Accounting for Land Use Changes in Projecting Future-Year Emissions Scenarios

Alison Eyth, Limei Ran, and Zac Adelman
Institute for the Environment, University of North Carolina at Chapel Hill
137 E. Franklin St., Chapel Hill, NC 27599-6116
eyth@unc.edu

David Theobold, Ph. D.
Natural Resource Ecology Lab, Colorado State University
Fort Collins, Colorado, 80523-1499, USA
davet@nrel.colostate.edu

ABSTRACT

Air quality modeling is performed at EPA for future-year time periods in support of regulatory analyses to assess the likelihood that areas will be in attainment with the National Ambient Air Quality Standards (NAAQS) after the application of proposed emissions controls. Typical years that are modeled include 2009, 2014, 2020, and 2030. Currently, changes in land use are not considered when the emissions for these future-year scenarios are generated. A premise of our research is that representing the impacts of land use changes on the geographic allocation of emissions may provide more realistic modeling results, especially in areas in which rapid development is underway. In this paper, we investigate using population and housing projections from the Integrated Climate and Land Use Change Scenarios (ICLUS) project, which includes outputs from the Spatially Explicit Regional Growth Model (SERGoM), to spatially allocate future-year emissions. We compare the results of this approach with a baseline set of emissions and present observations and preliminary conclusions regarding the utility of this approach for modeling applications involving different grid sizes.

INTRODUCTION

Future-year air quality is frequently modeled by both the EPA's Office of Air Quality Planning and Standards and Office of Research and Development. Examples of reasons for future-year air quality modeling include evaluating the impact of control strategies when performing the Regulatory Impact Analyses for proposed regulations. Future years that are commonly modeled include 2014, 2020, and 2030. States also perform future-year modeling when developing state implementation plans (SIPs) to demonstrate their projected compliance with existing regulations. One potential weakness of current modeling techniques is that location of emissions sources, and sometimes even the emissions amounts for some source types, are not updated when modeling future years. Population growth and development patterns make it unrealistic to assume that emissions in future years will be spatially distributed according to the same pattern that they are in the current day. Evidence of urbanization and suburban sprawl increasing is all around us. With these changes, residential, commercial, and transportation-related emissions likely will increase in the newly developed areas over current day levels, even if process improvements are successful at reducing per-capita emissions. As the residential and commercial emissions increase, agricultural and biogenic emissions in the newly developed areas will decrease.

The methods for spatially allocating emissions vary based on the type of emissions source. The emissions, as described in the emission inventories, need to be placed into the rectangular grid cells used by an air quality model such as the Community Multiscale Air Quality (CMAQ) model. Emissions processing systems such as the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system are used for this purpose. Point sources are allocated by determining the grid cell containing the location indicated by the x and y coordinates in the emissions inventory. Biogenic sources are allocated based on land use data. Area, nonpoint, and nonroad mobile sources are allocated using spatial surrogates, which

are discussed in more detail below. Onroad mobile sources are allocated using spatial surrogates for regional scale modeling, or road links for urban scale modeling.

Opportunities to refine the spatial allocation of emissions for future-year modeling exist if the spatial surrogates and land use data can be updated based on information about the location and amount of growth expected. The goals of this study are to: (1) define procedures by which the spatial allocation of emissions in future years can be made more realistic, and (2) assess the impact that updating the spatial allocation of emissions would have on future-year air quality model results.

APPROACH

To spatially allocate emissions in future years more realistically, it is important to anticipate where growth in population and housing will occur. The EPA's Integrated Climate and Land Use Scenarios (ICLUS) project provides such information. ICLUS includes two primary components: an algorithm that projects population growth at the county-level, and the Spatially Explicit Regional Growth Model (SERGoM), which uses population data to project housing density.

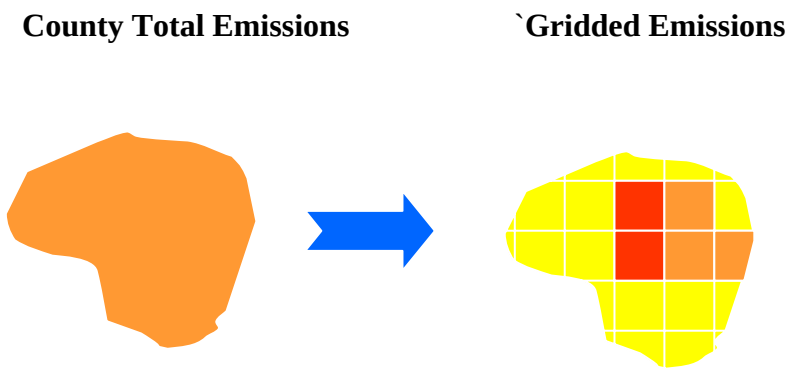
The ICLUS population forecasts grow county-level population using a cohort model that takes into account demographics and associated birth and death rates. Within the 2007 calendar year, the ICLUS population projections are expected to be updated, accounting for population migration. SERGoM estimates future-year housing density nationwide in 100m x 100m pixels across multiple decades. SERGoM inputs include housing data for the current time-step, future-year population forecasts by county, road networks, water bodies, local growth rates, and information about developable and undevelopable land. This information is then used to develop estimates of housing density at the next time step (Theobold, 2006). Typically, SERGoM is run at 10-year intervals to provide estimates of housing density each decade. The information about future-year housing density can be used to update the spatial surrogates related to housing that are used by SMOKE.

An evaluation of how the spatial surrogates are mapped to the emissions inventory, however, revealed that only a very small amount of the inventory emissions are currently allocated using housing-related surrogates. A much larger fraction of the emissions are allocated using population-related surrogates. Since county-level population projections are inputs to SERGoM, this information can also be used in developing population-related surrogates by using SERGoM housing outputs are used to map population data to grid cells. In addition, the onroad emissions are allocated according to surrogates that consider the urban and rural boundaries. These boundaries can be estimated in future years from SERGoM outputs by analyzing population densities.

The Role of Spatial Surrogates in Air Quality Modeling

Spatial surrogates are used to allocate county level emissions that appear in emission inventories into the rectangular grid cells used by urban and regional scale air quality models. Spatial surrogates are based on data at resolutions different from county-level, such as census tracts or road locations, and can therefore be used to allocate the emissions more specifically than assuming they are uniformly distributed through out the county. For example, motor vehicle emissions from interstate highways can be placed into the grid cells that intersect the highways themselves, and dry cleaning emissions can be allocated to parts of the county that have higher population, or the parts that have more dry cleaners – if data is available on specific locations of dry cleaners. Spatial surrogates consist of values between 0 and 1 that specify the fraction of the county emissions that should be allocated to each grid cell that intersects the county. The spatial surrogate fractions for each county sum to 1, except the sum may be less than 1 for some counties that intersect the edge of the modeling domain. The emissions for a specific county within a grid cell are computed by multiplying the surrogate fraction for that county and grid cell by the total emissions for the county. A diagram illustrating how spatial surrogates impact the emissions levels within the grid cells of a county is shown in Figure 1.

Figure 1. The impact of a spatial surrogate on emissions allocation within a county.



Currently, EPA OAQPS uses 65 different spatial surrogates for their modeling. Some examples of these surrogates are: population, urban population, housing, urban primary road miles, and rural secondary road miles. EPA computes the spatial surrogates using the Surrogate Tool, a free tool that is a component of the Emissions Modeling Framework, based on data in about two-dozen shapefiles. Within the 65 surrogates is a set of five surrogates based on the use of different fuels for residential heating, and another set of over 17 surrogates that are based on building square footages, broken down into several types of residential, commercial, institutional, and industrial uses within each census tract. The 65 surrogates must be generated separately for each modeling grid.

Selection of the Study Domain and Input Data Sets

Currently, the same surrogate datasets are used for present-day and future-year modeling. Therefore, no changes due either to population growth and migration or to housing changes are considered when spatially allocating emissions in future years. While this strategy simplifies emissions modeling, it ignores potentially important redistributions of emissions, which in turn can affect air quality.

During this project, we examined using SERGoM to update spatial surrogates for future-year modeling. To illustrate the potential impacts on a previously studied case, we selected a modeling domain modeled by the North Carolina Department of Environment and Natural Resources (NC DENR). A map of this modeling domain is shown as the shaded area in Figure 2. The major urban areas covered in this study in North Carolina are the Raleigh-Cary-Durham-Chapel Hill region, Greensboro and Winston-Salem, Charlotte, Fayetteville, and a portion of Wilmington. The modeling domain also includes the Greenville-Spartanburg area of South Carolina. We studied this domain using grids comprised of 4km x 4km grid cells, as well as 12km and 36km cells.

SERGoM output data for the years 2000, 2010, 2020, and 2030 was supplied in ArcGIS grid format by Dr. David Theobold of Colorado State University, along with the projected county-level population data that was used as input to SERGoM for each of the study years. We compared the 2000 population data to independently obtained census data and found minor differences of less than 1%. We also noted that population data was missing for 12 counties in Virginia that were within our modeling domain. We filled in the data for these twelve counties with our independently obtained 2000 census data. The shapefiles used as the basis for creating the EPA spatial surrogates provided the other major source of input data for this project (www.epa.gov/ttn/chief/emch/spatial/newsurrogate.html). Plots of the 2000 and 2030 housing density data output from SERGoM are shown in Figures 3 and 4. In these plots, the red portions represent urban areas (< 0.6 acres / unit), orange are suburban (0.6-1.7 acres/unit), yellow are exurban (1.7-40 acres/unit), light green are rural (> 40 acres/unit), and white and dark green are undeveloped private land and water bodies. Notice the enlargement of the red and orange urban and surburban areas, and the relative disappearance of the light green rural areas over the thirty years.

Figure 2. The modeling domain used for this study

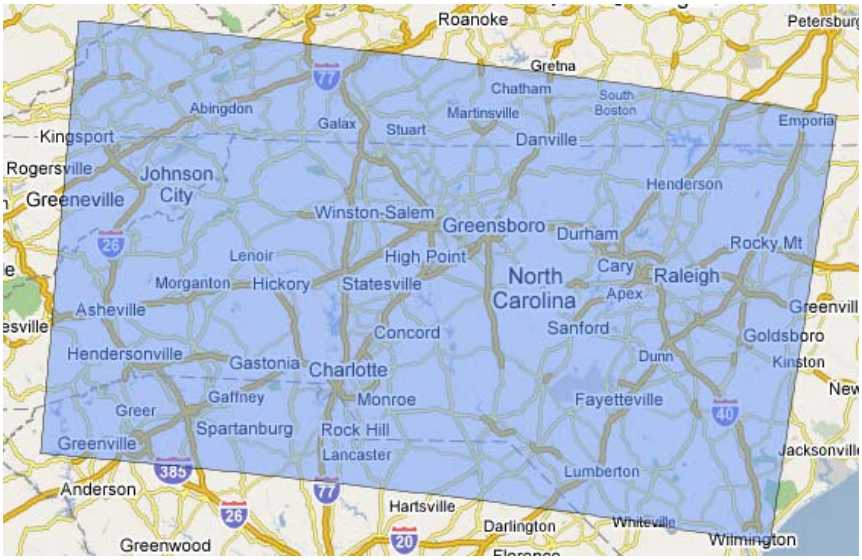


Figure 3. SERGoM Housing Density Outputs in the Year 2000

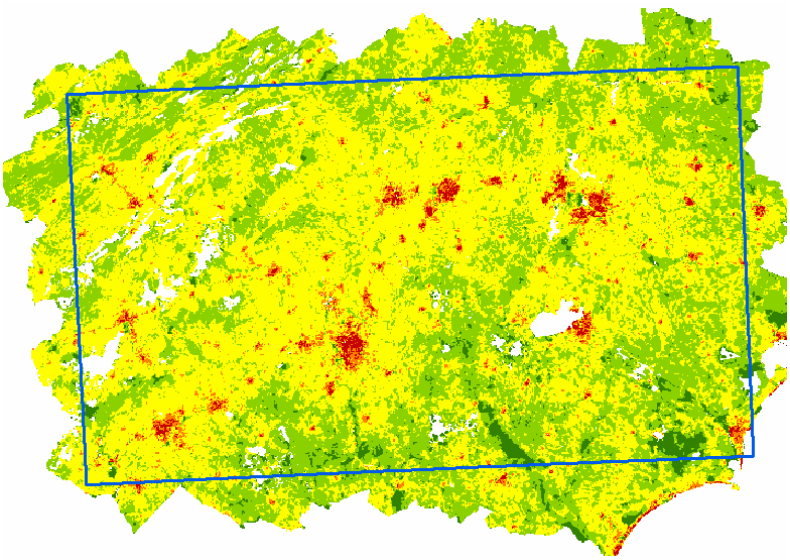
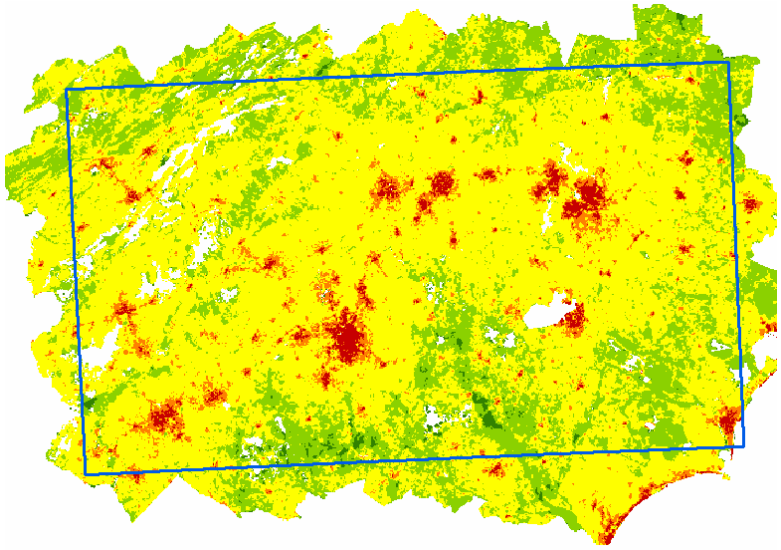


Figure 4. SERGoM Housing Density Outputs in the Year 2030



Using SERGoM Outputs to Create Input Files for Spatial Surrogate Generation

As preparation for creating updated spatial surrogates, the goal was to use the SERGoM outputs to create shapefiles that contained housing units, population counts, and urban-rural boundaries for at least the years 2000 and 2030. These shapefiles could then be input to the Surrogate Tool to create revised spatial surrogates that could be input to SMOKE to create gridded emissions. We would then be able to compare the gridded emissions for 2000 with those from 2030 to determine the impact of incorporating the population and housing growth information into the spatial allocation process. The process used to develop the new shapefiles was the following:

1. A shapefile that represented the modeling domain as a 4km grid using its Lambert Conformal conic map projection and a spherical earth was developed using ArcInfo.
2. The grid shapefile and the US county shapefile were projected into the Albers map projection with a NAD83 datum, which was the map projection of the SERGoM output data.
3. Using ArcMap, the US county shapefile in the Albers projection was clipped using the Albers projection shapefile for the modeling domain. This resulted in a shapefile that contained only the counties that intersected the modeling project grid. A plot of this shapefile that also shows the difference between the SERGoM population data for 2000 and US census data for 2000 is shown in Figure 5.
4. The shapefile of counties overlapping the modeling domain output from Step 3 was used to clip the housing density files output from SERGoM. This resulted in 100m raster files for the counties that overlapped the modeling domain.
5. Because the Surrogate Tool can take only shapefiles as input, the raster files were converted into polygon-based shapefiles of housing density in an Albers projection. This resulted in a shapefile that contained regions of constant housing densities.
6. The housing density shapefiles in the Albers projection using the NAD83 datum were re-projected into a Lambert conformal conic projection with a spherical earth so that they could be input to the Surrogate Tool, which can handle various map projections, but requires input files to use a Spherical earth.
7. The population data used as input to SERGoM was incorporated into the county shapefile as attributes, and the resulting shapefile was converted back into the Lambert conformal projection using a spherical earth.
8. The total housing units in each county were determined by aggregating the housing density over the county using an ArcGIS statistics function. This allowed us to compute the fraction of the housing in the county that was found in each county subregion.
9. In order to compute the population of each area within a county, the county shapefile of the study region was overlaid with each housing density shapefile using the ArcGIS Identity command. Housing units in each area were computed using the equation: GRID-VALUE * area in square meters / 1000000.0. The division was needed because the housing density grid files had values equal to housing units * 1000 per square hectare.
10. The population for each subregion within the county was then computed using the function: population[subregion]=county population * housing units [subregion] / county housing units. The subregion population density per square mile was then computed using the function: population density [subregion] = (population / subregion area) * 100000000.0 / 3.86102159 .Note: 1 square meter=3.8610x10E-7 square miles. As a result, we now have a shapefile that contains both housing units and population for each subregion within each county. Plots of the computed population density in 2000 and 2030 are shown in Figures 6 and 7, respectively.

Figure 5. SERGoM county population minus 2000 US census population for counties overlapping the modeling domain

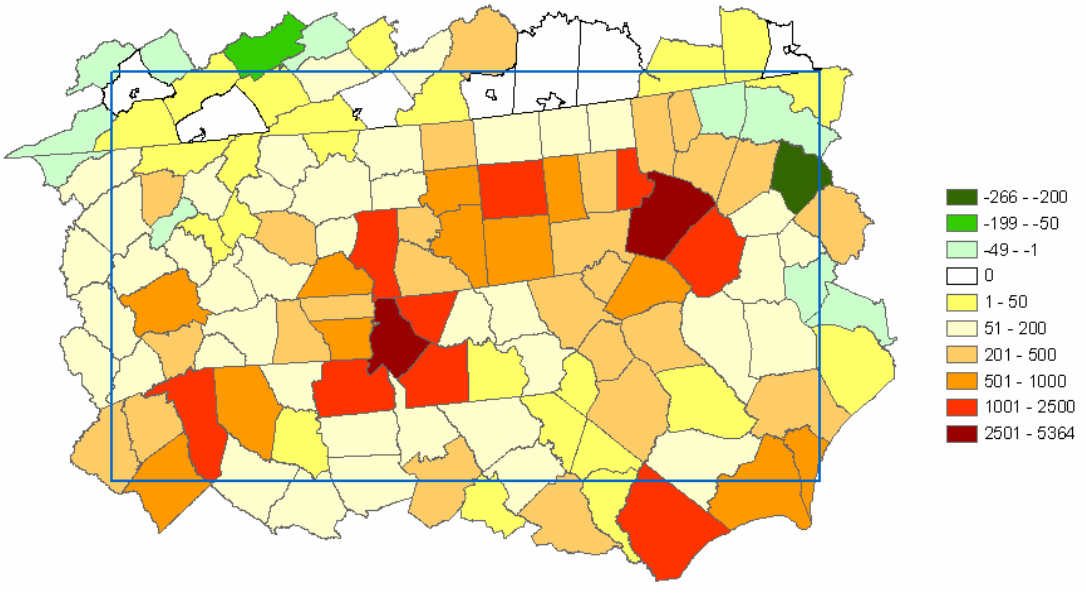


Figure 6. Computed population density for the year 2000 (people/square mile)

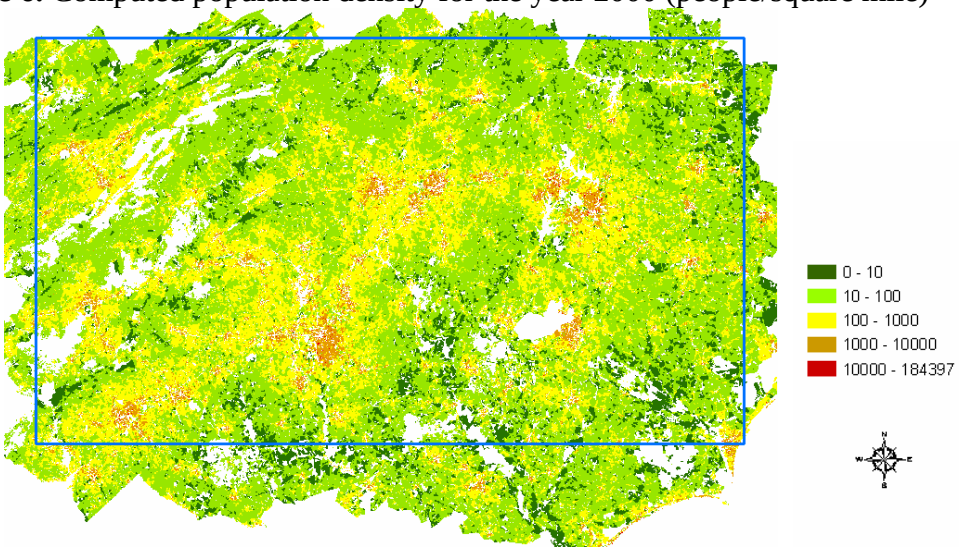
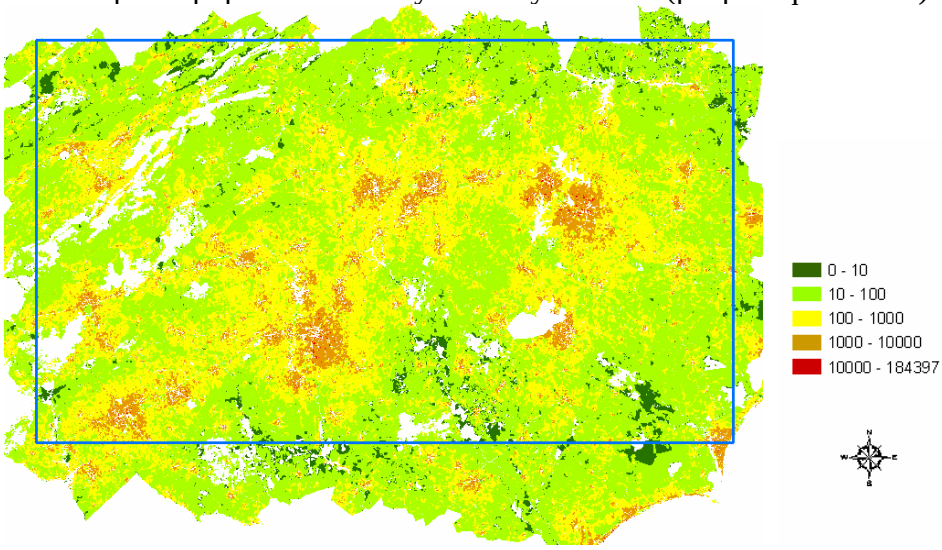


Figure 7. Computed population density for the year 2030 (people/square mile)



With the housing and population shapefiles now available, we still would like to revise the urban area boundaries based on the changing population density. A new shapefile containing the boundaries of urban areas in 2000 and 2030 was derived by classifying as urban any subregions of counties for which the population density exceeded 1000 people per square mile. This was consistent with the US Census Bureau's urban definition, in which each urban area has a population density generally exceeding 1000 people per square mile with a surrounding densely settled territory that has a total population of 50,000 or more. We did not take any currently defined city boundaries into account when we classified subregions as urban or rural. Plots of the computed urban boundaries in 2000 and 2030 are shown in Figures 8 and 9, respectively. Now with the urban boundaries available, it was possible to compute the values for an urban population attribute and add it to the population shapefile. The revised urban boundaries were also overlaid with 2000 road network shapefile to revise the classifications of the roads into urban and rural categories. The resulting shapefile was used as an input to the Surrogate Tool. *Note that we did not incorporate any changes to the road network itself, but planned changes to the road network could be incorporated for increased accuracy in the future year*.

Figure 8. Computed urban areas in 2000

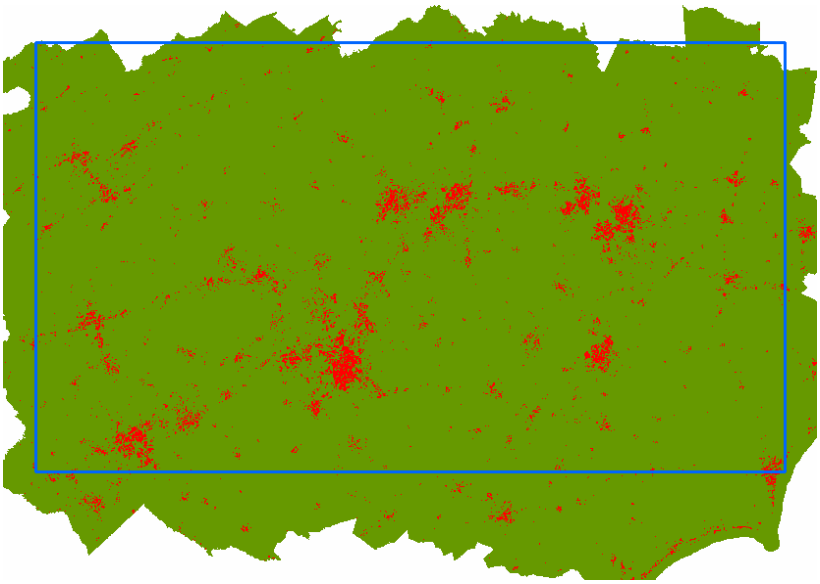
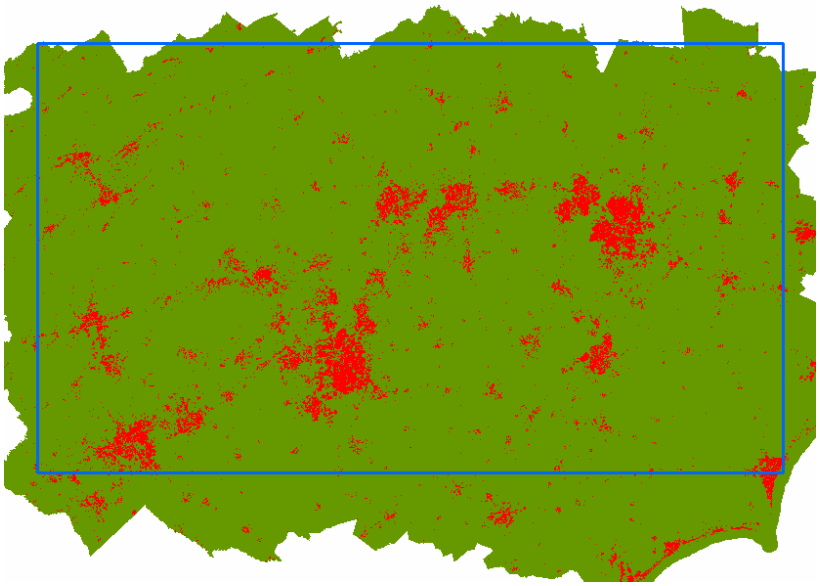


Figure 9. Computed urban areas in 2030



Computing Revised Surrogates for 2000 and 2030

The Surrogate Tool (see http://www.epa.gov/ttn/chief/emch/spatial/spatialsurrogate.html for more details) was used with the revised shapefiles to generate 11 revised spatial surrogates based on the shapefiles generated from SERGoM data. These calculations were performed for 4km, 12km, and 36km grids that covered the modeling domain. The revised surrogates are listed in Table 1. The standard 65 spatial surrogates used by EPA were also generated for each grid, so that a complete set of surrogates was available as input to SMOKE. SMOKE was then used to spatially allocate the emissions on the 4km, 12km, and 36km grids.

Tab le 1. Surrogates that were revised based on SERGoM outputs

SURROGATE NAME	SURROGATE CODE
Population	100
Housing	110
Urban Population	120
Rural Population	130
Urban Primary Road Miles	200
Rural Primary Road Miles	210
Urban Secondary Road Miles	220
Rural Secondary Road Miles	230
Total Road Miles	240
Urban Primary plus Rural Primary	250
0.75 Total Roadway Miles plus 0.25 Population	255

Several of the revised surrogates for 2000 and 2030 were compared with one another to evaluate the impact of accounting for the growth in population and changes in housing density on the surrogates. The percent difference in the population surrogate between 2030 and 2000 is shown for the 4km and 12km grids in Figures 10 and 11, respectively. Substantial differences can be seen at the 4km resolution. Note that when the surrogate fractions in one part of a county increase, they must decrease in another part of the county in order for the sum of the fractions to remain equal to 1. For example, in the Raleigh area, shifting of the relative allocation of emissions from the original urban area to new suburban and urban portions of the county is evident. The changes between 2000 and 2030 are not as noticeable at the 12km resolution. At both resolutions, most of the changes are between -25% and +25%, but larger differences are noted in some areas. The 36km differences are not shown because they are less evident than those at the 12km resolution.

The differences in the urban population surrogate are more pronounced, due to some grid cells changing from rural to urban. The road surrogates computed from the new road classes defined by new urban and rural areas also showed similar change patterns as those shown for the urban population surrogate.

Figure 10. Percent difference for the population surrogate at 4km (2030-2000)

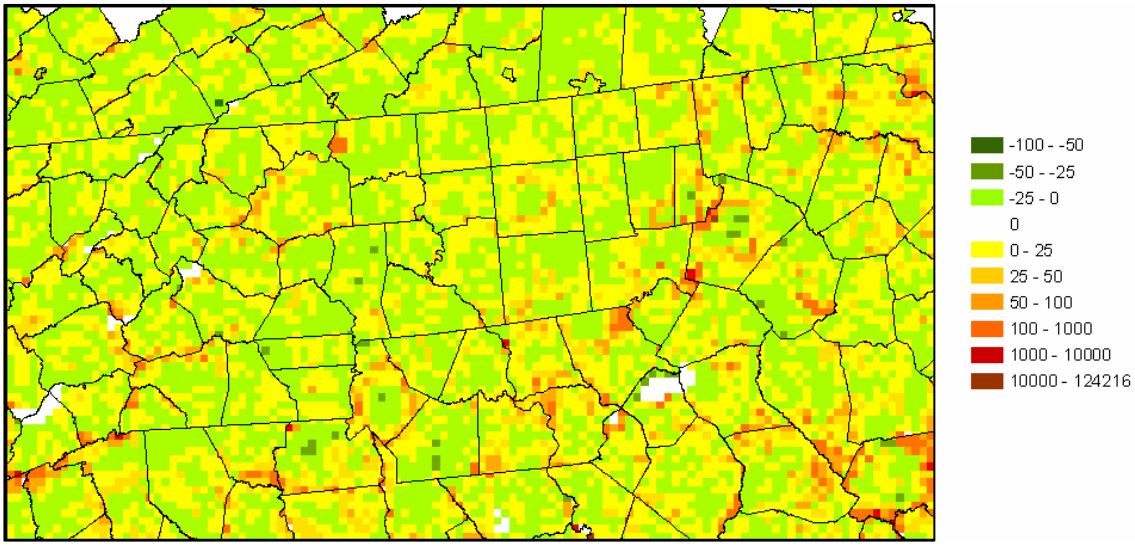


Figure 11. Percent difference for the population surrogate at 12km (2030-2000)

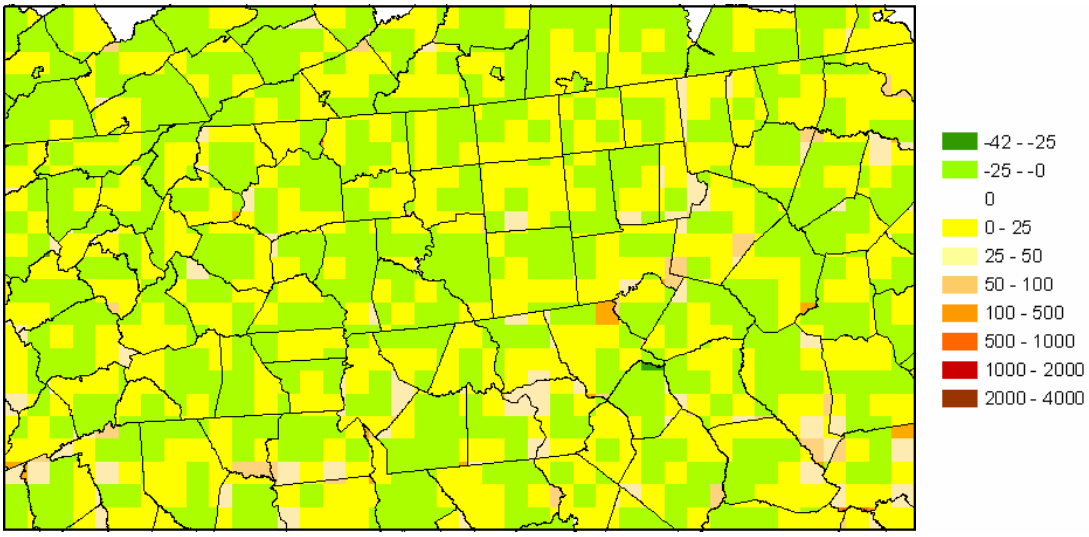


Figure 12. Percent difference for the urban population surrogate at 4km (2030-2000)

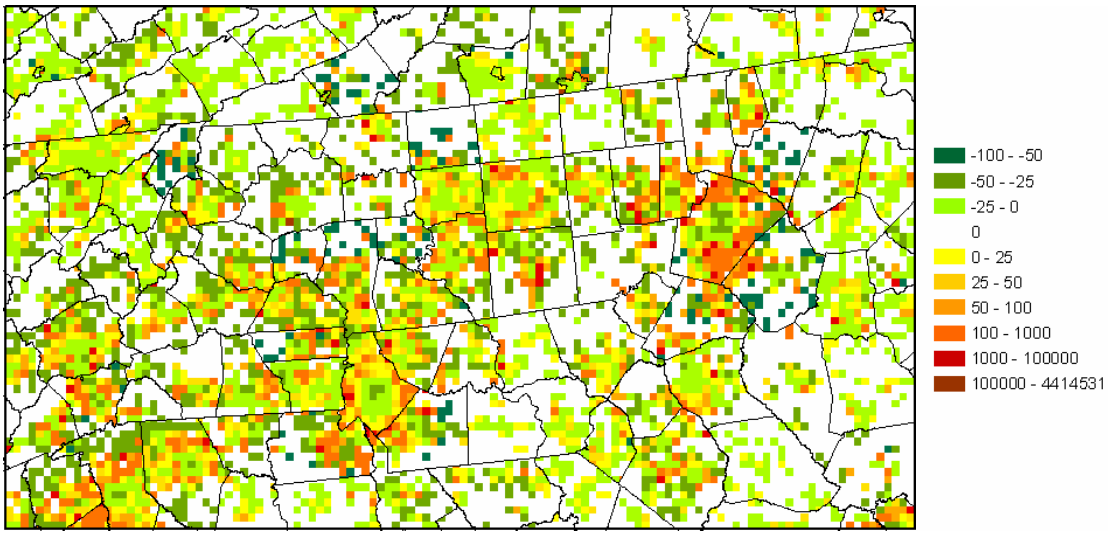
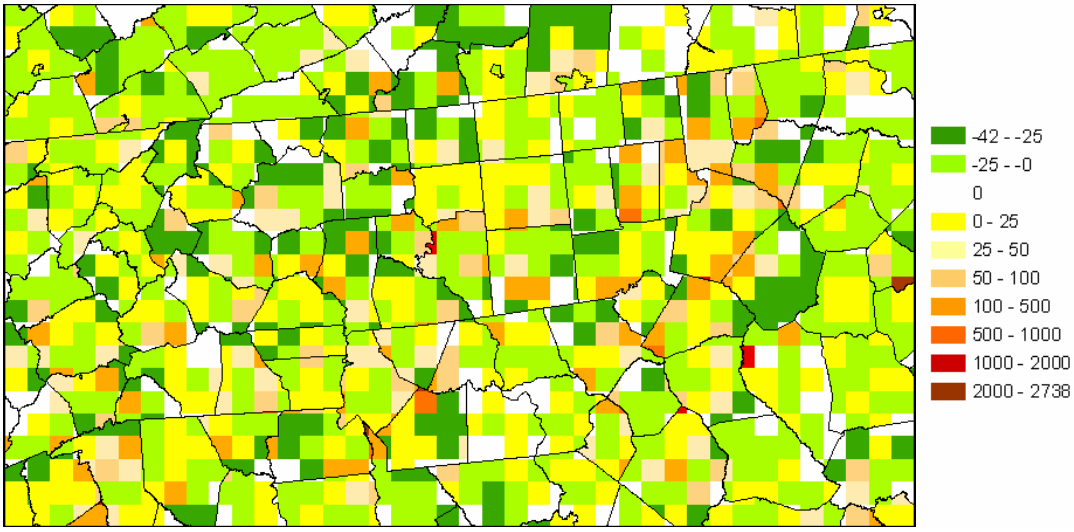


Figure 13. Percent difference for the urban population surrogate at 12km (2030-2000)



Emissions Processing

The 2002 EPA National Emissions Inventory (NEI) area and mobile source inventories were processed using SMOKE using the adjusted surrogates for 2000 and 2030.(http://www.epa.gov/ttn/chief/eiinformation.html) *Smkinven* was used to import the inventories, and *Grdmat* was used to grid the emissions. Chemical speciation and temporal processing were not performed because we were concerned with the spatial allocation of emissions. Reports of the gridded, unspeciated, annual emissions were created for inventory sector-pollutant combinations that made up a substantial component of the inventory, based on our analysis. No emissions growth was included in this analysis.

The 2002 NEI was evaluated to identify major source categories and their relationship to various surrogates. A summary of the area and mobile sources components of the 2002 NEI is shown in Figure 14. Notice the large contribution of onroad mobile source emissions to the total NOx, and the large contribution of nonpoint area sources to the total PM-related emissions. Figures 15, 16, and 17 show the surrogates that are most used for these pollutant-emissions category combinations. From Figure 15, all major surrogates have been updated using the outputs from SERGoM. From Figure 16, that the population surrogate has been updated, but the residential heating and other major surrogates were not. In Figure 17, the rural population surrogates was updated, but the others were not.

Figure 14. Summary of the 2002 criteria pollutant inventory

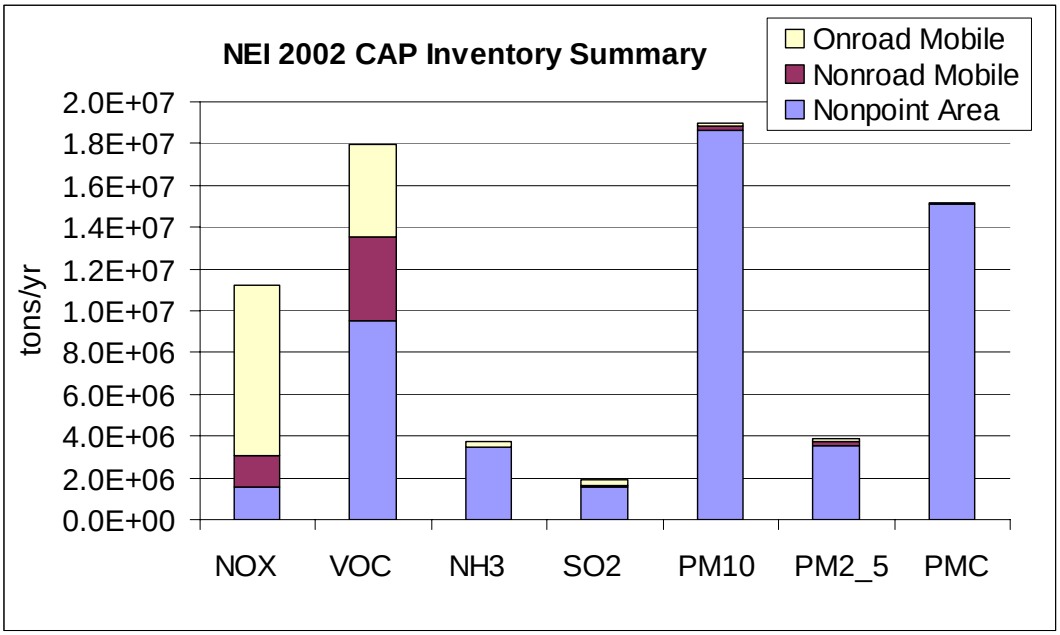


Figure 15. Major spatial surrogates used for onroad mobile NOx

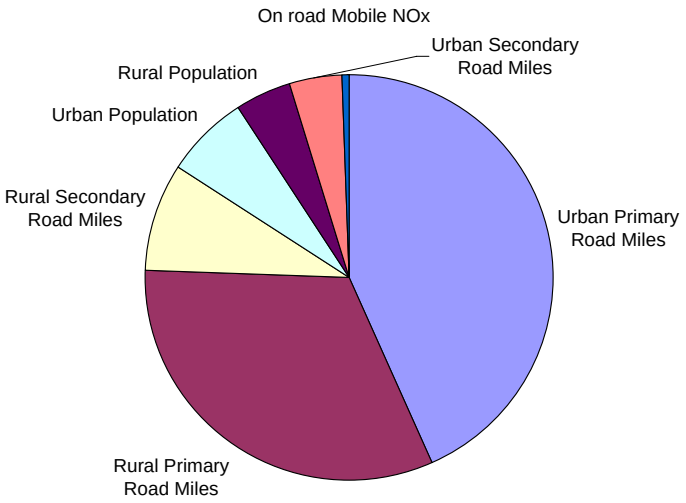


Figure 16. Major spatial surrogates used for stationary area VOC

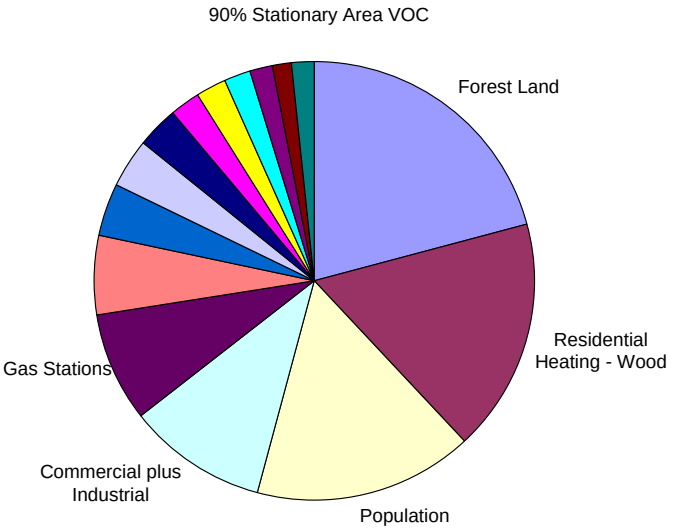
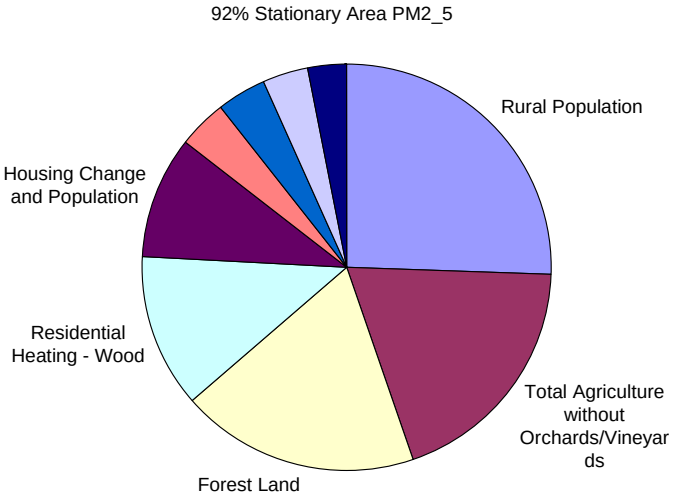


Figure 17. Major spatial surrogates used for stationary area PM2.5

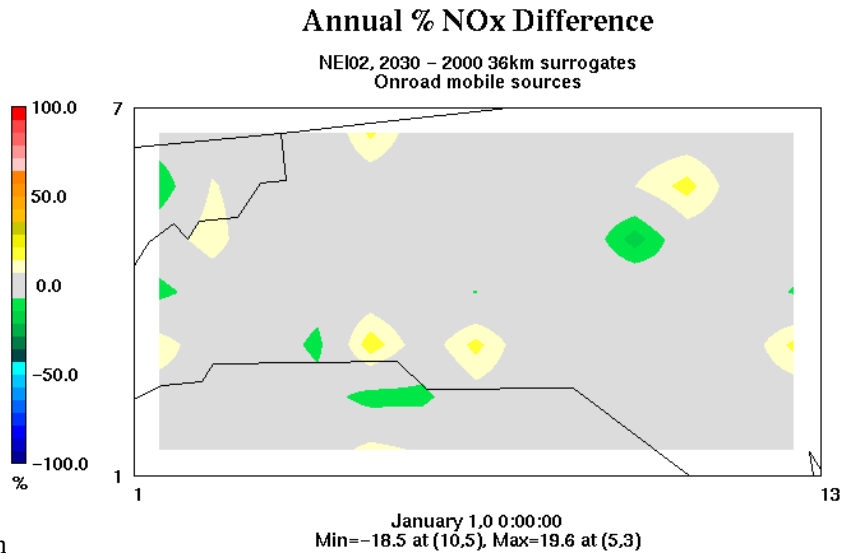


Changes in Emissions Resulting from Including Housing and Population Growth

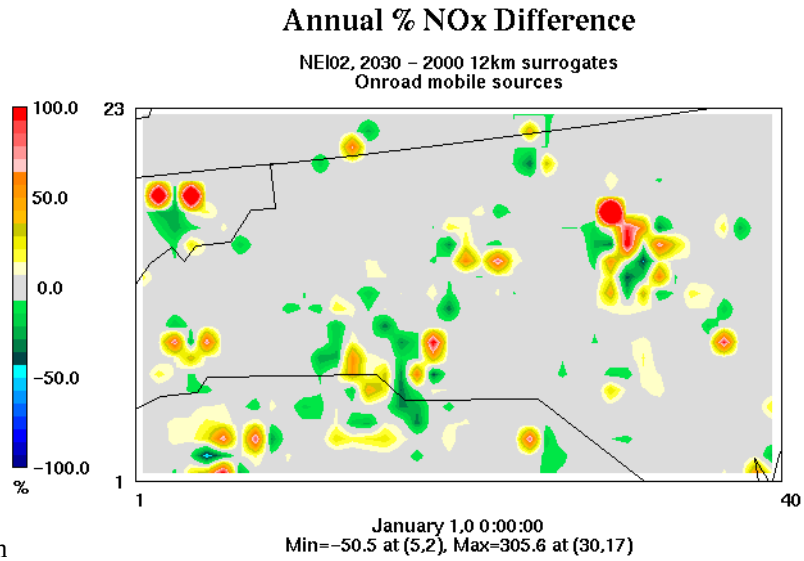
Once the emissions were processed through SMOKE, summary plots of the difference in emissions, from 2000 and 2030, for key sectors were created. The onroad mobile NOx emissions at 36km, 12km, and 4km are shown in Figure 18. The plots for stationary area VOC emissions are shown in Figure 19. The plots for stationary area PM 2.5 emissions are shown in Figure 20. Notice that the scale for Figure 18 is much broader than the scales for Figures 19 and 20. The impact of the updated surrogates on the onroad mobile NOx emissions was larger because all of the major surrogates used for those emissions were updated based on SERGoM outputs, whereas only one of the major surrogates was updated for the other pollutant-sector combinations. In each of these figures, we notice that the impact of updating the surrogates becomes more pronounced at finer grid scales. Substantial differences in emissions can be seen at the 4km resolution, moderate changes can be seen at the 12km resolution, and minimal changes can be seen at the 36km resolution. For example, the downtown Raleigh area has substantially less emissions at the 4km resolution in Figures 18 and 19. Currently, EPA typically models at 36km and12km resolutions, while states often perform their modeling over smaller regions at a finer resolution.

In Figure 20, the changes to stationary area source VOC emissions are due to the updates in the rural population surrogate. The differences between the level of use of the rural population surrogate in North Carolina and South Carolina become quite apparent at the 4km resolution. It should be noted that there was a spatial distortion that resulted from the changing of map projections, which displaced the North Carolina-South Carolina border by approximately 20km from where it should be. Recall that in this study, the housing density data from SERGoM was in the Albers projection with a NAD83 geographic datum. However, the Surrogate Tool and SMOKE can only support a spherical earth. Also, all EPA surrogate shapefiles are in the sphere geographic projection. During the data processing, the county shapefile had to be projected into the Albers with a NAD83 geographic datum in order to clip the housing density data. After the county shapefile was projected back to the spherical earth, the coordinates were shifted towards southeast. When generating surrogate shapefiles, we suggest using the same sphere projection for all shapefiles because the surrogate tool, SMOKE, and CMAQ only support a spherical earth. The Surrogate Tool could be updated to support different datums, but it is unlikely that SMOKE and CMAQ would be updated to support this. If the coordinate distortion is large, it can be partially corrected using control points based on coordinate transformation functions. It may also be worth studying alternative approaches for changing the map projection, with the hope that one of them will result in less distortion.

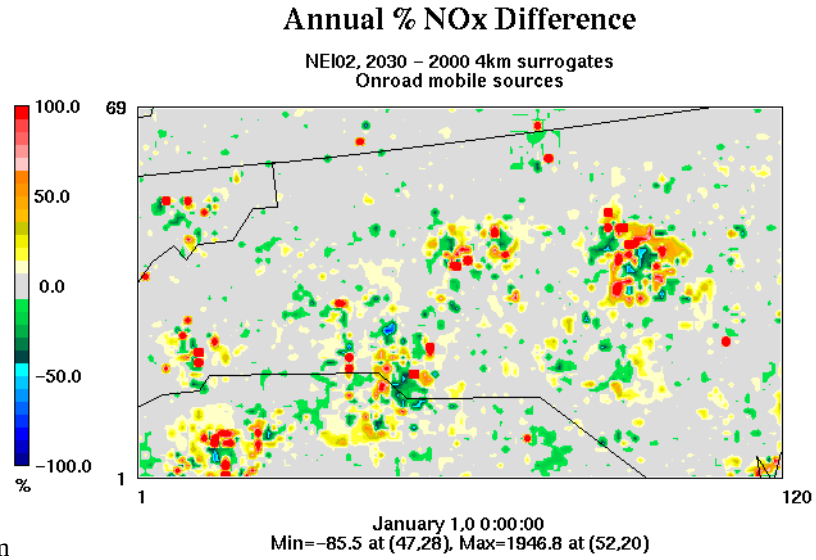
Figure 18. On-road mobile NOx emissions: annual differences due to surrogate changes (2030 – 2000)



36km

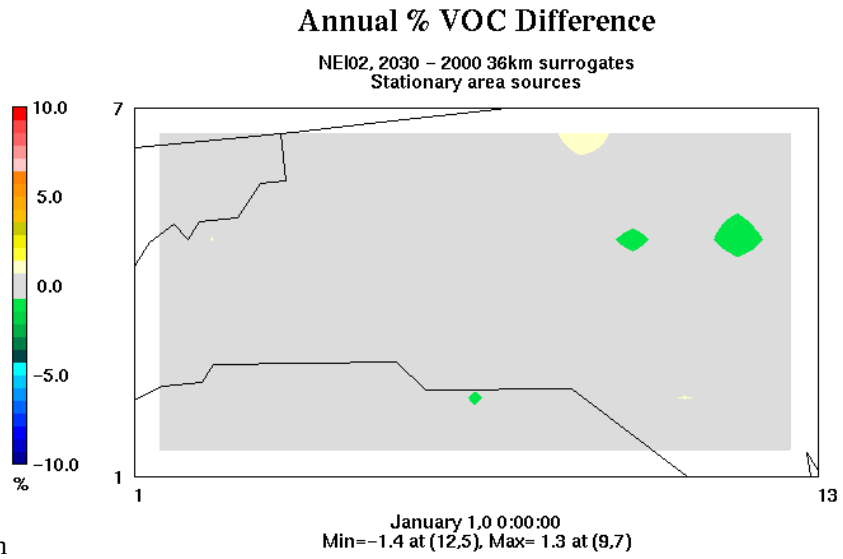


12km

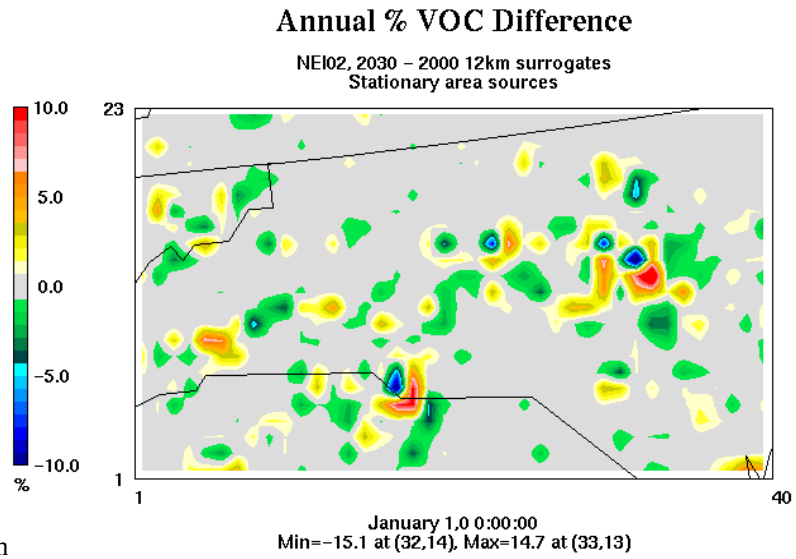


4km

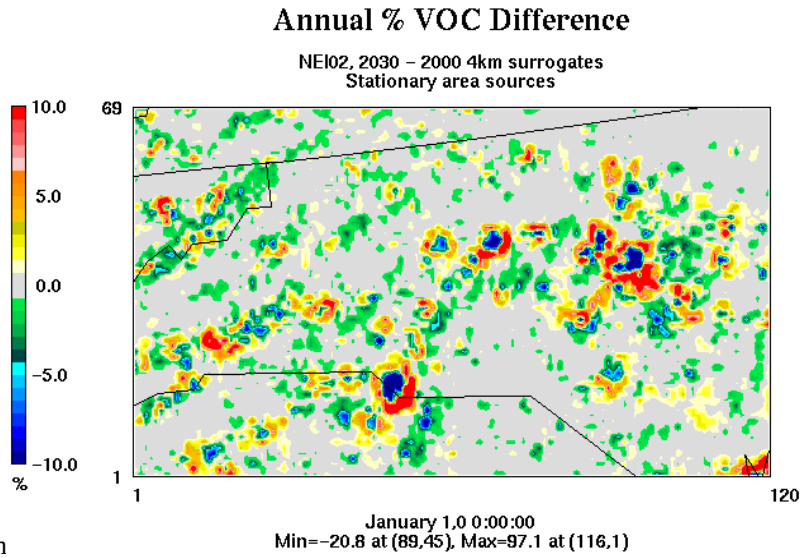
Figure 19. Stationary area VOC emissions: annual differences due to surrogate changes (2030 – 2000)



36km

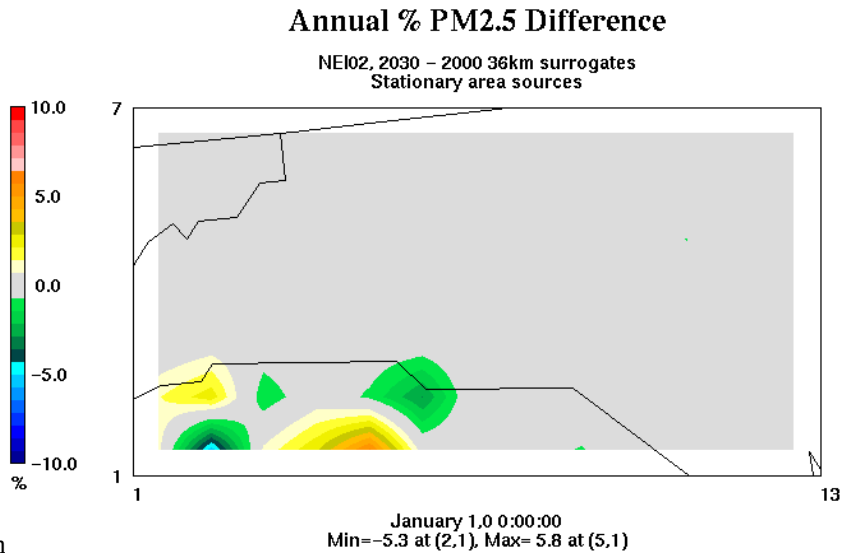


12km

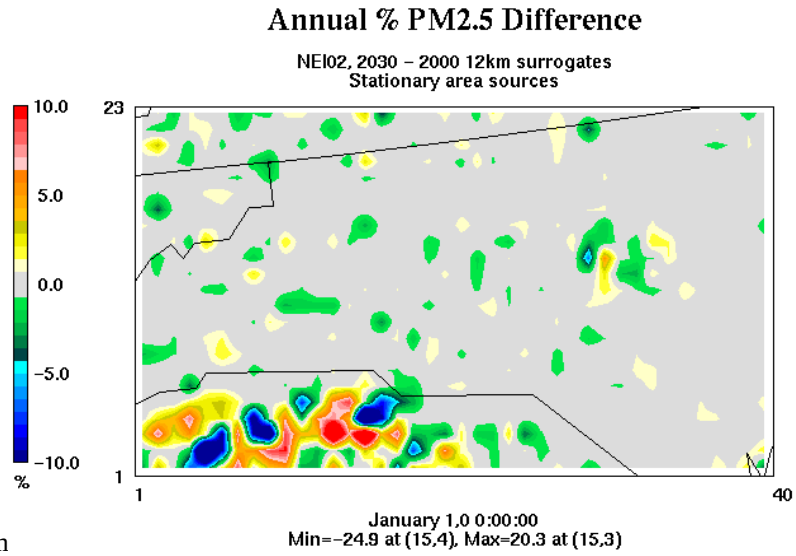


4km

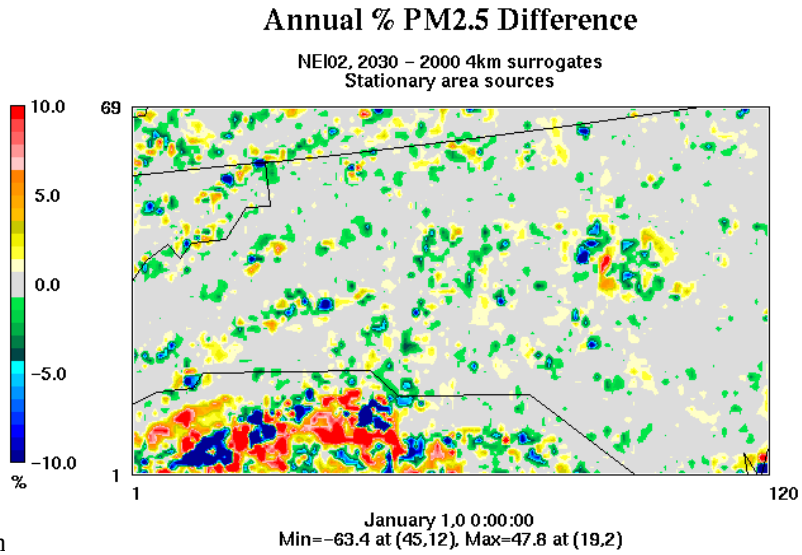
Figure 20. Stationary area VOC emissions: annual differences due to surrogate changes (2030 – 2000)



36km



12km



4km

CONCLUSIONS

From this study we have found that population changes housing density over a 30-year period are projected to be sufficiently large that they will result in a redistribution of pollutant emissions for fine- and medium-scale emissions modeling applications (e.g., 4 km and 12 km grids, respectively). Changes were less apparent when modeling at a 36 km grid, however the changes may be greater when population migration is considered.

The approach applied both population growth and housing density projections. In addition, housing density projections were analyzed to characterize future urban, suburban, and rural land use boundaries. For the case study application, urban and suburban boundaries were projected to grow in size, while rural areas diminished.

These changes in housing and population were reflected in the air quality modeling process by updating the related spatial surrogates used for emissions modeling. Use of the updated surrogates was shown to redistribute pollutant emissions, particularly from current urban areas to projected new urban and suburban areas. Allocation of emissions from on-road mobile sources was shown to change substantially since these are allocated using surrogates that are a function of urban and rural boundaries. The geographic allocation of other pollutants, such as PM2.5 and VOCs from stationary area sources and non-road mobile sources also showed some changes. These changes are expected to show some response in the air quality model outputs, but that has not been tested yet. It should be noted that currently, the population-related spatial surrogates are used more heavily in modeling than the housing related surrogates. Therefore, the fine-scale estimation of population in future years is an important component of this process.

There are a number of possibilities for future work in this area. First, it may be desirable to adjust additional surrogates using the SERGoM and projected population data. For example, comparing housing density and population projections from one decade to the next may inform the housing change and population surrogate. This surrogate is used for many stationary area sources of PM2.5. Similarly, SERGoM-projected rural boundaries may be used to update rural land area and total agriculture surrogates.

It also may be advantageous to explore updating to the spatial cross references between emissions and surrogates. For example, the cross-reference could be modified to make greater use of housing and population projections. For example, adding suburban area as a surrogate for many pollutants may provide better allocation than current surrogates that typically differentiate land as being either urban or rural. Similarly, when projecting many years in the future, housing may be a better surrogate for stationary area VOC and PM2.5 emissions than “residential heating – wood”.

Other potential changes that could be considered include updating road networks for mobile sources and land use data for biogenic sources. We might also consider updating the actual emissions in areas of high growth, and evaluating the impact of using the adjusted emissions on the air quality model outputs.

REFERENCES

Baek, B.H., A. Eyth, A. Holland, "Recent Updates to the SMOKE Modeling System", 15th Annual Emission Inventory Conference, New Orleans, Louisiana, May 16-18, 2006.

Theobold, D. M., "Mapping historical, current, and forecasted housing density in the US using the Spatially Explicit Regional Growth Model (SERGoM v2)", 2006.

Theobald, D. 2005. Landscape patterns of exurban growth in the USA from 1980 to 2020. *Ecology and Society* 10(1): 32. URL: http://www.ecologyandsociety.org/vol10/iss1/art32/

Eyth, A., L. Ran, R. Partheepan, G. Yarwood, M. Jiminez, and S. Rao. "New Tools to Generate Spatial Surrogate and Speciation Profile Inputs to SMOKE". 15th Annual Emission Inventory Conference, New Orleans, Louisiana, May 16-18, 2005.

Houyoux, M.R., Strum, M., Possiel, N., Benjey, W.G., Mason, R., Pouliot, G., Loughlin, D., Eyth, A.E., Seppanen, C. "EPA's New Emissions Modeling Framework", 14th Annual Emission Inventory Conference, Las Vegas, Nevada, April 11-14, 2005.

ACKNOWLEDGMENTS

The work described herein was funded under work assignment 2-06 of US EPA contract number EP-W-05-045. Dr. Dan Loughlin of EPA ORD's National Risk Management Research Laboratory (NRMRL) was the work assignment manager for this project. Ms. Alison Eyth managed the project at UNC. Ms. Limei Ran performed the spatial processing, developed the spatial surrogates, and performed related analyses. Mr. Zac Adelman performed the emissions processing and related analyses.

KEY WORDS

Projection
Spatial surrogates
Future-year
SMOKE
Surrogate Tool
Population
Housing
Urban
Rural