Daly, A. 2008. *Air Quality Modeling: Pre-Processing and Post-Processing*. Chapter 26 of *AIR QUALITY MODELING - Theories, Methodologies, Computational Techniques, and Available Databases and Software. Vol. III – Special Issues* (P. Zannetti, Editor). Published by The EnviroComp Institute (http://www.envirocomp.org/) and the Air & Waste Management Association (http://www.awma.org/).

Chapter 26

Air Quality Modeling: Pre-Processing and Post-Processing

Aaron Daly

EnviroComp Consulting, Inc., Fremont, CA (USA)
daly@envirocomp.com

Abstract: Environmental scientists now have an abundance of tools and data readily available to conduct and visualize air quality modeling simulations. Examples of using current tools for pre-processing and post-processing in air quality modeling are discussed, along with sources of data and the increasingly important role of Geographic Information Systems (GIS) in post-processing and visualization of modeling results.

Key Words: 2D visualization, 3D visualization, emission data, meteorological data, terrain data, land use/land cover data, aerial photography, satellite imagery, remote sensing, pollution roses, GIS, GUI, ESRI, ArcGIS, ArcExplorer, Global Mapper, Google Earth, SketchUp, georeferencing, geocoding.

1 Introduction

Since air quality models have been programmed in specialized computer languages such as FORTRAN, these programs require the input data to be in specific formats and contain certain meteorological parameters in order to run properly. After a user has collected the input data, or "raw data" that is required for running a model, the user needs to organize this data and also add additional inferred values so that the modeling program can read it and operate. The process of preparing raw data and inferring additional meteorological parameters to use in an air quality program is called the "Pre-processing of data"[1].

[1] http://www2.dmu.dk/AtmosphericEnvironment/cost/fisher.htm

After the data has been pre-processed and successfully read into the model, the output data from the modeling simulation needs to be evaluated for correctness, organized for clarity, and presented in a way so that the modeler and his/her colleagues and clients can understand the results. Sometimes, additional parameters and statistics need to be derived from the model output (e.g., averages or percentiles). This process of evaluating, organizing, presenting, and deriving additional values from the model output is called the "post-processing of data"[2]. Post-processing is also used as a diagnostic examination of the modeling simulation to determine whether the results are plausible and the model is working properly.

The air quality models that are available from the United States Environmental Protection Agency (US EPA)[3] and other regulatory agencies are usually composed of several modules. Besides the main module that performs the actual air dispersion simulation calculations, there are modules that perform pre-processing and post-processing tasks. These modules interact and share data with the main module.

Several private companies[4,5,6] have developed program suites that serve as "front-ends" to modules of several widely-used air dispersion models such as AERMOD[7] and CALPUFF[8]. These front-ends are usually graphical user interfaces (GUIs) that simplify the stages of pre-processing and post-processing and many other tasks such as setting up the modeling domain and assigning receptors.

2 Pre-Processing

The availability of front-ends for air dispersion models has reduced the amount of manual pre-processing work that is typically required for performing air dispersion simulations. However, users will usually still need to pre-process some data to enter into the front-ends. The following sections are examples and recommendations on how to handle data that may need pre-processing before it can be used in an air quality model.

2.1 Emission Data

Regulatory air dispersion models require the emission rate to be defined over regular time intervals that are typically one hour in length. If one is modeling a

---

[2] http://www.flame.org/~cdoswell/forecasting/human_role/future_forecasters.html
[3] http://www.epa.gov/scram001/
[4] http://www.weblakes.com/
[5] http://www.beeline-software.com/
[6] http://www.breeze-software.com/
[7] http://www.epa.gov/scram001/dispersion_prefrec.htm#aermod
[8] http://www.epa.gov/scram001/dispersion_prefrec.htm#calpuff

release scenario with emissions that vary in time or last for a short time, they should be averaged appropriately to fit the requirements of the model. For example, in simulating hourly averages, a release scenario that involves 2000 grams of pollutant released over a period of 6 minutes is averaged in the following way to produce an equivalent emission rate for the entire first hour of the release:

$$\text{Hourly Emission Rate} = \frac{2000\ \text{grams}}{3600\ \text{seconds}} = 0.5556\ \text{g/s} \tag{1}$$

Using an air dispersion model for accident reconstructions (e.g., fires or explosions) may require emission modeling as a separate step to characterize the substances that are being emitted and to develop a timeline of their emissions. The emission data used in air quality modeling are generally computed by an engineering evaluation of the release. Methodologies and guidelines are also provided by the US EPA[9,10,11].

2.2 Meteorological Data

Most regulatory air dispersion models require meteorological data from surface stations and upper-air stations. The front-ends to meteorological pre-processing modules, such as AERMET[12] and CALMET[13], are able to read meteorological data in many formats, but it is likely that data from representative surface stations may be in an unreadable format. For example, currently, the National Climatic Data Center[14] (NCDC) provides archived surface station data in different formats than SAMSON[15] and SCRAM/MET144[16], which are two conventional formats read by front-ends to meteorological pre-processors. If modelers require surface station data in formats such as SAMSON or SCRAM/MET144, they can try sources such as WebMET[17] and WorldGeoData[18]. Also, the Weather Underground[19] website has historical METAR records for weather stations worldwide. The advantage of purchasing data from a service like WorldGeoData is that the data has been quality-checked and missing data has been corrected in accordance with US EPA guidelines[20].

---

9 http://www.epa.gov/otaq/ap42.htm
10 http://www.epa.gov/ttn/chief/efpac/index.html
11 http://www.epa.gov/OMSWWW/models.htm
12 http://www.epa.gov/scram001/metobsdata_procaccprogs.htm#aermet
13 http://www.src.com/calpuff/download/CALMET_UsersGuide.pdf
14 http://www.ncdc.noaa.gov/
15 http://www.webmet.com/MetGuide/Samson.html
16 http://www.webmet.com/MetGuide/SCRAMSurface.html
17 http://www.webmet.com
18 http://www.worldgeodata.com/home.aspx
19 http://www.wunderground.com/
20 http://www.epa.gov/scram001/guidance/met/mmgrma.pdf

The most commonly used weather stations for regulatory air quality modeling in the US have been nearby National Weather Service (NWS) stations, which are usually located at major airports. These airport observations may not be the most suitable data for an air quality modeling application, because even though the data is reported every hour, the wind observations are not hourly averages. Instead, the airport winds are only 2-minute averages[21], and therefore may not properly characterize the winds over the entire hour.

Alternate sources of meteorological data that may have hourly-averaged winds or wind data collected over shorter intervals include monitoring stations maintained by environmental regulatory agencies (e.g., CARB[22] and LDEQ[23]); agricultural networks (e.g., CIMIS[24] and LSU AgCenter[25]); academic institutions; and private industry[26]. Data retrieved from these stations will usually have to be reformatted to a style compatible with the pre-processor. The front-end may have the ability to import data from a spreadsheet format (e.g., Microsoft Excel) and convert it to a common format like SAMSON[27]. If one needs to compute hourly averages or standard deviations from wind data reported in smaller time intervals, the US EPA guidelines should be consulted.

One should also make an effort to understand several characteristics about a weather station, including its geographic coordinates, terrain height of its location, height of its sensors, the type of land use in the surrounding area (e.g., rural, urban, residential, industrial, wetlands, grasslands, forests, etc.). It is also useful to know how the instrument is sited. For example, a wind sensor located in an open area with no buildings nearby may give more suitable wind data that one located close to or on top of a building.

Upper-air radiosonde data is commonly reported in an FSL format[28], as well as different NCDC formats. Manual pre-processing is usually not needed for radiosonde data since these formats are understood by front-end programs. Historical worldwide radiosonde data is available at the FSL website[29] or NCDC.

If mixing height data is needed, it can be calculated from radiosonde data by using a program[30] from the US EPA. The version of this program that is available currently (98340) is an older utility (December 1998), which is not Y2K compliant, and may have to be recompiled for computers that were built after the

---

[21] http://www.ofcm.gov/fmh-1/pdf/E-CH5.pdf
[22] http://www.arb.ca.gov/homepage.htm
[23] http://www.deq.louisiana.gov/portal/
[24] http://wwwcimis.water.ca.gov/cimis/welcome.jsp
[25] http://www.lsuagcenter.com/weather/
[26] Other U.S. states have similar websites for meteorological and air quality measurements. Here we cited California and Louisiana sites only.
[27] http://www.webmet.com/MetGuide/Samson.html
[28] http://raob.fsl.noaa.gov/intl/fsl_format-new.cgi
[29] http://raob.fsl.noaa.gov/
[30] http://www.epa.gov/scram001/metobsdata_procaccprogs.htm#mixing

year 2000. The source code and documentation are available with the program, so the mixing heights algorithm can be duplicated if the program does not work.

Mixing height data can also be purchased from NCDC and WorldGeoData[31], and some historical data is available from the US EPA SCRAM website[32].

2.3 Maps/Aerial Photography/Satellite Imagery

Front-end programs often allow the user to insert a basemap image into the modeling domain. The modeler can scan a paper document that contains a street map or industrial site map, and import the image into the program. In order to properly align the image in the domain, the modeler needs to have geographic coordinates of at least two points on the image. More details about this procedure (called georeferencing) are given in section 4.

Using a procedure similar to maps, aerial photography and satellite imagery can be scanned and imported into the modeling domain, provided that the modeler has accurate geographic information about them. It is preferred that imagery be "orthorectified", which means that any distortion due to terrain or camera angle is removed. This results in an image with a uniform scale that can be used as a map in a modeling domain. Sources of orthophoto imagery are numerous, and include the United States Geological Survey (USGS)[33], DigitalGlobe[34], Terraserver[35], GeoEye[36], and MapMart[37].

The most common type of imagery has the camera aimed directly down at the ground, but imagery where the camera is at an angle ("oblique imagery") is becoming more popular. Oblique imagery, also called "Bird's Eye View" imagery, can be used for a three-dimensional perspective. Pictometry[38] specializes in oblique imagery, and they sell imagery along with specialized tools that can give geographic coordinates of ground locations and measure horizontal and vertical distances in the images.

2.4 Terrain and Land Use/Land Cover

Terrain data is freely available on the internet in several formats. The common format for U.S. terrain data is Digital Elevation Model (DEM), which has a resolution up to 1 arc-second, which is about 30 meters for the contiguous U.S.

---

31 http://www.worldgeodata.com/home.aspx
32 http://www.epa.gov/scram001/mixingheightdata.htm
33 http://www.usgs.gov/
34 http://www.digitalglobe.com/
35 http://www.terraserver.com/
36 http://www.geoeye.com/
37 http://www.mapmart.com/
38 http://www.pictometry.com/

One source of DEM is WebGIS[39]. A source of highly accurate terrain data is the Shuttle Radar Topography Mission[40], which is available for the entire world. Depending on the location, this terrain data is in 90 meter or 30 meter resolution. Land use or land cover data is often used in conjunction with terrain data, and both are available from WebGIS, TRC[41], the National Geophysical Data Center[42], and the Global Land Cover Facility[43].

It can be challenging to convert terrain and land use data into the correct format required for a pre-processor. The program Global Mapper[44] is a useful tool for this task. It can retrieve, read, display, and convert many types of imagery and geophysical data into different formats, and it is relatively inexpensive and easy to use.

When a modeler imports any extra layers of data (i.e., imagery, terrain, land use/land cover) into pre-processors, he/she needs to be aware of the coordinate system and datum associated with the modeling domain and the data layers so that the data layers are placed properly in the domain. A common mistake is to confuse the two datums NAD27 and NAD83. NAD83 is a modern datum that was developed in 1983, but geographic data may still be archived in the older NAD27 datum developed in 1927. Not accounting for the correct datum could result in errors up to 100 meters. A program like Global Mapper or other Geographic Information Systems (GIS) programs are very useful in identifying the coordinate systems and datums of imported data and aligning them correctly. More information about GIS programs is given in section 4.

3 Post-Processing

After an air quality simulation has successfully run, it is often desirable to have a graphical representation of air pollutant concentrations. This visualization can be useful in determining the impact of a plume of pollutants on a nearby community and in determining if the simulation was a reasonable representation of the real world.

Many front-ends to air quality models have their own post-processing modules that display contours of concentrations in the modeling domain. If one has used an aerial photograph or streetmap as a basemap, one can then evaluate the impact and correctness of a plume.

Post-processing modules can be useful diagnostic and visual tools, but if a simulation project involves producing very accurately contoured data or high-

[39] http://www.webgis.com
[40] http://srtm.usgs.gov/
[41] http://www.src.com/datasets/datasets_main.html
[42] http://www.ngdc.noaa.gov/
[43] http://glcf.umiacs.umd.edu/index.shtml
[44] http://www.globalmapper.com/

quality graphical output, or knowing the relationships between the simulated plume and several other types of data, it is very useful to import the modeling results into a GIS workspace. This is a straightforward process if the modeling output is given as a text file with the coordinates and concentrations at each receptor used in the simulation. This is the case for models like AERMOD and CALPUFF. Other US EPA models, like ALOHA and SLAB, may give output in the form of the coordinates of each concentration contour curve, and extra steps are needed to import these contour levels into a GIS workspace.

Figure 1 below shows an example of a plume overlaid on an aerial picture in a GIS workspace.

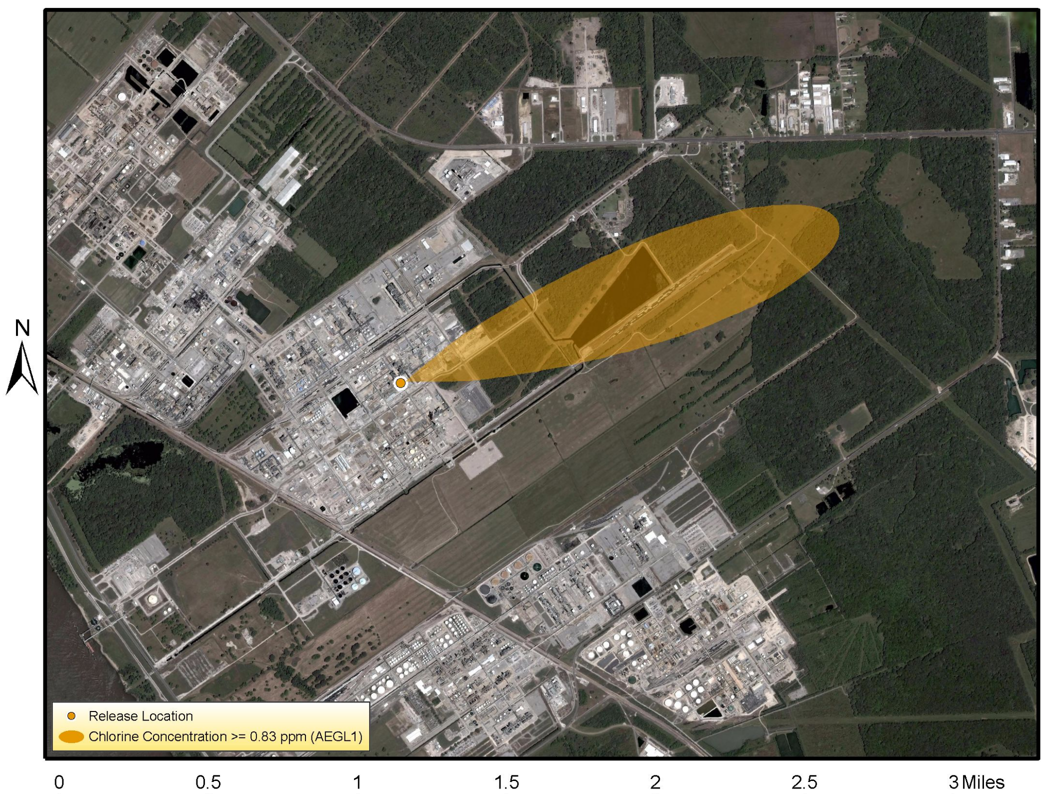

Figure 1. Post-processed plume in a GIS workspace.

The GIS environment allows the user to have significantly more control over how the plume is contoured and displayed than in typical post-processing modules, and the user can also overlay many other types of data along with the plume. Figure 2 below shows the same plume along with extra annotation that shows the time of the simulation and the wind direction.

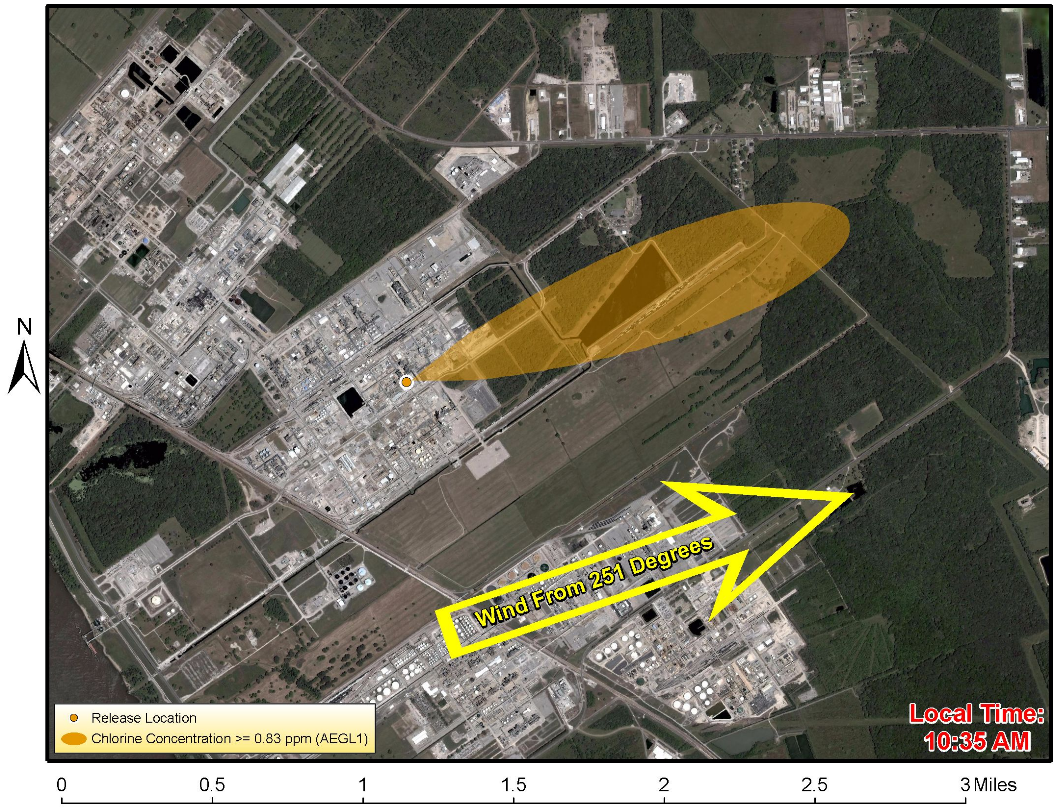

Figure 2. Post-processed plume in a GIS workspace with additional annotation.

Figure 3 shows the same plume with an overlay of the locations of several air quality measurements as a diagnostic exercise. Importing air quality measurements that have Global Positioning System (GPS) coordinates assigned to them is a straightforward task in the GIS environment.

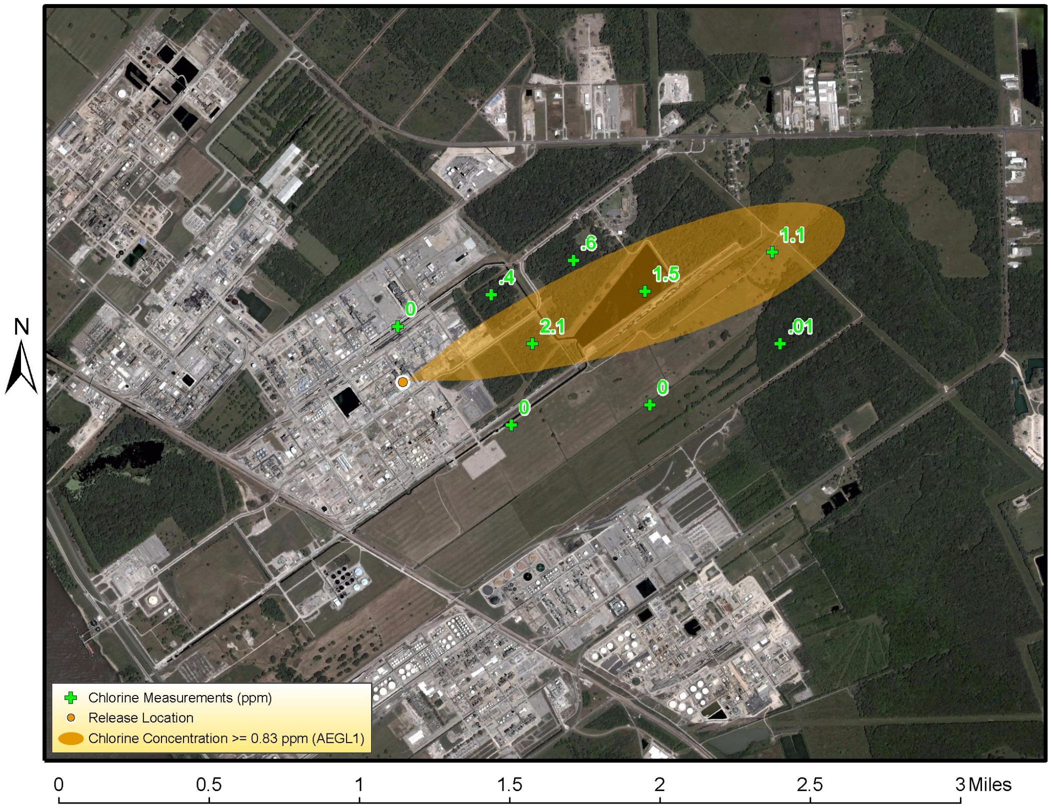

Figure 3. Post-processed plume in a GIS workspace displayed with locations of air quality measurements.

One may be interested in the subset of a data layer contained inside the plume. Figure 4 shows the same plume along with a layer that represents locations of claimants in a litigation case.

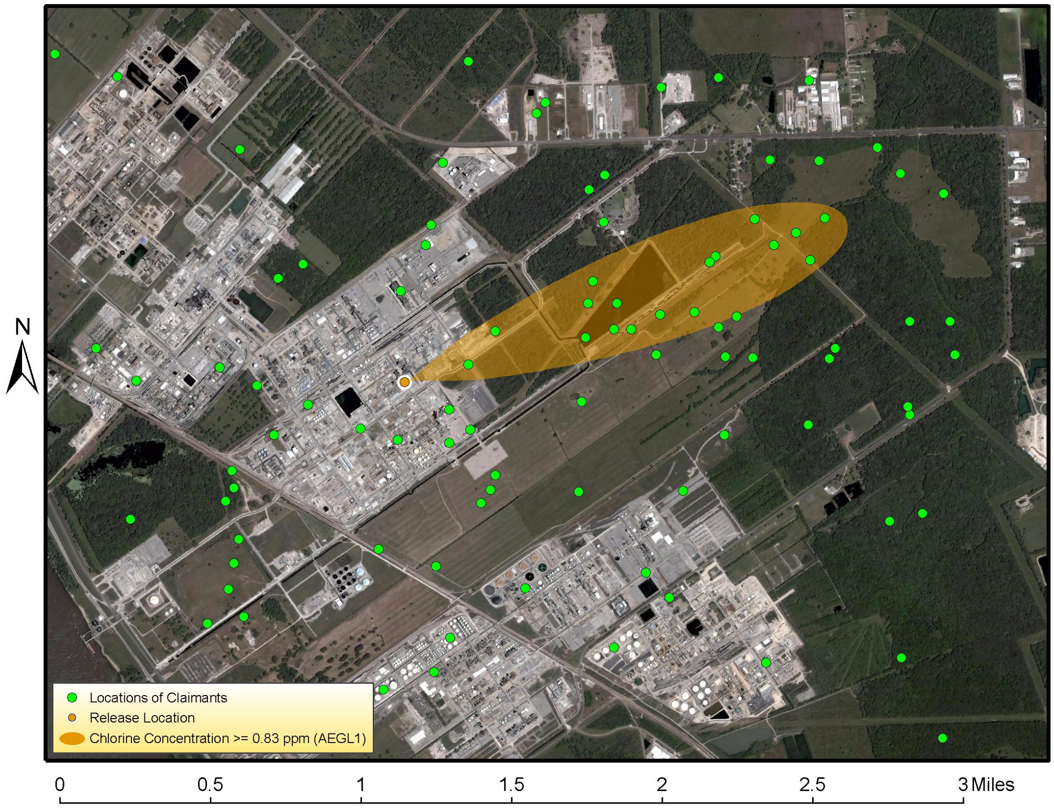

Figure 4. Post-processed plume in a GIS workspace displayed with claimant locations.

The GIS program can automatically extract the points inside the plume, which results in Figure 5. In this figure, the blue dots are the claimants inside the plume.

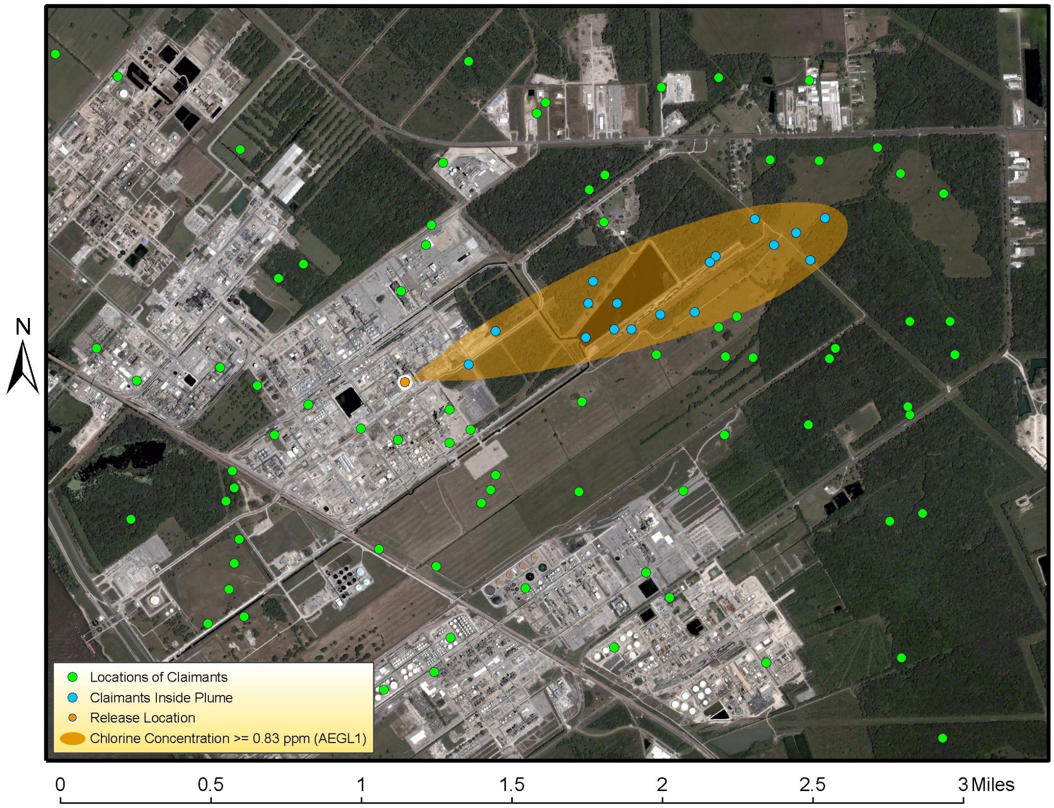

Figure 5. Post-processed plume in a GIS workspace displayed with locations of claimants in the area and also claimants inside the plume.

Since air quality models are often used to predict the impact of pollutants on people at ground level, two-dimensional (2D) representations like the preceding figures are often sufficient to tell the story. However, it may be useful to have a three-dimensional (3D) visualization of a plume if the simulation is over a very large area. Figure 6 is an image from a Google Earth[45] time animation of a 3D volcanic plume simulation of the Mount Etna eruption[46] on July 24, 2001. (Click on Etna Animation in the CD version of the book to view the full animation).

[45] http://earth.google.com/
[46] http://puff.images.alaska.edu/Google_Earth2/Etna_24_July_2001.gif

Figure 6. Post-processed plume of volcanic ash from the Mt. Etna eruption of July 24, 2001, in the Google Earth workspace.

Figure 7 is an image from a Google Earth time animation of a 3D volcanic plume simulation of the Mount St. Helens eruption[47] in May 1980. (Click on St. Helens Animation in the CD version of the book to view the full animation).

[47] http://puff.images.alaska.edu/Google_Earth2/StHelens_May_1980_side.gif

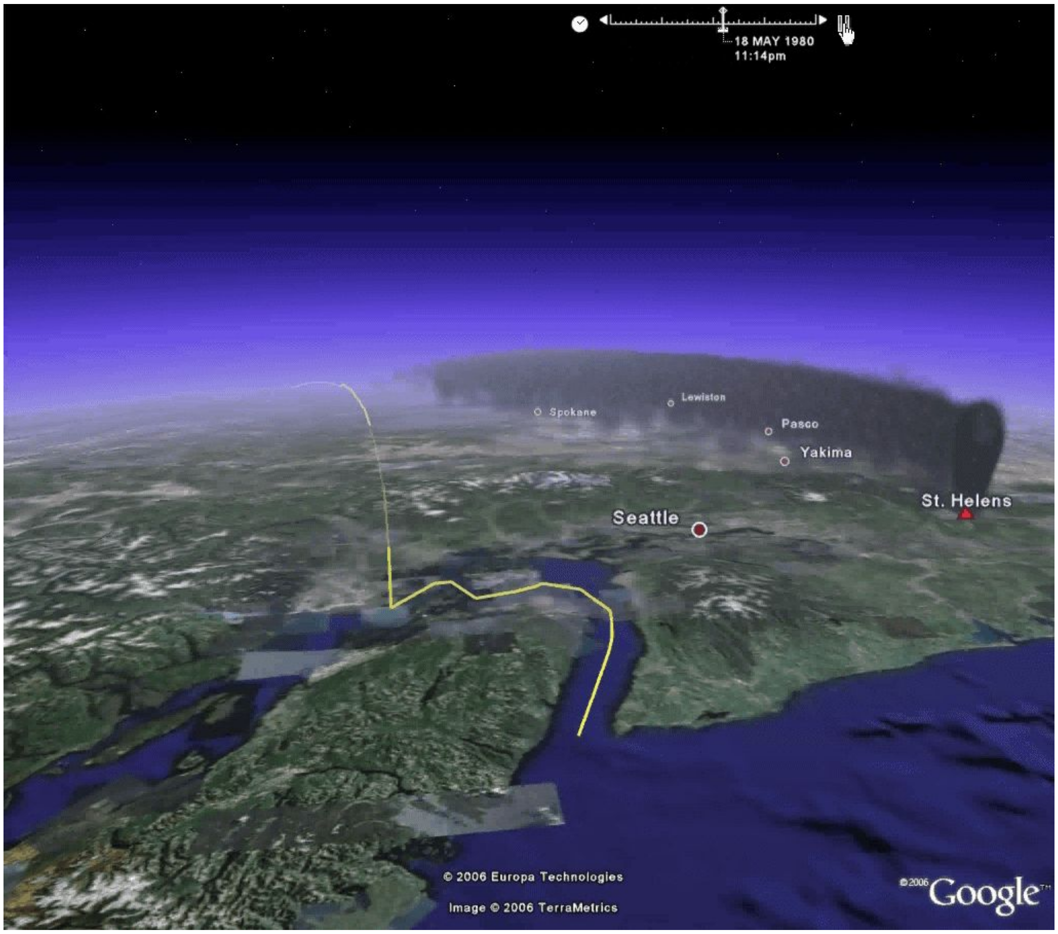

Figure 7. Post-processed plume of volcanic ash from the Mt. St. Helens eruption in May, 1980, in the Google Earth workspace.

Statistical analysis can be an important part of post-processing. For example, the output of air quality simulations are typically a sequence of air concentrations of a certain substance averaged over a period of time at each receptor. The average concentration values are used to create visualizations of plumes, but it can also be useful to look at concentrations over shorter intervals to see the range of concentration values. Figure 8 shows a timeline of concentrations at a single receptor in an air quality simulation. In this case, the peak 24-hour concentrations can be 25-30 times higher than the 5-year average concentrations. Also plotted is a Level of Concern (LOC) for this particular chemical. One can focus on days when the modeled concentrations exceeded this LOC as the next part of the modeling analysis.

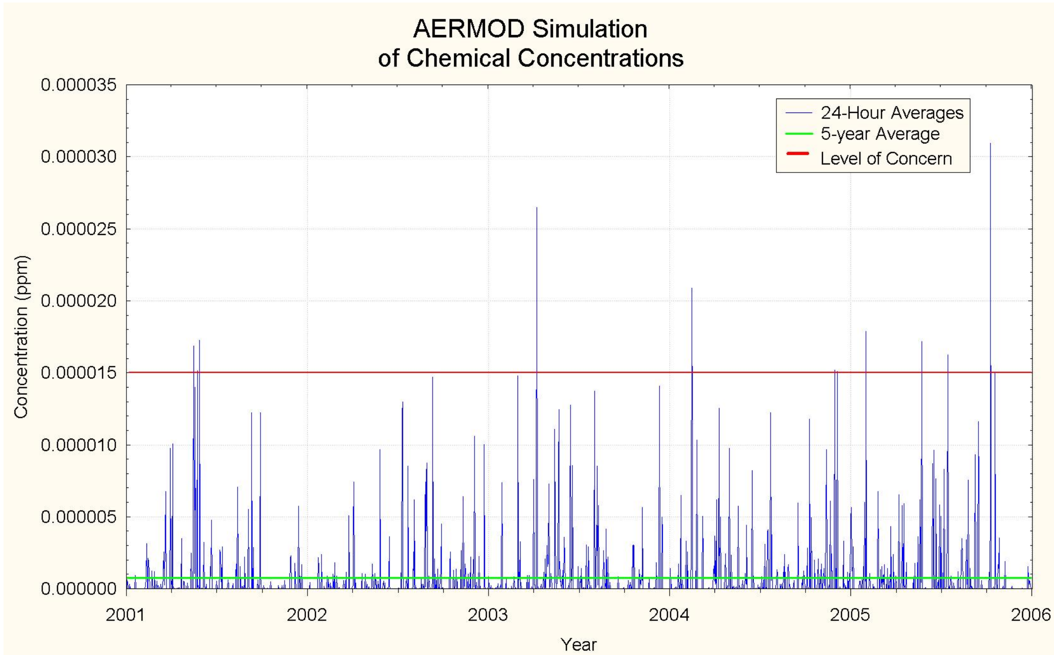

Figure 8. Timeline of simulated average concentrations from AERMOD.

Another useful statistical tool is a pollution rose. This tool combines wind direction data and air quality measurements to determine what directions the pollution is coming from. Figure 9 is a pollution rose for the tracer gas PTCH[48]. The petals of the rose point to the directions the tracer is coming from, and a larger petal means that more tracer gas is coming from that direction. In the figure, the major source of PTCH is from the south.

[48] perfluoro-trimethylcyclohexane.

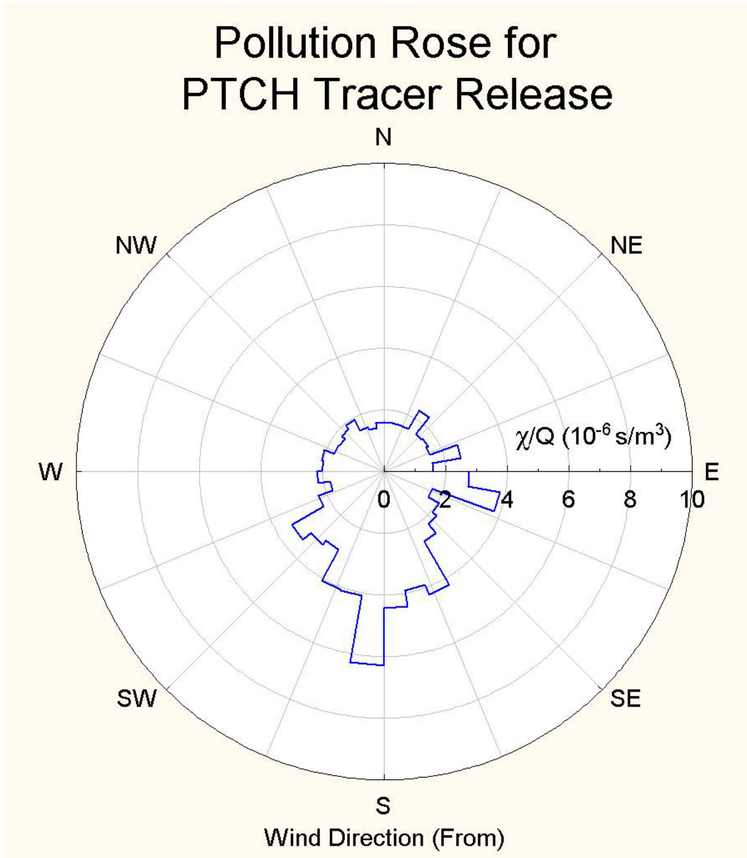

Figure 9. Pollution Rose for a Tracer Gas.

4 GIS in Air Quality Modeling

The field of GIS is constantly evolving and is gradually becoming more commonplace. The availability of free utilities like Google Earth and ArcExplorer[49] make GIS data and tools accessible to anyone with newer computers and fast internet access.

GIS can be used in an air quality modeling simulation from start to finish. For example, one may need to create an area source polygon for running AERMOD. A GIS user can view an aerial orthophoto, identify the source boundaries, and use tools to digitize the source polygon. The coordinates of the polygon source can then be entered into a front-end of an air quality model.

One can also use GIS to assign a scale and a coordinate system to an image. This process is called “georeferencing”. All that is needed are the coordinates of several locations in the image, and the GIS program georeferences the image so that it can be used as a basemap layer. Google Earth is useful for obtaining

[49] http://www.esri.com/software/arcgis/explorer/index.html

coordinates since it displays layers of georeferenced orthophotography for the entire Earth.

Some GIS programs can assign coordinates to street addresses. This process is called "geocoding". Geocoding is useful when one has a large list of locations that need to be displayed on a map. Both Google Maps[50] and Google Earth can geocode addresses. The program ArcGIS[51] developed by ESRI[52] also geocodes addresses, and one example of an online geocoding service is EZ-Locate[53].

GIS programs are also useful for "geoprocessing" tasks. These tasks include extracting data that lie inside of another layer (e.g., the claimants inside the plume in Figure 5), extracting data that are a certain distance from another layer, converting discrete point data into continuous data surfaces and vice versa, and interpolating and contouring data.

One very powerful aspect of a high-end GIS program like ESRI's ArcGIS is that it organizes many data layers such as discrete points, polygons, images, data surfaces, tables, and contours into a "geodatabase". This geodatabase is in Microsoft Access format, which means that it is accessible to non-GIS users who can interact with the database and update its information. Previously-created Access databases can also be imported into ArcGIS and integrated with spatial data to create geodatabases.

High-end GIS programs like ArcGIS can also be programmed with customized tasks. It is feasible to create tasks for air quality modeling directly in the GIS environment so that with a click of a button, an air quality simulation runs and plumes of pollutants are displayed over imagery and terrain.

Another advanced use of GIS is remote sensing where multi-spectral imagery can be classified into different types of land use/land cover, which may play an important role in air quality simulations. An example is shown in Figure 10 below. The top picture is a USGS Landsat Thematic Mapper (TM)[54] image of Tippecanoe County, Indiana from July 17, 1986. The bottom picture is the same image classified into different types of crops[55]. The program used for the classification is a freeware program called MultiSpec[56] that is developed by Purdue University.

[50] http://maps.google.com/
[51] http://www.esri.com/products.html
[52] http://www.esri.com/
[53] http://www.geocode.com/
[54] http://edc.usgs.gov/guides/landsat_tm.html
[55] Images are taken from http://cobweb.ecn.purdue.edu/~biehl/MultiSpec/Intro5_01.pdf
[56] http://cobweb.ecn.purdue.edu/~biehl/MultiSpec/

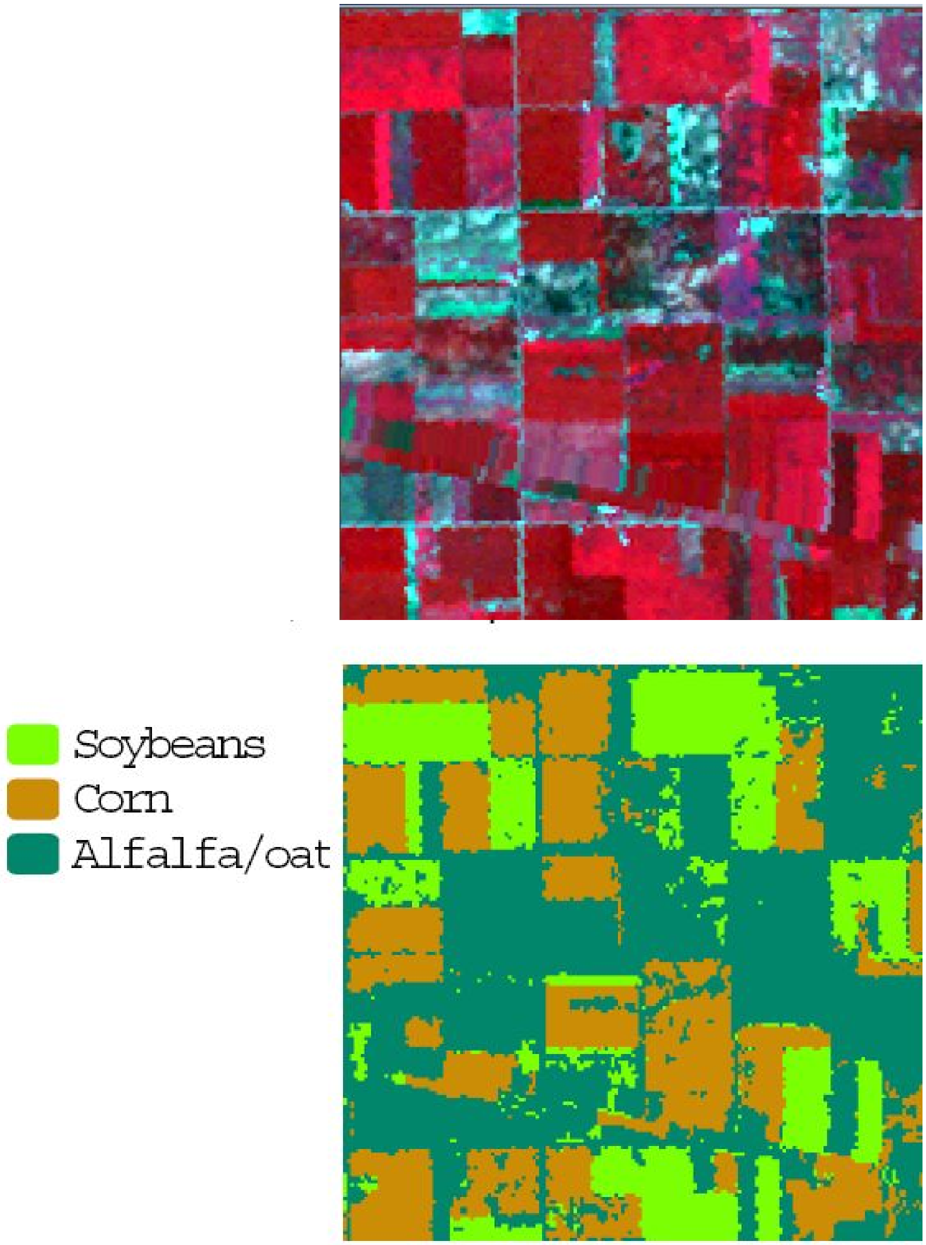

Figure 10. USGS Landsat TM image of fields in Tippecanoe County, Indiana (top), and the same image classified into different types of crop cover (bottom).

Computational Fluid Dynamics (CFD) models can also benefit from the use of GIS. The drafting tool SketchUp[57] can be used to create 3D buildings that can be georeferenced and exported to several formats, as well as visualized in Google Earth. Figure 11 shows an example of some buildings that were created in SketchUp, and later were georeferenced in Google Earth and used in FLUENT[58].

---

[57] http://www.sketchup.com/
[58] http://www.fluent.com/

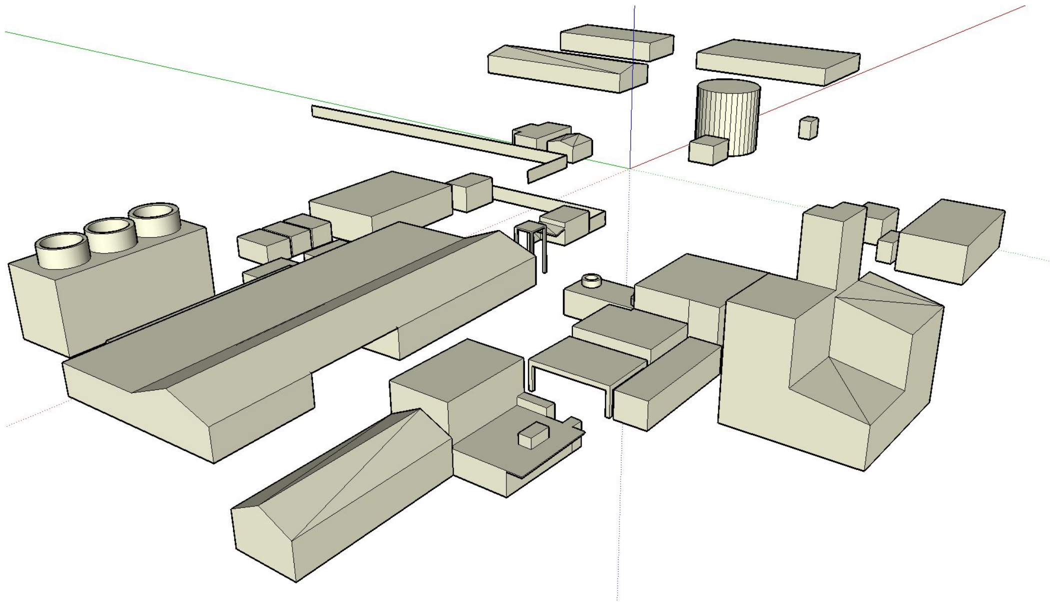

Figure 11. 3D buildings created with SketchUp.

5 Summary

Many tools and sources of data are available to an air quality modeler for pre-processing and post-processing. GIS is becoming indispensable as a tool, and with free programs like Google Earth and ArcExplorer, an air modeler can "visit" a site and look at the local terrain and land cover as a first step in their modeling project. The modeler can also "visit" the sites of nearby meteorological stations and determine which stations are most suitable for air quality modeling. GIS tools can also be used to create sources for air quality models and display the results of the simulations. If advanced spatial analysis like georeferencing, geocoding, creating geodatabases, or analyzing remotely-sensed imagery is needed, high-end programs like ArcGIS are available. Less expensive, "lightweight" GIS programs like Global Mapper are also available and can be very useful for visualization and small geoprocessing tasks.