15-413 Software Engineering Fall 1994
Carnegie Mellon University
School of Computer Science

Problem Statement

Overview

Many areas in the United States, and the world, are faced with serious air quality problems. In the United States, local, state, and federal agencies are engaged in efforts to identify where air pollution exists, the cause of air pollution, and how to control air pollution. The goal of the efforts is to bring an area that has been designated *out of compliance* with a air quality regulation into *compliance* with the regulation.

Air quality regulations are established by direct legislation, such as the 1990 Clean Air Act Amendments, and by rule development, such as the South Coast Air Quality Management District's (Los Angeles area) comprehensive rules on air emissions. The rules and regulations typically set limits on air emissions to the atmosphere for various sources of emissions (power plants, dry cleaners, automobiles, etc.).

Air quality models, such as the CIT airshed model, are used to identify the extent to which air quality is exceeded in an area. Furthermore, the air quality models are used to help identify measures that should be taken to abate air quality problems. Air quality models like CIT are advanced photochemical models. They simulate the physicochemical process that occur in the atmosphere. The air quality model simulates conditions in the atmosphere both in time and in space. An example of an urban-scale air quality study domain is the South Coast Air Basin centered on Los Angeles, California. An example of a regional-scale air quality study domain would be the Lake Michigan Airshed which encompasses most of Illinois, Indiana, Wisconsin, and Michigan. Air quality models like the CIT typically model a three to five day episode.

Emissions estimates over the study domain are a major input to the air quality models. Emissions estimates must be prepared for each hour of the days that the air quality model is going to simulate. Furthermore, air quality study domains are gridded; therefore, emissions must be estimated for each grid cell in the domain. The tool that is used to prepare emissions estimates for a particular study domain is an *emissions modeling system*. Examples of emissions modeling systems that are in current use include the following:

- FREDS -- the Flexible Regional Emissions Data System;
- EPS -- the Emissions Preprocessor System; and
- GEMAP -- the Geocoded Emissions Modeling and Projections System.

FREDS is a SAS® [1] based application that accepts only one type of emissions input and prepares only one type of output. FREDS accepts only emissions input files in the NAPAP format. FREDS prepares emissions estimates only for the Regional Oxidant Model another air quality model remotely similar to CIT. FREDS is comprised of numerous SAS® programs that are sequentially executed.
EPS is a FORTRAN based emissions modeling system with a SAS® visualization and query compo-

[1] SAS® is a statistical query and reporting system licensed and distributed by SAS Institute, Cary, North Carolina.

nent. EPS prepares emissions estimates files only for the Urban Airshed Model -- another air quality that is similar to CIT. EPS also only works with the carbon bond IV chemical mechanism. EPS is difficult to use over large domains because of the need to modify various FORTRAN programs and recompile when the air quality study domain changes.

GEMAP is a SAS® and ARC/INFO®[2] based emissions modeling system that has greater flexibility than either FREDS or EPS. GEMAP can work with either the carbon bond IV or SAPRC chemical mechanism. GEMAP can prepare emissions estimates input files for UAM, SAQM, and RADM -- all are air quality models.

FREDS and EPS operate on a narrow range of applications. Furthermore, FREDS and EPS can have excessive run times (on the order of days) over large domains. Although GEMAP is far more flexible than either FREDS or EPS, it too suffers from excessive run times over large domains. The cause of the excessive run times is due to inefficient data management during processing (inefficiently coded algorithms) and the size of the data sets that are managed (on the order of 10 MB to 400 MB)

In addition to excessive run times, GEMAP, FREDS, and EPS suffer from being based on proprietary software (SAS® for all and ARC/INFO® for GEMAP). Because proprietary software is used in the systems, a user must first purchase a license (or licenses) before the emissions modeling system can be used. A typical cost for SAS® on a Sun SPARCstation 1 to run GEMAP runs about $3000 (four licenses: Base SAS®, SAS/GRAPH, SAS/AF, and SAS/FSP). A typical cost for ARC/INFO on the same Sun platform is about $17,000. EPS has one other drawback in that its primary development tool is FORTRAN, and when a new study domain is being considered, EPS may require some modification and recompile effort.

You are asked to build a system called JEWEL that accomplishes the following two goals:

- Build an emissions modeling system that eliminates the need for users to purchase licensed software (SAS®, ARC/INFO®, etc.); and
- Build an emissions modeling system that has substantially faster run times than GEMAP.

Scenario: Lake Michigan Ozone Study

We will use the Lake Michigan Ozone Study (LMOS) as a test case for the new JEWEL system. LMOS is an ongoing air quality study (started in 1991) over the Lake Michigan area that involves the EPA; the air quality control agencies in Illinois, Wisconsin, Indiana, and Michigan; and the Lake Michigan Air Directors Consortium (LADCo). Alpine Geophysics (a client in this project) is under contract to LADCo to prepare emissions estimates input files for use in the Urban Airshed Model (UAM).

We will use the same input files that were used to prepare the first round of UAM-ready emissions estimates files. We will attempt to duplicate the UAM-ready emissions estimates files using JEWEL. We will be modeling an area the covers most of the four states that were just mentioned.

Functional Queries to be Answered

- Does JEWEL duplicate the UAM-ready emissions estimates?

[2] ARC/INFO® is a geographical information management system (GIS) licensed from ESRI, Redlands, California.

- What are the emissions estimates summaries after the completion of each major task?
- Can we display the entire modeling domain and subsets of the modeling domain with appropriate political boundary tags?
- Can we modify data via political tags?
- When emissions are lost (and they will be lost), why is JEWEL dropping them?
- Can the JEWEL system produce concentration maps? Emissions density plots? Temporal plots? (by domain and political tag?)
- Can the JEWEL system handle an increase of 19% power plant NOX, 13% mobile source NOX, 5% mobile source TOG, and an 8% reduction in biogenic ISOP?
- What are the new mobile source emissions estimates if VMT is increased by 15% in Illinois?
- Can the JEWEL system arbitrarily scale the emissions estimates up or down (halve the biogenics and double the mobile source emissions)?

Functional Requirements

Figure 1 provides additional detail on the functionality of an emissions modeling system. In general, an emissions modeling system must perform the following tasks -- that is, each of the boxes in Figure 1 must perform one or more of the following tasks:

- Prepare a modeling grid of the study domain;
- Estimate emissions either on an annual average (mass/year), average day (mass/day), or day-specific basis (mass/day);
- If necessary, apply regulatory factors (control efficiency, rule penetration, rule effectiveness) to the emissions estimates;
- Spatially allocate emissions estimates to the modeling domain;
- Temporally allocate emissions estimates in the modeling domain;
- Speciate emissions estimates; and
- Reformat the speciated emissions estimates for use by an air quality model.

In addition to the previously stated goals of this project, it would be desirable to develop a skeletal central repository in which the new emissions modeling system could utilize for its data needs.

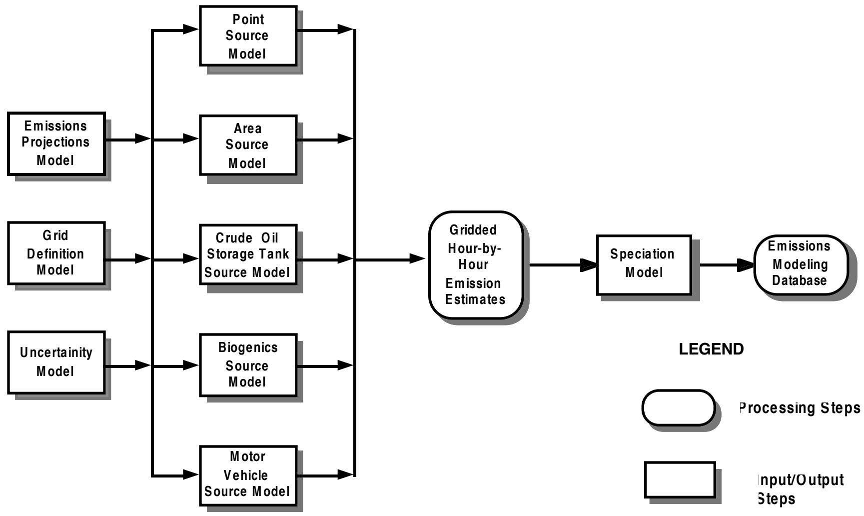


Figure 1. Simplified Overview of GEMAP

Global Requirements

The JEWEL system must be developed such that it eliminates the need for the user to purchase a software license. The system also must have substantially faster run times (on the order of hours) than current systems which operate over a period of days in some cases. Once again, the goal of an air quality study is to determine what the emissions limits must be in order to obtain a specified level of air quality. As stated earlier, current emissions models have excessive run times; therefore, estimating emissions is the *bottleneck* when it comes time to determine optimal emissions levels to obtain air quality goals. The emissions model *bottleneck* occurs because dozens of air quality model runs must be made, each with variations in the emissions estimates, to determine what constitutes *optimal emissions limits*.

The system should be able to operate on arbitrarily formatted input data sets. JEWEL must be able to not only operate on the inputs which are used to generate the emissions estimates, but the system must also be able to maintain a change history on any file that is modified. Input and corresponding output files must be tracked together.

The JEWEL system should provide a migration path from GEMAP and EPS. Many agencies are currently tied to the use of one or the other system. When the JEWEL system is functional and it provides better functionality than either GEMAP or EPS, the agencies must be able to migrate data from GEMAP or EPS into the JEWEL system.

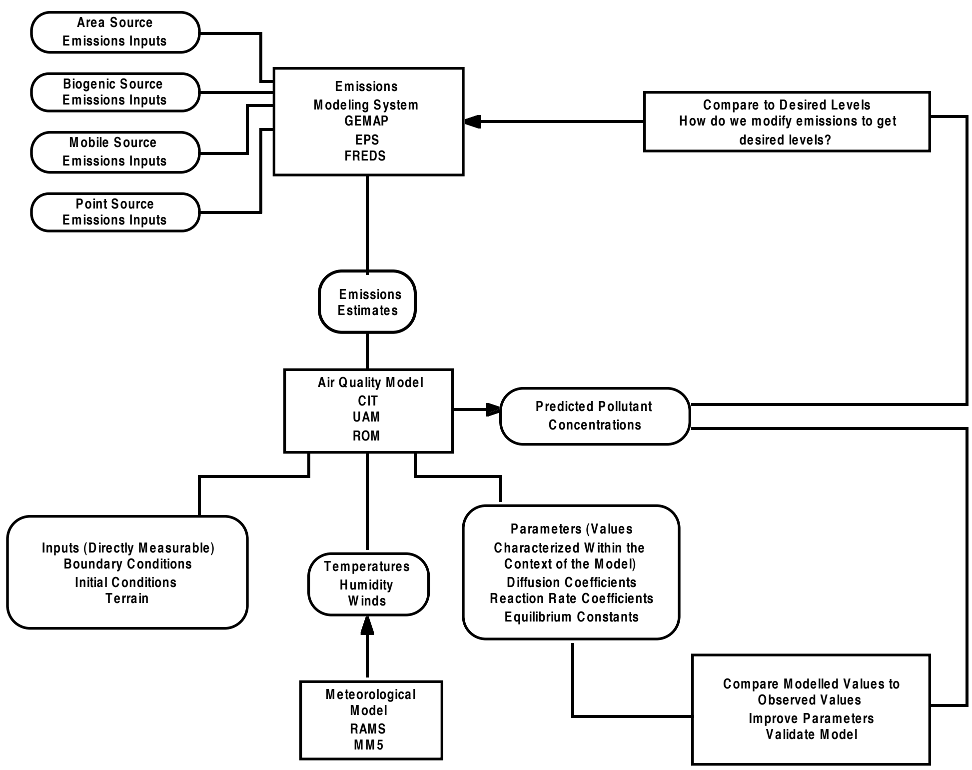
The information that is utilized in the JEWEL system must be able to be tagged by a limited hierarchy of political boundary information. In turn, the system must supply a means to visualize the data as tagged to political boundary objects.

Although the JEWEL system will be designed as a stand-alone system, the *high level design* must also be considered. It is envisioned that when the comprehensive environmental information management is built, JEWEL should be a virtual *plug-and-play* component. Furthermore, JEWEL should operate as a *plug-and-play* component of the GEMS framework.

JEWEL should have a user interface based on the X-Window standard -- X compatible windowing environments are delivered with virtually every variant of UNIX, and by requiring that the system interface be built with X compatibility in mind, we alleviate the need to purchase a license for proprietary software. The JEWEL system should be able to take advantage of a distributed computing environment if it is available -- as you will discover during the course of this project, many tasks that must be carried out in an EMS can be done simultaneously. The system must keep track of execution tasks so that the user does not try to rerun tasks. The user should be able to name and recall groups of related inputs and outputs that are maintained in JEWEL.

High Level Design

Emissions modeling is one component of an air quality study. Two other components of an air quality study are the meteorological modeling and the air quality modeling. The emissions model accepts input from the meteorological model, and it prepares inputs for the air quality model. Figure 2 provides an overview of the air quality modeling process.



In the grander scheme of things, an emissions modeling system is a component of a larger environmental information management system (Figure 3). The components of the integrated system read and write data to a central repository. The emissions modeling system (*Source-Specific Emissions Estimates Module* in Figure 3) is one of the many components of the envisioned system.

Client Contact

Dr. Ted Russell, a professor in Mechanical Engineering, and Jim Wilkinson, a doctoral student in Engineering and Public Policy and principal in Alpine Geophysics, will be the clients. We are also currently trying to secure commitments from an EPA staff member as an off-site client. Other members of Alpine Geophysics will likely be off-campus clients as well.

Client Acceptance

This is a broad problem statement, and as such, all the functionality that has been alluded to is not expected to be delivered in one semester. **Furthermore, the client is very willing to discuss any and all recommendations on changing the requirements specified in this document.** During the requirements analysis phase of the project, the functional and global requirements of an acceptable prototype will be established. An initial prototype of the JEWEL system is expected four to six weeks after the functional requirements have been established. The delivery of a second prototype of the JEWEL system is expected at the end of the semester.The client realizes that all components of JEWEL may not be functional by the end of the semester; however, the functional units of the prototype that are delivered at the end of the semester must reproduce the LMOS UAM-ready inputs.

Target Environment

The initial use of the JEWEL system will likely be for an ongoing air quality study in the Northeastern United State to determine the *optimal emissions limits* to set throughout the Northeast. The Northeast study will be run out of CMU and will be headed by Dr. Armistead (Ted) Russell of the Mechanical Engineering Department. The study will be undertaken on a network of DECstation 5000 computers running Ultrix.

Alpine Geophysics, an environmental technical service consulting firm, will also be retained for client design specifications and prototype acceptance. When successful, Alpine Geophysics will likely convert its current emissions modeling projects to the use of the JEWEL system. Alpine Geophysics' compute environment is totally Sun workstation based. Although Alpine Geophysics will likely be the first commercial entity to use the JEWEL system, it will by no means be limited to their use. The US Environmental Protection Agency (EPA) has a vested interest in seeing a project like this succeed, and they will be kept well aware of the progress of the project.

In addition to the US EPA, many local and state agencies will be interested in this product. The California Air Resources Board (CARB) will also be kept informed of the progress on this product. The CARB's compute environment is a dedicated IBM RS-6000 installation.

Most entities that perform emissions modeling tasks work in a mixed environment. However, the general trend is toward UNIX systems because most other air quality work (air quality modeling and meteorological modeling) is done in a UNIX environment. It has been the clients' experience that a typical UNIX workstation is configured as follows: 24 MB RAM, 4 GB hard disk (noncontiguous, maximum file size of 1.8 GB), 25 MHz CPU operating at 10 MIPS and 0.5 MFLOPS and 17" grayscale monitor (no graphics board).

Development Environment

It is envisioned that the JEWEL system will be developed on HP platforms. However, the JEWEL system must be readily portable to other UNIX platforms including but not limited to Sun SPARCstations, IBM RS, and DEC Unix-based computers. The implementation language should be C++.

GEMAP Demo

A demonstration version of GEMAP will be made available on Andrew and described in a separate handout 4. One copy of the GEMAP training and user/system documentation is available in WeH

3204 (Barbara Sandling's office).