

## A GIS BASED AIR QUALITY SYSTEM FOR THE APULIA REGION, SOUTHERN ITALY

I. Schipa<sup>1</sup>, C. Mangia<sup>1</sup>, A. Tanzarella<sup>1</sup>, D. Conte<sup>1</sup>, GP Marra<sup>1</sup> and U. Rizza<sup>1</sup>

<sup>1</sup>Institute of Atmospheric and Climate Sciences ISAC -CNR-LECCE, Italy

**Abstract:** Apulia region in the Southern Italy is frequently characterised by high photochemical pollution levels in the warm period and by high levels of PM<sub>10</sub> and NO<sub>2</sub> in the winter season. Emissions in the area derive essentially from urban, shipping and industrial activities. The main industrial activities are related to the iron and steel industry (one of the largest in Europe) on the western coast and to two coal power plants on the eastern coastline.

A GIS based air quality system has been developed to support local authorities in air quality management for the region. The proposed modeling system is based on RAMS (Pielke et al., 1992) and CALMET (Scire et al., 2000) meteorological models and on CALPUFF (Scire et al., 2000)/ CALGRID (Yamartino et al., 1989) dispersion models.

Diffuse emissions for the domain were obtained from the national CORINAIR data base ([www.sinanet.apat.gov.it](http://www.sinanet.apat.gov.it)) and were pre-processed by a new-developed tool GEM-PP (Gis EMISSION Pre-Processor) based on open source GIS. Point sources emissions are obtained by local inventory.

Meteorological and dispersion simulations were performed for the year 2005. Predictions have been compared with concentration data from the air quality monitoring network. Results evidence a good correlation between predictions and measurements for O<sub>3</sub>, NO<sub>x</sub>, SO<sub>2</sub> with most of data in factor of two of the measurements for rural stations and a tendency to underestimate measured data in urban stations. Overall the model tends to underestimate CO measurements. The uncertainty of the predictions are analysed and discussed in terms of the emission calculations, dispersion modelling and monitoring site.

**Key words:** modelling system, emission pre-processor, concentration data.

### 1. INTRODUCTION

The management of air quality has become, over the last few decades, a major problem for decision makers, because involves the knowledge of meteorological factors, chemical processes and emission inventories. Air quality models, combining meteorology and chemistry, are fundamental to understand the different mechanisms that lead to the accumulation of atmospheric pollutants. So they can play an important role in interpreting monitoring data and supporting strategies in emission control strategies (Russell and Dennis, 2000). Another critical point is the estimation of the emission inventory that must be very accurate when dealing with reactive pollutants, like O<sub>3</sub> and aerosols.

In this work a GIS based air quality system has been developed to support local authorities in air quality management for the Apulia region, in southern Italy. This region is characterized by the presence of two big industrial sites near the coasts, in the Taranto west coast and Brindisi areas in eastern coast (see Fig. 1). Moreover, the region is frequently characterised by high photochemical pollution levels in the warm period and by high levels of NO<sub>2</sub> in the winter season.

In this work the modelling system RAMS/CALMET/CALGRID has been applied to estimate the emissions and spatial distribution of the air pollutants over the south-east corner of Italy. The emission factors for the simulation domain have been used to prepare an emission inventory that will be computed and stored in databases of a GIS (Geographical Information System). The use of GIS technologies has enabled mapping of emissions and air quality levels that have been compared with measurements.

### 2. THE INVESTIGATED AREA

The investigated area is located in the Mediterranean sea in south-eastern corner of Italy. It is bordered by the Adriatic Sea to the east, the Ionian Sea to the southeast, and the Strait of Otranto and Gulf of Taranto to the south (Fig. 1). The region is 180 km long in NW-SE direction, with an average width of 60 km. It is generally flat, with small hills (less than 200 m) in the southeastern area and medium high hills (about of 500m ) in the northern part.



Figure 1. The investigated area.

Emissions in the area derive essentially from urban, shipping and industrial activities. The largest industrial emissions are placed near the coast, in Taranto and Brindisi areas. In Taranto the main industrial activities are related to the iron and steel industry (one of the largest in Europe) and oil refinery with hydrocarbons transformation processes. The above industrial activities use the harbour of Taranto to download primary goods and to delivery final products. In Brindisi the main industrial activities are due to two coal power plants and to a petrochemical plant. Table 1 shows the total emissions of NO<sub>x</sub>, CO, SO<sub>x</sub> and VOC coming from different emission sectors for year 2000 (CORINAIR). It is evident as the greater contributions to the emissions come from the industrial division and from the transport sectors that include vehicles, airports, harbours and railroads.

Actually, the air quality monitoring network of the investigated area is composed by 29 monitoring sites handled by Regional Environmental Protection Agency and other local public authorities. The position of each station is indicated in Figure 2. Eleven station are classified as industrial monitoring sites (I), seven station are classified as background monitoring sites (B), five stations are classified as traffic/urban (T/U) monitoring sites, three stations are classified as traffic/suburban monitoring sites (T/S) and, finally, three monitoring stations are classified as traffic/industrial (T/I) monitoring sites. All these monitoring sites provide mean hourly data of NO<sub>2</sub>, fourteen provide mean hourly data of CO, nine provide mean hourly data of O<sub>3</sub> and only one station provides mean validated hourly data of SO<sub>2</sub>.

Table 1. Total emissions (Mgy<sup>-1</sup>) of the main pollutants on the domain, from CORINAIR data referred to year 2000.

Emission/Sector	Industry	Transport	Biogenic	Others
NO <sub>x</sub>	35244	45402		3462
SO <sub>x</sub>	62204	2143		506
CO	327757	234534		54781
VOC	11131	50688	2036	19781

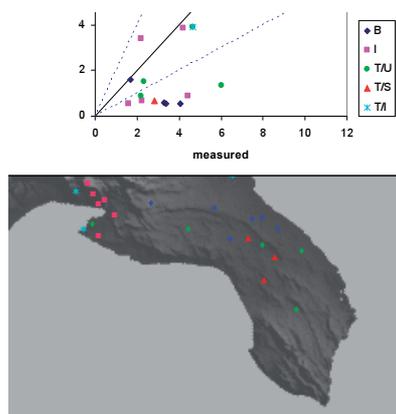


Figure 2. Simulation domain and location of monitoring site.

### 3. THE GEM-PP EMISSION PRE-PROCESSOR

Emissions from industrial point sources were derived from a regional database ([www.arpa.puglia.it](http://www.arpa.puglia.it)). Emissions from all other sectors have been estimated on the basis of last available Italian CORINAIR databases ([www.sinanet.apat.gov.it](http://www.sinanet.apat.gov.it)) and elaborated by means of an emission pre-processor, GEM-PP (Gis Emission PreProcessor, Marra et al., 2005).

The objective of the GEM-PP project (GIS Emissions Pre-Processor) is to develop a suitable GIS (Geographic information systems) tool to provide gridded emission input data using a top-down approach for air quality models. The tool has been developed using open source software for all GIS components, such as geo-processing libraries and digital mapping tools, and a database management system (RDBMS with GIS extensions). The GIS components are valuable for data manipulation of emissions over geographic entities, while the data management system is used as a backbone for the proposed processing emissions system (Janssen et al., 2003). It is worth noting that the GIS approach does not improve the accuracy of the emissions inventory, but does assist in obtaining gridded data and any other manipulations over geographic information. The utilization of a RDBMS system that can provide GIS functionalities improves the efficiency in such operations because it provides a structured workspace for the storage and retrieval of information. This permits the creation of a variety of input emissions for many different photochemical transport models, each with different requirements and file formatting. In the past few years, several commercial and open-source software solutions have become available to the GIS industry. In particular, the open-source GIS projects provide several powerful tools that can satisfy many major requirements of both scientific and the industrial applications. These open-source tools permit the development of affordable, low-cost software projects that can be utilized by the scientific community.

We selected components that matched the imposed functional requirements and that were able to satisfy technical constraints, such as flexibility, robustness, and interoperability with other software packages. In particular, we used the following key tools:

- As a GIS enabled enterprise database, we used PostgreSQL (<http://www.postgresql.org>) with PostGIS extensions (PostGIS documentation). PostgreSQL is a powerful, open-source relational database system (RDBMS), while PostGIS (<http://www.postgis.org>) [4] adds support for geographic objects. PostGIS follows the OpenGIS "Simple Features Specification for SQL" (OGC specifications, 1999) and provides GIS-like abilities by means of the SQL language (including buffering, intersection, within, and distance).
- As data format, geographic reference, and projection conversion utilities, we adopted the Geospatial Data Abstraction Library/Simple Features Library (GDAL/OGR, <http://www.gdal.org/>), which represents a translation library for rastered geospatial data formats. As a library, it presents a single abstract data model to the calling application for all supported formats. The related OGR library provides a similar capability for simple featured vector data. This project provides several command-line utilities for converting and projecting raster and vector data. The GDAL utilities handle raster data and the OGR utilities handle vector data. Both are bundled together under the GDAL project.
- As geo-processing libraries, we used the PROJ.4 Cartographic Projections Library (<http://www.remotesensing.org/proj>), which represents a complete package offering command-line facilities and a library for performing forward and inverse transformations of cartographic data to or from Cartesian data with a wide range of selectable projection functions.
- As a digital mapping tool, we used the University of Minnesota (UMN) MapServer (<http://mapserver.gis.umn.edu/>), an open source development environment that excels at rendering spatial data, such as maps, images, and vector data. Usually, MapServer is used for building spatially enabled Internet applications, but in this project, it has been used as a spatial data rendering engine. GEM-PP incorporates MapServer capabilities directly into the application through an application program interface (API) named MapScript (Mitchell, 2005). The MapScript API provides an assortment of powerful tools that are robust and ready to use, and exposes the functionality of MapServer to various scripting languages. In particular, we used the Simplified Wrapper and Interface Generator (SWIG, <http://www.swig.org/>) to port the MapScript API into our development environment to render maps.

Finally, the GEM-PP has been developed for UNIX-like systems to satisfy the requirements of the most commonly used photochemical models. It was implemented in Ruby (<http://www.ruby-lang.org>), an open-source object-oriented scripting programming language, because this language allows the development of robust applications in a time efficient manner, thanks to its object oriented approach.

### **The emission data processing**

The GEM-PP permits the emissions data processing of spatial disaggregation, time modulation, chemical speciation, and lumping.

The spatial disaggregation consists of dividing emissions at the province level down to a municipality level by using a specific set of spatial proxy surrogates. These are variables highly correlated with emissions, which can be obtained from local statistical data sources and land-use information.

The time modulation calculates the temporal modulation of emissions splitting annual emissions on an hourly basis, through a set of temporal profiles for each pollutant sector: a 12 element monthly profile, a 7 element weekly profile, and a 24 element daily profile. This step is performed using information about the activities responsible for the emissions, and considering the factors for each of them on working days and holidays, on a monthly and hourly basis. The NMVOC chemical speciation and lumping is necessary to determine the detailed composition of an emitted VOC associated to a single emission activity. For VOC splitting we refer to the methodology discussed in EPA (2002), Monforti and Pederenzoli 2004, and Carnevale et al., 2006.

The last step involves gridding the municipal level hourly emissions stored in the GIS database into a geo-referenced grid that overlaps the numerical grid. In the GIS context, this can be done in two phases. The first phase establishes an ensemble of geographical polygons to build vectorial grid cells using southwest grid angle coordinates, cell widths and heights, and grid dimension. The second phase performs a map overlapping (with a GIS operation) between the municipal level hourly emissions maps and the vectorial grid cells.

Figures 3 (a-b) depict the distribution of annual NO<sub>x</sub> and NMVOC emissions for all emission sources over the grid simulated domain. Figures evidence how the largest emissions are localized in Taranto and Brindisi areas due to industrial and energy production activities hosted there. As regard the NMVOC, Figure 3 (b) shows some peaks in the northwestern part in correspondence to the hill lands of Murges.

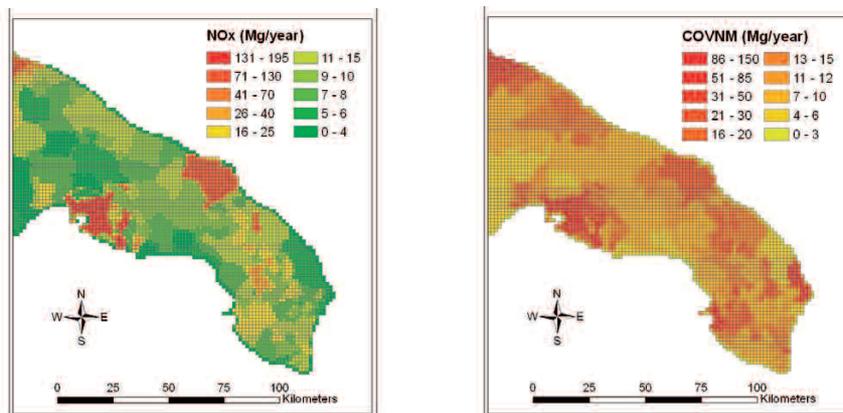


Figure 3. Total emissions distribution of NO<sub>x</sub> (a) and VOC (b) species.

#### 4. MODELLING SETUP

Numerical simulations have been carried out by using the two meteorological models RAMS (Pielke et al., 1992) and CALMET (Scire et al., 2000a), the emission pre-processor GEM-PP and the dispersion models CALGRID (Yamartino et al., 1992). RAMS is initialised with ECMWF data, while the boundary conditions for CALGRID model are provided using the chemical fields derived from the “CHIMERE” simulations at continental scale (<http://euler.lmd.polytechnique.fr/chimere>) with 0.5° of horizontal resolution and 6 vertical levels (Schmidt et al. 2001; Menu et al. 2007).

The simulations with the RAMS model have been performed in a two-way nested grids configuration with three grids. The coarsest grid has a resolution of 40x40 horizontal grid points with a grid spacing of 24 km; the medium grid has a resolution of 82x70 horizontal grid points and a grid spacing of 6 km and the inner grid has a resolution of 108x120 horizontal grid point with a mesh less than 2 km. The vertical atmosphere is subdivided into 25 levels with different thicknesses, from 100 m near the surface, gradually stretching up to a maximum of 1000 m at the top. For initial and boundary conditions, the Isentropic Analysis System (ISAN) package (the module of RAMS for the generation of data analyses) is used. Initially, analyzed fields are based on the ECMWF (European Centre Medium Weather forecast) grid datasets. Every 6 hours, the lateral and the top boundary conditions are updated in the coarsest grid, by using the ECMWF grid datasets. In the coarsest grid domain, a nudging toward the data is applied in the 3 grid points closest to the lateral boundaries and in the upper 5 grid levels. Horizontal domains and grid sizes have been designed taking into account both computational time limitations and the capability of the model to resolve essential mesoscale features. Meteorological fields were passed to the CALMET meteorological model which provides the boundary layer parameters necessary input for CALGRID. CALMET model has been run on the inner grid with the no-obs (no-observations) option. The domain size and grid spacing, used by the modelling system, are summarized in Table 2.

Table 2. Main characteristics of the three grids used in RAMS and of the one used in CALMET and CALGRID models; Lx, Ly and Lz are domain sizes in the x, y and z directions, respectively; Nx, Ny and Nz are the number of mesh points in the x, y and z directions, respectively; x and y are the mesh spacing in the x and y directions, respectively.

Grid	Lx (km)	Ly (km)	Lz (km)	Nx	Ny	Nz	x, y (km)
RAMS-1	960	960	15	40	40	25	24
RAMS-2	492	420	15	82	70	25	6
RAMS-3	162	180	15	108	120	25	1.5
CALMET/CALGRID	162	180	5	108	120	10	1.5

#### 5. RESULTS AND DISCUSSION

In order to demonstrate the level of representativeness of the system used in this study, the model predictions were compared with measured air quality data. Figure 4 (a) shows the predicted distribution of the hourly average concentration of NO<sub>2</sub> for 2005 year, while Figures 4 (b) and (c) show the scatter plots measured versus predicted concentrations for the yearly mean value and for yearly maxima values, respectively. The highest predicted concentration results from the effects of both industrial and traffic emissions. The scatter plots show that there is a light tendency of the model to underestimate the mean concentration values, especially for the traffic stations, related to the spatial resolution of gridded domain, and for the background stations, while there is a good correlation between predicted and measured maxima concentrations. Most of the data are in factor of two of the measurements, as recommended in the state of art model (blue broken line). The black broken line in the Figures 3 (b) and (c) denotes the range of 30% established by the Italian legislation (D.M. April 2, 2002): this range represents the target that the

data must observed as regard the yearly mean value. This target has not been respected by all monitoring sites and this is probably due to the incorrect temporal and spatial disaggregation of traffic flows over the simulation grid.

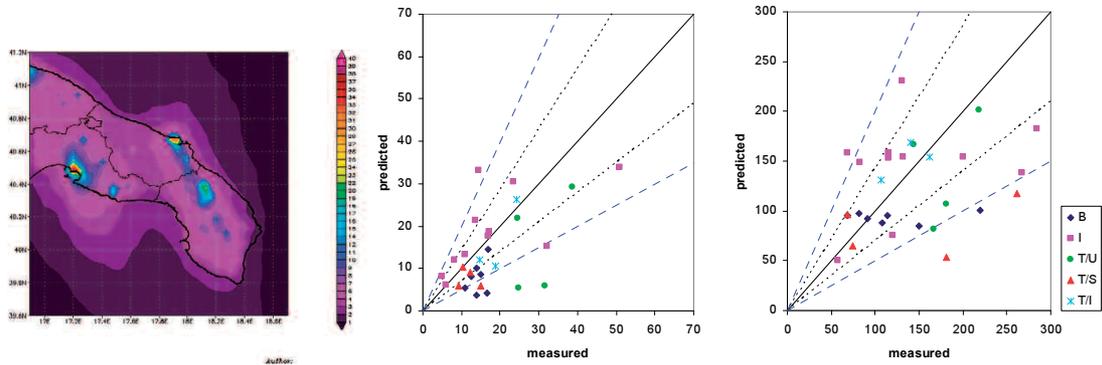


Figure 4. Yearly average concentration ( $\mu\text{g m}^{-3}$ ) of  $\text{NO}_2$  (a); scatter plots of predicted vs measured yearly mean (b) and maxima (c) values.

Figure 5(a) shows the predicted distribution of the hourly average concentration of  $\text{O}_3$  for summer 2005 (May to September), while Figure 5 (b) and (c) show the scatter plots measured versus predicted concentrations for the mean value and for the maxima values, respectively. The distribution of concentration levels is influenced by the presence of the two coasts and by the flow of the air masses from the sea. The highest predicted concentrations are observed along the coasts, resulting both from the presence of VOC and the minor concentration of  $\text{NO}_x$ . In general there is a good correlation between predicted and observed values and the concentration data are in factor of two of the measurements.

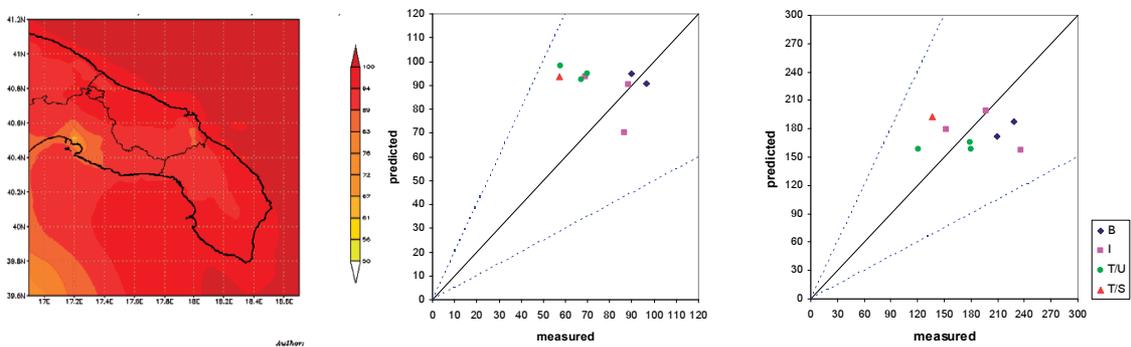


Figure 5. Yearly average concentration ( $\mu\text{g m}^{-3}$ ) of  $\text{O}_3$  (a); scatter plots of predicted vs measured summer mean (b) and maxima (c) values.

## 6. CONCLUSIONS

The air quality modelling system RAMS/CALMET/CALGRID has been used to simulate the distribution of pollutant species over the southern Apulia Region, in Italy, during the 2005 year. The accurate description of the emissions within the simulated domain has been obtain by using GEM-PP emission pre-processor.

The comparison between fields measurements and model results shows that the modelling system proposed is able to reproduce the mean values of the pollutants species over the entire 2005 over the simulated domain in a realistic way. As regard the  $\text{NO}_2$  data, there are some monitoring sites that do not observed the target value of 30% established by the Italian legislation (D.M. April 2, 2002), but this is probably due to the incorrect temporal and spatial disaggregation of traffic flows over the simulation grid. As regard the  $\text{O}_3$  data, he highest predicted concentrations are observed along the coasts and there is a good correlation between predicted and measured values.

*Acknowledgements:* This work was supported by the Province of Lecce and Apulia Region. The authors acknowledge the working group of the PREV'AIR system and particularly Cécile Honore of INERIS, for supplying concentration data used for the boundary conditions of the simulated domain, Dr. Lorenzo Angiuli and Dr. Alessandra Nocioni – Environmental Protection Regional Agency ARPA Puglia and Dr. Salvatore Francioso - Province of Lecce for supplying Environmental data.

## REFERENCES

- Carnevale, C., Gabusi V., Volta M., 2006: POEM-PM: an emission model for secondary pollution control scenarios. *Environmental Modelling and Software*, **21**, 3, 320-329.
- CORINAIR data: [www.sinanet.apat.gov.it](http://www.sinanet.apat.gov.it)
- EPA, 2002: SPECIATE 3.2. Available from: <http://www.epa.gov/ttn/chieff/software/speciate>.
- ESRI Shapefile format. Available from: <http://en.wikipedia.org/wiki/Shapefile>
- Geospatial Data Abstraction Library (GDAL) 1.4. Available from: <http://www.gdal.org>.
- Janssen, M. and Z. Wang, 2003: Open Emissions Model (OPEM): Re-Thinking the Emissions Modeling Paradigm. Report prepared by Lake Michigan Air Directors Consortium and University of California at Riverside. Available from: <http://www.ladco.org/tech/emis/misc/OPEM.pdf>.
- Mangia, C., Martano, P., Miglietta, M.M., Morabito, A. and Tanzarella, A., 2004: Modelling local circulations over the Salento Peninsula. *Meteorological Applications*, **11**, 231-244.
- Marra, G.P., I. Schipa, G. Aloisio, M. Cafaro, D. Conte, C. Elefante, C. Mangia, M. Miglietta, U. Rizza, A. Tanzarella, 2005: G-AQFS (grid Air Quality Forecast System): an experimental system based on GRID computing technologies to forecast atmospheric dispersion of pollutants. *Nuovo Cimento*, **028**, Issue 02, 192.
- Menut, L., 2007: <http://euler.lmd.polytechnique.fr/chimere/S>.
- Mitchell, T., 2005: Web Mapping Illustrated, O'Reilly & Associates Publisher.
- Monforti, F. and A. Pederzoli, 2005: THOSCANE: a tool to detail CORINAIR emission inventories. *Environmental Modelling & Software*, **20**, 505-508.
- Open Geospatial Consortium specifications (OGC), 1999: Available from: <http://www.opengeospatial.org>
- Pielke, R.A., Cotton, W.R., Walko, R.L., Tremback, C.J., Lyons, W.A., Grasso, L.D., Nicholls, M.E., Moran, M.D., Wesley, D.A., Lee, T.J. and Copeland, J.H., 1992: A comprehensive meteorological modeling system—RAMS. *Meteorology and Atmospheric Physics*, **49**, 69-91.
- PostGIS 1.3. Available from: <http://www.postgis.org>.
- PostgreSQL 8.2. Available from: <http://www.postgresql.org>.
- Proj.4 - Cartographic Projections Library. Available from: <http://trac.osgeo.org/proj>.
- Ruby programming language 1.8. Available from: <http://www.ruby-lang.org>.
- Russel, A. and R. Dennis, 2000: NARSTO critical review of photochemical models and modelling. *Atmospheric Environment*, **34**, 2283-2324.
- Schmidt, H., Derognat, C., Vautard, R. and Beekmann, M., 2001: A comparison of simulated and observed ozone mixing ratios for the summer of 1998 in Western Europe. *Atmospheric Environment*, **36**, 6277-6297.
- Scire, J.S., Insley, E.M. and Yamartino, R., 2005: Model Formulation and user's guide for the CALMET meteorological Model (Version 5.0). California Air Resource Board.
- Simplified Wrapper and Interface Generator (SWIG)1.3. Available from: <http://www.swig.org>.
- UMN MapServer 5.0. Available from: <http://mapserver.gis.umn.edu>.
- Yamartino, R.J., J. Scire, G.R. Carmichael, and Y.S. Chang. 1992: The CALGRID mesoscale photochemical grid model. *Atmospheric Environment*, **26A**:1493-1512.
- Yamartino, R.J.; J. Scire.; S.R. Hana; G.R. Carmichael, and Y.S. Chang., 1992: CALGRID: A Mesoscale Photochemical Grid Model. Volume I: Model Formulation Document Sigma Research Report No. A6-215-74. PTSD, California Air Resources Board. Sacramento, CA 94814.