Use of GIS and GPS as a QA Tool in Emission Inventory

Andrew Alexis
California Air Resources Board
1001 I Street, Sacramento, CA 95814
aalexis@arb.ca.gov
International Emission Inventory Conference, Atlanta, GA.
April 2002

ABSTRACT

Spatial data is increasingly important in emission inventories. Over the last 5 years, the Internet, public “right to know”, and environmental justice programs have greatly increased the visibility and use of emission inventory data. The increased visibility of inventory data in turn increases the need for accurate spatial inventory data. Geographic information systems (GIS) and global positioning system (GPS) technologies are invaluable tools that can be used by state and local government agencies to address spatial deficiencies of inventory data. This study addresses the spatial quality of large point source data. For the 1999 inventory year, large point sources accounted for 8.5% of NO_x emissions and 9.5% of Reactive Organic Gas (ROG) emissions in California. One GIS technology, address matching, can be used to assess the accuracy of point source coordinate data. The spatial data quality of a subset of point sources in the California Air Resources Board’s tabular California Emission Inventory Data and Reporting System (CEIDARS) was analyzed using geographic information systems. The reported coordinates of point sources were compared with coordinates determined by address matching the street address of the point source. The accuracy of reported coordinates varied widely. The net emission impact of these point sources on neighborhoods can be significant.

INTRODUCTION

In the recent past, inventory data was used primarily for regional State Implementation Plan (SIP) planning, and non-point source inventory data only had to be resolved spatially to a regional level. Grids used for modeling purposes were 5 kilometers on a side, so point sources only had to be accurate within a 5 kilometer grid cell for modeling to be reasonably accurate. The Global Positioning System (GPS) and Geographic Information Systems (GIS) were not widely available, and spatial errors as great as 2-3 kilometers did not significantly skew results. Greater spatial resolution was neither technically feasible nor important for most regulators. Inventory development focused on greater accuracy of estimates and greater emission category resolution.

The Air Resources Board’s (ARB) California Emission Inventory Data and Reporting System (CEIDARS) was designed in the early 1990s with SIP planning in mind. It is a relational database; it currently contains toxic and criteria data for over 17,000 large point sources; 460 area-wide source categories, and over 100 mobile source categories. Emissions data from sources other than large point sources are resolved spatially to the county-air basin-district level. Point source data from large sources are reported to local air pollution control districts (districts) and relayed to the ARB. The ARB and districts estimate emissions from small point sources at a regional level. ARB and districts also collaborate on area-wide sources emissions estimates, and ARB alone has responsibility for all other sources.

For the 2001 inventory, approximately 17% of NO_x emissions and 19% of reactive organic gas (ROG) emissions originate from stationary sources[1]. Of these emissions, approximately half, or 8.5% of NO_x emissions and 9.5% of ROG emissions, originate from large point sources[2] that are reported individually to districts. The remainder of stationary source emissions originates from small point sources.

This paper is primarily concerned with large point source data. All other sources, including small point sources, area-wide sources, and mobile sources, are estimated either manually with complex methodologies (such as AP-42 methodologies[3], or ARB's area-wide source methodologies[4]) or with equally complex estimation models (such are ARB's EMFAC mobile source model[5]). To achieve a greater spatial resolution from these sources either the methodologies or models must be redesigned, or the regional emissions they provide must be spatially disaggregated using spatial surrogates.

CURRENT STATUS OF SPATIAL DATA IN CALIFORNIA

Large point source inventory collection and reporting practices vary widely across the state. California's current emission inventory database system, the California Emission Inventory Data and Reporting System (CEIDARS), was designed approximately 12 years ago, before GIS technologies were prevalent. Universal Transverse Mercator (UTM) coordinates for point sources are stored both at the facility level and at the stack level. For the first 3 years of its existence, the CEIDARS database was only able to store UTM coordinates in kilometers and only at a maximum resolution of 100 meters. Coordinate data are not mandatory, and approximately 20 percent of facility level coordinates are missing. In addition, 36 percent of these coordinates are reported only to a 1,000 meter or greater resolution. Address data are available for point sources, but the rural nature of some point sources can make addresses not very useful. Stack data are not mandatory, and districts vary widely in reporting this information. Emissions without stack data are assumed to be fugitive emissions.

NEED FOR BETTER SPATIAL DATA

According to recent data, almost half of US households have access to the Internet[6]. The Internet has an expectation of immediacy and currency, and the public expects access to all governmental data and expects it to be correct, complete, and current. In particular, business and governmental agencies alike are creating web-enabled mapping applications that allows users to view maps of areas of their choice and overlay inventory and other environmental data. These applications highlight the spatial errors in the data. To the casual user, spatial errors can also suggest that other, non-spatial data may also be in error.

In November 1986, California voters approved Proposition 65, the Safe Drinking Water and Toxic Enforcement Act[7]. Proposition 65 requires the Governor to publish a list of chemicals that are known to the State of California to cause cancer, birth defects or other reproductive harm. More importantly, under this law businesses are required to provide a "clear and reasonable" warning before knowingly and intentionally exposing anyone to a listed chemical. Businesses are legally liable for their toxic emissions, and can be sued. Thus, under Proposition 65, the proximity of a business emitting toxic substances to potential receptors takes on great importance. Incorrect spatial data can erroneously place these businesses near “receptors”, when in reality there may be no receptors nearby.

In addition, ARB has developed and implemented airborne toxic control measures (ATCMs)[8] for control of emissions of toxic pollutants. Risk assessments are used in the development of ATCMs. These risk assessments require that spatial data be resolved to the neighborhood level to determine who is being affected by toxic emissions and where these emissions are occurring.

Finally, the ARB recently adopted policies for its Environmental Justice program[9] and is emphasizing a neighborhood-level approach to reducing air pollution. In order to accurately assess the neighborhood impacts of emissions, spatial data of greater accuracy are necessary. Also, one of the major policies of ARB’s Environmental Justice program is to strengthen outreach and education efforts in these communities. In order to implement this policy, ARB needs to make inventory data more accessible and understandable, which also necessitates obtaining greater spatial accuracy for inventory data.

TOOLS TO IMPROVE SPATIAL DATA

Inventory data from large point sources are usually reported individually; spatial data from these sources can be verified and supplemented where necessary using GIS and GPS technologies.

GIS - Address Matching
Geographic Information Systems (GIS), such as Environmental Systems Research Institute’s (ESRI) Arc-GIS series of products, provide many useful computer desktop tools that can be used to improve spatial data. Address matching is a GIS tool that can determine the exact location of an address from a master database of streets. (Address matching is also known as geocoding, which can also refer to the plotting of XY coordinates). Address matching is most useful in established urban areas that are mapped in the electronic street maps used by GIS to determine locations. New built-up urban areas and rural areas often are difficult to address-match. Also, address matching is useful for small point sources or where emission release point location data are not critical. Care must be taken with certain types of point sources, such as oil production fields and gravel pits, that often report a central office or a mailing address rather than the actual facility location. Address matching can be used to determine the “front door” location of point sources, but it cannot be used to determine emission release point locations.

Address matching can range in cost and quality. Standalone GIS software, such as the ESRI ArcGIS and Streetmap USA products, can cost upwards of $2,500, not including the Windows

2000 computer to run it on. The quality of the software and the coverage of the street data is excellent, although training costs are significant for this software. Air Resources Board staff have even used low-end consumer products such as Microsoft's Streets USA (approximately $30) to obtain coordinates of addresses for special projects where great precision is not necessary. At least one web site, www.mapsonus.com, allows users to perform address matching on-line. Although convenient for single address matching, low-end consumer products and web-based services do not have the batch processing capabilities of desktop GIS systems.

GIS – Other Data Sources

GIS can be also be used to link point sources to electronic GIS maps provided by other sources. One such example is the use by the Ventura County Air Pollution Control District of county assessor parcel maps. Many county tax assessor offices have created detailed county GIS parcel maps. These maps are potentially very useful, mostly because as a general rule point source locations are coded as polygons, rather than as single points. The point source polygons can later be supplemented with emission release point locations (as points) that have been determined by other methods. Ventura County requests assessors parcel numbers from the point sources it regulates, and links point source records to the Assessor maps it has been provided. GIS can also determine the centroids of assessors parcel polygons for reporting use. The one limitation of the use of county assessor maps is that many counties have invested very heavily in GIS maps of their areas and are trying to recoup this investment by reselling the map data to others, including other governmental entities. These data can be expensive.

GIS – Reported Coordinates

GIS can also plot point source locations using coordinates reported by either the point sources themselves or by local government. Self-reported coordinates are often of uncertain provenance, though, and California's experience (as will be discussed later) is that they can be wildly inaccurate. In the absence of other data, however, they can provide a fall-back location for point sources. For certain types of rural sources (such as oil production facilities) that do not have street addresses, reported coordinates may be the best available spatial data.

Global Positioning System

Global Positioning System (GPS) receivers are another useful tool in determining the spatial coordinates of point sources. The GPS was developed by the US Defense Department and consists of three elements: the satellite segment, the control segment, and the receiver segment. Receivers use triangulation of the positional data it receives from the satellite segment to determine location. To get an accurate location with a GPS receiver, the GPS receiver must be receiving signals from at least 4 satellites; 3 satellites to triangulate position and 1 to correct errors. These receivers range from consumer models costing less than $100 to surveyor-grade models with laser rangefinders costing upwards of several thousand dollars. Up until June of 2000, the Defense Department intentionally degraded the non-military GPS signal so that location was accurate only within 100 meters, a process known as selective availability. This was done for security reasons. In June, 2000, selective availability was discontinued, but current events suggest that selective availability could be restarted at any time. Location accuracy has improved somewhat, but still must be considered when using GPS receivers. Accuracy is improved proportionate to the cost of the GPS receiver.

GPS receivers are most useful in rural areas or urban areas without tall buildings. Accurate GPS locations require that the receiver have a clear view of the sky; obstructions such as trees or tall buildings can either block or reflect satellite signals, causing inaccurate readings. GPS receivers can be used to obtain emission release point locations if site access is available. If access to a point source site is not available, higher end GPS receivers can correct locations with offsets obtained with the use of laser rangefinders plugged into the receiver. For general use, consumer level GPS receivers are probably adequate for use in emission inventory where point sources accuracy of approximately 100 meters is adequate. GPS receivers can be used in conjunction with GIS systems; locations obtained with GPS receivers can be overlaid with street maps or terrain maps to crosscheck locations.

The ARB has instituted a program to loan GPS receivers to local air pollution control districts for use in obtaining point source location data. This program was aimed at small and medium sized districts that otherwise would not have the necessary resources to fund these devices. To date, 4 of California's 35 districts have borrowed GPS receivers.

WILMINGTON CASE STUDY

An example of the use of these tools in assessing the spatial data quality of California's emission inventory was an assessment of the accuracy of point source data in Wilmington, California. As part of the ARB's Community Health Program, the ARB has initiated a one-year study of air quality in Wilmington, California. The purpose of this study is to understand the cumulative impacts of air pollution on the residents of a community in order to better protect the health of all individuals. Wilmington was chosen for this study because of the close proximity of major industrial facilities, heavily traveled freeways, and the Ports of Los Angeles and Long Beach to local schools and residential neighborhoods. This study differs from regional air quality studies conducted previously because it is focusing on the neighborhood-scale impacts of air pollution within Wilmington.

Methodology

The CEIDARS database was queried for all point sources in the same county as Wilmington, Los Angeles County. These point sources were plotted using the reported coordinates (where available). In addition, point sources were address matched using ArcGIS and Streetmap USA. Reported locations and address-matched locations were then compared and mapped. Lines were drawn between the reported location and the address-matched location to estimate the locational error of the reported coordinates. The emissions were determined from CEIDARS for point sources in the study area, as determined either by the reported coordinates or by address matching. Emissions for point sources whose reported locations were within the study area were compared with emissions for point sources whose address-matched location were within the study area.

For this analysis only, the Wilmington study area was defined as a polygon containing ZIP Codes 90744, 90745, 90710, 90810, 90822, 90502, and portions of surrounding ZIP codes.

Results

The mean estimated location error of the reported coordinates was 1,759 meters (1.1 miles). Errors ranged from 18 meters to as much as 59 kilometers. Because the range of estimated error was so large, the standard deviation was 3,974 meters (2.4 miles), more than twice the mean.

When the reported coordinates were plotted, a number of "clusters" of point sources were apparent. These clusters consist of groups of from 2-15 unrelated facilities, all with the same or nearly the same set of reported coordinates. Later investigation suggested that these "clusters" might have resulted from the use of spreadsheets in the preparation of toxic emission inventories by district contractors. The inventory contractor may have inadvertently copied the same set of coordinates down the spreadsheet columns used to store coordinates.

Some point sources were reported to be within the study area, but had addresses that were not in the study area. The reverse was also true. Emissions from facilities whose address-matched locations fell inside the study area differed from emissions from facilities whose reported location fell inside the study area. From the perspective of the address matched locations, vinyl chloride emissions were over 2,000 percent higher in the study area for address-matched point sources as compared with point sources whose reported locations were within the study area. Likewise, emissions of methyl chloride were 1,296% higher. Table 1 summarizes the biggest differences.

This study shows the importance of accurate spatial data. Existing spatial data resulted in overestimates for some pollutants and underestimates for others. This could lead to erroneous conclusions and has significant implications for public policy.

The current study analyzed only differences in emissions resulting only from whether a given location was in or out of the study area. Emission totals from point sources were not verified with the point sources themselves or with the local air pollution control district. In June 2002, we hope to "ground truth" all 300 large point sources within the Wilmington study area by determining their location with GPS receivers and comparing these locations with reported and address-matched locations. Emissions estimates will be verified by contacting point sources and cross checking emissions estimates with local district files and any other emission estimates (such as US EPA Toxic Release Inventory data). This will permit further analysis of the effect of emissions on residential and school land uses within the study area. Additionally, we hope to supplement or correct point source location with spatial data from other sources, most notably the US Environmental Protection Agency's Enviromapper system.

These data will be used to crosscheck address-matched locations and reported locations. The data will also be used to ensure that the point source inventory in the Wilmington study area is complete and that the universe of point sources, beyond those reported in CEIDARS, are considered.

STATEWIDE ANALYSIS OF POINT SOURCE DATA

We have also used the same analysis developed for the Wilmington study to estimate the spatial error of all point sources in CEIDARS. For the 10,019 point sources that have reported coordinates and are address-matchable, the average distance between the address-matched coordinates and the reported coordinates is 6,243 meters (3.8 miles). This is deceiving, though, because it includes point sources with reported coordinates not within the reporting county. Also, some types of point sources, such as oil production facilities, report a mailing address rather than the actual facility address (which may not exist). When these two sources of error are removed, the average distance between the address-matched coordinates and the reported coordinates falls to 1,465 meters (0.9 miles).

GIS and GPS can not only be used to verify existing spatial data; they can also be used to supplement spatial data that are inaccurate. Some point sources may have a variety of sources of spatial data; the lessons of the Wilmington case study and future investigations will be used to develop a methodology which will be used to verify existing data and select the best alternate source of spatial data, if necessary. A composite point source spatial data layer will be created using this methodology.

CONCLUSIONS

In the past, emission inventories were primarily used for regional planning. Current technology and policy directions of the ARB are forcing a rethinking of inventory design, and it is likely that in the near future most inventories will have spatial data as a mission critical component. Emissions from large point sources may be significant and correct identification of the location of these point sources is critical. This paper presents one method used by the ARB to assess the location of large point sources in an environmental justice study area in Wilmington, California. The results of this study suggest that GIS and GPS technologies are valuable tools that allow verification of existing spatial data. Where needed, these tools can supplement inaccurate spatial data. The lessons learned from the Wilmington case study will be applied statewide to improve the overall inventory.

REFERENCES

1. California Air Resources Board. *The 2002 California Almanac of Emissions and Air Quality;* California Air Resources Board, in press. P. 58 ff.
2. Asregadoo, R. California Air Resources Board, Sacramento, CA., *personal communication.*
3. US Environmental Protection Agency, *Compilation of Air Pollution Emission Factors AP-42, Fifth Edition, Volume I: Stationary Point and Area Sources*; Research Triangle Park, NC, 1995; EPA-AP-42.
4. California Air Resources Board, *Emission Inventory Procedural Manual, Volume III – Methods for Assessing Area Source Emissions.* Sacramento, CA., October 1997.

5. California Air Resources Board, *EMFAC 2000 Model*. Sacramento, CA., May 2000. (see http://arbis.arb.ca.gov/msei/msei.htm)
6. CNN News Report, March 7, 2002. Found at *http://www.cnn.com/2002/TECH/internet/03/07/internet.reut/index.html*
7. California Office of Environmental Health Hazard Assessment, *Proposition 65 in Plain Language.* Sacramento, CA., October, 2001.
8. California Air Resources Board, *Airborne Toxic Control Measure for Emissions of Chlorinated Toxic Air Contaminants from Automotive Maintenance and Repair Activities;* Sacramento, CA. April, 2000.
9. California Air Resources Board, *Policies and Actions for Environmental Justice;* California Air Resources Board, December, 2001. P. 4

ACKNOWLEDGEMENTS

Special thanks to Michael Benjamin, Mike Fitzgibbon, Randy Pasek and Bob Fletcher for their invaluable comments.

DISCLAIMER

The opinions, findings, and conclusions expressed in this paper are those of the staff and not necessarily those of the California Air Resources Board. Mention of specific products or services is for informational purposes only and does not imply a recommendation or endorsement of the mentioned products.

KEY WORDS

Inventory, geographic information systems, GIS, global positioning system, GPS, resolution, address matching, geocoding, environmental justice

Table 1. Emissions Differences between Address-Matched Point Sources and Reported Coordinates Point Sources.

	Reported Location		Address Matched Location			
Pollutant	Number of Sources	Emissions (pounds per year)	Number of Sources	Emissions (pounds per year)	Difference (pounds per year)	Percent Difference
Vinyl chloride	2	29	5	664	636	2227%
Methyl chloride {Chloromethane}	1	869	2	12129	11260	1296%
Chloroform	10	1131	10	2324	1193	105%
Copper	20	1309	21	1961	652	50%
Sodium hydroxide	16	8165	18	11072	2907	36%
Trichloroethylene	8	3716	9	4920	1204	32%
Methyl bromide {Bromomethane}	1	10513	2	13009	2496	24%
Diethylene glycol monobutyl ether	4	4571	5	5577	1005	22%
Hexane	2	724	3	872	147	20%
Xylenes (mixed)	79	286223	88	331209	44986	16%
Ethylene glycol	3	832	2	751	-81	-10%
Nickel	27	1357	26	1190	-166	-12%
Particulate Matter	136	1396	132	1204	-191	-14%
Formaldehyde	50	39089	52	33625	-5464	-14%
Total Organic Gas	185	8884	181	7589	-1295	-15%
Ethylene	2	8215	1	6949	-1266	-15%
Ammonia	35	1193643	31	1001946	-191696	-16%
Sulfur Dioxide	140	4912	135	4073	-838	-17%
Oxides of Nitrogen	141	6172	137	5071	-1101	-18%
Hydrochloric acid	19	62656	21	51160	-11496	-18%
Methyl chloroform {1,1,1-TCA}	45	658720	47	504023	-154697	-23%
Manganese	14	1094	13	801	-293	-27%
1,3-Butadiene	11	718	10	447	-271	-38%
Acetaldehyde	30	3973	33	2447	-1527	-38%
Fluorocarbons	5	9409	3	5549	-3860	-41%
m-Xylene	2	540	1	101	-440	-81%
Fluorocarbons (chlorinated)	10	361125	9	19404	-341721	-95%