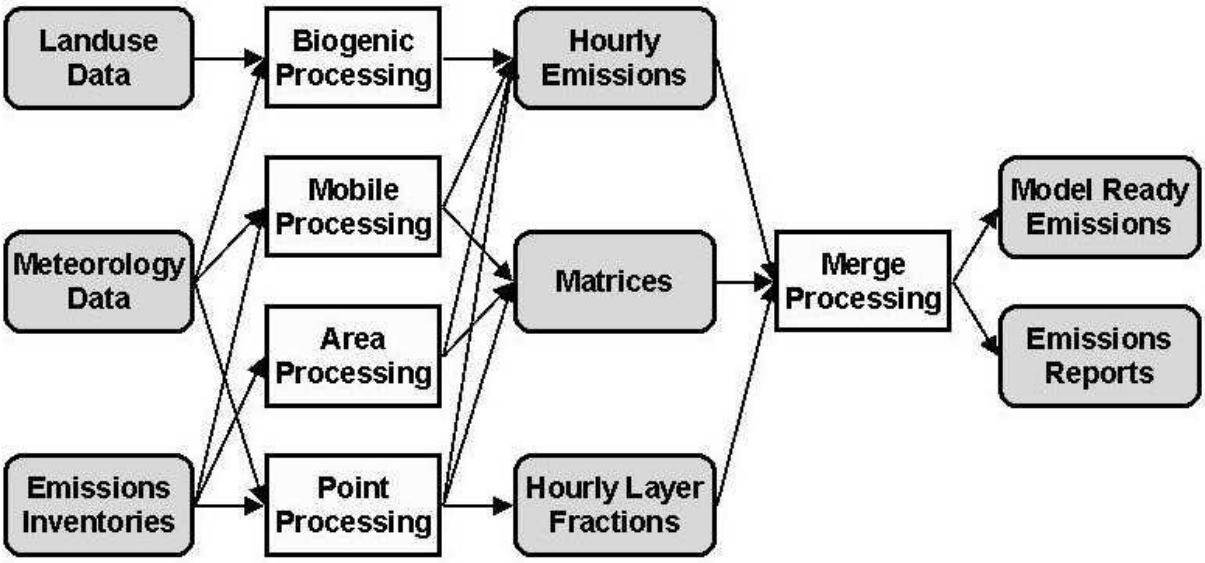
Chapter 1: Introduction

The environmental community has developed advanced numerical air quality models (AQMs) to understand the interactions among meteorology, emissions (both man-made and biogenic), and pollutant chemistry and dynamics. Emissions data from emissions models and regulatory inventories are one of the most important inputs for these air quality models. Scientist use air quality modeling for a number of purposes: for state and federal implementation plan development, for research on improved modeling methods, and most recently for air quality forecasting. In all of these cases, the trend has been to model larger regions, at a finer grid resolution, with more emissions sources, and for more purposes (e.g., ozone, particulates, toxics). These needs require a computationally efficient, user-friendly, and flexible emissions data processing system.

MCNC has created the Sparse Matrix Operator Kernel Emissions (SMOKE) Modeling System to allow emissions data processing methods to integrate high-performance-computing (HPC) sparse-matrix algorithms. The SMOKE system has greatly enhanced support for decision making about emissions controls for both urban and regional applications, has provided a mechanism for preparing specialized inputs for air quality modeling research, and has made air quality forecasting possible. SMOKE supports area, mobile, and point source emission processing and also includes biogenic emissions modeling through a rewrite of the Biogenic Emission Inventory System, version 2 (BEIS2) [http://www.epa.gov/ttn/chief/software.html#pcbeis].

The current version of SMOKE is primarily an emission processing system and not an emission inventory preparation system. This means that, with the exception of mobile sources, its purpose is to provide an efficient tool for converting emissions inventory data into the formatted emission files required by an AQM. For mobile sources, SMOKE additionally computes an emissions inventory from mobile-source activity data, using emission factors.

The sparse matrix approach utilized throughout SMOKE permits both rapid and flexible processing of emissions data. The processing is rapid because SMOKE utilizes a series of matrix calculations instead of less efficient algorithms used in previous systems. The processing is flexible because the processing steps of temporal projection, controls, chemical speciation, temporal allocation, and spatial allocation have been separated into independent operations wherever possible. The results from these steps are merged together at a final stage of processing using vector-matrix mathematics. The flow diagram below shows the relationship between data and processors in SMOKE. One example of how this approach is beneficial is development of a control strategy. In SMOKE, changing a control strategy requires only the control and merge steps to be processed again, whereas in other systems, most or all of the steps need to be redone.



Relationship between data and processors in SMOKE.

The SMOKE prototype has been available since 1996, and it has been an effective tool for emissions processing in a number of regional air quality modeling applications. In 1998 and 1999, SMOKE was redesigned and improved with the support of the U.S. Environmental Protection Agency (EPA), for use with EPA's Models-3 Air Quality Modeling System [http://www.epa.gov/asmdnerl/models3]. The primary purposes of the SMOKE redesign were support of (1) emissions processing with user-selected chemical mechanisms and (2) emissions processing for reactivity assessments (which will be described in more detail in Chapter 2).

The SMOKE modeling system can be used in a variety of ways. Currently, the most commonly used method is through the use of UNIX scripts. Although significant steps have been made to integrate SMOKE with Models-3, this process is ongoing. In addition, while the Environmental Decision Support System (EDSS) [http://envpro.ncsc.org/EDSS] is capable of providing an interface for using SMOKE without any changes to EDSS or SMOKE, the necessary steps for getting the two systems working together have not been made for all parts of the latest version of SMOKE. In addition, training is only available for the latest version of SMOKE using scripts.

Below, we summarize the major features of SMOKE.

- Support of Inventory Data Analyzer (IDA), Emissions Modeling System, '95 (EMS-95), and Emissions Preprocessor System 2.0 (EPS2.0) inventory input formats. The National Emissions Trends (NET) data will be available in IDA and EMS-95 formats.
- Supports both gridded and county total landuse for biogenic emissions modeling.
- Multi-country capability up to 10 countries.
- Fastest emissions processing tool available.

- Improved disk space requirements over other emissions processing software.
- *Any* pollutant can be processed by the system.
- *Any* chemical mechanism can be used to partition pollutants to model species, as long as the appropriate input data are supplied.
- Improved control strategy input formats and design over other systems.
- Control strategies can include changes in the reactivity of emitted pollutants. This is useful, for example, when a solvent is changed in an industrial process.
- Run-time memory allocation, eliminating any need to recompile the programs for different inventories, grids, or chemical mechanisms.
- No third party software is required to run SMOKE, although some input file preparation may require other software.
- Fewer SMOKE programs than the SMOKE prototype because programs were combined where possible to be used for multiple source categories.
- Support of plume in grid processing.
- Integrated with Models-3 file formats and settings.
- Improved data file formats.
- Support of CMAQ, MAQSIP, UAM-IV, UAM-V, and CAMx emissions input formats.

Some features that will be available in the near future are:

- Enhanced quality assurance pre- and post-processing.
- Fully integrated with Models-3, which will provide the SMOKE Tool for SMOKE input file preparation.
- Further testing of all routines with the latest inventory and data.
- Packaged with the National Emissions Trends (NET) inventory and other latest available data.
- Support of MAQSIP-PM, UAM-AERO, and REMSAD models.

About this document

This document is the most complete reference available on SMOKE. Chapters 2 and 3 provide background on emissions data processing, SMOKE-specific concepts, and other concepts specific to the four main source categories. Chapter 4, when available, will compare SMOKE methods, performance, and results to other emissions processing systems. Chapter 5 provides background on how SMOKE interacts with the environments in which SMOKE can be used: UNIX scripts, Models-3, and EDSS. Chapter 6 documents the file format converters and other SMOKE routines that can help prepare SMOKE inputs. Chapter documents the core emissions processing programs of SMOKE. Chapter 8 gives a small amount of documentation on the SMOKE quality assurance programs and utilities that are available at this time. Chapter 9 provides file format information for all of the SMOKE input files, but the SMOKE intermediate and output file formats are not yet documented. Chapter 10 provides documentation on the SMOKE scripts, and will eventual provide user instructions on how to run SMOKE from the scripts, from Models-3, or from EDSS. Chapters 11, 12, and 13 are not available at this time, but they will respectively provide documentation of the source code and include files, documentation on compiling SMOKE and SMOKE performance, and documentation on SMOKE errors and warnings. Chapter 14 contains the training materials: both copies of the slide presentations and the lab exercises. Finally, the Appendix provides a list of known SMOKE bugs and errors, and will provide a glossary of terms and an acronyms index. As the incomplete chapters become available, they will be released on the SMOKE web site.

This manual does not provide technical documentation on the code itself, but the code has extensive in-line documentation to assist Fortran-savvy users in understanding its workings.