

Final Report

METROPOLITAN EMISSIONS SIMULATION SYSTEM

EF78-23

February 1978

Contract No. A5-067-87

Prepared for

Air Resources Board
1709 - 11th Street
Sacramento, California

Prepared by

D. Gottlieb
H. Rabinowitz

Appendix by

T. N. Jerskey

Systems Applications, Incorporated
950 Northgate Drive
San Rafael, California 94903

LIBRARY
AIR RESOURCES BOARD
P. O. BOX 2815
SACRAMENTO, CA 95812

"The statements and conclusions in this report are those of the Contractor and not necessarily those of the State Air Resources Board. The mention of commercial products, their source or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products."

CONTENTS

I	INTRODUCTION	1
II	SYSTEM OVERVIEW	5
	A. The System Environment	7
	B. Recommendation of PL/I as Programming Language	9
	C. Design Considerations	10
III	SYSTEM FLOW	12
	A. CEF Generation	14
	B. Production of Geographical File	16
	C. Production of Basic RGEFs	17
	D. Production of Modified RGEFs	19
	E. Output Selected Emissions Data	20
	F. Implementation	21
IV	DESIGN SPECIFICATIONS	22
	1. Generate CEF from Source Files	30
	2. Produce Regional Gridded Emissions Files (RGEFs)	50
	3. Output Selected Emissions Data from RGEF	92
	4. Reference CEF for Background Information	121
	5. Produce Geographical File	122
	LIST OF ABBREVIATIONS	130
	APPENDIX: FORECASTING TECHNIQUES FOR THE EMISSION SIMULATOR	A-1

I INTRODUCTION

Emission inventories (EIs) attempt to categorize and account for all anthropogenic activities leading to injection of pollutants into the atmosphere. The data contained in an EI consist of either emissions rates for each source and source category, expressed as mass per unit time, or basic information from which emissions rates can be computed using an appropriate algorithm. Moreover, regardless of the form of the emissions data, in California at least, these data are collected principally at the local levels using various methods. Such data are intended to serve a variety of purposes at the federal, state, and local levels. Irrespective of the purposes of the EI, obtaining and maintaining an EI is a legal responsibility of the CARB prescribed by Section 39607(b) of the Health and Safety Code.

Recognizing the growing and expanding needs for an efficient and flexible EI, the CARB sought assistance about two years ago in the design of a computer software system to assimilate source emissions data from available and projected data banks so as to allow easy access for a variety of users. The system was named the Emissions Simulator (EMSIM). Systems Applications, Incorporated (SAI) was awarded a contract to provide this design.

Soon after initiation of SAI's study it became apparent that EMSIM should be broadened in scope and capabilities to:

- > Include both inventory reporting and modeling uses.
- > Fill the gap between the raw emissions data collected and the needs of various applications.

In addition, because of the likelihood of a broad community of users, EMSIM should melt simultaneously the twin goals of flexibility in choice of extractable information and communicability (ease of obtaining data). As originally

envisioned, EMSIM would store and present emissions data in a gridded map format. This basic concept turned out to be valid in achieving the expanded goals and in fact offered an optimum way of storing, retrieving, and using the EI information.

During the design of EMSIM, a variety of routine emissions reporting, policy/regulatory, emissions inventory maintenance, modeling, and control strategy "what if" kinds of questions were asked to determine EMSIM's practical characteristics:

- > Is EMSIM sufficiently flexible, efficient, easy to "talk to" and maintainable so that it does not degrade the performance criteria of the EI? Some of the essential EI performance criteria considered were:
 - EI update within 10 working days.
 - Consistency of source data throughout the EMSIM files.
 - Extension of source categories.
 - Confidentiality of point source data.
 - QA/QC functions.
 - English and/or metric units.
- > Can EMSIM support the state EI requirements which are federally mandated, to provide presentations of yearly state implementation plan (SIP) updates?
- > Can EMSIM support the federally mandated state EI requirements to provide emissions projections for all air quality maintenance areas (AQMA's) (see the Appendix)?
- > Can EMSIM support the state EI requirement to evaluate detailed current subregional emissions data related to clean water project proposals submitted by local agencies?

- > If these data are not available from the local agency, can EMSIM support the state in obtaining estimates of these data?
- > Can EMSIM support the EI activity of providing emissions projections by source types using various control technologies and alternatives?
- > Can EMSIM support the EI activity of providing emissions reports for AQCRs, AQMAs, and ACPs (air conservation programs) within one working day of the request?
- > Can EMSIM support the EI activity of providing emissions data and inventory reports within one working day of the request:
 - For gridded data with a resolution of at least 1 square km?
 - By county?
 - By air basin?
 - By air quality control region?
 - By air quality maintenance area?
 - By air conservation program?
 - By any user-defined area?
- > Can EMSIM support the EI activity of generating annual state-wide EI reports and a once per month compliance report?
- > Can EMSIM support the EI activity of retrieving and reporting historical data?
- > Can EMSIM provide data to and interface with CARB point source, area source, and mobile source simulation modeling activities of either the Gaussian or numerical modeling type?
- > Can EMSIM support CARB development and evaluation of statewide and basin-wide control strategies?
- > Can EMSIM support CARB and county evaluation of new-source review activities, emissions offset requirements, and prevention of significant deterioration review activities?

The answers to all of these questions are affirmative; however, it must be recognized that the effectiveness of EMSIM depends greatly on the existence of an efficiently designed and operating EI activity. EMSIM cannot overcome any fundamental limitations inherent in the EI data base itself.

Without the diligent and able guidance and assistance of a variety of state personnel, neither the scope nor the design of EMSIM would have reached its current state. Of particular importance were the roles played by Richard Bradley, Gary Knops, and Don Perrine, who helped formulate and define the users' requirements throughout the project; Richard Beers of General Services, who helped coordinate the finalization of the scope of the system and provided valuable insight into the data processing services that CARB could realize; Jack Paskind, who provided direction and encouragement in the critical planning stages of the contract.

II SYSTEM OVERVIEW

The Emissions Simulator System (EMSIM) is a set of computer files and programs designed to give easy access to emissions data for a wide variety of uses, including inventory and modeling. In the past, each of the many applications needing emissions data was required to laboriously recreate the information needed from a number of disparate sources, possibly gridding the information by hand. Thus, although there are vast quantities of emissions data collected throughout the state, it was quite expensive and time-consuming to gain access to the data actually needed for a particular application. The Emissions Simulator fills the gap between the raw collected emissions data and the needs of various applications. Moreover, EMSIM is designed to give great flexibility in choice of information extracted while remaining easy to "talk to" by a casual user. The user has a choice of geographical area, time frame, grid size, emissions species, and source categories to be collected and displayed.

The Emissions Simulator uses two major kinds of files, a Crude Emissions File (CEF) and a number of Regional Gridded Emissions File (RGEFs).^{*} The CEF is a large file which concentrates the crude data from existing emissions inventory files: the Point Source, Area Source, and Traffic Source files. The RGEF's correspond to particular user needs: each one represents a grid of a chosen region containing all requested emissions that fall within that region. An RGEF can be queried to provide information on emissions in that region. The information might be displayed in the form of a table of total emissions of each species in each grid cell, or it might be in the form of a file suitable for use with graphics packages, among other possible outputs.

^{*} See system flow diagram.

In creating a basic RGEF from the CEF, the emissions simulator collects all relevant emissions data by grid cell so that information is available organized geographically. In creating output from an RGEF the user has a wide choice in specifying the kind of output information to be extracted.

The Emissions Simulator also has the ability to create modified RGEFs from the basic ones for the purpose of making predictions or evaluating control strategies. The modification may consist in multiplying emissions from user selected source categories by a set of factors corresponding to projected changes in population and production distribution. Modification may also consist of adding new pollution sources or removing existing sources. The modified file representing the hypothetical situation can then be queried and reported on just as if it were a basic RGEF.

The reason why a gridded format is chosen for the regional files is central to understanding the Emissions Simulator. Gridding a region such as a county into, say, 1 km square grid cells provides a unifying method for the collection of emissions data from different types of sources. First, point sources can be placed in their appropriate grid cell. Then, countywide area sources, such as home heating, can be spread evenly over all the cells in the county. If the county is one for which land use data has already been developed, the area source emissions can be more precisely apportioned to the relevant grid cells. For instance, if a particular grid cell consisted of 40 percent residential area, it would be apportioned 40 percent of the home heating emissions allocated to a square km of residential area. Finally, traffic emissions, which have themselves been gridded by a system external to EMSIM, will be integrated into appropriate grid cells in the basic RGEF. When gridding of emissions is complete, the basic RGEF contains the best estimate available of all the emissions occurring in a given grid cell. This information can then be extracted either on a total basis ("what are the totals for each pollutant species in this square km?") or on a detailed basis ("what are the totals for certain production processes in this square km?").

Geographical gridding is thus a simplifying way for the user to conceptualize the structure of the regional files so as to integrate large amounts of emissions data. However, in reality the RGEFs are VSAM files which are keyed on a number of attributes besides geography. An advantage of using the VSAM organization to represent the RGEFs is that one can construct indexes to view the data from an alternative perspective. For instance, one could construct an index to enable the easy extraction of all emissions produced by petroleum refining processes, or by power plants. This information could then be collected and displayed by grid cell or as totals for each process. The point is that geographical gridding is only one of the possible organizing frameworks made possible by the flexibility of the VSAM access method.

In creating a modified RGEF from a basic one the user has a choice of temporal framework: either "snapshot" or "hourly-day." Snapshot gridded files will contain emissions for a single specified time period, which may be a particular day of the year, a month, or an entire year. There can also be snapshot files containing an average day's emissions or a worst day's emissions. Hourly-day files will be used by the modeling group and will contain 24 hourly sets of emissions for a particular day of the year. Fortunately, since modelers do not need great detail of source categories, these categories can be aggregated into a small number of source categories. The aggregation of source categories will enable the hourly-day files to remain reasonable in size despite the large volume of temporal information.

A. THE SYSTEM ENVIRONMENT

EMSIM is necessarily a complex system dealing with large volumes of data in a sophisticated way. It is therefore designed to be implemented on a large computer such as the IBM 370/168 at Teale Data Center. The output of the Emissions Simulator may be used as input to smaller systems. The Simulator itself cannot be designed for implementation on a minicomputer.

unless a PL/I-type compiler and a sophisticated access method are available. Given the expected IBM hardware environment, the designers of EMSIM have envisioned the use of PL/I as programming language and VSAM as file access method. VSAM provides both direct and sequential access to files as well as the possibility of alternate indices. A full scale data base management system (like ADATABASE) is not needed since the emission simulator is itself a data management system with its own peculiar needs which would not be served by the overhead of a general purpose DBMS.

Maintenance procedures form an integral part of the design of the EMSIM system. Each file internal to EMSIM can be created, modified, displayed, and purged through the use of utilities, programs, and JCL data-datasets (CNTL) which will be part of the EMSIM system. This systematic inclusion of maintenance procedures should allow the EMSIM user to handle most routine maintenance without the need of consulting a systems programmer.

Storage requirements for EMSIM were analysed applying generous upper bounds to all files. Assuming the number of processes in EIS to be 50,000 (there are currently fewer than 25,000) the size of the Crude Emissions File (CEF) will not exceed 19 megabytes (Mgb) which comprise less than 80 cylinders on an IBM 3330 disc pack or 50 cylinders on an newer 3350 disc packs. This estimate is for the entire state of California and does not assume any sophisticated data compaction efforts.

A typical snapshot gridded file (modified RGEF) for the Southern California Air Basin (SCAB) would not exceed 2 Mgb and therefore will not require more than 10 cylinders (on a 3330). This is a very generous estimate based on the assumption that the RGEF will be a VSAM variable-length-record file with an average number of 6 (non-zero) pollutant species per process.

An RGEF of SCAB with hourly-day temporal distribution can be contained in 27 cylinders (3330) and probably much less if careful aggregation of categories is applied.

Our estimates suggest that a single 3330 IBM disc pack should be more than adequate to meet storage requirements for EMSIM files for the next few years.

All the above estimates represent "worst case" analysis of storage requirements. In the actual design of the files close attention to compact representation of data will help achieve maximum utilization of storage space.

B. RECOMMENDATION OF PL/I AS PROGRAMMING LANGUAGE

As discussed above EMSIM is a sophisticated state-of-the-art computerized system whose complex data requirements call for a powerful modern high-level programming language. PL/I is such a language, allowing the ease of programming inherent in a high level language while maintaining a degree of flexibility available otherwise only in assembler languages. Neither FORTRAN nor COBOL is fully adequate to the programming task due to limitations which are described below.

EMSIM contains a large volume of data of different types and requires string-processing, floating point arithmetic, and large amounts of I/O to VSAM files. The complex data structures needed to describe the VSAM grid-cell records are easily handled in PL/I, less easily handled in COBOL, and virtually impractical to handle in FORTRAN. The system requires easily coded interaction with VSAM files which is easily done in PL/I, less easily done in COBOL, and impossible in FORTRAN. The geographical processing (point-in-polygon algorithm) requires floating point arithmetic--easy in PL/I or FORTRAN, more difficult in COBOL. In order to process the user-interface language, the string processing capabilities of PL/I are necessary, while neither FORTRAN nor COBOL has such capabilities. This difference could become crucial should the users want to put an on-line front-end on the system.

In addition to satisfying the data needs of EMSIM, PL/I offers several other advantages over the other widely available high level languages. PL/I naturally lends itself to structured programming and modularization, which have been shown to contribute to reliability of software, ease of maintenance and later changes. Only PL/I has the ability to allocate storage at run time; this flexibility of dynamic allocation is important in programs which are themselves large and must handle large quantities of data. It enables, for example, arrays to be allocated only to the smallest needed extent, as determined at run-time.

Because PL/I is easier for the programmer to use than other languages and is provided with excellent diagnostic compilers, the cost of coding and testing are likely to be lower than with other languages.

C. DESIGN CONSIDERATIONS

The EMSIM system was designed using top-down analysis following the guidelines of current wisdom on reliable software development (see "Reliable Computer Software: What it is and How to Get It," David Farnan, Defense Systems Management School, Department of Defense, 1975). The top-down functional charts of EMSIM and their descriptions follow in a later chapter. Here we note that top-down design insures a unified view of the system and the functions it must perform. Of course, it is an oversimplification to assume that the entire design can be done in a top-down sequence. One must have assurance, before reaching the bottom, that the lower levels will be feasible to implement. Therefore, the optimum design would involve top-down with some bottom-up programming. In this case, for example, it was necessary to develop the point-in-polygon algorithm and gridding algorithms bottom-up in order to assure that the top-down design will converge to practical programable modules.

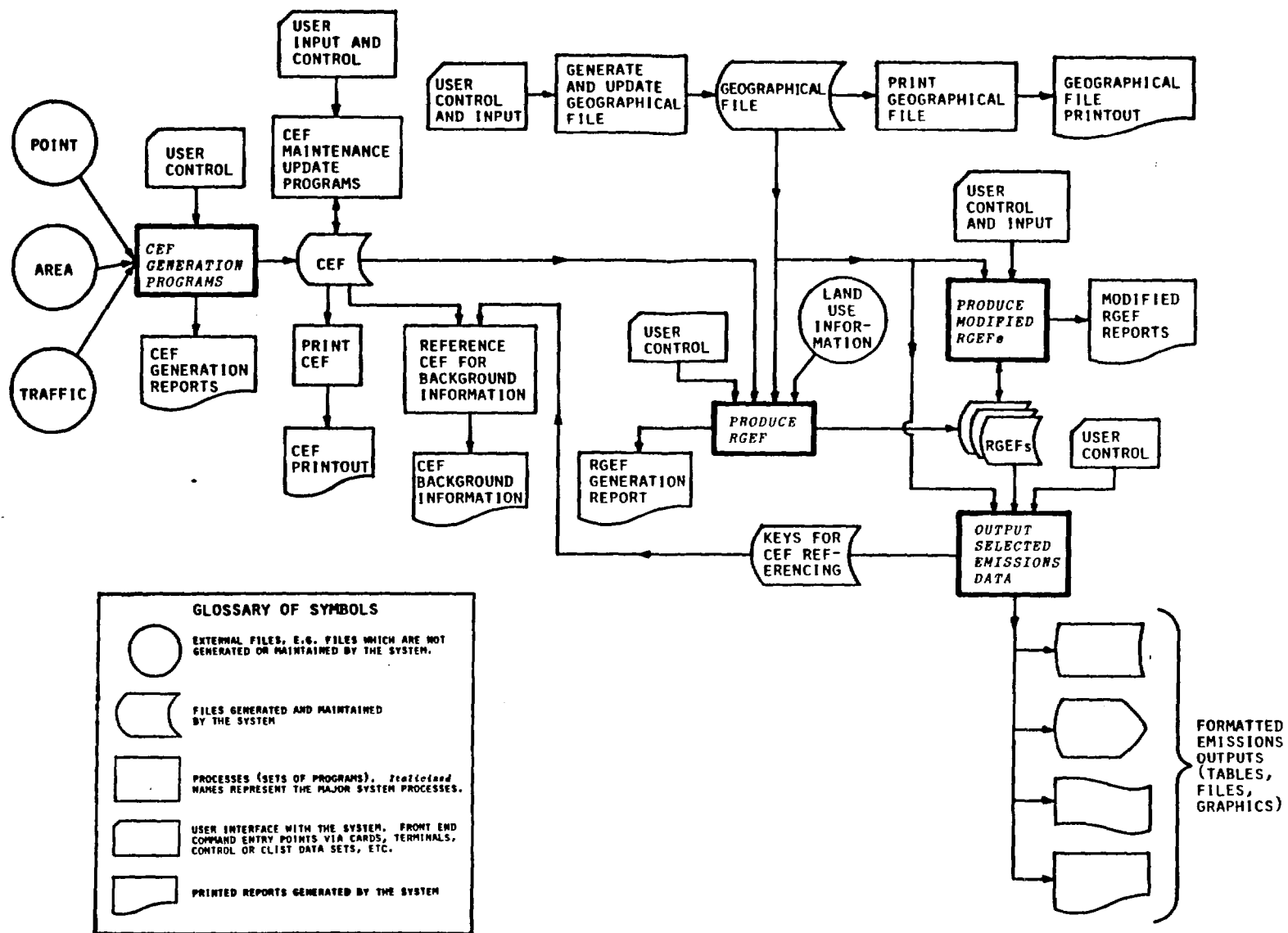
One benefit of the use of top-down analysis is the separation of input functions from processing functions. That is, in the case of EMSIM, some

of the raw source files are subject to possible changes in format. EMSIM is designed to minimize the impact of such changes in the input formats on the rest of the processing programs. Each module knows only about the data it directly needs; thus it can be independent of changes in the external format of the data and its collection method.

The user interface languages, especially the language for querying an RGEF, are designed to be largely free-format languages with tolerance for a variety of input styles. They are accompanied by effective error-analysis routines in case the user does make an error in input.

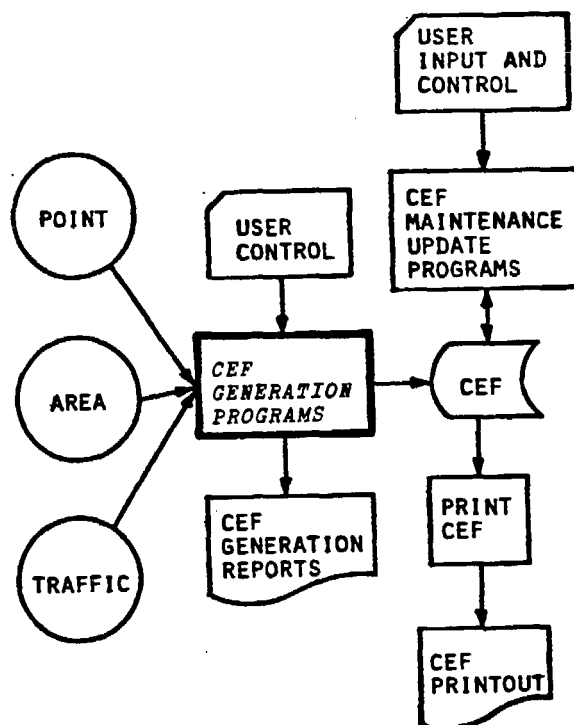
III SYSTEM FLOW

This chapter contains a description of the System Flow diagram. In order to facilitate a written description, the diagram is broken into several clusters centering around the major functions of the system, each of which is discussed in turn.



EMISSIONS SIMULATOR SYSTEM FLOW

A. CEF GENERATION



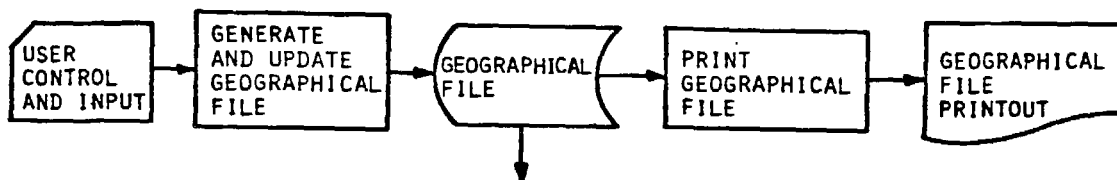
The CEF file is generated infrequently--once or twice a year--from the external source files. It is the master file for the EMSIM system and contains the data in a standardized format usable by the rest of the system. If one of the external source files undergoes a format change, which is quite likely, then only the front-end of the programs generating the CEF need change. The CEF's structural format should remain fixed, and hence the rest of the system should be insensitive to changes in source file format. The CEF is a VSAM file keyed on source type (point, area, traffic), county, air basin, industrial process code, Air Quality Control Region, and

containing geographical information, emissions data and temporal distributions as well as identifying information like plant identification, plant name and address. The identifying information is not passed on to the basic RGEFs but is retained as backup information in case it's needed. A separate standard record format will exist for point source, area source, and traffic source emissions. Generating the CEF consists of converting external source file records into the appropriate standard CEF record format, and of disaggregating point source emissions to the (industrial) process level. The latter function is necessary since current point source emissions in the EIS file are kept at the point source level and an algorithm is necessary to break the emissions down into estimated portions for each process within a plant.

The CEF is not a gridded file. It contains geographical information locating the source in each record, but it is not keyed on geographical coordinates. The EMSIM system is intended to have exactly one CEF, from which RGEF's are generated. When enough changes to the external source files have occurred to cause the user to want to create a new CEF, he has a choice of what to do with existing basic RGEFs. He can regenerate all the existing RGEF's from the new CEF or he can keep the old RGEF's knowing that they represent outdated data and that they cannot be backed up by the current CEF.

The CEF is intended to be a non-volatile file, that is, once created it will not be updated even if new external source emissions data are developed. The non-volatility of the CEF is intended to assure consistency with RGEF files. An allowance is made, however, for "maintenance update" of the CEF to correct any errors that may have occurred in its creation.

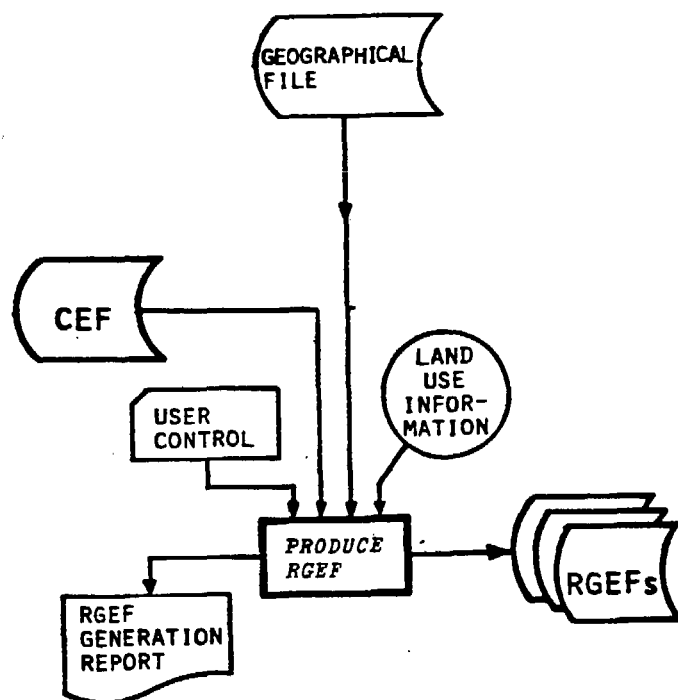
B. PRODUCTION OF GEOGRAPHICAL FILE



This portion of EMSIM is not part of the mainstream system flow but rather a necessary peripheral subsystem. The geographical file contains a set of records each representing an arbitrary geographical area described as a polygon in UTM coordinates. Each area is given a name so it can be referenced by other programs. An area may be a county, subcounty, or trans-county region. The polygon describing the area is assumed not to cross itself. The circumscribing rectangle of the polygon and the counties intersecting the polygon are also included in the record.

The geographical file will contain records describing all counties, air basins, and other frequently referred to areas in the state. It will be fully updatable so that new areas of any size can be added, or old records be purged.

C. PRODUCTION OF BASIC RGEFs



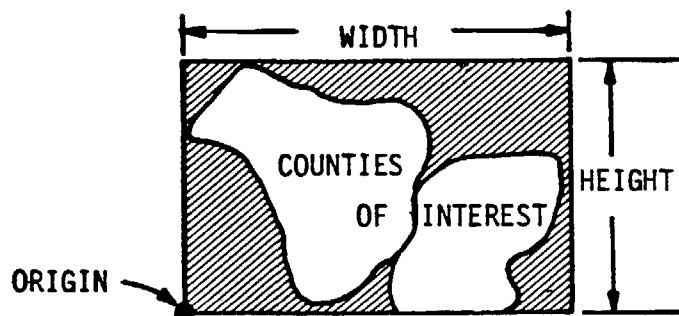
The RGEF is a VSAM file keyed on grid cell, county, air basin, and process code, and containing emissions data and a reference key to the CEF for background information. The RGEF has a header record which contains the grid specifications, species and category tables and aggregations, creation date, and other information.

For those counties with too little pollution data to make gridding worthwhile, it will be possible to make a temporary RGEF with just one "cell" containing totals for the entire county. The information can be printed using the RGEF generation report and then the file can be deleted.

Any number of basic RGEFs can be produced from the CEF. To produce an RGEF the user must enter the following user control parameters:

- > Grid origin--the UTM coordinates of the Southwest corner of the grid region.
- > Height and width of the grid region.
- > Size of the grid cells.

The conceptual structure of an RGEF is a rectangular grid containing at least one whole county. In fact, any county of interest to the user must be entirely included within the rectangular grid region. Otherwise it would be very difficult to allocate county-wide area emissions. Those grid-cells within the region belonging to counties not of interest will not actually be included in the file at all.

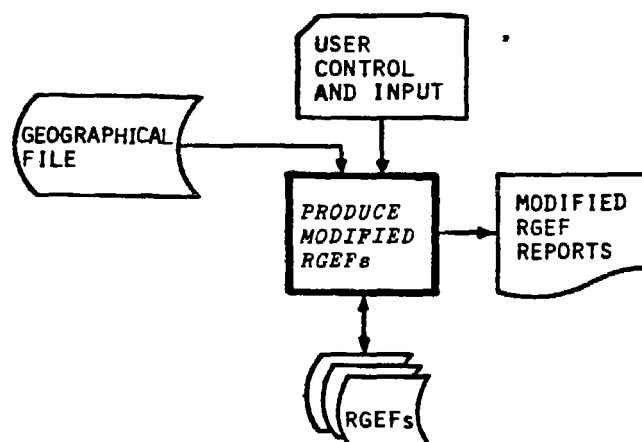


Although an RGEF is required to contain entire counties of interest, EMSIM is able to display information concerning subcounty or transcounty regions (see section on output of selected emissions data).

If a Land Use file is used to help apportion area emissions, it must be gridded on integer Km coordinates so as to match the grid region of the RGEF being produced.

As the RGEF is produced, an RGEF generation report is printed. This report can contain all information that has been placed in the RGEF, or just totals for the entire region, depending on the user's preference.

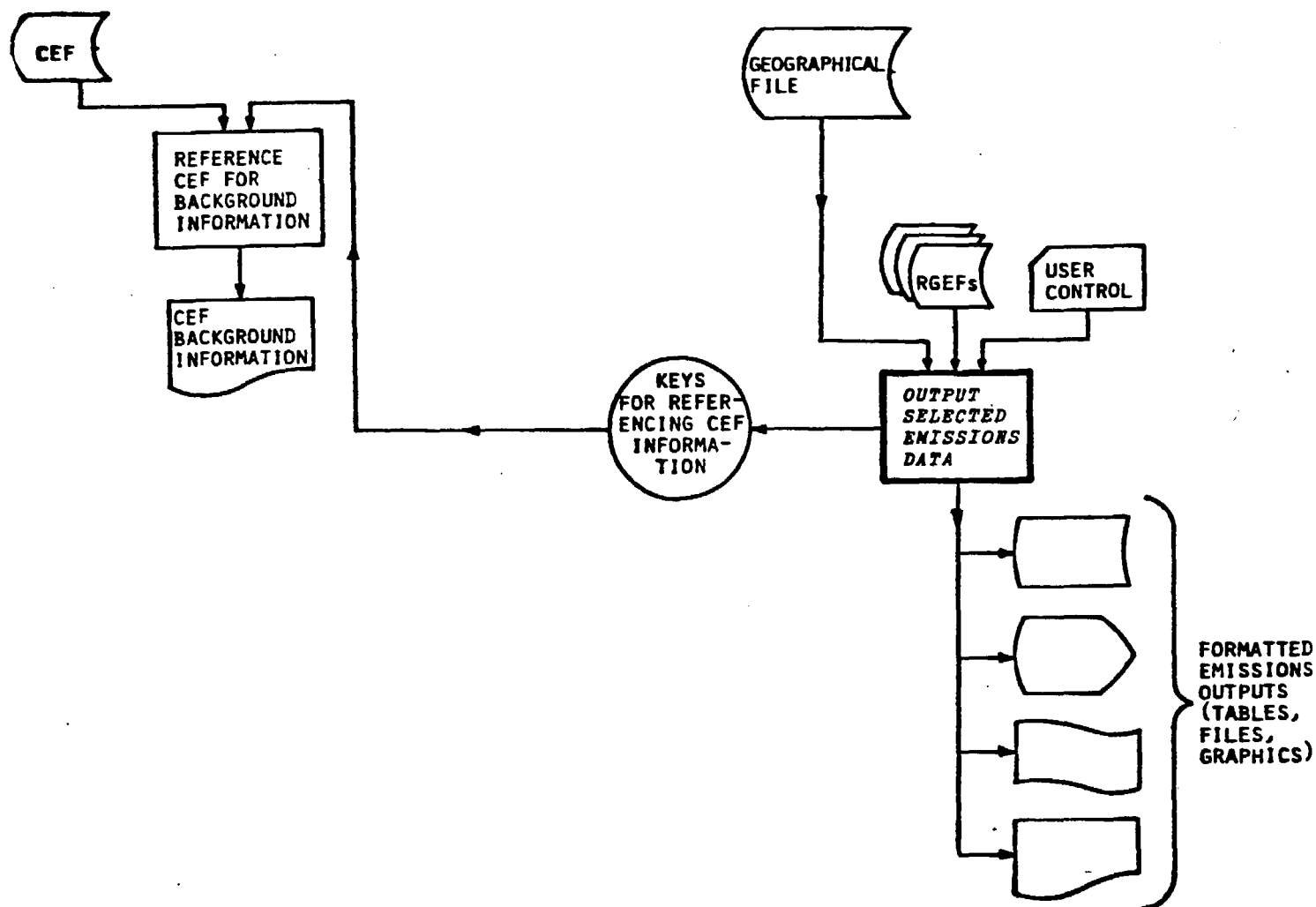
D. PRODUCTION OF MODIFIED RGEFs



For purposes of prediction and control strategy evaluation, EMSIM allows the creation of modified RGEFs to simulate projected scenarios. EMSIM does not produce future scenarios by itself. That is, EMSIM has no access to projected population or production figures for future years. That information is best assessed by human effort and is not worth the added complication of inclusion in the system. However, once these figures are assessed by hand, EMSIM makes it very easy to create the appropriate modified RGEF. This is done in a number of ways. Sources can be added to or deleted from the RGEF and a table of multipliers is set up allowing each pair of pollutant species/source-category emissions to be multiplied by a constant factor. For instance, production of hydrocarbon pollution by all service station operations in any user defined geographical area could be multiplied by .8 to allow for new pumping nozzles. The user could specify which pollutant species, production categories and time frame should be considered. At the same time the user could also instruct the system to aggregate pollutant species and process categories so as to produce new super-species and super-categories.

A modified RGEF, though representing a simulated reality, looks and behaves just like any RGEF. Moreover, the original RGEF is preserved unchanged when the modified RGEF is produced.

E. OUTPUT SELECTED EMISSIONS DATA



This section of EMSIM is the "payoff." All the rest of the system is designed so that the user can extract information about emissions and here is where he gets that information.

The user can specify any geographical region in the geographical file as a query area. He can specify an interest in any choice of pollutant species and process categories or combinations thereof. He can combine species into new superspecies and categories into new super-categories. He can even multiply certain emissions by output multiplier factors in a manner similar to that used in creation of a modified RGEF.

The information extracted can be presented on a process-by-process basis or as totals for all processes. It can be presented on a grid cell by grid cell basis or as totals for all grid cells in a particular area. It can be formatted in various ways suitable for printed output, graphics, or input to other systems.

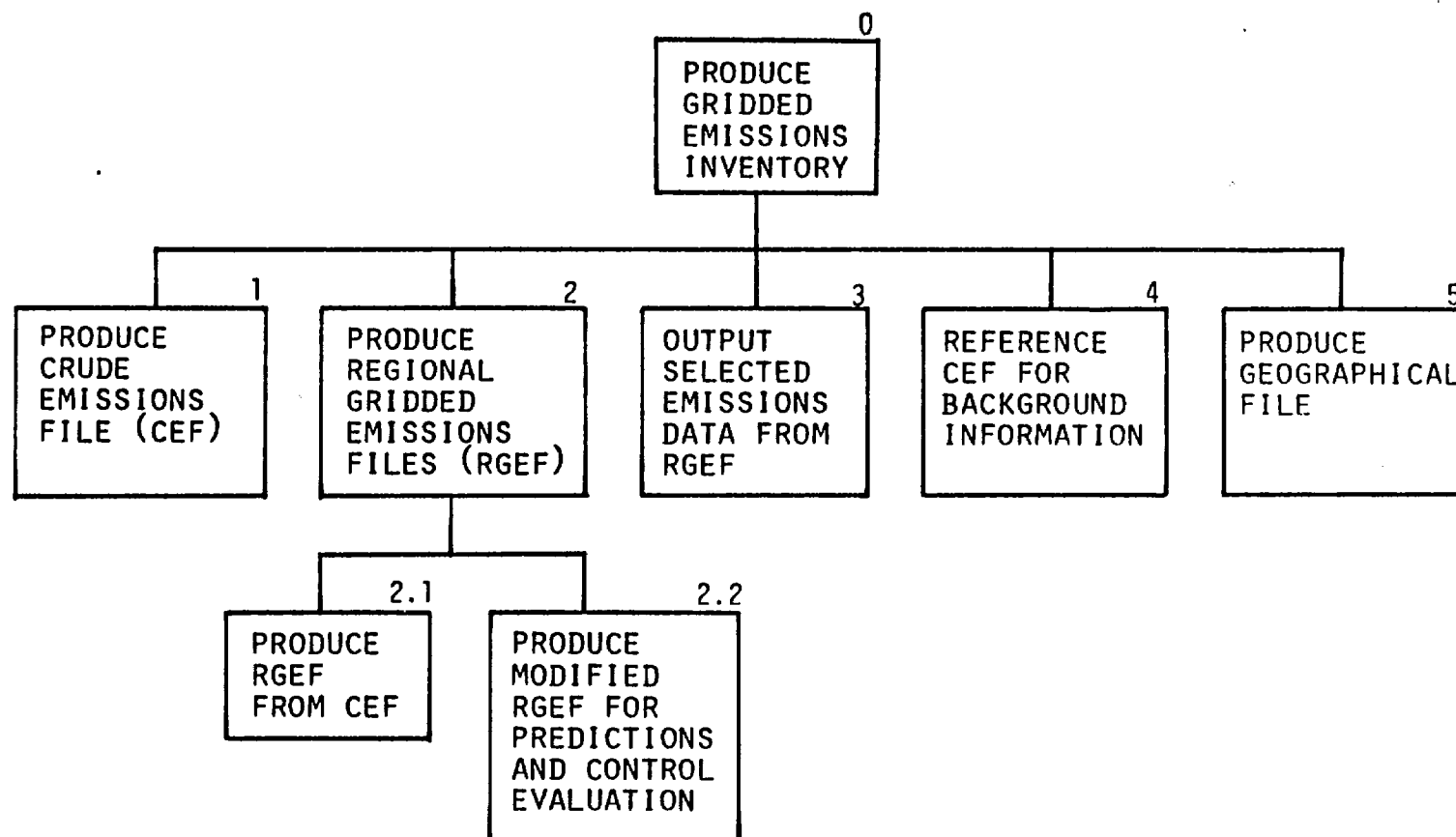
Sometimes a user will want to know more about the source of some emissions. Perhaps a particular plant is emitting large amounts of pollution and the name and address of the plant are desired. The user has two choices. He can go back to the printed output of the CEF and look up the offending plant, or he can use the file of keys for referencing CEF information automatically from the CEF. The extraction of background information from the CEF is carried out as a separate procedure from the rest of the output. Thus, the normal output procedure does not involve the CEF at all.

F. IMPLEMENTATION

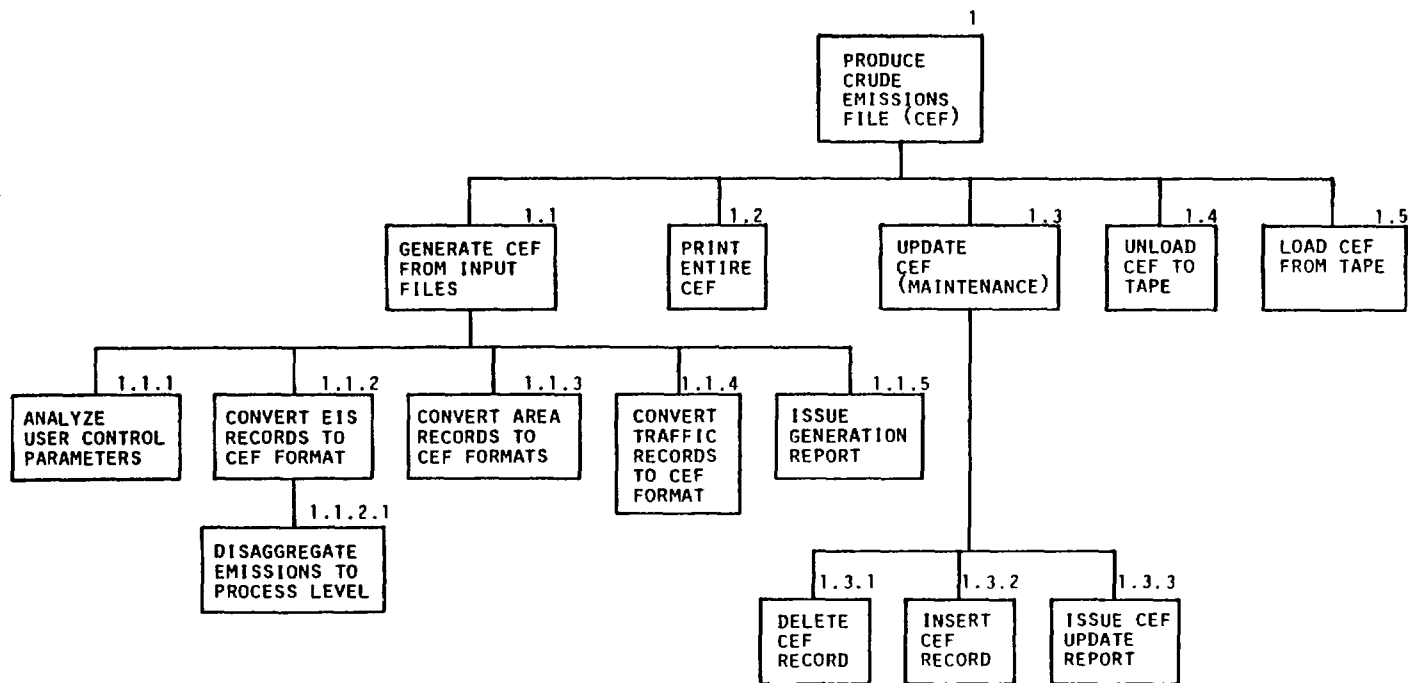
SAI envisions the completion of all implementation tasks within one year of elapsed time. SAI strongly recommends that the ARB assign one person to be involved full-time throughout that year. This will result in reduced implementation costs and will ensure that the ARB will have personnel intimately familiar with the Emissions Simulator system. That person could later assume the responsibility for system maintenance and could implement modifications should they become necessary.

IV DESIGN SPECIFICATIONS

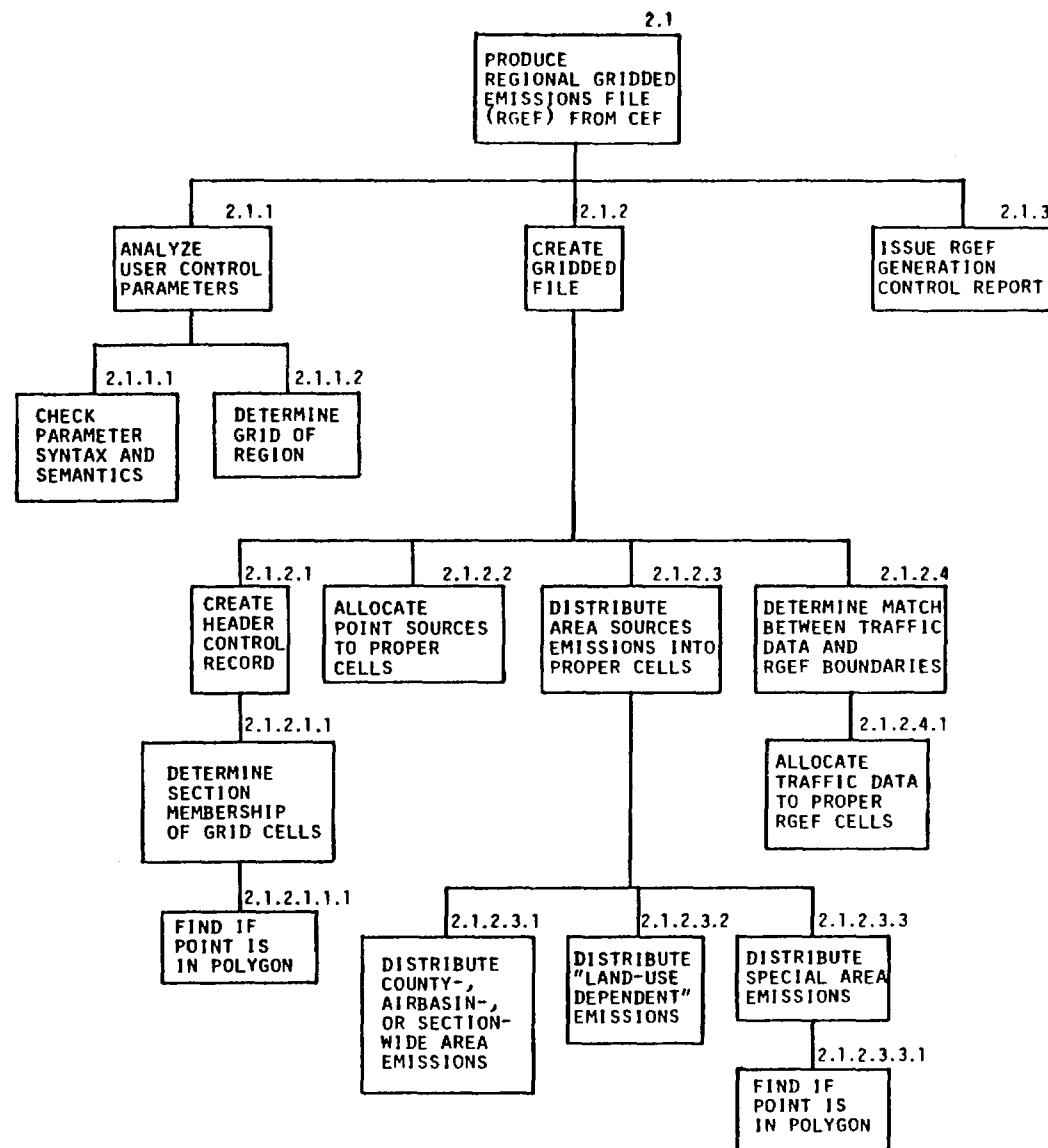
This chapter contains a detailed description of the computer programs that make up the Emissions Simulator system. A top-down functional chart is introduced and followed by a description of each major module. This description includes a general overview, file definitions, user command language and, for each program, a description of input/output files and transformation logic. The programs are related to the top-down functional chart and are numbered accordingly.



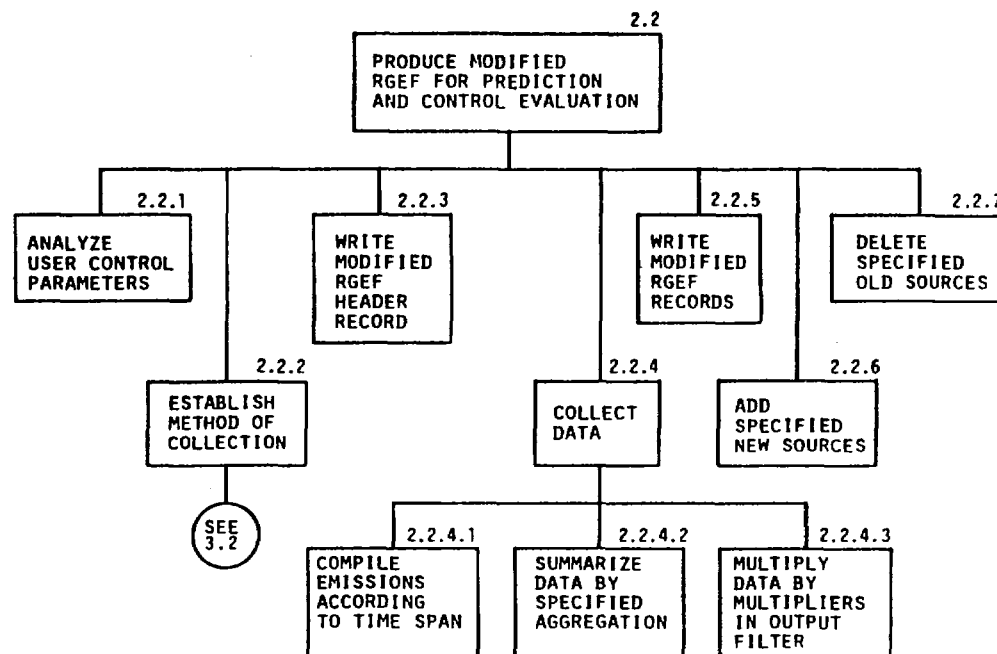
EMISSIONS SIMULATOR TOP-DOWN FUNCTIONAL CHART



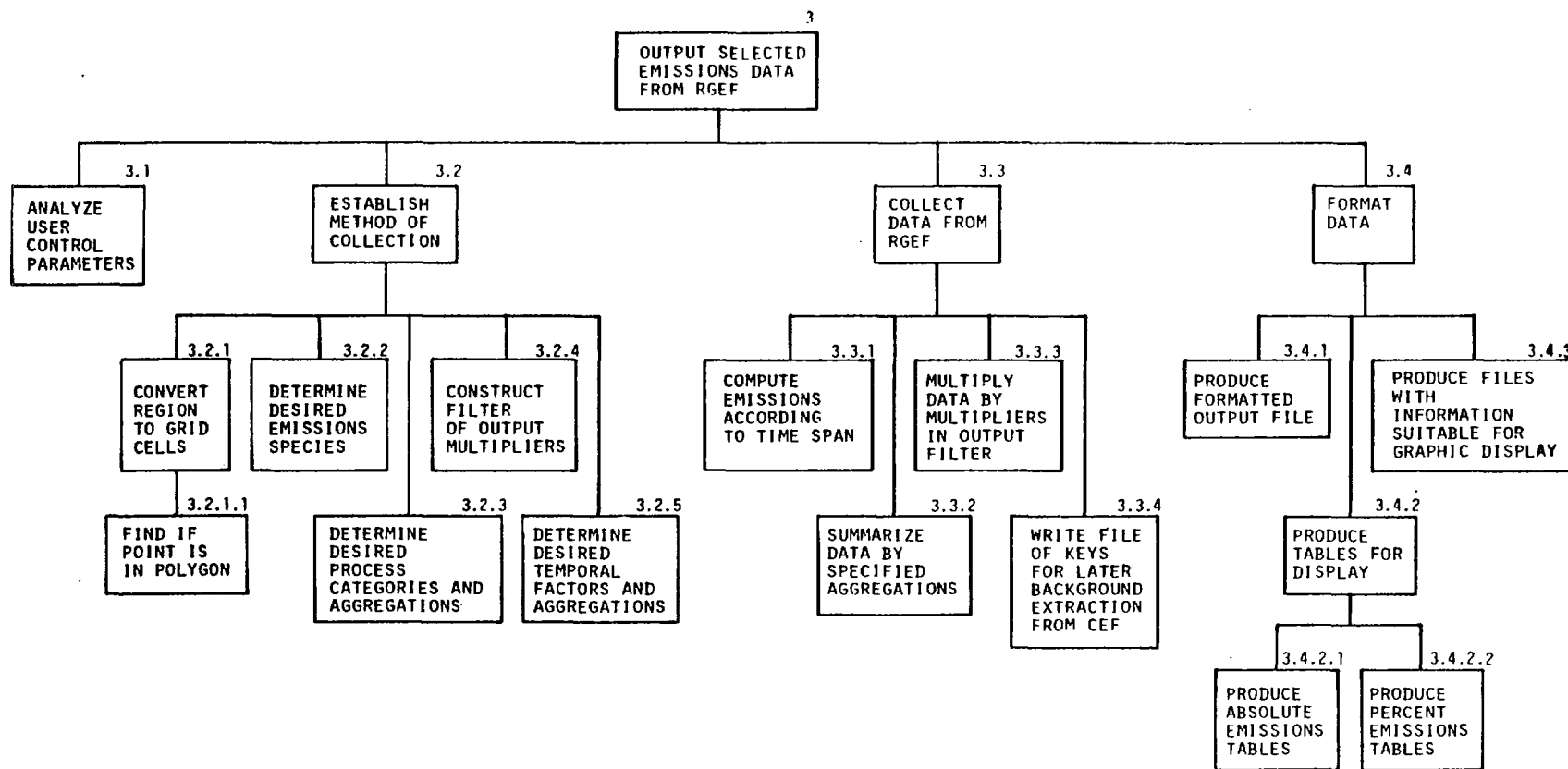
EXPANSION OF MODULE 1



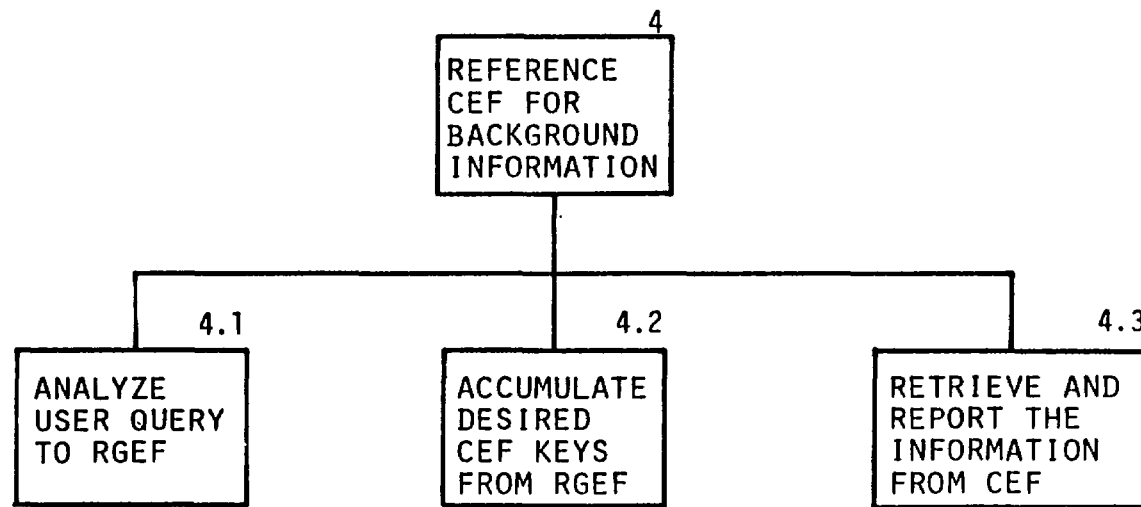
EXPANSION OF MODULE 2.1



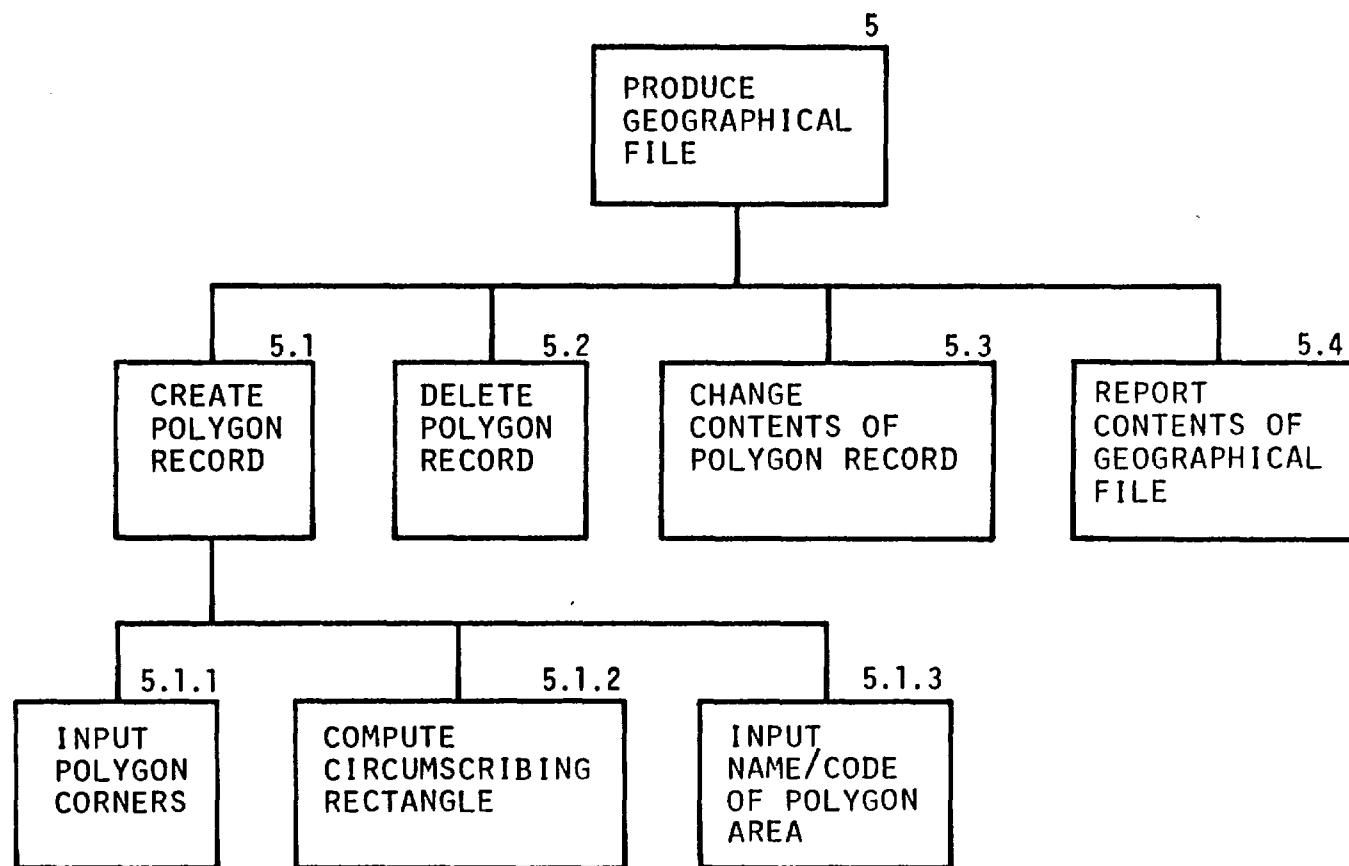
EXPANSION OF MODULE 2.2



EXPANSION OF MODULE 3



EXPANSION OF MODULE 4



EXPANSION OF MODULE 5

CEF FILE DEFINITION

Name: Crude Emissions File

Purpose: The main repository of emissions information in a standardized form, keyed for easy access. Also contains other information for reference purposes. Designed to make creation of gridded files a natural process.

Origin: "Generate CEF from Input Files" Program

Destination: "CEF Maintenance Update"
"Print CEF"
"Produce RGEF"
"Reference CEF for Background Information"
"CEF Maintenance Update"

Format: Variable length records

File organization: Key sequenced

Access method: VSAM

Medium: Disc

Estimated size:

Record types: Header Record
Point Source Record
Area Source Record
Traffic Grid Cell Record

CEF FILE DEFINITION

Record Descriptions

Record Name: Header Record

Field Name:	Description
-------------	-------------

KEY FIELDS

Record type	1 Byte. Value 0 to show it is a header record
Filler	

NON-KEY FIELDS

Creation time and date
Last update time and date
Last update performed by

CEF FILE DEFINITION

Record Descriptions

Record Name: Point Source Record (Note that each process within a point source has its own record.)

Field Name:	Description
-------------	-------------

KEY FIELDS

Record type	1 Byte. Value 1 to denote Pt. Source Record
County	2 Bytes. Binary
Air basin*	2 Bytes. Binary
Plant id	7 Bytes. Packed Decimal
Segment 1 id	2 Bytes. Zone Decimal
Segment 2 id	2 Bytes. Zone Decimal

} These 3 fields are for obtaining background reference to CEF from RGEF

NON-KEY FIELDS

AQCR

SIC-IPP Code	4 Bytes. Binary
SCC Code	4 Bytes. Binary
UTM Coordinates	
UTM zone	1 Byte. Binary
Horizontal	4 Bytes. Float Decimal
Vertical	4 Bytes. Float Decimal

Temporal Distribution information
(as in EIS file)

Other information desired from EIS
file: e.g., plant information,
stack data, AQCR

Emissions (repeated for each pollutant)

 Pollutant id

 Emissions (annual and hourly)

* The EIS file uses AQCRs instead of Air Basins, but this will have to be changed as Area files are based on air basins and counties.

CEF FILE DEFINITION

Record Descriptions

Record Name : Area Source Record (Note that each process within an activity has its own record.)

Field Name:	Description
KEY FIELDS	
Record type	1 Byte. Value 2 to denote Area Source Record
County	2 Bytes. Binary
Air basin	2 Bytes. Binary
Category/activity code	2 Bytes. Binary
File number	2 Bytes. Binary
Process code	2 Bytes. Zoned Decimal
<div style="display: flex; align-items: center;"> <div style="flex: 1;"> <p>This information is for obtaining background reference to CEF from RGEF.</p> </div> </div>	
NON-KEY FIELDS	
Type of area source	1 Byte. Possible values are: <ol style="list-style-type: none"> 1 County 2 Air basin 3 County-Air basin section 4 User defined--coordinates follow 5 User defined--geographical file area id follows
UTM coordinates and zones	(if type = 4)
Geographical file defined area id	2 Bytes. Binary
Distribution parameter	2 Bytes. Binary
Temporal distribution information (as in Area File)	
Emissions (repeated for each pollutant)	
Pollutant code	4 Bytes. Binary
Emissions (average annual)	
(maximum hourly)	
Other information desired from Area File--for backup purposes	

CEF FILE DEFINITION

Record Descriptions

Record Name: Traffic Grid Cell Record

The format of this record is unspecified since the format of the traffic source file is still unknown. However, it is safe to say that the following fields must be included:

Field Name:	Description
KEY FIELDS	
Record type	1 Byte. Value 3 to denote Traffic
County	Grid Cell Record
Air basin	
Process code	
Grid cell identification	
NON-KEY FIELDS	
Temporal distribution information	
Emissions (repeated for each pollutant)	
Pollutant code	
Emissions (hourly)	

CEF GENERATION REPORT FORMAT

CEF GENERATION REPORT

DATE: MM/DD/YY

PAGE XXX

CEF SUCCESSFULLY CREATED or

ERROR IN CEF GENERATION: CEF NOT CREATED

Error message

[and optionally]

CONTENTS OF CEF

POINT SOURCES

.
.

- listing of sources in key sequenced order
with all the information in each record
formatted nicely

AREA SOURCES

.
.
.

TRAFFIC SOURCES

Top Down Functional Breakdown

1.1 Generate CEF from Source Files

This function is a self contained program. It may also contain "Print Entire CEF" program, or invoke it via JCL. Its subfunctions will be subroutines in the program. They are described in detail below.

1.1.1 Analyze User Control Parameter

Input

PARM passed through the JCL.

Output

None. But this module sets up the control structure which can trigger the invocation of "Print Entire CEF" program.

Transformation

PARM is checked.

- > If PARM = 'PRINT' then entire CEF will be printed after it is created.
- > If PARM = something else, error raised, causing program termination and error message in CEF Generation Report.
- > If there is no PARM (PARM='') then CEF Generation Report will have the short form, no print of the CEF.

1.1.2 Convert EIS Records to CEF Format

Input

Emission Inventory System, Point Source File

Output

Point Source Records in CEF

Transformation

The module reads through a set of segment 0, segment 1, and segment 2 records on the EIS, and for each segment 2 record (process), constructs a corresponding CEF record.

The CEF Point Source Record contains identifying information from all three EIS segments, and is, in addition, a keyed record (see CEF record descriptions). It contains all the general information about the plant, point, and process that users may want to know for background information--possibly it may contain all the information currently stored in the EIS segments.

Since the emissions will be kept at the process level in CEF Point Source Record, they must be disaggregated from the Point Source segment (segment 1) of EIS, using emission factors from EIS process segment and ARB's general emission disaggregation algorithm (see next module). (Ultimately ARB may store emissions on the process level in the EIS.)

1.1.2.1 Disaggregate Emissions to Process Level

Input

EIS Point Source Segment Records and Process Segment Records

Output

Point Source (Process Level) Record in CEF

Transformation

The algorithm has been developed by ARB (see draft by J. Graham, 3/23/77, on "Disaggregation of Emissions"). The basic idea is to reconcile the point source emissions, measured in tons/year (and derived often through direct measurement) and the process emissions factors, measured in pounds per SCC units together with annual process weight in SCC units (and derived from theoretical calculation).

1.1.3 Convert Area Records to CEF Format

Input

Area Source File

Output

Area Source Records in CEF

Transformation

The module reads through a set of Level 1 (Activity) and Level 2 (Process) Records and for each Process Record, constructs a corresponding CEF record.

The CEF Area Source Record contains information from both Activity and Process Level Records, and is a keyed record. It will contain all information about the Area, Activity, and Process needed anywhere in the EMSIM system.

The Area File is designed so that Emissions are already disaggregated to the process level, so they can be used directly in the CEF record.

The Area described at the Activity Level is either defined by a set of UTM-coordinates in the Area File itself or by a reference to the geographical file. This description is preserved in the CEF.

1.1.4 Convert Traffic Records to CEF Format

Input

Traffic Source File

Output

Traffic Grid Cell Records in CEF

Transformation

The transformation performed by this module is not yet known in detail since the traffic file format is still unavailable. It will be necessary to convert each record so that its key suits the requirements of the CEF (see Traffic Grid Cell Record Description). It is hoped that the traffic file will be organized in grid cells of single kilometer square size, so as to fit easily into the EMSIM gridding method.

1.1.5 Issue Generation Report

Input

Error messages from other modules

System Time and Date

CEF

Output

CEF Generation Report

Transformation

At end of the "Generate CEF" program, this module prints a report of what has happened, listing any errors that may have occurred causing termination of the program, or reporting on successful completion time and date (the same time and date as are placed in the CEF Header Record).

If the "PRINT" parameter is specified, this module invokes the "Print Entire CEF" module to print its contents as part of the CEF Generation Report.

1.2 Print Entire CEF

This is a self contained program, though it may possibly be included as a subroutine of "Generate CEF from Source Files."

Input

CEF

Output

CEF contents report

Transformation

This module prints CEF, record by record in a straightforward, well-formatted manner, in key-sequenced order, organized by point, area, and traffic sources, and by county and air basin within those categories.

The printout produced by this program is an important reference document. Copies of it should be kept available to users of the EMSIM system for background information. Some display programs will print out lists of keys so that the user can look up relevant records in the CEF. Sometimes looking at the CEF report by hand will be faster than invoking the "Reference CEF for Background Information" program.

1.3 Update CEF (Maintenance)

Input

File of card images and commands
CEF

Output

CEF
CEF Update Report

Transformation

This function is a stand-alone program (not a subroutine). Its purpose is to correct any errors which may have occurred during the creation of the CEF. It should not be used for generalized updating of CEF as this will cause inconsistencies with RGEFs.

The command syntax is one line consisting of:

$$\left\{ \begin{array}{l} \text{INSERT} \\ \text{DELETE} \end{array} \right\} \quad \left\{ \begin{array}{l} \text{POINT} \\ \text{AREA} \\ \text{TRAFFIC} \end{array} \right\}$$

followed by a set of lines which describe the record to be inserted or deleted. To delete a record, the DELETE command is followed by a line describing the key of the record to be deleted. To insert a record, the INSERT command is followed by a set of records in the same format as the standard input file. So, for point source insertion, there must be a segment 0, segment 1, and any number of segment 2 records in EIS format. For area source insertion, there must be an activity record and any number of process records in area file format.

On deletion, an error is raised if key does not exist in file. On insertion, error is raised if key already exists on file. On completion of all insertions and deletions, the CEF header record "last update time and date" record is updated for consistency purposes.

1.3 Update CEF (Maintenance) (cont'd.)

To replace or change a record in the CEF, it is necessary to delete the record, then insert the changed version of the record.

The CEF update report contains each input command and input record followed by a successful completion message or an error message. An error in one insertion or deletion causes control to pass to the next insertion or deletion--the job is not terminated.

1.3.1 Delete CEF Record

Input

Key of a CEF Record

Output

Message to CEF update report

Transformation

Very straightforward. Read the key from input file, convert to internal form, attempt to delete the CEF record with said key. If delete goes through, write success message on CEF update report. If key not found on CEF file, write error message to that effect on CEF update report.

1.3.2 Insert CEF Record

Input

Records in source file format

Output

Message to CEF update report

CEF records

Transformation

This module will incorporate the same routines used by "Generate CEF from Source Files" program, since the input records will be in the same format as the original source files.

The information on the source records will be converted to CEF formats, and attempt will be made to write the new record(s) on CEF. If write is successful, success message is sent to CEF update report. If key already exists, error message is sent to CEF update report.

1.3.3 Issue CEF Update Report

Input

File of card images and commands

Error messages

Output

CEF Update Report

Transformation

Report will look like this:

CEF MAINTENANCE UPDATE REPORT

DATE MM/DD/YY

PAGE XXX

CEF CREATION DATE MM/DD/YY

UPDATE COMMANDS:

DELETE AREA

KEY = xxx...

*** AREA DELETED

INSERT POINT

seg 0 xxx...

seg 1 xxx...

seg 2 xxx...

*** ERROR: POINT SOURCE KEY ALREADY EXISTS.

.
. .
.

1.4 Unload CEF to Tape

1.5 Load CEF from Tape

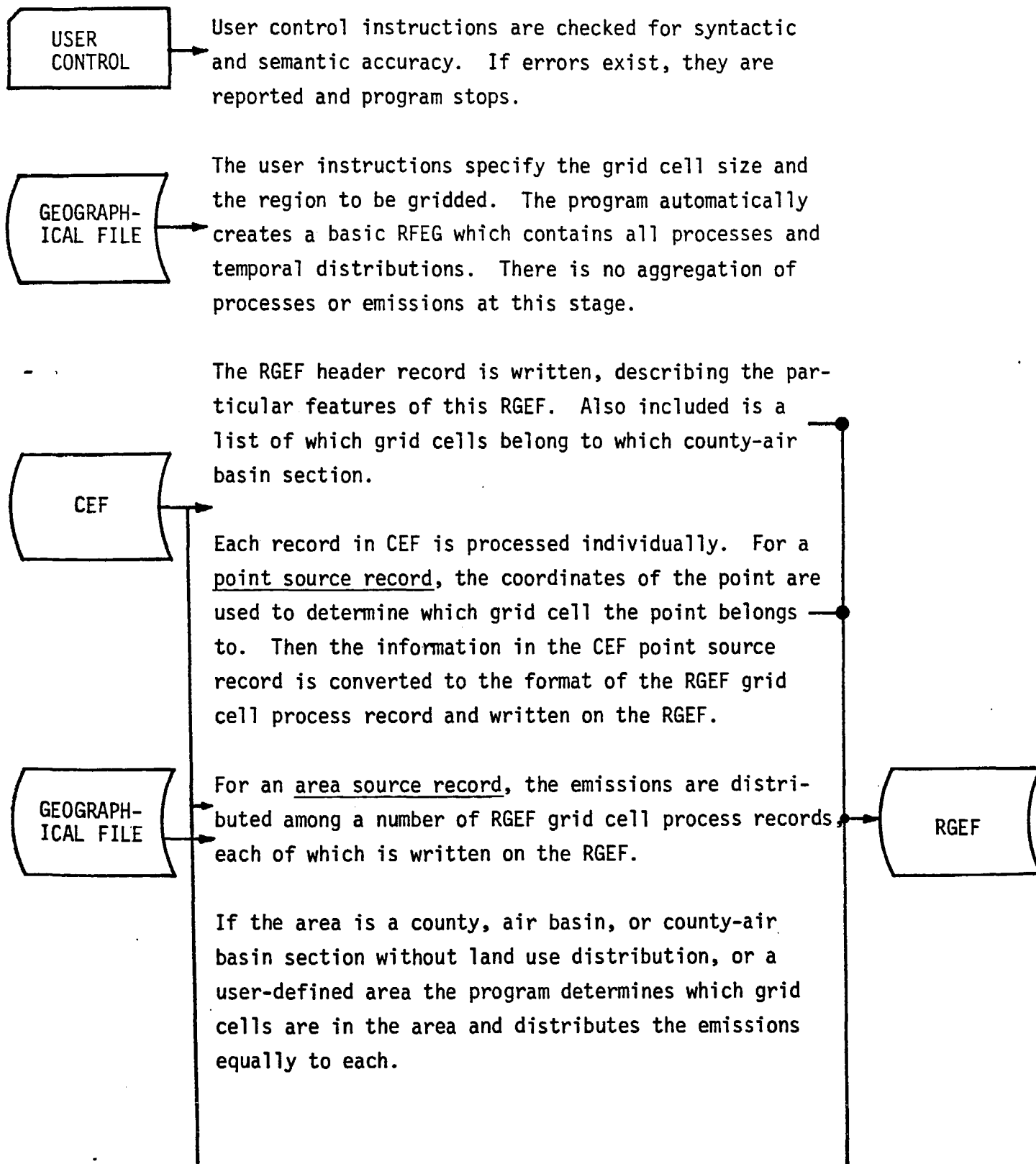
These functions can be carried out directly by VSAM utilities which convert a VSAM file to a sequential (tape) representation and vice versa. These utilities can be invoked through use of "canned" JCL.

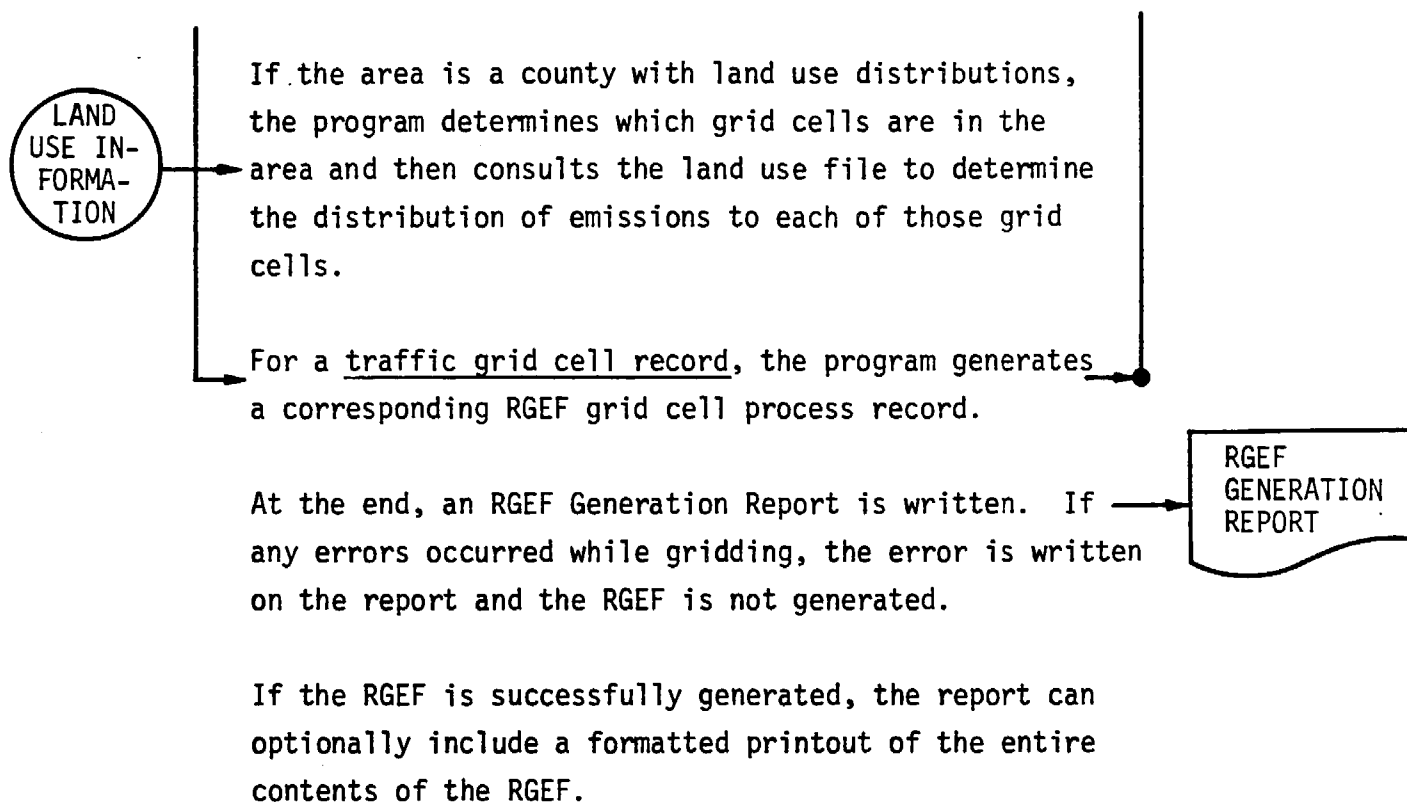
It will be convenient to keep the CEF on tape 1) as backup, 2) as a means of saving direct access storage space. If the user is satisfied to consult the printed contents of the CEF file, he or she can avoid keeping the CEF on disc once all the basic RGEFs have been created.

2. PRODUCE REGIONAL GRIDDED EMISSIONS FILES (RGEFs)

This function of the EMSIM system consists of two separate programs. One creates an RGEF from the CEF (2.1) and the other creates a modified RGEF from an existing RGEF (2.2). The first program (2.1) is used to create a number of basic RGEF's for standard areas of interest in the state. The second program (2.2) is used to create specialized "modified" RGEFs for short-term purposes, especially for use in predictions and control evaluations. It is very similar to the output program (3) and since many of their modules are identical they will be described under the output program.

2.1 Produce RGEF from CEF



2.1 Produce RGEF from CEF (cont'd.)

RGEF FILE DEFINITION

Name: Regional Gridded Emissions File (Note: There should be one basic RGEF for each region of interest in the state. The smallest region allowed is a county-air basin section.)

Purpose: Basic RGEF's contain all the emissions of a region, broken down by process and by geographical grid cell. They will be used as the data base to which emissions queries are addressed, and thus as the source for inventory reports and of data for modeling. Basic RGEF's are also used to create modified RGEF's, which are used for prediction and control evaluation purposes. An RGEF is either "basic" or "modified," but not both.

Origin: Produce RGEF from CEF
Produce Modified RGEFs

Destination: Produce Modified RGEFs
Output Selected Emissions Data

Format: Variable length records

File organization: Key sequenced

Access method: VSAM

Medium: Disc

Estimated size:

Record types: Header Record
Second Header Record
Grid Cell Process Record

RGEF FILE DESCRIPTION

Record Description

Record Name: Header Record

Field Name:	Description
-------------	-------------

KEY FIELDS

Record type	1 Byte. Value 0 denotes header.
Filler	

NON-KEY FIELDS

Modified status	1 Byte. Value M means it is a modified RGEF Value B means it is basic RGEF
-----------------	---

Creation time and date

Southwest corner UTM coordinates

Zone

Horizontal

Vertical

Grid cell size in km

Height of region in km

Width of region in km

Section breakdown (repeated for each
county-air basin section in region)

County id

Air basin id

Number of grid cells in section

Grid cell numbers of cells in
this section

RGEF FILE DESCRIPTION

Record Description

Record Name: Second Header Record (Note: This record may actually be written as a number of variable-length records.) This record embodies the definition of a modified RGEF. It only appears in modified RGEFs

Field Name:	Description
KEY FIELDS	
Record type	1 Byte. Value 2 denotes second header.

Filler

NON-KEY FIELDS

Switches indicating temporal span of file

Other definitional switches (e.g. for PROCESSES: ALL or OTHER)

PROCESS array

AGGREGATES array

VALUES array

Filter of output multipliers

List of species

RGEF FILE DEFINITION

Record Description

Record Name: Grid Cell Process Record

Field Name:

Description

KEY FIELDS

Record type 1 Byte. Value 1 denotes Grid Cell Process Record

Grid cell number

Type of source 1 Byte. Value: 1--Point, 2--Area, 3--Traffic

Process information*

Point SourceArea SourceTraffic Source

SIC IPP

Category/Activity Code

Process Code

SCC

Process Code

NON-KEY FIELDS

County

Air basin

Referback information (for referencing CEF)

Point SourceArea SourceTraffic Source

AQCR

Plant ID

Segment 1 ID

File number

Traffic source/grid cell
identification

Segment 2 ID

Stack Data

Temporal distribution information

Point SourceArea SourceTraffic SourceQuarterly thruput as
% of annualMonthly thruput as
% of annual

(To be determined)

Hrs/da, wks/yr, da/wk

Hrs/da, wks/yr, da/wk

Hourly thruput

Rate of thruput

RGEF FILE DEFINITION (Cont'd.)

Record Description

Record Name: Grid Cell Process Record

Emissions Information: Number of pollutants

Emissions data (repeated for each pollutant)

Amount/year

Amount/hour

*

The location of process information in the record can be handled in two possible ways, one of which is suggested above. Each method has advantages and disadvantages which ultimately must be tested by the implementer of the system against the actual use of the system. That is, it is impossible to predict in advance which method will be most efficient without knowing how the system will ultimately be used most of the time.

The method suggested here is to keep process information in the key. This approach makes use of the VSAM option of being able to read the VSAM index tables into core. Thus a display program or RGEF modification program can read the VSAM index into core (a chunk at a time) and quickly select those records with the desired grid cell numbers and process information; i.e., to select all dry cleaning processes in Alameda County, the keys could be scanned for those with dry cleaning process numbers and grid cells which belong to Alameda County. The advantage of this method is that it cuts down on disc I/O operations which can be time consuming, especially in cases where selection is based on region alone or on region and process together.

The second method is to use the VSAM alternate index facility to construct an alternate index based on process information alone, and to leave process information out of the original key field entirely. This method would yield two independent sets of keys, one based on grid cell numbers (and hence amenable to queries based on areas) and one based on process information. The advantage of this method is that access to records where selection is based on area alone or process alone is very rapid. However, access to records based on a combination of area and process is slower since it is necessary to actually read the records themselves in order to determine their eligibility for selection. Thus, to select all dry cleaning processes in Alameda County, the program would consult the alternate index which would identify all dry cleaning process records in the RGEF. Then each of those records would have to be read and its grid cell number checked to see if it belonged in Alameda County.

User Control Parameters for Producing RGEF, Syntax, and Semantics

One of these $\left\{ \begin{array}{l} \text{COUNTY} = \text{county}_1 (\text{county}_2) \dots \\ \text{BASIN} = \text{airbasin}_1 (\text{airbasin}_2) \dots \\ \text{SECTION} = \text{county} (\text{airbasin}_1) [\text{county}_2 (\text{airbasin}_2)] \dots \end{array} \right.$

CELL = c

PRINT = $\begin{cases} Y \\ \underline{N} \end{cases}$

where

() means optional word

... means previous term may be repeated any number of times.

county_i is a standard 4 digit county code,

airbasin_i is a standard 3 digit air basin code,

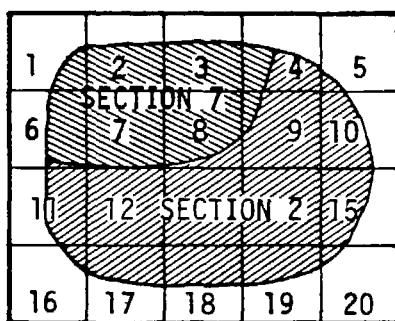
c is an integer specifying number of kilometers on
the side of a grid (square) cell. c may be 1, 2, 3 ...

Y means yes, print the entire RGEF. N means don't.

If this line is left out, the default is no print.

How the "Produce RGEF from CEF" Program Processes the User Control Parameters

After checking that the syntax of the instructions is correct, the program finds each county, air basin, or section in the Geographical File, and retrieves its circumscribing rectangle, which is part of the Geographical File. From this information, the program constructs a circumscribing rectangle for the entire region, rounded so it will be an integer multiple of c kilometers on each side. This is the gridded region.



Then the program determines which grid cells belong to each section and puts the list in the header record.

2.1.1 Analyze User Control Parameters

Input

User control parameters.

Geographical file.

Output

Error messages.

Grid and Region Definition.

Transformation

This module takes the user control parameters entered by the user to define the RGEF and determines from them the exact definition of grid and region for RGEF. A certain user control parameter also controls the amount of detail included in the RGEF Generation Report.

The grid and region definition is stored by another module in the header record of the RGEF, as the definition of grid and region is essential to the nature of the RGEF.

If there are any syntax or semantic errors in the user control parameters, an error message is written on RGEF generation control report and the program terminates.

The geographical file is used to find the polygonal representation of counties, sections, and air basins, and their circumscribing rectangles.

2.1.1.1 Check Parameter Syntax

Input

User control parameters.

Output

Error messages.

Internal representation of user control parameters.

Transformation

This module checks the syntax of user control parameters. (See "User Control Parameters for producing RGEF, syntax and semantics.")

If syntax is fine, then semantics are checked, e.g., check that county numbers represent actual counties, that cell size "c" is an integer, that PRINT = "Y" or "N."

If syntax and semantics are good, then user control parameters are stored in an internal form for use by other modules.

If there are syntax or semantics errors, they are written on the RGEF Generation Control Report, along with a reference to exactly where the errors occurred in the user control parameters. The RGEF is not created if there are any such errors.

Typical error messages are:

- > 09A35 NOT LEGAL COUNTY CODE
- > CELL MUST BE AN INTEGER
- > CANNOT HAVE COUNTY AND BASIN LISTS TOGETHER.

2.1.1.2 Determine Grid and Region

Input

Internal form of user control parameters.

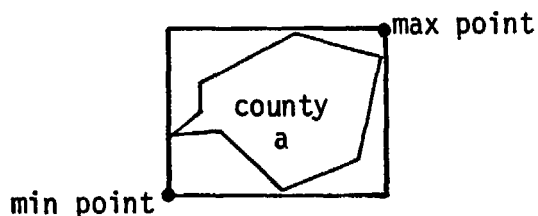
Output

Internal form of Region and Grid Definition.

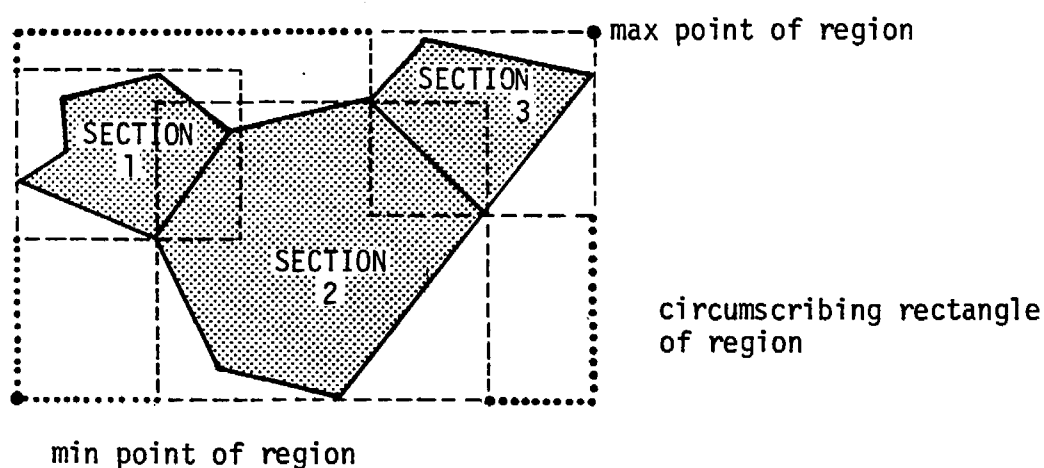
Transformation

The user control parameters specify the region as a set of counties, air basins, or county-air basin "sections." (The three types of area may not be mixed. This restriction does not limit possibilities, as every possible combination of the three types can be expressed as a list of sections.)

Each area (county, air basin, or section) is looked up in the geographical file, where its circumscribing rectangle is retrieved. The circumscribing rectangle is the smallest rectangle along UTM coordinate axes which entirely contains the area. It is determined by two points, the "min point," whose coordinates are (minimum x-coordinate of all points in area polygon, minimum y-coordinate of all points in area polygon) and similarly the "max point."



The circumscribing rectangle of the region B is then computed. Its min point is (minimum x-coordinate of min points of circumscribing rectangles of all areas in region, minimum y-coordinate of min points of circumscribing rectangle of all areas in region). Its max point is similarly defined,

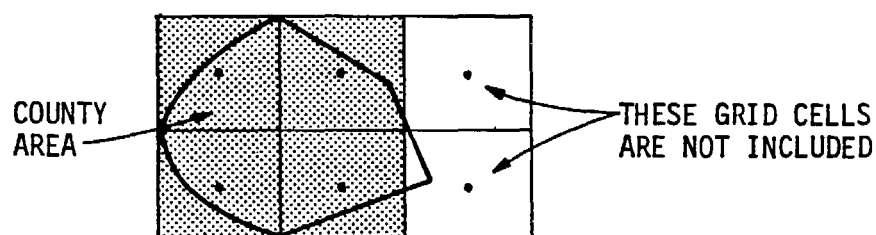


The circumscribing rectangle is then rounded in both height and width to the nearest multiple of c , the grid cell length, so that it can be divided into grid cells. The rounded circumscribing rectangle then becomes the region for the RGEF.

The grid cells within this region are numbered, from west to east, from north to south. Each grid cell is c km on each side.

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20

Further processing occurs in Subroutine 2.1.2.1 "create header control record," which computes which section each grid cell belongs to. If a grid cell belongs to none of the areas specified by user control parameters, it is not included in the RGEF. For example,



2.1.2 Create Gridded File

Input

Internal form of user control parameters.

CEF.

Geographical File.

Land Use Information File.

Output

RGEF.

This submodule is the heart of the "Produce RGEF from CEF" program. For each CEF record, one or more RGEF records are produced, indexed geographically. Each RGEF record represents a process within a grid cell. A point source process is allocated to a single RGEF record. An area source process is allocated possibly to many RGEF records, one for each grid cell in the area. A number of traffic source process grid cells are possibly allocated to one RGEF record (since the traffic source grid cells are presumably 1 km² while the RGEF grid cells may be larger).

2.1.2.1 Create Header Control Record

Input

Internal form of region and grid definition.

Geographical File.

Output

RGEF Header Record.

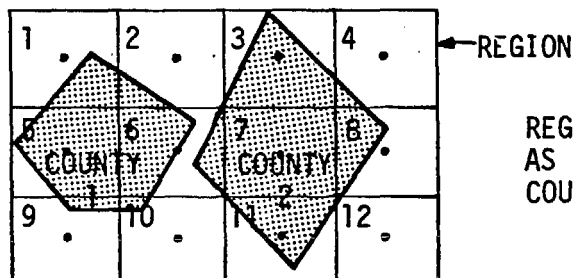
Transformation

"Modified status" field set to "B," meaning this is a basic RGEF. "Creation time and date" taken from system time and date.

The region's circumscribing rectangle is written on header record; it was already calculated in 2.1.1.2--"Determine Grid and Region"--module.

This module then examines the centroid of each grid cell in the region and checks if it is inside any of the sections defined to be part of the region (see point-in-polygon algorithm). If so, it includes the grid cell in the list of cells belonging to that section. When all grid cells have been checked, the list of cells belonging to each section is included in the header record. These lists will come in very handy in the gridding and displaying modules. Those cells that do not belong to any list are not included in the RGEF.

Cells belonging to
county 1: 5, 6
Cells belonging to
county 2: 3, 7, 11
Cells not belonging
to RGEF: 1, 2, 4,
8, 9, 10, 12



REGION IS DEFINED
AS CONTAINING
COUNTIES 1 AND 2

2.1.2.2 Allocate Point Sources to Proper Cells

Input

CEF Point Source Records.

Output

RGEF Grid Cell Process Records

For each air basin-county section in the region, this module retrieves all records from CEF having such a section in their key. The selected CEF records are processed one at a time. The UTM coordinates of the point source are examined to determine which grid cell the point belongs to. The calculation is simple and straightforward. It would probably help to use a standard subroutine (useful elsewhere) to convert from a two-dimensional grid cell numbering to a one-dimensional grid cell numbering, which is the official numbering scheme.

(1,1)	(1,2)	(1,3)	(1,4)
(2,1)	(2,2)	(2,3)	(2,4)
(3,1)	(3,2)	(3,3)	(3,4)

1	2	3	4
5	6	7	8
9	10	11	12

It is easiest to calculate the desired grid cell using the two-dimensional numbering, then to convert to the one-dimensional numbering.

When grid cell is determined, record contents are transferred from CEF format to RGEF format and the RGEF Record is written.

2.1.2.3 Distribute Area Source Emissions into Proper Cells

Input

CEF Area Source Records.

Geographical File.

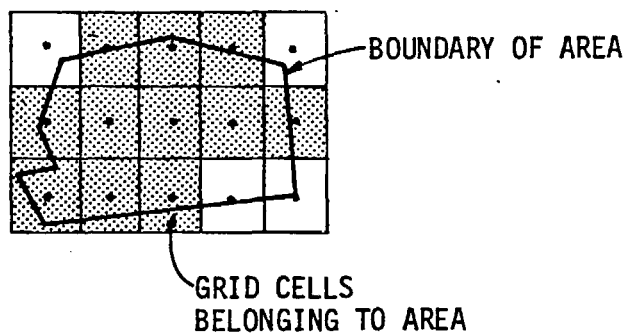
Output

RGEF Grid Cell Process Records.

Transformation

This module retrieves one at a time, all area source records in the CEF whose county-air basin section in the key is contained in the gridding region.

The basic transformation is intuitively clear: find those grid cells that belong to an area, and apportion the area-wide emissions among those grid cells.



A grid cell is said to belong to an area if its centroid is inside the polygon representing the area. For county, air basin, and section areas, the cells belonging to them are already listed in the header record. For user defined areas, it is necessary to retrieve the polygonal representation and to use the point-in-polygon routine to determine which cells belong to the area.

For user-defined areas apportionment of emissions is always done evenly over all grid cells belonging to the area. For county, air basin, and section areas without land use files, apportionment is also done evenly. For county, air basin, or section areas where land use files are available, apportionment varies according to the land use parameter of the area source and the distribution of that land use type among the grid cells. It is helpful if the land use file contains totals for each county and section of each land use type.

A separate record is created for each grid cell belonging to the area and participating in the given process. (All these records refer back to the single area source record in CEF). A possible space-optimization technique would be to store the temporal distributions for these records (which will all be the same) in a special keyed header record, thus saving some duplication of information.

2.1.2.3.2 Distribute "Land-Use Dependent" Emissions

Input

CEF.

RGEF header record.

Land Use File.

Output

RGEF grid cell process records.

This module starts like 2.1.2.3.1, finding the sections that make up the area and the grid cells that make up the sections. For each of these grid cells, the corresponding land use cell(s) are consulted. The land use cells give the amount of land in a given grid cell devoted to each land use type, e.g., .4 square km residential, .6 square km commercial.

The land use totals for each section are needed as well. For each grid cell, for each land use type within the grid cell, the amount of area in the grid cell devoted to the land use type in ratio to the total area of the land use type in the section is computed. Then all area sources having this section in their key are examined. If they have a land use parameter equal to one of the land use types in the grid cell, an RGEF grid cell process is written, with

$$\text{emissions} = \text{total section emissions from this area source} \\ * \frac{\text{amount of area in grid cell devoted to land use } x}{\text{amount of area in section devoted to land use } x} .$$

2.1.2.3.3 Distribute Special Area Emissions

Input

CEF.

Geographical File.

RGEF header record.

Output

RGEF grid cell process records.

Transformation

Special areas here means user-defined areas, which must be entirely within one county or air basin, as indicated in the key of the CEF area source record. If the coordinates defining the polygonal boundary of the area are contained in the CEF record, then they are retrieved. Otherwise, the polygonal boundary of the area is retrieved from the geographical file.

The RGEF header record is consulted to determine which grid cells belong to the county or air basin containing the user-defined area. The centroid of each cell is examined to see if it is inside the given area (using the Point-in-Polygon algorithm). When all eligible grid cells have been examined, the emissions are divided evenly among all cells inside the given area, and a new RGEF record is written for each grid cell inside the area.

2.1.2.1.1.1 and

2.1.2.3.3.1 Find if point is in Polygon

Input

UTM coordinates of a point.

Description of a polygon as list of points, each assumed connected to the next by a straight line, also a straight line assumed between last point and first point. The lines are assumed not to cross each other. The polygon need not be convex. No two consecutive lines may be horizontal.

Output

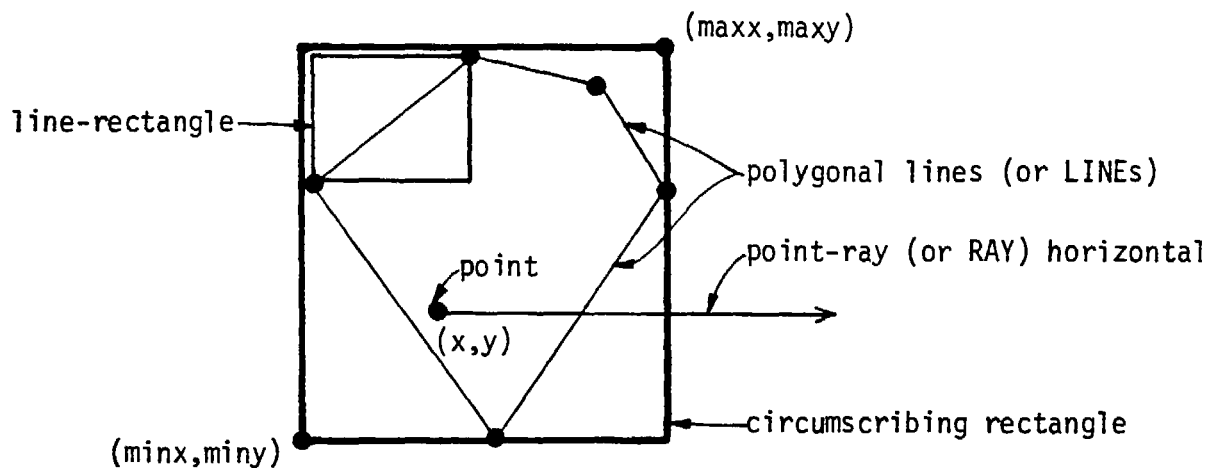
True or False, a flag showing whether the point is in the polygon.

Transformation

Use the "point in polygon" algorithm described below. The algorithm can be implemented in an efficient manner. Most of its time is taken up in the branching structure which may be optimized so that the most frequently traversed logical branches are encountered first. The algorithm takes more time as the number of points in the polygon increases.

Point in Polygon Algorithm

The algorithm uses the following terminology:



1. Check if point is within the circumscribing rectangle. If not, then point is not in polygon.
2. Set cross-count = 0.
3. For each polygonal line, with endpoints (x_1, y_1) and (x_2, y_2) do the following:

If point-ray intersects line-rectangle (see below for how to find out).

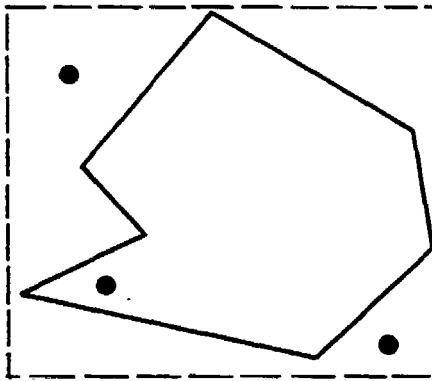
Then if point-ray crosses polygonal-line (see below for how to find out).

Add 1 to cross-count.

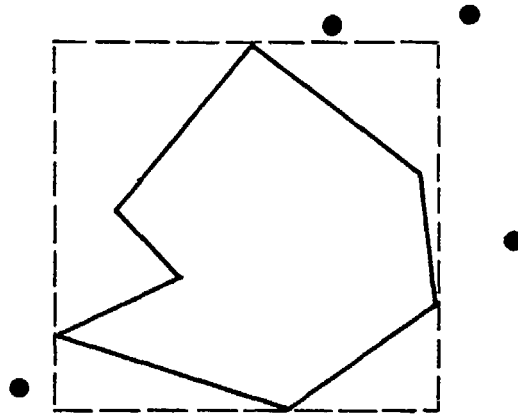
4. If cross-count is even then point is not in polygon,
 else (cross count is odd so) point is in polygon.

A more detailed explanation of point-in-polygon algorithm:

1. Point is within circumscribing rectangle of polygon if, and only if, $\text{minx} \leq x \leq \text{maxx}$, $\text{miny} \leq y \leq \text{maxy}$, where $(\text{minx}, \text{miny})$ is the min-point of the circumscribing rectangle and $(\text{maxx}, \text{maxy})$ is the max-point of the circumscribing rectangle.



points within
circumscribing rectangle



points outside
circumscribing rectangle

3. For each polygonal line p_1p_2 , where $p_1 = (x_1, y_1)$ and $p_2 = (x_2, y_2)$ do the following:

Let previous polygonal line be p_Lp_1 and next polygonal line be p_2p_N , where $p_L = (x_L, y_L)$, $p_N = (x_N, y_N)$. (The lines are viewed in a circle so that for the last polygonal line, p_N actually is the first point of the first line.)

```

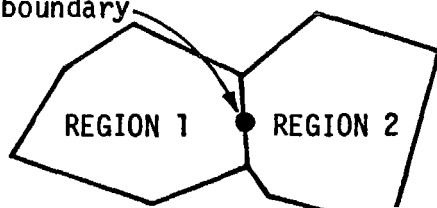
If case 1 (outside rectangle) then do nothing
else (Case 2 holds), (inside rectangle)
  if Case 3* (ray crosses line and line not horizontal)
    then if Case 5 (ray not at vertical extremes of rectangle)
      then add 1 to cross-count
    else (Case 6)(ray intersects an endpoint of polygonal line)
      if Case 7 ( $y = y_1$ )
        then do nothing (intersection dealt with elsewhere)
      else (Case 8)( $y = y_2$ )
        if Case 9 ( $y_1$  and  $y_N$  on opposite sides of  $y_2$ )
          then add 1 to cross-count
        else if Case 10 ( $y_1$  and  $y_N$  on same side of  $y_2$ )
          then do nothing
        else (Case 11)( $p_2p_N$  is horizontal)
          do nothing (intersection will be dealt with elsewhere)
    else (Case 4)(ray does not cross line or line is horizontal)
      if Case 12 [line is horizontal ( $y_1 = y_2$ ) and ray crosses
        line ( $y = y_1$ )]*
        then if Case 14 (previous and next lines same sided)
          then do nothing
        else if Case 15 (previous and next lines opposite sided)
          then add 1 to cross-count
        else (Case 16)(two horizontal lines in a row)
          signal error in input
      else (Case 13)(line not horizontal or ray does not cross line)
        do nothing.

```

* At these points, Case 3 and Case 12, for completeness the algorithm could check if the point actually lies on the polygonal line itself (not its point-ray). If so, then the algorithm is done immediately since the point is in the polygon.

If this extra check is not performed, some points lying on the boundary of the polygon will not be evaluated as lying in the polygon. However, even if this extra check is performed, boundary problems are not avoided since the point will also be evaluated as lying in the adjacent region as well.

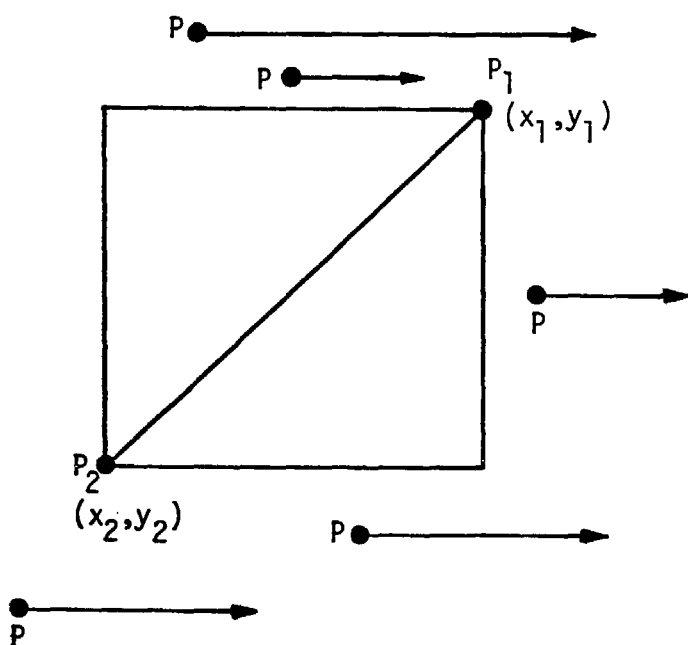
point lying on boundary



Case 1

Point-ray does not intersect line-rectangle
 i.e., $y > \max(y_1, y_2)$ or $y < \min(y_1, y_2)$
 or $x > \max(x_1, x_2)$

Examples of point-rays satisfying Case 1:



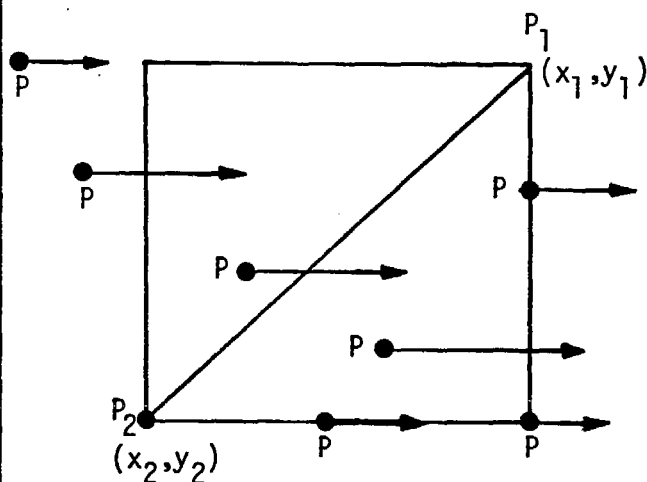
Note: The length of the arrow in the point-ray is immaterial; it is assumed to be infinite.

Note: Each diagram shows for each polygonal line, a number of possible point-rays satisfying the case.

Case 2 (complement of Case 1)

Point-ray intersects line-rectangle
 i.e., $\min(y_1, y_2) \leq y \leq \max(y_1, y_2)$
 and $x \leq \max(x_1, x_2)$

Examples of point-rays satisfying Case 2:



Case 3 (assuming Case 2)

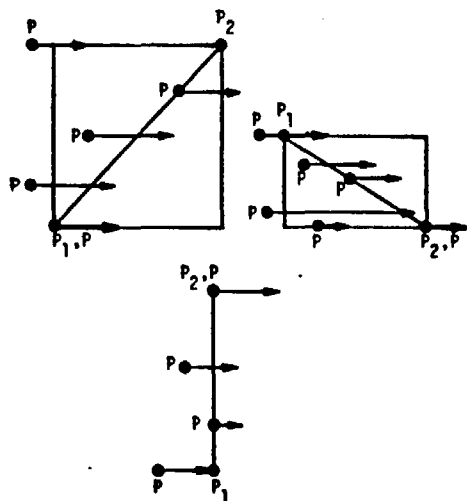
Point-ray crosses polygonal line and polygonal line is not horizontal; i.e.,

$$\frac{x_2^*(j_1 - y) + x_1^*(y - y_2)}{y_1 - y_2} \geq x$$

and $y_1 \neq y_2$.

Note that the left side of the inequality is the x-coordinate of the point where the line corresponding to the point-ray crosses the polygonal line.

Examples of polygonal lines and point rays satisfying Case 3:

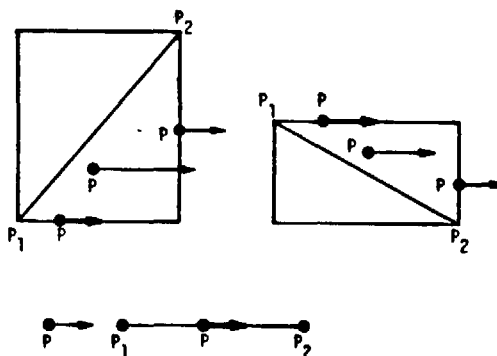


Case 4 (assuming Case 2) is the complement of Case 3 point ray does not cross polygonal line or polygonal line is horizontal; i.e.,

$$\frac{x_2^*(y_1 - y) + x_1^*(y - y_2)}{y_1 - y_2} < x$$

or $y_1 = y_2$

Examples satisfying Case 4:

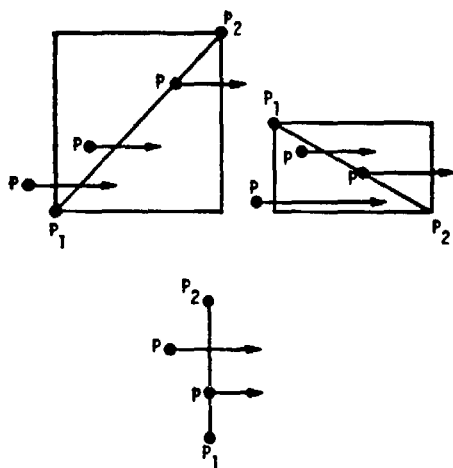


Case 5 (assuming Case 2, Case 3 hold)

Point-ray does not intersect vertical extremes of line-rectangle; i.e.,

$$\min(y_1, y_2) < y < \max(y_1, y_2) \quad .$$

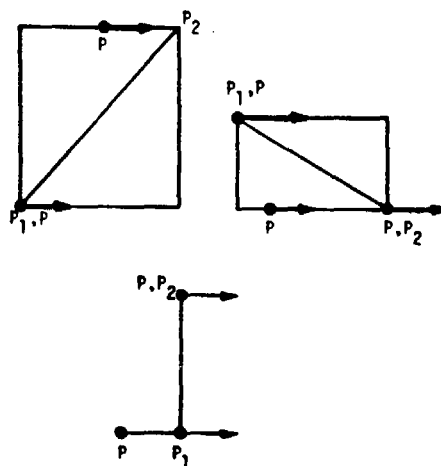
Examples of Case 5:

Case 6--complement of Case 5
(assuming Case 2, Case 3 hold)

Point-ray intersects a vertical extreme of line-rectangle; i.e.,

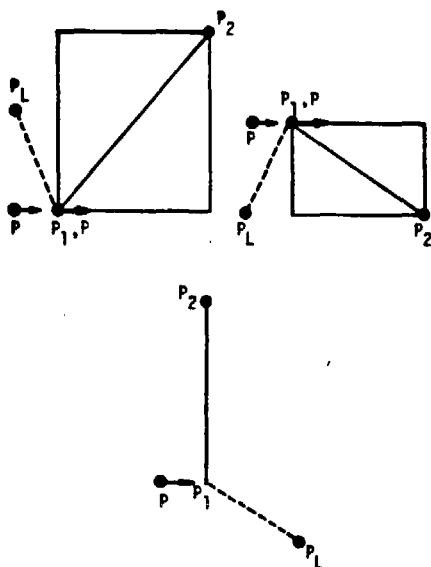
$$y = y_1 \quad \text{or} \quad y = y_2 \quad .$$

Examples of Case 6:



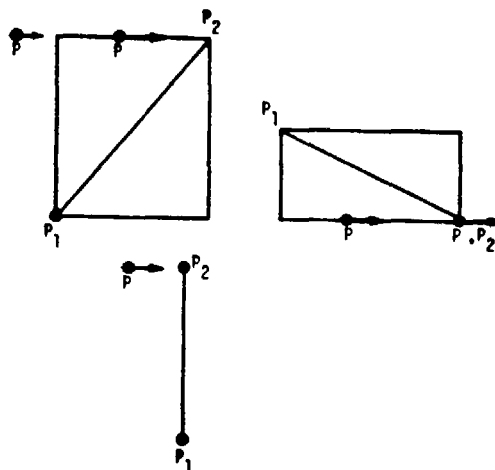
Case 7 (assuming Case 2, Case 3, Case 6 hold)

$$y = y_1$$



Case 8--complement of Case 7 (assuming Case 2, Case 3, Case 6 hold)

$$y = y_2$$



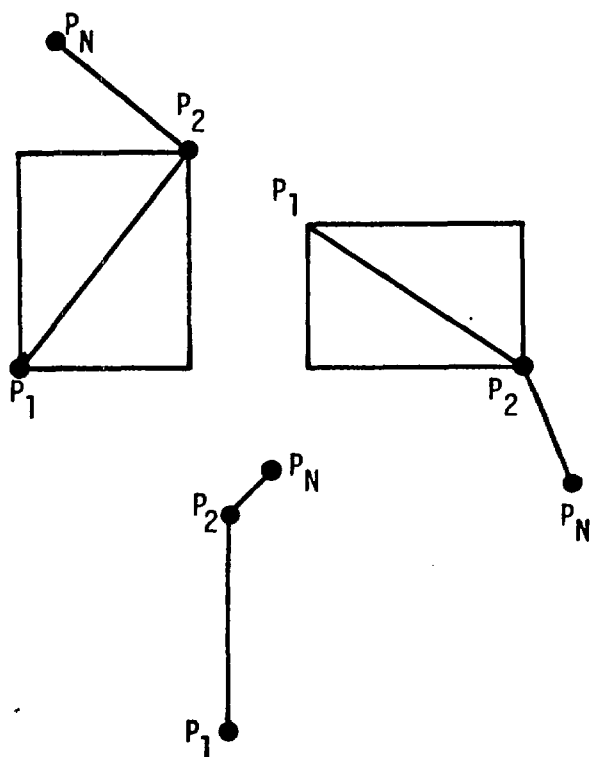
In this case we do not treat the intersection at all since it is handled for the segment $P_L P_1$, where it appears as Case 8.

Case 9 (assuming Case 2, Case 3,
Case 6, Case 8 hold)

y_1 and y_N are on opposite sides
of y_2 , i.e.,

$(y_1 < y_2 \text{ and } y_N > y_2)$ or

$(y_1 > y_2 \text{ and } y_N < y_2)$.

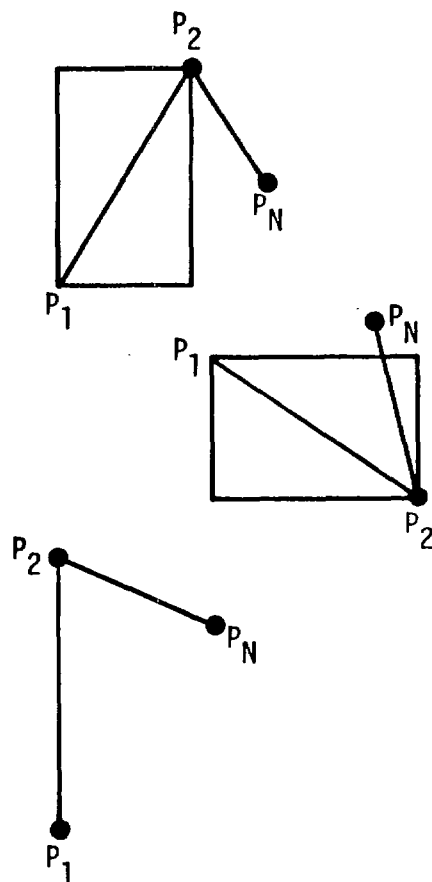


Case 10 (assuming Case 2, Case 3,
Case 6, Case 8 hold)

y_1 and y_N are on same side of y_2 ,
i.e.,

$(y_1 < y_2 \text{ and } y_N < y_2)$ or

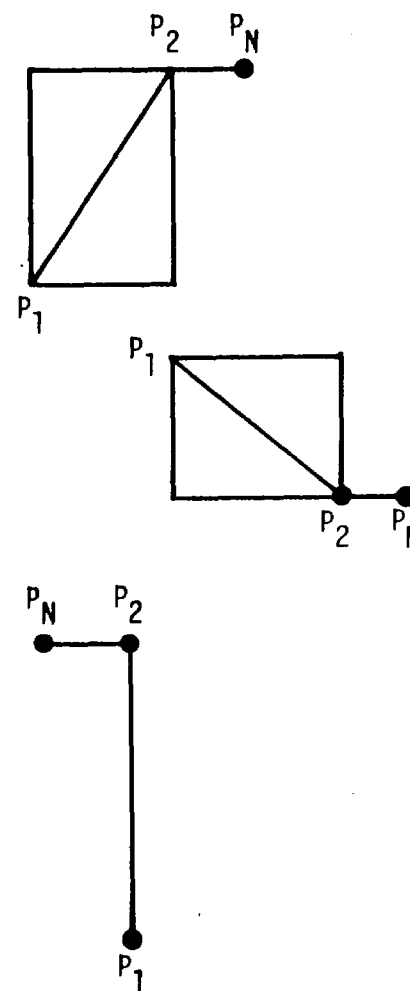
$(y_1 > y_2 \text{ and } y_N > y_2)$



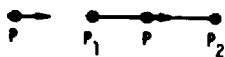
Case 11 (assuming Case 2, Case 3,
Case 6, Case 8 hold)

P_2P_N is horizontal, i.e.,

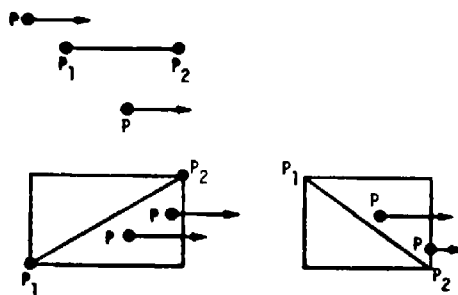
$$y_2 = y_N$$



Case 12 (assuming Case 2, Case 4 hold)



Case 13--complement of Case 12
(assuming Case 2, Case 4 hold)

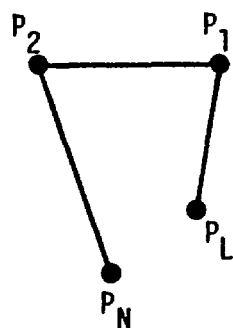
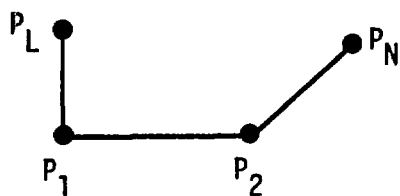


Case 14 (assuming Case 2, Case 4, Case 12)

Previous polygonal line and next polygonal line are same-sided, i.e.,

$$(y_L > y_1 \text{ and } y_N > y_1) \quad \text{or}$$

$$(y_L < y_1 \text{ and } y_N < y_1)$$

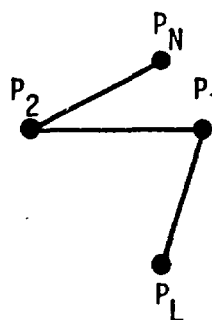
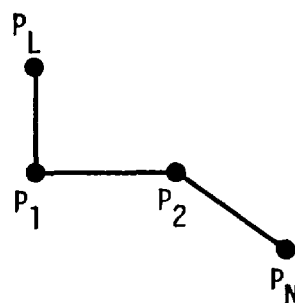


Case 15 (assuming Case 2, Case 4, Case 12)

Previous polygonal line and next polygonal line are opposite-sided, i.e.,

$$(y_L > y_1 \text{ and } y_N < y_1) \quad \text{or}$$

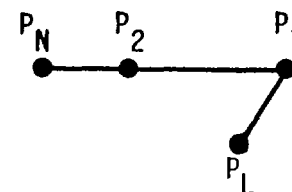
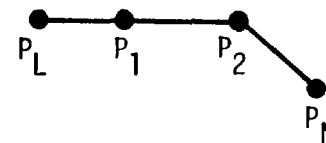
$$(y_L < y_1 \text{ and } y_N > y_1)$$



Case 16 (assuming Case 2, Case 4, Case 12)

Previous polygonal line or next polygonal line is horizontal, i.e.,

$$y_L = y_1 \quad \text{or} \quad y_N = y_1$$



2.2 Produce Modified RGEF for Predictions and Control Evaluation (cont'd.)

At the end, the Modified RGEF Report is written. This report contains any error messages arising during the execution of the program on a message signaling the successful creation of the Modified RGEF. It includes the user control instructions defining the Modified RGEF and possibly contains a printout of the entire new file.



MODIFIED
RGEF

2.2.1 Analyze User Control Parameters

2.2.2 Establish Method of Collection

2.2.4 Collect Data

See Sections 3.1, 3.2, and 3.3, respectively.

2.2.3 Write Modified RGEF Header Record

Input

Internal forms of user control parameters.

Output

Second RGEF Header Record.

Transformation

The Modified RGEF has a standard header record like any RGEF, which has a flag indicating Modified status. In addition it has a second Header Record which gives the definition of the Modified RGEF as specified in the user control parameters which created it.

The various internal arrays and switches representing the user control information are transferred to the second Header Record along with information on the size of each array. The second Header Record may in fact be written as a number of extra header records, one for each array, each of varying length. This choice is left to the implementer. The information embodied in the second Header Record is not actually used by any major program but is important as the stored definition of how the RGEF was created. The information contained in this record should be printed when the modified RGEF is printed by Program 3.

2.2.5 Write Modified RGEF Records

Input

Collect emissions data from 2.2.4.

Output

Modified RGEF Records.

Transformation

The data is formatted and written. The format needed for modified RGEF data may be slightly different from the Grid Cell Process Record format used in the basic RGEF.

2.2.6 Add Specified New Sources

Input

User Control Parameters.

Output

Modified RGEF Records.

Transformation

This module is used to add projected pollution sources to a modified RGEF. The user control parameters look like

$$\text{ADD } \left\{ \begin{array}{l} \text{POINT} \\ \text{AREA} \\ \text{TRAFFIC} \end{array} \right\}$$

followed by a list of all the fields in a typical record.

These are included with the regular user control parameters.

The record to be added is checked against the new modified RGEF to make sure it does not have the same key as an existing record. If it does, an error message is generated. If it does not, the new record is written.

2.2.7 Delete Specified Old Sources

Input

User Control Parameters
Modified RGEF

Output

Nothing.

Transformation

This module is used to delete sources whose demise is projected in a modified RGEF scenario. The user control parameters are

$$\text{DELETE } \left\{ \begin{array}{l} \text{POINT} \\ \text{AREA} \\ \text{TRAFFIC} \end{array} \right\}$$

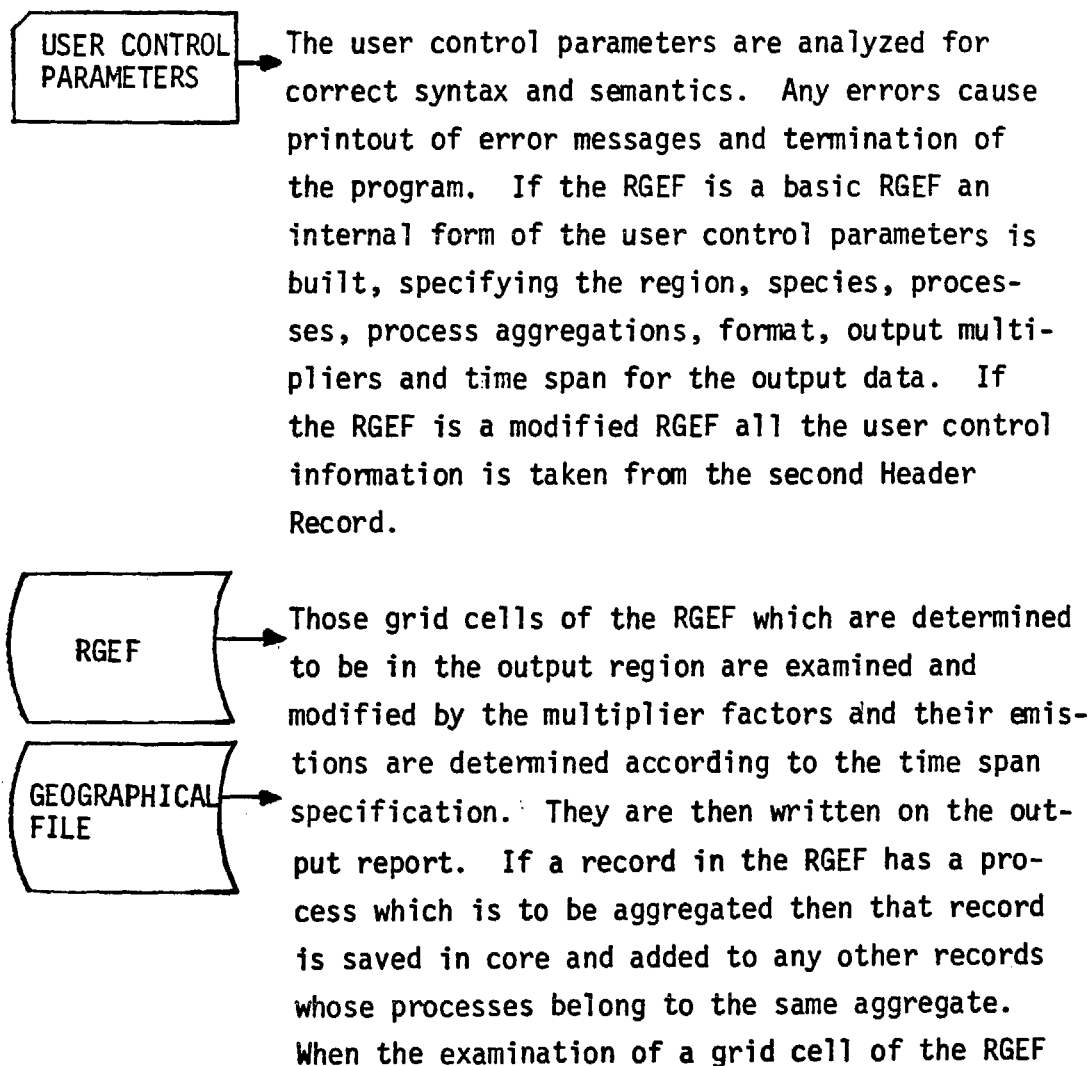
followed by a list of the key fields of the RGEF record. The process of the record cannot belong to an aggregate.

The key of the record to be deleted is checked against the new modified RGEF to make sure the record exists. If it does not, an error message is generated. If it does, the record is deleted.

3. OUTPUT SELECTED EMISSIONS DATA FROM RGEF

This function of the EMSIM system is the payoff of the entire system; it produces the tangible output requested by the user for inventory or modeling purposes. The user has a wide range of choices in specifying the type, amount, and format of output. The options available are expressed in the syntax of the User Control Parameters.

This function is a stand-alone program but it has many subroutines in common with 2.1. It can be used to print output from a basic RGEF or a modified RGEF. In the latter case the user does not have a wide choice of options; only a straightforward print is available.



is completed, the aggregated records are written on the output report. The geographical file may be used to determine which grid cells in the basic RGEF are relevant.

User Control Parameters for Producing Modified RGEF's or Printed Output from a Basic RGEF

In the following description underlined words represent default options, meaning that if the entire parameter is omitted, the underlined word is its understood value. Brackets [] enclose optional words. Elipses (three dots ...) mean the previous term may be repeated any number of times. Curly brackets { } indicate a choice of options.

$$\text{REGION} = \begin{cases} \text{COUNTY} & \text{county-number} \\ \text{BASIN} & \text{air-basin-number} \\ \text{SECTION} & \text{county-number air-basin-number} \\ \text{AREA} & \text{area-id} \\ \underline{\text{ALL}} & \end{cases}$$

Semantics: REGION parameter specifies the subregion of the RGEF from which data will be extracted. County-numbers and air-basin-numbers are the standard codes. Area-id is the id used for a user-defined area in the Geographical File. ALL means data is to be extracted from the entire RGEF region. The specified REGION should be checked to make sure it is contained in the basic RGEF region.

$$\text{SPECIES} = \begin{cases} \text{SPECIES}_1 [, \text{SPECIES}_2] \dots \\ \underline{\text{ALL}} \end{cases}$$

Semantics: SPECIES parameter specifies those pollutant species which are to be extracted from the RGEF records for output or inclusion in a modified RGEF record.

species_i is a standard pollutant code.

Example: SPECIES = 42101, 42401, 42602, 43101, 43201

$$\text{PROCESSES} = \begin{cases} [\text{PROCESS-OR-AGGREGATE},] \dots \begin{bmatrix} \text{all} \\ \text{other} \end{bmatrix} \\ \underline{\text{ALL}} \end{cases}$$

where

process-or-aggregate means process-value or aggregate-value

where

process-value means major-process-value or minor-process-value or
major-process-value: minor-process-value

where

major-process-value means $\begin{Bmatrix} \text{SIC} \\ \text{CAC} \end{Bmatrix}$ code-number

minor-process-value means $\begin{Bmatrix} \text{SCC} \\ \text{PC} \end{Bmatrix}$ code-number

aggregate-value means aggregate-name ([MAJOR-PROCESS-VALUE,] ... :
[MINOR-PROCESS-VALUE,] ...)

Semantics: PROCESSES parameter specifies those processes which are to be extracted from the RGEF records for output or inclusion in a modified RGEF record. The parameter is complicated since a point source process or area source process is identified by at least two different types of process code (and the coding for traffic sources is unknown). The EMSIM system is intended to allow the user to specify a process by either of its codes or by both.

The PROCESS parameter also allows the definition of aggregate processes. These are new processes created by adding up the emissions in the component processes.

Aggregate-name is an alphanumeric name of up to 20 characters to the newly-defined aggregate. This name will appear on any output which involves the aggregate process.

Code-number must be a number of digits equal to the length of the type of process code being defined (i.e. 8 for SCC, 8 for PC, etc.) Any of these digits may be an asterisk (*) which denotes an implicit aggregation. A code-number which has any asterisks denotes an aggregate process

containing all processes whose code numbers match its non-asterisk digits. Semantic checking on asterisks may be more precise. Certainly only the initial or terminal positions in the code-number may be asterisks (not the middle positions alone). Additionally, certain positions in the code-number may be meaningless if asterisked and in these places an asterisk should be prohibited.

The default option is ALL, which means extract all processes which are found in the basic RGEF. If ALL appears in the first format, following a list of process-or-aggregate entries, it means extract separately all the other processes of the basic RGEF which were not explicitly listed. If OTHER appears, it means extract all the other process of the basic RGEF which were not explicitly listed and aggregate them into a special category called OTHER.

An aggregate process must contain process of only one type (point, area, on traffic). These processes can be identified by their major-process-value, their minor-process-value or both. For example, the following aggregate definitions are all legal:

```
PESTICIDE-AGG (CAC 3000, 3101, 3102, 3109:)
ON-ROAD-VEHICLE (: PC P1100104, P1110104, P1111104, P1112104)
PESTICIDE-AGG2 (CAC 3000, 3101, 3102, 3109: PC P3572500, P3572500)
```

In the last example, the aggregate named PESTICIDE-AGG2 includes all area sources whose category-activity code is among those listed and whose process code is among those listed. In more abstract terms $\text{agname1}(\text{CAC } x_1, x_2: \text{PC } x_3, x_4)$ includes all area sources whose category-activity code is x_1 or x_2 and whose process code is x_3 or x_4 .

Examples:

```
PROCESSES = PC P3510302, PC P3510301, CAC 8301, SIC 2951:
            SCC 30500201,
            SCC 101001**, SCC 101002**, AGG1-MINING (SIC
            1100, 1400, 1000:), OTHER
PROCESSES = ALL
PROCESSES = CAC 3101 : PC P3572500, OTHER
```


$$\text{FORMAT} = \left\{ \begin{array}{l} \text{GRIDDED} \\ \text{REGIONAL} \end{array} \left\{ \begin{array}{l} \text{points} \\ \text{processes} \\ \text{totals} \end{array} \right\} \right\} \left\{ \begin{array}{l} \text{keys} \\ \text{emissions} \\ \text{all} \end{array} \right\}$$

Semantics: The FORMAT is only used for printed output. It is not allowed in the creation of a modified RGEF as it would be meaningless. This parameter specifies the format and level of summarization and detail in the printed output.

GRIDDED format means the output is separated according to grid cells. For GRIDDED format, POINTS means a separate listing is made for each RGEF record (except for aggregate processes). Each RGEF record represents a separate point (for point sources) which may have the same process number. POINTS means list each record separately, while PROCESSES means summarize all records with the same process numbers (SIC, SCC, PC and CAC). TOTALS means print only one set of emissions totals summarizing all the records in a grid cell.

REGIONAL format means the output is not separated by grid cells but listed all together for the entire file. In this case PROCESSES means a single entry is made summarizing all records in the file with the same process numbers. TOTALS means only one set of emissions totals is printed summarizing all the records in the RGEF.

Exactly what is printed is determined by the next option. EMISSIONS means to print the emissions totals for those pollutants specified in the SPECIES parameter. KEYS means to print the information in each RGEF record needed to find the key of the CEF record where additional background information is stored, the "referback" information. The KEYS option can only be used in conjunction with the POINTS option, since any aggregation beyond the point level loses some referback information. ALL means to print the emissions and the keys if applicable. It is the default option.

$$\text{MULT} \left(\left[\begin{array}{c} \{ \text{CAC} \} \\ \{ \text{SIC} \} \end{array} \right] \text{code} \left[\begin{array}{c} \{ \text{PC} \} \\ \{ \text{SCC} \} \end{array} \right] \text{code} , \left\{ \begin{array}{c} \text{all} \\ \text{species-number} \end{array} \right\} , \text{factor} \right)$$

Semantics: The MULT parameter (short for multiplier) is used to construct the filter of output multipliers, also known as adjustment factors. This parameter may be repeated many times to allow the entry of different factors for particular process-species combinations. The factors are used for projection and prediction purposes, either in the creation of printed output or in the creation of a modified RGEF. The factor is a floating point number used to multiply the specified emissions species of a specified process.

Example: MULT (CAC 8301 : PC P1322109, ALL, .5)
MULT (SIC 2951 : SCC 30500201, 11101, 1.9)

$$\text{TIME} = \left\{ \begin{array}{ll} \text{SNAPSHOT,} & \text{YEAR} \\ & \left\{ \begin{array}{l} 1 \\ 2 \\ 3 \\ 4 \end{array} \right\} \\ \text{QUARTER,} & \\ \text{MONTH,} & \text{month} \\ \text{DAY,} & \left\{ \begin{array}{l} \text{average} \\ \text{worst} \end{array} \right\} \\ & \text{DATE, date} \\ \text{ALL} & \end{array} \right.$$

Semantics: The TIME parameter specifies the temporal span of the output emissions data or the modified RGEF data.

The temporal component of the emissions data is a complicated phenomenon because different information exists for each type of source (point, area, and traffic) and it is necessary for programs to extrapolate the available information to produce the desired time frame. All this calculation should be invisible to the user.

The HOURLY option gives emissions for every hour of a particular day separately. Its use is in modeling applications. The SNAPSHOT option gives the emissions over one span of time, from as short as a day to as long as the entire year which the RGEF represents. The option ALL is like a dump option. For each selected record it prints out all the temporal information contained in the record. This information will differ according to the type of source, as already noted. The ALL option can only be used together with the POINTS option in FORMAT and if there are no aggregates under PROCESSES.

User Control Parameters for Producing Printed Output from a Modified RGEF

The user is allowed very few options for this printout compared to the many options allowed when printing a basic RGEF. The reason is that all the options are allowed in the creation of the modified RGEF and there would be too many possibilities of conflict if they were allowed again at the output stage. For instance, the specified REGIONS could be incompatible; multiple aggregation of PROCESSES could lead to all sorts of inconsistencies; TIME spans could be inconsistent. It is possible to insert complex checking procedures to insure against all such inconsistencies, but the benefits probably do not justify the added effort.

$$\text{FORMAT} = \left\{ \begin{array}{c} \text{GRIDDED} \\ \text{REGIONAL} \end{array} \right\} \left\{ \begin{array}{c} \text{PROCESSES} \\ \text{TOTALS} \end{array} \right\}$$

The modified RGEF is created at the PROCESSES level of summarization and therefore the POINTS and KEYS options cannot be used. They are generally not of interest in forecasting applications in any case.

3.1 and

2.2.1 Analyze User Control Parameters

This module is used in both (3) and (2.2) as a subroutine.

Input

User Control Parameters RGEF

Output

Error Messages

Transformation

The header record of the RGEF is checked to see if it is a basic or modified RGEF. The syntax and semantics of the User Control Parameters are then analyzed accordingly. (A detailed description of the syntax and semantics requirements is found on pages 93 to 98.) If there are any errors, the error messages are printed and the entire program terminates.

3.2 and

2.2.2 Establish Method of Collection

Input

User control parameters

RGEF

Geographical File

Output

Internal form of user control parameters

Transformation

Before data can be properly transferred from RGEF to printed output or modified RGEF, the definition of what is to be extracted must be translated from the user parameter syntax to internal data structures more easily manipulated by the program. This is done with the REGION, SPECIES, PROCESSES, MULT, and TIME parameters.

If a modified RGEF is being printed, the REGION, SPECIES, and PROCESSES parameters are assumed to equal ALL; there are no MULT parameters, and TIME is whatever time span was used in creating the modified RGEF (found in its header records).

3.2.1 and

2.2.2.1 Convert REGION to Grid Cells

Input

REGION parameter

Geographical File

RGEF header record

Output

A list of cell numbers

The number of cells in the list

Transformation

If the REGION parameter is COUNTY or BASIN, just find what sections make up the given county or basin, then consult the RGEF header record to find all the cells in these sections.

If the REGION parameter is SECTION, just consult the RGEF header record to find all the cells in the given section.

If the REGION parameter is ALL or if the RGEF being printed is a modified RGEF consult the RGEF header record to find all the cells in all the sections of the entire region.

If the REGION parameter is AREA, consult geographical file to find the record with the given area-id as its key. If there is no record, print an error message and terminate the entire program. Otherwise extract the polygon representation of the area, then use 3.2.1.1, the "point-in-polygon" algorithm to find which grid cells have centroids inside the polygon.

3.2.2 and

2.2.2.2 Determine Desired Emissions Species

Input

SPECIES parameter

RGEF header records

OUTPUT

SPECIES array

Transformation

Since there are not many pollutant species in all, it is possible to keep all possible codes in an array together with a flag for each indicating if it is to be included or not.

If the SPECIES parameter is ALL, turn on all the flags. If there is a list of species codes, compare each with the pollutant codes in the array and turn on the appropriate flags. If any species code is not in the array of pollutant codes print an error message and terminate the entire program.

If the RGEF is a modified RGEF to be printed, consult its modified header records to find which species flags to turn on.

3.2.3 and

2.2.2.3 Determine Desired Process Categories and Aggregations

Input

PROCESSES parameter

RGEF header records

Output

PROCESSES data structure

AGGREGATES data structure

Transformation

The internal data structures which contain process information consist of 3 arrays which will be called here PROCESSES, AGGREGATES, and VALUES. The actual code values are contained in the array VALUES, but the structural information is carried in the other two arrays, which point to VALUES.

The PROCESS array contains in each entry a source-type (point, area or traffic), a switch indicating major-processes-value, minor-process-value or both, and a pointer to the array VALUES, which may just be a number representing the index in the array VALUES where the code values begin.

Example: Suppose we have

PROCESSES = SIC x : SCC y

where x and y are point source codes. Then the data structure would look like

PROCESSES			VALUES		
Source Type	Major Minor Both	Pointer	Code Number	Number of Asterisks	Asterisks First, Last or More
			x	0	NONE
POINT	BOTH		y	0	NONE

The above example shows the structure of the VALUES array. The second two fields in the array are for easy handling of a code-number with asterisks. Suppose a code number is written by the user as

PROCESSES = ..., SCC 305001**

Then it would appear as

PROCESSES			VALUES		
Source Type	Major Minor Both	Pointer	Code Number	Number of Asterisks	Asterisks First, Last or More
POINT	MINOR				
			305001	2	LAST

The AGGREGATES array also points to the VALUES table. Each entry contains an aggregate-name (up to 20 characters), source type (which by definition must be the same for all processes in the aggregate), number of major-values, number of minor-values, and a pointer to the VALUES array. Thus if the user enters

PROCESSES = ..., AGNAME1 (CAC x_1, x_2, x_3 : PC y_1, y_2), ...

the aggregate is stored internally as

AGGREGATES					VALUES		
Aggregate Name	Source Type	No. of Major Values	No. of Minor Values	Pointer	Code Number	Number of Asterisks	Asterisks First, Last or More
AGNAME 1	AREA	3	2	→	x_1	0	NONE
					x_2	0	NONE
					x_3	0	NONE
					y_1	0	NONE
					y_2	0	NONE

It must be remembered that the three arrays just described will not be very big as they are used only to describe the user control parameters. If the user specifies the options ALL or OTHER, these will be set as separate switches, not included in the above arrays.

Note that the array structure just described also gives the program a shorthand internal code for dealing with any aggregate or process--namely its index in the AGGREGATE or PROCESS array.

If the RGEF is a modified RGEF to be printed, the three arrays are extracted directly from the RGEF header record rather than computed from the user control parameters.

3.2.4 and

2.2.2.4 Construct Filter of Output Multipliers

Input

MULT parameters

Internal arrays: PROCESSES, AGGREGATES, VALUES

Output

Filter of output multipliers.

Transformation

The filter of output multipliers, also known as adjustment factors, is used to alter the emissions of an RGEF in order to create a modified RGEF or printed output which embodies a predictive scenario. The filter can be thought of conceptually as a two dimensional array of floating point (unitless) numbers, where the first coordinate is determined by process codes and the second coordinate is determined by emissions species. If the PROCESSES parameter contains any aggregate definitions, these can be used as process codes in the filter, thus allowing a multiplier to apply at once to a large number of processes. In fact, the processes making up an aggregate may not themselves be used in the filter; only the aggregate is allowed to be multiplied, otherwise there would be consistency problems. The filter just described allows great flexibility in specifying adjustment factors, down to the individual process and species level.

If all possible processes were included in the filter, it would be quite large. So it is implemented as follows:

the array could contain a slot for each one, in the same order as the SPECIES array. Or it could contain just those species whose flags have been turned on in the SPECIES array. The internal code can be the ordering of the SPECIES array. In fact, possibly the entire EMSIM system could use a simplified coding for pollutants. There is no need to burden the user with 5-digit pollutant codes when there are only about 12 pollutants of interest and certainly not more than 40. It is also possible to use short alphanumeric codes for pollutants, like CO, SOX, NO2, etc. These would improve the readability of the user control parameters.

For process codes, unfortunately we cannot use the internal codes of the PROCESS array since this array might not specify all processes involved in the RGEF. So the filter should contain for each process code a flag indicating if the code is a point, area, traffic, or aggregate code. In the first two cases the 2 code numbers of the process are necessary. In the aggregate case, the internal code from the AGGREGATE array will be used.

When all the MULT parameters have been processed, the array should be sorted by process type and process code.

3.2.5 and

2.2.2.5 Determine Desired Temporal Factors and Aggregations

Input

TIME parameter

Output

Internal form of TIME parameter.

Transformation

The actual transformation of this module is straightforward. It involves setting a number of switches to indicate SNAPSHOT vs. HOURLY, YEAR vs. QUARTER vs. MONTH vs. DAY, etc.

It is the interpretation of the time span that will be laborious, actually part of the work of module 3.3.1 (and 2.2.3.1) which collects the emissions data. Unless the temporal information on the source files is standardized, each different TIME option will require a separate algorithm to compute its estimated value.

3.3 and

2.2.4. Collect Data from RGEF

Input

Internal Form of User Control Parameters

RGEF

Output

Data for forming into Printed output or modified
RGEF records

Transformation

This module is the heart of the main program (3) or (2.2), the actual extraction of desired data.

First the list of grid cells for the extraction region (produced by 3.2.1) is consulted. For each grid cell, all the RGEF records corresponding to that cell are read sequentially, and processed one at a time. For each record, its emissions are computed according to the TIME parameter specification. Its emissions are then multiplied by the output filter multipliers (produced by 3.2.4). If the process of the record is contained in an aggregate, the record's emissions are stored in a temporary array called TEMP for aggregation of output. This also occurs if the PROCESSES or TOTALS option is used in the FORMAT command. The emissions in the TEMP array are sent to the formatting routines when the processing of the grid cell (or the entire extraction region) is completed (depending on whether the FORMAT parameter is GRIDDED or REGIONAL). Otherwise, if the process of the record is not contained in an aggregate and the format option is POINTS, it is sent directly to the formatting routine.

3.3.2

2.2.4.2 Summarize Data by Specified Aggregations

Input

Emissions Data

TEMP array

Output

Same

If the process codes of an RGEF record being examined are part of an aggregate process (checked by consulting AGGREGATES and VALUES arrays) then the emissions data are stored in the TEMP array for temporary aggregation until processing of the grid cell or region is complete. The TEMP array is also used if the FORMAT option is PROCESSES or TOTALS. The TEMP array looks like

TEMP				
Process Codes	Species 1	Species 2	...	Species n
c_1			...	
c_2			...	
\vdots			...	

The representation of process codes and pollutant species are the same as in the Filter of output Multipliers array discussed in Section 3.2.4.

To store emissions in the TEMP array, the array is scanned to see if any existing process codes match those of the current record. If there is a match, the emissions of the current record are added to those in the matching row of the TEMP array. Otherwise a new row is created in the TEMP array.

3.3.3 and

2.2.4.3 Multiply Data by Multipliers in Output Filter

Input

Emissions Data

Output

Emissions Data

The process codes of the current RGEF record is checked to see if they belong to an aggregate. (This function can be combined in 3.3.2 and 3.3.3; it need not be repeated). Then the process codes are checked with the filter of output multipliers. If there is a match, the multiplications are performed on the emissions data. Otherwise the emissions data are unchanged.

3.3.4 Write File of Keys for Later Background Extraction from CEF

Input

RGEF

Output

Keys

Transformation

This module is used if the user control parameter `FORMAT = ... POINTS KEYS` or `FORMAT = ... POINTS ALL` is present. It cannot be used if `FORMAT = ... PROCESSES ...` or `FORMAT = ... TOTALS ...` as aggregation causes loss of at least some keys. For each record selected by the data collection routine as satisfying the user control parameters, its backward-reference information is written on the output report or on a separate file. The "keys" are not the keys of the RGEF records themselves. They are the information in the RGEF records which can act as keys to the CEF file.

3.4 Format Data

Input

Processed Emissions Data, and Key Information

Output

Formatted output file

Transformation

The raw information provided by 3.3 is formatted for printing. As much as possible, intelligible headings should be used, e.g., the pollutant names should be written out, not just their codes. This is probably not possible with processes, but should be done with aggregates, whose alphanumeric names are available to the system.

Following are some samples of printed reports from output programs 3.4.1 through 3.4.2.2.

Since data in the RGEF could be extracted and utilized in many different ways it is difficult to give, at this stage, all final report formats. As the system is being developed and specific user needs identified, new reports could be formatted and programmed. This task will not be difficult to the systems analyst due to the well organized structure of the RGEF files and due to the flexibility of the VSAM software.

EMISSIONS SIMULATOR

PAGE XXX

DATE: MM/DD/YY

RGEF ID


HEADER RECORD

THE ENTIRE HEADER RECORD

SECOND HEADER RECORD

THE ENTIRE SECOND HEADER RECORD

USER CONTROL PARAMETERS

THE USER CONTROL PARAMETERS SPECIFIED
FOR THIS PRINTOUT (SEE PAGES 93 TO 98)

EMISSIONS

PAGE XXX

GRID CELL

UTM COORDINATES SE CORNER:

HORIZONTAL XXXX

VERTICAL XXXX

POLLUTANT XX XX XX XX XX XX

POINT SOURCES

ID (if any) SIC,IPP,SCC or aggregate code

emissions

.
.
.
.
.
.AREA SOURCES

ID (if any) category or process code

emissions

.
.
.
.
.TRAFFIC PROCESS

Process code

emissions

.
.
.
.

GRID CELL TOTALS

XXX XXX XXX XXX XXX XXX

PERCENT OF TOTAL REGIONAL EMISSIONS

XXX XXX XXX XXX XXX XXX

For some summary type printouts the emissions could be listed as:

GRID CELL	XX	UTM COORDINATES OF SE CORNER			
		HORIZONTAL		XXXX	
		VERTICAL		XXXX	
POLLUTANTS		XXX	XXX	XXX	XXX

POINT SOURCE TOTALS BY PROCESS

PROCESS NAME	CODE	emissions
.		
.		
.		
.		

AREA SOURCE TOTALS BY PROCESS

CATEGORY/ACTIVITY	CODE	emissions
or PROCESS NAME	CODE	
.		
.		
.		
.		

TRAFFIC SOURCE TOTALS

emissions

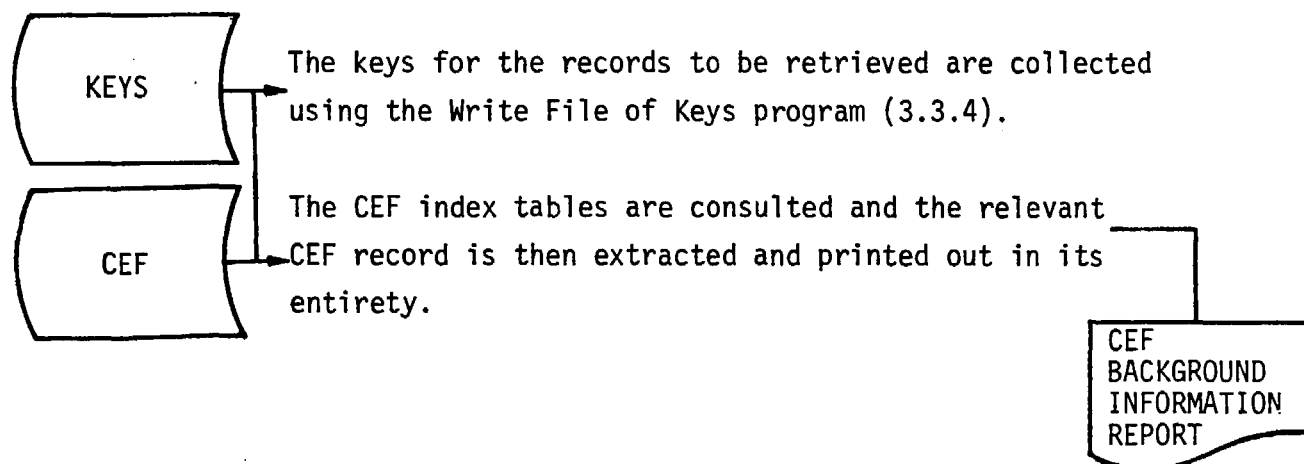
GRID CELL TOTALS	XXXX	XXXX	XXXX	XXXX
PERCENT OF TOTAL REGIONAL EMISSIONS	XX	XX	XX	XX

Tables for percent contribution of each grid cell would be formatted.

GRID CELL XX	POLLUTANTS	XX	XX	XX	XX	XX
1		XX	XX	XX	XX	XX
2						
.	Total emissions	{				
.	per grid cell					
.	or percent of					
.	total REGIONAL					
.	emissions					

4. REFERENCE CEF FOR BACKGROUND INFORMATION

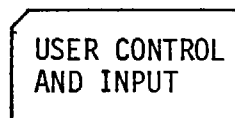
The function of this program is to retrieve detailed background information from the CEF. The user might need information about a selected number of sources that is not carried in the basic RGEFs (e.g., plant address).



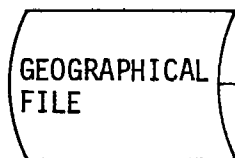
This program is a straightforward "read from VSAM" processing and therefore does not require a more detailed breakdown for the system implementor.

5. PRODUCE GEOGRAPHICAL FILE

The function of this program is to build, update, and display a file of geographical regions which are expressed as closed polygons.

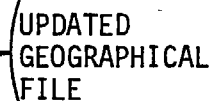


The user specifies the desired action (create, delete, or report) via cards, data set, or on-line terminal. The user also inputs the vertices of the polygon if needed.



The existing geographical file is needed to check correctness of create and delete commands. E.g., cannot delete a record that does not exist.

For each new area created, program computes its circumscribing rectangle, that is, its maximum and minimum coordinates. The geographical file emerges updated.



A report is produced detailing the transactions which occurred and any error messages. Since a number of transactions will occur in any job, an error does not cause job termination, only the printing of a message.



Optionally, if the user specifies the "report" command in the input, a printout of the entire geographical file will be included in the report.

GEOGRAPHICAL FILE DEFINITION

Name: Geographical File

Purpose: To hold descriptions of counties, air basins, and user-defined areas for use by gridding and displaying programs. EMSIM contains exactly one geographical file.

Origin: Produce Geographical File Program

Destination: Produce RGEF
Produce Modified RGEFs
Output Selected Emissions Data

Format: Variable length records

File organization: Key sequenced

Access method: VSAM or ISAM

Medium: Disc

Estimated volume: Several hundred areas

Record name: Area records

GEOGRAPHICAL FILE

Area Record Description

Field Name:

Description

KEY FIELDS

Area type

Size: 1 Byte, numeric

Value: 1 means record represents a county

2 means record represents an air basin

3 means record represents a county-air
basin section4 means record represents a user-defined
area

Area id

Size: 4 Bytes, packed decimal

Value: A number with maximum of 4 digits

NON-KEY FIELDS

Area name

Size: 30 Bytes, alphanumeric

Value: Description of the area

Area owner

Size: 30 Bytes, alphanumeric

Value: Name of individual or group responsible
for this area entry

Max x-coordinate

Max y-coordinate

Min x-coordinate

Min y-coordinate

Number of vertices

x-coordinate

Repeated
for each
vertex.

y-coordinate

Maximum
is

zone

Size: All 4 items have 4 Bytes, fixed
binary with 1 virtual decimal
place.

Size: 2 Bytes, binary

Size: 4 Bytes, fixed binary with 1 (virtual)
decimal place

Size: Same

Size: 1 Byte, binary

GEOGRAPHICAL FILE

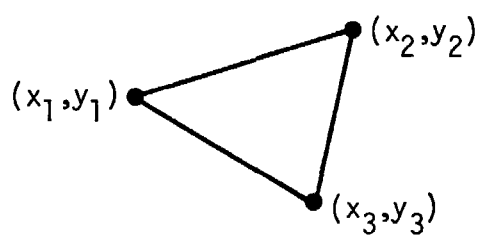
Input Format

Input to the Produce Geographical File Program is a file which may be created through cards or written on-line. The file consists of any number of transactions, which are processed one at a time. Each transaction begins on a new line with a slash followed by a command. Some transactions are followed by additional line(s) specifying vertices of a polygon.

Each area has an id number, by which it will be referenced throughout the system, and a name, which is just used for reporting on this file as a reminder of which area is which. Each area also has an area-type which can be county, air-basin, section, or user-defined. If the area is a county or air basin, its id number must be identical to the standard predefined county or air basin number. If the area is a county-air basin section ("section" for short), then its id number is county id followed by air basin id.

Security considerations should prevent the casual user from deleting counties, air basins, or those user-defined areas belonging to other user groups.

Example: a user defined area



```
/CREATE AREA 11 "PRACTICE-AREA"
```

```
x1    y1
```

```
x2          y2    x3    y3
```

GEOGRAPHICAL FILE

Commands

$$\text{/CREATE } \left\{ \begin{array}{l} \text{SECTION} \\ \text{COUNTY} \\ \text{BASIN} \\ \text{AREA} \end{array} \right\} \text{ area-id area-name [user-group-name]}$$

This command causes a new area to be added to the geographical file. Only a special systems user can create COUNTY, SECTION, or BASIN areas. These area-types should be created when the file is initiated and will remain fixed thereafter. An AREA area means a user-defined area. The user must supply an area-id (up to 4 digits) and an area-name (up to 30 characters). The area-id cannot already exist on the file unless it is associated with a different area type. If it does already exist, an error message is returned. The user-group-name is an optional security feature to show which user group has responsibility for a given area. Only the identified user group (or special systems user) can delete its own areas.

The CREATE command can be abbreviated by its first letter, C; similarly with area-types.

The CREATE command is followed by a new line or lines containing the coordinates of the vertices of the polygon which approximately encloses the area. The vertices are assumed to be in sequential order going around the edge of the polygon. It is assumed that there is an edge between each vertex and the next vertex, and between the last vertex and the first vertex.

Each vertex is represented by its UTM x-coordinate (horizontal) and its UTM y-coordinate (vertical). Each coordinate is separated by spaces but number of spaces is free form. The user may choose to enter just one x-coordinate and one y-coordinate per line. X-coordinates may have up to 3 places to the left of the decimal point and one place to the right. Y-coordinates may have up to 3 places to the left of the decimal point and 1 place to the right.

/DELETE $\left\{ \begin{array}{l} \text{SECTION} \\ \text{COUNTY} \\ \text{BASIN} \\ \text{AREA} \end{array} \right\}$ area-id [user-group-name]

This command causes the deletion of an area already existing in the geographical file. Only a special systems user can delete COUNTY, SECTION, or BASIN areas. AREA areas (user-defined) can only be deleted by the user-group that created them, or by a special system user. The area-id must exist on the geographical file; if not, an error message is returned.

/REPORT [ID] [OWNER]

This command causes a listing of the entire geographical file to be printed out. If ID is not specified, the file is printed out sorted by area-name (alphabetic sort). If ID is specified, the file is printed out sorted by area-type, then by area-id within area-type (the natural key ordering of the file). If OWNER is specified, the owner-group of each area is listed; this option is only available to the special systems user.

In order to sort the file by area-name, the best way for the program to proceed is to move the area-name, area-type, and area-id of every record into a core array, then sort them by area-name, then use area-type and area-id as key to reaccess each record for printout in the desired order.

GEOGRAPHICAL FILE

Report Format

EMISSIONS SIMULATOR
GEOGRAPHICAL FILE REPORT

DATE: MM/DD/YY

PAGE XXX

TRANSACTION SUMMARY:

for each CREATE transaction

$\left\{ \begin{array}{l} \text{SECTION} \\ \text{COUNTY} \\ \text{BASIN} \\ \text{AREA} \end{array} \right\}$	area-id	$\left\{ \begin{array}{l} \text{CREATED.} \\ \text{ALREADY EXISTS. TRANSACTION IGNORED.***} \end{array} \right.$
--	---------	--

for each DELETE transaction

$\left\{ \begin{array}{l} \text{SECTION} \\ \text{COUNTY} \\ \text{BASIN} \\ \text{AREA} \end{array} \right\}$	area-id	$\left\{ \begin{array}{l} \text{DELETED.} \\ \text{NOT FOUND IN FILE. TRANSACTION IGNORED.***} \end{array} \right.$
--	---------	---

for REPORT transaction, a new page is turned after the transaction summary
and the contents of the file are listed in the following form:

GEOGRAPHICAL FILE CONTENTS

AREA NAME	AREA TYPE	AREA ID
-----------	-----------	---------

x(30)...	x(7)	x(8)
----------	------	------

$\left. \begin{array}{l} \text{000000.000.9 00000.9 00000.000.9 00000.9 etc.} \\ \text{000000.000.9 00000.9 00000.000.9 00000.9 etc.} \end{array} \right\}$	(Listing of coordinate pairs)
---	-------------------------------

⋮		
x(30)	x(7)	x(8)

⋮

LIST OF ABBREVIATIONS

AQCR	Air quality control region
AQMA	Air quality maintenance area
CARB	California Air Resources Board
CEF	Crude Emissions File
EI	Emissions inventories
EIS	Emissions inventory system
EMSIM	The Metropolitan Emission Simulator System
JCL	Job Control Language (IBM)
PL/I	A high level programming language currently available at the Teale Data Center
RGEF	Regional Gridded Emissions File
SAI	Systems Applications, Incorporated
UTM	UNIVERSAL TRANSVERSE MERCATOR--a geographical coordinate system
VSAM	Virtual Storage Access Method--an IBM file access software package

APPENDIX
FORECASTING TECHNIQUES FOR THE EMISSION SIMULATOR

APPENDIX

FORECASTING TECHNIQUES FOR THE EMISSION SIMULATOR

In this appendix we present our review of forecasting techniques and outline what we consider to be a reasonable forecasting tool for incorporation into future emission simulator systems.

Forecasting methodologies for the spatial scales of interest, (the size of a state or a country) are very similar and are based on forecasting techniques outlined in the EPA document "Guidelines for Air Quality Maintenance Planning and Analysis." For stationary sources other than power plants and other large facilities the basic approach is to forecast emissions at the level of a Standard Metropolitan Statistical Area (SMSA) using OBERS economic projections. An SMSA normally comprises an area of a few counties. Once projections are made to this level the emissions are then distributed based on other factors such as land use planning, population distribution, or other factors which in some way can be associated with the location of stationary sources. Power plants and other major point sources are normally forecast based on individual projections of energy needs for an area and are not normally based on economic factors. Traffic models are normally based on population distribution and the location of industries or businesses that attract traffic. An important aspect of traffic models is the controls and regulations that are placed on new vehicles or retrofitted to old vehicles. In the following section, we briefly summarize the methodologies that are reviewed in this study.

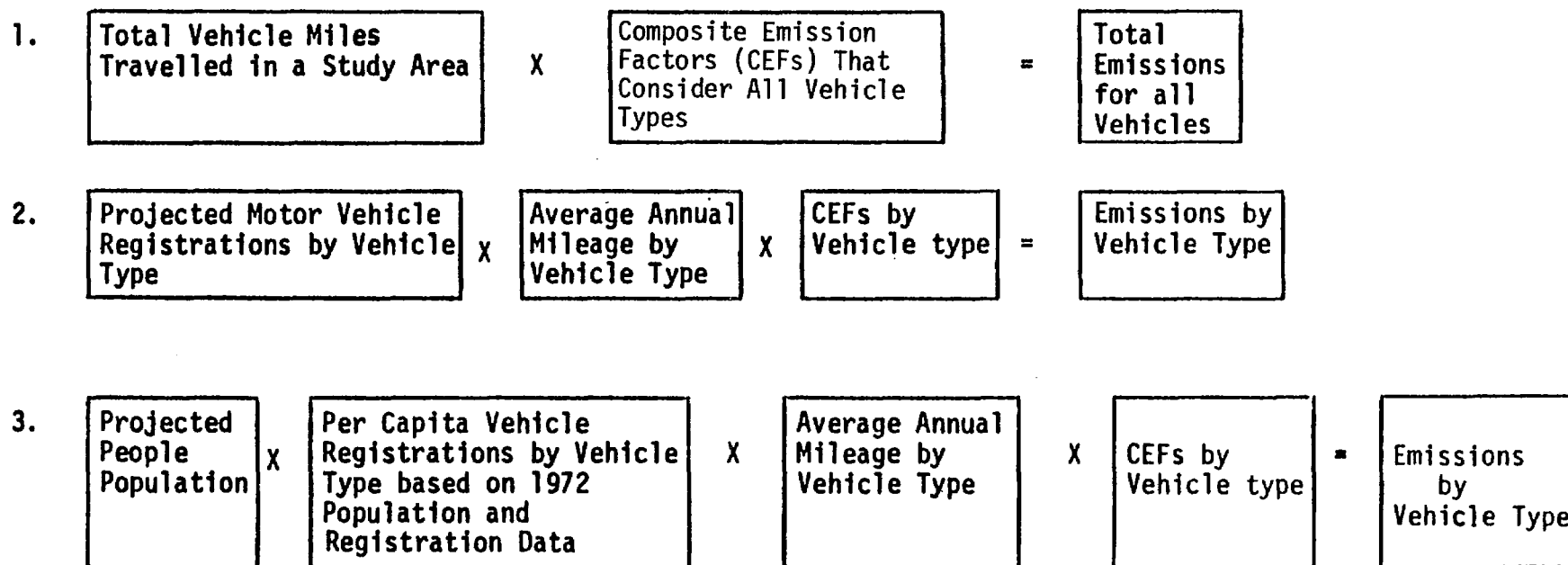
A. REVIEW OF FORECASTING METHODS

California Air Resources Board (1974). In a document prepared in 1974, the State of California Air Resources Board discusses emissions forecasting methodologies for on-road motor vehicles, aircraft, off-road

mobile sources, trains, ships, fossil fuel power plants, and other stationary sources. Three approaches are considered for estimating emissions from on-road motor vehicles. These approaches are summarized in Figure A-1. In the first approach, the total vehicle miles travelled (VMT) for the study area are used in conjunction with a composite emission factor representative of total vehicle activity. The VMT estimates for a study area are developed for transportation studies using data from origin-destination surveys or transportation models. The composite emission factors (CEFs) representative of total vehicular activity are estimated from past or anticipated distributions of VMT by vehicle type and CEFs tabulated for each vehicle type. In the second approach, data compiled by the California Department of Motor Vehicles are used to project motor vehicle registrations by county and vehicle type through 1985. In the third approach, the 1972 values for the average number of vehicles in use by county and county population data are used to calculate per capita vehicles in use by vehicle type. Future year population projections are then used with the per capita values to estimate the number of future year vehicles.

For a given study area, the emissions from aircraft are a function of the number of airports, the type of aircraft using the airport, and the number of landings and takeoffs (LTOs) by different aircraft classes. Emissions factors per LTO cycle for future years have been estimated by CARB (1974) based on schedules for implementation of emission standards and anticipated aircraft replacement rates. If projections of the number of LTO for future years are not available, future year emissions can be projected using population growth factors and emissions contained in the 1972 ARB Emissions Inventory.

For off-road vehicles, trains, and ships, future year emissions were estimated by multiplying the 1972 emissions from these sources by population growth factors.



Source: CARB (California Air Resources Board), 1974.

FIGURE A-1. THREE APPROACHES FOR ESTIMATING EMISSIONS FROM MOTOR VEHICLES

To project emissions from fossil fuel power plants, it is necessary to know the percentages of the production to be generated by gas, fuel oil, and coal. Furthermore, it is necessary to know how the use of these fuels will vary by season and what the expected sulfur content of the fuel used will be. If adequate information on future year power plant construction and operation is not available, per capita energy generation values and projected population for a study area can be used to estimate future power generation. To estimate emissions, the amount of power generated by the various energy sources must be determined in order to apply the appropriate emission factor. Per capita energy demand values were based on 5 scenarios developed by Rand (1972) and population projections of the California Department of Finance.

CARB (1974) assumed that the emissions from other stationary sources is a function of the industry type and the geographical location. Projections of total earnings were developed by the Bureau of Economic Analysis for Standard Metropolitan Statistical Areas throughout the nation. These projections are used as indications of growth for industrial activity and the associated emissions.

Jensen, Loscutt, and Manger (1976). The California Air Resources Board published a document in April 1976 in which a methodology was outlined for forecasting emissions in counties and AQMA's throughout the state. The methodology outlined differs very little from that discussed in the 1974 project.

Booz, Allen and Hamilton (1975). Four methodologies for projecting emissions were reviewed and compared based on projections made for the St. Louis Metropolitan Area. Two of the methodologies, the Regional Emissions Projection System (REPS) and the Plan Revision Management System (PRMS) are generally applicable to any geographical region within the United States. The other two methodologies, the Attainment Study, and the Trial

Air Quality Maintenance Plan (AQMP) were developed specifically for application to St. Louis. Table A-1 summarizes the key features of the four projection methodologies. Besides differing in the geographical regions to which they are applicable, the methodologies differ in the pollutants considered, the base year emissions inventory, and the projection years to which the projection models have been applied.

All four projection methodologies involve the following major steps:

- > Determination of the base year emissions inventory.
- > Determination of regional growth factors for future years that reflect the expected change in pollution-predicting activity.
- > Projection of present regional emissions inventories to future years using these growth factors.
- > Adjustment of the emission projections to include the effects of present and future control regulations.

Emissions projections for REPS, the Trial AQMP and the Attainment Study are given for the following source categories:

- > Power plants.
- > Other point sources (excluding power plants).
- > Highway vehicles (excluding diesel vehicles for REPS and the Attainment Study, including diesel vehicles for the Trial AQMP).
- > Other area sources (excluding highway vehicles).

The PRMS projections are given in a different category source format:

- > Fuel combustion
- > Industrial process
- > Solid waste
- > Transportation
- > Miscellaneous.

TABLE A-1. KEY FEATURES OF THE PROJECTION METHODOLOGIES

Feature	Projection Methodology			
	REPS	PRMS	The Attainment Study	Trial AQMP
Geographic Region	Any AQCR or the nation	Any geographic region	St. Louis AQCR	St. Louis AQMA
Pollutants Considered	HC CO TSP SO _x NO _x	HC CO TSP SO _x	HC CO	HC CO TSP SO _x
Base Year	1974*	1970 [†]	1972	Various years adjusted to 1975
Projection Years	1980*	Continuous to 1977 in semi-annual increments [†]	1975 1977 1980	1980 1985
Emission Projections	Yes	Yes	Yes	Yes
Air Quality Projections	No	No (procedures given but not numerical results)	Yes	Yes
Publication Date of Reference Report	May 1975	March 1974	March 1974	December 1974

* Base year and projection year for REPS are the option of the system user. The base year may be 1974 or any subsequent year; projections may be developed for any year between the base year and 2000.

[†] The PRMS methodology may be applied to any base year and projection period.

Source: Booz, Allen and Hamilton, 1975.

Table A-2 summarizes the sources of data for the base year emissions inventory, growth projections, and pollution abatement programs for the four methodologies. The regional growth factors used in REPS were based on economic and demographic forecast data including EPA-developed national economic growth projections and Department of Commerce regional activity projections. The SEAS and OBERS projections have been supplemented in REPS by a special analysis of growth and relocation trends for five industries that are among the heaviest industrial polluters. These critical industries are electric power generation, steel, chemicals, pulp manufacturing, and petroleum refining. The output of this analysis is a file of data on new plants expected to become operational in the future. The REPS system includes the effect of control regulations in two ways. First, if any point source has been granted a control variance that will have expired by the projection year, projected emissions are reduced to the level allowable under these regulations. Second, Federal New Source Performance Standards which govern new and retrofit industrial equipment are included in the REPS system. In general, future area source fuel combustion is estimated by projecting the BTU demand for each customer category and allocating that demand to the fuel mix for the projection year. The REPS system does not include emission controls for any area source category except gasoline highway vehicles.

The general projection method used for point sources in the PRMS computer program was as follows:

- > Identify the individual point sources that account for 90 percent of the point source emissions.
- > Project emissions for each of these sources using
 - Allowable emissions and compliance dates for the point source from NEPS.
 - Applicable local emission regulations.
 - A growth factor for the emission source category in which the source is included.

- > Project emissions for the remaining sources by aggregating emissions for each source category and applying the appropriate growth factor for that category.

The PRMS projection method used for area sources was as follows:

- > Aggregate emissions for each area source category from all jurisdictions in the AQCR.
- > Identify applicable local regulations that will effectively reduce projected area source emissions. The regulations that were used in the emission projections affected the following source categories and pollutants.
 - Cool combustion: TSP, SO_x
 - On-site incineration and open burning: all pollutants
 - Land vehicles: all pollutants.
- > Apply appropriate growth factors and control regulations to the geographically aggregated area source emissions to estimate future emissions.

The source of the growth factors was the OBERS projections with two exceptions. The growth factor for power plants was developed from published FPC data on new plants or modifications to existing plants. Growth was included for land vehicle transportation in adjusted transportation emission reduction curves from the Federal Motor Vehicle Emission Control Program.

Booz, Allen and Hamilton (1975) list the following capabilities, which are important for any projection methodology, which were included in at least one of the four methodologies:

- > Stationary source emissions should be estimated on an individual point source basis. This is critical because individual sources often have unique emission rates and compliance schedules. Emission projections in REPS and PRMS are developed on an individual source basis, while

in the Attainment Study and the Trial AQMP, emissions are forecast on a plant basis. For the latter approach plant-wide averages for growth and extent of control must be estimated for the various process and fuel combustion point sources within each plant.

- > A computerized methodology greatly facilitates individual source calculations, and allows alternative forecast scenarios to be evaluated quickly and efficiently. REPS is a computerized model; in addition to containing complete "default scenario" data, REPS has extensive capability for override of these default data with more accurate local information concerning emission rates, growth and compliance schedules for existing and proposed stationary sources.
- > All point and area sources must be identified by legislative jurisdiction. Computerized systems should be able to access all source data for each jurisdiction, in order to apply emission regulations for each jurisdiction to the appropriate sources. PRMS, the Attainment Study, and the Trial AQMP considered all emission sources by jurisdiction. Although only NEDS information and Federal New Source Performance Standards were input to the REPS projections presented in this report, the effect of other local emission control regulations may be input to REPS by the system user.
- > Local information and studies, such as plant survey data and detailed transportation studies, increase the accuracy of emissions (and air quality) forecasts. Plant surveys identify specific growth and expansion plans, compliance schedules and process changes. Transportation studies and forecasts define average speed, mileage, and other traffic characteristics which have a significant effect on emission and air quality estimates. Both the Attainment Study and the Trial AQMP utilized plant surveys and detailed transportation forecasts.

Goodrich (1973). An emissions projection methodology of general applicability has been discussed as part of the evaluation of alternative land use plans for the New Jersey Hackensack Meadowlands. The following five-step procedure was outlined for projecting emissions from the base-line year of 1969 to the year of 1990:

- > Step 1--activities. For all land use and transportation data, the level of activity is specified. The activity refers generally to the intensity of use--vehicles per day on a highway, acres of residential land use, square feet of industrial plant space.
- > Step 2--activity indices. For each category of activity, develop the parameters that translate the activity levels into demand for fuel for heating purposes--for instance, BTUs per square foot for industrial plant space.
- > Step 3--fuel use. For each category of activity (and geographical area), parameters for the propensity to use different fuels are applied to the fuel requirements.
- > Step 4--emission factors. For each category of activity, engineering estimates of fuel and process source emission factors are developed and applied to fuel use and process rates.
- > Step 5--emissions. Emissions calculated from fuel and process sources are adjusted for seasons, based on temperature variations and parameters representing the percent of fuel used for heating purposes.

The methodology used to get activity data for 1990 was extremely simplified. The following criteria were used:

- > Activity data for power plants and incinerators were calculated separately based on engineering calculations of projected growth.

- > No new point sources were considered outside the Meadowlands.
- > For existing point sources, regional and local projections of employment changes from the New Jersey Board of Labor and Industry and the Tri-State Transportation Commission were used to project forward employment for existing firms; this parameter was then used to project forward heating demand for 1990.
- > Changes in process emissions were made in consultation with the New Jersey Department of Environmental Protection on an industry-by-industry basis.
- > All activity data was obtained directly from the New Jersey Department of Transportation.
- > For area sources, regional and national fuel use projections, population, employment, and square feet of floor space data from the Tri-State Transportation Commission for 1985, together with parameters from the current inventory were used.

The activity indices were obtained from the following numbers:

- > BTUs per dwelling unit, square foot, or classroom.
- > The schedule.
- > Percentage of heat used for space heating versus process heating.
- > Relative propensity of fuel use for each of the categories.

Emissions factors for 1969 were taken from the Compilation of Emission Factors published by the Public Health Service. Emissions factors for automobiles were obtained from EPA. The projection of emissions factors was based on anticipated application of more restrictive emission controls. The methodology applied in this particular application is not well documented.

The baseline used for 1990 fuel consumption was the "Energy Model for the United States" prepared by the U.S. Department of the Interior in July

1968. Known cases of fuel switching currently being made were incorporated into the projection for 1990.

Mikolowsky, et al. (1974). In this study by the Rand Corporation, various levels of technological controls were examined as part of a study to look at projected air quality trends in the Los Angeles Basin. Emissions were projected rather simply by the use of the annual growth rates for motor vehicles and other sources shown in Tables A-3 and A-4.

TABLE A-3. ESTIMATED VMT: LOS ANGELES AIR QUALITY
CONTROL REGION
(Used to Calculate Average Daily Emissions)

Vehicle Class	Average VMT (mi/day)		Annual Increase in VMT, 1977 to 1990(a) (Percent)
	1970	1977	
Light-duty motor vehicles	128,275,600	148,856,200	1.8
Heavy-duty gasoline-powered trucks	4,468,400	4,933,100	1.8
Heavy-duty diesel-powered trucks	1,377,200	1,520,400	1.8
Diesel-powered intracity buses	206,400	261,900	1.8
Four-cycle motorcycles	1,538,600	1,785,500	4.6 (b)
Two-cycle motorcycles	943,100	1,028,500	-10 (b)

(a) Using current SCAG population growth forecasts and assuming that vehicle miles traveled grow at a rate that is 50 percent greater than the population growth; compounded annually.

(b) Assuming that new two-cycle motorcycles are substantially replaced by four-cycle motorcycles or that the emission characteristics of the new two-cycles become essentially the same as those of the four-cycles (EPA, 1974).

Source: Mikolowsky et al. (1974).

TABLE A-4. ANNUAL GROWTH RATES FOR EMISSION SOURCES OTHER THAN MOTOR VEHICLES

Source Type	Annual Growth Rate (percent), 1970 to 1990	Source Type	Annual Growth Rate (percent), 1970 to 1990
Petroleum industry	2.0	Incineration	1.2
Organic solvent users	1.2	Combustion of fuels	1.2
Chemical industry	1.6	Agriculture	-1.1
Metallurgical industry	1.6	Aircraft operation	1.4
Mineral industry	-0.4	Ships and railroads	1.4

Source: Hikołowsky et al. (1974).

In computing the average emissions per mile from motor vehicles, Rand considered, in addition to the VMT, the following factors:

- > Emission factors by vehicle model year.
- > Emission factors deterioration functions by vehicle age and model year.
- > Average speed correction functions for each species.
- > Vehicle population distribution by age of vehicle.
- > Vehicle mileage distribution by age of vehicle.
- > Effectiveness of technological emission control tactics on each model year.

The principal data sources for this information were the EPA for the new vehicle emission factors and speed correction functions, the CARB for the age deterioration functions, effectiveness of the technological control devices, and vehicle mileage distributions by age of vehicle, and the Department of Motor Vehicles for vehicle age distribution.

Three different cases of emission controls were considered: the Reference Case, the Nominal Case, and the Full Technological Controls Case. In the Reference Case, stationary sources were considered to be controlled at the levels existing in 1970. The emission control strategy for motor vehicles consists of the Federal Motor Vehicle Control Program for new vehicles.

The Nominal Case was the same as the Reference Case except that additional control measures for both stationary and mobile sources have been included. For stationary sources, the only additional control tactic reduces evaporative emissions from surface coating operations. Emission controls on LDMVs consists of two programs that were being implemented in California at the time of the study. The first applied to 1955-1965 vehicles. Beginning in 1972, these vehicles were required to have an exhaust emission control device installed upon transfer of ownership or at the time of initial registration in California. It was assumed that 80 percent of the 1955-1965 vehicles would have the retrofit installed by 1975. The second program applies to 1965-1970 vehicles and is intended to reduce NO_x emissions. Presently, this program requires that the vehicles be retrofitted at the time of initial registration or transfer of ownership. However, in 1975 the program required the mandatory installation of the retrofit on all 1966-1970 vehicles in the LA AQCR. Finally, an additional reduction in emissions from commercial aircraft is included in the Nominal Case. The modification is due to an agreement between EPA and the airlines and was intended to reduce visible emissions from commercial aircraft.

The Full Technological Controls Case consisted of the Nominal Case plus all of the technological emission controls specified in the implementation plan for the LA AQCR. Additional emission reductions from stationary sources were obtained in the Full Technological Controls Case by further controlling petroleum operations and organic solvent users. Furthermore, two additional tactics were included in the Full Technological Controls Case to further reduce the emissions from LDMVs. All in-use vehicles are

required to undergo an annual inspection to ensure that excessive exhaust emissions are not occurring. In addition, the implementation plans provided for the retrofit of oxidizing catalytic exhaust gas converters to reduce emissions of hydrocarbons and carbon monoxide.

Reynolds and Seinfeld (1975). As part of an evaluation of strategies for meeting the ambient air quality standard for photochemical oxidant, emissions of hydrocarbons, HO_x , and CO were projected for Los Angeles for 1977. Source categories included on-road motor vehicles, aircraft, power plants, oil refineries, and other miscellaneous fixed sources. Emissions were first computed for 1977 taking into account only growth factors that would affect sources. After this had been accomplished, emissions were adjusted for motor vehicle emission controls in the 1977 EPA control strategy.

The 1977 emissions from motor vehicles were calculated directly from the factors that affect these emissions. Average hot-start, cold-start, or steady cruise emission rates may be obtained by averaging the emissions from each type of motor vehicle using the following equation:

$$E_l = \frac{\sum_{i=1}^8 \mu_i \sum_{j=65}^{77} x_{ij} m_{ij} l_{lij}}{\sum_{i=1}^8 \mu_i \sum_{j=65}^{77} x_{ij} m_{ij}}$$

where l , i , and j are indices denoting chemical species, vehicle class, and model year, respectively, and

- E_l = average emission rate of species l (g/mi),
- μ_i = fraction of motor vehicles belonging to class i ,
- x_{ij} = fraction of class i vehicles manufactured in model year,
- m_{ij} = annual or daily mileage (mi/year or mi/day) for class i ,
model year; vehicles,
- l_{lij} = emission rate of species l from class i ; model year;
vehicles (g/mi).

The statewide distribution of motor vehicles, by model year, on 1 July 1972 was obtained from the State of California ARB. It was assumed that this distribution applied to 1 July 1977 and that the statewide distribution is applicable in the Los Angeles Basin.

The 1972 vehicle registration data for Los Angeles and Orange Counties listing the total number of automobiles, trucks, and motorcycles were used to calculate the relative distribution of these three types of vehicles. Statewide registration data for each class of light- and heavy-duty vehicles were then employed to further refine the automobile, truck, and motorcycle distributions cited above.

Annual mileages for automobiles and trucks as a function of vehicle age were obtained from the State of California ARB.

The cold- and hot-start emission rates for pre-1972 vehicles for hydrocarbons, CO, and NO_x were calculated from data obtained by AESi (1973). Emissions from 1972-77 vehicles were obtained from calculations based on consideration of existing emissions standards. Exhaust control device deterioration was considered for hydrocarbons and CO in post-1965 vehicles and for NO_x emissions from post-1969 model year vehicles.

The daily vehicle miles traveled for both freeways and surface streets in 1969 have been given on a 2-mile grid by Roth, et al. (1974). Liu (1975) estimated that the VMT in the Basin would increase by 17 percent between 1969 and 1977. The VMT increase in each 2x2 mile grid square was computed by apportioning the increase in total mileage according to projected population growth in each square.

To estimate aircraft emissions for 1977, three changes had to be made in the 1969 aircraft inventory reported by Roth, et al. (1974), including the addition of another class of aircraft (wide-body), the revision of jet aircraft emission factors, and the projection of the number of daily aircraft operations at each airport in 1977.

Projected NO_x emissions at each of the power plants operated by Southern California Edison Company and the Los Angeles Department of Water and Power were furnished by the power companies. Emissions from other plants were adjusted according to projections by the Los Angeles APCD. The temporal distribution of power plants emissions for 1977 was taken to be the same as in 1969.

Refinery emissions were assumed to increase proportionately to the increase in fuel consumption.

Emissions from other sources were estimated by Liu (1975).

Nordsieck (1976). Emissions projections at five-year intervals from 1980 through 1995 have been calculated for the Santa Barbara region by Nordsieck (1976). Predictions indicate that reactive hydrocarbon emissions from petroleum operations in the South Coast Air Basin will surpass the emissions from automobiles by 1985.

The emissions from on-road motor vehicles were obtained from a variety of sources. Vehicle mileage estimates were supplied by the Santa Barbara County Transportation Study (SCOTS), vehicle emissions factors were calculated from Supplement 5 to EPA's Compilation of Air Pollutant Emission Factors (AP-42), and vehicle mix data were supplied by CALTRANS. The projected increases in VMT given in SCOTS were allocated by grid square on the basis of population projections.

To calculate aircraft emissions, the mix of aircraft was assumed not to change. The number of LTO cycles was assumed to be proportional to the population. EPA regulations were also taken in consideration in calculating emissions.

Hydrocarbon emissions from petroleum production and transport were based on the projections of crude oil production rates in the vicinity of Santa Barbara and estimates of vapor loss rates from a given volume of petroleum products.

Stationary source emissions were allocated to land-use categories (industrial, residential, commercial) for the base year and for land use representative of the General Plan for Santa Barbara County (Nordsieck and Martinez, 1974). Intermediate year distributions were obtained by assuming that changes in land-use occur in proportion to population growth. Controls required in the State Implementation Plans were assumed to reach full application by 1977. Stationary source inventories of CARB were used.

EPA (1974). Techniques for projecting emissions to future years have been discussed in two volumes of the EPA document entitled "Guidelines for Air Quality Maintenance Planning and Analysis." In Volume 7 methodologies for projecting county emissions are discussed while in Volume 13 techniques for allocating the projected emissions to subcounty areas are presented.

Future emissions from point sources can be projected in the following manner:

$$\begin{aligned} (\text{future emissions}) = & (\text{base-year activity}) (\text{growth factor}) \\ & \times (\text{emission factor}) (1 - \text{equivalent control efficiency}) \end{aligned} .$$

Depending on the type of source growth factors can be calculated from population projections for the county or from employment projections. Alternatively, the growth factors may be calculated from statewide population projections using historical trends for the county. As a last resort, the OBERS projections may be used to calculate growth factors. It is recommended that projections of emissions from large point sources be treated individually--the information required for calculating growth factors being obtained from the individual facility.

Three levels of emission projections are defined based on the availability of resources, the degree of accuracy required and the amount of resources expended in upgrading the baseline inventory. At each level, six categories of emissions were considered:

- > Industrial process emissions
- > Fuel combustion
- > Transportation
- > Electric generation
- > Incineration
- > Miscellaneous area sources.

Level 1 projections require the least amount of input data. However, it is recommended that the projection of industrial process emission be carried out for individual sources. If this cannot be accomplished, growth factors can be obtained from the OBERS economic projections. A weighted control factor should be calculated based on control regulations on each source type. To calculate emissions from fuel combustion, it is recommended that the county baseline fuel use be converted to BTU equivalents. The projected BTU demand from the residential sector is determined from population growth, from the industrial sector by the OBERS industrial growth projections, and from the commercial sector from the OBERS commercial growth projections. The future fuel mix is to be determined from interviews with local authorities. Emission factors are to be determined from AP-42 or the NEDS emission file. It is recommended that the growth factors for off-highway vehicles be determined by the expected change in agricultural activity. Growth factors for rail transport are to be obtained from transportation studies or directly from the railroads. Growth factors for commercial vessels are to be determined from consultation with officials, while the growth of recreation vehicles is assumed to follow population growth. Sources of data for projecting aircraft operations include transportation and individual airport studies, state air traffic projection data from the FAA, and national air traffic forecasts from the FAA. Emissions from electricity generation requires several factors to be determined, including the amount of electricity to be generated in the county during the projection years, the fuels to be burned, the efficiency of fuel conversion (kWh/unit fuel), and emission factors. To perform a Level 1 forecast of the emissions from incineration,

it is recommended that the manufacturing sector growth factor from OBERS be used. The major miscellaneous area sources are evaporation from retail gasoline sales and from solvent operations. Future gasoline sales can be determined from the county or state energy office, the county or state tax agency, transportation studies, or gasoline dealer association surveys. Population growth factors are to be used in determining the growth of solvent operations.

Level 3 projection estimates differ primarily from Level 1 estimates in the treatment of industrial process emissions and emissions from fuel combustion. Industrial process emissions are to be determined from:

- > Growth factors determined from an industrial interview program for each SCC process.
- > The determination of equivalent control efficiency required for each pollutant based on applicable control regulation.
- > An emissions factor calculated for each point source.

To forecast emissions from fuel combustion, the county baseline fuel use is again to be converted to BTU equivalents. In the residential sector, growth factors are to be determined from the projected increase in the number of dwelling units. For industry, growth factors are to be determined from interviews of projected major industrial polluters as determined from county growth plans. Commercial fuel combustion and fuel use in all sectors are to be determined from interviews with the appropriate authorities. Emission factors are to be determined from AP-42 or from NEDS.

To allocate emissions projections to subcounty areas requires a knowledge of the spatial distribution of existing sources as well as knowledge of preferred sites for future sources. Specifically, projected county emissions are based on the anticipated emissions arising from the fulfillment of a need for growth in a particular sector. On the county level, no consideration was given to whether the growth could be satisfied by increased activity of existing sources or whether these sources

would be saturated requiring the construction of additional sources. Furthermore, the increased emissions output from existing sources and the placement of new sources may be constrained by federal, state, and local regulations on emissions from new sources. Hence, allocating county emissions to subcounty regions requires a knowledge of the location of sources, their current activity, their saturation level, the location of future source sites, and regulations on new and existing source emissions.

Volume 13 of the EPA document "Guidelines for Air Quality Maintenance Planning and Analysis" discusses the allocation of emissions to subcounty areas at three orders of complexity. The principal data requirements for allocating projected emissions to the subcounty area for each of five source categories are shown in Table A-5.

TABLE A-5. ACTIVITIES FOR ALLOCATING PROJECTED EMISSIONS TO SUBCOUNTY AREAS

Source Category	Analysis Order	Principal Data Requirements
Residential	1	Population for each area
Fuel Combustion	2	Dwelling units for each area
	3	a. Dwelling units for each area
		b. Building size distribution for each area
		c. Fuel use distribution for each area
Transportation		d. Total degree-days for each area
	1	Population for each area
	2	a. VMT for 3 vehicle classes for each area
		b. Location and activity data for all major airports, railroad terminals, port facilities, and concentrations of miscellaneous sources both existing and projected
Commercial/ Institutional Fuel Combustion	All	Location and fuel use of large commercial/institutional sources from point source file
	1	Population for each area
	2	Commercial/institutional employed persons for each area
		a. Number of commercial/institutional establishments for each area
		b. Commercial/institutional building size distribution for each area
		c. Commercial/institutional fuel use distribution for each area
Industrial Emissions		d. Total degree days for the region
	1	a. Growth rates for each of the major process sources:
		1) OBERS by process source category or SCC process source categories
		2) Special planning studies
		3) Chambers of Commerce
		4) Plant interviews
		b. File of existing point sources
		Sources:
		1) State air pollution control agency
		2) USEPA-NEDS file
	2	a. As in 1
		b. As in 1
		c. Employment or land use for existing sources
		d. Employment levels or land use by process source category and geographic area
		e. Estimates of fractions of existing and new source growth
Solid Waste Disposal	1	a. Actual and projected population and/or employment for each area
		b. Estimated county totals of solid waste disposed by on-site incineration and open burning
	2	a. Actual and projected populations and/or employment for each side
		b. County totals of solid waste disposal of by on-site incineration and open burning
	3	a. Actual and projected population and/or employment for each area
		b. County totals of solid waste disposed of by on-site incineration and open burning

TABLE A-6. EMISSION SOURCE CATEGORY AND CORRESPONDING
GROWTH INDICATOR CATEGORY

<u>EMISSION SOURCE CATEGORY</u>	<u>GROWTH INDICATOR CATEGORY¹</u>
PETROLEUM	
Production	Mining-crude petroleum and natural gas
Refining	Manufacturing - petroleum refining
Marketing	Population ²
ORGANIC SOLVENT USERS	
Surface Coating	Manufacturing - composite index
Other	Population ²
Degreasing	Manufacturing - composite index
Dry Cleaning	Population ²
CHEMICAL	Manufacturing - chemical and allied products
METALLURGICAL	Manufacturing - primary metals
MINERAL	Mining - non-metallic, except fuels
FOOD & AGRICULTURAL PROCESSING	Manufacturing - food and kindred products
PESTICIDES	Agriculture
WOOD PROCESSING	Manufacturing - lumber products & furniture
COMBUSTION OF FUELS	
Power Plants ³	-----
Industrial	Manufacturing - composite index
Domestic and Commercial	Population ²
Orchard Heaters	Agriculture
WASTE BURNING	
Agricultural Debris	Agriculture
Forest Management	Forest and fisheries
Range Improvement	Agriculture

TABLE A-6 (Concluded)

<u>EMISSION SOURCE CATEGORY</u>	<u>GROWTH INDICATOR CATEGORY¹</u>
WASTE BURNING (cont.)	
Dumps ⁴	-----
Conical Burners	Manufacturing - lumber products & furniture
Incinerators	Population ²
Other	Population ²
MISCELLANEOUS AREA SOURCES	
Wild Fires	Constant
Structural Fires	Population ²
Farming Operations	Agriculture
Construction and Demolition	Contract construction
Unpaved Roads	Population ²
Other	Population ²

¹ Growth indicator based on earnings or production data from [37,39,49] unless otherwise specified. Specific indices may be found in Tables B.3.5 through B.3.21 of Appendix B. A specific example of relating emission source category to growth indicator category is presented in Table B.3.22 of Appendix B.

² Population growth factors based on projections of California Department of Finance, D-100 series.

³ Special study - Please refer to section on Fossil Fuel Electric Generating Plants.

⁴ Future emissions are assumed negligible.

The ARB has also expressed the need for the projection methodology to have the capability of allowing the forecaster to decide subjectively how the emissions from a given source will change or whether new sources will be added. This capability should be available both at the SMSA or county level and at the subcounty level. That is, the user may want to add a source to the projected county emissions and let the allocation of the subcounty level be done objectively. On the other hand, there may be times when it is necessary to determine the location of the new source at the subcounty level after projections have been made objectively at the county level.

Finally, we consider only major stationary sources, other stationary sources, and off-road motor vehicles. The emissions from on-road motor vehicles will be determined from the ARB transportation models, which can be integrated into the forecasting model. A flow diagram of the general forecasting algorithm that could be used in an emissions simulator is shown in Figure A-2. We have included an option for the forecasting to be done in a subjective manner by the forecaster either at the county level or at the subcounty level. The flow diagram follows the general pattern of forecasting at the county level and allocation at the subcounty level described previously. For the stationary source emissions, each source category is treated independently as is each category of off-road vehicles. A more specific flow diagram can be put together for the ARB source categories given in Table A-6; however, often the source categories may be quite different from the ARB categories, which would require the user to determine new relationships between sources and growth categories. Furthermore, the allocation procedures will vary depending upon the county in which the allocation of emissions is being made. In general, the allocation of emissions will have to be done by population, but the forecasting algorithm should be sufficiently flexible so that other factors such as employment categories, land use, or housing unit distribution can be used to allocate emissions when this information is available. This can, of course, be done, but it will require considerable interaction of the forecaster with the forecasting model that he is using.

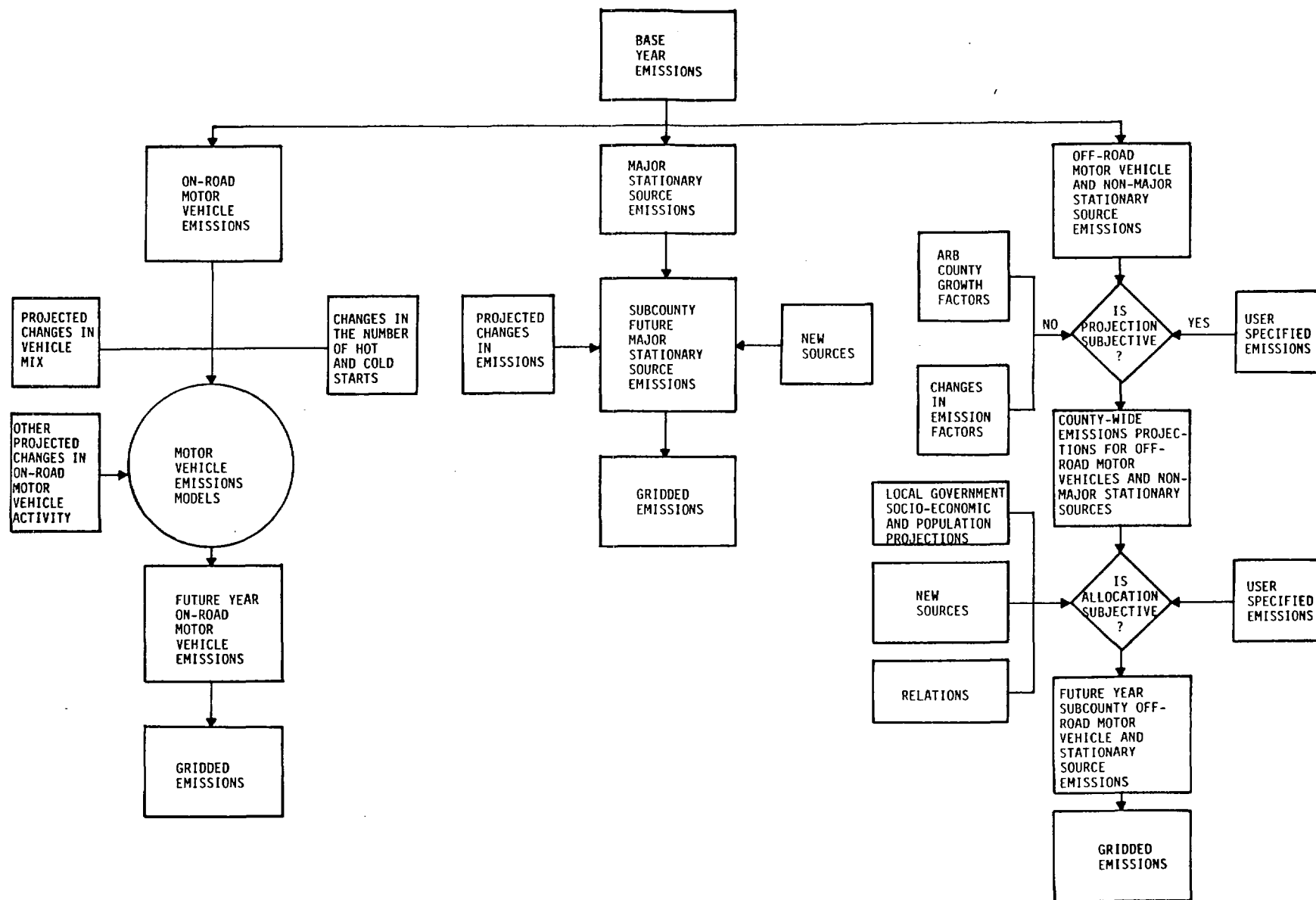


FIGURE 2. FLOW DIAGRAM OF A FORECASTING ALGORITHM FOR THE ARB EMISSIONS SIMULATOR

For the current emissions simulator, we recommend incorporation of the ARB emissions projections at the SMSA level with the option left open for other emissions categories to be included at the user's discretion. At this time, we recommend that allocation at the subcounty level be subjective. Putting together an algorithm for doing the objective allocation would require considerable data collection and analysis from the various agencies beyond what has already been done.

REFERENCES

- Air Resources Board (1974), "Emissions Forecasting Methodologies," Air Resources Board Evaluation and Planning, Sacramento, California.
- _____ (1976), "Emissions and Air Quality Assessment," Air Resources Board Technical Resources Unit/Land Use Planning Program, Sacramento, California.
- Booz-Allen and Hamilton (1975), "Comparison of Four Methodologies To Project Emissions for the St. Louis Metropolitan Area," EPA-450/3-75-074, Office of Air and Waste Management, Research Triangle Park, North Carolina.
- Goodrich, J. C. (1973), "Hackensack Meadowlands Air Pollution Study--Emission Projection Methodology," EPA 450/3-74-056-b, Office of Air and Waste Management, Research Triangle Park, North Carolina.
- Mikolowsky, et al. (1974), "The Regional Impacts of Near-Term Transportation Alternatives: A Case Study of Los Angeles," R-1524-SCAG, Rand Corporation, Santa Monica, California.
- Reynolds, S. D., and J. H. Seinfeld (1975), "Interim Evaluation of Strategies for Meeting Ambient Air Quality Standard for Photochemical Oxidant," Environ. Sci. Tech., Vol. 9, pp. 433-447.
- Nordsieck, R. A. (1976), "Air Quality Impact Analysis for the Santa Barbara County (California) Regional Transportation Plan," Santa Barbara County-Cities Area Planning Council, Santa Barbara, California.
- Environmental Protection Agency (1974), "Guidelines for Air Quality Maintenance Planning and Analysis," Volumes 7 and 13, EPA-450/4-74-008 and EPA-450/4-74-014, Office of Air and Waste Management, Research Triangle Park, North Carolina.