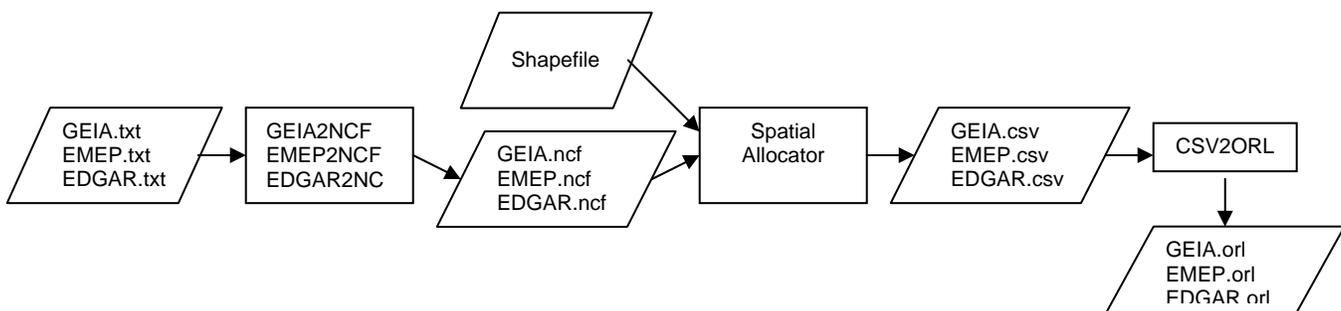
We tested this approach on three different gridded inventories. The Global Emissions Inventory Activity (GEIA) is a 1° resolution global inventory of greenhouse gases, ozone precursors, air toxics, and ozone depleting pollutants. The European Monitoring and Evaluation Program (EMEP) inventory is a global scale dataset that is provided as gridded, country total and sector total data (Vestreng et al., 2006). The Emissions Data for Global Atmospheric Research (EDGAR) 3.2 fast track 2000 is a global inventory of greenhouse gases, ozone precursors, and ozone depleting pollutants that is distributed as a 1° resolution gridded inventory and as country total inventories (Olivier et al., 2005). We used a year 2002 country boundaries Shapefile at a resolution of 1:1,000,000 meters as the spatial attribute for allocating the emissions data to administrative units.

We input the resulting ORL inventory from each gridded dataset to SMOKE to test the format compatibility of this approach. We used the SA to generate spatial surrogates using a 2.5 minute resolution, year 2005 global population data set (SEDAC, 2008) to spatially allocate the inventories to different model grid resolutions. We compared the inventories gridded in SMOKE with the global population surrogate to the inventories on their native grids to demonstrate the spatial allocation improvements of this approach.

Results

The conversion of the gridded data to the ORL area source format resulted in a loss of 5-10% of the domain total emissions. One possible explanation for the loss of emissions is that there are missing polygons in the administrative unit boundary Shapefile for locations where emissions exist in the gridded inventory. For example, shipping lane emissions in the gridded inventories were dropped because the administrative unit Shapefile that we used did not contain polygons for the marine areas. Another explanation for the loss of emissions is that a portion of the gridded emissions for coastal regions is also lost by not using a marine polygon for the inventory grid cells that straddle land-ocean boundaries.

All of the examples shown here use the GEIA original version 1 NOx inventories to illustrate the results from this approach for preparing gridded inventories for air quality modeling. Figure 5 is a graphic of the GEIA NOx inventory shown as country totals. This graphic illustrates the results of converting a gridded inventory to a Shapefile of country emissions totals.

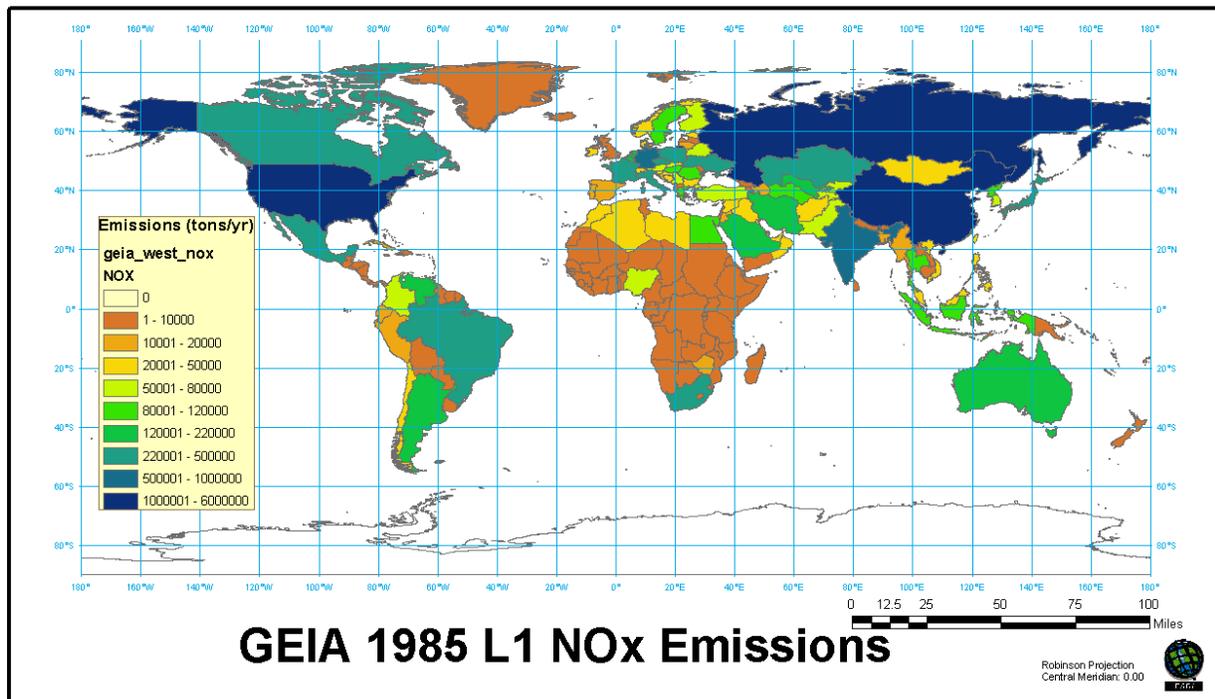**Figure 5.** Map of GEIA NOx emissions by country

Figure 6 shows the 2005 world population dataset used to develop the population spatial surrogate for allocating the gridded inventories to different model grids. The 2.5 minute gridded population data are suitable for allocating emissions data to grid resolutions as high as 1 km.

Figure 6. 2.5 minute resolution 2005 global population

SEDAC Gridded World Population v3 (2005)
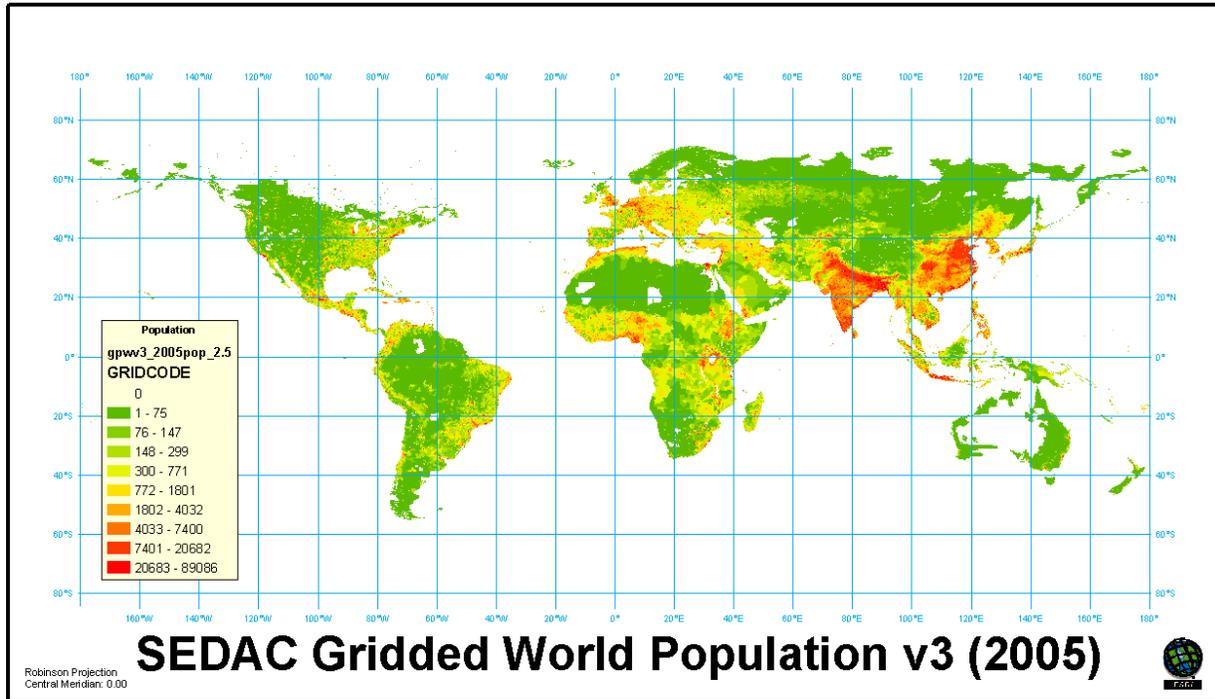
The following figures show the benefits of re-gridding the GEIA dataset for urban and regional scale modeling applications. Figures 7 through 9 are comparisons of the GEIA native 1° resolution NOx emissions with the GEIA emissions gridded with the population surrogate for different modeling grids. These figures illustrate the differences between using fine scale spatial surrogates to re-grid the emissions versus using the lower resolution gridded inventories.

Figure 7. GEIA US NOx emissions: native 1° grid (L), 12 km resolution population surrogate (R)
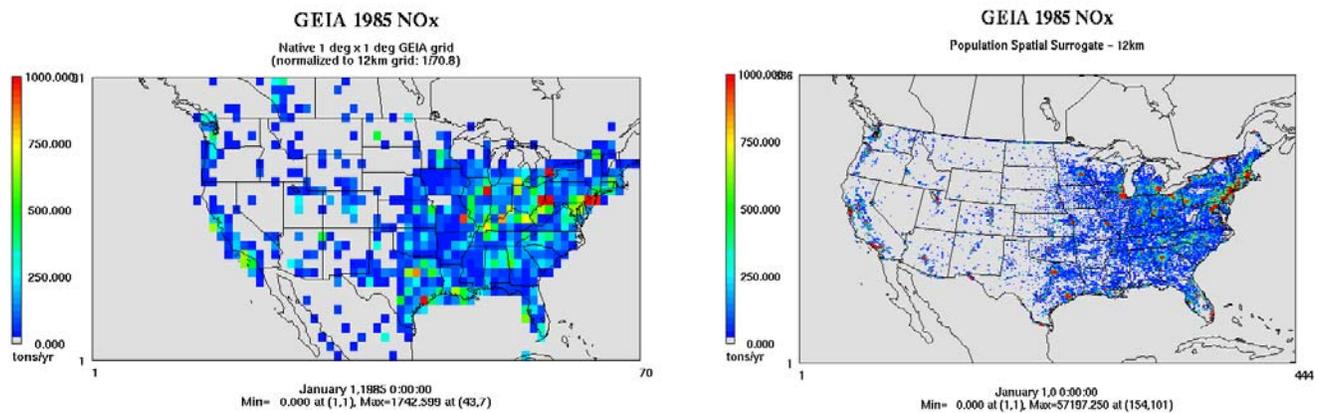
Figure 8. GEIA India NOx emissions: native 1° grid (L), 24 km resolution population surrogate (R)
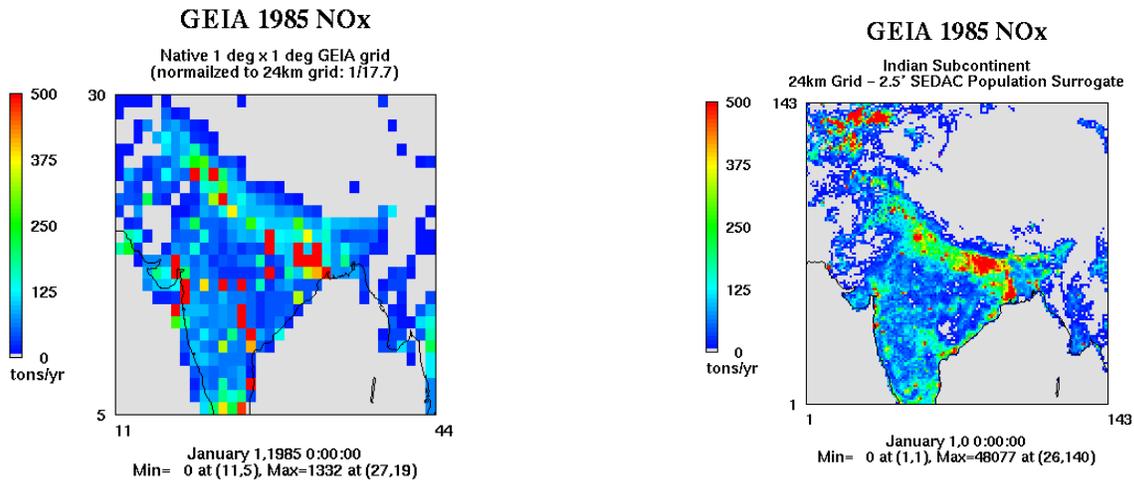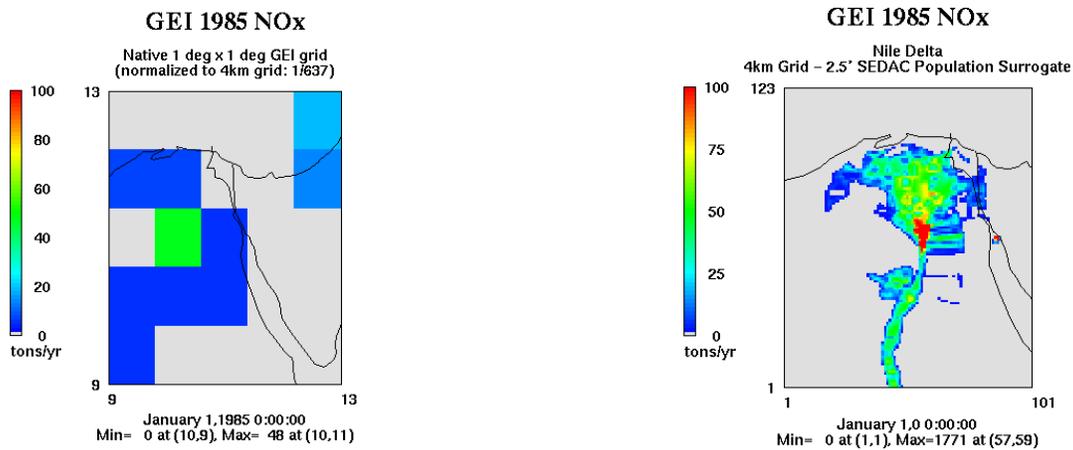

Figure 9. GEIA Nile Delta NOx emissions: native 1° grid (L), 4 km resolution population surrogate (R)

Discussion

Figures 7 through 9 demonstrate the improvements that this approach offers to the spatial representation of emissions derived from low-resolution gridded inventories. The emissions that are allocated to the example model grids using a population surrogate show more highly resolved spatial features around population centers than do the emissions on the native GEIA grid. Figures 7 and 8 compare emissions on the GEIA grid against lower resolution model grids. While there are clear difference in how well the emissions patterns are resolved when using the population surrogate to grid these data, the same general features appear in each pair of plots. Figure 9 illustrates a comparison of the GEIA gridded emissions to a very high-resolution model grid. The features of the Cairo metropolitan area and surrounding Nile Delta that are resolved by gridding the data with the population surrogate are not captured in the plot of the emissions on the native GEIA grid. These plots illustrate that the impacts of using this approach on air quality model results will become more significant at higher grid resolutions.

The EMEP and EDGAR emissions inventories are ideally suited for evaluating this approach because they are distributed as both gridded and country total datasets. Preliminary comparisons between the EMEP country totals and the country totals derived from the gridded EMEP inventory using the approach presented here showed differences on the order of 1-80% with an average difference across all of the EMEP countries of –6% (i.e. on average, 6% of the emissions were dropped by converting the gridded inventories to country totals). Further analysis is required to determine why there are differences between the EMEP country totals and the country totals derived from the gridded

inventories. One possible explanation is a difference in the definition of the country boundaries that we used compared to the EMEP country boundaries. We also used a fairly low-resolution country boundary Shapefile (1:1,000,000) to calculate the administrative unit totals. A higher resolution dataset may provide better fidelity with the EMEP country totals. One indication that this may be a cause of the discrepancy is that the countries that have the largest emissions losses from our approach either have long coastlines (i.e. Greece, Norway, Croatia) or they are small islands (i.e Malta, Cyprus). The low-resolution Shapefile cannot resolve the country boundaries relative to the gridded inventory and may result in a portion of the emissions for these countries being allocated offshore. Further analysis of the country totals derived from the gridded inventories and the country totals provided with these inventories is required to fully explain the observed differences.

CONCLUSIONS

This approach for extracting administrative unit inventories from gridded datasets is a way to make preparing gridded inventories for air quality modeling more accessible and efficient for areas of the world where only gridded emissions estimates are available. The resolution of the administrative unit for which emissions can be calculated is determined by the availability of Shapefiles. For example, a Shapefile of county boundaries can be used to calculate county total inventories, which then allow modelers a high level of control over how temporal profiles, chemical speciation, projection information, and spatial allocation are applied to the emissions. One issue with this approach is that the parts of the world that lack detailed inventory data also lack the detailed ancillary emissions data (i.e. temporal, chemical, spatial, etc). Even in the absence of detailed ancillary information, the conversion of gridded inventories to area source inventories provides a way for modelers to take advantage of the multi-source processing capabilities of SMOKE and it allows the flexibility of gridding inventories to any spatial resolution, regardless of the resolution of the original inventory data.

The availability of Shapefiles and other land use information will also determine the accuracy of downscaling the gridded inventories to fine-scale modeling grids. In the examples presented here we used a population surrogate to re-grid the emissions data to finer scale modeling grids. Additional land cover/land use data sets can be used to develop spatial surrogates for downscaling the gridded emissions. For example, the US Geological Survey 1 km resolution global landcover characterization project provides high resolution land cover data for the entire globe (http://edcsns17.cr.usgs.gov/glcc/). High-resolution land cover surrogates, such as agricultural, urban, rural, water, and forest, can be derived from this dataset for downscaling global gridded emissions inventories to fine-scale modeling grids.

REFERENCES

Institute for the Environment (IE), *Sparse Matrix Operator Kernel Emissions system User's Guide*, Available at http://www.smoke-model.org. Chapel Hill, NC, 2008a.

Institute for the Environment (IE), *Spatial Allocator User's Guide*, Available at http://www.ie.unc.edu/cempd/projects/mims/spatial. Chapel Hill, NC, 2008b.

Olivier, J.G.J., Van Aardenne, J.A., Dentener, F., Ganzeveld, L. and J.A.H.W. Peters, "Recent trends in global greenhouse gas emissions: regional trends and spatial distribution of key sources". *Non-CO$_2$ Greenhouse Gases (NCGG-4)*, Millpress, Rotterdam, 2005, ISBN 90 5966 043 9, pages 325-330.

Socioeconomic Data and Applications Center (SEDAC), Gridded population of the world version 3 (GPWv3). Available at http://sedac.ciesin.columbia.edu/gpw/.

Streets, D. G., et al., "An inventory of gaseous and primary aerosol emissions in Asia in the year 2000," J.Geophys. Res., 2003, 108(D21), 8809, doi:10.1029/2002JD003093.

Vestreng, V., Rigler, E., Adams, M., Kindbom, K., Pacyna, J. M., Denier van der Gon, H., Reis, S., and Travnikov, O., "Inventory review 2006, Emission data reported to LRTAP and NEC Directive, Stage 1, 2 and 3 review and Evaluation of inventories of HM and POPs". Prepared for EMEP/MSC-W Technical Report 1/2006 ISSN 1504-6179, 2006.

Winiwarter, W., Dore, C., Hayman, G., Vlachogiannis, D., Gounaris, N., Bartzis, J.G., Ekstrand, S., Tamponi, M., Maffeis G., "Methods for comparing gridded inventories of atmospheric emissions—application for Milan province, Italy and the Greater Athens Area, Greece," The Science of The Total Environment, Volume 303, Issue 3, 1 March 2003, pages 231-243

KEY WORDS

SMOKE, emissions, gridded inventories, global emissions, Spatial Allocator