CONTRACT NO. A933-157
FINAL REPORT
JANUARY 1998

Development of a Model for Assessing Indoor Exposure to Air Pollutants

CALIFORNIA ENVIRONMENTAL PROTECTION AGENCY

AIR RESOURCES BOARD
Research Division

DEVELOPMENT OF A MODEL FOR ASSESSING INDOOR EXPOSURE TO AIR POLLUTANTS

Final Report A933-157

Prepared for:

California Air Resources Board
Research Division
2020 L Street
Sacramento, CA 95814

Prepared by:

Michael D. Koontz
William C. Evans
Charles R. Wilkes

GEOMET Technologies, Inc.
20251 Century Boulevard
Germantown, Maryland 20874

GEOMET Report No. IE-2631

January 1998

For more information about the ARB's Research Division,
its research and activities, please visit our Web site:

http://www.arb.ca.gov/rd/rd.htm

DISCLAIMER

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

ACKNOWLEDGMENTS

The primary author of this report, Michael Koontz, served as co-investigator for the project (with Niren Nagda), developed the conceptual design for the model software, and conducted the model validation effort. The second author, William Evans, developed the Monte Carlo simulation routines for the model, the routines for sampling values from various types of distributions, and the mass-balance calculation routine. The third author, Charles Wilkes, assisted in the review of existing models and description of the mass-balance calculation routines. The success of this project was the product of the authors' work and the efforts of a number of other individuals. The authors wish to thank all those who contributed to the design, implementation, analysis, and reporting of this study.

Overall project management was provided by Ms. Susan Lum of the California Air Resources Board with the assistance of Ms. Peggy Jenkins, Manager of the Indoor Exposure Assessment Section at the California Air Resources Board.

Various staff members of GEOMET Technologies, Inc., provided programming, technical and word processing support during the course of the project. Sang Lee developed the user interface for the model as well as methods for saving/accessing user inputs and passing all inputs to computational routines. Lai Huynh developed the data base on activity profiles and the routine for selecting profiles that match user inputs, and also provided general programming support for computational programs and their outputs. Harry Rector provided useful suggestions at various points in the project and graphics support for the final report. Leon Pocinki was primarily responsible for the literature review. Laura Niang assisted in preparation of model inputs. Word processing support was provided by Joanne Bittner, Joy Warren, Glenda Heredia, Marilyn McLaughlin, and Kris Hinkley.

This report is submitted in fulfillment of California Air Resources Board Contract No. A933-157, "Development of a Model for Assessing Indoor Exposure to Air Pollutants," by GEOMET Technologies, Inc., under the sponsorship of the California Air Resources Board. Work on this research project was completed as of January 1998.

ABSTRACT

The primary mission of the ARB Indoor Program is to identify and reduce Californians' exposures to indoor pollutants. To fully consider indoor exposures in assessing risk, the ARB needs estimates of average and peak indoor exposures for the general California population as well as certain subgroups of that population such as individuals who may be highly sensitive to indoor pollutants. The model described in this document--California Population Indoor Exposure Model (CPIEM)--is a software product that has been designed to expedite the exposure-assessment process by providing a user interface and calculation tools for supplying and integrating all required information.

The primary function of the CPIEM software is to combine indoor-air concentration distributions with Californians' location/activity profiles to produce exposure and dose distributions for different types of indoor environments. This function, referred to as Level 1-2 of the model, is achieved through a Monte Carlo simulation whereby a number of location/activity profiles that were collected in prior ARB-sponsored surveys are combined with airborne concentrations for specific types of environments (e.g., residences, office buildings). For many compounds, the concentration data are either limited or nonexistent. Consequently, a second function of the model (Level 3) is to estimate indoor-air concentration distributions based on distributional information for mass-balance parameters such as indoor source emission rates, building volumes and air exchange rates.

The CPIEM software has several unique features:

- Ability to estimate distributions of Californians' inhalation exposure and potential inhaled dose, with accumulation across multiple indoor environments
- Estimation of exposure/dose distributions using concentration distributions that are measured through field studies or simulated within the model itself
- Dynamic modeling of hourly and daily indoor-air concentrations in an indoor environment, taking into account various types of indoor sources as well as outdoor concentrations, air exchange rates, and losses to indoor sinks.

- Capability within the model for saving inputs and for future expansion, allowing the repository of input data to grow as new information becomes available.

CPIEM was subjected to various verification and evaluation efforts under this project. Verification steps for Level 1-2 indicated that inputs (activity profiles, concentration distributions) are properly accessed and used by the model, and that exposures and doses are correctly computed and accumulated across locations. Model estimates were evaluated utilizing data from an NO_2 study in Los Angeles. The average personal exposure estimated from the simulation agreed closely with that based on the field study, but the standard deviation of the modeled exposure distribution was lower than that for the field study. This downward bias was expected because of the limited information available for constructing concentration inputs for environments other than the residence.

Estimates from Level 3 of the model were evaluated for three pollutants--chloroform, benzo[a]pyrene (BaP) and nitrogen dioxide (NO_2)--for which sufficient source-related information and field studies for comparison were both available. The modeled standard deviation for chloroform, relative to the mean, initially was lower than that for field measurements. This is believed to be due to initial description of each type of water use--toilets, faucets, showering/bathing, dishwashing and clothes washing--as a separate source in the model. When the chloroform sources were treated in this manner, the model sampled a different emission rate for each type of water use in a given residence, whereas the emission rates are likely to be very similar. When these sources were combined in order to use a common emission rate within each residence, the modeled ratio of the standard deviation to the mean better reflected the ratio based on measurements. There was a similar finding for NO_2--the mean of the modeled distribution matched the measurements well but the modeled standard deviation initially was low relative to measurements. When separate sources that were initially defined for cooking breakfast, lunch and dinner were subsequently combined into one source, the modeled concentration distribution matched the measurement data very well. In both cases--chloroform and nitrogen dioxide--the combined sources accounted for the same consumption (liters of water for chloroform, Btu of fuel for nitrogen dioxide) as the individual sources when

aggregated. These modeling outcomes suggest that similar types of indoor sources should be combined whenever possible.

The principles on which both components of the model are based are scientifically and mathematically sound, but the accuracy of the outputs is limited by that of the inputs. Data on concentration distributions, needed for Level 1-2 of the model, are not yet available for many of the environments. Even in cases where measured concentrations are available, there can be inaccuracies due to biases inherent in monitoring devices or sampling strategies. For Level 3 of the model, there is a notable lack of information at present for many of the indoor sources as well as pollutant-specific penetration factors and decay rates.

TABLE OF CONTENTS

		Page
1.0	INTRODUCTION	1-1
	1.1 Background and Objectives	1-1
	1.2 Review of Existing Models	1-2
	1.3 Estimates Provided by the ARB Model	1-7
	1.4 This Report	1-9
2.0	SUMMARY AND CONCLUSIONS	2-1
3.0	RECOMMENDATIONS	3-1
4.0	MODEL REQUIREMENTS AND STRUCTURE	4-1
	4.1 Model Requirements	4-1
	4.2 Model Structure and Capabilities	4-2
5.0	EXPOSURE AND DOSE DISTRIBUTIONS	5-1
	5.1 Modeling Approach	5-1
	5.2 Model Inputs	5-9
	5.3 Key Algorithms	5-20
	5.4 Assumptions and Limitations	5-24
6.0	CONCENTRATION DISTRIBUTIONS	6-1
	6.1 Modeling Approach	6-1
	6.2 Model Inputs	6-3
	6.3 Key Algorithms	6-18
	6.4 Assumptions and Limitations	6-29
7.0	MODEL VERIFICATION AND EVALUATION	7-1
	7.1 Exposure/Dose Calculations	7-1
	7.2 Concentration Calculations	7-8
	7.3 Discussion of Evaluation Results	7-27
8.0	REFERENCES	8-1

Appendix A SOFTWARE REQUIREMENTS DOCUMENT
Appendix B ACTIVITY CODES
Appendix C BIBLIOGRAPHY RESULTING FROM LITERATURE REVIEW
Appendix D MODEL VALIDATION PLAN

LIST OF FIGURES

		Page
4-1	Conceptual Hierarchy for the Model	4-3
5-1	Overview of Exposure/Dose Calculation Routine	5-2
5-2	Environments/Activities Loop for 24-Hour and 12-Hour Integration Periods	5-5
5-3	Environments, Activities and Hours Loops for 8-Hour and 1-Hour Integration Periods	5-7
6-1	Overview of Concentration Calculation Routine	6-2
6-2	Routine for Sampling and Using Input Parameters for Long-term Sources	6-7
6-3	Routine for Sampling and Using Input Parameters for Episodic Sources	6-9
6-4	Routine for Sampling and Using Input Parameters for Frequent Sources	6-12
6-5	Episodes Loop for Frequent Sources	6-13
6-6	Schematic Representation of Single Compartment Indoor-air Model	6-19
6-7	Flowchart of Routine for Calculating Hourly-Average Indoor Concentrations Over a 48-Hour Period	6-25
7-1	Histogram of Values Sampled from a Normal (75, 25) Distribution	7-3
7-2	Histogram of Values Sampled from a Lognormal (75, 50) Distribution	7-4
7-3	Histogram of Values Sampled from a Triangular (0, 75, 200) Distribution	7-6

LIST OF TABLES

		Page
2-1	Inputs Required for Each Level of the Model	2-3
5-1	Location Codes and Percent of Time Associated with Nine Types of Environments	5-10
5-2	Specific Types of Locations Associated with Each Type of Environment	5-11
5-3	Activity Levels Used for Dose Calculations	5-13
5-4	Assignment of Activity Codes to Activity Levels	5-13
5-5	Inhalation Rates Assigned to Four Activity Levels for Three Population Subgroups	5-15
5-6	Current Sources of Information on Concentration Distributions for Selected Pollutants and Indoor Environments	5-16
5-7	Summary of Concentration Data Available for Model Level 1-2	5-17
5-8	Ways in Which Concentration Distributions for Various Locations Can Be Described as Inputs to Exposure/Dose Calculations	5-18
5-9	Summary of Exposure and Dose Calculations	5-21
6-1	Examples of Indoor Sources for Pollutants Provided with the Model	6-5
6-2	Inputs for Long-term Sources	6-5
6-3	Inputs for Episodic Sources	6-8
6-4	Dependence of Calculations for Episodic Sources on Rate of Decline in Emission Rate and Duration Since Use	6-10
6-5	Inputs for Frequent Sources	6-11
6-6	Current Sources of Information on Outdoor Concentration Distributions for Selected Pollutants	6-14
6-7	Current Sources of Information on Indoor Volumes and Air Exchange Rates	6-17
7-1	Comparison of Personal Monitoring Measurements and Model Outputs for Nitrogen Dioxide (in $\mu g/m^3$)	7-7

LIST OF TABLES (continued)

		Page
7-2	Summary Statistics on 24-hour-average NO_2 (in $\mu g/m^3$) for Individual Model Runs	7-8
7-3	Summary of Existing Information on Rates of Water Use and Volatilization Coefficients	7-9
7-4	Inputs for Five Types of Indoor Water Use Modeled as Frequent Sources	7-11
7-5	Time-of-Day Distribution (%) for Five Types of Indoor Water Use	7-12
7-6	Model Inputs for Outdoor Concentrations, Penetration Factors, Indoor Sinks, Volumes, and Air Exchange Rates	7-13
7-7	Comparison of Field Measurements and Initial Modeling Outputs for Chloroform (in $\mu g/m^3$)	7-14
7-8	Model Inputs for Combined Chloroform Sources	7-15
7-9	Time-of-Day Distribution (%) for Combined Chloroform Sources	7-15
7-10	Comparison of Field Measurements and Refined Model Outputs (Combined Sources) for Chloroform (in $\mu g/m^3$)	7-17
7-11	Summary of Model Inputs for Benzo[a]pyrene	7-19
7-12	Comparison of Field Measurements and Model Outputs for Benzo[a]pyrene (in ng/m^3)	7-19
7-13	Summary Statistics for Individual Model Runs for Benzo[a]pyrene (in ng/m^3)	7-20
7-14	Summary of Model Inputs for Cooking Meals	7-21
7-15	Model Inputs for Outdoor Concentrations, Penetration Factors, Indoor Sinks, Volumes, and Air Exchange Rates	7-22
7-16	Comparison of Field Measurements and Initial Modeling Outputs for Nitrogen Dioxide (in $\mu g/m^3$)	7-23
7-17	Summary of Model Inputs for Combined Cooking Events	7-24
7-18	Comparison of Field Measurements and Refined Modeling Outputs for Nitrogen Dioxide (in $\mu g/m^3$)	7-25
7-19	Summary Statistics for Individual Model Runs for Nitrogen Dioxide (in $\mu g/m^3$)	7-26

GLOSSARY

Air exchange rate -	the rate at which air is exchanged between the airspace in an indoor environment and the surrounding outdoor airspace, expressed as the indoor/outdoor airflow rate (m^3/h) divided by the indoor volume (m^3), with resultant units of inverse time (1/h).
Breathing rate -	the rate at which an individual breathes or inhales air, expressed in units of volume/time (e.g., m^3/h); also called inhalation rate.
Concentration -	the extent of occurrence of a pollutant in air, expressed in terms of pollutant mass per unit volume of air (e.g., $\mu g/m^3$), or as parts of the pollutant per billion parts (ppb) or per million parts (ppm) of the air-pollutant mixture, both by volume.
Conservation of mass -	a principle stating that decreases in pollutant mass in a defined airspace (e.g., indoors) are equally compensated by corresponding increases in mass in other airspaces (outdoors) or media (indoor sinks) to which the pollutant is transported.
Decay rate -	the first-order rate of reduction in the indoor-air concentration of a pollutant, due to physical/chemical reactions in air or with indoor sinks, expressed in units of inverse time (e.g., 1/h).
Dose (inhaled) -	the integral over time of the product of a pollutant concentration times a breathing rate, expressed in units of mass (e.g., μg).
Emission rate -	the rate at which an indoor source emits a pollutant into the indoor airspace, expressed in units of mass/time (e.g., $\mu g/h$).
Environment -	a type of place or building where an individual spends time, such as a residence, school or office building.
Exposure -	the contact at one or more boundaries (e.g., mouth or skin) between a human and a pollutant at a specific concentration for a period of time; the three principal routes of exposure are inhalation of air (the subject of this report), ingestion of food or liquids, and dermal contact.
Indoor-air model -	an equation, algorithm, or series of equations/algorithms used to calculate the average or time-varying pollutant concentration in an indoor environment for a specific situation.

Indoor sink -	a material, furnishing or appliance used or installed indoors that causes the indoor-air concentration of a pollutant to decrease by processes such as absorption, adsorption or chemical reaction.
Indoor source -	a product, material, appliance or activity indoors that causes the indoor air concentration of a pollutant to increase.
Integration period -	the time period (e.g., hour or day) over which inhalation exposure or inhaled dose is mathematically integrated.
Location/activity profile -	an accounting of how an individual's time over a defined period (e.g., hour or day) is spent in various types of environments at various levels of activity or exertion.
Mass-balance equation -	a differential equation, based on the conservation of mass, stating that changes in pollutant mass indoors over time are related to gains (from indoor sources or from outdoor mass transported indoors) and losses (indoor mass transported to outdoors or to indoor sinks).
Mitigation -	an action or series of actions intended to reduce the concentration of a pollutant indoors.
Model evaluation -	a series of steps through which a model developer or user assesses a model's performance for selected situations.
Model parameter -	a mathematical term in an indoor-air model that must be estimated by a model developer or user before model calculations can be performed.
Model validation -	a series of evaluations undertaken by an agency or organization to provide a basis for endorsing a specific model (or models) for a specific application (or applications).
Model verification -	a series of checks to ensure that model logic has been correctly programmed and that model calculations are mathematically correct.
Monte Carlo simulation -	a stochastic process by which values are repeatedly sampled from distributions for various model parameters and used to estimate an indoor-air or exposure model; each repetition of the process is called a trial.
Penetration factor -	the fraction of the outdoor pollutant concentration that bypasses the outer envelope of an indoor environment and enters the indoor airspace.

Section 1.0

INTRODUCTION

1.1 BACKGROUND AND OBJECTIVES

Under Section 39660.5 of the California Health and Safety Code, the State of California Air Resources Board (ARB) is charged with assessing human exposure to toxic air contaminants in indoor environments and identifying the relative contribution of indoor exposures to total human exposure, taking into account both ambient and indoor air environments. The Indoor Air Quality/Personal Exposure Assessment Program (Indoor Program), established within ARB's Research Division in 1986 to investigate indoor and personal exposures to air pollutants, has the primary responsibility for obtaining and analyzing information related to this section of the code.

Indoor exposure assessment is a critical component of assessing total exposure to toxic air contaminants because (1) residents in California and other areas of the country spend a significant majority of their time indoors, and (2) for many pollutants, indoor-air concentrations are typically higher than those outdoors. Assessment of indoor exposures can be particularly complicated due to the considerable variety in (1) structure types and building-construction characteristics, (2) indoor sources such as appliances, construction and interior finishing materials, furnishings, and consumer products, and (3) occupant activities including movement patterns, uses of various types of sources, operation of space-conditioning equipment, and opening or closing of windows.

To fully consider indoor exposures in assessing risk, the ARB needs estimates of average and peak indoor exposures of the California population as well as estimates of indoor exposures for sensitive subgroups of that population. However, the quantity and type of data available for use in developing such estimates vary from compound to compound and are usually very limited. Personal exposure data, not necessarily restricted to time spent indoors, are available for some compounds whereas for others only indoor-concentration data or source-emissions data (or none of these) may be available.

Calculating indoor exposures for a given population involves the integration of information on individuals' location/activity patterns (i.e., how and where people spend their time) and concentrations encountered by these individuals at various indoor locations. If concentration measurements have not been collected in some of the locations of interest, then some estimate or prediction of the concentration distribution is also needed. In the past, ARB staff have gone through a fairly involved and time-consuming task of locating, developing and interfacing various pieces of information needed to support indoor exposure assessments. The objectives of the project described in this report were to develop and validate a software product (model) that can expedite the exposure-assessment process by providing a user interface and calculation tools for supplying and integrating all required information. Two key requirements of the software are that it be (1) user friendly and (2) reasonably accurate without being overly complex.

1.2 REVIEW OF EXISTING MODELS

The primary objective of the model is to produce estimates of exposure and dose distributions by combining indoor-air concentration distributions with location/activity profiles of California residents. In order to accomplish this objective, it was determined that the model should have the following three levels of capabilities:

- Level 1 - provide estimates of total indoor exposure distributions by aggregating environment-specific exposure estimates (see Level 2)
- Level 2 - provide estimates of indoor exposure distributions for specific indoor environments by combining indoor concentration data with population activity data
- Level 3 - provide estimates of indoor-air concentration distributions based on factors such as indoor source emissions, building volumes, and indoor-outdoor air exchange rates.

A review of existing models prior to development of this model revealed no models with Level 1 capabilities. The absence of this type of model is probably due to the relative paucity of personal exposure or concentration measurements that have been made in indoor environments other than residences. Most personal exposure measurements have been made for periods of 12 hours or longer with time-integrated samplers; thus, the

sampling results reflect contributions from multiple environments. It may be possible to deduce the average indoor contribution to the personal exposure estimates, but estimating the underlying distributions associated with different averaging times would involve a number of tenuous assumptions.

Several models with level 2 capabilities currently exist. Such models have been used predominantly to estimate population exposure distributions for ambient criteria pollutants such as carbon monoxide and ozone. Probably the best known such model is the Simulation of Human Activity and Pollutant Exposure (SHAPE) model, developed at EPA (Ott 1984). In brief, the model uses Monte Carlo sampling techniques to build exposure frequency distributions by combining population activity data with concentration distributions for different microenvironments.

The original SHAPE model simulated activity patterns for each individual by sampling from probability distributions for the chance of entry, time of entry, and time spent in different microenvironments. In more recent modeling exercises, observed activity data (times when participants entered and left each microenvironment) from EPA studies in Washington, D.C., and Denver (Akland et al. 1985) were used. The microenvironmental concentrations used as model inputs are segregated into a component due to sources within each microenvironment and an ambient background component (i.e., assuming no indoor sources are present). Although SHAPE was developed primarily for estimating total (24-hour) exposure distributions, it can be used to estimate similar distributions for indoor environments contributing to the total exposure.

The Benzene Exposure Assessment Model (BEAM) (Behar 1989) utilizes the basic SHAPE model for application to a different contaminant. Activity patterns from the Washington, D.C., and Denver studies by EPA, and national studies of time use conducted by Robinson, are being considered as one set of model inputs. The other primary inputs are microenvironmental concentrations. Background atmospheric concentrations are estimated through air dispersion modeling. Difficulties have been encountered in developing concentration distributions for microenvironments such as "indoors" or "in transit" due to the relative paucity of appropriate data. The most extensive data base

available, resulting from the TEAM study, consists of 12-hour personal exposure measurements that cannot be readily disaggregated into microenvironmental components.

A number of other "physical-stochastic" models exist that utilize Monte Carlo techniques to simulate human behavior, represent time-varying pollutant concentrations in various microenvironments, and address uncertainty in the knowledge of model parameters. The earliest of these models, the NAAQS Exposure Model (NEM) (Johnson and Paul 1983), was developed to estimate and analyze population exposures under current or proposed ambient air quality standards. Consequently, the model is "driven" primarily by ambient concentration data. NEM operates at a higher level of aggregation than SHAPE, simulating hour-by-hour movements of representative population groups (cohorts) through different exposure districts or neighborhoods and through different microenvironments, and accumulating the resultant exposure over a period of one year. Each cohort is identified by district of residence, district of employment where applicable, age-occupation group, and activity-pattern subgroup. Although noteworthy for its recognition and relatively sophisticated treatment of serially dependent data through an autoregressive model, the NEM treatment of indoor or in-transit microenvironments relies on relatively simple empirical adjustment factors.

More recent alternatives to NEM, such as the regional human exposure (REHEX) model (Lurmann et al. 1989) and the personal air quality model (PAQM) (Hayes 1989), are predominantly attempts to expand aspects such as the number of microenvironments or time-activity patterns or to provide a more refined use of certain data bases to generate model inputs. Like NEM, these models share the limitation that indoor sources of pollutants are not explicitly treated.

Thus, of the commonly used Level 2 models, only SHAPE and its spinoff (BEAM) explicitly incorporate indoor concentration data, and in the case of BEAM, suitable data are extremely limited. In the absence of exposure or concentration data to support the use of Level 1 or Level 2 models, the primary recourse is use of Level 3 models to predict indoor concentrations for various situations defined by factors such as building dimensions, indoor source activity, outdoor concentrations, and indoor-outdoor air exchange rates.

Models with Level 3 capabilities include (1) the Computerized Consumer Exposure Models (CCEM) developed by Versar (1987) for the EPA Office of Toxic Substances, (2) the CONTAM model (Axley 1988), developed at the National Institute of Standards and Technology (formerly the National Bureau of Standards) with support from EPA and the Department of Energy, (3) the INDOOR model (Sparks 1988) developed by the Indoor Air Branch of EPA's Air and Energy Engineering Research Laboratory in Research Triangle Park, since updated as the EXPOSURE model (Sparks et al. 1993), (4) the Multichamber Consumer Exposure Model (MCCEM) developed by GEOMET (Koontz and Nagda 1991) for EPA to address some of the limitations of CCEM, and (5) the MIAQ model developed at the California Institute of Technology (Nazaroff and Cass 1989).

With the exception of MIAQ, all models were developed for use in an IBM-PC/compatible environment. The models differ significantly in user-friendliness and aspects such as (1) the number of indoor zones/spaces, contaminants, and sources that can be handled in a single run and (2) the level of sophistication with which release mechanisms, time-varying emission profiles, sink effects, and infiltration/interzonal airflows can be treated. It must be recognized that the relative strengths or weaknesses of these models are strongly tied to their originating needs. Unlike many other applications, indoor air quality models typically are not software products that can be purchased as "off-the-shelf" items. Rather, most existing software models are research tools that were originally developed for a specific purpose and are in a near-continual stage of development and refinement. Despite differences in level of sophistication and relative strengths and weaknesses, the similarity in fundamental underpinnings of the models developed for the IBM-PC environment is evidenced by the virtually identical results they produce for a relatively simple single-chamber situation (Koontz and Rector 1989).

The models listed above were generally intended to estimate indoor concentration profiles and/or associated indoor exposures for a specific situation or exposure scenario. As such, they were not developed in anticipation of ARB's current need to model the distribution of population exposures for specific types of indoor environments. MCCEM does have the capability (as does EXPOSURE) to address indoor exposure by "assigning" an individual to different zones of a building at different times, but this capability was

provided to enable calculation of exposure for specific scenarios rather than a wide variety of situations. Similarly, MCCEM provides Monte Carlo and sensitivity options, but these tools were provided primarily as an aid to uncertainty analysis for a specific situation rather than estimation of an exposure distribution across many situations (i.e., different indoor environments and factors affecting indoor concentration profiles).

A model that was developed to simulate concentration distributions across residential environments is the LBL "macromodel" (Traynor et al. 1989). The initial focus of the model has been on combustion pollutants such as carbon monoxide, nitrogen dioxide, and respirable suspended particles, for which the underlying mathematical relationships are reasonably well understood and sufficient input data exist. Key inputs to the model include relevant building-stock parameters such as house volume and air exchange rate, market penetration of indoor combustion appliances, and source-usage models. Monte Carlo simulation is used to create a housing stock of 3,500 residences described in terms of physical characteristics (e.g., volume, leakage area), the number and type of sources, source emission rates, and venting factors. Submodels are included for determining usage rates for heating appliances, calculating air exchange rates, and predicting indoor concentrations.

Although the "macromodel" has some features that are related to ARB's current needs, it also has some clear limitations. For example, the model addresses only combustion sources and pollutants. More importantly, the model uses a mass-balance equation with steady-state assumptions to predict average indoor concentrations over a one-week time interval. Although the steady-state assumption was made as a practical decision to decrease the complexity of the modeling effort, it represents a significant shortcoming vis-a-vis ARB's need to estimate concentration/exposure distributions for various averaging times so that both peak and average exposures can be addressed.

As evident from the above review, some models in use at the outset of the project individually provided some of the capabilities to meet ARB's specific needs. At the same time, however, none of the existing models come close to provide all the capabilities desired by the ARB, nor was it apparent that they could be readily adapted to do so. This

outcome is largely due to the fact that each existing model was developed with a specific purpose in mind that did not coincide with ARB's modeling objectives.

1.3 ESTIMATES PROVIDED BY THE ARB MODEL

The modeling software developed for ARB under this project, called the California Population Indoor Exposure Model (CPIEM), two major components--calculation of exposure/dose distributions and calculation of indoor-air concentration distributions. Inputs to the model are supplied through a user interface

The exposure/dose module is called the Level 1-2 module because it simultaneously calculates exposure and dose for all selected environments while providing estimates of "total indoor air" exposure and dose across these environments. Thus, Level 1-2 is a single module with all outputs produced at the same time; Level 1 is not a separately accessible function. Level 1-2 of the model uses measured or modeled concentration distributions for one or more environments, together with location/activity patterns (i.e., amount of time spent in each environment at specific activity levels), to calculate exposure and inhaled-dose distributions (Level 2) for the chosen environment(s). Concentration values for a given environment are sampled from user-specified distributions and multiplied by time durations in the environment, as sampled from surveyed Californians' location/activity profiles, to simulate time-integrated exposure. Multiplication by breathing rates determined from the location/activity profiles and pulmonary ventilation data yields an estimate of the potential-inhaled-dose distribution for each modeled environment. The model then aggregates the environment-specific exposure and dose estimates to develop distributions of "total indoor air" exposures and doses (Level 1), that is, the portion of total (24-hour) exposure/dose associated with time spent indoors. Because the outdoors is included as one of the environments for the model, it is also possible to simulate total (24-hour) exposure and dose distributions.

The concentration module (Level 3 of the model) utilizes a mass-balance equation, based on the principle of conservation of mass, to estimate concentration distributions for specific types of indoor environments such as residences, offices and schools. This module samples values from user-specified distributions for parameters such as emission

rates for indoor sources, building volumes, outdoor-air concentrations and indoor-outdoor air exchange rates, which are used as inputs to the mass-balance equation. The output from the concentration module can be used as one of the inputs to the exposure/dose module, and would be particularly useful for pollutant-environment combinations for which concentration data are not available from field monitoring studies.

The model can provide estimates of distributions for both inhalation exposure and potential inhaled dose. Inhalation exposure is defined as the time-integrated concentration encountered by an individual while in an indoor environment (National Academy of Sciences 1991; USEPA 1989). The mathematical form is given as:

$$C_T = \int_0^T C(t)\ dt$$

where C(t) is the concentration in the environment at time t, T is the amount of time spent in the environment, and C_T is the time-integrated concentration. If the concentration is measured in $\mu g/m^3$ and time in hours, then the units for C_T are $\mu g\text{-}h/m^3$. The average concentration in the environment, $\overline{C}_T$ ($\mu g/m^3$), is equal to C_T divided by T. $\overline{C}_T$ is not included in the summary statistics reported by the model; however, both C_T and T are included in the detailed results that are provided, enabling the user to calculate $\overline{C}_T$ if desired.

Potential inhaled dose is defined as "an exposure multiplied by rate and assumes total absorption of the contaminant" (National Academy of Sciences 1991; USEPA 1989). This can be mathematically represented as the product of the time-integrated concentration and the individual's breathing rate (i.e., amount of air inhaled per unit time while in the environment):

$$D_T = \int_0^T B(t)\ C(t)\ dt$$

where B(t) is the breathing rate at time t and D_T is the potential inhaled dose over the time duration T. If the breathing rate is assumed to be constant and this constant rate is expressed as $\overline{B}_T$, then the potential inhaled dose can be expressed as:

$$D_T = \overline{B}_T \int_0^T C(t)\, dt = \overline{B}_T \cdot \overline{C}_T \cdot T$$

If the breathing rate is in units of m^3/h and the units for $\overline{C}_T$ and T are as above, then D_T is expressed in μg (i.e., $m^3/h \bullet \mu g/m^3 \bullet h$). For the model, the average breathing rate while in the environment is assigned from activity codes contained in each location/activity profile; this assignment is conditional on the individual's age/sex category--adult male, adult female or child (i.e., under age 12).

1.4 THIS REPORT

The sections that immediately follow provide conclusions (Section 2.0) and recommendations (Section 3.0) stemming from the project. Because the project was designed to create a software product, the conclusions are rather limited. The software requirements for the model are described in Section 4.0, which also provides an overview of the model. Section 5.0 provides a description of the exposure/dose module and Section 6.0 describes the concentration module; these two sections have a parallel structure that covers the modeling approach, inputs, key algorithms, and assumptions and limitations. A user's guide produced as a companion to this report also covers various features of these modules, from the standpoint of the user interface. Section 7.0 of the report addresses efforts within the project to verify and evaluate the two major components of the software, and Section 8.0 lists references cited throughout the report.

Section 2.0

SUMMARY AND CONCLUSIONS

The primary mission of the ARB Indoor Program is to identify and reduce Californians' exposures to indoor pollutants. A general program goal related to this mission is to obtain additional data and information needed to improve exposure analyses, thereby providing a stronger foundation for risk assessments and future decisions related to risk management. To fully consider indoor exposures in assessing risk, the ARB needs estimates of average and peak indoor exposures for the general California population as well as certain subgroups of that population such as individuals who may be highly sensitive to indoor pollutants. The model described in this document is a software product that has been designed to expedite the exposure-assessment process by providing a user interface and calculation tools for supplying and integrating all required information.

The primary function of the model is to combine indoor-air concentration distributions with Californians' location/activity profiles to produce exposure and dose distributions for different types of indoor environments. This function, referred to as Level 1-2 of the model, is achieved through a Monte Carlo simulation whereby a number of location/activity profiles that were collected in prior ARB-sponsored surveys are combined with randomly chosen airborne concentrations for specific types of environments (e.g., residences, office buildings). For many compounds, the concentration data are either limited or nonexistent. Consequently, a second function of the model (Level 3) is to estimate indoor-air concentration distributions for an environment based on distributional information for mass-balance parameters such as source emission rates, building volumes and air exchange rates.

The model developed under this project has several unique features. First, it has the capability to estimate <u>distributions</u> of Californians' inhalation exposure and potential inhaled dose, where the simulated exposure/dose for each individual is accumulated over time and across <u>multiple indoor environments</u>. Second, the exposure/dose distributions can be estimated using environment-specific concentration distributions that are either <u>measured through field studies or simulated within the model itself</u>. Third, distributions for <u>hourly and daily indoor concentrations</u> in a given type of environment are modeled

dynamically, taking into account various types of indoor sources as well as outdoor concentrations, air exchange rates and losses to indoor sinks. The user has the ability within the model to use the concentration outputs directly as inputs to the simulation of exposure/dose distributions. Lastly, there is a capability within the model for saving inputs and for future expansion. Thus, the repository of available input data within the model can grow as more field monitoring studies are conducted or as measurements to support modeling (e.g., chamber studies of emission rates) are collected. A list of ten pollutants is initially provided with the model, but the user can readily expand this list. All inputs saved within the model are automatically linked to the pollutant for which they are provided by the user.

The types of inputs required from the user for each level of the model are indicated in Table 2-1. In many cases, the model provides defaults for choices or input data sets. The first step for the user, in using the model, is to select a pollutant. The model provides a list of ten pollutants, but the user may add to the list. The next step is to choose the model level--exposure/dose distributions (Level 1-2) or concentration distributions (Level 3). If Level 1-2 is selected, then the user also must choose an integration period--24 hours, 12 hours, 8 hours, or 1 hour. Input data (concentration distributions) are provided with the model for some, but not all, integration periods for each of the ten pollutants (except total PAHs, for which data were not available). If data are not available for the integration period chosen by the user for the pollutant selected for a Level 1-2 model run, then the user must either supply new data or choose a different integration period for which data are available.

The next step for Level 1-2 is to choose which of 2,962 activity profiles available in the model are to be used for a specific run. By default, all profiles are available for sampling/use by the model, but a subset can be chosen by the user based on factors such as age, sex, income, location of residence, and time of year. The user can use all profiles that match his or her selection criteria or instruct the model to draw a sample from the matching profiles. The activity profiles provide information on the amount of time spent by each individual in each of nine different types of environments at each of four different activity levels. The model calculates time-integrated exposure as the product of the

Table 2-1. Inputs Required for Each Level of the Model

Exposure/Dose Distributions (Level 1-2)	
Pollutant	Indoor Concentrations[a] per environment[b]
Integration Period	Breathing Rates[c]
Population Subgroup	Random Number Seed
Number of Trials	
Indoor Concentration Distributions (Level 3)	
Pollutant	Volumes[a]
Indoor Sources[a]	Air Exchange Rates[a]
Outdoor Concentrations[a]	Number of Trials
Penetration Factors[a]	Random Number Seed
Indoor Sinks[a]	

[a] Expressed as distributions, from which the model samples a value for each trial.

[b] Nine types of environments are available--residence, office, industrial plant, school, travel in vehicle, public access building, restaurant/lounge, other indoor, and outdoors.

[c] Specific to age/sex (adult males, adult females and children) and activity level (heavy, moderate, light and resting).

sampled concentration for an environment times the amount of time spent in that environment for a sampled activity profile. Inhalation, or breathing, rates provided by the model are specific to three age/sex groups and the four activity levels. These rates are used by the model to calculate the potential inhaled dose received by each individual in each environment (product of concentration times amount of time spent in the environment times breathing rate while in the environment). The model reads the age/sex group, quantity of time and activity level from an activity profile and samples the concentration from a distribution.

The "total" exposure and dose estimated by the model do not necessarily correspond to the entire integration period, but rather to the fraction thereof comprised by the chosen environments. The number of chosen environments can range from one to nine; if only one environment is chosen, then the total exposure and dose for the model run will be equivalent to the environment-specific exposure and dose. A special feature of the model is that more than one concentration distribution can be described by the user for a given

environment. Such a situation could arise, for example, in cases where data on concentration distributions are available from more than one field study. In this instance, the user indicates the distributions to be used and assigns weights to each, with the sum of the weights equal to 100 percent.

For Level 3 of the model, user inputs are required for all factors--indoor sources, outdoor concentrations, building penetration factors, indoor sinks, building volumes, and indoor-outdoor air exchange rates--that are used in an indoor-air model based on the principle of conservation of mass. Some of the inputs are relatively complex. For indoor sources, for example, inputs are required for each type of pollutant-specific source that could be used in a building. Such sources include, for example, interior finishing materials and furnishings, combustion sources, office equipment, and various types of consumer products. Each source can be categorized as long-term, episodic or frequent; the nature of the inputs required and the model's handling of information varies by category. The source-specific information is used in determining an initial indoor-air concentration and in constructing a vector of hourly emission rates that is used in simulating hourly indoor concentrations over a time sequence of 24 hours. The outdoor-concentration inputs are also used in determining the initial concentration and in constructing a vector of hourly outdoor concentrations for the 24-hour simulation. Information on mass-balance parameters is provided for a subset of the pollutants that initially appear in the model (the user can add pollutants to this list).

Although the initial indoor concentration for the model conceivably could be sampled from some distribution, in many cases a randomly chosen concentration would be inconsistent with the impacts of long-term, episodic and frequent sources that are sampled by the model as being present or used in a particular building. Instead, the model explicitly considers the impacts of the sampled long-term and episodic sources in initializing the indoor concentration for each building. Residual impacts of frequent sources are taken into account by modeling the day before the 24-hour period of interest, in addition to modeling frequent sources on the day of interest to capture their concurrent impacts. Although the model calculates hourly concentrations for the day before (referred to as "the previous day"), these values are not output; rather, they are used as part of the inputs to determine

the starting (midnight) concentration for the day of interest ("the current day"), for which sequential hourly-average concentrations and the daily-average concentration are output from the model.

Because this project is concerned with development of a software product, there are no conclusions per se. The only statements of a conclusive nature relate to model verification and evaluation efforts. Verification steps for Level 1-2 indicated that inputs (activity profiles, concentration distributions) are properly accessed and used by the model, and that exposures and doses are correctly computed and accumulated across environments. Model estimates were evaluated utilizing data from an NO_2 study in Los Angeles, in which personal exposures were measured throughout a 24-hour period as well as during the subsets of time when participants were home versus not at home. The average total (personal) exposure estimated from the simulation agreed closely with that based on the field study, but the standard deviation of the modeled exposure distribution was about 11 percent lower than that for the field study. This downward bias was expected because of the limited information available for constructing concentration inputs for environments other than the residence.

Estimates from Level 3 of the model for the residential environment were evaluated for three pollutants--chloroform, benzo[a]pyrene (BaP) and nitrogen dioxide (NO_2)--for which sufficient source-related information and field studies for comparison were both available. Initial model estimates for chloroform had a higher mean than the field measurements, possibly due to conservative model inputs of 100 percent penetration of outdoor concentrations and no losses to indoor sinks. The modeled standard deviation, relative to the mean, initially was lower than that for field measurements. This is believed to be due to initial description of each type of water use--toilets, faucets, showering/bathing, dishwashing and clothes washing--as a separate source. When the chloroform sources were treated in this manner, the model sampled a different emission rate for each type of water use in a given residence, whereas the emission rates are likely to be very similar. When these sources were combined, in order to use a common emission rate within each residence, the modeled ratio of the standard deviation to the mean better reflected the ratio based on measurements. There was a similar finding for

NO_2--the mean of the modeled distribution matched the measurements well but the modeled standard deviation was low relative to measurements. When separate sources that were initially defined for cooking breakfast, lunch and dinner were subsequently combined into one source, the modeled concentration distribution matched the measurement data very well. These modeling outcomes suggest that similar types of indoor sources should be combined whenever possible.

For BaP, a single long-term source was modeled that had a lognormal distribution with a broad tail. Model-based estimates for the mean and percentiles through the 75th were very consistent with those based on measurements, but there were differences toward the upper tail of the distribution. Because the percentiles near either tail of the distribution tend to be less stable statistically than measures of central tendency such as the mean, the model should be run with a sufficient number of trials (e.g., 500) to increase the stability of certain parameter estimates.

Although the principles on which both components of the model are based are scientifically and mathematically sound, the accuracy of the outputs is limited by that of the inputs. Data on concentration distributions, needed for Level 1-2 of the model, are not yet available for many of the environments. Even in cases where measured concentrations are available, there can be inaccuracies due to biases inherent in monitoring devices or sampling strategies. For Level 3 of the model, there is a notable lack of information at present for many of the indoor sources as well as pollutant-specific penetration factors and decay rates. The model is further constrained by the assumption of a single, well-mixed indoor compartment, but this assumption is appropriate given that the model is to be reasonably accurate without being overly complex.

All estimates provided by the model are subject to some degree of uncertainty, but the model offers a means to estimate this uncertainty. Once the user has specified all inputs, the model can be run several times with all inputs the same except the random number seed. Variability for each output parameter (e.g., average exposure, 95th percentile of the exposure distribution) across repeated model runs can then be characterized through a measure such as the coefficient of variation.

Section 3.0

RECOMMENDATIONS

The recommendations stemming from this project are along three lines--additional data for inputs, model evaluation, and further extension/enhancement of the model. For Level 1-2 of the model, data on concentration distributions are not yet available for many of the environments, although inputs are available for residential settings for most pollutants. Any future field monitoring studies in California should include as an objective the collection of information on concentrations in environments such as offices, industrial plants, schools, public access buildings, and travel in enclosed vehicles. If these inputs are collected, together with measurements of total (24-hour) indoor exposure, then a more thorough evaluation of Level 1-2 can be performed.

For Level 3 of the model, many pollutants are lacking information on indoor sources as well as pollutant-specific penetration factors and decay rates. The greatest initial focus should be on the acquisition of source-related information; this includes not only estimation of emission rates (e.g., through chamber studies) but also collection of information on patterns of product use and installation of interior finishing materials and furnishings. For frequently used products, information is needed on frequency, duration, quantity and time of day of use. For long-term and episodic sources, the needed information includes duration since installation or use and the quantity installed or used.

Although Level 3 of the model is constrained by the assumption of a single, well-mixed compartment, it is not recommended at this time that this component of the model be further refined. Model refinements to accommodate multiple compartments or phenomena such as reversible sinks will result in additional input requirements, and the base of information to meet such requirements is not expected to be available for many years to come. Instead, it is recommended that any future work on the model focus on expanding it to incorporate dermal and ingestion routes of exposure. Each of these areas should have a series of input screens along the lines of those for inhalation exposure/dose, and the model should have the capability (as it currently does for inhalation exposure) to model selected inputs in cases where measurement results are not available.

Section 4.0

MODEL REQUIREMENTS AND STRUCTURE

4.1 MODEL REQUIREMENTS

The primary function of the model is to combine indoor-air concentration distributions with Californians' location/activity profiles to produce exposure and dose distributions for different types of indoor environments. This function is achieved through a Monte Carlo simulation whereby a number of location/activity profiles that were collected in prior ARB-sponsored surveys are combined with airborne concentration distributions for specific types of environments such as residences and offices. For many compounds, data on measured concentrations are either limited or nonexistent. Consequently, a second function of the model is to estimate indoor-air concentration distributions based on distributional information for mass-balance parameters such as indoor source emission rates, building volumes and air exchange rates.

In the early stages of the project, a software requirements document (SRD) was developed to turn the general requirements stated above into a more specific set of requirements related to the types of model inputs, capabilities and outputs. The SRD is reproduced in Appendix A of this report and summarized here. Model calculations for either exposure/dose distributions or concentration distributions are to be performed for a 24-hour time period. In calculating exposure/dose distributions, the model uses integration periods of 1, 8, 12 or 24 hours, depending on the specific pollutant for which calculations are requested. An initial list of ten pollutants is provided with the model, but the software allows the user to expand this list. The exposure/dose calculations are performed for either the general California population or a subset of that population, and are performed for up to nine different types of environments. The model outputs include (1) graphs of the differential (density) and cumulative probability distribution for exposure/dose (Level 1-2) or concentration (Level 3), (2) summary statistics for the exposure/dose (Level 1-2) distribution or concentration (Level 3) distribution, and (3) an output file containing environment-specific and total exposure/dose values for each trial (person) in the simulation (Level 1-2), or daily and hourly average concentration values for each trial (building such as a residence) in the simulation (Level 3).

4.2 MODEL STRUCTURE AND CAPABILITIES

The conceptual hierarchy of the model is illustrated in Figure 4-1. Level 3 of the model utilizes a mass-balance equation to estimate concentration distributions for specific types of indoor environments such as residences, offices and schools. Level 2 of the model uses measured or modeled concentration distributions for one or more environment, together with location/activity patterns (i.e., amount of time spent in each environment at specific activity levels), to calculate exposure and dose distributions for the chosen environment(s). Level 1 of the model aggregates the exposure estimates across the chosen indoor environments to develop a distribution of "total indoor air" exposures, that is, the portion of total (24-hour) exposure associated with time spent indoors. If outdoors is also selected as one of the environments to be modeled, then the model can estimate a total exposure or dose distribution that accounts for all time spent by Californians during a 24-hour period. Level 1-2 calculations are performed together as one integrated module within the model.

For each sampled location/activity profile, Level 1-2 of the model estimates inhalation exposure (i.e., the time-integrated concentration encountered while in an environment) and potential inhaled dose (i.e., the product of time-integrated concentration and amount of air inhaled per unit time while in the environment). The average breathing rate while in the environment is assigned on the basis of (1) activity codes contained in each location/activity profile and (2) studies by the ARB and other organizations on breathing rates associated with various levels of exertion. The assignment of breathing rates to activity codes is conditional on the individual's age/sex category--adult male, adult female or child (i.e., under age 12). Given the functional roles of the model levels, Level 1-2 is used interchangeably with "exposure/dose distributions" throughout the remainder of this document and Level 3 equates to "concentration distributions."

The model provides a user-friendly interface for making choices and providing inputs at each level. This interface, described in detail in the user's guide for the software, features a variety of pull-down and pop-up menus. It also provides defaults where possible and is equipped with summaries of distributional information for various inputs, in cases where such information is known to exist. Inputs for exposure/dose calculations (see

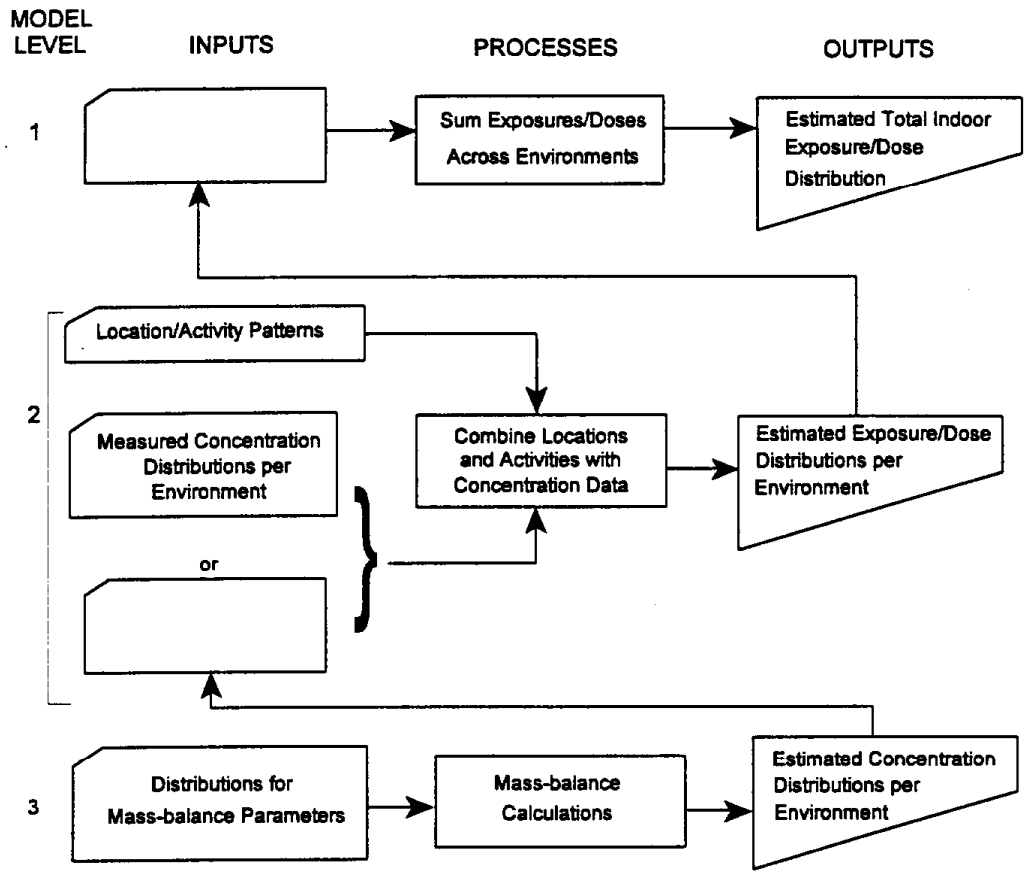

Figure 4-1. Conceptual Hierarchy for the Model

Section 5.0) include the population subgroup for which exposures are to be estimated, environment-specific concentration distributions, and breathing rates for adult males, adult females and children. Inputs for concentration calculations (see Section 6.0) include indoor sources, outdoor concentrations, building penetration factors, indoor sinks, building volumes and indoor-outdoor air exchange rates.

The model software is written in Microsoft QuickBASIC (Version 4.5). Three additional software products were used, each of which is compatible with the BASIC programming environment and offers special features important to the development efforts for this project. QuickWindows Advanced (Version 2.1), a user interface system, was used in designing the screens for the user interface; although useful in this regard, the software also had a number of limitations that could be overcome only through detailed programming efforts. A library of relational database management routines for QuickBASIC, called db/LIB (Version 2.0), was used to manipulate information in files containing activity profiles and to allow users to store and later access choices/inputs related to population subgroups, location-specific concentration distributions, and various mass-balance parameters. GRAFLIB (Version 6.0), a video and printer graphics library, was used for saving and printing graphical outputs from the model.

Section 5.0

EXPOSURE AND DOSE DISTRIBUTIONS

5.1 MODELING APPROACH

Figure 5-1 provides an overview of the approach used to model exposure and dose distributions for California residents. Through the user-friendly interface provided by the model, the user makes a number of choices and provides a variety of inputs. Where inputs are required, as in the case of concentration distributions, the model provides access to those inputs that are currently available from existing studies. Where choices are required, the model may indicate one choice as the default, but the user is free to override the default choice. The inputs and choices, described in detail in the accompanying user's guide for the model, are briefly summarized below.

The first step for the user is to choose a pollutant and integration period. A list of the following pollutants is provided by the model (along with environment-specific concentration distributions for these pollutants, where available): benzene, benzo[a]pyrene, carbon monoxide (CO), chloroform, formaldehyde, nitrogen dioxide (NO_2), inhalable particles (PM_{10}), perchloroethylene, trichloroethylene, and total polycyclic aromatic hydrocarbons (PAHs). The user has the option of adding more pollutants to the list, but must provide concentration-distribution inputs if a new pollutant is chosen for a model run. There are five choices for the integration period: 24 hours, 12 hours during daytime, 12 hours during nighttime, 8 hours and 1 hour. Although 24 hours is expected to be the usual choice, the 12-hour period is provided for compatibility with certain concentration data that have been collected, and the 1-hour and 8-hour periods are provided so that exposures to CO can readily be assessed relative to ambient air quality standards for CO.

The user next chooses which of 2,962 activity profiles available in the model are to be used for a specific run. By default, all profiles are used, but a subset can be chosen based on factors such as age, sex, income, location of residence, and time of year. The

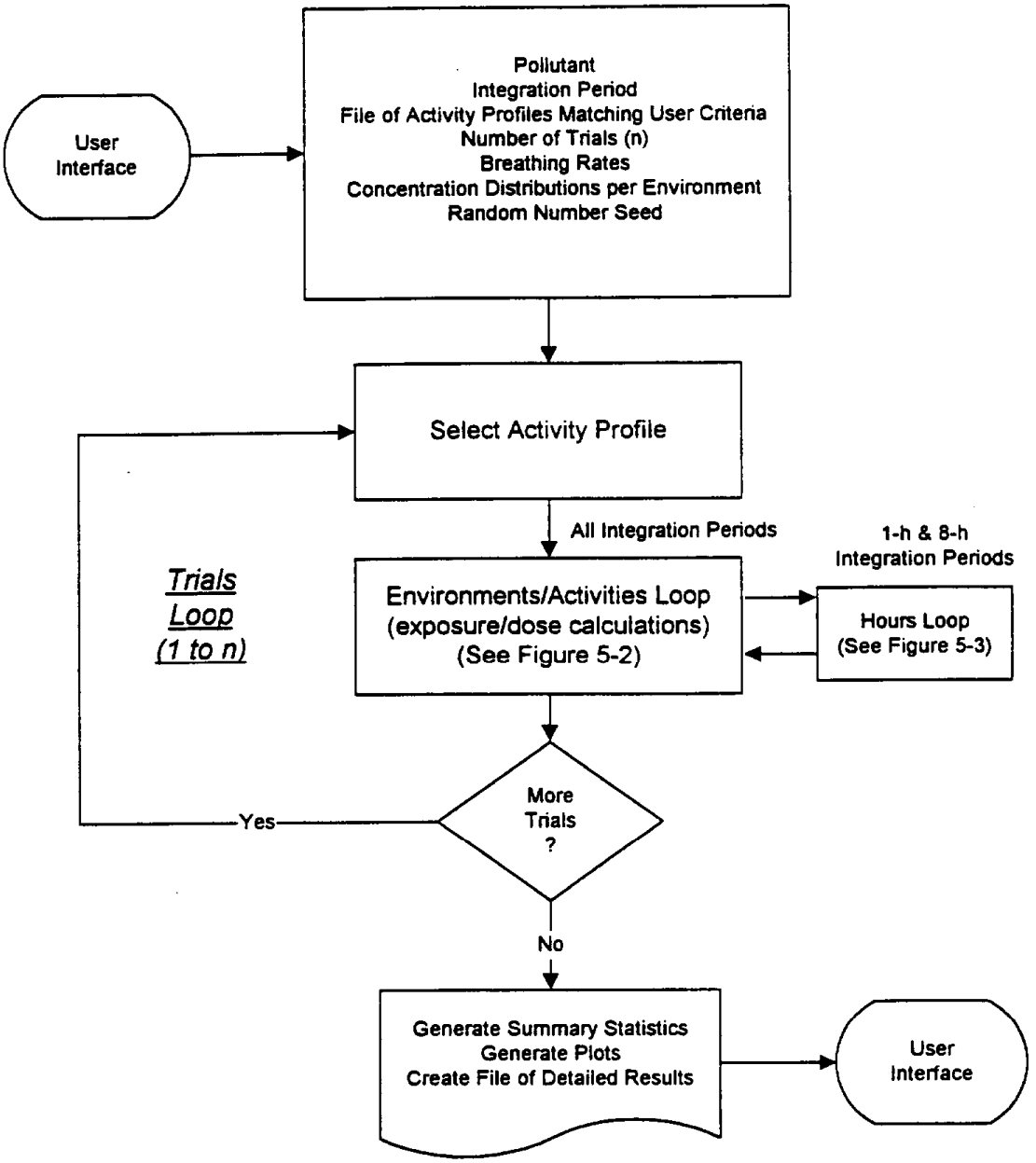

Figure 5-1. Overview of Exposure/Dose Calculation Routine

model identifies all profiles that match the user's criteria, places these in a file for access during model execution, and reports the number available to the user. The user can use all matching profiles or instruct the model to draw a sample; if drawing a sample is chosen, then the user inputs the number of profiles to be selected. The activity profiles provide information on the amount of time spent by each individual in each of nine different environments (residences, offices, industrial plants, schools, enclosed vehicles, public access buildings, restaurants/lounges, other indoor locations and outdoors) at each of four different activity levels (heavy, moderate, light and resting).

Inhalation, or breathing, rates provided by the model are specific to three age/sex groups (adult males, adult females, and children under age 12) and the four activity levels. These rates are used by the model to calculate the potential inhaled dose received by each individual in each environment. The other inputs for the dose calculation are the quantity of time spent in each environment and the concentration in each environment. The model reads the quantity of time from the activity profile and samples the concentration from a distribution. A variety of formats for describing the concentration distribution is allowed by the model, including provision of a data file containing concentration values. A random-number seed for the simulation can be selected by the user; optionally, the system clock can be utilized by the model to determine the seed.

Once the user is finished with choices/inputs and signals completion to the model, the exposure/dose calculation is initiated. The core of the calculation is a "trials loop" whereby the model (1) samples an activity profile at random from those matching the user's criteria (or goes through all matching profiles sequentially if instructed by the user to use all available profiles) and (2) calculates the time-integrated exposure and potential inhaled dose for each profile; each sampling or use of an activity profile constitutes a "trial." Nested within the trials loop are an "environments/activities" loop, which applies to all integration periods, and an "hours" loop, which applies only to the 1-hour and 8-hour integration periods; these loops are described below. At the completion of the trials loop, the model generates summary statistics and plots, produces a file containing detailed results for each trial, and returns control to the user. The user can then view the plots and

statistics and choose (1) whether to print or save (or both) the statistics and plots and (2) whether to save the file of detailed results.

The "environments/activities" loop, depicted in Figure 5-2, applies to each environment chosen by the user (inputs on concentration distributions are required for each chosen environment). In general, the loop for each environment consists of (1) sampling a concentration value from the user-selected distribution, (2) calculating the exposure and dose for the time spent in that environment, and (3) accumulating the exposures/doses across the chosen environments to calculate the total exposure and dose for all environments combined. Details on these calculations are provided in Section 5.3. After the model has cycled through all applicable environments for one trial (activity profile), the calculation results are stored to an array and the model proceeds to the next trial.

The total exposure and dose do not necessarily correspond to the entire integration period, but rather to the fraction thereof comprised by the chosen environments. The number of chosen environments can range from one to nine; if only one environment is chosen, then the "total" exposure and dose for the model run will be equivalent to the environment-specific exposure and dose. A special feature of the model is that more than one concentration distribution can be described by the user for a given environment. Such a situation could arise, for example, if concentration distributions are available from more than one field study. In this instance, the user indicates the distributions to be used and assigns weights to each, with the sum of the weights equal to 100 percent. For example, if distributions from two field studies are available and the two studies are considered equally valid and representative, then a weight of 50 percent could be assigned to each. In such cases, the model will select one of the distributions at random (with probabilities proportional to the user-assigned weights) within each trial. As shown in the figure, the concentration distribution(s) chosen by the user can be described in any of six different ways, and the model is equipped with routines for sampling concentrations in accordance with each of the descriptions.

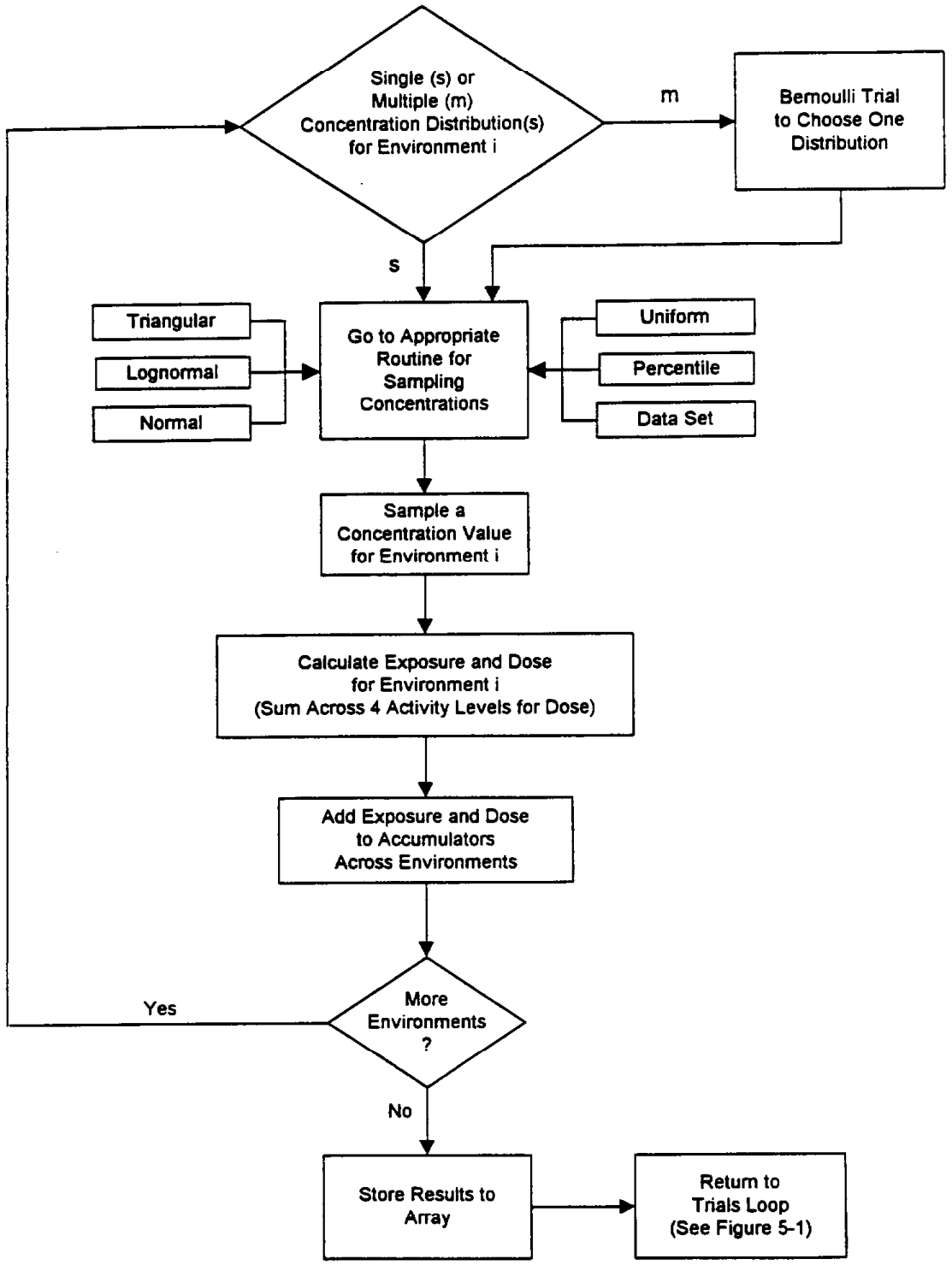

Figure 5-2. Environments/Activities Loop for 24-Hour and 12-Hour Integration Periods

The environments/activities loop shown in Figure 5-2 indicates that, in cases where the user has input multiple concentration distributions for a given environment, the model performs a Bernoulli trial to choose one distribution from which to sample a concentration value. Suppose, for example, that the user has input two concentration distributions, the first with a weight of 75% and the second with a weight of 25% (the weights must add to 100%). The Bernoulli trial in this case involves first sampling a number at random between 0 and 1 using the uniform distribution, and then choosing user distribution 1 if the random number is between 0 and 0.75 or distribution 2 otherwise. This procedure can readily be extended to cases where the user has input three or more distributions for an environment. For example, for three input distributions with the first weighted at 50% and the second and third weighted at 25% each, the model chooses the first distribution if the random number is between 0 and 0.5, the second distribution if the random number is between 0.5 and 0.75, and the third distribution otherwise.

For the 8-hour and 1-hour integration periods, an "hours" loop is nested within the environments/activities loop (Figure 5-3); exposure/dose calculations that require information on locations and activities are contained within the hours loop. In contrast to the 24-hour and 12-hour integration periods, for which only one concentration value is sampled per trial (individual) for each environment, the routine for the 8-hour and 1-hour periods samples a vector of 24 concentration values for each trial. For the 1-hour period, the first concentration value corresponds to the period from midnight to 1:00 a.m., the second from 1:00 a.m. to 2:00 a.m., and the last (24th) from 11:00 p.m. to midnight. For the 8-hour period, the concentration values relate to 24 running, or overlapping, time periods; the first is from midnight to 8:00 a.m., the second is from 1:00 a.m. to 9:00 a.m., and the last is from 11:00 p.m. to 7:00 a.m.

Information in the activity profiles, on time spent in each environment at each activity level, has been stored for corresponding time periods. For the 1-hour integration period, for example, there are 24 sequential records for each individual; each record contains information on the amount of time spent in each environment at each activity level

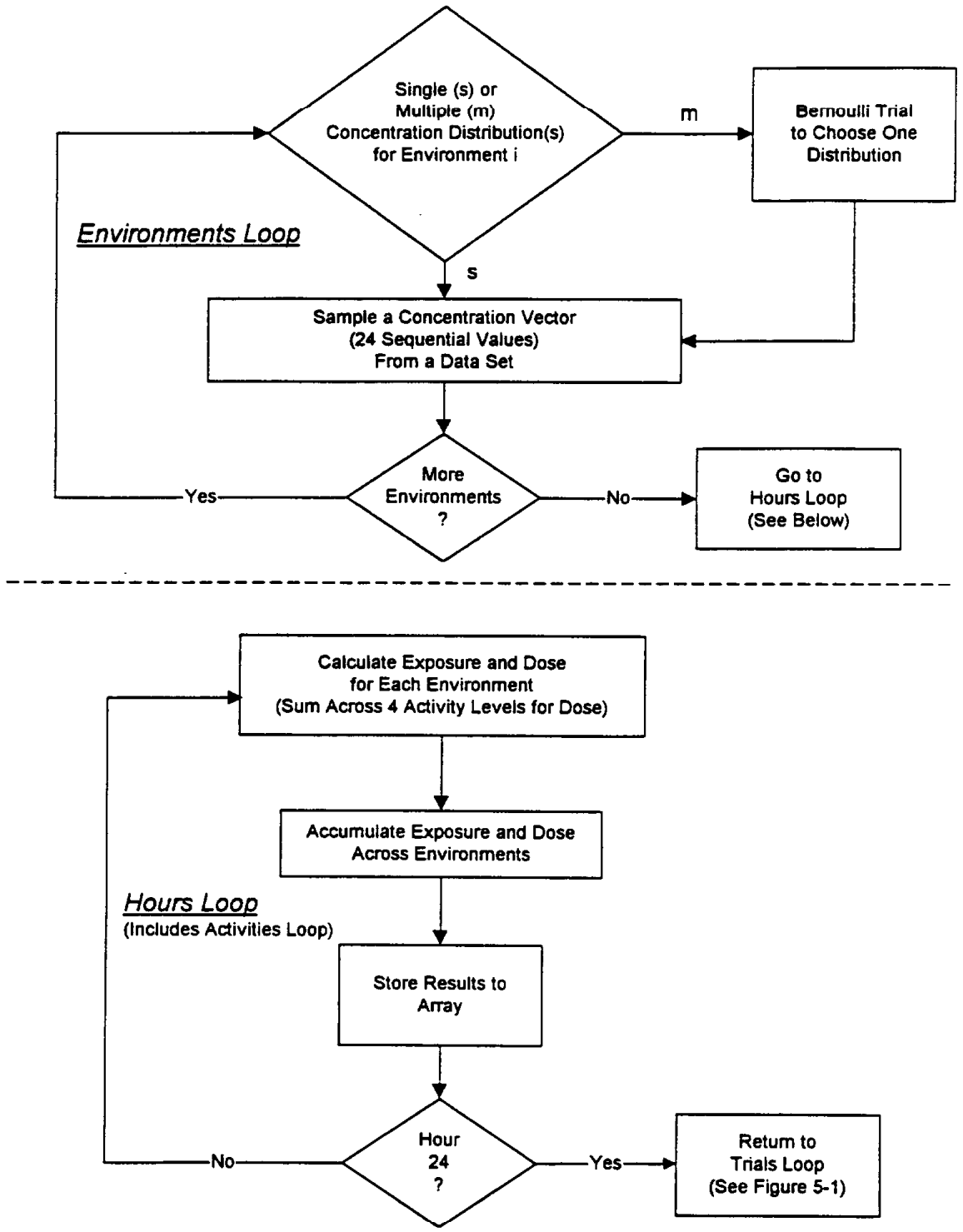

Figure 5-3. Environments, Activities and Hours Loops for 8-Hour and 1-Hour Integration Periods

during an hour. For the 8-hour integration period, there are again 24 sequential records; in this case, however, the amount of time spent in each environment at each activity level is stored for successive 8-hour periods.

For the 8-hour and 1-hour periods, exposures and doses are calculated for each environment (and then summed across environments) within the hours loop, by multiplying the sampled concentration in each environment for each hour (or 8-hour period) in the loop by the corresponding time spent in the environment (and by the breathing rate, for dose), as contained in the activity profile for that hour. The calculation results for each environment, and for the accumulated exposure/dose across environments, are stored, hour by hour, to an output array. After calculations are performed for all 24 hours using one set of 24 activity profiles, the model returns to the trials loop to select the next set of 24 activity profiles. In contrast to the 24-hour and 12-hour periods, for which the unit of analysis (basis for statistics output by the model) is a single activity profile per individual, the unit of analysis for 8-hour and 1-hour periods is one of the 24 sequential profiles per individual. Thus, for example, if 100 trials are chosen by the user, then the model will output 2400 records (100 trials times 24 sequential records per trial), and all plots and statistics will be based on these 2400 records. Because of the greater underlying detail, the model will take considerably longer to execute when the 8-hour or 1-hour integration period is selected; as noted previously, these time periods customarily apply only to carbon monoxide.

A set of 24 sequential concentration values is needed for each trial (individual) when the user selects an 8-hour or a 1-hour integration period. Consequently, the only allowable way to describe concentration inputs is through a data set, with each record in the data set containing a vector of 24 sequential values. Procedures for creating such a data set are described in the accompanying user's guide for the model. Because the 24 values in a given vector typically have some degree of serial correlation in real life, such concentration vectors should not be "made up." Instead, they should be taken from field studies where sequential hourly measurements were recorded, or they should be simulated using Level 3 of the model. The user can choose more than one data set, if there are data sets resulting from multiple field studies or from multiple runs of level 3 of the model.

Weights can then be assigned to each data set, and the model will sample vectors from the data sets in proportion to their respective weights.

5.2 MODEL INPUTS

Aside from the choice of pollutant and integration period, the primary inputs for the exposure/dose calculations are the activity profiles, breathing rates, concentration distributions, and random number seed. Sources of information and key model features relating to each of these inputs are described in this section. The next section (5.3) describes the algorithms within the model that make use of these inputs.

5.2.1 Activity Profiles

Sources of Information. All activity profiles used in the model are derived from two ARB-sponsored surveys whereby 24-hour time diaries were administered to a modified random sample of California residents. The first (Jenkins et al. 1992), which involved a target population of adults (ages 18 years and above) and adolescents (ages 12 to 17 years), provides 1,762 location/activity profiles. The second (Phillips et al. 1991), which involved a target population of children (ages 11 years and below), provides an additional 1,200 location/activity profiles, for a total of 2,962 profiles. The two surveys are complimentary in that virtually identical schemes were used for recording both location/ activity information and selected characteristics of the participants. In preparation for the data processing steps described below, the ARB developed a combined activity file for adults and children, with some recoding of original data in order to make the codes from the two surveys completely compatible.

The location/activity information for each survey participant was organized by the ARB contractor in a series of records, with the number of records equal to the number of time-sequenced combinations of distinct locations and activities. Under the project that is the subject of this report, a computer program was developed to (1) group the location codes into the nine types of environments shown in Table 5-1 and (2) sum the time spent across locations within each environment type. This program was applied separately to the entire 24-hour period for each participant, to 12-hour daytime (6:00 a.m. to 6:00 p.m.) and nighttime periods, to 24 sequential 1-hour periods, and to 24 running (overlapping)

Table 5-1. Location Codes and Percent of Time Associated with Nine Types of Environments

Environment Type	Applicable Location Codes	Average Minutes Per Day[a]	Percent of Subjects' Time
1. Residences	01-09, 12, 13, 32	954.0	66.3
2. Offices	21	70.1	4.9
3. Industrial Plants	22	34.9	2.4
4. Schools	25	40.4	2.8
5. Travel in Enclosed Vehicles	51, 52, 55-58	97.8	6.8
6. Stores and Other Public Buildings	23, 24, 26, 27, 30, 31	84.2	5.8
7. Restaurants/Lounges	28, 29	36.0	2.5
8. Other Indoor Locations	33, 35-39, 61, 99	36.4	2.5
9. Outdoors	10, 11, 34, 40, 53, 54, 59, 60, 63	86.3	6.0

[a] Weighted estimates based on activity profiles for adults and adolescents (1,762 cases).

8-hour periods. As a result, five location/activity files were generated; the 24-hour and 12-hour files each contain 2,962 records of time spent in each of the nine environments, whereas the 8-hour and 1-hour files each contain 24 records per participant, or 71,088 (24 x 2,962) records in total.

The descriptions of the codes associated with each environment are given in Table 5-2. Time in the residence includes time spent in the participant's residence as well as other persons' residences. Some locations/activities associated with transportation, but not involving travel in an enclosed vehicle, are assigned to the outdoor-environment; examples include walking, bicycling and waiting at a bus stop. The public-buildings environment includes retail outlets, hospitals, churches, and other public buildings such as libraries. Restaurants and lounges were broken out as a separate category because special exposure situations (e.g., relating to tobacco smoke) can arise in these environments. The

Table 5-2. Specific Types of Locations Associated with Each Type of Environment

Environment Type	Associated Location Code	Description of Location
Residences	01	Kitchen at home
	02	Living room, family room or den at home
	03	Dining room at home
	04	Bathroom at home
	05	Bedroom at home
	06	Study or office at home
	07	Garage at home
	08	Basement at home
	09	Utility room or laundry room at home
	12	Room to room at home
	13	Other household room
	32	Other's home
Offices	21	Office building, bank or post office
Industrial Plants	22	Industrial plant or factory
Schools	25	School
Travel In Enclosed Vehicles	51	Car
	52	Van or pick-up truck
	55	Bus
	56	Train or rapid transit
	57	Other truck
	58	Airplane
Stores and Other Public Buildings	23	Grocery store
	24	Shopping mall
	26	Other public place
	27	Hospital or doctor's office
	30	Church
	31	Indoor gym or health club
Restaurants or Lounges	28	Restaurant
	29	Bar or night club
Other Indoor Locations	33	Auto repair, parking garage, gas station
	35	hotel or motel
	36	Dry cleaners
	37	Beauty parlor/barber shop
	38	Work: varying places
	39	Other Indoor Locations
	61	Other transportation
	99	Unknown location
Outdoors	10	Pool or spa at home
	11	Yard or patio at home
	34	Playground or park
	40	Other outdoor locations
	53	Walking
	54	At bus or train or ride stop
	59	On bicycle
	60	On motorcycle or scooter
	63	In a stroller or carried by an adult

"other indoor" environment represents a mixture of special exposure situations (e.g., parking garages, dry cleaners), environments that do not clearly fit any other category, and unknown (but presumed indoor) environments.

Within each of the nine environment types, each individual's time was further disaggregated according to four activity levels. The four activity levels, and illustrative types of activities for each level, are listed in Table 5-3. As described later (Section 5.3), breathing rates assigned to each activity level are used for dose calculations. Specific activity codes assigned to each level are listed in Table 5-4; these activity codes are described in Appendix B. Thus, the time (1440 minutes) associated with the 24-hour activity profile for each participant was disaggregated into a total of 36 categories (9 environments times 4 activity levels). The same disaggregation procedure was applied to the 12-hour daytime and 12-hour nighttime portions of the 24-hour profile, and to the sequential 1-hour profiles and running 8-hour profiles.

In addition to the files containing time spent in various environments at different activity levels, a second file was developed that contains characteristics of each participant in the time-diary studies. These characteristics include (1) the age and gender of the participant, (2) the region (South Coast, San Francisco Bay area, or remainder of the state), county and type of area (city, suburban, or rural) where the participant resided at the time of the study, (3) the month of year and day of week for which the participant's locations and activities were recorded, (4) the income category for the participant's household, and (5) the education-level category and activity status (e.g., working, retired, school, unemployed) for the participant. Because the last two characteristics are not strictly applicable to children and adolescents, two additional categories--youth and teen-- were created so that all participants could be assigned to a category for these characteristics.

The file containing participant characteristics was sorted in a manner identical to the location/activity file, according to a unique identifier within each of three age groups (first adults, then adolescents, then children). As an expediency for linking information from the two files, an index number (from 1 to 2,962) was assigned sequentially to each record in each file.

Table 5-3. Activity Levels Used for Dose Calculations

Activity Level	Examples of Activities
Resting	Sleeping, napping
Light	Sitting, eating, travel in vehicle, relaxation
Moderate	Gardening, house cleaning, active play with children
Heavy	Vigorous physical exercise, active sports

Table 5-4. Assignment of Activity Codes to Activity Levels

Activity Level	Associated Activity Codes[a]
Resting	45,46
Light	1,2,5-8,10,11,14,18-24,26-28,30-38,40-44,47,48,50-58,60-68,70-78,83-88,90-99
Moderate	12,13,15-17,25,81,82
Heavy	80
Variable[b]	3,9,29,39,49,59,69,79,89

[a] In the case of three-digit codes (finer level of activity description,) only the first two digits were used; descriptions of each activity code are given in Appendix B.

[b] Levels assigned to these travel-related activities were dependent on the mode of transportation utilized, as follows:

car - light
pickup/van - light
walking - moderate
at transit stop - light
bus - light
rapid transit - light
truck - light
airplane - light
bike/skateboard - moderate
motorcycle/scooter - moderate
carried by adult - resting
all other - moderate

User Inputs and Options. A routine provided with the user interface allows the user to select a subgroup of the 2,962 profiles. For example, the user could select females between the ages of 18 and 45 who reside in the South Coast area in households with an annual income below $20,000. The matching routine provided by the model (1) identifies the index number for each participant whose characteristics match the user's criteria, and (2) uses the index number for the matching participants to create a new file containing environment/activity information for this subgroup.

At the completion of the matching routine, the model reports the number of matches found to the user. The user can then elect to re-run the matching routine (e.g., if the number of matching records is considered too low), instruct the model to use all matching profiles for exposure/dose calculations, or ask the model to randomly select activity profiles from the matching subgroup for use in the calculations. In the latter case, the user also specifies the number of trials (activity profiles) to be selected; if this number exceeds the number of matching profiles, the model issues a warning message but still allows the user to retain this choice.

5.2.2 Inhalation Rates

Inhalation, or breathing, rates were assigned to each of the four activity levels (see Table 5-5) by ARB staff. The assignments were made separately for three population subgroups--adult males, adult females, and children (the adult values are assumed to apply to adolescents as well). Breathing rates were based primarily on an ARB-sponsored study in which Adams (1993) measured the pulmonary ventilation rates of 160 individuals, ages six to seventy, during various activities. Data from this study were especially useful because the investigator measured actual ventilation rates during common activities (such as yard work, house work, driving a car, and so on) as well as during traditional lab protocols such as sitting, standing, and running at various speeds. When assigning breathing rates in the model, consideration also was given to values presented in other pulmonary ventilation literature and published data (GCA 1985; International Commission on Radiological Protection 1975; Layton 1993; U.S. EPA 1989; U.S. EPA 1995).

The inhalation rates are used in the model for calculation of potential inhaled dose (see Section 5.3). The rates listed in Table 5-5 are the current defaults for the model, but these values may be modified in the future as new pulmonary ventilation data become available. The model allows the user to change the default values and to save any array of inhalation rates (including the default values) in a data file for subsequent access.

Table 5-5. Inhalation Rates Assigned to Four Activity Levels for Three Population Subgroups

Activity Level	Inhalation Rate (m^3/h) for Population Subgroup		
	Adult Males	Adult Females	Children (<12 years)
Resting	0.5	0.4	0.3
Light	1.0	0.8	0.8
Moderate	2.5	2.0	1.4
Heavy	5.0	3.0	2.2

5.2.3 Environment-Specific Concentrations

Sources of Information. One of the key inputs to the model is a description of the concentration distribution for various types of environments (residence, offices, etc.). In the early stages of the project, a literature review was conducted to identify field monitoring studies conducted in California, from which descriptions of concentration distributions could be developed for the target pollutants for this study. The literature review was also intended to provide information for various types of inputs to the mass-balance model (Level 3 of the software model) described in Section 6.0. The information reviewed as part of this effort included a published search by the National Technical Information Service, an indoor-air reference bibliography prepared by the U.S. Environmental Protection Agency (USEPA 1988), proceedings of international indoor air conferences held in 1984, 1987 and 1990, and peer-reviewed articles oriented toward indoor air in journals such as Environmental Service and Technology, the Journal of the Air Pollution Control Association (in recent years, the Journal of the Air and Waste Management Association), Environment International, Atmospheric Environment, and

Indoor Air. Project staff also contacted researchers at organizations, such as the U.S. Environmental Protection Agency, Research Triangle Institute, Harvard School of Public Health, and Integrated Environmental Services, known to be actively conducting indoor-air monitoring studies in California. A bibliography resulting from the review appears in Appendix C.

Sources of information on concentration distributions for various pollutant-environment combinations are given in Table 5-6; these sources were identified through the literature review and personal contacts described above. Information for most pollutants is largely restricted to residences, but some limited data are available for office settings (formaldehyde) and for travel in enclosed vehicles (formaldehyde and VOCs). For all pollutant-environment combinations shown in the table, information on concentration distributions has been entered in the model using one or more of the description methods described below. The pollutants and integration periods for which distributions have been entered are summarized in Table 5-7. Further details are provided in Appendix B of the user's guide. As summarized in Table 5-6 and further detailed in Appendix B of the user's guide, some of the concentration data provided with the model are based on averaging periods longer than 24 hours and, therefore, are best treated as "practice" data sets.

The information sources in Table 5-6 were used as inputs to the model on concentration distributions, in the form of percentiles of the cumulative frequency distribution or as the mean and standard deviation for an assumed lognormal distribution. Some measurement results were summarized in the literature as means and standard errors rather than means and standard deviations. Because the standard error is defined as the standard deviation divided by the square root of the sample size ($\sqrt{n}$), any reported standard error was multiplied by $\sqrt{n}$ to obtain the standard deviation. For the 1987 TEAM study conducted in California (Wallace et al. 1991), published results included summary statistics for 12-hour daytime and nighttime samples, but did not include 24-hour statistics. Individual sampling results from the study were obtained through the EPA Environmental Monitoring Systems Laboratory in Las Vegas, and the daytime/nighttime values were averaged to develop a 24-hour average from which the mean and standard deviation were computed. For consistency, the means and standard deviations for the 12-hour samples were also computed from the same set of data.

Table 5-6. Current Sources of Information on Concentration Distributions for Selected Pollutants and Indoor Environments

Pollutant	Environment	Reference	Comments
Formaldehyde	Residences	Sexton et al. (1989)	Mobile homes, weekly
		Sexton et al. (1986)	Mobile and conventional homes, weekly
		Rogozen et al. (1984)	Conventional homes, weekly
	Offices	Rogozen et al. (1984)	Weekly, limited sample
		Turk et al. (1986)	Pacific Northwest area, biweekly
	Travel	Shikiya et al. (1989)	Trips averaging 33 minutes
Volatile organic compounds	Residences	Wallace et al. (1991)	Los Angeles, 24-hour and 12-hour[a]
- benzene		Sheldon et al. (1992a)	Woodland, 24-hour
- chloroform	Travel	Shikiya et al. (1989)	Trips averaging 33 minutes
- perchloroethylene			
- trichloroethylene			
Benzo[a]pyrene	Residences	Sheldon et al. (1992b)	Riverside, 24-hour
PM_{10}	Residences	Pellizzari et al. (1992)	Riverside, 24-hour
		Colome et al. (1990)	Orange County, 24-hour, asthmatics, limited sample
Nitrogen dioxide	Residences	Wilson et al. (1986)	Los Angeles, weekly
		Spengler et al. (1992)	Los Angeles, 48-hour
Carbon monoxide	Residences	Wilson et al. (1993)	Three utility service areas, 48-hour
			- Pacific Gas & Electric
			- San Diego Gas & Electric
			- Southern California Gas

[a] Limited published data; raw data were obtained from U.S. Environmental Protection Agency and analyzed to develop distributional information (see text).

Table 5-7. Summary of Concentration Data Available for Model Level 1-2

Pollutant	Residence (Environment 1)	Office (Environment 2)	Travel in Vehicle (Environment 5)	Outdoors (Environment 9)
Benzene	24AP[a]	--	24AP	24AP
Benzo(a)pyrene	24AP	--	--	24AP
Carbon Monoxide	24	--	24	24
Chloroform	24AP	--	24AP	24AP
Formaldehyde	24	24	24	24
Nitrogen Dioxide	24	--	--	24
PM10	24AP	--	--	24AP
Perchlorethylene	24AP	--	24AP	24AP
Trichloroethylene	24AP	--	--	24AP
Total PAHs	--	--	--	--

[a] 24 refers to 24-hour integration period, A refers to 12-hour daytime period, and P refers to 12-hour nighttime period.

User Inputs and Options. The manner in which information on concentration distributions can be input to the model is summarized in Table 5-8. In most cases, use of the lognormal or percentile descriptions would be most appropriate. The triangular distribution provides a useful approximation of the normal or lognormal distributions when only limited information is available, such as minimum, maximum and most frequently occurring concentration. Actual measured concentration values can also be input using the "data set" description; when this description is used, however, the only possible outcomes in the simulation for the pollutant-environment combination described in this manner are the values contained in the data set. For the 1-hour and 8-hour integration periods, which require a vector of 24 sequential concentration values as input for each trial of the simulation, the only allowable description is the data set.

Table 5-8. Ways in Which Concentration Distributions for Various Locations Can Be Described as Inputs to Exposure/Dose Calculations

Description of Distribution	Inputs Required from the User
Normal	Arithmetic mean and standard deviation
Lognormal	Arithmetic mean and standard deviation
Triangular	Minimum, mode and maximum
Uniform	Minimum and maximum
Percentile	Various percentiles[a] of the cumulative frequency distribution and associated concentration values
Data Set[b]	Name of data file (.DBF format) and number of records in the file

[a] Minimum, 5th, 10th, 25th, 50th, 75th, 90th, 95th and maximum are preferred.

[b] Only description allowed for 1-hour and 8-hour integration periods, because vectors of 24 sequential concentration values must be input to the model.

As noted in the introduction to Section 5.1, more than one concentration distribution can be described by the user for a given environment; in such cases, the user assigns weights to these distributions (the sum of the weights must equal 100 percent).

For each trial of the simulation, the model then chooses one of these distributions at random (in proportion to the assigned weights), from which a concentration is randomly selected for that environment. The following example for formaldehyde in residences illustrates how the weighting option can be used. Concentration data are available from separate surveys of mobile homes (Sexton et al. 1989) and conventional homes (Sexton et al. 1986; Rogozen et al. 1984). The measurements in mobile homes have been summarized separately for summer and winter. The measurements for conventional homes from Rogozen et al. (1984), although somewhat dated, are for a statewide random sample, whereas those from Sexton et al. (1986) are from a localized pilot study of volunteers. Mobile homes account for approximately 4 percent of all residential housing units in California. Given this information, the user might opt to (1) weight the summer/winter results equally (2 percent each) for the mobile-home subset and (2) give the Rogozen results for conventional homes twice the weight (64 percent) of the Sexton results for conventional homes (32 percent). The sum of these weights (2 + 2 + 64 + 32) is 100 percent.

5.2.4 Random Number Seed

The random number seed initiates the sequence of all random sampling done as part of the model's simulation routines. This seed can either be input by the user or, at the user's option, determined from the system clock. The user's ability to control the seed across related runs may prove useful for some investigations. For example, in the case of concentration inputs for one environment only, the user could keep the seed constant to help assess the difference in estimated exposure/dose distributions due to the manner in which data from a single study are described (e.g., lognormal vs. percentile). In the case of multiple environments with concentration data, one could assess the sensitivity of selected output parameters (e.g., mean or median of the total exposure/dose distribution) to selected inputs by varying the input values for one environment while keeping all others constant. Such investigations could also be carried out while varying the random number seed from one run to the next, but in such cases the changes in the output distribution could be partly due to differences in time allocations for a new sequence of sampled activity profiles. The user can also assess the stability of various parameters describing the

estimated exposure or dose distribution by making repeated runs whereby all inputs are kept the same except the random seed. Examples of this approach are given in Section 7.0 of this report.

If the number of environments for which concentrations are to be sampled is changed from one run to the next, then the sequence of sampled activity profiles will be altered, even if the seed is kept the same. The sequence changes because the number of random choices within each trial is altered (a random choice is required to select an activity profile as well as a concentration value for each environment). Similarly, keeping the number of environments constant but changing from a single description of the concentration distribution for one environment to multiple descriptions would also alter the sequence of sampled activity profiles, because an additional random number would be needed within each trial to choose a concentration distribution. For example, with one concentration distribution, the first random number would be used to select the first activity profile, the second to select a concentration, the third to select a second activity profile, and so on. With multiple concentration distributions for an environment, the first random number would be used to select the first activity profile, the second to select a concentration distribution, the third to select a concentration from the chosen distribution, and the fourth to select a second activity profile.

5.3 KEY ALGORITHMS

The key algorithms used in this segment of the model relate to exposure and dose calculations that are performed for each trial of the simulation; these calculations are summarized in Table 5-9. For each chosen environment for which the user provides input on the concentration distribution, the model (1) reads the amount of time (in hours) spent in the environment from each sampled activity profile (trial) and (2) samples a concentration value from the distribution (in $\mu g/m^3$ for most pollutants). The time-integrated exposure received by the individual in that environment is calculated as the product of time spent and concentration in the environment, in units of $\mu g\text{-}h/m^3$ (i.e., $\mu g/m^3 \times h$). The total exposure across all environments chosen for a given model run is the sum of the environment-specific exposures. For each run, the model calculates and can report both the environment-specific exposure distributions and the total exposure distribution.

Table 5-9. Summary of Exposure and Dose Calculations

Notation	Description	Calculation	Unit of Measurement
C_{ij}	Concentration for trial i in environment j	None (sampled from concentration distribution)	$\mu g/m^3$
T_{ij}	Time spent for trial i in environment j	None (obtained from sampled activity profile)	h
T_i	Total time[a] spent in the chosen environments for trial i	$\sum_{j=1}^{9} T_{ij}$	h
E_{ij}	Exposure for trial i in environment j	$C_{ij} \cdot T_{ij}$	$\mu g\text{-}h/m^3$
E_i	Total exposure[a] for trial i	$\sum_{j=1}^{9} E_{ij}$	$\mu g\text{-}h/m^3$
T_{ijk}	Time spent for trial i in environment j at activity level k	None (obtained from sampled activity profile)	h
BR_{ijk}	Breathing rate for trial i in environment j at activity level k	None (obtained from sampled activity profile and lookup table[b])	m^3/h
D_{ij}	Inhaled dose for trial i in environment j	$\sum_{k=1}^{4} C_{ij} \cdot T_{ijk} \cdot BR_{ijk}$	μg
D_i	Total inhaled dose[a] for trial i	$\sum_{j=1}^{9} D_{ij}$	μg

[a] Although the calculation is shown across nine environments (the maximum possible), only the subset chosen by the user will be used in the actual calculations. Total indoor exposure and dose can be simulated by providing inputs for the first eight environments (the ninth is outdoors).

[b] See Table 5-5, which provides the default values in the model for the lookup table.

Calculation of potential inhaled dose for any environment involves individual breathing rates associated with the activity in that environment. Within the activity record, an individual's time spent in each environment is segmented into the quantities of time spent at each of four activity levels (see Section 5.2.2). The potential dose received by an individual in any environment at a given activity level is the product of the sampled concentration ($\mu g/m^3$), time spent at that activity level (h), and breathing rate at that activity level (m^3/h); thus, the potential inhaled dose is expressed in units of μg. The environment-specific dose is then the sum of potential inhaled doses across the four activity levels, and the total inhaled dose across all environments is the sum of the environment-specific doses.

A file of detailed results that is available as an optional output contains the total time for each individual across all environments involved in the simulation, together with total exposure and total dose. At the user's option, environment-specific times, exposures and doses can also be stored in this file. The file can be used, for example, to develop estimates of the proportional contributions of each environment to total exposure/dose or to develop customized plots of exposure/dose distributions. An example of this type of file is provided in the user's guide for the model software. The file can also be used to calculate the average concentration in each environment (or across environments) for each trial, by dividing the time-integrated exposure by the time spent in the environment.

A variable labeled TIMEWT, which is contained in the file of activity profiles, is used in constructing statistics to summarize the exposure or dose distributions. This variable, developed by another contractor as part of the ARB-sponsored activity surveys, is used to compensate for unequal selection probabilities across survey participants. Without use of these weights, the mean of the exposure/dose distribution ($\bar{x}$) would be calculated as

$$\bar{x} = \left(\sum_{i=1}^{n} x_i \right) \Big/ n \qquad (5\text{-}1)$$

where x_i is the exposure or dose received by the ith individual and n is the number of individuals (trials) in the simulation. With weighting, as is done in model computations, the weighted mean is calculated as

$$\bar{x} = \left(\sum_{i=1}^{n} w_i \cdot x_i \right) \Big/ \sum_{i=1}^{n} w_i \tag{5-2}$$

where w_i is the value of the TIMEWT variable for each activity profile.

The formula for the weighted variance (s^2) of the distribution is as follows:

$$s^2 = \frac{\sum_{i=1}^{n} \left[w_i \cdot \left(x_i - \bar{x} \right)^2 \right]}{\left(\sum_{i=1}^{n} w_i \right) - 1} \tag{5-3}$$

An alternative formula, which is more efficient computationally, is used in model calculations (see, for example, Byer 1987, p. 519):

$$s^2 = \frac{\left[\sum_{i=1}^{n} \left(w_i \cdot x_i^2 \right) \right] - \dfrac{\left[\sum_{i=1}^{n} \left(w_i \cdot x_i \right) \right]^2}{\sum_{i=1}^{n} w_i}}{\left(\sum_{i=1}^{n} w_i \right) - 1} \tag{5-4}$$

The weighted standard deviation reported by the model as one of the summary statistics is the square root of the weighted variance.

The other statistics reported by the model are percentiles of the cumulative frequency distribution for exposure or dose. The TIMEWT variable is also utilized in computing these statistics. Individual exposure or dose values are first sorted from lowest to highest, and the weights are "carried along" with each value during the sort. The cumulative relative frequency associated with each value is then defined as the sum of the weights across all values less than or equal to that value, divided by the sum of the weights across all values. More specifically, the percentile associated with a specific ordered value (e.g., k^{th}) is calculated as

$$\frac{\sum_{i=1}^{k} w_i}{\sum_{i=1}^{n} w_i} \tag{5-5}$$

A specific percentile (e.g., 90th percentile) of the cumulative frequency distribution is computed by linear interpolation between the values associated with the nearest cumulative frequencies above and below the percentile of interest.

5.4 ASSUMPTIONS AND LIMITATIONS

Although the principles on which the exposure/dose component of the software model are based are scientifically and mathematically sound, the accuracy of the outputs is obviously tied to that of the inputs. Data on concentration distributions are not yet available for many of the environments. Although concentration distributions certainly can be assumed in such instances, such assumptions pose obvious limitations on the modeling results.

Even in cases where measured concentrations are available, there can be inaccuracies due to biases inherent in monitoring devices or sampling strategies. For example, concentration measurements in the residential environment are commonly taken in stationary environments such as bedrooms or living rooms. However, occupant

activities may be associated with close proximity to certain indoor sources (e.g., when cooking, showering, or using consumer products), leading to higher exposures than the well-mixed assumption (i.e., assumption of a homogeneous concentration throughout an indoor airspace) would suggest. As a result, exposure estimates tied to results of stationary indoor monitoring would tend to underestimate actual exposures to some degree. Prior to using the model, the user is advised to review articles or reports from which input data have been derived, as a basis for making judicious use of the model.

An implicit assumption in combining location/activity patterns with environment-specific concentrations in the model is that concentrations in different environments on the same day are independent. However, if outdoor concentrations are uniformly high over a broad area on a given day, and if the outdoor concentration has a significant impact on the indoor concentration, then concentrations across environments may be correlated to some degree. If there are such occurrences, the independence assumption should not bias the mean of the exposure/dose distribution but is likely to misrepresent the shape of the distribution; for example, exposures toward the upper tail of the distribution are likely to be underestimated by the model. We are not aware, at present, of any results of detailed studies that would suggest the extent of correlation that can be expected across environments for the various types of pollutants included in the model.

All estimates provided by the model are subject to some degree of uncertainty, and parameters such as the 90th or 95th percentile generally will be less stable than parameters such as the mean. That is, if one were to repeat a model run, keeping all inputs the same except the random number seed, then the mean values most likely would differ little between runs whereas the 95th percentile values could differ more substantially. As the sample size (number of trials) becomes larger, estimates for all parameters should become progressively more stable. An estimate of parameter stability can be obtained, for example, by running the model at least several times with all inputs constant except the random number seed. The variability for each parameter across repeated runs can then be assessed through a measure such as the coefficient of variation (i.e., the ratio of the standard deviation across runs to the average value across runs).

The representativeness of the activity profiles could be a potential limitation in cases where a relatively small subgroup is selected. Similarly, the weights associated with each profile may be questionable when relatively small subgroups are utilized. It is possible to add additional profiles to the data base for the model should additional time-diary surveys be performed in the future, but these will need to be compatible with the environment/activity grouping schemes used in the model and with the participant characteristics that are used as a basis for selecting population subgroups.

Section 6.0

CONCENTRATION DISTRIBUTIONS

6.1 MODELING APPROACH

Figure 6-1 provides an overview of the approach used to model indoor-air concentrations for specific types of structures (e.g., residences). Through the user-friendly interface provided by the model, the user provides a variety of inputs. In selected cases, the user makes some choices, but these choices are largely related to the manner in which the inputs are provided (e.g., manner in which to describe a distribution of structure volumes). The model provides access to some of the inputs that are available from existing studies, but in several cases these inputs are quite limited. The inputs and choices, described in detail in the user's guide for the model, are briefly summarized below.

As with the exposure/dose component of the model, the first step is to choose a pollutant. A choice of integration period is not needed here, as the model automatically calculates both hourly and daily average concentrations, thereby providing compatibility with all integration periods used for Level 1-2 calculations. User inputs are required for all factors--indoor sources, outdoor concentrations, building penetration factors, indoor sinks, building volumes, and indoor-outdoor air exchange rates--that are needed as inputs to an indoor-air model based on the principle of conservation of mass.

Some of the inputs are relatively complex. For indoor sources, for example, inputs are required for each type of pollutant-specific source that could be used in a building. These include interior finishing materials and furnishings, combustion sources, office equipment, and various types of consumer products. As discussed in Section 6.2, each source can be categorized as long-term, episodic or frequent; the nature of the inputs required and the model's handling of information varies by category. The source-specific information is used in determining an initial indoor-air concentration and in constructing a vector of hourly emission rates. The outdoor-concentration inputs are also used in determining the initial concentration and in constructing a vector of hourly outdoor concentrations.

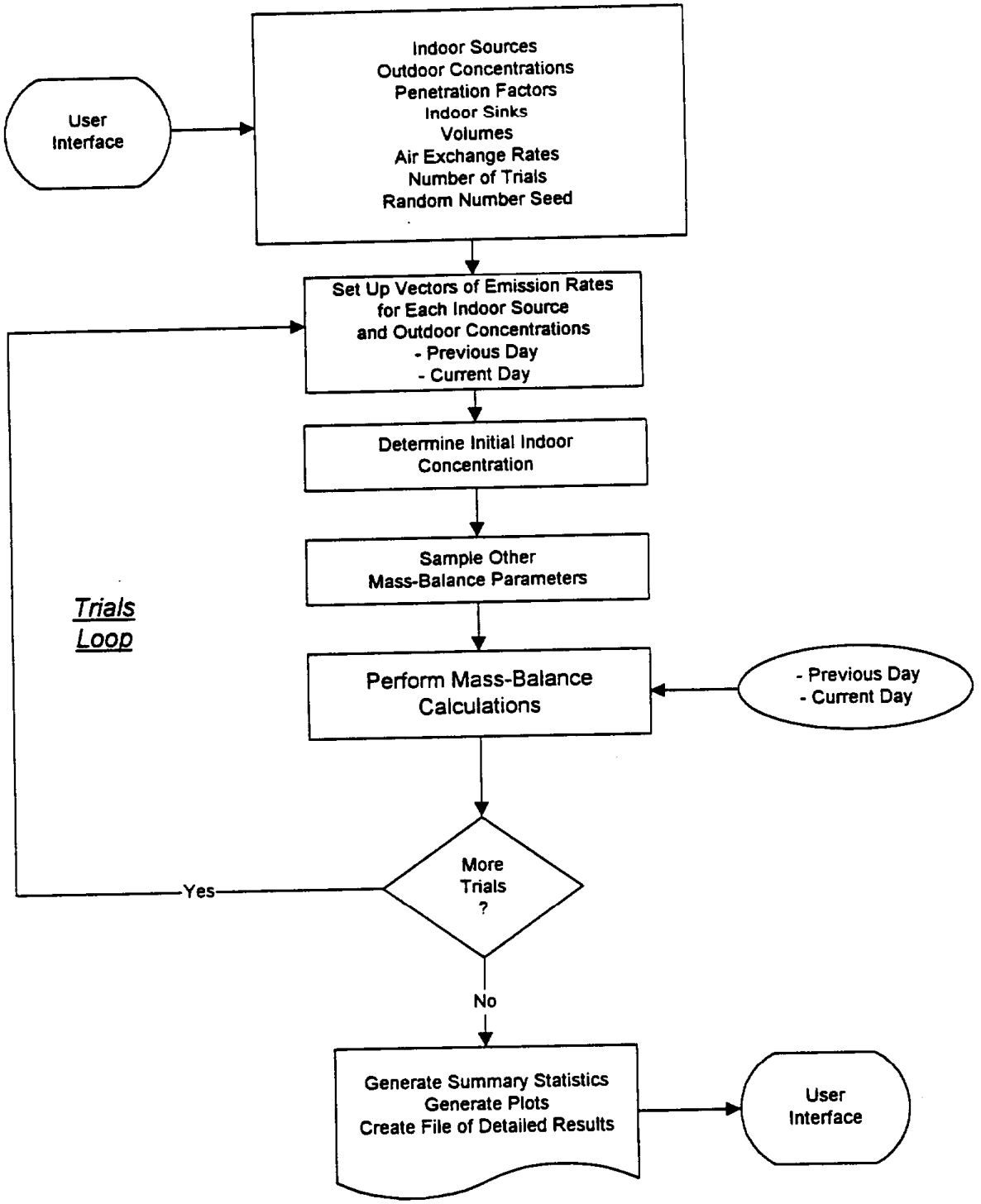

Figure 6-1. Overview of Concentration Calculation Routine

One key aspect of the model is the manner in which the indoor concentration is initialized at the start of the 24-hour period to be modeled. Although the initial concentration conceivably could be sampled at random from some distribution, in many cases a randomly chosen concentration would be inconsistent with the impacts of long-term, episodic and frequent sources that are sampled by the model as being present or used in a particular building. Instead, the model explicitly considers the impacts of the sampled long-term and episodic sources in initializing the indoor concentration for each building. In addition, the residual impacts of frequent sources are taken into account by modeling the day before the day of interest, in addition to modeling frequent sources on the day of interest to capture their concurrent impacts. Although the model calculates hourly concentrations for the day before (hereafter referred to as "the previous day"), these values are not available as model outputs; rather, they are used as part of the inputs to determine the starting (midnight) concentration for the day of interest ("the current day"), for which sequential hourly-average concentrations and the daily-average concentration are available as outputs from the model.

Once the user is finished with all inputs and signals completion to the model, the "trials loop" is initiated. The first part of this loop calculates an initial concentration for the previous day and computes vectors of hourly emission rates and hourly outdoor concentrations for the previous day and the current day. These values are passed, along with randomly chosen values for other input parameters (e.g., volume, air exchange rate), to a routine that performs mass-balance calculations for 24 hours on the previous day and 24 hours on the current day. At the completion of the trials loop, the model generates summary statistics and plots, produces a file containing detailed results for each trial, and returns control to the user. As with the exposure/dose portion of the model, the user can then view the plots and statistics, and choose whether to print or save statistics and plots and whether to save the file of detailed results.

6.2 MODEL INPUTS

Aside from the choice of pollutant, the primary inputs for concentration calculations are the indoor sources, outdoor concentrations, penetration factors, indoor sinks, volumes, air exchange rates, number of trials and random number seed. Sources of information and

key model features relating to each of these inputs are described in this section. The next section (6.3) describes the algorithms within the model that make use of these inputs.

6.2.1 Indoor Sources

For modeling purposes, indoor sources are classified into three types--long-term, episodic and frequent. Long-term sources, such as interior finishings, furnishings and some appliances, tend to be relatively static features of buildings. Episodic sources typically are used or present on a weekly, monthly or less frequent basis; some examples are carpet cleaning, painting and bringing home dry-cleaned clothes. Frequent sources tend to be used on a daily basis, often more than once a day; cooking, showering and tobacco smoking are good examples. As described below, the model requires different types of inputs for these three categories of sources (although there are some commonalities across categories), and the manner in which the inputs are used vary by category. For many sources, key information such as emission rates is very limited. Examples of indoor sources for the ten pollutants provided with the model are given in Table 6-1. Source-related inputs for three of these pollutants--chloroform, benzo[a]pyrene, and nitrogen dioxide--for which emission rates and other source inputs have been fairly well characterized are described in Section 7.2 and are included with the model software.

Long-term Sources. Inputs required for long-term sources are listed in Table 6-2. Many sources, because they contain a finite amount of material that can be emitted (as in the case of interior finishings that offgas VOCs), emit at a declining rate over time as the reservoir of available material is gradually depleted. Two of the inputs for such sources are the initial emission rate and the rate of decline in the emission rate over time. This rate is fairly well characterized, for example, for pressed-wood products or materials that emit formaldehyde. For some sources, such as pilot lights on a gas range, the rate of decline is zero; that is, the source emits at a nominally constant rate over time.

Table 6-1. Examples of Indoor Sources for Pollutants Provided with the Model

Pollutant	Primary Sources	Source Type
Benzene	Tobacco smoking Unburned natural gas (leaks) Various consumer products	Frequent Long-term Frequent/episodic
Benzo[a]pyrene	Tobacco smoking Wood burning	Frequent Frequent (winter)
Carbon monoxide	Range cooking Range pilot lights	Frequent Long-term
Chloroform	Chlorinated water supply	Frequent
Formaldehyde	Pressed wood products	Long-term
Nitrogen dioxide	Range cooking Range pilot lights	Frequent Long-term
PM_{10}	Tobacco smoking Wood burning Vacuuming/sweeping (resuspension)	Frequent Frequent (winter) Frequent
Perchloroethylene	Dry-cleaned clothes	Episodic
Total PAHs	Tobacco smoking Wood burning	Frequent Frequent (winter)
Trichloroethylene	Various consumer products	Frequent/episodic

Table 6-2. Inputs for Long-term Sources

INPUT	UNIT OF MEASUREMENT
Percent of buildings with source present	%
Initial emission rate	μg/h per quantity
Rate of decline in emission rate	$months^{-1}$
Quantity present	quantity[a]
Duration since installation	months

[a] At the user's option, this quantity can be expressed as a load factor in relation to volume of the building, in units of quantity/m^3.

Both the emission rate and the rate of decline are input to the model in the form of parameters describing a distribution of such rates across structure types such as residences (only one type of building, corresponding to an environment in the Level 1-2 model, can be modeled at a time). The forms of input are the same as those described previously (see Section 5.2.3) for environment-specific concentrations. As shown in Figure 6-2, the model samples whether or not a given source is present in a particular building (each building constitutes a trial in this model), based on the user input denoting the percent of buildings with this type of source present. If the source is sampled to be present, then the model samples the initial emission rate and the rate of decline from their respective distributions. A constant rate of decline for all buildings (e.g., zero decline in the case of pilot lights) can be described in the form of a normal distribution with a mean equal to the constant rate (zero in the example case) and a standard deviation equal to zero.

The model also samples the quantity of material present. The user can input this quantity either in absolute terms (e.g., square feet of finishing material) or in relation to building volume (e.g., square feet per unit volume). The choice is at the user's discretion--for some sources, such a pilot lights, the quantity (e.g., amount of fuel used) is not logically related to volume, whereas for sources such as carpeting the quantity (e.g., square feet installed) would be related. In the latter case, the model multiplies the sampled volume (in m^3, see Section 6.2.3) by the loading factor (quantity of material per m^3 of volume) to determine the quantity present in a given building.

After all inputs are sampled, the model uses these inputs to calculate (1) the initial indoor concentration (at midnight, at the start of the previous day) due to an indoor source and (2) hourly emission rates for the previous day and the current day. These contributions are ultimately added to those from other indoor sources (whether long-term, episodic or frequent) and passed to the mass-balance routine as a vector of 48 hourly emission-rate values--24 for the previous day and 24 for the current day.

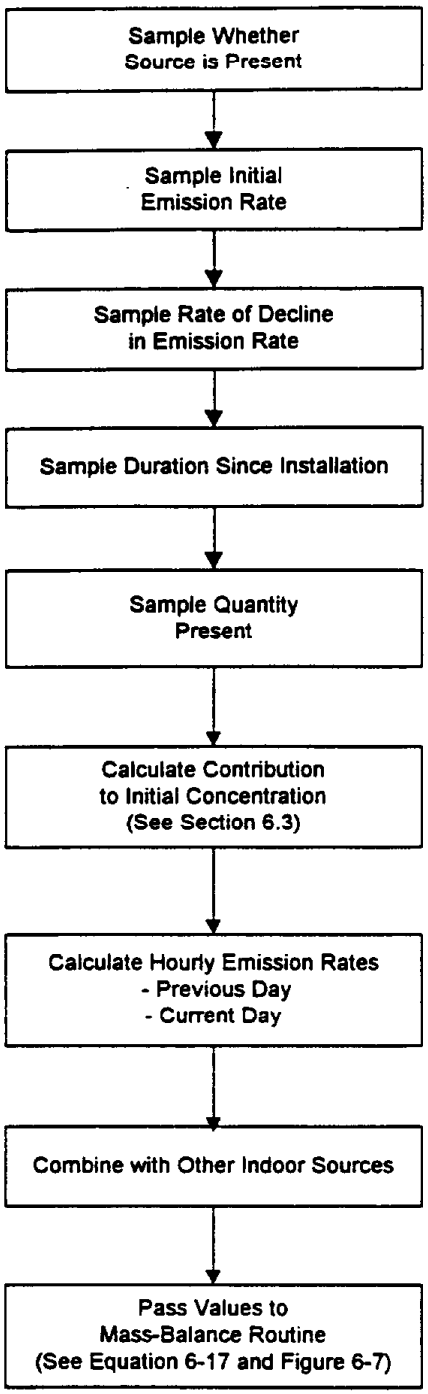

Figure 6-2. Routine for Sampling and Using Input Parameters for Long-term Sources

The inputs for episodic sources (Table 6-3) are essentially the same as those for long-term sources, except that the time scale for rate of decline and duration since use is in days rather than months. An additional distributional form that can be used to describe duration since use is the exponential, whose distribution can be characterized by a single parameter--mean time since last use. Several surveys (e.g., Westat 1987) have estimated this parameter as one of the descriptors of product usage patterns.

Table 6-3. Inputs for Episodic Sources

INPUT	UNIT OF MEASUREMENT
Percent of buildings with source present	%
Initial emission rate	μg/h per quantity
Rate of decline in emission rate	days^{-1}
Quantity present/used	quantity[a]
Duration since use	days

[a] At the user's option, this quantity can be expressed as a load factor in relation to volume of the building, in units of quantity/m^3.

Episodic Sources. The routine for sampling and applying the input parameters for episodic sources (Figure 6-3) is more complicated than for long-term sources. The calculations for contributions to the initial concentration (at midnight on the start of the previous day) and hourly emission rates are conditional on the rate of decline in emission rate and duration since use (Table 6-4). The value sampled for the duration since use is viewed from the perspective of a respondent being queried at midnight at the end of the current day. If the sampled value is zero days, then use occurred on the current day, with no contribution to the initial concentration or emissions on the previous day. If the sampled value is one day, then use occurred on the previous day; there is no contribution to the initial concentration (because this concentration occurs at the start of the previous day), but there may still be residual emissions on the current day (in addition to those on the previous day) if the rate of decline is other than zero. If the sampled value is two days or greater, then use occurred prior to the previous day. Depending on the duration since use and the rate of decline, there may be contributions to the initial indoor concentration and to the vectors of emission rates for either day. The other parameter used in this routine

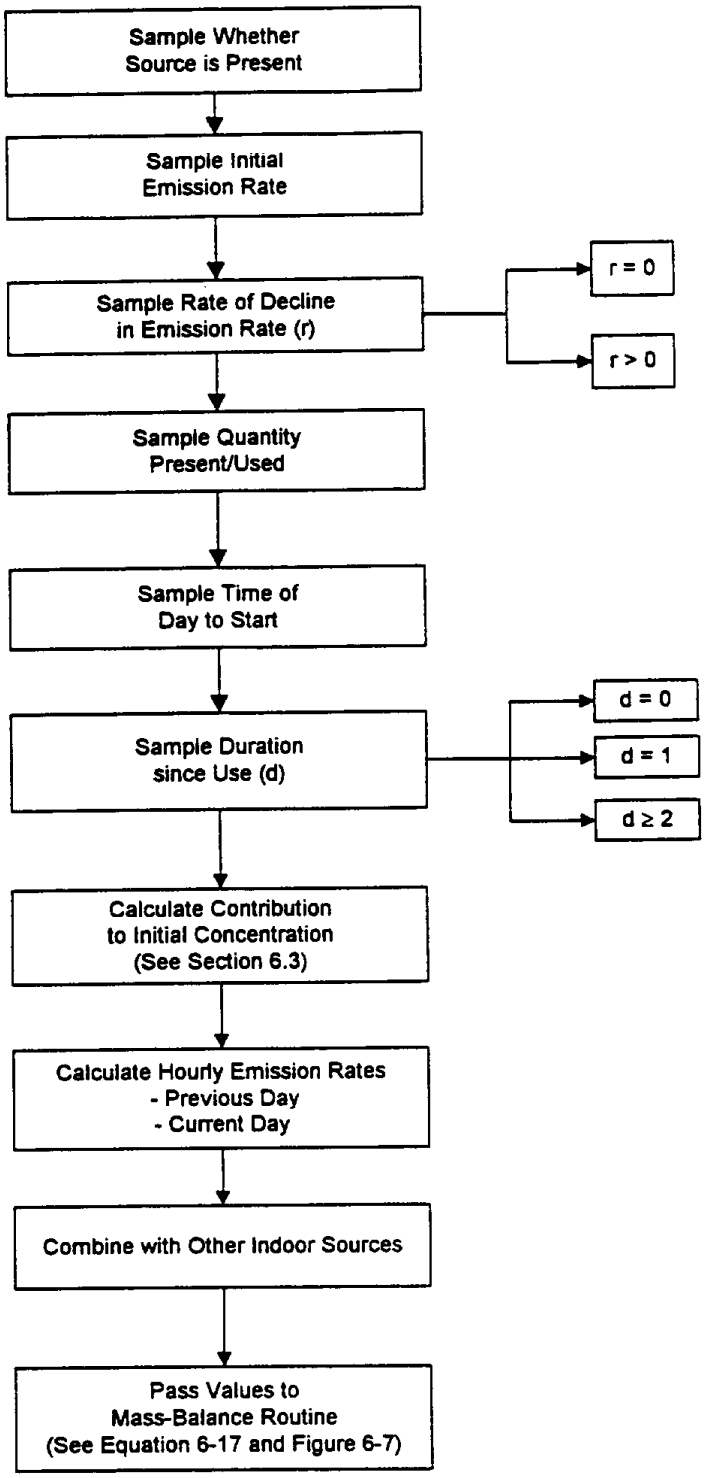

Figure 6-3. Routine for Sampling and Using Input Parameters for Episodic Sources

Table 6-4. Dependence of Calculations for Episodic Sources on Rate of Decline in Emission Rate and Duration Since Use

Rate of Decline (r)	Duration Since Use (d)	Impact on Calculations
0	0	Emission rates for previous day all = 0 Calculate emission rate for hour of use on current day
0	1	Calculate emission rate for hour of use on previous day Emission rates for current day all = 0
0	≥2	Calculate initial indoor concentration due to use Emission rates for previous day all = 0 Emission rates for current day all = 0
>0	0	Emission rates for previous day all = 0 Calculate emission rates for hour of use on current day and for subsequent hours
>0	1	Calculate emission rates for hour of use on previous day and all subsequent hours (including current day)
>0	≥2	Calculate initial emission rate at midnight of previous day Calculate emission rates for all hours on previous day and current day

is the time of day (hour) when the source is used. There is no user input for this parameter; the model samples an hour at random between 9:00 a.m. and 8:00 p.m.

Frequent Sources. More inputs are required for frequent sources (Table 6-5) than for the other two types. Three key additions are distributions for the number of episodes per day, time of day (start hour) when episodes can occur, and the duration of episodes. The rate of use can be expressed in terms of quantity used per unit time (i.e., quantity per minute) or per unit volume (quantity per m^3 of volume). In addition to emission rates and rates of use, the model independently samples the number of episodes that occur on the previous day and the current day (Figure 6-4). The model then goes through an episodes loop (Figure 6-5) whereby a start hour and duration are sampled for each episode, the hours affected (in terms of the period of use) are determined, and hourly emission rates are calculated and summed across episodes.

Table 6-5. Inputs for Frequent Sources

INPUT	UNIT OF MEASUREMENT
Percent of buildings with source present	%
Initial emission rate	μg/h per quantity
Rate of decline in emission rate	hours^{-1}
Rate of use/quantity used	quantity[a]/minute
Whether or not overlapping episodes are allowed	yes/no
Number of episodes per day	n
Distribution of start hour per episode	% share for each hour
Duration of episode	minutes

[a] At the user's option, this quantity can be expressed as a load factor in relation to volume of the building, in units of quantity/m^3.

A key feature of this routine relates to the notion of overlapping episodes. Overlapping episodes can be allowed by the user if it is plausible for more than one episode to occur simultaneously, or at least during the same hour. Overlapping episodes could occur, for example, for activities such as showering or smoking tobacco products, but are considerably less likely for cooking; a decision in this regard is left to the user. If overlapping episodes are not allowed, then the hour(s) associated with the first sampled episode for a given day are "blocked out" when sampling the second episode, and so on. When overlapping episodes are allowed, there is no such constraint. The start hour for each episode is sampled using probabilities proportional to the relative frequencies input by the user (see Table 6-5).

Linked Sources. A special feature of model inputs for indoor sources relates to the notion of linked sources. An example of linked sources, for a pollutant such as NO_2, is use of a gas range for cooking and pilot lights on the range. Cooking with the range is a frequent source whereas a pilot light is a long-term source (i.e., always in operation). A subset of gas ranges have pilot lights. In this case, after providing inputs for gas-range

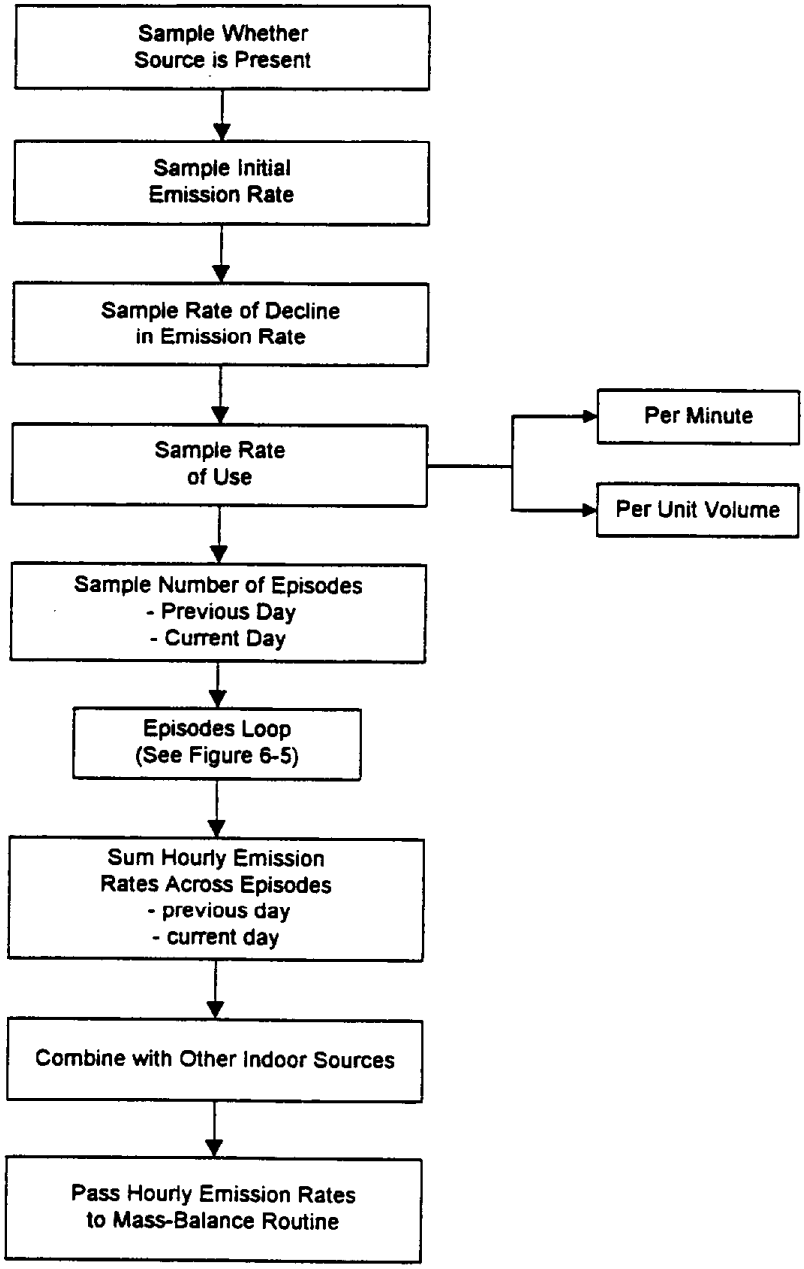

Figure 6-4. Routine for Sampling and Using Input Parameters for Frequent Sources

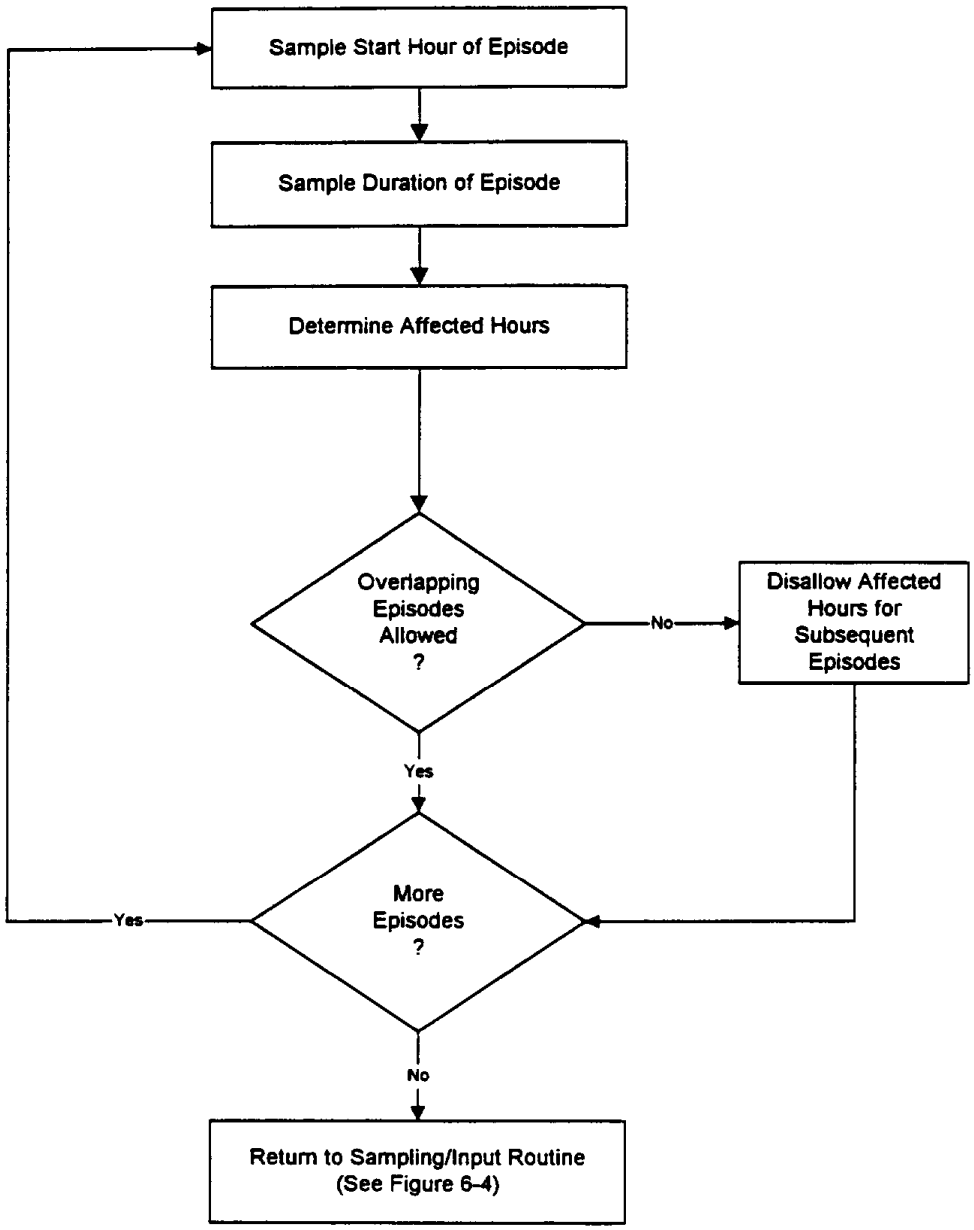

Figure 6-5. Episodes Loop for Frequent Sources

cooking and then saving these inputs, the user can provide inputs for pilot lights and link this source to the cooking source. The model will first sample whether a gas range is present in the building and, if so, will then sample whether the gas range has pilot lights. The guiding theme here is that the source present in a subset of cases should be linked to the source that is present in a greater percentage of cases. If both sources are present in the same fraction of cases, the direction of the linkage is arbitrary.

6.2.2 Outdoor Concentrations

There are two options for the user in providing inputs for outdoor concentrations--daily-average values or hourly-average values. In most cases, only daily averages are available from existing data sources. Sources of information (Table 6-6) for input parameters include state/local ambient air monitoring networks and some indoor-air monitoring studies wherein outdoor concentrations were measured in parallel with the indoor measurements.

Table 6-6. Current Sources of Information on Outdoor Concentration Distributions for Selected Pollutants

Pollutant	Source/Reference	Comments
Formaldehyde	ARB Air Toxics Monitoring Network	Statewide, 24-hour[a]
Volatile organic compounds - benzene - chloroform - perchloroethylene - trichloroethylene	ARB Air Toxics Monitoring Network Wallace et al. (1991) Sheldon et al. (1992a)	Statewide, 24-hour[a] Los Angeles, 24-hour and 12-hour[a] Woodland, 24-hour
Benzo[a]pyrene	Sheldon et al. (1992b)	Riverside, 24-hour
PM_{10}	ARB Monitoring Network Pellizzari et al. (1992)	Statewide, 24-hour[a] Riverside, 24-hour
Nitrogen dioxide	Wilson et al. (1986) Spengler et al. (1992)	Los Angeles, weekly Los Angeles, 48-hour
Carbon monoxide	Wilson et al. (1993)	Three utility service areas, 48-hour

[a] Limited published data; raw data were obtained from ARB and U.S. Environmental Protection Agency and analyzed to develop distributional information (see text).

Data from ARB's air toxics monitoring network (ARB 1988) for the years 1985 to 1990 were obtained and analyzed to develop distributional information for selected VOCs and benzo[a]pyrene. The database contained one record for each daily pollutant measurement at each monitoring site. After eliminating sites with limited data whose inclusion might have biased the results (e.g., due to seasonal variations), the records for each pollutant were statistically summarized across all monitoring sites. This procedure was followed for the state as a whole and for three regions (South Coast, San Francisco Bay area, and remainder of the state). For the TEAM study data, 12-hour daytime/ nighttime values obtained from EPA were averaged to derive 24-hour values, as previously described for the indoor concentration data (see Section 5.2.3).

The various alternatives for describing input distributions (normal, lognormal, etc.) apply to the case of daily-average outdoor concentrations. If the user elects hourly values, then the only alternative is provision of a data set (in .DBF format), with each record consisting of an index value (one, two, etc.) followed by 24 concentration values ordered from midnight to midnight. Further details on constructing data sets in this fashion are provided in Section 7.1.5 of the user's guide.

For daily values, the model samples a single value that is assumed to apply throughout the period to be modeled. For hourly values, the model samples one vector that is assumed to apply equally to the previous day and current day. The daily value is used as one of the inputs for calculating the initial indoor concentration, and the first value in the hourly vector is used for this purpose. In either case, the outdoor-concentration values are passed to the mass-balance routine as a vector of 48 values (24 for the pervious day, 24 for the current day) so that the routine can be applied without unnecessary branching.

6.2.3 Other Mass-Balance Parameters

For all other mass-balance parameters -- indoor sinks, building penetration factors, volumes, and air exchange rates -- a single value is sampled for each structure. Each of these parameters is invoked through the usual array of options for describing distributional information.

Indoor Sinks. Aerosols and reactive gas-phase pollutants disappear more rapidly than dilution by air exchange would suggest. Such processes are expressed in the model in terms of a pollutant-specific first-order decomposition rate with dimensions of inverse hours. In addition to gravitational settling, aerosols are subject to surface deposition processes. For fine particles (< 1 μm), loss rates to deposition are generally somewhat smaller than loss rates for air exchange; loss rates for particles approaching 10 μm, however, can be ten or more times larger (see, for example, Nazaroff et al. 1993, Nazaroff and Cass 1989). Reactive gases like NO_2 are subject to gas-phase chemistry as well as interactions with fixed surfaces (walls, floors, furnishings). Loss rates are determined by the nature and distribution of surfaces. Numerous experimental studies of NO_2 deposition suggest a range of 0.2 to 1.3 h^{-1} (see, for example, Nazaroff et al. 1993). Most indoor materials will act as reversible sinks for organic vapors (see, for example, Colombo et al. 1993, Tichenor et al. 1991). Under these circumstances, the sinks will affect the time needed to achieve equilibrium or steady-state conditions, but the steady-state concentration is virtually identical to that which would have prevailed in the absence of the reversible sink. Carbon monoxide is nonreactive under normal conditions.

Building Penetration Factor. As air travels through intentional and unintentional openings in the building envelope, gaseous and particle-bound pollutants have an opportunity to interact with surfaces encountered along the way. We are generally unconcerned with describing pollutant losses to the building envelope for air that is leaving the building. For air that enters the building, however, this scrubbing effect can remove substantial amounts of some pollutants so that, in the absence of indoor sources, indoor levels can be significantly lower than those outdoors. This process is incorporated into the model as the building penetration factor, a dimensionless fraction representing the amount of pollutant of outdoor origin that survives the journey through the building envelope. The penetration factor is pollutant-specific, and ranges between zero (no penetration/complete scrubbing) and one (complete penetration/no scrubbing). The penetration factor is not measured directly because air enters the building through an extremely large number of openings distributed throughout the building envelope. Rather, it is estimated through the indoor/outdoor concentration ratio from data collected at sites that are free of indoor

sources and where outdoor levels are relatively high (see references listed in Table 6-6). For relatively stable pollutants like CO, and moderately reactive pollutants like NO_2, the penetration factor is one (see, for example, Traynor et al. 1989). Observed penetration factors for PM_{10} near 0.7 are explained by removal of larger particle sizes (Colome et al. 1990). For organic vapors, the ubiquity of indoor sources generally forces indoor levels to exceed outdoor levels (see, for example, Brown et al. 1994, Sheldon et al. 1992a), making it difficult to estimate the penetration factor. For benzo[a]pyrene, modeling efforts in one study (Sheldon et al. 1992b) suggested a penetration factor of 0.6.

Volume and Air Exchange. Separate measurement surveys are not generally commissioned to directly evaluate the range and distribution of residential volumes and air exchange. Such data are normally collected through indoor air quality and exposure studies. Table 6-7 summarizes published studies that constitute relevant data sources.

Table 6-7. Current Sources of Information on Indoor Volumes and Air Exchange Rates

Reference	House Volume[a]	Air Exchange[a] Rate	Comments
Wallace et al. (1991)	X (259)	X (98, 90)	Los Angeles, February and July
Wilson et al. (1986)	X (416)	X (571, 426, 371)	Los Angeles, January, March and July
Wilson et al. (1993)		X (287)	Three utility service areas, January to March
Sheldon et al. (1992b)		X	Riverside, fall
ADM Associates (1990)	X (30)	X (29)	Sacramento, Riverside, July/August, single-family houses built before 1975
Berkeley Solar Group and Xenergy (1990)	X (40)	X (40)	North Coast, South Coast, South Valley, North Valley, August-January, single-family houses built between 1984 and 1988

[a] Number of measurements in parentheses, where reported.

6.2.4 Number of Trials /and Random Number Seed

For the Level 3 model, the maximum number of trials that can be run is constrained by available conventional memory, due to limitations on the size of various arrays used in the calculation routines. For example, if 585K of conventional memory is available, then the maximum number of trials is approximately 1,300. This number should be sufficient for most applications. For example, the hourly statistics would be based on 31,200 (1300 x 24) values, because 24 sequential hourly concentrations are modeled for each trial. For daily concentrations, the estimates for the mean and standard deviation should be quite stable, whereas for an upper-percentile value such as the 95th percentile there may be some instability. If desired, a larger number of trials can be generated by making two or more separate runs, keeping all inputs the same except the random number seed.

As with the Level 1-2 model, the random number seed for the simulation can be input by the user or determined by the model from the system clock. Outputs from the Level 3 model include summary statistics and plots of the concentration distribution for the modeled environment, as well as a file containing detailed (trial-by-trial) results. The summary statistics can be imported directly into the Level 1-2 model.

6.3 KEY ALGORITHMS

The key algorithms in the model relate to estimation of emission rates, calculation of the initial indoor concentration, and calculation of hourly-average indoor concentrations over a 48-hour period. The estimation of hourly emission rates, involving summation of contributions from long-term, episodic and frequent sources, was discussed previously (see Section 6.2.1).

The model uses a single indoor compartment. For the calculation of hourly-average concentrations, this compartment is "fed" by internal and external (outdoor) sources that are held constant over each hour. The sources can change from hour to hour, but do not change within an hour. A first-order loss also is allowed in the compartment. This system can be represented as shown in Figure 6-6.

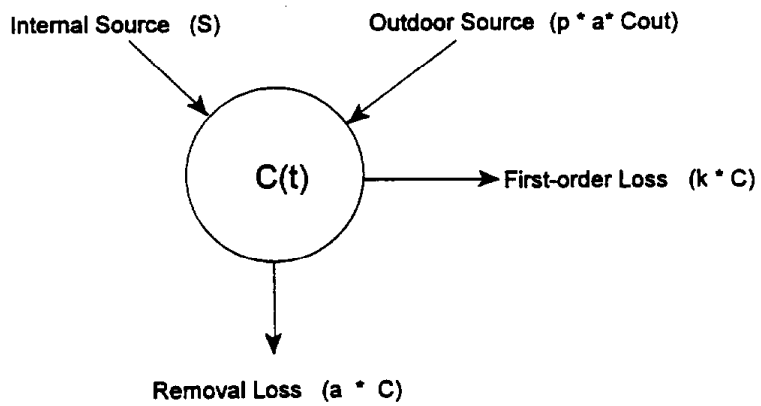

Figure 6-6. Schematic Representation of Single Compartment Indoor-air Model

From first principles, the mass-balance equation across the compartment boundaries shown in Figure 6-6 is as follows:

$$V\frac{dC}{dt} = \sum \text{Gains} - \sum \text{Losses} \qquad (6\text{-}1)$$

Where
V = compartment volume
C = pollutant concentration (mass per volume)
t = time
Gains = rate of pollutant gains (mass/time)
Losses = rate of pollutant losses (mass/time)

Substituting the gain and loss mechanisms shown in Figure 6-6, equation (6-1) becomes:

$$V\frac{dC}{dt} = S + pQC_{out} - kVC - QC \qquad (6\text{-}2)$$

Where
S = source rate (mass/time)
p = penetration factor (dimensionless)
Q = Airflow rate into and out of compartment (volume/time)
k = first-order loss rate (1/time)

Rearranging:

$$\frac{dC}{dt} = \frac{S}{V} + paC_{out} - kC - aC \qquad (6\text{-}3)$$

Where
a = air exchange rate (hr^{-1})
= Q/V

Let $G = \frac{S}{V} + paC_{out}$

and $L = -(k+a)$.

By substitution equation (6-3) becomes:

$$\frac{dC}{dt} = G + LC \qquad (6\text{-}4)$$

Rearranging and integrating both sides of equation 6-4 yields:

$$\int \frac{dC}{(G + LC)} = \int dt \tag{6-5}$$

Since the model assumes that the source strength, air exchange rate, first-order loss rate and outdoor concentration are constant over each hour, equation (6-5) can be integrated to obtain the indoor concentration as a function of time.

The integration formula used to integrate equation (6-5) can be found in any standard table of integrals:

$$\int \frac{dx}{ax + b} = \frac{1}{a} \ln (ax + b) \tag{6-6}$$

Applying the integration formula in equation (6-6), equation (6-5) yields:

$$\frac{1}{L} \ln |LC + G| \Big]_{C_o}^{C(t)} = t \Big]_{o}^{t} \tag{6-7}$$

Solving for the boundary conditions yields:

$$\frac{1}{L} \left[\ln \left(LC(t) + G \right) - \ln \left(L\ C_o + G \right) \right] = t \tag{6-8}$$

Or:

$$C(t) = \frac{e^{Lt}(L\ C_o + G) - G}{L} \tag{6-9}$$

Rearranging:

$$C(t) = \frac{-G}{L}\left(1 - e^{Lt}\right) + C_o e^{Lt} \tag{6-10}$$

A basic theorem from the calculus states that the average of a function f(t) over a time interval [O,T] is:

$$\bar{f} = \frac{1}{T}\int_o^T f(t)\ dt \tag{6-11}$$

Substituting C(t) in equation (6-10) for f(t) in equation (6-11), we get:

$$\overline{C} = \frac{1}{T}\int_o^T C(t)\ dt = \frac{1}{T}\int_o^T \left(-\frac{G}{L}(1 - e^{Lt}) + C_o\ e^{Lt}\right)dt \tag{6-12}$$

Rearranging:

$$\overline{C} = \frac{1}{T}\left[\int_o^T \frac{-G}{L} + \frac{G}{L}\int_o^T e^{Lt}dt + C_o\int_o^T e^{Lt}dt\right] \tag{6-13}$$

The integration formula used to integrate equation (6-13) is as follows:

$$\int e^{ax}\,dx = \frac{1}{a}\,e^{ax} \tag{6-14}$$

Applying the integration formula given by equation (6-14) to equation (6-15) yields:

$$\overline{C} = \frac{1}{T}\left[-\frac{GT}{L} + \frac{G}{L^2}\left[e^{LT} - e^{0}\right] + \frac{C_o}{L}\left[e^{LT} - e^{0}\right]\right] \tag{6-15}$$

Rearranging:

$$\overline{C} = \frac{1}{T}\left[\frac{-GT}{L} - (1 - e^{LT})\frac{G}{L^2} - (1 - e^{LT})\frac{C_o}{L}\right] \tag{6-16}$$

Adapting equation (6-14) to this problem by using the endpoint concentration of the previous hour as the initial concentration for the current hour, the average concentration for hour I is expressed as:

$$\overline{C_i} = \frac{-C_{i-1}}{LT}\,(1 - e^{LT}) - \frac{G}{T(L^2)}\,(1 - e^{LT}) - \frac{G}{L} \tag{6-17}$$

Where $\overline{C}_i$ = average concentration for hour I
C_{i-1} = concentration at the endpoint of the previous hour
$L = -(k + a)$
$G = S/V + pa\,C_{out}$
T = averaging period (1 hour)

The average hourly concentration can be calculated for each hour by using equation (6-10) to calculate the endpoint concentration of the previous hour and equation (6-17) to calculate the average concentration for the current hour. A flowchart summarizing the calculation routine is given in Figure 6-7. The second step in the flow chart, labeled "Set up factors to save execution time," refers to calculation of the exponential term (e^{LT}) in

Equation (6-17). Execution time is saved by calculating this term only once, rather than repeatedly (i.e., 48 times) for each hour in the loop where hourly concentrations are calculated (shown as the shaded box in Figure 6-7).

For each trial, the model must be initialized with a realistic estimate of the initial indoor air-concentration. This is accomplished by considering indoor sources that are potentially active during the time period prior to the modeled period. For purposes of this discussion, the modeling period is divided into three distinct periods: (1) period of interest, (2) the initialization period, and (3) the pre-initialization period. The period of interest is the 24-hour period for which modeling results are reported. The initialization period is defined as the 24-hour period immediately preceding the period of interest. The pre-initialization period is defined as any time preceding the initialization period. The initial concentration is defined as the concentration at the start of the period of interest, and the pre-initial concentration is defined as the concentration at the start of the initialization period.

For purposes of initialization, the sources are divided into two categories: (1) long-term or episodic sources likely to have an impact on the indoor concentration more than 24 hours after the source is activated, and (2) short-term or frequent sources for which the impact is negligible after 24 hours. The model initializes for a case (1) source by estimating the concentration at the start of the initialization period and then modeling the source's impact on the indoor concentrations during the initialization period to obtain an initial concentration prior to the period of interest. The model initializes for a case (2) source by modeling the source's impact during the 24-hour initialization period.

The concentration at the start of the initialization period is impacted by two of the three types of indoor sources defined in section 6.2.1-- long-term and episodic. If any of these sources are active during the 24-hour period immediately before the modeled interval, the impact of the source on the indoor concentration is taken into account by estimating its contribution to the initial indoor concentration. The impacts of these sources on the initial indoor concentration are additive; therefore, they are addressed separately, and their contributions are summed to estimate the initial indoor contributions as follows:

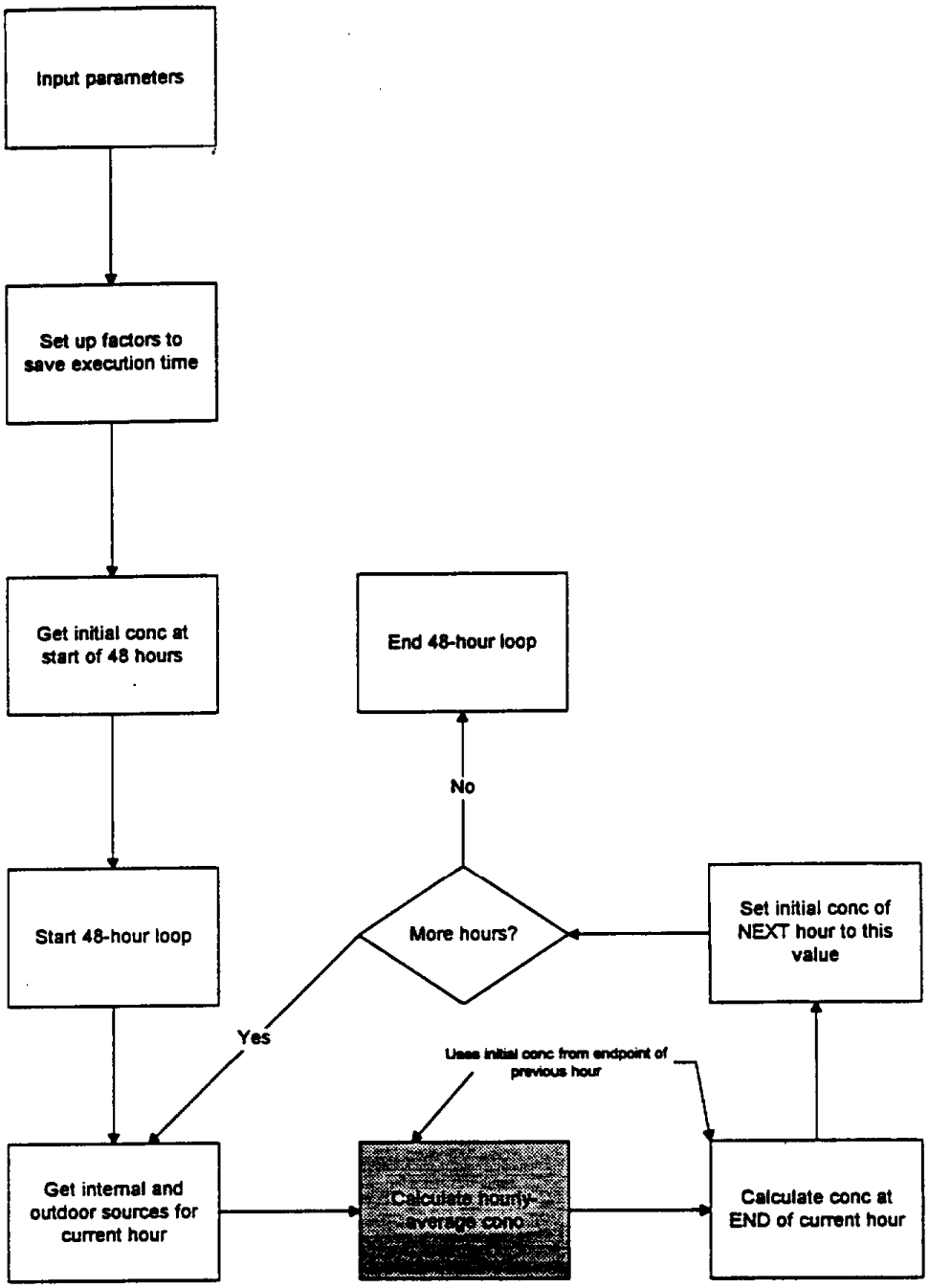

Figure 6-7. Flowchart of Routine for Calculating Hourly-Average Indoor Concentrations Over a 48-hour Period

$C_{pre\text{-}init}$ = (Contribution of Long-term Sources) + (Contribution of Episodic Sources) + Outdoor Contribution (6-18)

Where $C_{pre\text{-}init}$ = concentration at the start of the initialization period.

The outdoor contribution is the impact of the outdoor concentration on the indoor concentration in the absence of indoor sources. This is represented mathematically by the following equation:

$$\frac{dC_{OC}}{dt} = -(k+a)C_{OC} + paC_{out} \tag{6-19}$$

Where C_{OC} = contribution of the outdoor concentration to the initial indoor concentration.

Assuming that the outdoor concentration is relatively constant, the steady-state approximation provides the best estimate of the impact of the outdoor concentration on the indoor concentration. Therefore, equation (6-19) is solved by setting dC/dt = 0 to yield the following equation:

$$C_{OC} = \frac{paC_{out}}{k+a} \tag{6-20}$$

The steady-state approximation also provides the best estimate of the contribution of each long-term source. This contribution is estimated by using the emission rate of each source at the time of initialization and solving the steady-state equation in the absence of all other sources, as follows:

$$\frac{dC_{LT}}{dt} = \frac{S}{V} - (k+a)C_{LT} \tag{6-21}$$

Where S = source strength at the start of the initialization period (mass/time)
C_{LT} = contribution of the long-term source to the initial indoor concentration.

Solving for the steady-state solution by setting dC/dt = 0 yields the following equation:

$$C_{LT} = \frac{\frac{S}{V}}{k+a} \tag{6-22}$$

Episodic sources that are simulated to occur during the pre-initialization period also are taken into account in determining the concentration at the start of the initialization period. Two types of episodic release are (1) those that occur in a relatively short time period (on the order of an hour or two), and (2) releases that exponentially decline with age. Carpet cleaning is an example of an episodic source with a burst release, and painting is an example of an exponentially declining release.

The episodic source with a burst release is represented by a 1-hour constant release period, where the total mass is released at a constant rate. The resulting indoor concentration is then allowed to decline due to air exchange and first-order losses. The release period is mathematically described by equation (6-2) and equation (6-10). For this case, the outdoor contribution is calculated separately, the initial concentration is zero, and the release period is 1 hour. Therefore, equation (6-10) reduces to:

$$C_{BR}(t=1\ hour) = \frac{-S_{BR}}{LV}\,(1-e^{L}) \tag{6-23}$$

Where S_{BR} = Burst release source strength (mass/time)

$$= \frac{\text{Total mass released by source}}{1\ \text{hour}}$$

$L = -(k + a)$

Calculating the decrease due to air exchange and first-order losses is also mathematically represented by equation (6-10), by setting the source strength to zero, and using C(t = 1 hour) for the initial concentration (C_o). Again, for this case, the outdoor contribution is not included. Therefore, equation (6-10) becomes:

$$C_{BR}(t=T) = C(t=1\,hr)\ e^{LT} \tag{6-24}$$

Where C_{BR} (t = T) = the contribution of the burst-release episodic source to the initial indoor concentration
T = the time between the end of the release and the start of the initialization period.

The episodic source that exponentially declines with age uses the same technique as the long-term source (described earlier). The source's contribution to the initial concentration is estimated by the steady-state solution shown in equation (6-22) as follows:

$$C_{ED} = \frac{\frac{S}{V}}{k+a} \tag{6-25}$$

Where C_{ED} = the contribution of the exponentially declining episodic source to the initial indoor concentration.

Therefore, the indoor concentration at the start of the initialization period is:

$$C_{pre-init} = C_{OC} + \sum_{i=1}^{M} C_{LT_i} + \sum_{j=1}^{N} C_{BR}(t=T)_j + \sum_{k=1}^{P} C_{ED_k} \tag{6-26}$$

Where i, j, k = variables representing individual sources
M = total number of long-term sources
N = total number of burst-release episodic sources
P = total number of exponentially declining episodic sources.

At this point, the model proceeds with calculation of hourly concentrations for the 24-hour initialization period and the 24-hour period of interest, for which the results are reported.

6.4 ASSUMPTIONS AND LIMITATIONS

As with the exposure/dose calculations, the principles on which the emission and concentration calculations are based are scientifically and mathematically sound, but the accuracy of the outputs can be no better than that of the inputs. At the present time, there is a notable lack of information for many of the indoor sources as well as the pollutant-specific penetration factors and decay rates. The model is further constrained by the single-compartment and well-mixed assumptions, although both assumptions are reasonable in light of the requirement that the model be reasonably accurate without being overly complex. Indeed, addressing multiple compartments, departures from the well-mixed assumption and phenomena such as reversible sinks would require a base of information that cannot be expected for many years to come. On the other hand, a steady stream of new information is being generated for more basic requirements such as emission rates, and patterns of time-related decline in rates, for various types of products and materials.

One of the key features of the concentration model is the ability to use it in examining potential impacts of alternative mitigation strategies such as reformulation or even banning of certain materials or products. In the case of a ban, the model could simply be re-run with the banned source removed from the input stream. In the case of reformulation, a new distribution of emission rates (if available) could be input for the source of concern and the model re-run. Comparisons of outputs from related model runs (e.g., with versus without mitigation options) need to consider the statistical uncertainty in estimated parameters of the indoor concentration distribution. As with the exposure/dose model, the stability of parameter estimates can be characterized through repeated model runs whereby all input parameters are kept constant from run to run except the random number seed.