# Redesign of California's Emission Forecasting System (CEFS)

Martin E. Johnson
California Air Resources Board
Technical Support Division
P.O. Box 2815
Sacramento, CA 95812
E-Mail: mjohnson@arb.ca.gov

Dr. Bharat Lakhanpal
Institute for Economic and Environmental Studies
California State University, Fullerton
Fullerton, CA 92634-9480
E-Mail: blakhanpal@fullerton.edu

## ABSTRACT

Because of the policy decisions currently being based on future emission inventory projections for regional attainment planning and attainment demonstrations, the staffs of the California Air Resources Board (ARB) and California State University Fullerton (CSUF) have been involved in a three-year joint project to design a computer analysis program capable of responding to these critical tasks of California's air pollution program.

The purpose of the project was to redesign the algorithms and data management components to provide a state-of-the-art system for predicting future emission levels (particularly for stationary sources) in an optional and flexible data management environment. The program incorporates new forecasting logic, new methods for estimating seasonal emissions for planning purposes, day-specific emissions for air quality models, and for spatial analysis of emissions. The scope of the project was large and complex, requiring the integration of various data models, such as the CEIDARS base year inventory, the rule tracking subsystem, socioeconomic activity, temporal activity, and Geographic Information System (GIS) data. ARB staff was responsible for overall project management and conceptual design, and CSUF staff was responsible for software construction.

Although the principle purpose of the redesign was technical in nature and will be the focus of much of this paper, the success of this project is largely attributed to the attention given to developing working relationships between the staffs of ARB and CSUF. Thus the theme of diligent teamwork will be woven throughout the pages to follow and is a key underlying message to those planning similar design efforts in the future.

## INTRODUCTION

Air pollution programs have always depended on predictive models for gaining a better understanding of what the emission volumes will be in the future. The results of these models assist in the development of air quality plans; determine how and where air pollution can be reduced most efficiently; track progress toward meeting the requirements of air pollution control mandates; and are used to construct emission trends. The California Air Resources Board (ARB) has performed computerized emission projections since 1981. The first forecasting program operated on an IBM 3270 mainframe in a RAMIS hierarchical database environment and was designed to operate in conjunction with the base year emission inventory management system (known as the Emission Data System (EDS)). The RAMIS based program has operated reliably and has served as a valuable emission analysis tool for over fifteen years.

In California, local air pollution control and air quality management districts (hereafter referred to as districts) are responsible for providing much of the stationary source emission inventory information. For this reason, the EDS was redesigned earlier this decade to enable district end-users the remote access necessary to add/update emission inventory information electronically. The program was designed using the industry standard Oracle Relational Data Base Management System (RDBMS) and operates in the UNIX environment. The new base year emission inventory management system is referred to as the California Emission Inventory Development and Reporting System (CEIDARS)[1]. With the CEIDARS system becoming operational, there was a high degree of incompatibility between the RAMIS based forecasting programs and the inventory data in CEIDARS. By having a new forecasting system developed in an Oracle RDBMS environment, we would ensure a tight robust linkage with CEIDARS now and into the future.

Another change in the ARB's air pollution tracking activities occurred as a result of the heightened interest in emission analyses due to the 1990 Federal Clean Air Act (CAA) Amendments, and the need for tracking the progress of the 1994 State Implementation Plan (SIP). It became apparent that the forecasting program would require a major redesign effort to meet client needs. Since districts are responsible for adopting local emission control regulations for many stationary source sectors, districts need the ability to input control rule information into the forecasting system. As such, remote on-line access capabilities were a major criteria in the redesign of the forecasting program. The principle design objectives for the new forecasting system, California Emission Forecasting System (CEFS), were the following:

- Achieve full compatibility with CEIDARS
- Improve the forecasting logic
- Improve the capabilities for tracking CAA emission reduction requirements
- Improve the temporal algorithms used for seasonal and day/hour-specific inventories
- Develop forecasted day/hour-specific emission inventories for modeling
- Enable on-line remote access to districts

This paper will describe the forecasting algorithms employed by CEFS; will provide a look into the life-cycle management of the project as a system; and will describe the attributes of the software and user interface.

## METHODOLOGY AND ALGORITHM DESIGN

### Background

The conceptual design of a new emission forecasting algorithm began in 1991 with the development of a statewide Rule Tracking Subsystem (RTS)[2] which was developed to quantify the impacts of emission control strategies on future emission inventories. The ARB's prior forecasting system performed emission projections based on a broad growth and control category definition defined by ARB—there are approximately 100 growth categories and 200 control categories[3]. The RTS was designed to track the effects of emission control rules on stationary sources at the process level (identified by Source Classification Code (SCC) and Standard Industrial Classification (SIC) or Emission Inventory Code (EIC)—which comprises about 30,000+ possible process/industry categories statewide. The development of the RTS and its associated control analysis logic would become the core of the design of the CEFS forecasting algorithm. As with the control logic, the growth logic was also redesigned to link growth activity directly to the source categories impacted. As such, the prior growth categories have also been abandoned.

### Key Logic Features

CEFS includes two separate forecast algorithms: 1) TREND forecast module in which the emissions are aggregated (pre-summed by source categories (SCC/SIC; EIC)) for each region (defined by District/Air Basin/County); and 2) GIS forecast module which performs emission projections at the facility/device/process level for developing gridded inventory inputs for modeling. The general relation for the TREND algorithm is shown in equation (1) below. Typically, the forecasting workload at ARB has been limited to criteria pollutants (TOG, $NO_X$, CO, $SO_x$ , PM) and this will continue to be the emphasis in the future. However, the CEFS algorithm is capable of projecting emissions for toxic pollutants as well. As shown in equation (1), multiple control measures can be applied to a particular emission category. This is invaluable when doing control measure analysis and "what-if" scenarios. The effects of particular control measures can be isolated by modifying the control data prior to running a forecast scenario. In addition, CEFS differentiates between *adopted rules* and *proposed measures*, so the user can run forecast scenarios to compare baseline projections to planning projections. Figure 1 shows a typical graphical representation of this type of emission analysis which is used to track the progress of air quality plans and for attainment demonstration assessments.

CEFS employs a sophisticated region and emission category selection hierarchy when selecting and applying growth and control factors. Figure 2 shows the region and category assignment options which are available to the user. Considering, for example, a rule which applies to the entire universe of a particular process type (identified by SCC), the category selection "option 7" would be assigned to these control data. Now, if that rule is more stringent for a particular industry (identified by SIC), then "option 4" could be assigned for that process/industry (i.e. SCC/SIC combination). CEFS will cycle through the control data records applying the highest level of the selection hierarchy to that category (i.e. the control data can be "layered" to target the source categories exactly as the rule calls for without double-counting). Figure 3 displays how a boiler rule can be applied to various levels of process/industry groupings. The same type of logic is used for the growth data.

The logic used for applying the growth and control factors in the TREND and GIS modules are the same—the only difference being that the factors are applied at the facility level in the GIS program. The main difference between the algorithms is how the temporal logic is applied. There are two distinct temporal algorithms that drive CEFS: 1) Seasonal Inventory Temporal Algorithm employed in the TREND module[4] and 2) Daily/Hourly Temporal Algorithm employed in the GIS module[5]. Equation (1) uses the "seasonal" temporal algorithm. The Daily/Hourly algorithm will be discussed later in this section. Each factor in equation (1) will be discussed separately. Each of the primary variables are functions of other variables which are denoted with subscripts.

**TREND Forecast Module**

$$E\_fy_{(r,s,p,y)} = E\_by_{(r,s,p,o)} * TF_{(r,s,p)} * FRAC_{(r,s,p,y)} * GF_{(r,s)} * [\ CF_{(r,m_1,s,p)} * CF_{(r,m_2,s,p)} * \ldots * CF_{(r,m_i,s,p)}\ ] \qquad (1)$$

where

Primary Variables:

E_fy = Emissions in the future year (tons/day; annual average or seasonal)

E_by = Emissions in the base year (tons/year)

TF = Temporal Factor (or "Seasonal Adjustment Factor")

FRAC = ROG, VOC, $PM_{10}$ , or $PM_{2.5}$ fraction (if applicable)

GF = Growth Factor (is the ratio of two activity levels at the end-point years--explained later)

CF = Control Factor (is the ratio of two control levels at the end-point years--explained later)

Subscript Variables:

r = region (district, air basin, county)

s = the source category (SCC/SIC or EIC for the TREND algorithm; The GIS algorithm performs the projections at the facility level.)

p = the pollutant

m = the control measure (i = the ith control measure)

y = the year to be projected

o = the base year

**<u>The Temporal Factor.</u>** In order to characterize emissions during typical air quality exceedance periods, seasonal adjustment factors (or "temporal factors") are used to apportion emissions into the periods under consideration for air quality planning purposes. The temporal algorithm used for calculating seasonal emission inventories was rewritten to better approximate the seasonal variation of emissions [4]. In the prior algorithm, emission estimates for <u>point sources</u> represented an "average annual operating day". The assessment of this algorithm showed that it did not accurately estimate seasonal emissions because the algorithm was driven by the number of operating days a point source process operates. For processes that operated intermittently, this resulted in exaggerated seasonal emission estimates. For <u>area sources</u>, the algorithm estimated emissions based on an "average seasonal operating day". The new seasonal inventory algorithm offers an improved method for estimating seasonal average emissions and is shown in equation (2) below:

$$TF_{(r,x)} = SEAS\_FRAC_{(r,x)} / (\#\ days/season) \qquad (2)$$

where

SEAS_FRAC = Sum of fractional monthly throughputs for the emission process
Summer: May-October
Winter: November-April
#days/season=184 for Summer, 181 for Winter

or

= [# operating days / season] / [# operating days / year]

Subscript Variables:

x = denotes the different levels at which the temporal data are stored and utilized. For point source processes, if data are provided at the *facility/device/process* level, these data are used otherwise default data maintained by *EIC* are utilized. For area-wide sources, all temporal profiles are maintained by EIC.

A single emission-weighted temporal factor, $TF_{(r,s,p)}$, is calculated for each source category (SCC/SIC combination for point sources, EIC for area-wide and off-road mobile sources) and pollutant in the emission inventory as shown in equation (3).

$$TF_{(r,s,p)} = \sum (EMS_{(r,x,p)} * TF_{(r,x,p)}) / \sum EMS_{(r,x,p)} \qquad (3)$$

**<u>ROG, VOC, $PM_{10}$ and $PM_{2.5}$ Fractions.</u>** The reactive portion of the Total Organic Gas (TOG) emissions (expressed as either ROG or VOC) are calculated by applying reactive fractions which are maintained in CEIDARS. Districts may supply fractions at the *facility/device/process* level for point source processes. If these data are not provided at this level, default fractions maintained by ARB at the SCC level are invoked[6]. In like manner, the portion of Particulate Matter (PM) falling within the ten-micron and 2.5 micron size ranges (i.e. $PM_{10}$ and $PM_{2.5}$ respectively) are estimated from district-supplied fractions or by applying size fractions maintained by ARB in CEIDARS.

As with the temporal factor, an emission-weighted fraction is developed for each SCC/SIC pair. Although equation (1) shows FRAC as a variable in the algorithm, the factor is actually carried to the TREND database and stored. The fractions are applied at the reporting step.

$$FRAC_{(r,s,p,y)} = \sum (EMS_{(r,x,p,y)} * FRAC_{(r,x,p,y)}) / \sum EMS_{(r,x,p,y)} \quad (4)$$

where
x = denotes either *facility/device/process* level or *SCC* level data values

Note: The capability exists for default fractions to be maintained by SCC for all years in a separate lookup table in CEIDARS. These fractions are used for the projection years. When such data become available, this will enable CEFS to capture the time-variation of reactive fractions when performing forecasts and backcasts (e.g. changes in organic composition due to reformulation of gasoline).

**The Growth Factor.** Growth factors are derived from county-specific economic activity profiles, population forecasts, and other socio/demographic activity. Growth profiles are typically associated with the type of industry and secondarily to the type of emission process. For point sources, economic output profiles by industrial sector are typically linked to the emission sources via SIC[7]. For area sources, other growth parameters such as population, dwelling units, fuel usage etc. may be used. The growth factor is the ratio of the growth level in the future year, $GL_y$ to the growth level in the base year, $GL_o$. These growth levels are also region and source category dependent.

$$GF_{(r,s)} = GL_{(r,s,y)} / GL_{(r,s,o)} \quad (5)$$

where
o = denotes the base year for growth and control

**The Control Factor.** Control factors are derived from "adopted" ARB regulations, district rules, and "proposed" measures which impose emission reductions or a technological change on a particular emission process. Control factors comprise three components: Control Efficiency, Rule Effectiveness, and Rule Penetration. Control factors are closely linked to the type of emission process and secondarily to the type of industry. Control levels are assigned to emission categories which are targeted by the rules via emission inventory codes (SCC/SIC, EIC etc.) used in CEIDARS. The control factor is the ratio of the control level in the future year, $CL_y$ to the control level in the base year, $CL_o$. These control levels are also region, control measure, source category, and pollutant dependent.

$$CF_{(r,m,s,p)} = CL_{(r,m,s,p,y)} / CL_{(r,m,s,p,o)} \quad (6)$$

where

$$CL = 1 - \{(1-CE) * RE * RP\}_{(r,m,s,p,y)} \quad (7)$$

where
*Control Efficiency,CE* = The technological efficiency of the abatement device or method (expressed as the fraction of remaining or uncontrolled emissions)
*Rule Effectiveness,RE* = Correction factor to adjust for "real-world" operating conditions (i.e. equipment breakdowns, operator variance etc.)
*Rule Penetration,RP* = The degree a given rule penetrates the universe of emission processes coming under the purview of the rule (taking into account rule exemptions, etc.)

**GIS Forecast Module**

The GIS forecast module develops a day/hour specific emission inventory for a particular day of the week in a given month (e.g. a typical Wednesday operating day in July) as well as projects the emissions into the future. These features enable CEFS to predict emissions levels under episodic conditions which are then fed into photochemical models.

As mentioned in the opening paragraph of this section, the GIS module relies on a Daily/Hourly temporal algorithm[5] which was designed specifically for this module (these temporal data are stored in the CEFS Temporal data module). The general forecasting logic is the same as the TREND module, therefore, we will only treat the temporal and spatial aspects of the algorithm here. Due to the complexity of the algorithm, only a partial treatment of the logic will be covered here. Readers interested in more detailed information on the logic (i.e. linkages between the program and particular data elements in the CEFS Oracle tables, etc.) may refer to references.

**Case 1: When Day/Hour-Specific Temporal Data are not Available.** In this case, the daily and hourly emissions are estimated based on weekly and daily operation cycle data (denoted by DPWK and HPDY). Formerly, a single weekly and daily operating cycle were collected and maintained in CEIDARS for a given year. In the new temporal module, this operation data will be month-specific which will provide a much improved method for generating modeling inventory inputs. Equations (8) and (9) below are used to estimate the daily and hourly emissions for case 1.

$$\text{DEMS} = \frac{\text{EMS}}{365} * \frac{\text{Monthly Throughput}}{(1/12)} * \frac{\text{DAYFACTOR}}{(1/7)} * 907.18474 \qquad (8)$$

$$\text{HEMS} = \frac{\text{EMS}}{365} * \frac{\text{Monthly Throughput}}{(1/12)} * \frac{\text{DAYFACTOR}}{(1/7)} * \frac{907.18474}{\text{HOURS}} \qquad (9)$$

where

DEMS = Emissions of a particular day of the month (kg/day)

HEMS = Emissions per hour of the particular day (kg/hour)

EMS = Base year emissions (tons/year)

Monthly Throughput (expressed in terms of monthly fraction or ratio of operating days)

= MONTHFRAC

or

$$= \frac{\text{(Operating Days/month, OP\_DAYS)}}{\sum \text{OP\_DAYS for all 12 months}}$$

DAYFACTOR = decoded weekly operating cycle (DPWK)

907.18474 = conversion factor (tons to kilograms)

HOURS = decoded daily operating cycle (HPDY)

**Case 2: When Day/Hour-Specific Temporal Data are Available.** In these cases, data are maintained by EIC category in two separate Oracle tables in CEFS. In the process of designing this algorithm and associated temporal data module, we considered designing the algorithm to accommodate day/hour specific information at the process and device level for point sources. However, generally the data would only be available by conducting a survey of particular point sources for particular operating days. The inclusion of this temporal information at the process and device level, although desirable, would add substantial complexity to the algorithm. Due to the general unavailability of these data, adding this additional complexity was not advisable. In cases when data become available at this level, these data can be used to calculate the emissions manually and overwrite the data in the GIS output files. Equations (10)-(13) are employed to calculate the daily and hourly emissions for case 2.

$$DEMS = EMS * \text{Monthly Throughput} * DAYFRAC / 4 \tag{10}$$

where

DEMS = The daily emissions based on a 28 day month
EMS = Annual emissions in tons/year
Monthly Throughput = (same as defined in Case 1 above)
DAYFRAC = The fraction of weekly activity occurring on that day

$$DEMS\_adj = \{DEMS - [\sum DEMS_{i=1,31} - (EMS*\text{Monthly Throughput})] / \#DAYS \} * 907.18474 \tag{11}$$

where

DEMS_adj = The daily emissions adjusted for the actual number of days in the month
#DAYS = The total number of days in the month

If hour-specific data are not available by EIC, equation (12) is used to calculate the hourly emissions.

$$HEMS=DEMS\_adj * (1/HOURS) \tag{12}$$

where

HOURS=The decoded hourly operation cycle HPDY

$$HEMS=DEMS\_adj * HOURFRAC \tag{13}$$

where

HOURFRAC= The fraction of daily activity occurring in that hour

Once the daily and/or hourly emissions are calculated, equations (14) and (15) below are employed to do the projection. The output of these equations give projected day/hour specific emissions. For point sources, these emissions are by facility/device/process along with associated UTM coordinates which are carried from CEIDARS. For area-wide sources, the emissions are spatially resolved using spatial surrogate parameters assigned to the area source categories in CEIDARS. There is no FRAC variable in these equations because CEFS provides TOG emissions as the input to modeling inventories. The speciation of organics are handled as part of the photochemical modeling using the same basic profiles as those used for annual and seasonal inventories.

$$E\_fy_{(r,s,p,y)} = DEMS_{(r,s,p,o)} * GF_{(r,s)} * [\, CF_{(r,m_1,s,p)} * CF_{(r,m_2,s,p)} * \ldots * CF_{(r,m_i,s,p)} \,] \tag{14}$$

$$E\_fy_{(r,s,p,y)} = HEMS_{(r,s,p,o)} * GF_{(r,s)} * [\, CF_{(r,m_1,s,p)} * CF_{(r,m_2,s,p)} * \ldots * CF_{(r,m_i,s,p)} \,] \tag{15}$$

The first major test for the GIS Forecast module will be to prepare the point source modeling inputs for the 1997 Southern California Ozone Study (SCOS97)[8]. This work will include an assessment of temporal variations of all emission processes, the results of which will be loaded and used as a basis for developing a day/hour-specific inventory for gridding purposes. These gridded inventories will be primary inputs to the photochemical modeling exercises to be conducted for this study.

## PROJECT MANAGEMENT AND COORDINATION ISSUES

As mentioned in the introduction, the overall objectives for CEFS included:

- Interfacing seamlessly with CEIDARS, the Oracle RDBMS based emission inventory system for obtaining inventory data required by the forecasting algorithms
- Operationalize the forecasting, tracking, and modeling inventory algorithms including recent improvements made by ARB staff
- Provide on-line access to the districts
- Accomplish all the above in a flexible Oracle RDBMS based system operating on ARB's DEC machine using the ULTRIX operating system

As can be surmised from these objectives, developing a system to accomplish all the objectives was a challenging task. The task was daunting because of the following reasons:

- The scope of the project was large not only in size but also in its breadth
- CSUF did not have a DEC machine running ULTRIX to use during software development

From a system design perspective, it would be most expedient to develop the system in the same environment as the one in which the production system would be run upon completion. Accordingly, CSUF proposed that development be carried out on the DEC machine at Teale Data Center in Sacramento. However, the DEC system hosted several other production-level applications (CEIDARS, Air Toxics Emission Data System (ATEDS), Reformulated Gas Tracking etc.). The system support personnel at ARB were understandably reluctant to allow a production machine to be used for development work. After several safeguards were put in place, the decision was made to develop the system on the DEC machine itself.

The next hurdle for beginning software development was the fact that Sacramento and Fullerton are several hundred miles apart! Here, we made use of the Internet and all development and testing was done on the DEC machine from Fullerton using Telnet. For any software development work, access to a machine is needed at the highest security level (i.e. close to that of the system administrator). Since we were going to access the machine over the Internet, we had to be extremely careful about security issues. Only three accounts and corresponding passwords were issued to CSUF. Further, log-in times and activities/processes were monitored carefully.

In addition to the general challenges mentioned so far, other challenges emerged during the design, development, and testing of CEFS: the recent importing of CEFS from the DEC environment to the SUN environment as well as the redesign of CEIDARS itself. These will be highlighted in the following sections.

### Project Phases

**Background.** The CEFS project began with a series of telephone meetings between ARB and CSUF. The purpose of these initial meetings was for CSUF to understand the domain knowledge, the capabilities of the existing system, the high level user requirements (of ARB and district staff), the hardware and systems software available, and the overall time line. The first step after this was to prepare an initial proposal that included a technical approach and a project plan. To allow the project to proceed in an orderly manner, CSUF proposed that system development proceed in accordance with the System Life Cycle methodology. This well accepted systems design methodology views software projects as consisting of several phases: planning, analysis, design, implementation/development/testing, and use/maintenance.

**<u>Customer Requirements Specifications.</u>** The first order of business was to understand the functionality of the existing system and to analyze the "customer requirements". For project success, it was critical that this first step be done thoroughly since much of the subsequent design work depends on it. Through a very thorough process which involved face-to-face interviews, document analysis, as well as examination of existing programs and new algorithms, data files, and reports, CSUF developed two documents: 1) Existing Systems Document and 2) Customer Requirement Specification Document[9]. This step took the most amount of time because of the steep learning curve in developing a clear understanding of emission analysis techniques employed by CEFS. This is where the Customer (ARB) and the Contractor (CSUF) established the foundation for this project. The subsequent phases of the project (described below) built off of this early momentum.

**<u>High Level Design.</u>** The next step was to develop a design that would meet the user needs and the system objectives identified earlier. The design phase began with a high level design[10] which included an overall topology (figure 4) that identified the main modules of CEFS. The linkages to CEIDARS were planned right from this level. For each module, an Entity-Relationship Diagram was developed to capture the data structures, and Data Flow Diagrams were developed to capture the processing aspects. This design document also included a module on system administration for tasks such as adding users and changing access rights of users.

**<u>Detailed Design.</u>** The next step was to develop a detailed design document[11]. Though the high level design task is important because the inner workings of a system depend on it, the detailed design is important because it determines the users' interface with the system. This part of the design included screen design, report structure, data dictionary, and special function keys. The approved design document formed the basis for system implementation/development which included:

- Creating the Oracle tables
- Writing programs for data input, forecasting, and reporting
- Testing the system in a modular fashion based upon test plans

**Summary**

Once the system level testing was completed, the system was placed under user testing. This was where ARB staff could now operate the system in test mode. All problems found during this time were systematically reported and fixed until the system was deemed to be ready for production.

Three factors were most noteworthy for the whole project:

- Aside from two face-to-face meetings and a couple of review meetings, all the coordination between ARB and CSUF was done through the use of E-mail
- The whole system was created remotely at Fullerton
- All the development work was done over the Internet using Telnet

This speaks particularly well of E-mail as a project management tool and of the future of the Internet as a means to support off-site systems development. This becomes even more impressive considering the large scope and complexity of CEFS. In developing CEFS, we truly leveraged the advantages offered by the networked environment!

## SYSTEM DESIGN

The System Life Cycle methodology was used to effectively manage CEFS development. However, an enormous effort went into coming up with a design that allowed all the system objectives and user requirements to be met.

## Overall Design

The overall design of the system is shown in figure 5. As can be seen from this figure, very early on, we came up with the concept of a “working database” and an “approved database”. The working database is one which contains growth, control, and temporal data that is accessible to the districts in read and write mode. Thus, they can make changes to this database. ARB staff also have the same type of access to this database. The working database is subject to post validation QA checks and reports are generated for invalid records. Once the invalid data have been taken care of, the growth and control datasets must be interpolated for the years to be forecasted. These interpolated data records are flagged to differentiate them from the original data. This is accomplished with a pre-processing program in the forecast module. Once the “working database” is stable, ARB staff can create a snapshot of the working database to create an “approved database”. All forecasting and reporting would be done from the data in the approved database. Access to the approved database is read-only for all users.

The forecast process involves taking the data from the approved database, as well as selected data from CEIDARS, to produce a “Trend Database” and/or GIS input files. Finally, all the necessary forecasting reports are run from data in the Trend Database. Figure 4 provides a view of all the system modules: RTS module, Growth Module, Temporal Module, CEIDARS interface Module, Input Parameter Module, Forecast Module, and Report Module.

## Data Inputs

As can be seen from Figure 4, the heart of CEFS is the Forecasting Module. This module contains inputs from the RTS, Growth, Temporal, and CEIDARS interface modules. In addition, the Input Parameter Module allows the users (ARB staff) to enter the parameters for running the specific forecast. In addition to being used in the forecasting module, these parameters are also stored in CEFS and provide permanent documentation for each forecast run. Each of the RTS, Growth, and Temporal Data Modules present two methods for data entry: 1) screen-based input and 2) batch file based input. Batch file based input was designed for use by ARB staff while screen based input was designed primarily for district users. Data entry screens were designed to allow data to be entered in a user friendly manner while performing validation checks wherever necessary. Further, the screens enforce all mandatory fields. Batch file based input expects the user to provide data in a predefined format.

## Processing

The main processing, done in CEFS, is by the Forecasting Module. Data are also processed for generating QA/QC reports on growth and control data as well as for generating standard system reports. The forecasting process generates entries for the trend database and also provides an input into the GIS system. It relies on a series of steps that include reading the base year data from CEIDARS, calculating the temporal factor, speciation factors, growth data generation, control data generation, emission forecast calculation for the trend database and GIS system, and finally storing forecast data into the Trend Database and GIS input files (Note: The Trend and GIS forecast modules are independent processes which are invoked with a switch in the Input Parameter Module).

## Reporting

As mentioned above, CEFS has a separate report module with seventeen built-in standard reports that are generated from the Trend Database. These reports include emission trends summarized by major source category—these standard summary reports offer the user flexibility to report “growth-only” or “control-only” scenarios as well. There are also several special analysis reports (See the sample report shown in figure 6). Reports can also be generated for growth and control data.

## System Administration and Security

System Administration is an important function in all information systems. CEFS contains a system administration module that allows the administrator to manage user accounts and corresponding levels of data access.  To make the task easier, administrators can create groups with specific data access profiles and then add users to suitable groups.  For example, district users can make changes to their own district's data but not to the data for other districts.  ARB users can, however, make changes to all data.

**System Maintenance**

Every information system is a living system.  Requirements, hardware, software, procedures, etc. are always in a state of change.  Thus, there is always a need to modify the system.  For example, CEFS has already seen the need for modifications because of several reasons:  changes to user requirements (most minor changes were accommodated during system development while major ones were set aside for a subsequent enhancement project), changes to the hardware (from DEC to SUN), changes to the system software (from ULTRIX to SunOS), and changes to the data that are input into CEFS (CEIDARS is now CEIDARS-II).

Modifications require a significant commitment of resources and it is therefore useful to plan for the same at the beginning of the project.  In developing CEFS, these resources were planned for the first year after system completion.  Further changes are foreseen but have not been planned at the present time.

**CEFS Location**

CEFS is a stand-alone application operating on a SUN SPARCserver 1000  using the SunOS 5.5 UNIX operating system and is located at Teale Data Center in Sacramento, California.  The main forecasting processing module is written in C.  Auxiliary reporting programs are written using a mix of C and PL/SQL.  CEFS has been in production since September 30, 1997.

## TESTING RESULTS AND CUSTOMER ACCEPTANCE

CEFS was delivered to the ARB in October 1995.  The original design defined in the Customer Requirements Specifications called for the application to be operated on a DEC5000 using the Ultrix operating system.

**Phase I: DEC 5000 Machine**

With the size and complexity of the application turning out to be much greater than originally conceptualized, the level of effort required to perform the acceptance testing was accordingly much greater than planned.  ARB staff conducted preliminary review of all CEFS modules (i.e. inputs, data processing, and output modules) over the first three months following delivery of the application.  Several iterations of testing/comments/corrections had to be made before CEFS was ready to perform a "real" test case scenario.  This was particularly true of the TREND module which had many systemic problems which had to be identified and corrected.  In order to stress the system to a "production-level" workload, it was decided to run a test scenario on the South Coast Air Basin (SCAB) in attempt to replicate the emission projections ARB developed for the 1994 South Coast Air Quality Management Plan—which was based on similar forecasting logic applying source specific growth and control factors at the SCC/SIC level.  This test involved significant front-end work in converting the growth and control data into the new CEFS data formats, but a successful test on SCAB would provide a high confidence level that the system would be able to handle just about any forecasting task.  The test was successful and the TREND forecast module was approved by ARB in February 1996.

Over the next six months, ARB staff conducted testing on each input data module (i.e. Growth, Rule Tracking System, and Temporal), Report module, and the GIS forecast processing module which performs the more intricate day/hour specific emission projections.  In addition, the system administration functions of CEFS (e.g. snapshot procedures, user account security etc.) all had to be tested and corrected.  Although the SCAB test (described above) served as a good test in that the database suit was very large, the test was relatively simple in terms of the logical links set up between the growth, control, and emissions data records.  As mentioned in the Methodology and Algorithm Design section, CEFS is capable of performing much more complex layering of growth and control schemes (i.e. emission category and region selection hierarchy) that were not used in the SCAB test. This logic needed to be validated in isolated tests by ARB staff. In July 1996, ARB approved the CEFS/DEC5000 application for acceptance.

**Phase II: SUN SPARCserver Machine**

The CEFS application never had the notoriety of performing “production-level” work on the DEC5000 machine. Within weeks after approving the application, the ARB was informed by the Teale Data Center that the DEC5000 machine would be taken out of production in the first quarter of 1997. Therefore, CEFS would need to be migrated to the replacement system—a SUN SPARCserver 1000 which also operates in a UNIX environment (but a different version). Any person who has attempted to move a Windows application from one PC to another without executing the installation programs knows that sometimes it works, and other times it doesn’t. Now imagine doing it with two “slightly” different operating systems. Now you have the picture! Fortunately, the modifications required to CEFS were relatively minor and the migration was successful. It took approximately two months of work by CSUF staff to make the system operational. The application was delivered to the ARB in March 1997 for testing. Each test which was conducted on the DEC5000 machine had to be repeated on the SUN system to verify functionality. As of July 1997, with the exception of a few standard report programs, the SUN version of CEFS was fully tested and is in production. The remaining work on the CEFS application (e.g. debug reporting programs etc.) was placed on hold due to a reprioritization of work to make more design changes as described below.

**Phase III: CEFS-II / CEIDARS-II**

In parallel with the testing of the SUN version of CEFS, ARB and CSUF staff had to gear up for a software design change-order to accommodate the recent redesign of CEIDARS. CEIDARS was recently redesigned to allow a merger of the criteria and toxic emission inventories into a single system. The new base year emission inventory system would be named CEIDARS-II and CEFS would have to be modified accordingly (the new application will be known as CEFS-II). These design changes were defined by ARB and conveyed to CSUF staff in April 1997. The CEFS-II application is nearing completion and this third phase of testing began in September with a target date for full production of November 1997.

## FUTURE STUDY AND ENHANCEMENTS TO CEFS

CEFS is a complex system with many users supplying input data and many clients with unique data product demands (approximately 200 special forecast data requests are processed each year for ARB’s clients). Although CEFS is in production, the ARB staff is planning to perform several modifications to meet client needs. The following enhancements are currently being considered:

**Expansion of SIP Tracking Capabilities**

From a data processing standpoint, CEFS is fully capable of tracking the progress of the California SIP. No changes to the forecast processor logic are anticipated. Up until now, the emission inventory staff at ARB have been responsible for executing baseline emission projections (accounting for “adopted” control measures only). As discussed in the Methodology and Algorithm Design section, CEFS has the necessary logical switches built-in to account for “proposed” measures—enabling “what-if” planning scenarios to be run. In order to enable ARB planning staff the ability to manage proposed measure inputs and link them to the emission inventory, we anticipate a need to modify the RTS user interface enabling planners to add and update “proposed” control measure data and SIP commitments in CEFS. These data will then be used to run special planning forecast scenarios by emission inventory staff. In addition to the input user interface modification, there are some special reporting programs which will be developed for assessing attainment demonstrations.

**District End User Enhancements**

Instruction on CEFS has been given to the districts via annual emission inventory workshops, however, the use of CEFS by district personnel has been limited due to the unforeseen design changes over the course of the project (i.e. the migration of CEFS to a new platform, the contract change orders to accommodate CEIDARS II etc.). We are now at a point where emphasis will be shifted to hands-on training of district end-users. As districts use CEFS, unique needs will surface and CEFS will need to be enhanced to meet these needs. It is still too early to know the extent of the enhancement. The comments received to date have been positive.

**Quarterly Seasonal Inventories**

Ozone planning inventories are typically developed for a six-month period running from May to October and CO planning inventories are developed for the period November to April. CEFS prepares seasonal inventories based on these planning periods. During the development of the 1997 $PM_{10}$ SIPs in California, it became apparent that quarterly and monthly planning periods can be more appropriate due to the high precipitation during certain months of the year. The modification of CEFS to accommodate monthly and quarterly planning periods will be a minor modification. (Note: CEFS can develop (day/hour specific) inventories for episodic conditions using the GIS forecast module but this is not for planning inventory preparation—it is for preparing modeling inputs.)

**Replacement of SCC Code System**

The ARB has developed an alternative coding system to replace the EPA's Source Classification Code (SCC) system[12]. The SCC codes are considerably lacking in detail and definition. If the proposed replacement code system is implemented on a national level, CEFS will have to be modified to accommodate this new system.

## CONCLUSIONS

The CEFS application is a comprehensive emission analysis system. CEFS serves as a central hub for the collection, management, processing, and reporting of all emission forecast related data, planning inventories, and day/hour specific emission inventories at the ARB.

With the newly developed rule tracking system, growth, and temporal data management models at its core along with a completely revamped forecasting algorithm, CEFS offers a much improved algorithm for predicting future emission inventories. As such, it will provide a sound basis for assessing progress of California's SIP and attainment demonstrations. The application also provides a seamless interface with CEIDARS which was a principle objective of this project. The application is a multi-user application which can be accessed remotely by other ARB staff and district staff via modem or telnet connection thus satisfying another key design objective of opening up the forecasting program to outside users. With these design objectives met, the program will serve as a valuable tool for emission inventory and planning staff for many years to come.

When the conceptualizing of CEFS began with the development of the rule tracking system in 1991, there was a subtle warning given by caring on-lookers that this program would become a monster. That wisdom was heeded. In the process of designing such a system to meet the ever demanding and changing needs of ARB's clients, we soon found that the warning was accurate. The question we often found ourselves asking was: Are we trying to do too much with this system? In hindsight, we can say without hesitation: No! The requirements which were anticipated at the outset of the project are in fact a reality now and are met by the system. In fact, more enhancements are on the horizon. When considered on the whole, we approached this project with an abandon, but each step was carried out with a certain sense of caution. The success of this project can be largely attributed to the early recognition of the complexity and breadth of the task and the need to work as a team. CEFS is the product of that team.

**ACKNOWLEDGMENTS**

The design and development of CEFS was a joint effort of many people. We want to thank those who contributed vitally to this project. Without your contributions, this project would never have succeeded. Thank you all!

**Air Resources Board Staff**

Technical Support Division

Terry McGuire
Gary Agid
Bertha Bache

Emission Inventory Branch

Linda Murchison
Dennis Goodenow
Dale Shimp
Vivian Lerch
Yosh Yajima
Patricia Velasco
Cheryl Taylor
Andy Delao
Jon Taylor
Mena Shah
Don Rake
Skip Campbell
Vijay Bhargava
Beth Schwehr
Chris Nguyen

Information Systems Branch

Bill Welty
Harry Ng
Louise Moy Woo

Administrative Services Division

Linda Quider

APCD and AQMD emission inventory staffs

**California State University Fullerton Staff**

Anil Puri
Ram Singhania
Hemant Kumar

## DISCLAIMERS

The opinions, findings, and conclusions expressed in this paper are those of the staff and not necessarily those of the California Air Resources Board.

## REFERENCES


1. Bickett, C., California Air Resources Board, *"Redesign of the California Emission Inventory System"*, paper presented at the Emission Inventory International Specialty Conference, October 1993.

2. Johnson, M., Rake, D., California Air Resources Board, *"Tracking the Effects of Rules on Emissions in California"*, paper presented at the Emission Inventory International Specialty Conference, October 1993.

3. Lerch, V., Yajima, Y., California Air Resources Board, *"The California Air Resources Board's Emission Forecasting System"*, paper presented at the Emission Inventory International Specialty Conference, October 1993.

4. Bickett, C., Johnson, M., California Air Resources Board, *"Proposed Modification to ARB's Seasonal Inventory Algorithm"*, Internal documentation on the methodology employed by CEFS to calculate temporal factors for seasonal emission inventory preparation, March 1995.

5. Bickett, C., Johnson, M., California Air Resources Board, *"Proposed Modification to ARB's Daily/Hourly Emission Inventory Algorithm and File Preparation Methods for GIS Emission Gridding Model"*, Internal documentation on the methodology employed by CEFS to calculate daily and hourly emission inventories, May 1995.

6. Schwehr, B., California Air Resources Board, *"Latest Definitions of VOC and ROG"*, Technical Support Division Fact Sheet, September 5, 1997.

7. DRI/McGraw-Hill, *"A Study to Develop Statewide and County-Level Economic Projections"*, contract 92-326, 1993.

8. Desert Research Institute, *"1997 Southern California Ozone Study (SCOS97)--Draft Field Study Plan"*, contract 93-326, 1996.

9. California State University, Fullerton, *"Customer Requirement Specification (CRS) Document for CEFS"*, June 1995.

10. California State University, Fullerton, *"High Level Design (HLD) Document for CEFS"*, June 1995.

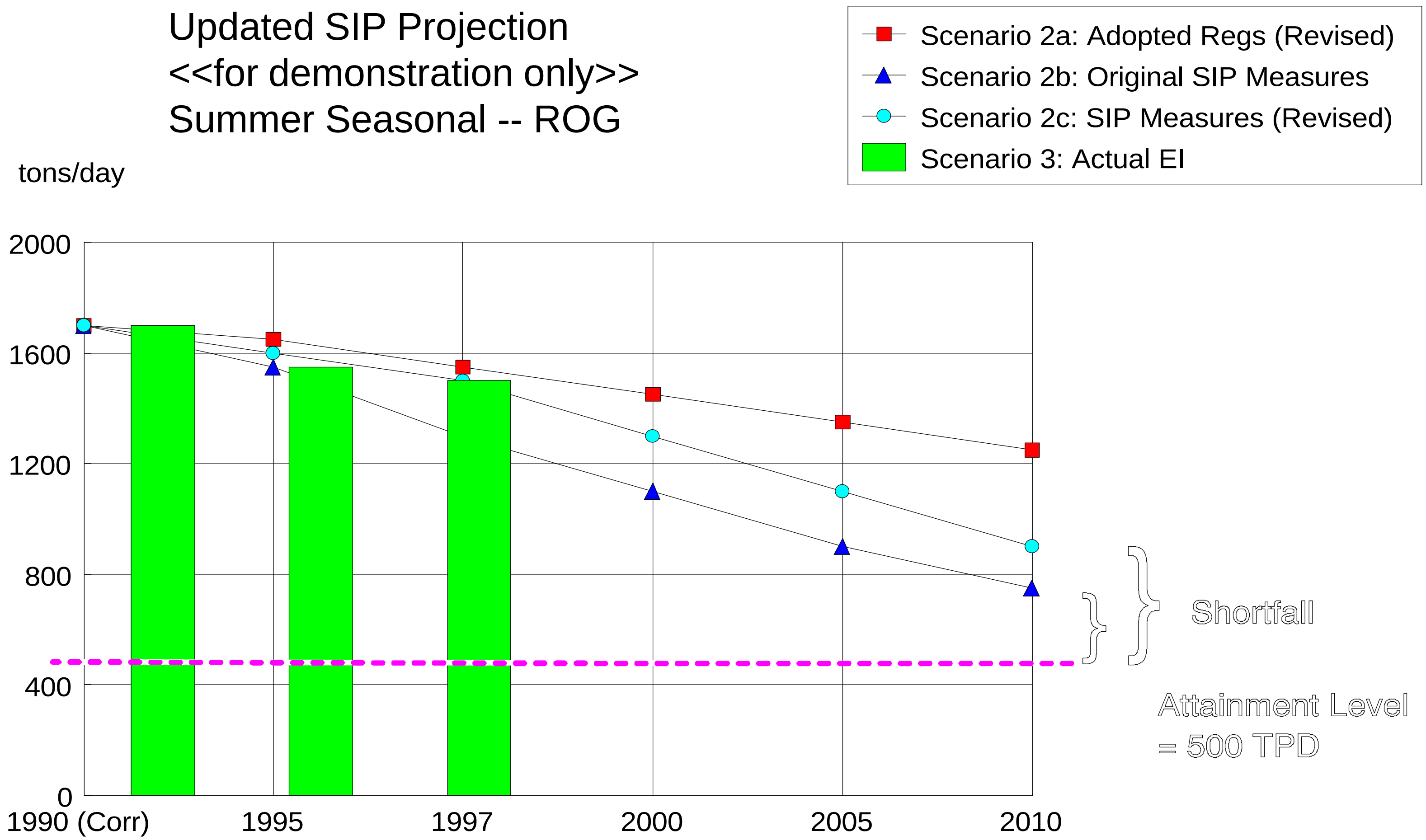
11. California State University, Fullerton, *"Detailed Design Document for CEFS"*, June 1995.

12. Nitsos, A., *"New Coding System to Replace the Source Classification Codes (SCC's)"*,paper presented at the Emission Inventory International Specialty Conference, October 1997.

**KEY WORDS**

California Emission Forecasting System
Rule Tracking
Seasonal and Planning Emission Inventories
Information Management Systems
Computer Tools and Database Design

# Figure 1: SIP Emission Analysis




Reporting Parameters:
REGION
POLLUTANT
SOURCE_CAT
YEAR

* The 1990 inventory is reconstructed to account for methodology improvements.

**Figure 2: Growth and Control Data Hierarchy**

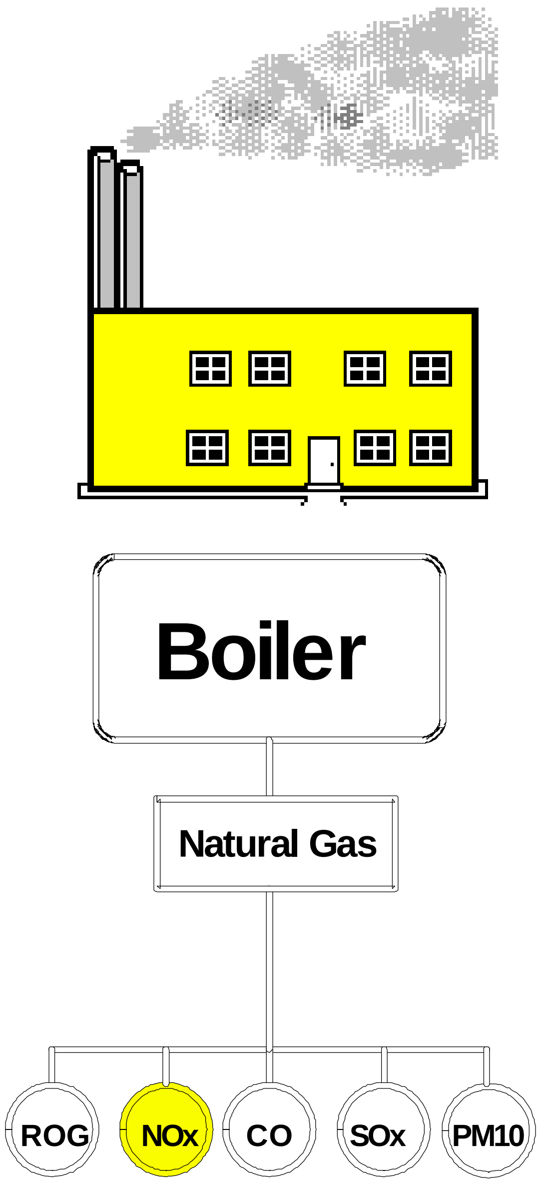
Region Selection:

1. District, Air Basin, County, Sub-County
2. District, Air Basin, County
3. Air Basin, County
4. County
5. Air Basin
6. District
7. California

Category Selection:

1. Facility, SCC, SIC
2. Facility
3. Facility, EIC
4. SCC, SIC
5. SCC6, SIC
6. SCC3, SIC
7. SCC
8. SIC
9. EIC, SIC
10. EIC
11. CES
12. SIC2
13. Facility, EIC, SIC

# Figure 3: Control Data Layering

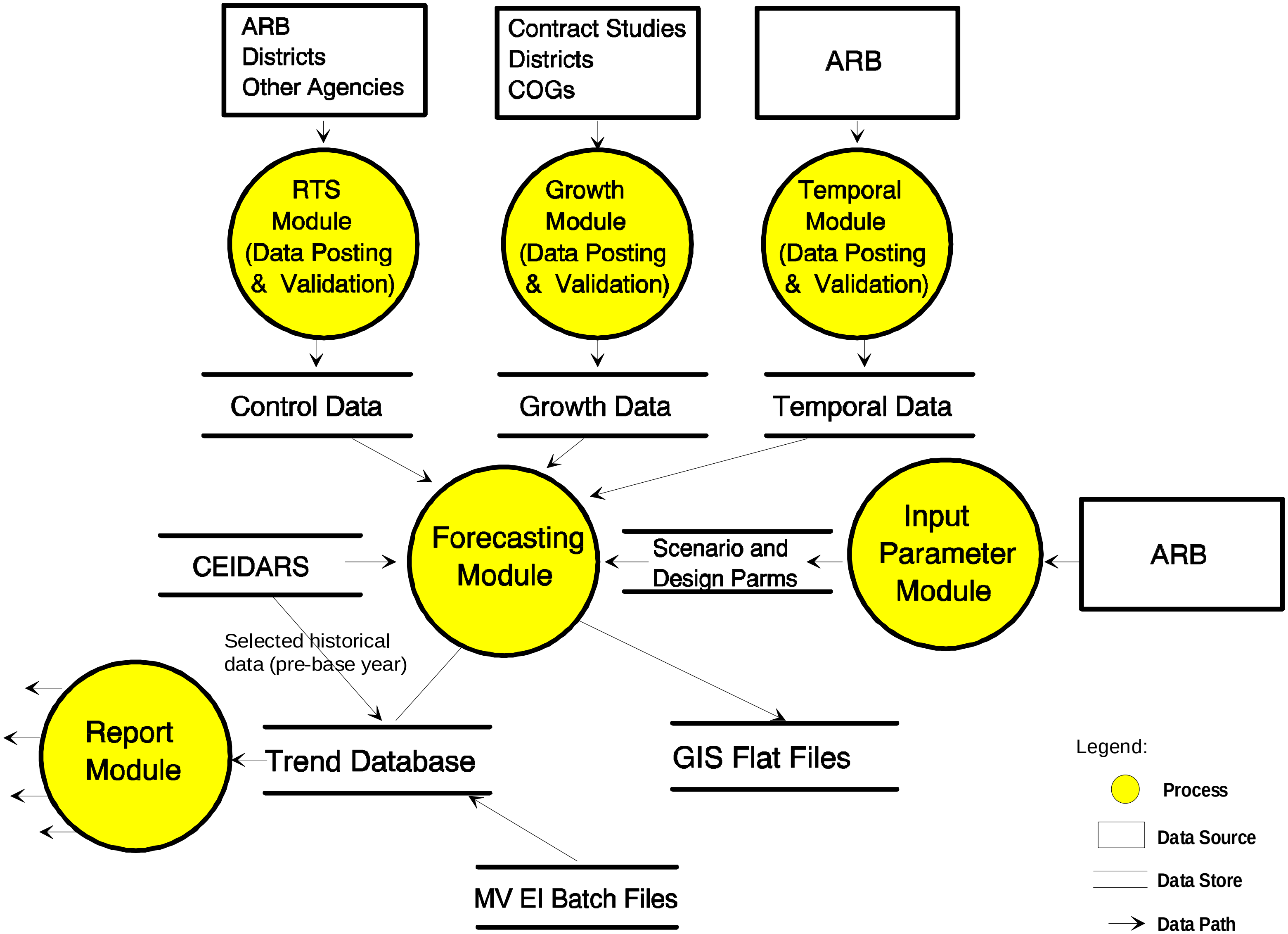



**Rule X places NOx limits on natural gas fired boilers and heaters > 100 MMBTU/HR (SCC: 10300601)**

| SCC | SIC |
|---|---|
| 10300601 | 3678 (Electronics) |
| | 4961 (Steam Supply) |
| | 8062 (Medical Services) |
| | 8211 (Elementary Schools) |
| | 8221 (Universities) |
| | 9711 (National Security) |

- Category Selection “Option 7” used if rule applies to the entire universe of SCC: 10300601.
- Category Selection “Option 4” used if the rule targets a particular industry sector more or less stringently than the other industries.

# Figure 4: CEFS Context Flow Diagram



# Figure 5: CEFS User Access