Quality Assurance of Emission Inventories Using Visual and Geographical Techniques

Roy Huntley
U.S. Environmental Protection Agency, OAQPS (D205-01), Durham, NC 27711
huntley.roy@epa.gov

David Misenheimer
U.S. Environmental Protection Agency, OAQPS (C339-01), Durham, NC 27711
misenheimer.david@epa.gov

ABSTRACT

Because of the large quantity of data in emission inventories, the potential exists for many types of inconsistencies, anomalies, and errors. The varied methods for calculating emissions, and the number of sources of information on which the calculations are based, make identification of errors a challenge. There are several approaches that may be used to identify the bulk of the errors in an emissions dataset, including visual techniques such as mapping of emissions data and visual comparisons to corresponding data sets for recent years.

This paper will present examples of visual and geographical techniques for identifying errors in emission inventories and will discuss methods for automating some GIS quality assurance approaches.

INTRODUCTION

One of the most important steps in preparing a complete, reliable emission inventory is quality assurance. It is also one of the steps most often overlooked. Often, resources available to the reviewer are very limited, so it is important to have techniques that can be used to quickly identify gross errors and their underlying causes. Such techniques stand the greatest chance of actually being used, especially under the pressures of tight deadlines for release of the inventory. There are several techniques that act as "reality checks" that can be employed to identify most of the large anomalies in an emissions dataset, including visual techniques such as mapping of emissions data and visual comparisons to corresponding data sets for recent years. Generally, it is best to start with summary data and when potential problems are identified, "drill down" to determine the specific reason for the problem.

This paper will present examples of visual and geographical techniques for identifying errors in emission inventories and will discuss methods for automating some GIS quality assurance approaches. The target audience for this paper is the emission inventory reviewer whose job it is to identify and resolve errors and data gaps in the inventory.

BODY

Important Points

Past experience shows that prior to a thorough quality review, emission inventories typically contain errors and data gaps. This is due to the nature of assembling large and complex data sets. You, the reviewer, should begin the emission inventory review process with the assumption that there are discrepancies in your data set, and your job is to identify and resolve as many as you can, especially the largest ones. We write this paper not to place blame for past errors but to describe techniques that the inventory reviewer can use to identify and resolve errors in the future. We are using real examples because we think they are more meaningful than using fictional ones. We make mistakes too, and for this paper, we will use examples of our own past mistakes, as well as those of other organizations, to illustrate how these techniques can be used in real situations.

In this paper, we discuss the use of existing "off-the-shelf" software that represents basic technology and has been around for many years. For mapping, we used the ArcGIS product ArcView 9 (single user license currently lists for $1,500) and for charts we used the Microsoft product Excel (currently lists for $229).

While ArcGIS requires more than a trivial skill level, we are demonstrating techniques involving the more basic functions of the software. These could be picked up with a one day introductory ArcGIS training course, and such courses are available online. A note of caution: When using the mapping functions of ArcGIS, one must realize that what one sees is often a function of how the maps are drawn and there are some important things you can miss. The inventory reviewer should not map the data one time and assume because there are no obvious anomalies that no problems exist. It is often necessary to view data different ways in order to highlight outliers.

For example, the map will look different if you divide the data into "quintiles" (i.e., dividing the data into a range of 5 sets with an equal number of data points in each set) compared to a map divided by "natural breaks". Another approach to viewing data from different perspectives is to plot emission densities or normalized emissions. Emission densities are emissions normalized by a related set of data such as population (e.g., tons of emissions per person) or land area (e.g., tons of emissions per square mile). Emission density maps can negate the effect that population or land area has on emissions and allow you view a more equal comparison of emissions across geographic areas. Emission densities are especially useful when emissions are a function of population, such as the category of consumer solvents.

Graphing emissions data (e.g., bar charts or pie charts) is another useful technique for identifying anomalies. We will offer examples of how graphs can be used in emissions data quality assurance.

Finally, we recognize that there are many different ways to look at the EI data beyond what is discussed in this paper. We hope that you find our paper useful, and we encourage you to try these techniques, and to develop others that are equally or more effective, while performing your own quality assurance of the emissions data. Certainly more detailed analysis can be done and we encourage others to identify techniques that work for their situation.

General Assessment of Starting Data Sets

Our starting data set is usually our composite national EI which is made up of data from State, Local, and Tribal agencies, other EPA offices, other federal agencies, and industry. We believe it is necessary to make a few comments on how the inventory comes together in order to understand the nature of some of the errors we find.

Where State/Local/Tribal data is missing or the pollutant coverage is incomplete, we "gap-fill" with our own EPA estimates. For the NOx and SO2 emissions from electric generating units (EGU), we use the data from the EPA's Emission Tracking System. Incorporating and integrating the data can be tricky because despite the common reporting format, there are still issues with the data. Issues include format errors, invalid codes, missing source categories and facilities, incomplete data coverage, and missing information. Double counting of emissions can happen easily if we are not vigilant, particularly when there is more than one source category code (SCC) per source category. For example, in the nonpoint sector, emissions from consumer solvents can be reported under two SCC's:

- SCC 2465100000 (Personal care products from consumers), and
- SCC 2460100000 (Personal care products from consumer and commercial sources)

Beyond these issues, there are cases where the emission estimates are calculated or entered into a database in error, such as the real case in which a facility reported 135,358 tons of VOC instead of the correct value of 0.135358 tons. This was found by comparing county level emissions with a previous year. The large increase of VOC was noted, the reporting agency was contacted, and corrections were made.

Another issue to include in the general assessment of input data for an inventory is consistency of pollutants reported. One technique for assessing this issue is to plot the various forms of the pollutants reported across the geographic area of the inventory. See **Figure 1**, Reporting of PM from Residential Wood Combustion. This technique allows you to visually assess the various forms of the pollutant reported by different agencies and to identify the agencies that did not report the pollutant.

Data Inconsistencies and Trends in Emission Data

Visual depictions of emissions or emission densities from nonpoint sources can be used to identify inconsistencies across geographic boundaries and over time. In general, you should not see definite patterns in emissions that highlight state or county borders. For example, emissions from consumer solvents should be present in all counties and should be proportional to population, due to the nature of the source. If you were to map emission density (emissions divided by population), you should see some consistency from state to state. For this source category , you would not expect to see large differences between states, or between urban areas and rural areas. Not all source categories are like this. For instance, agricultural tilling is done in agricultural rural areas and would not be proportional to population. So emissions density maps would not be recommended and furthermore, emissions in urban counties would be suspect. Another example is emissions from “offshore” oil and gas production equipment. For this category, emissions should only be in counties that have a shore. See **Figure 2**, Offshore Emissions. In this case, we see emissions from offshore oil and gas production equipment in landlocked counties. It was identified that the incorrect SCC was used and this error was corrected by the reporting agency.

Figure 3, PM2.5 from Wildfires, shows the county level PM2.5 primary emissions from wildfires. (Note that for the 2002 final EPA National Emission Inventory, or NEI, we will track wildfires as point sources; however, we included this slide as an example of inconsistencies across geographic boundaries). From this figure, the wildfire emissions appear to stop at the California-Arizona border. Other borders are highlighted as well. One interesting cross check is to compare the emissions from wildfires to the forested areas of the U.S. See **Figure 4**, Fraction Forested for Each County. You could also use this map to compare with emissions from Residential Wood Combustion, assuming a relationship between available forests and the burning of wood in homes. Note in **Figure 4**, state boundaries are not highlighted by the data. This technique could be used for regional studies to compare emission levels across geographic boundaries (states, counties, Indian lands, etc) to ensure a reasonable level of consistency.

Another valuable tool to quality assure emissions data is to compare current year estimates to data for previous years, and to compare data for one area to data for surrounding areas. This technique can be used to identify gross errors in emissions. See **Figure 5**,VOC Emissions Differences between 2002 & 1999 NEI. In this figure, we mapped the differences between the county VOC emissions in the nonpoint sector between 1999 and 2002. We removed emissions from wildfires and prescribed burning from both years because we did not want the large VOC emissions differences from those sources to mask changes for other source categories. This map highlights counties where the 2002 VOC emissions have changed since 1999. The largest increases are in Texas, New York, Florida, New Mexico, and Maine. Large decreases occurred in some counties in Illinois, Arizona, California, Oregon, and Washington. This type of map would not identify the specific reason for the large changes, but it will identify the geographic areas for further investigation. As we mentioned earlier, mapping emissions like this doesn’t necessarily identify outliers. This figure is a good example. Note the scale of emissions for the dark blue goes from 10,001 ton to 7,000,000 tons. The map legend tells us that somewhere in the dark blue counties is a source emitting 7,000,000 more tons of VOC in 2002 than in 1999. After further investigation, we found that for Terrell County, TX, the state submitted a value of 6,759,000 tons of VOC using SCC = 2461022999 (emulsified asphalt). We identified this as an error and contacted the reporting agency who concurred and corrected the numbers.

Note that even if changes were small compared to 1999, this does not necessarily mean that the emissions are accurate and all appropriate sources are inventoried. Also note that even though we used 1999 as a benchmark, 1999 emissions are not necessarily correct.

Paved Roads

Figure 6, PM2.5 Primary Emissions from Paved Roads, presents the data summed to the state level. More than these 8 states submitted emissions for this source category, but for illustrative purposes, we are only showing these 8 states, and we are comparing the 2002 state submittals to both the 2002 EPA numbers and the 1999 NEI emissions. This figure includes a trend analysis (by comparing 2002 emissions to 1999), a comparison of state estimates to EPA estimates, and a comparison between states. Several interesting items should be noted. First, for all 8 states shown in the figure, the 2002 EPA emissions are much less than the 1999 emissions. The reason is that EPA changed its methodology for paved roads resulting in a decrease in the emissions estimates. Most of the 1999 NEI emissions for this category must have been based on the EPA methodology.

Secondly, note that for many states, their submittals were higher than the EPA estimates and quite comparable to the 1999 estimates. We called each of these states to determine if they used the older EPA methodology to estimate 2002 emissions from paved roads, having not realized the EPA methodology had changed. California and Colorado used their own methodology, although Colorado acknowledged a mistake and sent us revised estimates. Four southern states used the same contractor, and that contractor had done the estimates before EPA released the new methodology. So for those states, we used the EPA estimates in the NEI.

Thirdly, New York is an example of a state whose emissions estimates were very close to EPA's emissions. We concluded that New York used the revised EPA methodology.

Lastly, New Jersey is an example of a state whose emissions do not match either the old or the new EPA methodology. We concluded that New Jersey has their own methodology which resulted in a lower emission estimate than the EPA methodology.

Asphalt Paving

This category presents an example of a technique where once you identify erroneous values which skew the data, you eliminate them and replot. By replotting, you allow other anomalies to be revealed. For asphalt paving, as noted earlier, there was an erroneously high value in Texas. **Figure 7**, VOC Emissions from Asphalt Paving, presents VOC emission estimates for SCC = 246102XXXX (asphalt paving), summed to the state level, and compared to 1999 estimates. Note that the high VOC value in Texas scales the graph so that other states cannot be viewed. **Figure 8** presents the same graph with the high 2002 Texas VOC emissions removed. Note how the software re-scales the data so that it now becomes apparent that values for Idaho and Indiana are higher than emissions for the remaining states. **Figure 9** is the same graph with values for Idaho, Indiana, and Texas removed. Note that comparisons that can be made more easily for the remaining states. This technique of comparing emissions to surrounding areas (i.e., states) as well as comparing data to past inventories can be very useful as a "reality check".

We have already discussed the error identified in the Texas data. For Idaho, we contacted the agency personnel and they informed us that they use cutback asphalt for paving, which has higher VOC emissions than the emulsified asphalt that many states use.

Residential Wood Combustion

We mentioned previously the technique of mapping emission densities (i.e., emissions normalized by a related data set such as land area). **Figure 10** demonstrates the mapping of emissions of PM2.5 from the burning of wood in fireplaces and woodstoves, normalized by population. If we plotted these emissions without normalizing by population (tons PM2.5/year), the map would highlight areas of high population (i.e., major metropolitan areas) because of the corresponding high numbers of woodstove and fireplace use. To look at the geographic variation of residential wood combustion and how it relates to the "per person" burning of wood, we normalize by population. Rather than highlighting major metropolitan areas, the population normalized map should highlight the areas of the country with colder climates, increased access to wood, and a culture that supports the burning of wood in residences.

Figure 10, which presents state reported estimates for residential wood burning, demonstrates a general trend of higher emissions in the northern climates, which is expected. It also presents some possible anomalies, such as higher emissions in Colorado and Alabama than in surrounding states. **Figure 11** presents the same information, except that it is averaged across each state. This type of map can be used to highlight differences from one state to another. For example, the differences between Colorado and Alabama compared to their surrounding states are more apparent in this map.

Locational Errors

Mapping is especially useful to quickly identify gross locational errors. At the national level, this is of limited use. **Figure 12** present locations of Electric Generating Units (EGU's). Some are noticeably wrong because they plot in the ocean. **Figure 13** shows the map with the locations corrected. This would be a more useful technique at the state or local level. **Figure 14** shows a plot of NOx point sources in North Carolina. Personnel familiar with the point sources at the state, local or tribal level could use this to quickly determine if sources are located correctly. Points of reference, such as rivers, highways and lakes, can be helpful in determining if facilities are located properly. You can plot the sources such that the more emissions, the larger the circle, so that the large point source emitters can be easily identified.

CONCLUSION

Visual techniques can be very useful in quality assurance of emissions data. Time should be set aside, and resources (tools, personnel, etc) should be acquired to support an adequate review of the inventory. You should assume that a significant number of errors are contained in the data and that reviewing the data from a number of perspectives is necessary to highlight the errors and allow for their correction. As part of planning for the review, you should document the techniques that you plan to use. After completion of the review, you should document the techniques that were actually used - this will be beneficial for the next inventory review.

Figure 1. Reporting of PM from Residential Wood Combustion

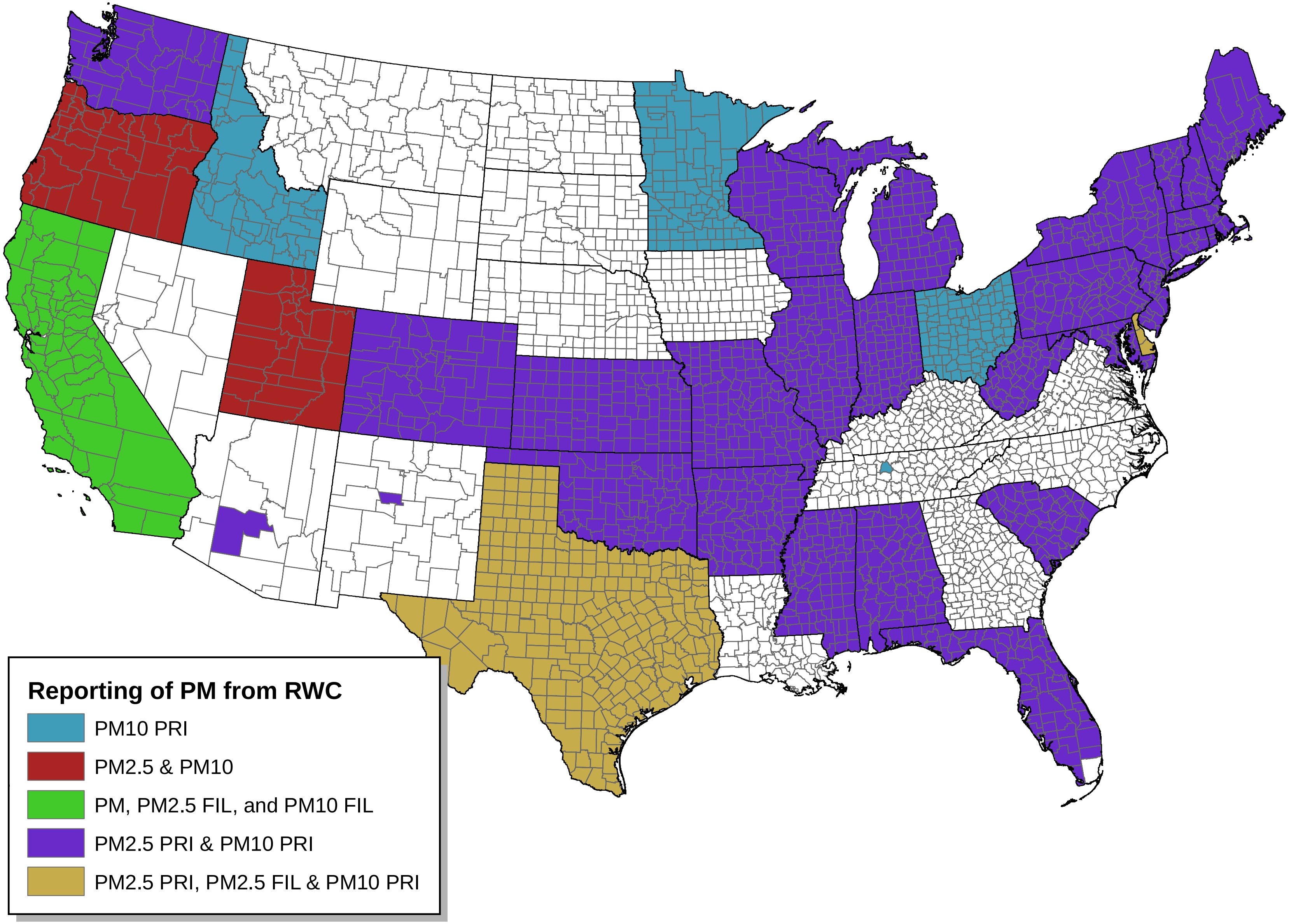


Figure 2. Offshore Emissions

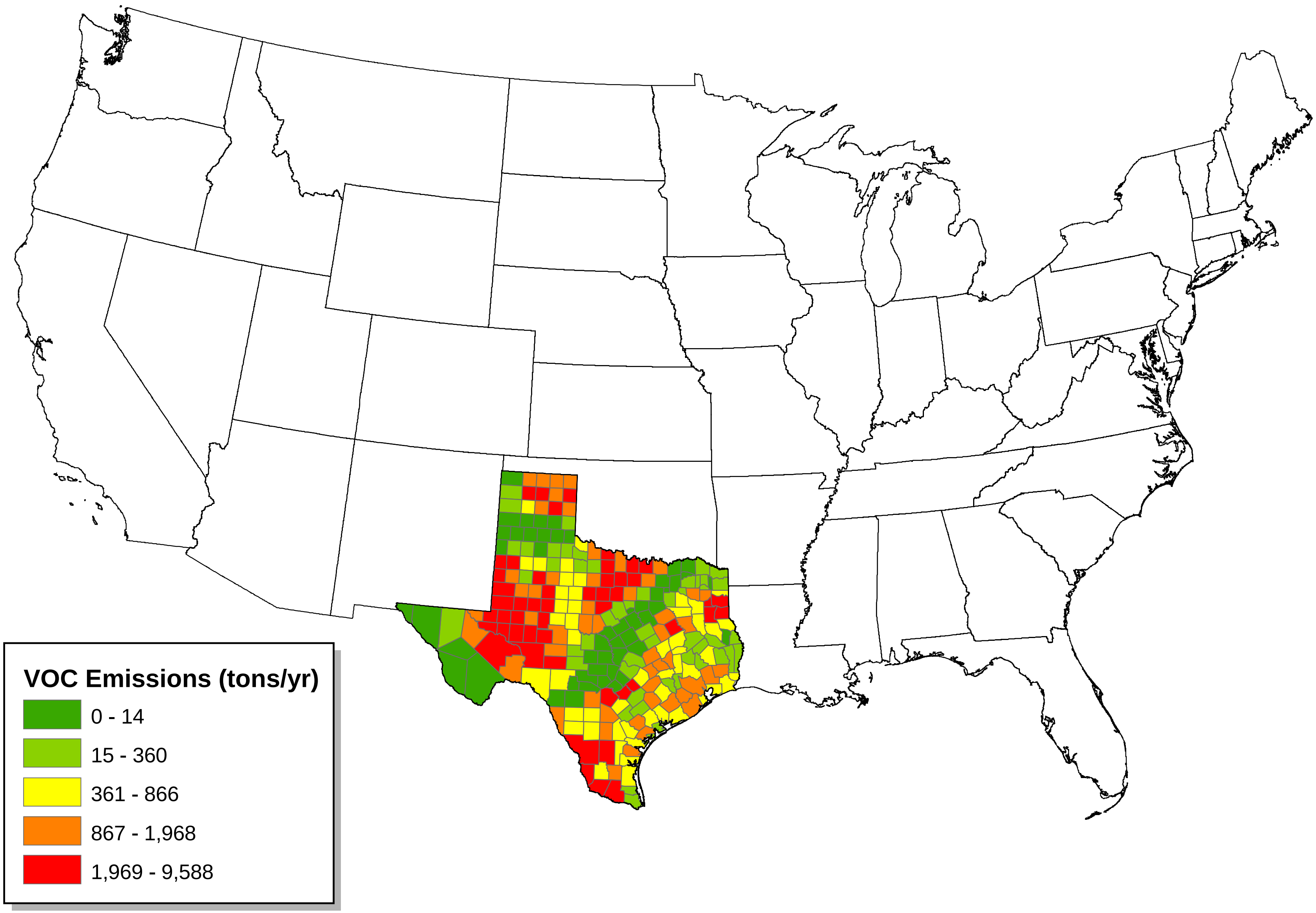


Figure 3. PM2.5 from Wildfires

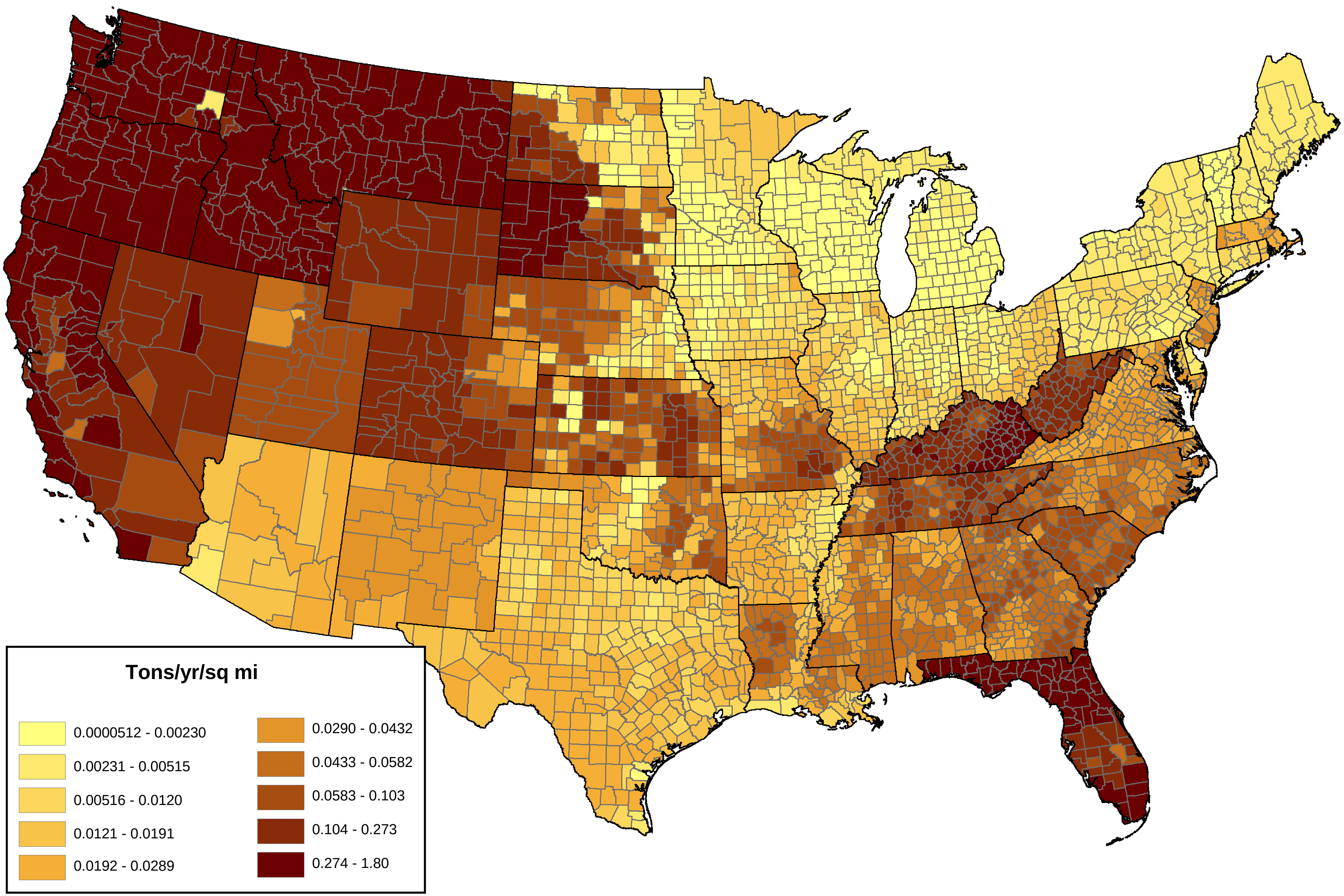


Figure 4. Fraction Forested for Each County

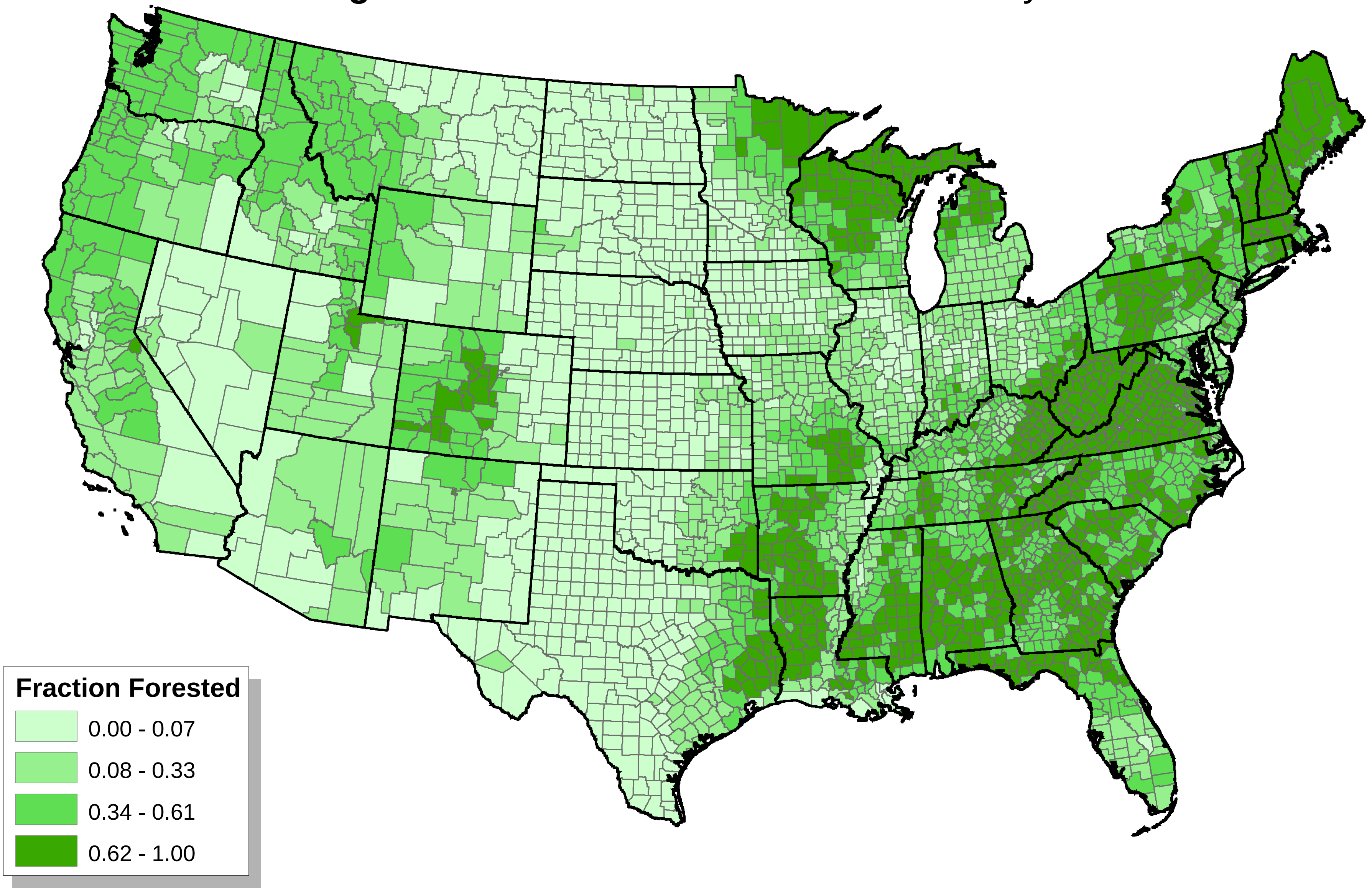


Figure 5. VOC Emissions Differences between 2002 & 1999 NEI

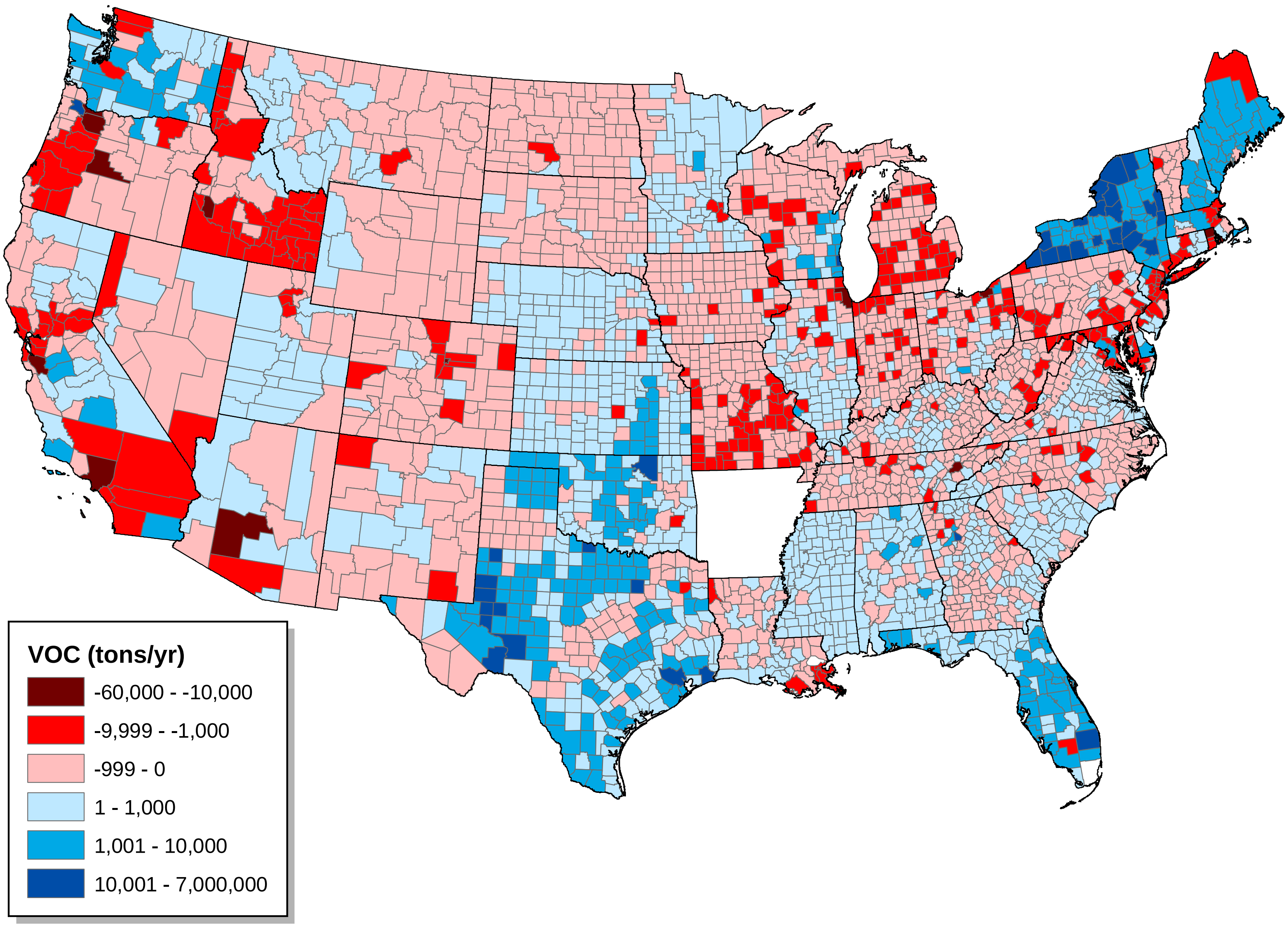


Figure 6. PM2.5 Primary Emissions from Paved Roads

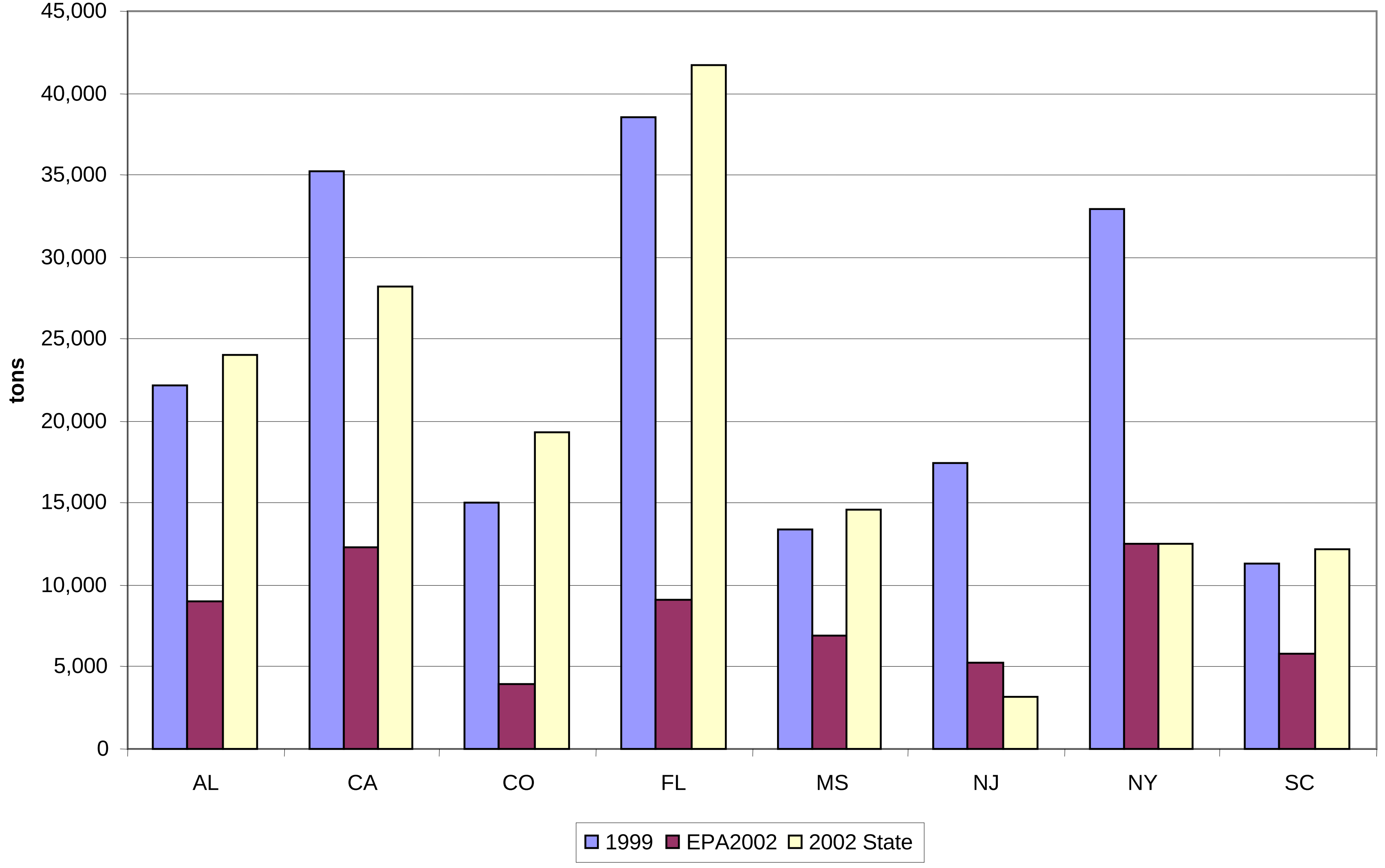
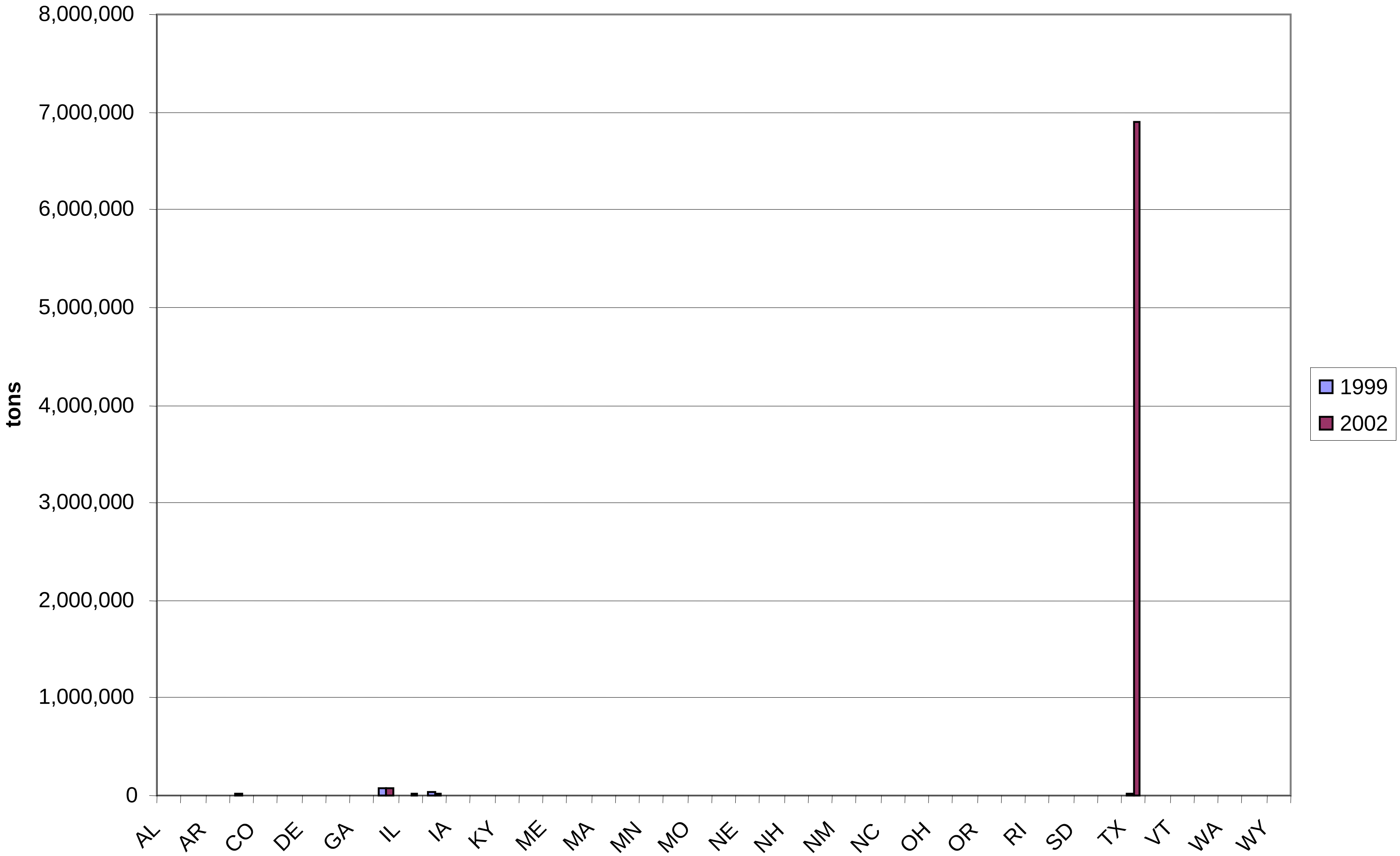


Figure 7. VOC Emissions from Asphalt Paving



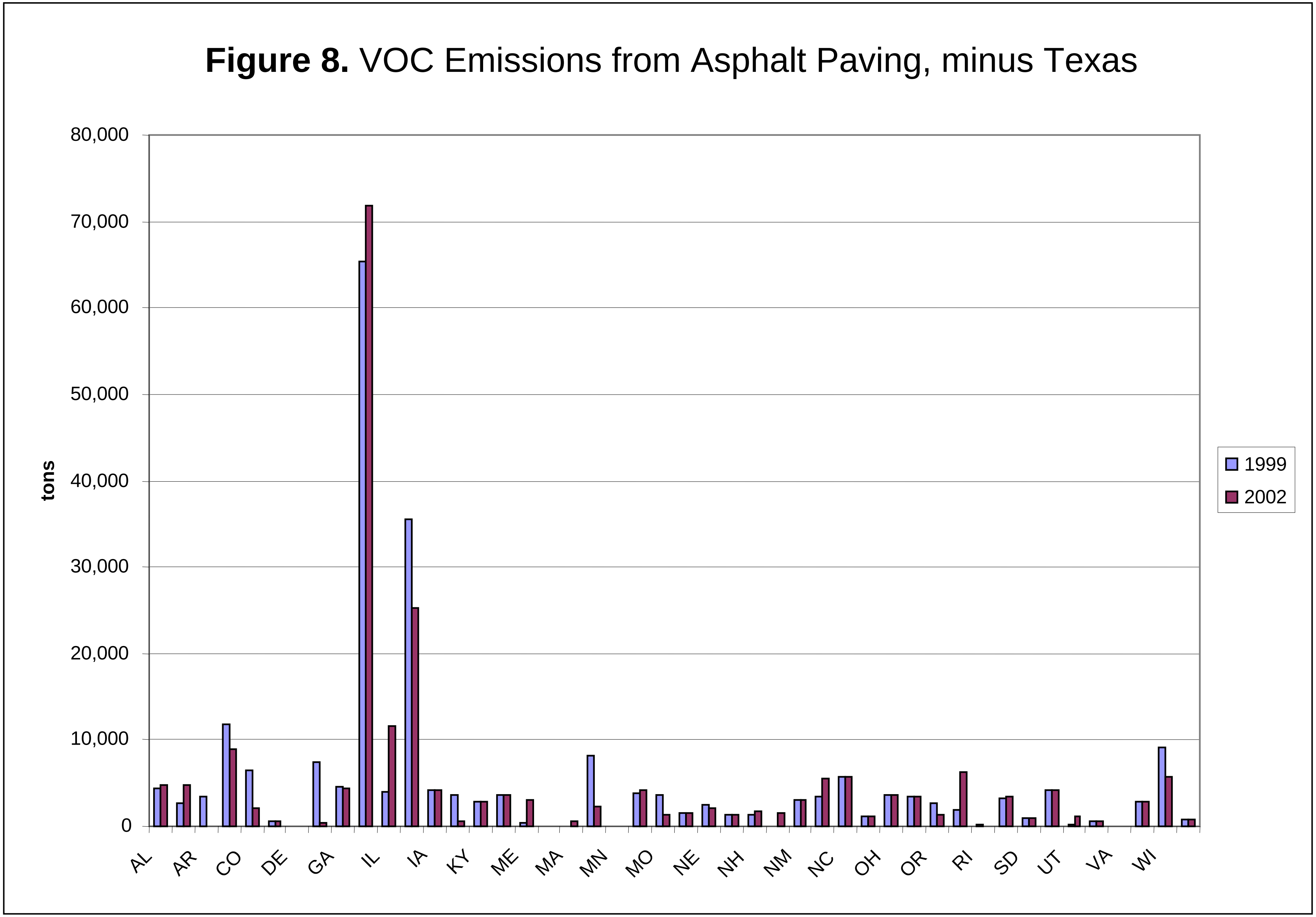


Figure 8. VOC Emissions from Asphalt Paving, minus Texas
tons
80,000
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
AL
AR
CO
DE
GA
IL
IA
KY
ME
MA
MN
MO
NE
NH
NM
NC
OH
OR
RI
SD
UT
VA
WI
1999
2002

Figure 9. VOC Emissions from Asphalt Paving, minus Texas, Idaho & Indiana

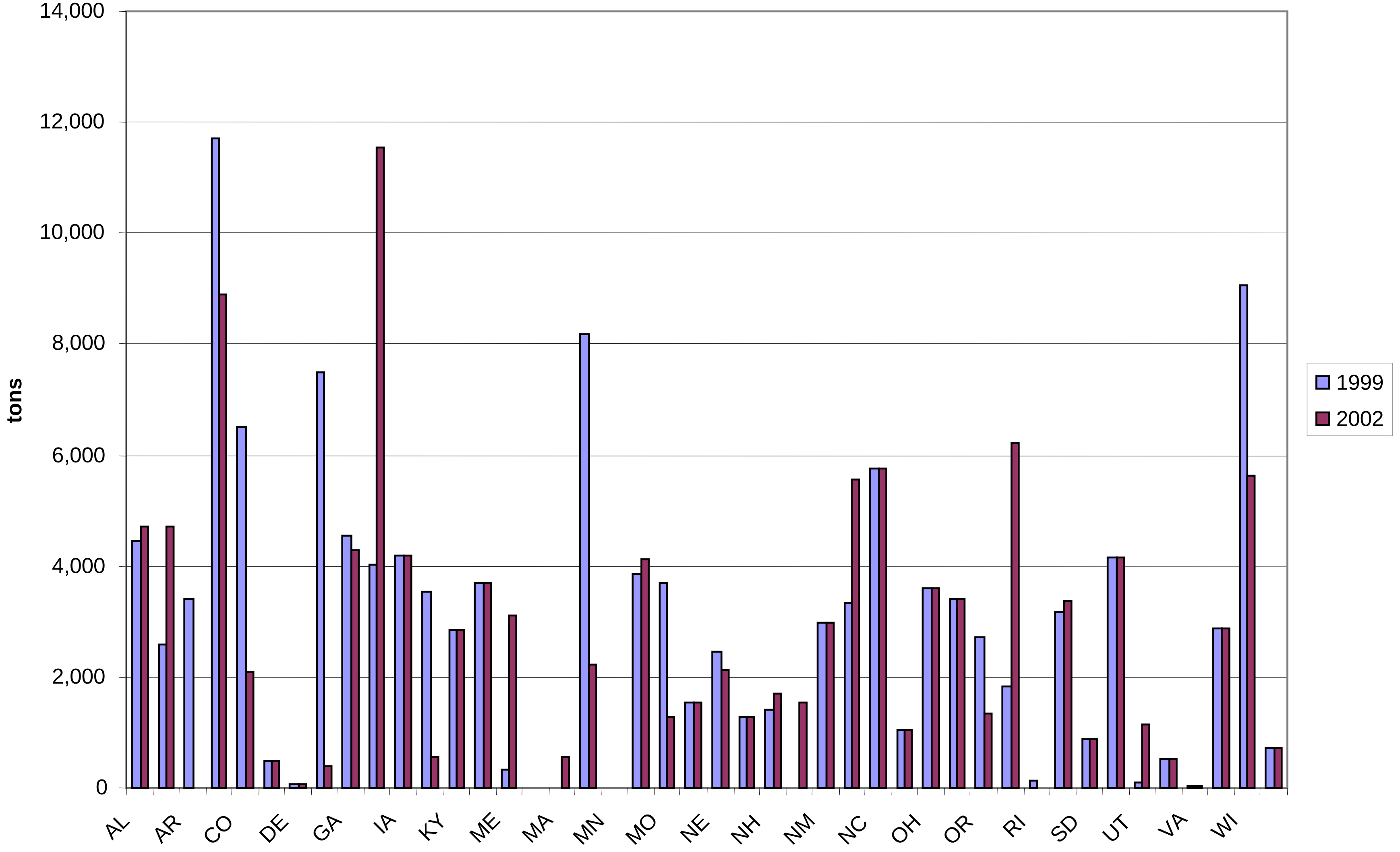


Figure 10. PM2.5 Emissions from Residential Wood Combustion

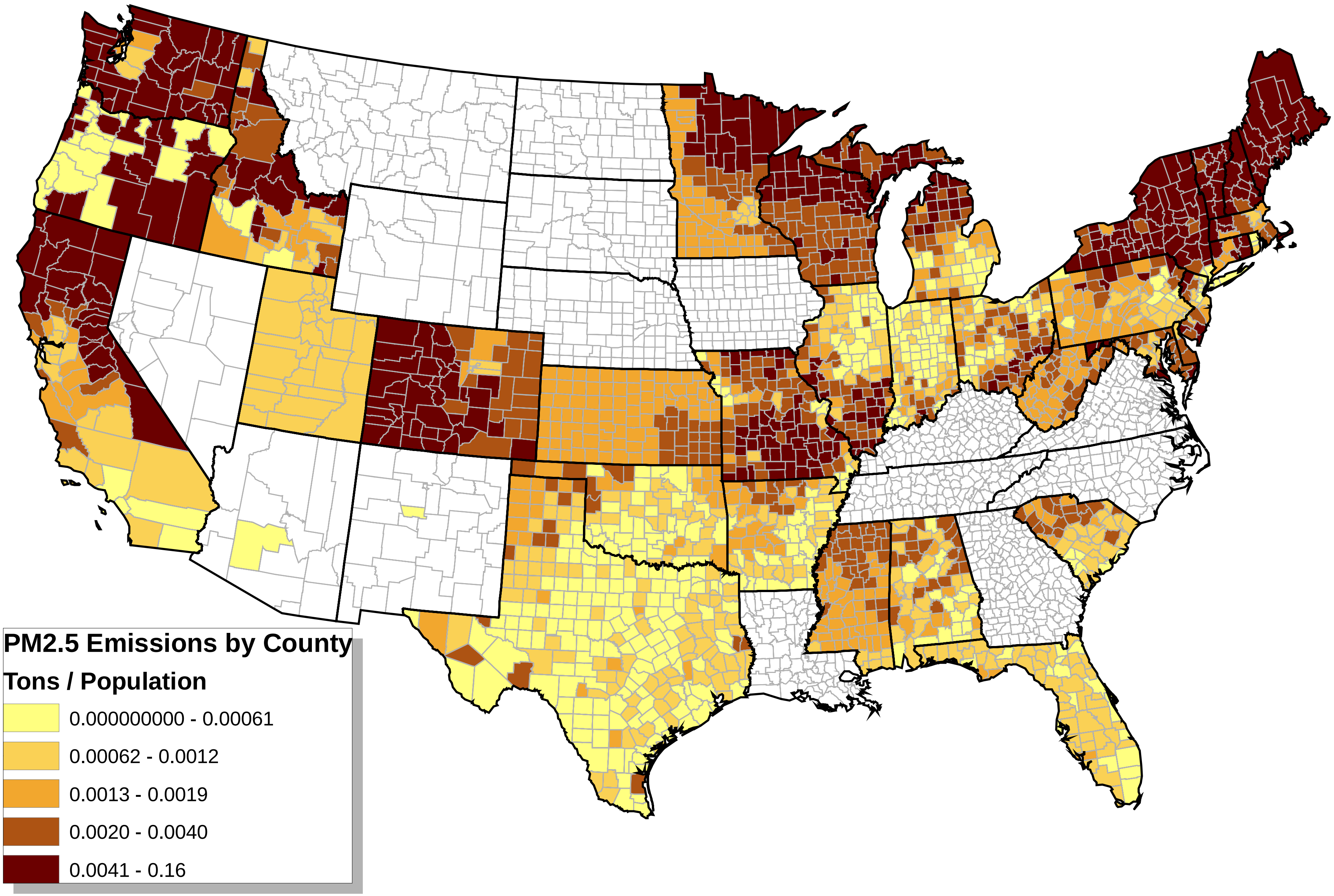


Figure 11. PM2.5 Emissions from Residential Wood Combustion

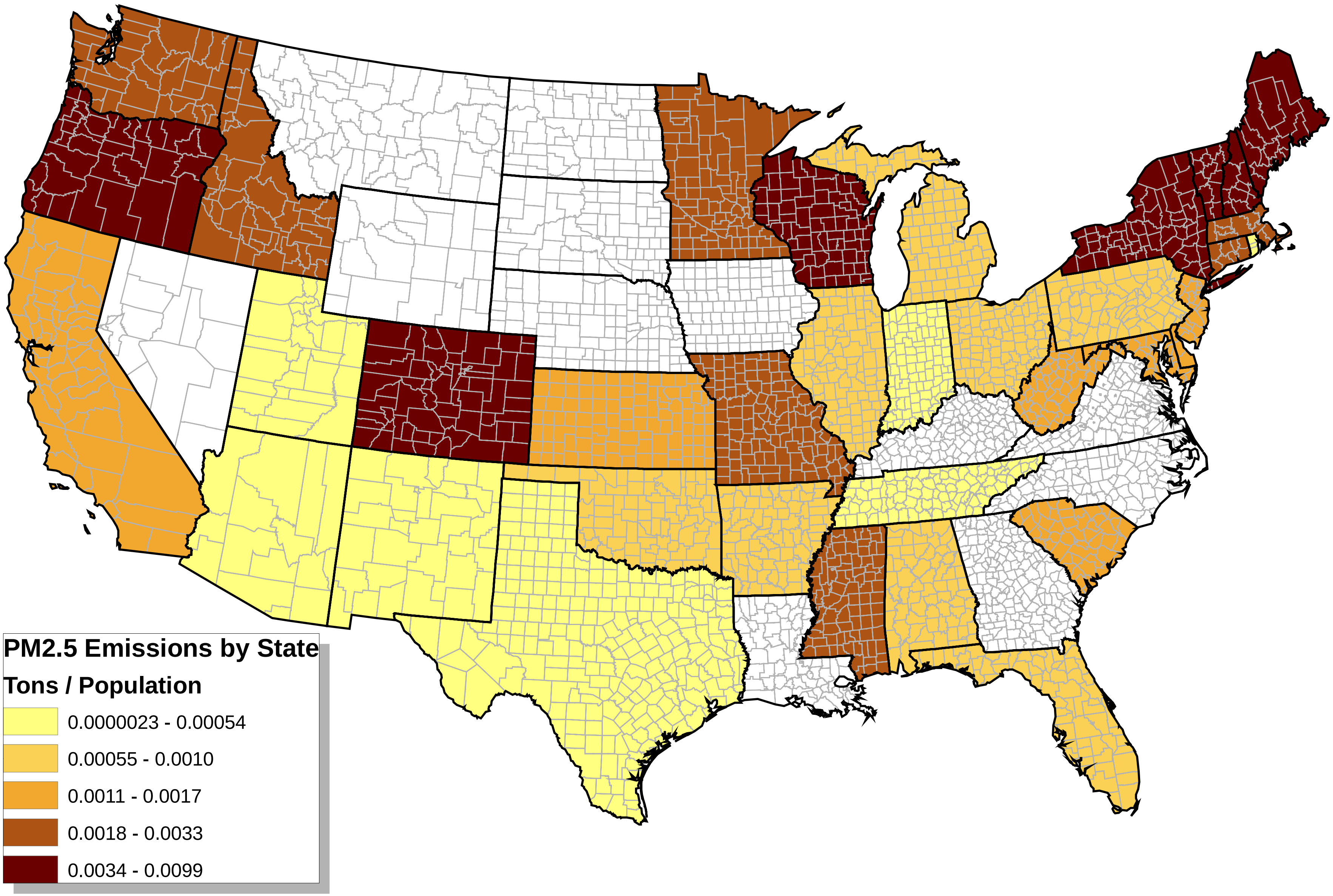


Figure 12. 2002 NEI Draft – EGU Locations

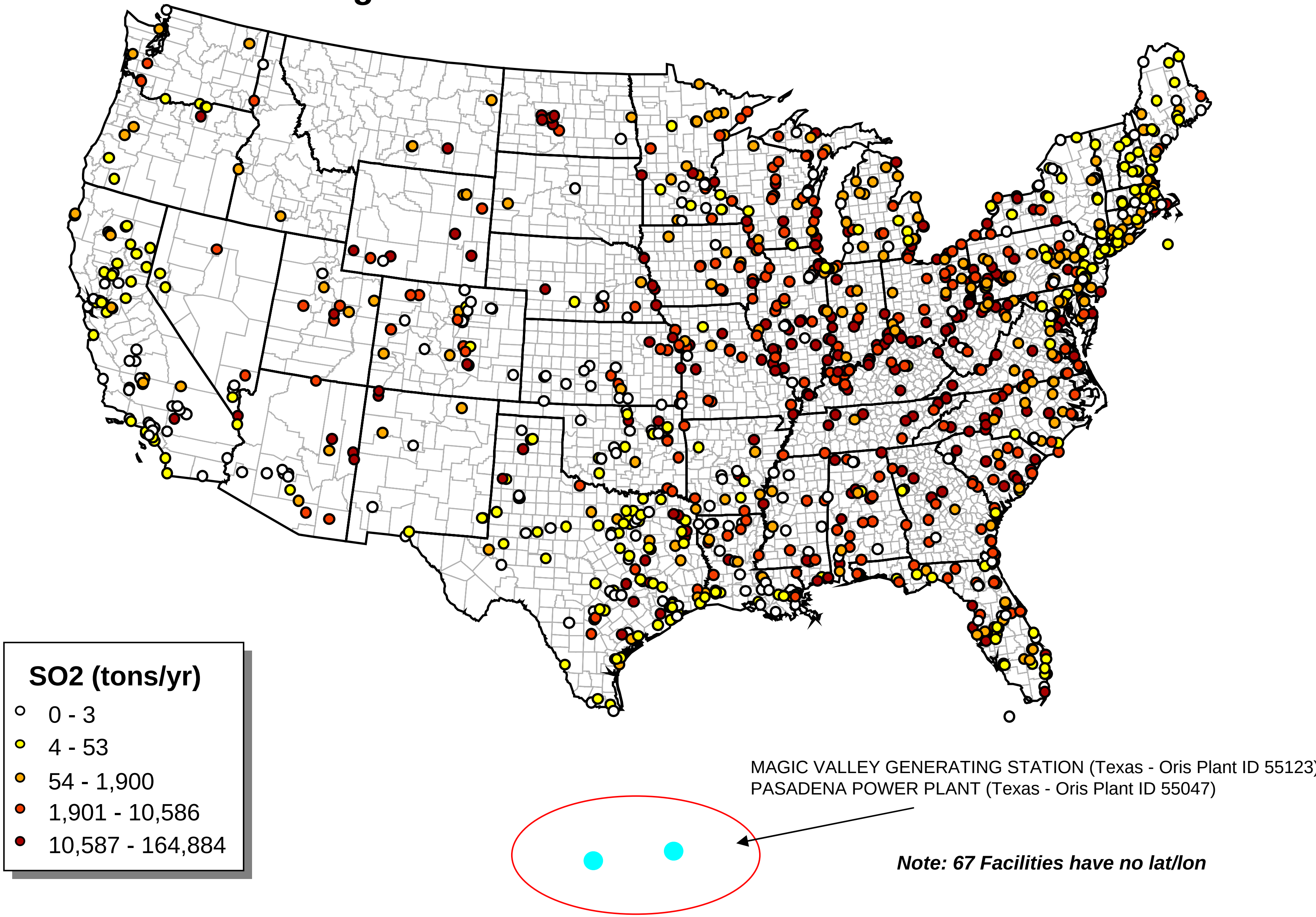


Figure 13. 2002 NEI Draft – EGU Locations, corrected

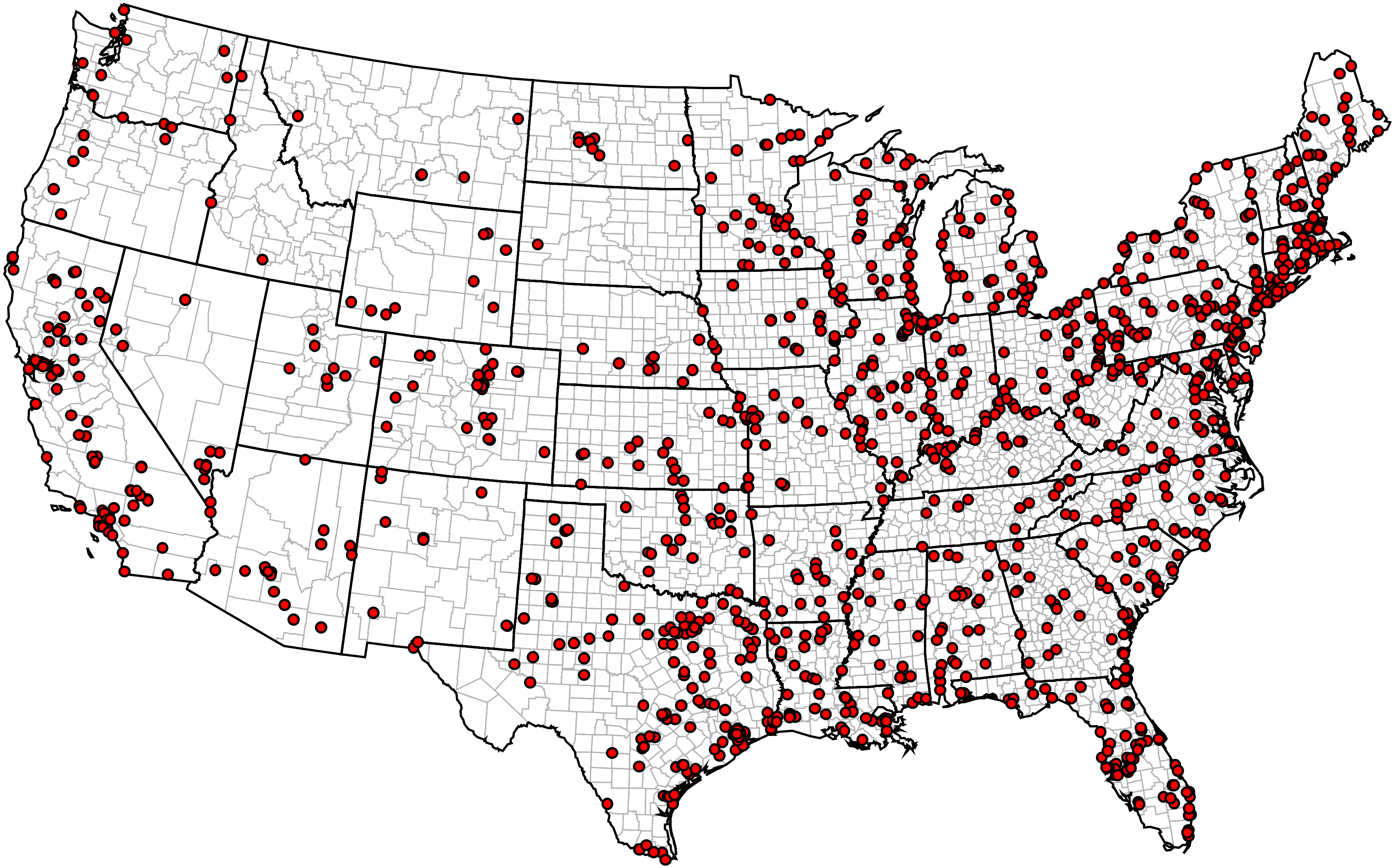
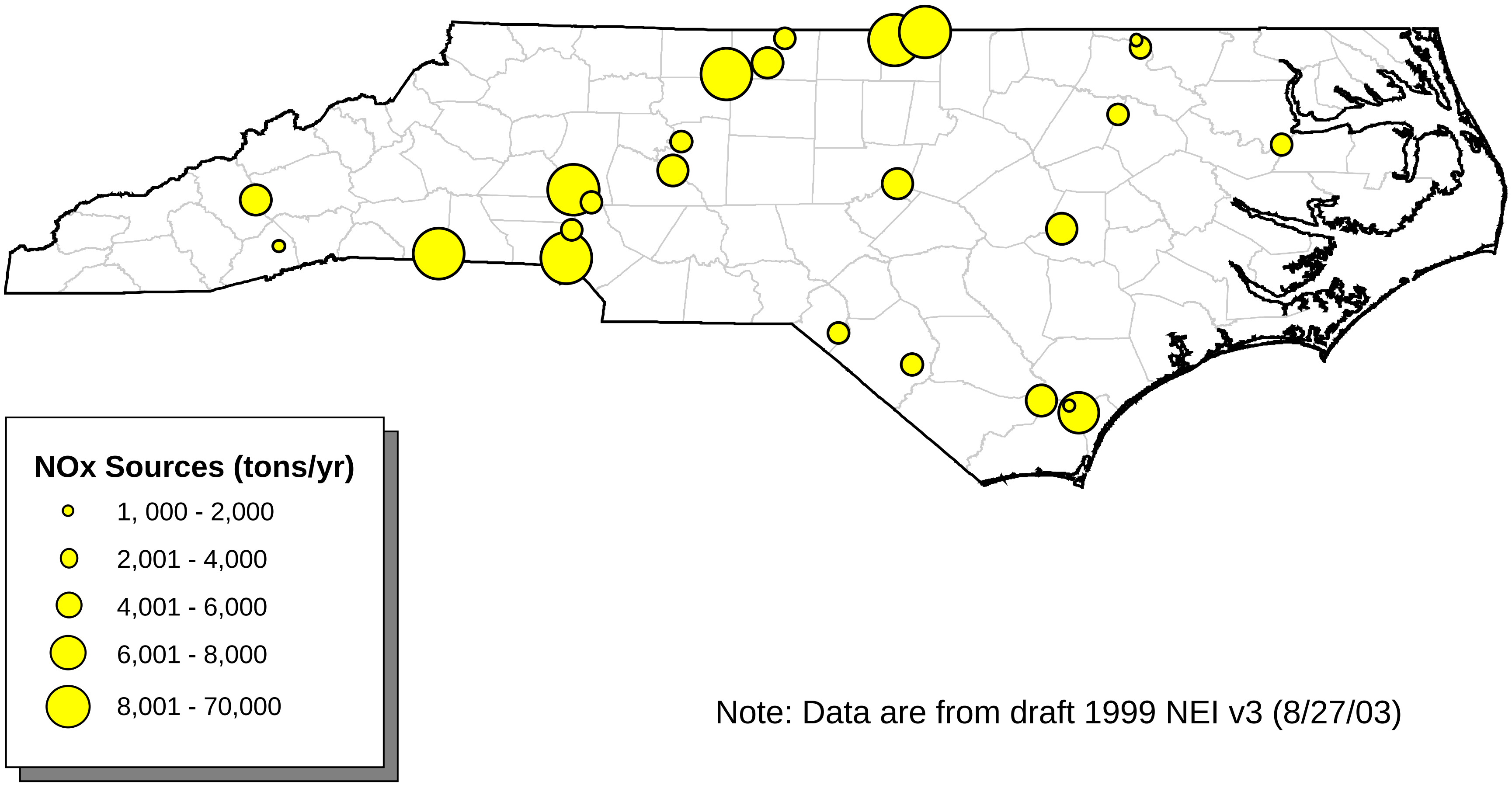


Figure 14. NOx Point Sources in North Carolina



Note: Data are from draft 1999 NEI v3 (8/27/03)