



Draft Design and Implementation Plan for EPA's Multi-Scale Motor Vehicle and Equipment Emission System (MOVES)

Draft Design and Implementation Plan for EPA's Multi-Scale Motor Vehicle and Equipment Emission System (MOVES)

John Koupal, Mitch Cumberworth, Harvey Michaels, Megan Beardsley, David Brzezinski

Assessment and Standards Division
Office of Transportation and Air Quality
U.S. Environmental Protection Agency

NOTICE

*This technical report does not necessarily represent final EPA decisions or positions.
It is intended to present technical analysis of issues using data that are currently available.*

*The purpose in the release of such reports is to facilitate the exchange of
technical information and to inform the public of technical developments which
may form the basis for a final EPA decision, position, or regulatory action.*

Table of Contents

Table of Contents	i
List of Tables	ii
List of Figures	ii
1. Introduction	1
2. Use Cases	4
3. Overview of MOVES Design.....	7
3.1. Design Principles.....	8
3.2. Core Model and Implementations	11
4. Software Design Layers.....	12
4.1. Control 1: System Specifications	12
4.2. Control 2: Run Specification.....	13
4.3. Control 3: Monte Carlo Controller.....	13
4.4. Control 4: Master Loop	14
4.5. Application 1 and Database 1.....	14
4.6. Utilities.....	16
4.7. Database 2	16
4.8. Application 2 and Database 3.....	16
5. Data Flow Design.....	17
5.1. The Core Model and Output Processing	18
5.2. The Master Generator Loop, Internal Emission Control Strategies, and Data Generators	21
5.3. The User Interface	25
6. MOVES Databases.....	29
6.1. Input Databases	29
6.2. Execution Databases	31
6.3. Output Database	31
7. Mathematical Formulation and Data Inputs.....	33
7.1. On-Road Implementation	35
7.2. Off-Road Implementation	55
8. Software Implementation	57
8.1. Software Development Methods Chosen for MOVES	58
8.2. Software Development Tools Selected for MOVES.....	58
8.3. Distributed Processing.....	60
References	60
Appendix A. MOVES GUI Details (separate document)	
Appendix B. MOVES output Database Design (separate document)	

List of Tables

Table 2-1. MOVES Use Cases	5
Table 3-1. MOVES Emission Processes	7
Table 3-2. Total Activity Basis by Process	9
Table 3-3. Example Source and Operating Mode Bin	10
Table 4-1: MOVES Software Design Layers	12
Table 7-1: Implementations Proposed for MOVES	34
Table 7-2: Proposed MOVES On-Road Vehicle Classification Scheme.....	37
Table 7-3: Overview of Total Activity Generator (TAG) Steps	39
Table 7-4: Operating Mode Parameters By Emission Process.....	51
Table 7-5: Proposed VSP Bin Definitions	52
Table 7-6: Overview of Operating Mode Distribution Generation For VSP	53
Table 7-7: Level of Disaggregation for Soak Time and Time Since Start	55

List of Figures

Figure 5-1. Data Flow Design: MOVES Core Model.....	18
Figure 5-2. Data Flow Diagram: MOVES Master Generator Loop, Internal Emission Control Strategies, and Data Generators	23
Figure 5-3. Data Flow Diagram: MOVES User Interface.....	26
Figure 6-1. MOVES Output Database Design	32

1. Introduction

The Clean Air Act requires EPA to develop and regularly update emission factors for all emission sources. Pursuant to this charge, EPA's Office of Transportation and Air Quality (OTAQ) has developed a number of emission and emission factor estimation tools for mobile sources, including MOBILE (for highway vehicles) and NONROAD (for off-road mobile source pollutants). EPA is proposing to update these tools with the Multi-Scale Motor Vehicle and Equipment Emission System (MOVES), which is intended to include and improve upon the capability of these tools and, eventually, to replace them with a single, comprehensive modeling system.

The National Research Council (NRC) published a thorough review of EPA's mobile source modeling program in 2000.¹ The NRC provided several recommendations for improving EPA's mobile source modeling tools, including (a) the development of a modeling system more capable of supporting smaller-scale analyses; (b) improved characterization of emissions from high-emitting vehicles, heavy-duty vehicles, and off-road sources; (c) improved characterization of particulate matter and toxic emissions; (d) improved model evaluation and uncertainty assessments; and (e) a long-term planning effort coordinated with other governmental entities engaged in emissions modeling.

A particular focus of the NRC report was the need to provide emission factors and tools that will allow the estimation of emissions at finer analysis scales. Historically, EPA's mobile source emission estimation tools and underlying emission factors have been focused on the estimation of mobile source emissions based on average operating characteristics over broad geographical areas. Examples of this scale of analysis are the development of SIP inventories for urban nonattainment areas and the estimation of nationwide emissions to assess overall trends. In recent years, however, analysis needs have expanded in response to statutory requirements that demand the development of finer-scale modeling approaches to support more localized emission assessments. Examples include "hot-spot" analyses for transportation conformity and evaluation of the impact of specific changes in transportation systems (e.g., signalization and lane additions) on emissions.

We have adopted most of the NRC's recommendations as our objectives in designing MOVES. The MOVES design objectives include:

- applicability to a wide range of spatial and temporal scales,
- inclusion of all mobile sources and all pollutants,
- addressing model validation and the calculation of uncertainty,
- ease of updating the model,
- quality assurance,
- ability to interface with other models, and
- ease of use.

These objectives have shaped our development plan for MOVES, as detailed in the rest of this paper.

We have begun and will continue to develop MOVES with extensive coordination and outreach. A cross-agency team representing OTAQ, the Office of Air Quality Planning and Standards (OAQPS), the Office of Research and Development (ORD), and Regional EPA offices produced an issue paper containing an initial proposal for the model in April 2001.² Since then, we have coordinated with states, metropolitan planning organizations, the U.S. Department of Transportation, consultants, academics, and the “FACA Modeling Workgroup”, which is the Modeling Workgroup of the Mobile Source Technical Review Subcommittee (MSTRS) of the Clean Air Act Scientific Advisory Committee (CAASAC), a committee of experts from government, industry, and academia established under the Federal Advisory Committee Act (FACA). This continuing coordination will enable MOVES to benefit from advances in mobile source emission modeling and to meet the needs of the user and stakeholder community.

This report presents the proposed MOVES design and implementation plan and includes our analysis of “use cases,” the design and implementation plan proposed to meet these use cases, specific mathematical formulations and data input requirements to implement this design, and software choices which are proposed to implement the design. The MOVES design is centered on the three primary components of mobile source emission modeling: characterizing the fleet, characterizing the fleet’s activity, and characterizing the emissions associated with that activity. Detailed design aspects for the fleet and activity components are presented in this document, primarily in the mathematical formulations presented in Section 7. The characterization of emissions is discussed only generally in this document, with our detailed proposal for this element to be published as a separate report by the end of 2002.

The current implementation schedule for MOVES is to first release a greenhouse gas implementation for on-road sources in 2003, termed MOVES GHG. The first iteration of MOVES GHG would support development of national inventories and projections at the county-level for Fuel Consumption, CO₂, N₂O, and CH₄, by Fall 2003. A second iteration would add future projections, policy evaluation capabilities, and the ability to account for “life-cycle” (i.e., well-to-pump) effects in the estimate of emissions, by the end of 2003. A third iteration would add local inventory (roadway link/traffic zone) analysis capability by mid-2004.

MOVES GHG will serve important needs within EPA and in the model user community. A comprehensive model that allows bottom-up estimation of greenhouse gas emissions from on-road sources is currently not available and Intergovernmental Panel on Climate Change guidelines recommend the development of such a model.³ It will also give the broader user community an opportunity to work with MOVES while the full-scale version of the model is developed. Choosing a greenhouse gas model as the first implementation of MOVES allows us to start with a small scope relative to all of the considerations that go into modeling of ozone-precursor and criteria pollutants. For example, the model will not need to address evaporative emissions or implement finer scale emission analysis.

The next phase in implementing MOVES, planned for Fall 2005, would add HC, CO, NO_x, SO₂, PM, NH₃, and air toxics for on-road sources and would allow the estimation of emissions at all analysis scales. This implementation would replace the current MOBILE6

model. The time between MOVES GHG and the full MOVES implementation would be focused on the development of emission algorithms for these pollutants, since the software aspects of the model will have been significantly developed for the greenhouse gas implementation.

In order to manage the scope of the project, new estimates of off-road emissions would not be integrated into MOVES until after the on-road implementations were developed. Prior to the development of new estimates for off-road emissions, we are considering integrating the NONROAD model into the MOVES framework in conjunction with the on-road releases. This would allow the estimation of emissions from all mobile sources within one modeling system, an important objective of MOVES.

2. Use Cases

The preparation of use cases is a necessary first step in software design and follows from a basic assessment of model needs. In developing a design for MOVES, we first assessed the many ways in which the model would be used, informed by the NRC report and interviews with subject-matter experts. With the help of a contractor (MCNC), we interviewed expert users for each use case. In addition to this formal interviewing process, we also consulted closely with researchers and users within EPA's Office of Research and Development (ORD) and Office of Air Quality Planning and Standards (OAQPS), as well as with the FACA Modeling Workgroup, comprised of experts from industry, academia, government, and consulting firms. From this assessment we developed a list of essential use cases, which explore, at a high level and independent of particular programming languages and hardware, how users need to or would like to use the software.

The result of this analysis was a list of fundamental use cases which have driven the MOVES design. Table 2-1 lists the use cases together with the subject matter experts interviewed on these areas.

In addition to these broad model purposes, through consultation with model users we also addressed use cases from the perspective of user interaction:

- A powerful and versatile GUI and a batch, command-line interface.
- Flexibility of input and output formats.
- Multiple options for output processing
- Ease of comparison of results from multiple model runs
- Ease of transition from MOBILE/NONROAD to MOVES

These reflect overarching design considerations that have been considered in the development of a design which serves the broad purposes listed in Table 2-1.

An important, but difficult-to-define, aspect of use cases is performance specification. All users would like the software to run as quickly as possible. County-level national inventories currently take weeks to run and process, including time-saving short cuts such as representing all U.S. counties by 175 representative counties and representing a month by a single day. Similarly, air quality simulations require hourly emissions values for thousands of grid cells over several to tens of days, so current mobile source modeling efforts require similar approximations to be able to complete execution within weeks. Because national inventories and gridded air quality simulations are major undertakings, inventory preparation requiring a few weeks is acceptable. A time requirement longer than roughly a month would not be acceptable. In current practice, compromises and approximations such as previously described are used to assure timely results. While model execution times of weeks are acceptable for these demanding applications, shorter times and less approximation are always desirable. The ability to quickly evaluate alternative control scenarios (say, in an hour to a day) would aid in the development of rule makings and control strategies.

Table 2-1. MOVES Use Cases

Use Cases	Experts Consulted
<p>National inventory development for EPA reports and regulations</p> <ul style="list-style-type: none"> • National Emissions Inventory (NEI) • Inventory of U.S. Greenhouse Gas Sources and Sinks • International Reporting • Regulatory analyses 	<p>Alison Pollack, Environ Corporation Greg Stella, EPA/OAQPS</p>
<p>Local inventory development</p> <ul style="list-style-type: none"> • SIP development • Conformity • NEI input • Trading programs 	<p>Kip Billings, Wasatch Front Regional Council (Salt Lake MPO) Mike Keenan, New York State Department of Environmental Conservation Mohammed Khan, Md. Dept. of Env. Mark Janssen, LADCO Alison Pollack, Environ Corporation</p>
<p>Hot-spot and project level analysis</p> <ul style="list-style-type: none"> • PM and CO Conformity • Toxic exposure analysis 	<p>Harold Nudelman, New York City Department of Environmental Protection Alan Huber, EPA/ORD</p>
<p>Model interaction</p> <ul style="list-style-type: none"> • Air Quality Pre-Processors • Travel Models <ol style="list-style-type: none"> 1. Travel Demand Models 2. Microscopic Models 3. TRANSIMS • Dispersion Models 	<p>Mark Janssen, LADCO Rob Ireson, Consultant Rick Dowling, Consultant Matt Barth, UC Riverside Fred Ducca, US DOT Bruce Spear, US DOT Greg Stella, EPA/OAQPS Alan Huber, EPA/ORD</p>
<p>Policy evaluation</p> <ul style="list-style-type: none"> • Cleaner vehicles and equipment • Cleaner fuels • Less travel and equipment use • Reducing in-use emissions • Modifying vehicle and equipment operation 	<p>Alison Pollack, Environ Corporation David Korotney, EPA/OTAQ Ken Adler, EPA/OTAQ Roger Gorham, EPA/OTAQ</p>
<p>Model Analysis</p> <ul style="list-style-type: none"> • Validation • Uncertainty analysis • Sensitivity 	<p>Alison Pollack, Environ Corporation Mark Janssen, LADCO</p>
<p>Model updates and expansion</p>	

We intend to develop MOVES in an iterative fashion. The earliest releases of the model will focus on the most important use cases, as reflected in our implementation plan discussed previously. In keeping with our iterative paradigm of software development, we do not consider the use case issue closed. We feel that we have obtained a good overview, but we expect refinements, additions, and, perhaps, removals as we proceed with model development.

3. Overview of MOVES Design

The MOVES design is modular, general purpose, data driven, easy-to-use, and high performance. All emission scales and processes are incorporated into a flexible framework of time spans and locations. The MOVES design provides several means of modeling the effects of emission controls. MOVES emission rates and activity information are derived from databases, easily updated as needed.

Understanding MOVES requires learning some new terms or new definitions for existing terms. The term *source* is used to encompass both on-road vehicles and off-road equipment pieces. A *source use type* is a specific class of on-road vehicles or off-road equipment defined by unique activity patterns. *Source bins* are a subset of use types: subcategories that differentiate emission levels within a use type, covering categorizations such as weight class, fuel type, technology, standard, horsepower range, etc. *Total activity* is defined for a given use type as the product of the population of that use type and the per-source activity by time and location. We subdivide total activity into categories that differentiate emissions, known as *operating mode bins*; the intersection of source bins and operating mode bins results in a unique *source and operating mode bin*. By *emission rates*, we mean the most disaggregated rates the model produces internally by source and operating mode bins. By *emission factors*, we mean emission rates aggregated in various ways over source and/or operating mode bin and normalized by some activity basis, such as mass of pollutant per time or per mile.

An *emission process* is a unique emissions pathway. The proposed emission processes for MOVES are shown in Table 3-1. These processes are being defined the same as in MOBILE6 and NONROAD, except that resting loss emissions would now include estimates of offgassing, or hydrocarbons escaping from car materials such as upholstery.

Table 3-1. MOVES Emission Processes

Combustion Products	Hydrocarbon Evaporation	Other
Running Exhaust Start Exhaust Crankcase	Diurnal Hot Soak Resting Loss Running Loss Refueling	Brake Wear Tire Wear

These emission processes are generally associated with different source bins and operating modes and they may also produce different pollutants. Therefore, each emission process is generally handled separately from the others, using common data when appropriate. From a design perspective, this separation creates several submodels within the model. For example, running exhaust emissions are affected by a different set of fleet and activity characteristics than are diurnal evaporative emissions.

Three basic analysis scales are defined for MOVES, following from the NRC report and the April 2001 issue paper. *Macroscale* analyses are appropriate for developing large-scale (e.g., national) inventories, for which the basic spatial unit would be the county. *Mesoscale* analyses are appropriate for generating local inventories at a finer level of spatial and temporal resolution. For on-road vehicles, mesoscale analyses would use as spatial units roadway links and traffic analysis zones or would use vehicle trips consistent with output from standard travel demand models. *Microscale* analyses allow the estimation of emissions for specific corridors and/or intersections, which is appropriate for assessing the impact of transportation control measures and for performing project-level analyses. For off-road equipment, the appropriate spatial and activity units for mesoscale and microscale analyses have not been fully delineated, but would likely be sub-county units of some sort for mesoscale and specific locations for microscale.

3.1. Design Principles

The proposed MOVES design has been developed to address each of the use cases listed in Table 2-1. This section provides an overview of the general design for the MOVES model and the rationale for these design decisions.

We have established two fundamental design principles, which have guided the development of the MOVES design, as follows:

- The model design needs to be flexible enough to accommodate the calculation of emissions from a wide variety of emission sources (e.g., on-road and off-road), over multiple scales, and multiple emission processes using a similar design framework.
- The model must be developed in a modular way to allow ease of update.

These two principles have led to a design that provides a generic framework for addressing the first principle and is implemented in a modular and transparent fashion to address the second principle.

To address the first principle, a generic framework has been developed which provides significant flexibility for applying the model across emission source, scale, and process. The basic concept of this generic framework is the following: for a given time, location, use type and emission process, total emissions can be calculated by the following four steps:

1. Calculate the **Total Activity**, expressed in units of the activity basis (explained in following section) for the given emission process.
2. Distribute the total activity into **Source and Operating Mode Bins**, which are defined as having unique emissions for that emission process.
3. Calculate an **Emission Rate**, which characterizes emissions for a given process, source bin, and operating mode bin and which accounts for additional effects such as fuel and meteorology.
4. **Aggregate** emission rates across these modes using the source bin and operating mode distribution from Step 2.

These steps are characterized mathematically as follows:

$$TotalEmissions_{UseType} = TotalActivity_{UseType} \times \sum_{n=1}^{Number\ of\ Bins} EmissionRate_{UseType, Bin} \times BinDistribution_{UseType, Bin}$$

In this equation, *bin* refers to source and operating mode bin. Each of these elements is elaborated below.

3.1.1. Total Activity

For a given use type, time, and location, total activity is the product of population and per-source activity. How *activity* is defined will depend on the emission process being modeled. For most processes, we propose to characterize total activity by source-time; i.e., source hours operating (SHO) or source hours parked (SHP). Source-time is the product of the population of vehicles/equipment and the analysis time span. Source-time is an attractive way of characterizing activity, because it is common to all emission processes and operating modes. For on-road sources, source-time is more broadly applicable than the more conventional Source Miles Traveled (VMT), since diurnal or idle emissions are produced when a source is not moving. For non-transportation equipment, such as bulldozers and cranes, SHO and SHP can be applied as with on-road sources, while VMT cannot. It is important to note, however, that this use of source-time mainly applies to the calculation of emissions within MOVES and does not preclude the use of VMT to express the activity of on-highway sources, since SHO and VMT are easily inter-convertible if average speed is known. The model design accepts a wide variety of inputs, including VMT, and produces a wide variety of outputs, including emission factors in grams/mile. Section 7 on mathematical formulation provides more detail on this approach.

Some processes have total activity for which non-time based activities are more directly estimated. Specifically, start exhaust will use number of starts per time and location as the total activity basis, and source refueling will use gallons of fuel burned. The total activity basis proposed for each process is shown in Table 3-2.

Table 3-2. Total Activity Basis by Process

Emission Process	Total Activity Basis
Running Exhaust, Brake Wear, Tire Wear, Running Loss, Crankcase	Source Hours Operating (SHO)
Start Exhaust	Number of Starts
Diurnal, Hot Soak	Source Hours Parked (SHP)
Resting Loss	Source Hours (SH)
Refueling	Gallons of Fuel Used

3.1.2. Source and Operating Mode Bins

MOVES differs from the previous MOBILE series of models in that it is explicitly designed to use the same set of emission rates at a very large scale (such as national-level inventories) and at a very small scale (such as intersection modeling). It also combines into a single modeling framework the ability to generate (1) emissions for a wide range of pollutants and processes and (2) emissions from both on-road and off-road vehicles and equipment.

To handle this diversity of use, and to enable ease of updates, we propose that MOVES store emission rates as average emissions within discrete source and operating mode bins. Each discrete bin will be dictated by factors that are found to result in unique emissions, based on subcharacteristics of the use type and operating modes. Parameters which define source bins could be based on fuel type, accumulated use (which accounts for deterioration), source weight (on-road), horsepower range (off-road), technology types, and standard levels. The concept of source bin is similar to methods used to distinguish emission rates in MOBILE or NONROAD. One key difference to note is that accounting for vehicle mileage using the bin concept is a departure from the traditional use of a continuous deterioration rate correction factor. This new approach to addressing vehicle mileage has been employed in UC Riverside’s Comprehensive Modal Emission Model.⁴

Examples of source and operating mode bins are shown in Table 3-3. A broader range of the on-road use types and source bins are presented in Section 7.

Table 3-3. Example Source and Operating Mode Bin

Use Type	Type of Bin	Bin Parameters	Example
Passenger Car	Source Bin	Fuel Type Mileage Technology Standard Emitter Category	Gasoline High Fuel Injection/3-Way Catalyst Tier 1 Normal
	Operating Mode Bin	Power Bin	Vehicle Specific Power =13 to 16 kW/ton

To facilitate the calculation of model uncertainty, we propose that each source and operating mode bin would contain a mean value and either the distribution of uncertainty around the mean or the variability of the entire sample used to generate the mean. This approach will allow the calculation of uncertainty in model prediction using either propagation of error or Monte-Carlo analysis techniques. This will be discussed in more detail in a separate report detailing the proposed emission analysis method for MOVES.

3.2. Core Model and Implementations

Two basic components of the MOVES design have evolved from the generic framework: the core model and implementations. The core model would perform the four-step calculation outlined in the previous section for the times and locations being modeled, resulting in outputs of emission factors and total emissions. At the heart of the core model is the emission calculation function, a large part of this function being a direct database lookup of emission rates by source and operating mode bin.

We refer to the “implementations” as the body of applications, controls, and utilities that would use the core model to address many of the use cases presented in Table 2-1. For example, macroscale inventory generation would operate the core model in conjunction with a database of fleet and activity information and front-end applications, which produce necessary core model inputs from these data. Uncertainty analysis using Monte Carlo simulation would require a module that would iterate the core model many times, while populating the input parameters with random samples from the distributions of their means and, then, analyze the combined outputs to produce uncertainty estimates.

Fundamental to the implementations is the concept of “data generators,” which would convert data from a multitude of sources to the standard input form acceptable to the core model. For each emission process, the core model would be developed to use a standard input form for total activity and operating mode distributions regardless of analysis scale or the ultimate source of that information. It is envisioned that we would develop data generators for MOVES to support essential use cases, while parties other than EPA, using standard interfaces in MOVES, would develop other generators.

The rest of this document provides more detail about the core model and implementations that make up the MOVES design. Sections 4, 5, and 6 offer specific perspectives of the design: software design layers, data flow design, and database design. Section 7 looks at specific mathematical formulations and data inputs for the proposed MOVES implementations. Section 8 offers information about the specific software tools and techniques involved in creating and operating MOVES.

4. Software Design Layers

The MOVES Software Design Layers are shown in Table 4-1 and provide a view of MOVES software structure. The layered structure is described in more detail in the subsequent text.

Table 4-1: MOVES Software Design Layers

Layer	Components		
Control 1	System Specifications		
Control 2	Run Specification		
Control 3	Monte Carlo Controller		
Control 4	Master Loop		
Application 1	<u>Input</u>	<u>Core</u>	<u>Output</u>
	Data Manager Data Importers Control Strategies Data Generators	Core Model	Output Aggregator OLAP Reporter Exporters
Database 1	Input Databases Execution Location Database Emission Rate Database	Execution Emission Rate Database	Output Database
Utilities	Examples: Visualization Tools, DBMS Tools, Data Browser, API, MIMS utilities		
Database 2	Archived runs: runs specs and output databases for later comparisons		
Application 2	Data Crank		
Database 3	Supporting Data for Data Crank		

4.1. Control 1: System Specifications

The system specifications describe the operating environment for one or more MOVES runs. These include a specification of whether or not distributed processing (distributing the model execution over multiple computers, discussed in Section 8.3) is being used and, if so, identification of computers involved in distributed processing and directories for shared files.

4.2. Control 2: Run Specification

The Run Specification, or Run Spec, is the next highest control and specification layer in MOVES. It is both a set of user inputs that define a model run and a piece of code that produces the inputs and initiates lower levels of control. It is usually produced and modified with a graphical user interface (GUI), as explained in Section 5.3. The User Interface.

The Run Spec defines the situation that MOVES will model, including pollutants, emission processes, location and time period, source use types, and emission control strategies. It also allows the user to define the scale of the analysis and to identify alternate sources for input data. It defines the level of detail to be included in the MOVES output and allows the user to specify if a Monte Carlo uncertainty analysis is desired. The Run Spec initiates the Monte Carlo Controller and Master Loop (Control Layers 3 and 4).

The Run Spec data will be stored as a binary file. The input portion of the Run Spec may be edited with a graphical user interface, as mentioned above, or it may be created or edited in extensible markup language (XML) by an external program. The GUI will make the model relatively easy to use, even for beginning modelers, while the XML option will allow sophisticated users to batch process multiple runs and to link MOVES into a broader system of transportation and air quality models.

4.3. Control 3: Monte Carlo Controller

An important element of MOVES will be the quantification, where possible, of model uncertainty. As described in Section 3.1.2, we propose to characterize emission rates as distributions. One way to implement the calculation of uncertainty would be through the use of Monte Carlo analysis. This would require the development of a utility in MOVES that would apply Monte Carlo analysis to generate emission-based uncertainty estimates of model results. The Run Spec would allow the user to choose a Monte Carlo analysis and to specify the number of iterations and simulations per iteration. The Monte Carlo Controller would then run the model through the required numbers of repeat emission estimates, selecting appropriate emission rates from the emission rate distribution.

In the future, this approach could be extended to estimate other sources of uncertainty in the model results. Probability distributions for activity estimates, meteorology, and other parameters may be added later if information is available. The importance of these other factors, especially activity, in emissions estimates argues for their eventual inclusion in model output uncertainty calculations.

One disadvantage of using a Monte Carlo approach to calculate uncertainty is that we expect Monte Carlo runs to take a significant amount of processing time. Since we expect these runs to be performed primarily for diagnostic and assessment purposes, they will be relatively infrequent and generally performed only by users with relatively sophisticated computing resources. Consequently, the processing time will be acceptable.

Other approaches to estimating uncertainty include non-numerical methods involving properties of the mean and variance, transformation of variables, and Taylor series expansions.⁵ These methods have the advantage of not requiring large numbers of iterations. They have the disadvantage of being only approximate, especially in cases where the uncertainty or is large, the distributions are unusual, or the supporting data are limited.

If a non-numerical method of error propagation were employed, the Monte Carlo controller would not be required in MOVES, and the calculation of uncertainty would take significantly less time. The feasibility of each approach is being considered as part of the emission analysis evaluation.

4.4. Control 4: Master Loop

MOVES users will often desire results for multiple times, locations, and emission processes. And, even when consolidated reporting is desired, more precise results can generally be obtained by calculating results from smaller areas, time periods, and sub-processes and, then, summing them. Thus, a basic function of MOVES will be looping over all the times, locations, and processes defined in the Run Spec. The Master Loop code will control this looping in the underlying application layers of the model.

4.5. Application 1 and Database 1

Two basic components of the MOVES design that have evolved from the generic framework are the core model and the implementations, as explained in Section 3.3. The core model is to be as generic as possible. It will accept inputs of (a) total activity, (b) source bin distributions, (c) operating mode distributions, (d) meteorological information, and (e) fuel information. It will look up emission rates from a database. It will output emission factors and total emissions.

The implementations will incorporate the core model, as well as pre- and post-processors that handle inputs and outputs. This enhanced system will allow the generic core model to perform various specified use cases.

The following describes in more detail the input, core, and output functions, as well as the data on which each function acts.

4.5.1. Input

MOVES users will come to the model with a wide range of inputs, ranging from none (defaults only) to detailed information on meteorology, fuel composition, source activity, source populations, etc. In general, both default data and user data will be stored in *Input Databases* in specified database formats. To maximize flexibility in the use of MOVES, a major task for the MOVES model is selecting and /or generating the appropriate core model inputs from the user inputs and model defaults.

The *Data Manager* has primary responsibility for sorting through the available user inputs and default values to choose the appropriate values for the time/location/process indicated by the Run Spec and Master Loop. These are fed into the *Execution Location Database*. The Data Manager may also replace MOVES default emission rates from the *Emission Rate Database* with user inputs. These are stored in the *Execution Emission Rate Database*.

The Data Manager works only with data in very specific MOVES input formats. In many cases it will be the user's responsibility to develop inputs in the correct format. However, to accommodate specific use cases, we plan to develop a set of *Importers* that will translate data from other well-defined formats into acceptable MOVES inputs. Other parties could develop additional importers as desired.

Many MOVES use cases involve modeling the world with and without various emissions control strategies. In some cases, such as an entirely new idea for emissions control, these control strategies and their effects will be defined entirely by the user in the user inputs. We refer to these as *External Control Strategies*. However, some control strategies are quite common and—to meet the need for ease of use—MOVES will have some *Internal Control Strategies*. That is, MOVES will include processors that use user data on control strategy parameters to modify data in the Execution Location Database (ELD) and the Execution Emission Rate Database (EERD) to account for these programs. When complete, MOVES will include four basic types of internal control strategies: fuel programs, emission standards, inspection/maintenance programs, and controls of fleet population and activity.

Generally, the inputs commonly available will not exactly match that needed by the core model. In some cases, this data will need to be filtered to select only the information needed by the core model for a specific run. In other cases, the data will need to be grown in time, allocated in space or time, or otherwise adjusted to provide the level of detail needed by the core model. These processes will take place in *Generators* specific to the kind of data being processed. Separate generators will exist for meteorological data, fuel data, operating modes, technology and standards and total activity.

4.5.2. Core Model

The core model is designed to provide estimates of total emissions, as well as emission factors. The core model will look up appropriate emission rates for each source and operating mode bin and apply corrections to account for conditions not in the database: for example, adjustments to account for local fuel and meteorological conditions.

The core model will use structured query language (SQL) to define the relevant source and operating mode combinations (bins) to apply the emission rates and appropriate correction factors. With distributed processing, the SQL look-ups and calculations may be performed on remote computers.

4.5.3. Output

The *Output Aggregator* sums emissions across bins into larger groupings that can be stored in an *Output Database*. The Output Database is in a form to be manipulated by the relational database software in the model into the specific data format requested in the Run Spec. The output table(s) may then be further processed by the user with an Online Analytical Processing (OLAP) tool.

For specific implementations of the model, we also will develop *Exporters* that process the MOVES output into the formats needed for common post-processors such as the EPA's SMOKE model.

4.6. Utilities

We expect that much of the processing required at the Application 1 layer will make use of utility software including visualization tools, Database Management Software (DBMS) tools, data browsers, and utilities from EPA's Multimedia Integrated Modeling System (MIMS).

4.7. Database 2

Many of the MOVES use cases require comparisons between runs of the model under different conditions. To facilitate this, MOVES will create a database of archived runs, linking the inputs defined in the Runs Specs and the output stored in the output database.

4.8. Application 2 and Database 3

An important use case for MOVES is frequent and easy updating. As data becomes available and as resources permit, we will need to update default data. This function will be outside of MOVES, but is important to consider in the design of the system. We plan to develop a standard algorithm for updating defaults to reflect new data, and to automate them as much as possible. We call this set of standard algorithms the *Data Crank*. One application of the Data Crank under consideration is the use of a physically-based emission model, which would be updated with raw emission data and used to populate the bins in the Emission Rate Database. Systematic storage of underlying data, through the Mobile Source Observation Database (MSOD) and other tools, will also be an important part of the MOVES system.

5. Data Flow Design

The MOVES design is modular, general purpose, data driven, easy-to-use, and high performance. The core model design applies to all emission scales and processes and incorporates a flexible framework of time spans and locations. It is also designed to facilitate rapid execution, given the many combinations it must process. This core model (which may be used directly by providing Core Model Input Data to the Input Data Manager) is preceded by several processing stages which make it progressively more easy to use. These components provide a graphical interface and means of accepting a variety of input data in the forms actually available to users. The MOVES design provides several means of modeling the effects of emission controls. MOVES emission rates and activity information are derived from databases during model execution. These databases may be updated as new information becomes available to users and to developers of the model.

This section presents the MOVES software design in terms of its data flow. It extends the discussion of the software design layers in Section 4 by going into greater detail regarding MOVES data and processing elements and the order in which these elements are executed. We break this discussion into three sections: Core Model, Implementations, and User Interface. The first two were defined in Section 3. The data and process flow diagrams discussed in this section focus on the Application 1 and Database 1 software layers presented in Section 4, while also including Control Layer elements, such as the Run Specification. The design presented here applies to all modeling scales and all emission processes.

Data flow diagram sections are explained, beginning with one for the core model, which produces emission results in a generic format. Subsequent diagram sections show the earlier processing stages usually necessary to prepare data for the core model.

Some relatively standard conventions used in these diagrams are:

- Parallelogram symbols represent data in a directly accessible form; e.g., in memory.
- Cylinders represent databases.
- Rectangles open on the right side represent relatively simple data storage elements.
- Round-cornered rectangles represent processes that operate on data, generally producing further data.
- Square boxes represent external interfaces to the system.
- Solid line arrows represent data flow.
- Dashed line arrows represent control flow (if separate from data flow).

Some special conventions used in these diagrams are:

- Diagram elements shaded gray are produced by previous processing stages.

- Diagram elements with a gray shadow may have different versions for different emission processes (tailpipe running exhaust, start, HC diurnal evaporation, etc.).
- The “RTC” notation on a data flow arrow means that the process producing the data flow “runs to completion” before any of the data is processed by the subsequent step.

5.1. The Core Model and Output Processing

A data flow diagram for the Core MOVES Model is shown in Figure 5-1. It is the most generic part of MOVES. Because much of the core model will be implemented with a Relational Database Management System (RDBMS) in a fashion which supports distributed processing, this diagram is more a conceptual representation than a literal one.

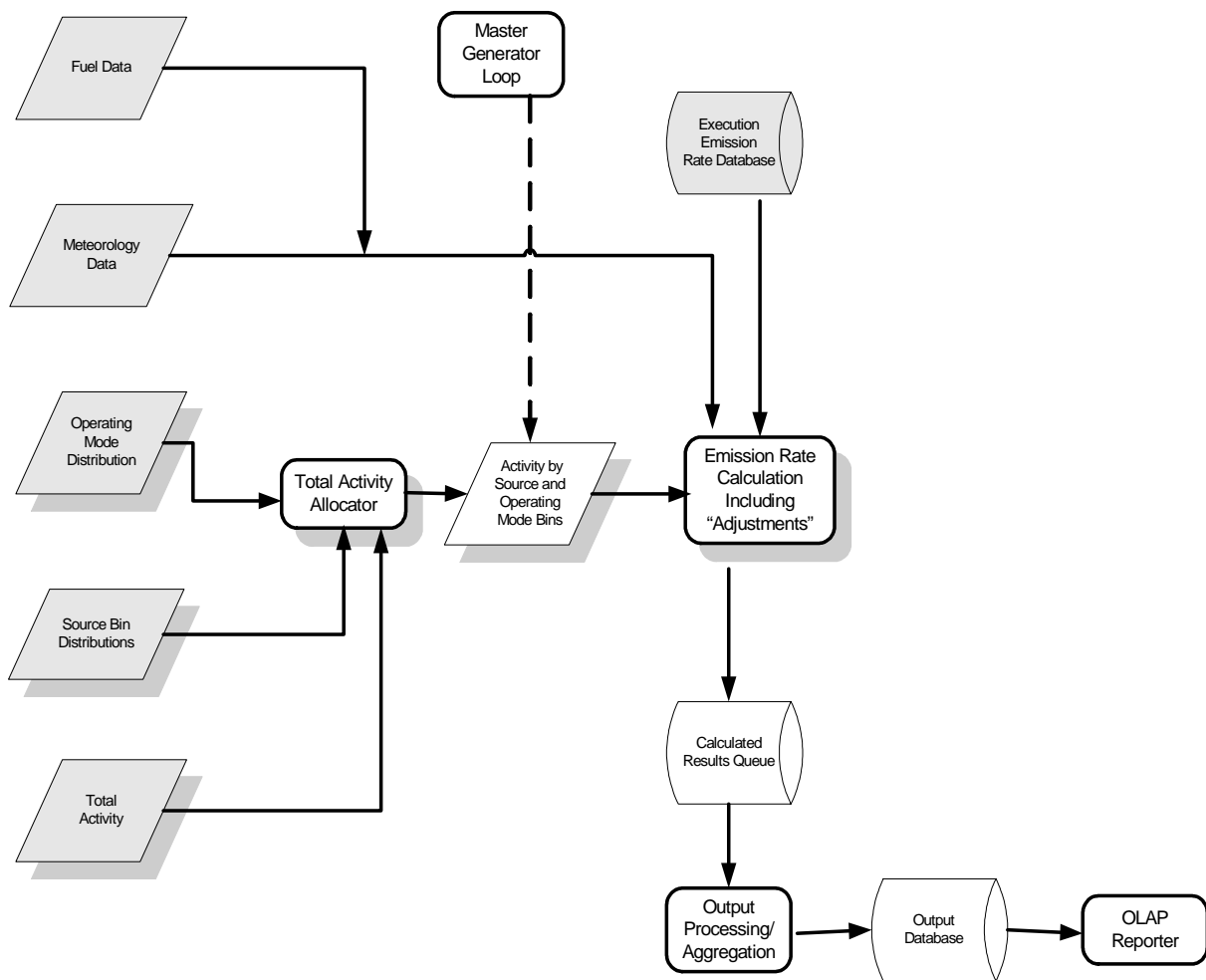


Figure 5-1. Data Flow Design: MOVES Core Model

5.1.1. Core Model Input Data

Input data for the core model is represented by the five parallelograms on the left hand side of Figure 5-1. Because of the volume of information involved, at any one time during program execution, data being processed may include only a portion of the times, emission processes, and locations being modeled, as determined in a flexible manner by the Master Generator Loop (described later in this section). This *Core Model Input Data* includes:

- Fuel parameter information for the available supply of fuel by time, location, and kind of fuel (e.g., gasoline, diesel, natural gas, etc.).
- Meteorology information (temperature, humidity, solar load, etc.) by time and location.
- Total Activity by source use type, time, location, and (for some processes) roadway type. The activity basis can depend on the emission process, as presented in Table 3-2.
- Operating mode distributions by source use type, time, location, and (for some processes) roadway type. These distributions subdivide total activity by operating mode. The activity basis and set of modes used may depend upon the emission process; e.g., modes for exhaust running emissions are likely to be vehicle specific power bins, while for hot soak emissions they are likely to be soak time bins.
- Source bins distributions for each source use type. These allow the core model to separate the Total Activity of each use type into the source bins necessary to estimate emissions. These distributions vary by model year and, thus, the age of the source, but are assumed not to vary by location and not to vary otherwise with time.

One way to use MOVES is to simply input into the model this Core Model Input Data, along with corresponding emission rate information (which corresponds to default emission rate information built into MOVES). This method would provide ultimate flexibility to the user for modeling any case for which the required input data was available, from parking lots to entire nations. However, we envision that most users will not use MOVES in this way, because the data requirements are so high; hence, the concept of “implementations” has been developed based on input of more commonly available data, as discussed in Section 7.

Our goal is to keep the core model as general purpose as possible. Core model procedures must know which activity basis and set of operating modes apply to their emission process, but they do not need to know what the different kinds of sources, kinds of fuel, technology types, locations, operating modes, etc., represent. An operating mode might be a source power bin or soak time prior to start. Some processes; e.g., crankcase emissions, may not need operating modes at all.

The core model, beginning with the Total Activity Allocator, is designed to be implemented primarily in relational database management software in a potentially distributed fashion. This means that the Total Activity Allocator, Source and Operating Mode Bins, and Emission Rate Calculator are logical entities whose physical implementation may assume a

different form. The Source and Operating Mode Bins, for example, may never actually be physically produced, but the results will be the same as if they were.

5.1.2. Total Activity Allocator

The *Total Activity Allocator* applies the Source Bin Distributions and the Operating Mode Distributions (which do not interact with each other) to the Total Activity Information to produce very detailed “Activity by Source and Operating Mode Bins,” within a given source use type.

5.1.3. Emission Rate Calculator

The *Emission Rate Calculator* is a logically separate process, which removes Source and Operating Mode Bins from their pool as it determines emission results for them. The Emission Rate Calculator has access to a large database of emission rates, information about the meteorology and fuel supply applicable to the times and locations being modeled, and relevant correction factors. Different versions of the calculator may be used for each emission process.

The Emission Rate Calculator will first look up a base emission rate from a large database of such factors. Assuming uncertainty analysis is not being done, the mean of the emission rate distribution will be used. If uncertainty analysis is being done, then a pseudo-random sample from the emission rate distribution will be generated (Monte Carlo approach) or error propagation equations applied to the uncertainty on the mean (error propagation approach).

Adjustments specific to each emission process may then be applied to these basic rates, also in a probabilistic fashion if uncertainty estimates are being generated. Calculation of running exhaust emissions, for example, are likely to include adjustments for temperature, air conditioning use (which in turn will depend on weather conditions), and fuel properties.

5.1.4. Execution Emission Rate Database

The *Execution Emission Rate Database* contains base emission rates for different pollutants, as produced by different emission processes, by source and operating mode bin. It also contains correction factors (or coefficients necessary to calculate correction factors) for factors requiring these corrections, such as fuel or meteorology. The version of this database used in the core model has been adjusted to reflect User Input Data and the effects of any Internal Control Strategies (which will be described later). It has also been reduced in size, when possible, to contain only the rates needed to fulfill the needs of a particular modeling run.

5.1.5. Calculated Results Queue

Result output is produced by the core model in the form of a single generic table format that contains emission rate and mass information distinguished by pollutant, emission process, location (a link being the most detailed location), time (an hour normally being the smallest unit of time, though 15 minute periods are allowed at microscale), roadway type (for highway

vehicles), kind of highway vehicle/ nonroad equipment, power class (of nonroad equipment), and model year.

5.1.6. Output Processing/Aggregation

Output produced by the core model must generally be aggregated: first, because portions of the output may have been produced piecemeal by different processes and computers and, second, because the user may not desire the full extent of disaggregation produced by the Emission Rate Calculator. MOVES allows the user to aggregate link and zone level results to counties, states, or the entire modeling domain. It allows results to be aggregated in time up to the hour, day, month, season, or year. It also allows distinctions by emission process, kinds of vehicles/equipment, model year, roadway type (of highway vehicle activity) and power class (of nonroad equipment) to each be included or removed from the results. As more of these distinctions are removed, the more manageably sized the result data will be.

5.1.7. On-Line Analytical Processing (OLAP) Reporting Utility

MOVES will include an OLAP reporting utility program, which the user may employ after the core process has been completed, to produce a variety of reports. Such an utility allows the user to specify a hierarchy of fields that define a set of categories, subcategories, sub-sub categories, etc. whose sums are calculated and reported. Specific reporting utilities will likely be developed to produce output in the format necessary for input into air quality pre-processing models. These format-specific utilities are also referred to as “Exporters.”

5.2. The Master Generator Loop, Internal Emission Control Strategies, and Data Generators

The MOVES core model described previously is incomplete relative to many use cases. For example:

- It implies that the model run can be executed in a piecemeal fashion, but contains no mechanism for doing this.
- The generic input data it requires is not in a form that would be easy for many users, except in a microscale situation, to provide.
- It contains no internal provisions for modeling the effect of emission control programs or strategies (e.g., source inspection/maintenance), unless the user can determine outside the core model how the strategies he or she wishes to model would affect the core model input data.

Because of this, “implementations” (discussed in Section 3.2) will be developed to drive the core model for specific use cases. Implementations will take the form of front-end components, which perform preliminary processing stages and higher layer control components, performed immediately prior to execution of the core model. The front-end components are the Master Generator Loop (which controls looping of the core model); Data Generators (which provide data needed by the core model based on default or user-supplied input data); and Internal

Emission Control Strategies (which enable the user to select and modify emission control strategies the user wishes to model). These steps produce the Core Model Input Data and Execution Emission Rate Database required by the core model. Figure 5-2 shows the design components involved in this stage.

5.2.1. Master Generator Loop

The *Master Generator Loop* is the mechanism that manages execution of the core model in a piecemeal fashion to accomplish the overall model run specified by the user. It controls execution of all the processes executed in this processing stage (i.e., the Control Strategies and Data Generators), as well as the core model itself. The MOVES design provides for this to be done in a very flexible fashion. All these processes will be instances of classes (i.e., objects) which implement an interface enabling them to “sign up” with the Master Generator Loop to be executed at any of a number of Master Loop Levels. At the topmost level, a process, such as a fuel control strategy which changes fuel properties the same way across the entire modeling domain, can sign up to be executed once for the entire run. Alternatively, a process can sign up to be executed once per location, once per emission process per location, or once per time period for each emission process per location. At the lowest Master Generator Looping Level, a process can sign up to be executed for each hour for each emission process at each detailed location. Within each level, control strategy processes will be executed before generators, with the core model running last. (The core model will probably not be allowed to sign up at a level higher than any of the other processes do.)

A process can sign up at a high level if the amount of data it produces is relatively small. In this case, it can take advantage of efficiencies that result from processing a large part of the run together. If, however, the amount of data produced prohibits this, a process should sign up at a lower level. This allows for processing efficiency and data storage tradeoffs to be made in a flexible fashion.

5.2.2. Execution Location Database

The *Execution Location Database* provides, along with the Execution Emission Rate Database described earlier, the input data for the processing steps shown in Figure 5-2. These execution databases are created by still earlier processing stages (to be described later) and exist only for the duration of a MOVES model run. The Execution Location Database is capable of storing the Core Model Input Data. It is also capable, however, of storing data more readily available to model users for use by the Data Generators to derive core model data. For example, unweathered fuel Reid Vapor Pressure (RVP) might be stored in the execution location, whereas the core model requires weathered RVP information. Another notable example is that the Execution Location Database will contain vehicle activity information in terms of vehicle miles

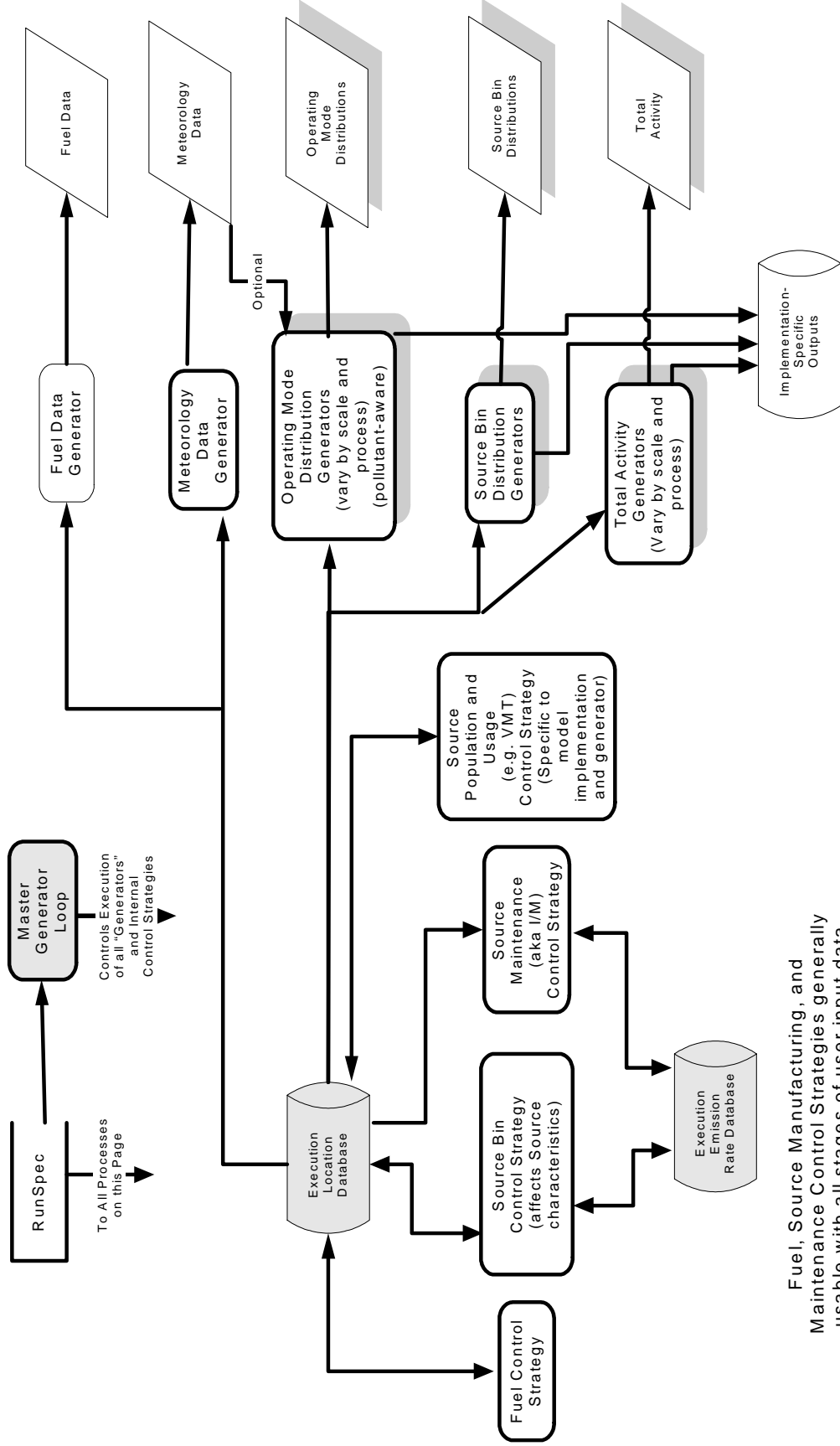


Figure 5-2. Data Flow Diagram: MOVES Master Generator Loop, Internal Emission Control Strategies, and Data Generators

traveled (VMT), which is then converted to source hours operating (SHO) by a Data Generator. All data inputs detailed in Section 7 would be stored in the Execution Location Database.

5.2.3. Data Generators

Data Generators derive Core Model Input Data from information in the Execution Location Database. They implement the mathematical formulations, described in Section 7, as required for different versions of MOVES. Each of the five types of Core Model Input Data has a distinct generator. As indicated in the diagram, some types of generators may have different versions for different emission processes, and different implementations or releases of MOVES may include different Data Generators. (Object inheritance will be very useful in this area of the design.) Each generator is responsible for signing up with the Master Loop at an appropriate level and, when executed, for supplying the Core Model Input Data for the corresponding portion of the run. In the simplest case, if Core Model Input Data is already available in the Execution Location Database, all a Generator needs to do is to select out of the database the data needed for this portion of the run. In the more usual case, however, the Generators must process data from a portion of the Execution Location Database to produce the Core Model Input Data. This process is described in Section 7, Mathematical Formulations and Data Inputs.

5.2.4. Internal Emission Control Strategies

Prior to executing the Data Generators for each portion of the model run, another set of processes can be executed to take into account the effects of various *Emission Control Strategies*, whose detailed characteristics are built into the model. Different implementations of MOVES will contain a different and, generally, expanding set of these internal control strategies. These will be needed in MOVES both to represent programs actually in existence; e.g., an I/M program being operated in a particular state, and to represent hypothetical measures being evaluated. In the first instance, the default situation would be to include the control strategy, unless the user removed it. In the second instance, the default would be not to include the control strategy, unless the user added it.

All internal Emission Control Strategies have a scope which specifies a set of use types, calendar years, and locations to which it applies. The MOVES design envisions that there will be four kinds of internal Emission Control Strategies:

- **Fuel Control Strategies:** This kind of strategy modifies the parameters of the fuel available at certain times and locations in some fashion, within a given fuel type. For example, this strategy would be used to specify that the gasoline in a given location was RFG rather than conventional gasoline. The effects of such a strategy are taken into account as modifications to the Fuel Data Input to the core model via the Fuel Data Generator.
- **Source Bin Control Strategies:** This kind of strategy results in actions that modify the source bin distribution, which focus on specific vehicle and equipment attributes.

The effects of such a strategy are taken into account as modifications to the Source Bin Distribution Input to the core model via the Source Bin Distribution Generator. Examples of this would be the creation of a new emission standard, which would also require supplying emission rates for the new standard in the Execution Emission Rate Database. Other examples of these control strategies include increased penetration of advanced fuel technology or shifts in on-road vehicle weight to improve fuel economy. If emitter category becomes a source bin parameter, it is possible that I/M could be included in this family of control strategies, since it would fundamentally redistribute vehicles across the emitter bins.

- **Source Maintenance (e.g., Vehicle Inspection/ Maintenance) Control Strategies:** Modeling of I/M in MOVES will depend on the method of characterizing high emitters in the model, to be discussed in more detail in a separate document. The two primary options for this are (a) characterizing the spread of emitters in the fleet as a single distribution in the Emission Rate Database or (b) defining discrete source bins to represent emitter classes such as “high” and “normal.” If the first approach were used, modeling of I/M would modify emission rate distributions in the Execution Emission Rate Database to reflect the effects of Control Strategies that modify the frequency or the manner that vehicles and equipment are repaired. If modeled as discrete bin categories, I/M would be modeled as an extension of the Source Bin Control Strategies.
- **Source Population and Usage Control Strategies:** This kind of strategy modifies either the number of vehicles and equipment in existence and/or the amount and nature of operation that these vehicles and equipment experience. This group of control strategies is defined as such because it would modify inputs to the Total Activity Generator, which (as discussed in Section 7) works with population by age and activity of this population, and/or the Operation Mode Distribution Generator. A wide variety of strategies fall into this category; e.g., a vehicle scrappage program, a program which encourages carpooling, a program which optimizes traffic signal timing so that vehicles accelerate less frequently, etc. The effects of this type of control program will be taken into account by modifying the Operating Mode Distribution and Total Activity Information in the Execution Location Database, which is used by Operating Mode Distribution Generators and Total Activity Generators. Because this kind of control strategy will modify preliminary data used to calculate the Core Model Input Data, not just information which is directly passed to the core model, this kind of control strategy will not be available to the user who wishes to supply Core Model Input Data directly.

5.3. The User Interface

The MOVES model described so far, even with its Control Strategies and Data Generators is still incomplete. In particular:

- It contains no mechanism for the user to indicate which results (e.g., time periods, pollutants, etc.) are of interest and should be modeled.

- The Execution Location Database it requires, though easier for a user to provide than the Core Model Input Data, is still not in a form that would be easy for users to provide in many cases.
- One aspect of this is that no mechanism has been provided for users to combine their input data with default information in the MOVES model.

These capabilities are addressed by preliminary processing stages and higher layer control components described in this final section of the MOVES data flow and shown in Figure 5-3. These stages supply the Execution Databases required by the rest of the MOVES model.

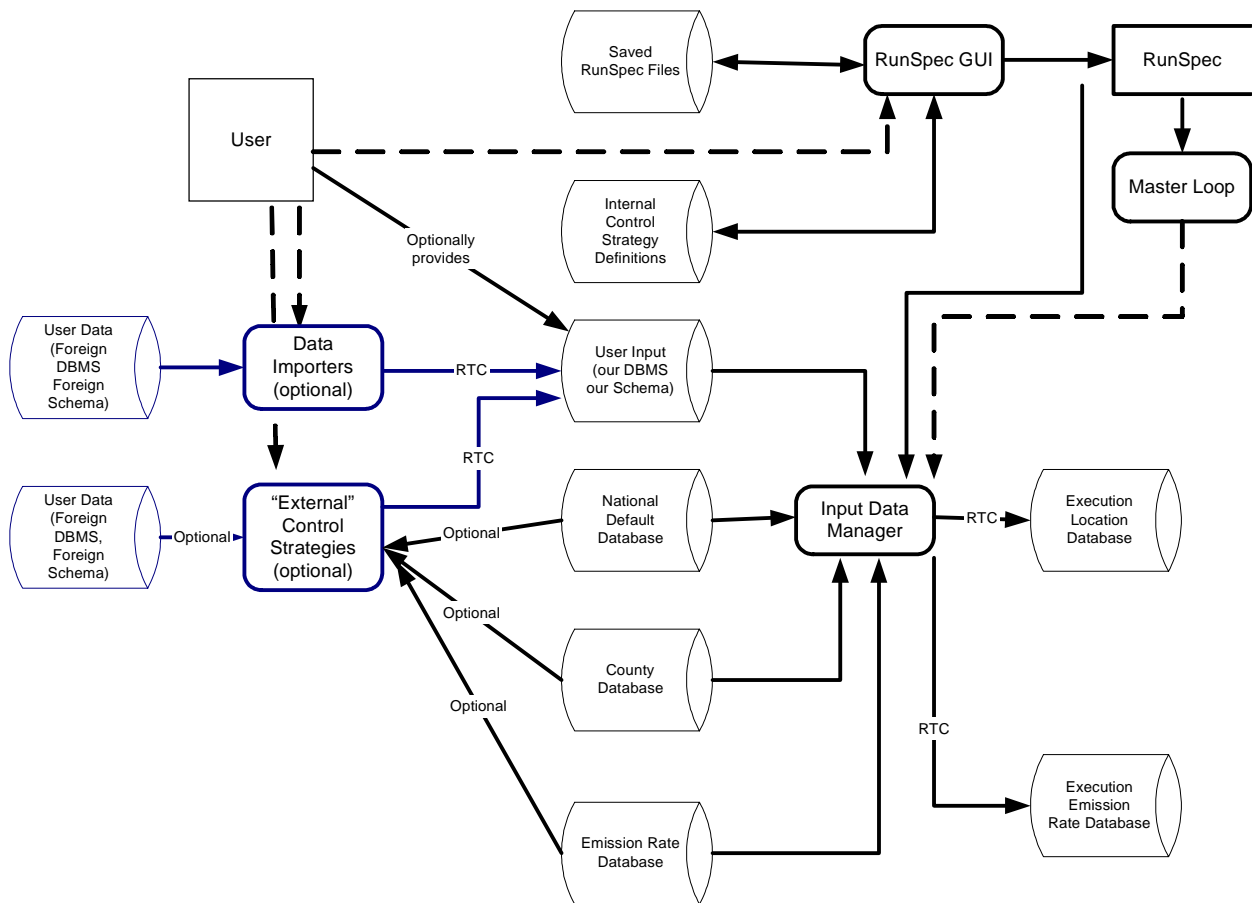


Figure 5-3. Data Flow Diagram: MOVES User Interface

5.3.1. The Run Specification and Graphical User Interface (GUI)

MOVES will have a graphical user interface (GUI) by which the user specifies various parameters for the model run desired. These parameters include the locations and time periods to model, the vehicle classes and nonroad equipment types of interest, the pollutants, the control strategies to include, output format options, etc. This set of information is collectively called a

Run Specification or Run Spec. This GUI is described in detail in Appendix A, so the description here will be brief and concentrate on how it participates in the MOVES data flow.

The MOVES Run Spec GUI will allow the user to specify a model beginning with a default Run Spec or allow the user to modify Run Spec object files saved previously. (It also allows the user to save created Run Specs for future use.) It has available a set of Internal Control Strategies and knows how to locate the MOVES User Input Data, to be discussed later in this section. It produces or retrieves a complete Run Spec before allowing the user to transfer control to the Master Generator Loop described previously, which runs the Input Data Manager described later in this section.

5.3.2. Default Databases

The MOVES model will contain databases of default information pertaining to the modeling domain as a whole (the National Default Database), particular counties of the U.S. (the County Database), and emission rates (the Emission Rate Database). These default databases are read by the Input Data Manager in the process of producing the Execution Location Database, but they are never modified by running the model. These three logical databases are all stored as a single relational database schema, which contains all of the information needed to model the counties of the United States at the macroscale with default information.

5.3.3. User Input Database

Users of MOVES may supply their version of any of the information in the default databases. Indeed they are required to do this to some extent if they wish to model locations other than the states and counties of the United States or if they wish to run the model at the mesoscale or microscale.

MOBILE and NONROAD both allow the user to provide input through external files which conform to a specific format. The User Input Database of MOVES continues this approach. MOVES will accept input data as relational database tables of a specific database software and schema. The User Input Database is stored in a single relational database, which is an extension of the schema of the default databases. Tables that can be used to submit alternative information for everything in the default database schema are included in the User Input Database. Additional tables are included in the User Input Database for the user to supply Core Model Input Data directly. The User Input Database, therefore, has a data structure similar to the Execution Location Database and the Execution Emission Rate Database.

5.3.4. Input Data Manager

The Input Data Manager is run once to completion near the beginning of every MOVES model run. Its job is to produce Execution Location and Execution Emission Rate Databases for the model run which:

- Contain only the subset of information needed to satisfy the Run Spec

- Override Default Database information with User Input Data to the extent it has been supplied.

5.3.5. Data Importers

The User Input Data, while generally easier to provide than the core model input data, still must be provided in a specific database and schema native to MOVES (e.g., macroscale highway running activity information can be in terms of vehicle miles traveled and average speeds rather than vehicle hours operating in vehicle specific power bins). Specific Data Importers may be written, however, by EPA and others, to convert input data from other forms to the database and schema required for MOVES input. For example, EPA will write an Importer to accommodate the data input to MOBILE6. Other importers could accept data from specific travel demand models. Importers normally will accept data in XML or other ASCII text formats. Importers will be run by the user to produce MOVES User Input Data prior to running MOVES itself.

5.3.6. External Control Strategies

External Control Strategies are a special kind of Importer which can be used to model the effects of control measures by supplying MOVES User Input Data that captures their effects. They provide an additional means to model the effects of various emission control measures that are not built into MOVES via the internal control strategies discussed in Section 5.2.4. Processes which implement External Emission Control Strategies are similar to Data Importers except that they may read (but not directly alter) MOVES Default Database information. Instead they produce MOVES User Input data, which the Input Data Manager uses to override MOVES Default Database information. External Control Strategies are not planned for the initial implementations of MOVES, though they could be added later as needed by EPA and others.

6. MOVES Databases

To calculate mobile source emissions at all three scales, MOVES will need to store and process huge amounts of information. Unlike the MOBILE model, which stores data externally in flat text files and internally in Fortran block data modules, MOVES will store data in relational databases. This will allow MOVES to store data efficiently and to take advantage of standard Structured Query Language (SQL) to access and process the data.

Previous sections of this report have described the role of the databases in the context of the software design layers and the data flow. This section focuses on the information that the databases contain. Some of the databases are temporary, existing only for the duration of a MOVES model run. Others are permanent sets of default values that can be used for all runs of a given MOVES implementation.

MOVES uses three basic categories of databases, detailed in this section. Input databases include default databases supplied by MOVES and user-supplied databases. Execution databases are temporary databases created by MOVES during execution. The output database contains MOVES output aggregated per user instructions.

6.1. Input Databases

Input databases include all information needed to calculate emission inventories and extensive default information. This information comprises the following:

- **Total Activity Information:** How much activity (e.g., source hours operating, source hours parked, etc.) occurs for sources of different types (i.e. types which matter for activity) and ages. The MOVES design considers source age to be an activity discriminator and allows distributions of such source activity discriminating characteristics to vary by time and location. Each emission process may have its own activity basis and information.
- **Operating Mode Distributions:** How much of this activity occurs in each of various operating modes. Each emission process is allowed to have its own operating modes which may be as simple as “operating” vs. “not operating” or more complex such as source-specific-power (VSP) bins. The MOVES design allows these operating mode distributions to vary by time and location and by type of roadway.
- **Source Bin Distributions:** How is source activity distributed among sources having different characteristics; i.e., characteristics that matter for emissions but not for activity? Such characteristics might include the emission standard to which the source was built, its type of fuel injection, catalyst type, etc. MOVES may also treat fuel type and source weight or engine displacement (within some classification constraints), as being characteristics of this type. The MOVES design allows such distributions to vary by type of source and model year, but not by time or location. As always each emission process may have its own activity basis and information.

- **Meteorology Data:** Information such as temperature, humidity, etc. that influence emission rates by time and location. MOVES does not plan to include effects of meteorology on activity.
- **Fuel Data:** Information about the fuel parameters of different kinds of fuel (gasoline, diesel, etc.) associated with times and locations. Fuel characteristics are considered by the model to influence emission rates.
- **Emission Rate Information:** Emission rates and statistical distribution information for the various pollutants and emission processes and kinds and model years of sources (including technology and standards discriminators) for each operating mode. Both mean emission rate and emission rate distribution information is needed in order to generate estimates of uncertainty. These rates are not a function of time or location.

Finally the default databases include all information needed by all implemented Generators to derive total activity, source bin, meteorology, fuel and operating mode information in the form required for the core model, as detailed in Section 7. This implies very significant elaborations to the default database. In particular, the macroscale implementation will include extensive default input information.

Conceptually the default input information is stored in several separate but related databases as follows:

- **Emission Rate Database:** Because of its size, unique structure, and major role in the model, default emission rate information is considered its own database.
- **National Default Database:** Those portions of the default database (other than emission rates) which do not depend upon locations or counties are considered to constitute a “National Default Database” or “Domain Default Database.”
- **County Default Database:** Those portions of the default database (other than emission rates) which depend upon county are considered the “County Default Database.”

In addition, the user may provide User Input Databases, directly or indirectly via *Data Importers*, for particular run(s) of the MOVES model. Their structure is the same as the corresponding Default Databases. Providing this information is optional at the macroscale if the user wishes to model the states and counties of the United States using default information. It is required for mesoscale or microscale modeling. It may be used to:

- Replace or override any of the information in the default databases.
- Supply the same kinds of information as the default database for geographic locations other than the states and counties of the U.S.; e.g., grid cells or smaller scale geographic units such as traffic analysis zones or roadway links.

6.2. Execution Databases

During execution of the MOVES model several databases will be created on a temporary basis. These are shown in the data flow diagram (Figure 5-3) as the Execution Location Database and the Execution Emission Rate Database. They have the same structure as the default and user input databases already described. These databases need not concern the end user. Their role is to facilitate efficient execution of the model. They store information which:

- Resolves differences between the user input and default data.
- Contain only information needed for a particular run of the model; e.g., only the source types, pollutants, locations, etc for which the user has requested that emission estimates be calculated.
- Facilitate the storage of intermediate results.

6.3. Output Database

MOVES will produce its core model output in the form of a simple database, as shown in Figure 6-1.

The MOVES Run table contains one record pertaining to the entire run of the model. Such as the time of the run, a reference to the Run Spec used, unit scaling factors, etc.

The MOVES Output table may contain many records, each reporting an individual emission factor and/or mass emission result for one pollutant. The level of aggregation of this table is governed by user input parameters in the run specification. In its most detailed form, it can contain hourly result records for each of the most detailed locations being modeled for each SCC code and power class (for nonroad) or source use type, fuel type, and roadway type (for highway vehicles) for each model year and emission process and pollutant. At the most disaggregate level the output of the MOVES output table would be comparable to the database output from MOBILE6. More detail may be found in Appendix B: MOVES Output Database Design.

It is up to the user how long these output databases are preserved and, in the initial implementations of MOVES, for managing their storage and retrieval. The database format facilitates aggregating or summarizing the results of individual runs, as well as combining or comparing the results of multiple runs. Initial implementations of MOVES will include a general purpose OLAP reporting tool for aggregating and totaling results and displaying them in a flexible category hierarchy, as well as “exporters” such as a specific tool to convert the generic MOVES output into a form suitable for input to SMOKE. Later use case implementations of MOVES are likely to include additional tools to facilitate the storage, retrieval, and comparison of MOVES run results for particular purposes.

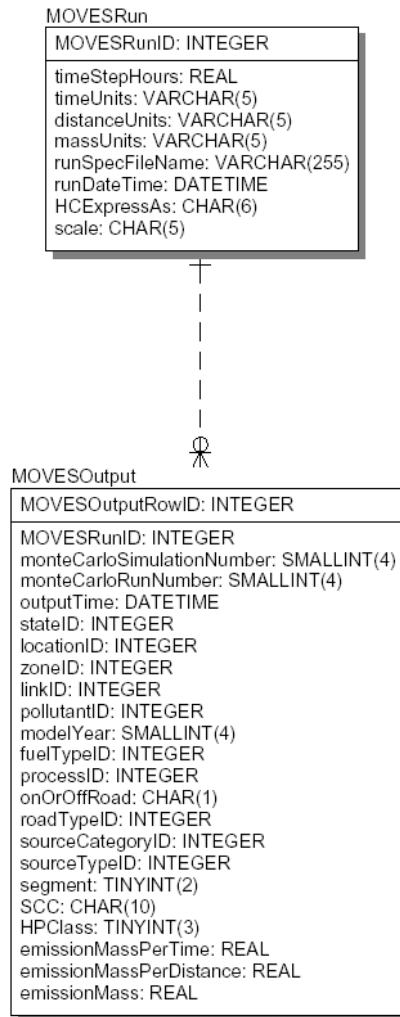


Figure 6-1. MOVES Output Database Design

7. Mathematical Formulation and Data Inputs

As discussed in Section 3, we are proposing that meeting the use cases presented in Table 2-1 will require specific “front-end” implementations be developed. These implementations will provide the necessary information to the core model using a large body of default data and best practice methods currently employed in developing mobile source emission inventory estimates.

In order to go from generally available information to input needed by the core model, MOVES will employ data generators, which are necessary specifically for the calculation of total activity and the distribution of operating modes. The nature of the MOVES framework is such that the majority of mathematical calculation will occur within the generators, specifically the Total Activity and Operating Mode Distribution Generators, which are linked to specific MOVES implementations. This section presents the proposed mathematical formulations for specific MOVES implementations.

We propose to develop total activity and operating mode generators for the following implementations:

- Macroscale on-road inventory development at two levels: the national level or a user-defined domain level
- Mesoscale on-road inventory development at the user-defined domain level
- Microscale on-road analysis in conjunction with CAL3QHC at the user-defined domain level
- Macroscale off-road inventory development at the national level

We also believe that the advanced transportation model TRANSIMS could be linked to MOVES through the development of specific generators. We do not propose to take on the development of these generators, but would support this development with developers of TRANSIMS.

As stated in Section 3, a primary objective of the MOVES design is to develop a generic framework that can apply to different analysis scales and emission processes. To maintain this design objective a generic approach to addressing the three analysis scales is required. The core model design is based on the idea of time and location, without regard to what the times and locations are; this is the generic feature of the core model. What primarily defines the difference between macroscale, mesoscale, and microscale are the definitions of time and location prior to implementation by the core model.

Location is defined by two elements: the “Domain,” which is the entire area being modeled, and “Zones,” which are subdivisions of the domain. MOVES will be flexible enough to work with any definition of domains and zones, as long as the user can supply necessary information to support that definition. Typical Domain/Zone combinations could be Nation/County, Region/Grid Cell, or County/Traffic Analysis Zone. For certain on-road emission processes, a third location element is also required: “Links,” which can either be

assigned to a Domain or a Zone. Links are defined more concretely depending on the scale being modeled, but the core model will fundamentally apply the concept of links across all scales.

An overview of the primary implementations are shown in Table 7-1, which highlights the definitions of *domain*, *zone* and *link* for each implementation.

Table 7-1: Implementations Proposed for MOVES

Implementation	Description	Use Cases Applications	Domain	Zone Locations	Link Locations
On-Road Macroscale Inventory (National Default)	Produces county-level inventory for entire nation.	National Inventory Development Policy Evaluation AQ Pre-Processor Interaction	U.S plus territories	Counties	HPMS Roadway Types
On-Road Macroscale Inventory (Domain-Specific)	Produces inventory for user-defined domain	Local Inventory Development (e.g., SIP/Conformity analysis) Policy Evaluation AQ Pre-Processor Interaction	User-defined; likely examples are counties, MSA, nonattainment areas, states	User-defined; likely examples are grid cells, zip codes, census blocks	HPMS Roadway Type
On-Road Mesoscale Inventory	Produces inventory for user-defined domain at link/zone level	Local Inventory Development Policy Evaluation Travel Demand Model Interaction	User-defined; likely examples are counties, MSA, nonattainment areas, states	User-defined; likely example is Transportation Analysis Zones (TAZ)	User-defined on-Network links Off-network HPMS road way types
On-Road Microscale Analysis w/ CAL3QHC	Produces necessary emission factor inputs for microscale analysis using CAL3QHC	Hot Spot Analysis Dispersion Model Interaction	CAL3QHC analysis area	N/A	Idle and non-idle links as defined in CAL3QHC input file

Table 7-1: Implementations Proposed for MOVES

Implementation	Description	Use Cases Applications	Domain	Zone Locations	Link Locations
Off-road Macroscale Inventory (National)	Produces county-level off-road inventory for entire nation.	National & Local Inventory Development Policy Evaluation AQ Pre-Processor Interaction	U.S plus territories	Counties	N/A

The total activity and operating mode generators will require significant mathematical processing and input data to convert from standardly available input data (e.g., VMT data) to the inputs necessary by the core model. Indeed, given the simplicity of the core model structure, the generator steps represent the most significant mathematical formulation elements of MOVES. The remainder of this section focuses on walking through the proposed mathematical formulation for the total activity and operating mode generators for the specific implementations supported by MOVES.

7.1. On-Road Implementation

For the on-road implementations, the generator functions depend heavily on how vehicles are grouped. We describe vehicle categorization in MOVES before going into the total activity and operating mode generators.

7.1.1. MOVES Vehicle Types

In MOBILE6, a single definition of vehicle groupings was developed and each was assumed to have the same emission rate and activity patterns. One shortcoming of this approach is that categories which are important for distinguishing unique emission rates (e.g., vehicle weight class) may not be the best category descriptions in terms of activity or may not match available fleet and activity data. Therefore, we are proposing to approach vehicle categorization in MOVES from the viewpoint of both activity and emissions, which will enable MOVES to take into account what is important for activity separately from what is important for distinguishing unique emission rates.

In terms of emissions, the parameters used to define source bins are based on factors determined to explain variability in emission production, such as fuel type, emission standard, technology type, vehicle weight, and engine size. In terms of activity, the MOBILE6 approach would be to also characterize activity according to these bin parameters; e.g., all trucks of a given fuel type and weight are similar in terms of usage patterns. This approach does not account for the difference in use patterns for vehicles which may fall within a given emission bin, for example a short-haul delivery truck and a garbage truck. Instead, for MOVES we are proposing a shift from characterizing vehicle activity based on emission-based parameters to characterizing vehicles by how they are used.

For characterizing activity, we are proposing MOVES categories known as *use types*, which are more closely aligned with vehicle use patterns. Use types will be defined by unique activity patterns. The link between use types and the emission-based source bins will be made by the source bin distributions, to enable the core model to take total activity from the use types and distribute it across the bins defined by unique emission rates. The important distinction between use types and source bin parameters is that activity is assumed to be constant across the source bin parameters. For example, under this approach diesel and gasoline passenger cars would be assumed to accrue mileage at the same average rate and travel at the same average speed for a given roadway type.

For on-road vehicles, we propose that MOVES use types be subsets of Highway Performance Monitoring System (HPMS) vehicle classes, which are important to retain in the MOVES design because VMT data is provided from the Federal Highway Administration (FHWA) by these classes. The VMT data provided by HPMS form the basis of macroscale on-road activity estimates in the total activity generator for all emission processes, and, hence, drive the structure of activity generation in MOVES.

Table 7-2 shows a breakdown of the HPMS vehicle classes, the proposed MOVES use types, and examples of source bin parameters. The use types shown in Table 7-2 are initial proposals which will likely need to be refined based on analysis of classification schemes in available data sources, most notably national registrations databases and the Vehicle In-Use Survey (VIUS), published by the U.S Census Bureau. A subset of source use type which will be accounted for in MOVES, but is not shown in Table 7-2, is source age, because activity does vary by age. Specifically, older vehicles accrue fewer annual miles than newer vehicles, an activity characteristic that is modeled in MOBILE6.

Table 7-2: Proposed MOVES On-Road Vehicle Classification Scheme

Activity Classifications		Emission Classifications Examples (Source Bins)					
HPMS Class	Use Type	Weight Range (pounds)	Fuel Types	Standard Groups	Technologies	Mileage Range	Emitter Class
Passenger Cars	Passenger Cars	< 4000 > 4000	Gas, Diesel, CNG, LPG, M85, E85, EV, HEV, FCV	Pre-control, Tier 0, Tier 1, NLEV, Tier 2	No catalyst Oxidation catalyst 3-way catalyst Carbureted Fuel injection EGR secondary air	Low Medium High	Normal High
Other 2-axle / 4-tire Vehicles	Passenger Trucks (Pickup, Minivan, SUV)	< 4000 < 6000 < 10,000					
	Work Trucks	< 6000 < 10,000 < 14,000					
Single Unit Trucks	Service Trucks	<14,000 <16,000 <19,500					
	Local Delivery Trucks	<26,000 <33,000 <60,000					
	Short-Haul Delivery Trucks						
	Motorhomes						
Buses	Interstate Buses		Gas, Diesel, CNG, LPG, M85, E85, EV, HEV, FCV	Gas: Pre-control, 1987, 1988, 1991, 1998, 2004, 2007 Diesel: Pre-control, 1985, 1988, 1990, 1991, 1998, 2004, 2007			
	Urban Buses						
	School Buses						
Combination Trucks	Long-Haul Delivery Trucks	< 33,000 < 60,000 > 60,000	Diesel				
Motorcycles	Motorcycles		Gas, EV				

7.1.2. Total Activity Generator

The macroscale on-road inventory implementations (national default and domain-specific) use estimates of vehicle miles traveled (VMT) as a starting point for generating total activity for most emission processes. This approach recognizes that emission production in a given location depends on the activity in that location, which may not be captured by what is registered in that location. This is particularly true of heavy-duty vehicles, for which the registered fleet and the on-road fleet can be quite different for a given location.

The basic process of total activity generation is the same for the national default and domain-specific implementations. The primary difference is that for the national case, total activity will be generated for the entire country (the default domain) and allocated to the county level (default zone location) using default geographic allocation factors. For the domain-specific case, total activity will be generated for the modeling domain based on user-supplied VMT data, which would override activity calculated from the default geographic allocation factors. If the user wishes, domain-level emissions could be allocated to the zone level (e.g. to facilitate calculation of emissions by grid cell) with user-supplied geographic allocation factors.

The steps necessary to produce total activity for these levels are detailed step-by-step in this section. First it is necessary to define several terms and acronyms found in the mathematical formulation. HPMS Vehicle Types and MOVES Use Types are listed in Table 7-2. The National Default Database (NDB) contains default fleet and activity information, which will be applied across the entire modeling domain unless overridden by user-supplied data. The Location Default Database (LDB) contains default location-specific information, which will be applied at the location level and can also be overridden by user-supplied data.

The first set of steps for total activity generation is to allocate VMT to use type according to the fleet and mileage accrual rates. This requires estimating the makeup of the fleet in a known base year (calendar year 2002 is proposed as the base year for MOVES GHG, the first release of MOVES), growing and scrapping the fleet to the desired analysis year, calculating the proportion of aggregate VMT traveled by use type and age in the analysis year (known as the “travel fraction”), and applying this to an estimate of aggregate VMT in the analysis year. Years prior to 2002 would be treated as base years as well, in that the databases will contain data from these years for performing the necessary calculations. The growth steps will only apply to years after the latest base year, beginning in 2003 in the case of MOVES GHG.

The mesoscale level of analysis proposed to be supported by MOVES requires that the user supply sufficient information to calculate bottom-up emissions at the link/zone level for running exhaust, start, brake wear, and tire wear processes. The most likely source of this information would be output from a traditional travel demand model. All other processes are handled using macroscale allocation approach down to the zone level, including off-network emissions for running exhaust and brake and tire wear processes. All fleet inputs are applicable across the entire modeling domain. The user must define the modeling domain, the road network (i.e., which roadways will be modeled as individual links), and analysis zones. More disaggregated approaches to the mesoscale model, such as the MEASURE framework developed

by Georgia Tech and EPA’s Office of Research and Development,⁶ could be run in MOVES with direct interface with the core model, but would not be supported as a specific implementation of MOVES.

The microscale level of analysis proposed to be supported by MOVES is based on necessary input for the dispersion model CAL3QHC, which is EPA’s recommended model for intersection analysis,⁷ applicable primarily to CO, PM and air toxics. Through the CAL3QHC input file, the user would supply sufficient information to calculate emission rates for running exhaust, brake wear, and tire wear at the link and intersection level. Other processes would need to be modeled through direct interface with the core model. All fleet inputs are applicable across the entire modeling domain. The user must define the modeling domain and the road network (i.e., which roadways will be modeled as individual links) through the CAL3QHC input file.

Table 7-3 lists all of the steps necessary for total activity generation at the macroscale, mesoscale, and microscale levels. Each of these steps is detailed in following text. All steps apply across each of the three scales, unless the mesoscale/microscale cases are specifically separated out.

Table 7-3: Overview of Total Activity Generator (TAG) Steps

Step		Description	Vary by Scale?
TAG-1	Calculate Base Year Vehicle Population	Establish base-year vehicle population in modeling domain by use type, age	No
TAG-2	Grow Vehicle Population to Analysis Year	If analysis year is in future, establish analysis year population through growth and scrappage	No
TAG-3	Calculate Analysis Year Travel Fraction	Calculation travel fraction across domain as function of age mix and annual miles traveled, by use type and age	No
TAG-4	Calculate Analysis Year VMT	If analysis year is in future, establish analysis year aggregate VMT across domain from base year VMT and growth	Yes
TAG-5	Allocate Analysis Year VMT By Roadway Type, Use Type, Age	Allocate aggregate VMT to roadway type, use type and age	Yes
TAG-6	Temporally Allocate Annual VMT to Hour by Roadway Type, Use Type, Age	Allocate annual VMT to hourly VMT	Yes
TAG-7	Convert to Total Activity Basis By Process	Convert to total activity basis at domain level: SHO, SHP, Starts, etc.	Yes
TAG-8	Allocate Total Activity Basis By Zone Location	Allocate total activity to zone locations using geographic allocation factors	No

Following the overall steps outlined above, the macroscale generator will be able to provide total activity data at the level needed by the core model: by emission process, time (one-hour increments), location (roadway “link” and zone), use type, and age.

TAG-1: Calculate Base Year Vehicle Population

Vehicle population by use type and age would be generated from national databases such as those compiled by R.L. Polk company and/or the HPMS system and from the Vehicle In-Use Survey (VIUS). User-supplied data could be developed from local registration databases, Inspection/Maintenance programs, and/or roadside surveillance programs such as remote sensing. Calendar year variation applies to years prior to the base year, for which historical data will be used directly in the calculation of emissions for those years. For example, MOVES GHG.1 will produce emissions for historical calendar years 1990 through 2001, with the base year being 2002.

Calculation	Base Year Population *	Age Distribution
Possible Sources	NDB, User	NDB, User
Can Vary By	Calendar Year, Use Type	Calendar Year, Use Type, Age

TAG-2: Grow Vehicle Population from Base Year to Analysis Year

The growth step is only necessary when the year being modeled is later than the base year. Starting with the base year vehicle population, the population must be grown to the analysis year using a year-by-year process. MOVES will allow for a dynamic registration process, meaning that age distributions will vary from year to year due to fluctuations in new sales. The calculation for this step depends on whether the age being modeled is zero, which must account for new sales, or non-zero.

For age zero vehicles, new sales are allowed to vary by year and will be expressed as a function of previous year sales (which are the same as the population of age zero vehicles in the previous calendar year, calculated from TAG-1). Sales rates will vary by use type and calendar year, which will provide the flexibility to input alternate sales penetrations for various use types (such as passenger cars and passenger trucks). National default growth rates will be developed based on historical trends and economic projections. User-supplied local data could be developed based on local trends and projections.

Survival rates the probability that vehicles will remain on the road, i.e. not be scrapped. Likely sources for national default levels are scrappage rates such as those published by Oak Ridge National Laboratory. Migration rates reflect the relative change of vehicle population due to vehicles entering or leaving the modeling domain. At the national default level it would likely be assumed that the migration rate is zero. Local rates could be developed from local registration records.

Age Zero (New Vehicles)

Calculation	Age Zero Vehicle Population In Previous Calendar Year *	Sales Rate *	Survival Rate *	Migration Rate
Possible Sources	TAG-1	NDB, User	NDB, User	NDB, User
Can Vary By	Calendar Year, Use Type	Calendar Year, Use Type	Use Type, Age (fixed)	Use Type, Age (fixed)

Other Years

Calculation	Vehicle Population In Previous Calendar Year *	Survival Rate *	Migration Rate
Possible Sources	TAG-1	NDB, User	NDB, User
Can Vary By	Calendar Year, Use Type, Age	Use Type, Age	Use Type, Age

To enable the dynamic registration feature, these steps would be repeated in every year until the analysis year was reached.

TAG-3: Calculate Analysis Year Travel Fraction

The travel fraction is defined as the proportion of VMT traveled by a given use type and age. This depends on the age distribution as well as per-vehicle annual mileage accumulation rates (MAR) by age. For future analysis years, the age distribution is calculated by dividing the population by use type and age (calculated in the previous step) by total population summed over use type and age. For historical years, the age distribution is already contained in the default or user-supplied database. The age distribution is then multiplied by the relative MAR in each year and divided by the product of these terms over all years to produce the travel fraction. The relative MAR is the relative per-vehicle mileage accumulation by use type and age across each HPMS class, compared to a fixed use type / age combination. We propose to use the same relative MAR for all years under the assumption that the relative level of travel for an average vehicle of a given use type will remain constant across all ages.

Calculation	(Vehicle Population / Total Vehicle Population Within Mapped HPMS Vehicle Class) *	Relative MAR / Sumproduct (Previous terms, over all ages)
Possible Sources	TAG-2	TAG-2 summed by use type & age within HPMS class
Can Vary By	Calendar Year, Use Type, Age	Calendar Year, HPMS Vehicle Class (Table 7-2)

TAG-4: Calculate Analysis Year VMT

For the base year and historic years at the macroscale level, VMT will be stored directly in the database (national VMT for the default case, user-supplied VMT for the domain-specific case), so a growth step is not required. Future years require the projection of VMT growth. It is important to note that under the proposed approach, VMT growth is treated as an external variable, rather than being calculated internally. Steps TAG-1 through TAG-3 are undertaken to develop the travel fractions necessary to allocate aggregate VMT generated in this step by use type and age.

Historic and base year VMT will be provided by HPMS estimates. Specifically, the FHWA report Highway Statistics (Table VM-1) would provide default national information by HPMS Class. Domain-specific estimates input by the user would be derived from HPMS or aggregated travel model input.

Overall VMT growth for the entire modeling domain is applied to base year VMT by HPMS vehicle class. Default VMT growth would be on the national level and would likely be developed on a calendar year basis, taking into consideration historical trends, economic projections, and current estimates of VMT growth. User-supplied local data should consider the same factors for the local domain being modeled. VMT growth is expressed here as linear growth, i.e. a percentage increase (which can vary by year) relative to VMT in the base year.

Evaluation of policies or scenarios affecting aggregate VMT growth would be employed with user input at this stage. The user would enter, by calendar year, aggregate VMT growth rates on a domain-wide level as part of the Source Usage Control Strategy.

Macroscale TAG-4

Calculation	Base Year VMT	$1 + [(\text{Projected annual linear VMT growth}) * (\text{Analysis Year} - \text{Base Year})]$
Possible Sources	NDB, User	NDB, User
Can Vary By	HPMS Vehicle Class	Calendar Year, HPMS Vehicle Class

The primary difference between macroscale and the meso/microscale is that, for the latter, estimates of VMT are generated from link-level activity estimates. For on-network travel (the network represented by links), VMT estimates for these scales are calculated on a link-basis from user input of and link length from travel demand models, CAL3QHC and/or auxiliary sources such as TIGER. Volume is defined as the number of vehicles traveling over a link during the analysis time period.

Mesoscale/Microscale On-Network TAG-4

Calculation	Analysis Period Volume *	Link Length *
Possible Sources	User	User
Can Vary By	Roadway Link	Roadway Link

Off network VMT for mesoscale is handled in much the same way as macroscale VMT. The user would be required to input base year off-network VMT, likely from HPMS estimates, and provide growth estimates to allow the calculation of analysis year off-network VMT.

Mesoscale Off-Network TAG-4

Calculation	Base Year Off-Network VMT	1+ [(Projected annual linear VMT growth) *]	(Analysis Year - Base Year)]
Possible Sources	User	User	Calculated
Can Vary By	HPMS Vehicle Class	Calendar Year, HPMS Vehicle Class	

TAG-5: Allocate Analysis Year VMT By Roadway Type, Use Type, Age

At the macroscale level, this step allocates aggregate VMT in the analysis year (calculated in the previous step) to roadway type, use type, and age based on analysis year travel fractions and roadway allocation factors. The new input for this step is roadway allocation factor, which allocates aggregate VMT to the 12 HPMS roadway types. It is assumed that roadway allocations in the analysis year are the same as the base year. National default roadway allocation factors would be derived directly from national summaries of HPMS VMT data, as summarized in Table VM-2 of FHWA’s Highway Statistics report. User-supplied information at the local level could be derived from state or county-level HPMS data.

The allocation to roadway type at the domain level is necessary at this stage because the conversion to SHO from VMT in later steps relies on average speed, which varies by roadway type. Variations in roadway allocation at the location level can be addressed through the development of geographic allocation factors in step TAG-8, which vary by roadway type as well.

Macroscale TAG-5

Calculation	Analysis Year VMT *	Roadway Allocation *	Analysis Year Travel Fractions
Possible Sources	TAG-4	NDB, User	TAG-3
Can Vary By	Calendar Year, HPMS Class	HPMS Roadway Type, Use Type	Calendar Year, Use Type, Age

The allocation of on-network VMT at the Mesoscale/Microscale levels would be calculated by applying the domain level travel fractions at the link level.

Mesoscale/Microscale On-Network TAG-5

Calculation	Analysis Year VMT *	Analysis Year Travel Fractions
Possible Sources	TAG-4	TAG-3
Can Vary By	Roadway Link	Calendar Year, Use Type, Age

The allocation of off-network VMT at the Mesoscale levels would be calculated by applying the domain level travel fractions to the off-network VMT.

Mesoscale Off-Network TAG-5

Calculation	Analysis Year VMT *	Analysis Year Travel Fractions
Possible Sources	TAG-4	TAG-3
Can Vary By	Calendar Year, HPMS Class	Calendar Year, Use Type, Age

TAG-6: Temporally Allocate Annual VMT to Hour by Roadway Type, Use Type, Age

Step TAG-5 produced aggregate annual VMT estimates. Step TAG-6 then allocates this down to the hourly level, the basic unit of time for macroscale and mesoscale analysis. This requires a series of temporal allocation factors: year to month, month to day, and day to hour. These will require an awareness of the calendar to reflect idiosyncrasies such as variation in days per month. Monthly allocation factors will take into account differences due to season and length of the month and will need to take into account leap years. National defaults have been developed for EPA’s National Emission Inventory (NEI). Daily allocation factors will be developed for each day in the week, would take into account differences such as weekday versus weekend travel, which will vary by vehicle class and roadway type. Daily allocations will depend on the month, since a given month has a different number of days. Hourly allocation will account for hourly differences within a day, which will depend on the day of the week, vehicle class, and roadway type.

National defaults for monthly allocation factors have been developed for the NEI, which we propose to use for MOVES as well. Default daily allocation factors would need to be developed based on HPMS VMT estimates. Hourly allocation factors have been developed from driving survey data as part of MOBILE6, and we propose to use these as the default for MOVES as well.

Macroscale TAG-6

Calculation	Analysis Year VMT *	Monthly Allocation Factor *	Daily Allocation Factor *	Hourly Allocation Factor
Possible Sources	TAG-5	NDB, User	NDB, User	NDB, User
Can Vary By	Calendar Year, Roadway Type, Use Type, Age	Month, Use Type, Calendar Year (Leap Year / Non-Leap Year)	Month, Day of Week, Use Type, Roadway Type	Day of Week, Hour, Roadway Type, Use Type

The primary difference in temporal allocation between mesoscale and macroscale is that the mesoscale activity from travel demand models is typically expressed over periods less than a day, such as am peak, pm peak, and off-peak. Thus the hourly allocation factors would reflect a period shorter than one day, although they could be derived from the national default daily-to-hourly allocation factors.

Mesoscale TAG-6

Calculation	Analysis Period VMT *	Hourly Allocation Factor
Possible Sources	TAG-5	NDB, U
Can Vary By	Network/Off-Network, Roadway Link, Use Type, Age	Month, Day of Week, Use Type, Network/Off-Network, Roadway Link

TAG-7: Convert to Total Activity Basis By Process

The next series of steps relates to producing total activity for each of the processes based on VMT. The total activity to be calculated includes: (a) source hours, (b) source hours operating, (c) source hours parked, (d) number of starts, and (e) gallons refueled. This conversion is performed at the domain level primarily to maintain consistency between the different activities. The method outlined in this step assumes that the modeling domain is a “closed box”, in which all operation and parking occur. In particular this assumption greatly simplifies the calculation of source hours parked, since source hours and source hours operating are known. The “closed box” approach is more applicable to large modeling domains than specific locations, hence the generation of total activity at the domain level.

Total Source Hours

Total source hours (SH) is a necessary input in the computation of Source Hours Parked (SHP), the activity basis for diurnal, hot soak, and resting loss. Total source hours in the modeling domain for the hour being analyzed amounts to the vehicle population in the domain for that hour, split by use type and age, which was calculated under a previous step. The key assumption in this approach is that all vehicles which exist within a modeling domain are reflected in the VMT estimates, which were used to generate the population. In other words, all vehicles which produce emissions drive at some point within a year.

Calculation	Analysis Year Vehicle Population *	1 Hour
Possible Sources	TAG-2	
Can Vary By	Calendar Year, Use Type, Age	

Source Hours Operating

Source Hours Operating (SHO) is proposed as the activity basis for running exhaust, crankcase, running loss, brake wear, and tire wear processes. SHO is calculated by dividing VMT on a given roadway type, use type, age, and hour by the average speed, which could vary by day of week, hour, use type, and roadway type. National defaults for average speed would be developed by instrumented vehicle surveys or speed studies. User-supplied data could come from similar sources at the local level.

Macroscale On-Road SHO

Calculation	Analysis Year VMT /	Average Speed
Possible Sources	TAG-6	NDB, User
Can Vary By	Calendar Year, Roadway Type, Use Type, Age, Month, Day, Hour	Day of Week, Hour, Use Type, Roadway Type

The primary difference with the mesoscale/microscale is that average speed will be provided by the user at the link level, based on input from a travel demand model and/or post-processing.

Mesoscale/Microscale On-Road SHO

Calculation	Hourly VMT /	Average Speed
Possible Sources	TAG-6	User
Can Vary By	Roadway Link, Use Type, Age, Hour	Roadway Link, Use Type, Hour,

Mesoscale off-network SHO is generated by dividing off-network VMT by the average off-network speed, by use type and hour.

Mesoscale Off-Network On-Road SHO

Calculation	Hourly VMT /	Average Speed
Possible Sources	TAG-6	NDB, User
Can Vary By	Calendar Year, Use Type, Age, Hour	Use Type, Hour

An emerging concern in the on-road vehicle inventory is extended idling, primarily from heavy-duty vehicles, for which *hoteling* (overnight idle) is common. Extended idle is not reflected in the on-road VMT estimates used to generate on-road SHO and must be accounted for

separately. MOVES will explicitly account for extended idling activity through the estimation of extended idle SHO, which will be treated within the model as a 13th roadway type (allowing it to fit into the allocation scheme discussed in the next section). We are proposing to do this by developing an “Idle SHO Factor” which is expressed as “Idle SHO per On-Road SHO.” We propose that this factor is best applied at the daily level, rather than the hourly level, since extended idle is more likely to happen in a daily cycle.

<i>Extended Idle SHO</i>		
Calculation	Daily On-Road SHO *	Idle SHO Factor (Idle SHO / On-Road SHO)
Possible Sources	TAG-7 On-Road SHO Summed By Day	NDB, User
Can Vary By	Calendar Year, Use Type, Age, Hour	Use Type, Hour

Total SHO is calculated as the sum of On-Road SHO and Extended Idle SHO.

Source Hours Parked

Source Hours Parked (SHP) is proposed as the activity basis for evaporative emission processes, specifically diurnal, resting loss, and hot soak. As proposed, SHP would be calculated as the difference between Source Hours and Source Hours Operating. The primary benefit of this is that it will ensure a direct link between activity basis for exhaust and evaporative emissions. Because roadway type is not relevant to evaporative emissions, source hours operating would be summed across roadway type to reflect the operation in the entire modeling domain.

Calculation	Source Hours -	Total Source Hours Operating
Possible Sources	TAG-7 SH	TAG-7 Total SHO
Can Vary By	Calendar Year, Use Type, Age, Hour	Calendar Year, Use Type, Age, Month, Day, Hour

Number of Starts

Number of starts is the activity basis for one process: start exhaust. To ensure consistency across exhaust emission processes, we are proposing to link the number of starts to source hours operating by developing a “Starts per SHO” factor, which would vary by day of week, hour, and use type. Whether the SHO for this factor is based on the total SHO or only on-road SHO will need to be determined. This factor is not assumed to vary by age; but linking starts to SHO ensures that the variation in travel by vehicle age is reflected in the number of starts as well. National default levels for this factor would be derived from driving survey data, such as that used in the development of start per mile estimates in MOBILE6. User-supplied local estimates could be derived from driving survey data as well.

Calculation	SHO *	Starts Per SHO
Possible Sources	TAG-7	NDB, User
Can Vary By	Calendar Year, Use Type, Month, Day, Hour	Day of Week, Hour, Use Type

Gallons Refueled

Gallons of fuel used (and, hence, refueled) is the activity basis for the refueling emission process. This is directly tied to VMT using an average miles per gallon estimate by vehicle class and age. VMT would be aggregated by day and roadway type.

Calculation	Daily VMT *	Gallons Per Mile
Possible Sources	TAG-6	NDB, User
Can Vary By	Use Type, Age, Month, Day	Use Type, Age

TAG-8: Allocate Total Activity to Zone Location

Once total activity has been calculated for the modeling domain, geographic allocation is necessary to estimate activity within each zone location. This will require the development of geographic allocation factors for each zone location and activity basis. For the default level, allocation factors will be developed to allow allocation from the national level to the county level, and stored in the default county database (CDB). For domain-specific modeling, these factors would be supplied by the user to allow allocation from the domain to the user-defined zone level.

Source Hours Operating (Macroscale Only)

The allocation of domain-level total activity to the zone location level is a critical step for SHO macroscale emission estimation. It will require allocation factors to be developed for each zone location and emission process. Allocation factors for all emission processes will vary by zone location and calendar year, which will allow them to account for location-specific differences in growth.

SHO allocation factors will also vary by roadway type. This will enable location-specific adjustments to be made to the roadway allocations previously. At the national level, default allocation factors could be developed using a similar process employed in the development of county & roadway level VMT estimates in the NEI, which relied on roadway length data. Allocation factors would also allow for locality-specific growth, for which defaults could likely be developed from census population growth estimates. Because extended idle will be treated as an additional “roadway type”, specific geographic allocations can be tailored to where concentrations of extended idle are more likely (e.g., truck stops).

Allocation factors would only be necessary for domain-level modeling if zone locations smaller than the modeling domain were desired. One example of this might be the modeling of grid cells within a modeling domain for input into emission processors like SMOKE.

Calculation	SHO *	SHO Geographic Allocation Factor
Possible Sources	TAG-7	CDB, User
Can Vary By	Calendar Year, Use Type, Age, Month, Day, Hour	Calendar Year, Zone Location, Roadway Type

Source Hours Parked

Allocation factors for SHP are independent from those for SHO, a significant shift from the MOBILE modeling framework (which links evaporative emissions to miles traveled). This change in approach for MOVES enables a decoupling of where vehicles travel and where they park, which may vary greatly depending on whether one is considering an interstate in a rural county or a urban area with heavy commuter influx. Whereas SHP depends on SHO at the domain level, at the zone location level SHP will not depend on miles traveled or SHO in that zone. This will allow SHP to be determined by variables which are more pertinent to vehicle parking, such as number of parking spaces, office space, number of residences, etc. National default allocations could be developed using these types of surrogates, from census data at the county level.

Calculation	SHP *	SHP Geographic Allocation Factor
Possible Sources	TAG-7	CDB, User
Can Vary By	Calendar Year, Use Type, Age, Month, Day, Hour	Calendar Year, Zone Location, Hour

Source Hours

For a given location, Source Hours is simply the sum of SHO and SHP within each zone location; no allocation factor is necessary.

Calculation	SHO +	SHP
Possible Sources	TAG-7	TAG-7
Can Vary By	Calendar Year, Use Type, Age, Month, Day, Hour, Zone Location	Calendar Year, Use Type, Age, Month, Day, Hour, Zone Location

Number of Starts (Macroscale Only)

Start allocation factors, as with SHP, are independent from SHO. This allows more flexibility than MOBILE6, which links starts to miles traveled. The MOBILE6 approach assumes that starts happen in proportion to where miles are traveled. By separating the allocation factors, starts can be estimated at the local level based on parameters more directly related to starts (which are generally the same as SHP), such as parking spaces and residential density. National default allocations could be developed using these types of surrogates, from census data at the county level.

Calculation	Number of Starts *	Start Geographic Allocation Factor
Possible Sources	TAG-7	CDB, User
Can Vary By	Calendar Year, Use Type, Month, Day, Hour	Calendar Year

Gallons Refueled

Refueling allocation factors are also independent from SHO. This allows more flexibility from MOBILE6, which links refueling to miles traveled. The MOBILE approach assumes that refueling happens when miles are traveled. By separating the allocation factors, refueling can be estimated at the local level based on parameters more directly related to refueling, primarily the concentration of refueling stations. National default allocations could be developed using gas station density. Allowing allocation factors to vary by hour will allow flexibility in assessing programs such as Ozone Action Days, which encourage shifts in source refueling to evening hours.

Calculation	Gallons of Fuel Refueled *	Refueling Geographic Allocation Factor
Possible Sources	TAG-7	CDB, User
Can Vary By	Calendar Year, Use Type, Month, Day	Calendar Year, Hour

7.1.3. Operating Mode Distribution Generator

For on-road emissions, the second generator requiring significant mathematical processing to calculate core model input is the Operating Mode Distribution Generator (OMDG). The output of this generator is the distribution of operating modes as defined for each emission process, for those emission processes which will break total activity into modes. A significant design feature of MOVES is that the same operating mode distribution generation process can be applied to each of the analysis scales.

Operating modes are defined as breakdowns of total activity necessary to reflect differences in emission rates. The Emission Rate Database will contain emission rates broken down by operating mode bin. Table 7-4 shows our proposed operating mode parameters for

emission processes requiring further activity breakdown. Emission processes not listed will not require the breakdown of total activity into modes.

A central feature of MOVES is that the design has been developed such that the same definition of operating modes and the same operating mode-based emission rates will be applied at all three analysis scales of the model. The Operating Mode Distribution Generator and processes within the generator will function independent of scale.

Table 7-4: Operating Mode Parameters By Emission Process

Emission Process	Operating Mode Parameter(s)
Running Exhaust, Brake Wear, Tire Wear	Average Speed and Vehicle Specific Power (VSP)
Start Exhaust, Hot Soak	Soak Time
Diurnal	Tank Pressure
Running Loss	Time Since Start

Distributions for soak time and time since start will be stored directly in the Execution Location Database; default values will be developed primarily from driving surveys. User-supplied information from local surveys can be input to replace these defaults. For these modes, the OMDG will simply transfer the distributions from the Execution Location Database to the core model.

Vehicle Specific Power and Tank Pressure will require mathematical processing within the OMDG to go from generally available input data to the distribution of modes according to the parameters specified in Table 7-4. The step-by-step processes proposed for calculating operating mode distributions for these parameters are discussed in the following.

7.1.3.1. Vehicle Specific Power

Vehicle Specific Power, or VSP, is generally defined as power per unit mass of the source. The calculation of absolute power generally centers on the forces a vehicle must overcome when operating on the road, including: acceleration, the force of gravity due to positive road grade, tire rolling resistance, and aerodynamic drag. Normalizing this power by mass to calculate VSP allows for this metric to be estimated based only on the instantaneous speed of the vehicle and road grade, if assumptions for the coefficients of rolling resistance and drag are made and no wind speed is assumed, as discussed by Jimenez.⁸ The basic concept of VSP has been applied in numerous studies in different forms and has shown to be a useful metric for characterizing vehicle emissions. As discussed by Jimenez, similar metrics have included “Positive Kinetic Energy” (PKE) proposed by Watson et al.,⁹ and “Specific Power” employed in studies such as EPA’s FTP Revision Study¹⁰ and the development of MOBILE6 driving cycles,¹¹ which were based on the product of speed and acceleration without consideration for road load effects.

We are proposing that MOVES use Vehicle Specific Power (VSP) to characterize “modal” emission rates. The primary benefit of this metric is that it combines into a single parameter numerous physical factors influential to vehicle fuel consumption and emissions: vehicle speed, acceleration, road grade, and road load parameters such as aerodynamic drag and rolling resistance. By normalizing the vehicle weight, we can directly calculate VSP knowing only the driving pattern of a vehicle and road grade, if assumptions are made for road load coefficients. An in-depth evaluation of the merits of using VSP will be presented in a separate document detailing the MOVES emission analysis plan.

The implementation of VSP will take the form of bins defined by ranges of VSP. An initial proposal for the definition of these bins has been developed for EPA by North Carolina State University under contract to evaluate the VSP binning approach.¹² These are shown in Table 7-5. More detail on the development of these bins can be found in the report.

Table 7-5: Proposed VSP Bin Definitions

VSP Bin	Definition (VSP in kW/Metric Ton)
1	$VSP < -2$
2	$-2 \leq VSP < 0$
3	$0 \leq VSP < 1$
4	$1 \leq VSP < 4$
5	$4 \leq VSP < 7$
6	$7 \leq VSP < 10$
7	$10 \leq VSP < 13$
8	$13 \leq VSP < 16$
9	$16 \leq VSP < 19$
10	$19 \leq VSP < 23$
11	$23 \leq VSP < 28$
12	$28 \leq VSP < 33$
13	$33 \leq VSP < 39$
14	$39 \leq VSP$

Further work by EPA, to be presented in a separate document, has determined that model accuracy can be improved by using these VSP bin definitions in conjunction with average cycle speed; from this finding we are proposing that operating mode bin definitions for the running exhaust emission process be first subdivided by average cycle speed (for example: less than 20 mph, between 20 and 40 mph, and greater than 40 mph), then by the VSP bins shown in Table 7-5. The operating mode distribution generator will produce the distribution of time spent in each of these bins for a given roadway link. Default driving cycles will be used to calculate these distributions within MOVES. We are proposing to base the calculation of average cycle speed

and VSP bin distributions in MOVES on the MOBILE6 framework for facility-specific driving patterns. MOBILE6 uses driving cycles to reflect varying Levels of Service by aggregated facility type: freeway, arterial, and local roads. MOVES would extend these cycles to represent travel on the 12 HPMS roadway types. With this approach, the user would interface with MOVES at the level of average speed and facility type, which maintains consistency with the MOBILE6 input structure. These driving cycles (expressed as second-by-second vehicle speed) can be converted to VSP in each second, along with input factors such as road grade, rolling resistance, and wind speed. We can then determine the distribution of time falling in the average speed and VSP bins. This distribution would then be passed to the core model. The specific steps for accomplishing this within the operating mode distribution generator are shown in Table 7-6:

Table 7-6: Overview of Operating Mode Distribution Generation For VSP

Step		Vary by Scale?
OMDG-1	<i>Determine Distribution of Driving Cycles</i>	No
OMDG-2	<i>Calculate Second-By-Second VSP</i>	No
OMDG-3	<i>Calculate VSP Bin Distribution</i>	No

OMDG-1: Determine Distribution of Driving Cycles by Use Type, Roadway Type

For the given roadway link being analyzed, the initial inputs to the OMDG will be average speed (or speed distribution) by vehicle use type and HPMS roadway type. National defaults will be available for macroscale analysis, likely derived either from the roadway-specific speeds used in the NEI or those derived as defaults for MOBILE6. Users can supply local information on average speeds from sources such as speed studies or transportation models.

The MOVES Execution Location Database (ELD) will store driving cycles that represent fixed average speeds on specific roadway types. As mentioned, the national defaults will be available to all analysis scales based on the facility-specific driving patterns developed for MOBILE6. These will be stored as second-by-second speed trajectories. Users can input alternate second-by-second speed trajectories by average speed and HPMS roadway type from sources such as instrumented vehicles or output from transportation models.

Under this step the average speed or speed distribution will be mapped to the driving cycles stored in the ELD, using a “bracketing” approach. The determination of “bracket” speeds on a given roadway link would take the average speed and find the next highest and next lowest speeds for which driving patterns are stored, by the roadway type being modeled. Weights would then be applied to the “bracketed” driving cycles to match the average speed being modeled. If a speed distribution was supplied, this process would be repeated for each individual speed point and the overall driving cycle mix reweighted to match the input speed distribution.

Because it is proposed that average cycle speed bins will be used in conjunction VSP bins to further refine model, this process would be repeated within each average speed bin.

OMDG-2: Calculate Second-By-Second VSP For Relevant Driving Cycles

For each driving cycle mapped in the previous step, Vehicle Specific Power would be calculated on a second-by-second basis using the following equation:

$$\text{VSP} = v * (a*(1+\epsilon) + g*\text{grade} + g*C_R) + 2\rho*C_D*A*v^3/m$$

where:

- v: is vehicle speed (assuming no headwind) in m/s
- a: is vehicle acceleration in m/s²
- ε: is mass factor accounting for the rotational masses
- g: is acceleration due to gravity
- grade: is road grade
- C_R: is rolling resistance
- ρ: is air density
- C_D: is aerodynamic drag coefficient
- A: is the frontal area
- m: is vehicle mass

Speed and acceleration in each second would be supplied directly from the drive cycle used in the calculation. Default values of the other variables would be stored in Execution Location Database, with the opportunity for users to supply custom information. The default value for grade would be zero; we propose that users could enter grade information by roadway link for mesoscale or microscale analysis only. Defaults for vehicle-related parameters such as rolling resistance, drag coefficient, frontal area, and vehicle mass would be available by vehicle use type category and could be replaced by user-supplied input.

OMDG-3: Calculate VSP Bin Distribution

For each driving cycle, second-by-second VSP calculations from the previous step would then be binned, according to the definitions in Table 7-5, by calculating the frequency of VSP occurrences in each bin. Within each average speed bin the overall VSP distribution for a given roadway link and use type would then be calculated by multiplying the VSP bin distributions from each relevant driving cycle with the distribution of each driving cycle calculated in Step 1. These VSP bin distributions would be multiplied by the average speed bin distribution to generate the the final operating mode distribution across average speed/VSP bins. The extended idle roadway type will, by definition, have 100 percent operation in VSP Bin 3, which includes idle operating. The resulting VSP bin distribution by use type and roadway link would be forwarded to the core model.

7.1.3.2. Tank Pressure

We propose using tank pressure as the basic parameter for distinguishing diurnal emission rates. Diurnal emissions are generally caused by an increase in tank pressure due to temperature rise. Hence, a simple division of tank pressure modes could be increasing pressure, stable pressure, or decreasing pressure.

7.1.3.3. Soak Time and Time Since Start

As mentioned, Soak Time and Time Since Start will not require calculations within the OMDG. The level of disaggregation is shown in Table 7-7 for each related process.

Table 7-7: Level of Disaggregation for Soak Time and Time Since Start

Non-Calculated Operating Mode Parameter(s)	Can Vary By
Soak Time	Use Type, Day of Week, Hour
Time Since Start	Use Type, Day of Week, Hour

7.2. Off-Road Implementation

The fundamental underpinnings of the core model apply to off-road sources as well as on-road sources: all sources have total activity, a distribution of that activity into specific modes, and emission associated with those modes. The differences between on-road and off-road sources will be handled through unique total activity and operating mode distribution generators.

We are proposing, as an initial step, that these front-end generators be developed specifically to support macroscale analysis of off-road sources (including commercial marine, aircraft, and locomotives), at the same two sublevels applied for on-road. The national default level would function similar to the current NONROAD model, producing county-level emissions for the entire country. The domain level would allow users to model specific areas, but would require the input of domain-specific inputs for calculating total activity. Finer scale analyses of nonroad would be possible with direct interaction of the core model if the user had fleet and activity data for the area being modeled.

7.2.1. Total Activity Generation

Because the development of an off-road implementation will follow that of an on-road implementation, at this stage we are merely presenting options for how total activity would be accounted for with off-road sources. For starters, we propose that off-road emissions be modeled in a more aggregate way than on-road emissions in several ways. First, the emission processes presented in Table 3-1 for on-road would likely be reduced to (a) exhaust, (b) crankcase, (c) diurnal, (d) resting loss, (e) running loss, and (f) refueling. The total activity basis for each process defined in Table 3-2 would remain the same, reflecting the broad applicability of the core model design. A primary reason for adopting the time-based activity approach is, in fact, to accommodate off-road emissions.

Three options we are aware of for generating total activity for off-road emissions are as follows:

- **Population Method:** Continue the approach currently used in NONROAD, which is based on estimates of equipment population and per-vehicle hours operating from survey data. The total activity generator would take these inputs, apply growth and scrappage as necessary, and multiply analysis year population and per-vehicle hours to generate Vehicle Hours Operating across the modeling domain, by source class. Other activity basis (e.g., SHP) would be linked to population and SHO.
- **Surrogate Method:** Develop relationships between surrogate inputs, such as tillable acres, housing starts, etc., and total activity for specific source classes. The total activity generator would take these surrogates as input and apply these relationships to generate SHO across the modeling domain. Other activity basis (e.g., SHP) would be linked to SHO.
- **Density Method: Develop** estimates of activity (SHO) density for specific source classes and land use types. The total activity generator would take these density estimates and extrapolate them to the modeling domain based on domain size and land-use mix. Other activity basis (e.g., SHP) would be linked to SHO or would have unique activity density estimates.

Each of these three methods would generate total activity across the entire modeling domain. The total activity generator would then allocate to specific locations (county or user-defined, such as grid) using geographic allocation factors stored within the Execution Location Database. The current NONROAD model uses this approach.

7.2.2. Off-Road Operating Mode Distribution Generation

We propose the application of operating modes for off-road be at a more aggregate level than for on-road. Primarily, we are proposing not to break SHO into modal (power-based) bins. The level of disaggregation is likely less necessary for off-road than for on-road. At the far extreme, it may be viable for a first iteration to not split total activity into operating modes at all (as done in the current NONROAD model), which would mean that an OMDG was not needed for the off-road implementation. An in-between option would be to consider very coarse operating modes, such as idle and non-idle.

8. Software Implementation

MOVES will be executed using specific software tools and techniques. In this section, we explain our key decisions and requirements in choosing a specific programming language and database. We also briefly describe our work on a distributed processing system, which will increase the speed of doing many runs and be easy for users to implement.

To actually guide the selection of software development methods and tools, it is desirable to focus on the specific requirements that have been discussed. From these we know that the software methods and tools chosen to develop MOVES should:

- Be state of the art.
- Encourage, via conventions and standards related to the tools, good design and programming practices and documentation.
- Promote an object-oriented approach to software design and development.
- Operate on, and produce model products that operate on, current and future EPA desktop computers that are Intel™ style PCs using the Windows™ software operating system.
- Produce model products which are easily ported to other platforms, (e.g., Linux and Unix-based systems)
- Produce products that do not require the end user to purchase software licenses beyond a software operating system.
- Facilitate development of graphical user interfaces in addition to more traditional ASCII text file input/output.
- Facilitate, or at least be compatible with, use of MIMS, a modeling framework being developed by EPA/ORD.
- Execute in an acceptable amount of time. We know from the current MOBILE and NONROAD models that MOVES is intended to replace, and from the use cases intended for MOVES, that model execution time performance is an important consideration.
- Employ standard methods and tools. It is always a good idea that the methods and tools chosen be widely used in the industry and have a promising future. (Though it is often difficult to see very far ahead.) There isn't anything so demanding or specialized about the MOVES software requirements that one would expect unusual methods or approaches to be needed.

Because MOVES represents a significant and long term departure from current EPA/OTAQ models, it is not of high importance, as it would be for many projects, that the methods and tools chosen be ones with which EPA/OTAQ and its contractors are currently familiar.

8.1. Software Development Methods Chosen for MOVES

The scope of intended use cases for MOVES is sufficiently broad that a framework for MOVES and its implementations should be explicitly designed, and these designs documented, in some detail. On the other hand MOVES is not so complex that a highly elaborate or formal design approach is needed. Indeed, this is not desirable since detailed requirements are apt to evolve throughout the project and cannot be fully known at the outset.

This consideration has led us to an approach which attempts to blend the best aspects of traditional, structured software development with those of a newer approach called “extreme programming.” This new approach is adapted to the reality that software requirements cannot be known in advance and are often discovered by repeated iterations of development and testing in which the end user is directly involved. For more information on this approach the reader is referred to *Extreme Programming Explained* by Beck.¹³ Automated testing is an aspect of extreme programming which we intend to use for MOVES, writing tests concurrently with the software, and maintaining a suite of tests that can be run at any time to verify that the software has not been “broken” by subsequent changes.

Perhaps of most general significance, we have adopted the practice of iterative software design and development. This means that the usual development life cycle of design, programming, and testing—leading to eventual product release and use—will be repeated many times, allowing EPA to gain expertise and for MOVES to develop incrementally. This will allow us to work closely with customers, for requirements to change, to get rapid feedback, and to produce frequent internal and external product versions.

Additional detail regarding the coding standards being adopted for MOVES and plans for software documentation, version control, and testing will be documented in the Quality Assurance Program Plan for the project.

8.2. Software Development Tools Selected for MOVES

Of the many software tools used in this project, we felt two areas needed some discussion about what we choose and why: programming language and Database Management Software (DBMS).

8.2.1. Programming Language

While Fortran has been used for past and current EPA/OTAQ models and is still an excellent language for mathematical computations, it is not strong with regard to object orientation, is not particularly platform independent, and is certainly not a good choice for graphical interfaces. So our requirements really narrow the choice of reasonable language alternatives for MOVES to either Java or C++.

We have made a preliminary decision to program MOVES with the Java programming language as defined by SUN’s Java System Development Kit (SDK). It is a state-of-the-art,

widely used, object-oriented, programming language, which has enjoyed considerable success over the last several years and appears to have a promising future, independent of the fate of any particular company. With its “write once, run anywhere” philosophy, it offers the highest degree of platform independence of any language available. Java comes with a set of associated tools, such as Javadoc, which lead to relatively uniform coding and documentation practices. Java is the native programming language of MIMS and so using Java maximizes our ability to take advantage of MIMS. (An initial example of this is that we have made a preliminary decision to adopt the MIMS Java Programming Standard as the basis of the MOVES programming standard.) Sun’s implementation of Java is the most widely used and clearly the most committed to remaining industry standard.

Java is ideal with respect to all of the MOVES software requirements except execution speed performance. Because it is a partially compiled and partially interpreted language, its execution time performance will never be as rapid as fully compiled languages such as Fortran or C++. On the other hand we don’t know exactly where MOVES performance bottlenecks will occur or how much of a problem they might be. Also Java performance has improved and is likely to continue to improve with future versions. We realize that it may eventually be necessary, if portions of MOVES written in Java prove to be performance bottlenecks, to convert them to another programming language, such as C++.

The SDK includes Javadoc, a utility program that produces HTML-based software documentation from comments embedded in the Java programming code. The MOVES project has incorporated Javadoc commenting conventions in its coding standard and intends to use Javadoc as its principal software documentation tool.

8.2.2. Database Management Software (DBMS)

As described earlier in this document, MOVES is data driven, meaning that its emission rates are primarily determined by looking them up in a database. Input and output data are also in databases. Furthermore, it is a fundamental objective of MOVES that it be able to incorporate new vehicle population, vehicle activity, and emission measurement data easily. Given the complexity of such data, and the limited input/output capability of the Java and C++ programming languages, there is a strong requirement that MOVES include relational database management software (RDBMS) for the loading, storage, and (especially) retrieval of information. For this RDBMS to be state-of-the-art, platform independent, and in widespread use further dictates that it allow the use of SQL (Structured Query Language) for database operations.

We investigated the two principal open-source RDBMS in widespread use, MySQL and PostgreSQL. PostgreSQL requires an additional, separate product to implement as a server on Windows-based systems and, so, is not available in a form that would allow the entire MOVES model to run on the EPA desktop platform. MySQL is written in C++ and, so, is reasonably platform-independent. Executable versions are also available for a variety of Windows and Unix platforms. While lacking some high-end features, run time performance has been given major consideration in MySQL, so it might be able to compensate for the relatively slow performance

of Java programs. MySQL, therefore, appears to meet all of our relevant software requirements and is the tool we have selected.

8.3. Distributed Processing

Distributed processing is a method used to speed model execution by dividing up a lengthy calculation and distributing separate portions of it to multiple computers or computing systems. The multiple separate results are then collected together into a final, collated result. In MOVES, a run for a single time and location is expected to execute quickly. There are, however, important use cases for which thousands of time-location runs are required. Examples include generating national annual inventories at the county level, preparing inputs for gridded air quality models, and assessing uncertainty using Monte Carlo methods. Such lengthy calculations take weeks to run using MOBILE and NONROAD, and are anticipated to take similarly long periods using MOVES. Fortunately, the calculation for any time-location combination is completely independent of all the other time-location combinations, making it straightforward to divide these calculations between many computers.

Two methods under consideration for distributed processing in MOVES are *grid computing* and *Beowulf clusters*. Grid computing requires only a network of computers with access to a shared hard drive. An example would be a typical office network with many desktop computers with access to a common folder on a hard drive on a server computer. A Beowulf cluster requires Beowulf software and a dedicated group of machines that are all connected together on their own network. There is a master machine which distributes software and data to worker machines. The workers run the calculations and return the data to the master. Much software to run Beowulf clusters is free, and many organizations have set them up to handle various intensive computing tasks.

Because both of these methods for distributed processing are widely used, we plan to program MOVES to be able to use either one or both of them together. Our general scheme is for there to be a Master MOVES, which would reside on analysts' desktop machines and would set up and initiate the run, collect and collate the results, and produce final output. The Master would also divide up the calculation and distribute it to Worker MOVES via a shared hard drive on a server. The workers would perform the calculations and return the results to the shared drive. This basic scenario would work with both the grid computing and Beowulf clusters, although small differences would exist between the Beowulf workers and the grid computing workers.

References

1 National Research Council. 2000. *Modeling Mobile Source Emissions*. National Academy Press; Washington, D.C.

2 U.S. EPA. April 2001. *EPA's New Generation Mobile Source Emissions Model: Initial*

Proposal and Issues. EPA420-R-01-007. Office of Air and Radiation, Ann Arbor, MI.

³ Houghton, J.T.; Meira Filho, L.G.; Lim, B.; Treanton, K.; Mamaty, I.; Bonduki, Y.; Griggs, D.J.; Callendar, B.A.; editors. 1996. *Revised 1996 IPCC Guidelines for National Greenhouse Gas Inventories Reference Manual*. Intergovernmental Panel on Climate Change (<http://www.ipcc.ch/pub/guide.htm>).

⁴ Barth, M.; An, F.; Younglove, T.; Scora, G.; Levine, C. April 2000. *Development of a Comprehensive Modal Emissions Model: Final Report*. Prepared for National Cooperative Highway Research Program under NCHRP Project 25-11.

⁵ Cullen, Alison C.; Frey, H. Christopher. 1999. *Probabilistic Techniques in Exposure Assessment. A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*. Plenum Press, New York.

⁶ Bachman, William H. August 1998. *A GIS-Based Modal Model of Automobile Exhaust Emissions*. EPA-600/R-98-097. Prepared by School of Civil and Environmental Engineering, Center for Geographic Information Systems, Georgia Institute of Technology, Atlanta, Georgia 30332, under EPA Cooperative Agreement CR823020.

⁷ U.S. EPA. September, 1995. *User's Guide to CAL3QHC Version 2.0: A Modeling Methodology for Predicting Pollutant Concentrations Near Roadway Intersections*. EPA-454/R-92-006. Office of Air and Radiation, Research Triangle Park, NC.

⁸ Jimenez-Palacios, J. February 1999. *Understanding and Quantifying Motor Vehicle Emissions and Vehicle Specific Power with TILDAS Remote Sensing*, MIT Doctoral Thesis.

⁹ Watson, H.C.; Milkins, E.E.; Preston, M.O.; Chittleborough, C.; Alimoradian, B. 1983. *Predicting Fuel Consumption and Emissions—Transferring Chassis Dynamometer Results to Real World Driving Conditions*, SAE Technical Paper Series, 830435.

¹⁰ U.S. EPA. May 1993. *Federal Test Procedure Review Project*. EPA420-R-93-007. Office of Air and Radiation.

¹¹ Carlson, T.R.; Austin, T.C. October 1996. *Development of Speed Correction Cycles*. Prepared for EPA by Sierra Research, Inc.

¹² Frey, H. Christopher; Unal, Alper; Chen, Jianjun; Li, Song; and Xuan, Chaoting. August 31, 2002. *Methodology for Developing Modal Emission Rates for EPA's Multi-scale Motor Vehicle & equipment Emission System*. Prepared for OTAQ, U.S. EPA by Department of Civil Engineering, North Carolina State University, Raleigh, NC.

¹³ Beck, Kent 2000. *Extreme Programming Explained*. Addison-Wesley, New York.

Appendix A to Draft Design and Implementation Plan for EPA’s Multi-Scale Motor Vehicle and Equipment Emission System (MOVES)

EPA MOVES RunSpec GUI Details

September 23, 2002
Revision 3

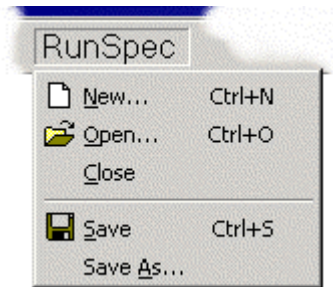
Cimulus, Inc.
US EPA

1. Introduction

This document describes the MOVES Design Team’s detailed design of the GUI used to setup a MOVES Run Specification (“RunSpec”). Like other aspects of this model, the GUI design is iterative and likely to change as the model is developed.

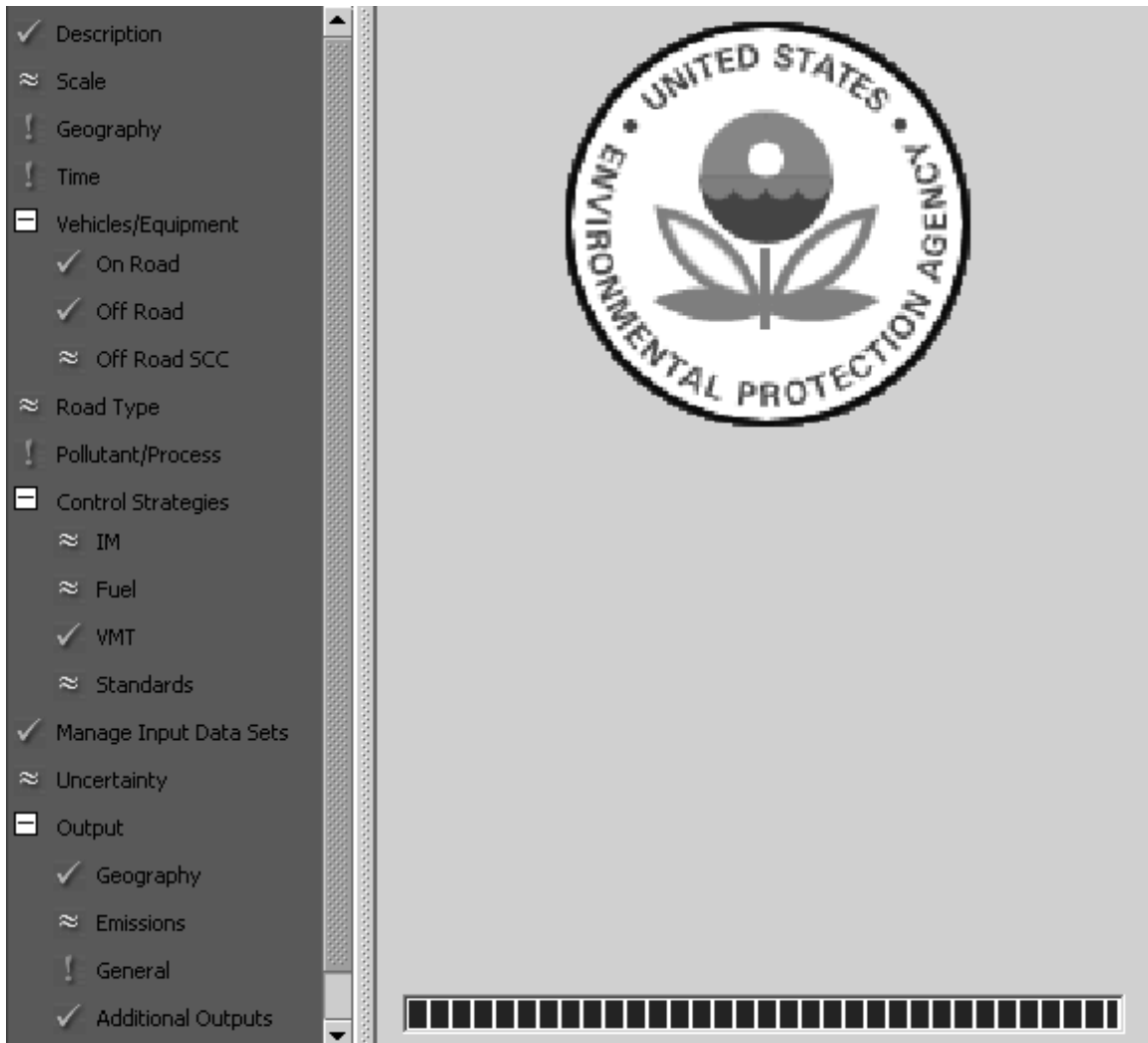
1. Introduction	1
2. RunSpec UI Overview	2
3. Description	3
4. Scale	4
5. Macroscale Geographic Bounds	4
6. Mesoscale Geographic Bounds.....	5
7. Time Spans	6
8. On Road Vehicle Equipment	7
9. Off Road Vehicle Equipment	7
10. Off Road SCC Vehicle Equipment	8
11. Road Type	8
12. Pollutants and Processes	9
13. Internal Control Strategies	11
14. Manage Input Data Sets.....	12
15. Uncertainty	12
16. Geographic Output Detail	13
17. Output Emissions Breakdown	14
18. General Output	15
19. Additional Outputs.....	16
20. Other UI Notes.....	16
21. Revision History	16

2. RunSpec UI Overview



There will exist a top-level menu that will contain a section for the RunSpec. The options will be typical of those used when manipulating documents: New, Open, Close, Save, and Save As. These will allow RunSpec objects to be created, loaded from disk, and saved to disk.





When editing a RunSpec object, the RunSpec UI will appear, typified by the display shown below. It is likely that this UI will be shown below the top-level menus just as documents are shown in other standard programs such as Excel, Word, and WordPerfect.



The UI shown above is conceptually split into two halves: the navigation list (in green on the left) and the detail panel on the right. Selecting an item from the navigation list will place that item's detailed UI into the right hand side. The detail panels are shown in the following sections. For example, the above graphic depicts the selection of the "General" subsection within the "Output" section.

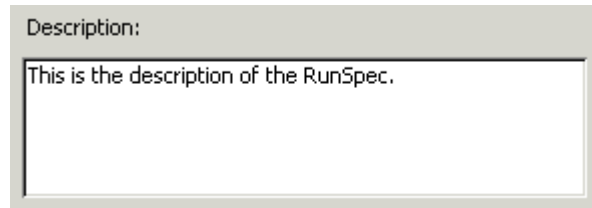
The RunSpec navigation list depicts a tree-like structure of areas of the RunSpec's information. Some sections, such as "Control Strategies" and "Output", contain subsections. These sections are shown with an icon (☐) that allows the list of subsections to expanded or collapsed.

RunSpec navigation list items are shown with an icon that indicates the completeness of the RunSpec in that section, as shown in the table below:

Icon	Meaning
	Needs additional user supplied data.
	Filled in sufficiently to run.
	Default data present, but otherwise filled in sufficiently to run.
	Tree close/expand

Note the icons shown on the sample UI are not necessarily indicative of which sections/subsections will have default data available.

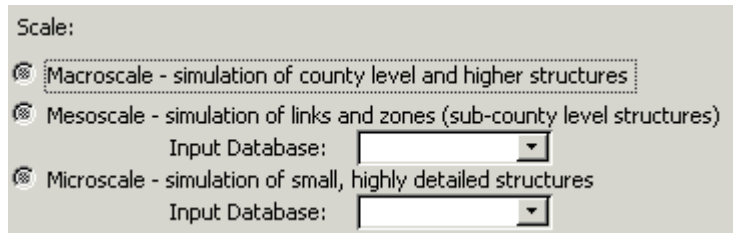
3. Description



Every RunSpec can have a description, which is a free form block of user-supplied text. This text can be up to 5000 characters in length.

The default description is blank and the model can be run without a description.

4. Scale



Scale:

Macroscale - simulation of county level and higher structures

Mesoscale - simulation of links and zones (sub-county level structures)
Input Database:

Microscale - simulation of small, highly detailed structures
Input Database:

The user must choose a scale at which the model will operate. In fact, the selection of scale determines which geographic panel will be presented and may affect availability of other panels as well.

When the user selects “Mesoscale”, the “Input Database” controls for it are enabled. In order to use mesoscale, the user must provide a database with the links and zones that they may want to use. The geography panel(s) will let the user select the subset of those in the database, but the database must be provided.

Microscale operates similarly to Mesoscale with regard to the need for the user to supply input data that, in turn, provides bounds for geography (and perhaps time span) selections.

There is no default scale and the model cannot be run without a scale selection.

5. Macroscale Geographic Bounds



Region: States: Counties: Selections:

Nation

State

County

Select All Add Select All Add Delete

When using the model at the Macroscale level, the geographic selection must yield a list of county-state combinations. This can be quickly accomplished by selecting “Nation”, in which case the additional controls shown on this screen, including the Selections, will be hidden.

Selecting “State” will present the user with the list of states but will hide the Counties controls. The user can then highlight and Add individual states one at a time or can select all States at once.

Selecting “County” will present the user with the list of states (without its “Select All” and “Add” buttons) as well as the list of Counties (with its “Select All” and “Add” buttons). The user can then select a State to see the list of Counties within and Add them individually or in large groups using the County’s Select All button.

There is no default selection for this panel and the model cannot be run without a selection on this panel.

6. Mesoscale Geographic Bounds

The screenshot shows a software interface for selecting geographic bounds at the mesoscale level. It features a 'Region:' section with two radio buttons: 'County' (unselected) and 'Zone/Link' (selected). Below this are five main columns: 'States:', 'Counties:', 'Zone:', 'Zone Links:', and 'Selections:'. Each of the first four columns contains an empty rectangular box. Below the 'States' and 'Counties' boxes are two buttons: 'Add' and 'Select All'. Below the 'Zone' and 'Zone Links' boxes are two buttons: 'Add' and 'Select All'. Below the 'Zone Links' box is a section labeled 'County Links:' with an empty rectangular box and two buttons: 'Select All' and 'Add'. At the bottom right of the interface is a 'Delete' button.

When operating the model at the Mesoscale level, the geographic selection must yield a list of links and zones along with their respective counties. This can be accomplished quickly by selecting “County”, in which case the Zone, Zone Links, and County Links controls will be hidden and the user can Add all the zones and links within individual counties. Note that lack of a “Select All” button for Counties because it is believed that Mesoscale runs will not be performed over all zones and links within a State.

Selecting “Zone/Link” will show the controls as depicted above, except for the “Add” Counties button, and allow the user full control over which links and zones will be included. Note that both Zones and Links are contained within Counties but not all Links are contained within Zones. This necessitates the two list approach shown above in which the user can select a county then links within the county without first selecting a zone.

The list of states, counties, zones, and links will be drawn from the user’s data provided when they selected the scale.

There is no default selection for this panel and the model cannot be run without a selection on this panel.

7. Time Spans

The screenshot shows a graphical user interface for defining time spans. It features a 'Start' section with dropdown menus for Year, Month, Day, Hour, and Time Zone. Below this is a 'Duration' section with two input fields and a dropdown menu. To the right is a 'Selections:' list box containing the text '2002-Jan-1 00:00,Local,3 days'. At the bottom are three buttons: 'Add', '<<', and 'Delete'.

A single RunSpec can specify one or more time spans to be modeled. Each time span has a starting time and a duration. No two time spans can overlap in any fashion.

The starting time of each individual time span can be absolute (that is, relative to GMT) or determined from the local time of the location it is applied to. The “Local” time zone is the default time zone. The start time’s resolution is hourly for Macroscale and Mesoscale. That is, the start time is specified as the start of a single hour on the clock. Thus, a start time of 2002/Jan/01 15:54 is not possible. Microscale allows start times at ¼ hour increments and would list the hours as “00:00”, “00:15”, “00:30”, etc. rather than display yet another dropdown list.

The duration of each individual time span is set as an integral number of hours, days, months, or years from the start. Thus, a duration of 0.5 years is not possible but 6 months is allowed. Internally, the model will choose a temporal looping method that is compatible with the durations and output detail selected. Additionally, Microscale models will be allowed to choose a duration of ¼ hour increments.

The “Add” button will give an error message if an attempt is made to add a time span that overlaps the existing selections.

In the above panel, the user sets the start and duration then uses the “Add” button. Selecting an existing time span in the “Selections” list enables the “<<” and “Delete” buttons but does not otherwise affect the Start and Duration controls.

The “<<” button populates the Start and Duration controls with a copy of the information contained in the current selection. Subsequently changing the Start/Duration controls then using Add does not, however, modify the existing selection. This feature is intended to be used by those wishing to model a single month within several years so that they may vary the year with each “Add” operation.

There is no default selection for this panel and the model cannot be run without a selection on this panel.

8. On Road Vehicle Equipment

The screenshot shows a software interface for selecting on-road vehicle equipment. It features three main columns: 'Fuels:', 'Vehicle Category & Type', and 'Selections:'. The 'Fuels:' column has a 'Select All' button below it. The 'Vehicle Category & Type' column is divided into two sub-columns: 'Category:' and 'MOVES Type:', each with its own 'Select All' button. The 'Selections:' column has a 'Delete' button below it. At the bottom of the interface is a wide button labeled 'Add Fuel/Vehicle Category & Type Combinations'.

This panel allows the user to select which, if any, on road vehicles are to be modeled. The user must select from two distinct dimensions (“Fuels” and “Vehicle Category & Type”) then choose to model their cross product.

Within the “Vehicle Category & Type” dimension, there is a hierarchy of “Category” and “MOVES Type”. Choosing the cross product of a Fuel and a single Category is modeled identically to choosing the cross product of a Fuel and all the MOVES Types within the Category.

The default selection on this panel is to choose everything and the model can be run without a selection on this panel as long as some vehicle type has been selected on one of the Vehicles panels.

9. Off Road Vehicle Equipment

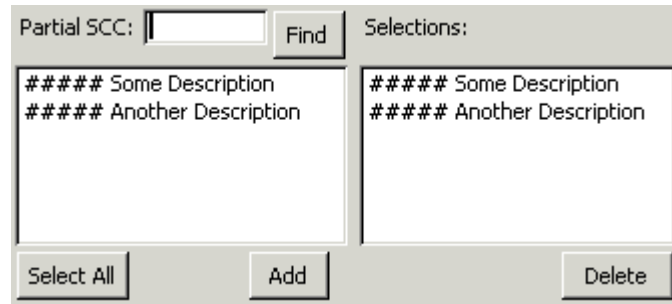
The screenshot shows a software interface for selecting off-road vehicle equipment. It features three main columns: 'Fuels:', 'Segments:', and 'Selections:'. The 'Fuels:' column has a 'Select All' button below it. The 'Segments:' column has a 'Select All' button below it. The 'Selections:' column has a 'Delete' button below it. At the bottom of the interface is a wide button labeled 'Add Fuel/Segment Combinations'.

This panel allows the user to select which, if any, off road vehicles are to be modeled. The user must select from two distinct dimensions (“Fuels” and “Segments”) then choose to model their cross product.

It is expected that the selections on this panel will overlap somewhat with those on the “Off Road SCC Vehicle Equipment” panel (see next section). Internally, the model will combine the selections on this panel and the other vehicle panels to arrive at the unique set of vehicles to be modeled.

The default selection on this panel is to choose nothing and the model can be run without a selection on this panel as long as some vehicle type has been selected on one of the Vehicles panels.

10. Off Road SCC Vehicle Equipment



This panel allows the user to select which, if any, off road vehicles are to be modeled. The user must input a partially or fully complete SCC code for off road sources then select one or more fully complete SCC codes.

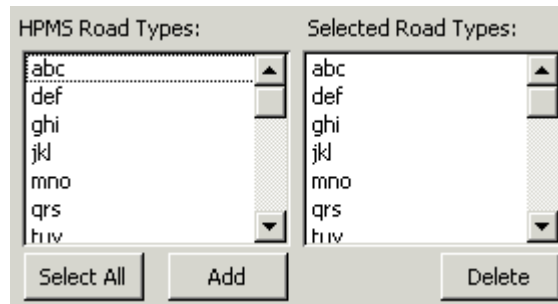
Due to the extensive nature of the full SCC listing, the list of SCC codes will only show those matching the user's partial search and will not initially be populated with the entire SCC code list.

Both of the displayed SCC code lists will show the SCC code and its description.

It is expected that the selections on this panel will overlap somewhat with those on the "Off Road Vehicle Equipment" panel (see previous section). Internally, the model will combine the selections on this panel and the other vehicle panels to arrive at the unique set of vehicles to be modeled.

The default selection on this panel is to choose nothing and the model can be run without a selection on this panel as long as some vehicle type has been selected on one of the Vehicles panels.

11. Road Type



This panel is available if the user has selected any on road vehicles to be modeled. All controls are disabled otherwise.

The list of HPMS Road Types is taken from a master list of roadway types in the database and is **not** filtered to only those roadway types actually present in the geographical range selected.

The default selection on this panel is to choose all roadway types and the model cannot be run without a selection on this panel if there are on road vehicle types selected.

12. Pollutants and Processes

	Exhaust ▼	Evaporative ▼	Brake Wear ▼	Tire Wear ▼
	Running ▼	Starting ▼	Running ▼	Resting ▼
Ozone >				
HC >	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
CO >	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
NOx >	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
Greenhouse >				
CO2 >	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
CH4 >	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>
Carbon Equiv. >	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
Aerosols >				
Tire Particles >				<input checked="" type="checkbox"/>
Brake Particles >			<input checked="" type="checkbox"/>	
Lead Particles >	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Elemental Carbon >	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
Organic Carbon >	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
Sulphate >	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
Gaseous SO2 >	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
Gaseous NH3 >	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
Toxics >				
Benzene >	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Formaldehyde >	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

HC:

PM Size: microns

Note that this panel is illustrative, rather than comprehensive. As MOVES is developed, the exact choice of pollutants and processes may change.

This panel, while complex in general appearance, is used denote which processes and pollutants the user wishes to have calculated. It shows the relationship between process and pollutant and allows the user to perform single selections (via) as well as group selections (via and)

When any particulate aerosol is selected, the PM size controls will be enabled. The user can select 2.5 micron, 10 micron, or "Other" and then provide a custom micron size. The PM size selected applies to all particulates selected.

When HC is selected, the HC controls will be enabled. The user must select a method to express hydrocarbons as from the list: NMHC, NMOG, THC, VOC, ROG, and TOG.

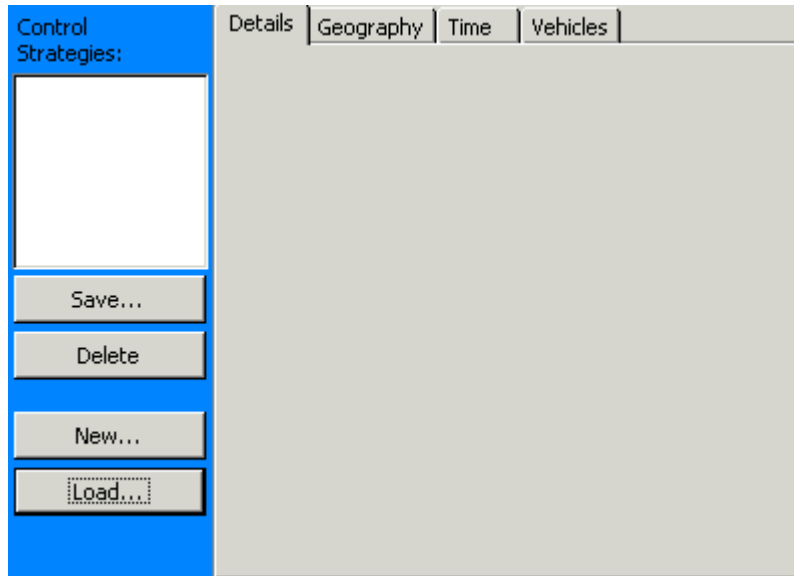
When selecting an entire row (➤) or an entire column (▼), the likely algorithm to be implemented will examine the row or column and determine if there is a mixture of selected (☑) and unselected items (☐).

- If there is no mixture, then all checkboxes in the row/column will be inverted (selecting those that were unselected and unselecting those that were selected). This allows entire rows/columns to be toggled quickly.
- If there is a mixture, then unselect all selections outside of the row/column. This allows intersections of rows/columns to be made quickly.

Thus, clicking “Exhaust ▼” will select the entire columns for “Running” and “Starting”. Then, selecting “Greenhouse ➤” will reduce the selection to only the intersection of Exhaust and Greenhouse.

Note that the above display is only meant to illustrate the desired UI principle. It is not meant to be scientifically accurate or complete. It is also expected that the grids will scroll both vertically and horizontally to accommodate larger listings of processes and pollutants.

13. Internal Control Strategies



“Internal Control Strategies” are actually objects stored within the RunSpec. As objects, they have individual values for member variables and belong to a class.

The finite set of classes supported by the model are listed on the RunSpec navigation list. The user picks one of these classes and is then presented with the above panel which shows the Control Strategy objects of that class that are actually loaded into the current RunSpec.

When the user selects an object, a class-specific UI is shown in the “Details” tab with object-specific values. Object-specific values are also shown in the tabs “Geography”, “Time”, and “Vehicles” which are common to all classes of Control Strategies. The UI for these common tabs is currently unspecified but should take the form of equivalent panels elsewhere in this GUI.

The user can save individual objects to disk with the Save button.

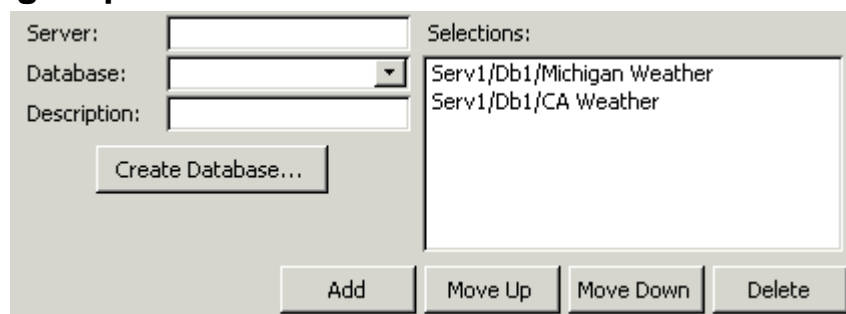
The user can remove an object from the RunSpec using the Delete button.

The user can instantiate an object of the current Control Strategy class using the New button.

The user can load a previously saved Control Strategy object of any supported class by using the Load button.

By using the Save and Load features, users can create control strategies that are used by many RunSpecs but without the risk of changing data for use in one RunSpec and having an unexpected change occur in another, as would happen if the objects were external to the RunSpec.

14. Manage Input Data Sets



This panel is used to specify additional databases to be read by the model during execution. These databases must adhere to the MOVES schema and use the DBMS used by MOVES.

Users can create a new database on an existing local or remote server by using the “Create Database” button.

The data in these databases will “overlay”, that is augment or fully/partially replace, data from the MOVES default data for the duration of the model run. Thus, the order in which these databases are applied to the default data is important. The “Move Up” and “Move Down” buttons are used to modify the order in which a particular database selection is used. Databases at the bottom of the list are applied last.

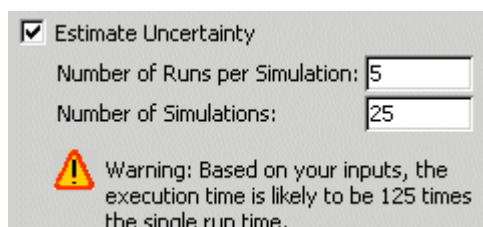
Technically, the input databases on the Scale panel are input data sets, but they are required for user input whereas contents of databases listed on this panel are not read by the UI but only during model execution.

The input databases on the Scale panel are always read first followed by the selections on this panel.

The “Add” button validates that the combination of server and database is unique within the selections.

There are no default selections for this panel and the model can be run without any selections on this panel.

15. Uncertainty



This panel is only an example of how we might present an iterative, Monte Carlo-type method of estimating uncertainty. We may also implement alternative, non-numerical methods of error propagation.

The “Estimate Uncertainty” checkbox defaults to unchecked/Off. When unchecked, all other controls on this panel are disabled.

A warning icon and wording is shown to the user whenever their selections would run the model more than 1 time. Due to time variances of computing hardware, no time estimate for completion will be given.

The model can be run without enabling uncertainty.

16. Geographic Output Detail



Nation
 State
 County
 Road Type
 Link & Zone

This panel sets the level of geographic detail used for the “Location” dimension of the output data. The number of output records generated is directly related to the level of detail chosen on this panel. For instance, should a user select “County”, then the database will contain an aggregated record for each county (among other output dimensions of course) and information about the portions that Road Types contributed to the county total will not be available. However, it is expected that choosing “County” will still allow reporting to be aggregated to the State and National levels.

For Macroscale operation, the “Link & Zone” option is disabled.

For Macroscale operation, the default on this panel is “County”. For Mesoscale operation, the default is “Link & Zone”. The model cannot be run without a selection on this panel.

17. Output Emissions Breakdown

The screenshot shows a GUI panel for 'Output Emissions Breakdown'. It is organized into several sections:

- Always:** Three checked options: Time, Location, and Pollutant.
- All:** Four options: Model Year (checked), Fuel Type (unchecked), Emission Process (unchecked), and Distinguish Particulates (checked).
- On Road/Off Road:** A checked option.
- On Road:** Four options: Road Type (unchecked), Category (unchecked), MOVES Vehicle Type (checked), and SCC (unchecked).
- Off Road:** Three options: Segment (unchecked), SCC (checked), and HP Class (checked).

This panel allows the user to specify additional dimensions to the output data. The output data will always contain dimensions for time, location, and pollutant. The granularity of the time dimension is set on the “General Output” panel as the “Output Timestep”. The location dimension is set on the “Geographic Output Detail” panel. The pollutant dimension is implicit and is not otherwise set.

The dimensions that apply to both on road and off road sources are Model Year, Fuel Type, and Emission Process. Technically, “Distinguish Particulates” is not a dimension but rather causes the particulates pollutants to be reported as separate pollutants rather than as a single aggregate particulate count. Any combination of the “All” dimensions may be selected.

Should the user decide to distinguish on road sources from off road sources, additional options become available. On road allows Road Type, Category, MOVES Vehicle Type, and SCC to be selected. Off Road allows Segment, SCC, and HP class. Each of these two groups has internal interdependencies between the allowed dimensions:

- If On Road SCC is used, no other On Road option may be used.
- Zero or one of On Road Category and On Road MOVES Vehicle Type may be used. Thus, checking On Road’s Category would disable On Road’s MOVES Vehicle Type and vice versa.
- Zero or one of Off Road Segment and Off Road SCC may be used.
- Off Road HP Class is independent of the other Off Road options.
- If any On Road or Off Road option is selected, the On Road/Off Road option is automatically implied.
- However, selecting the On Road/Off Road option does not imply that any On Road or Off Road options must be selected.

The default on this panel is to use the implied “Time”, “Location” and “Pollutant” dimensions. The model can be run without user selections on this panel since there are always implied output dimensions.

18. General Output

The screenshot shows the 'General Output' panel. It has a title bar 'Output Database'. Below it are two dropdown menus: 'Server:' and 'Database:'. A warning icon (yellow triangle with exclamation mark) is next to the text 'Data is already in this database.' To the right of this text is a button labeled 'Create Database...'. Below this is another dropdown menu labeled 'Output Timestep:' with 'Month' selected. The main section is titled 'Output' and contains three checkboxes: 'Time Factors (mass/time)', 'Distance Factors (mass/distance)', and 'Emissions (mass/time step)'. The 'Emissions' checkbox is checked. Below these are three dropdown menus: 'Time Units:', 'Distance Units:', and 'Mass Units:'.

This panel is conceptually the final panel completed prior to running the model.

The user must specify the server and database into which the results should be placed. There is a local default server shown but no default database selected. A warning is shown if the database already contains rows within the output tables. The “Create Database” button is used for creating new databases and is likely identical to that used on the Manage Input Data Sets panel.

The Output Timestep is the “time” dimension of the output data. The setting for this value directly affects the number of output rows in the database. The options available for the Output Timestep are potentially: Hour, Day, Month, Season, and Year. Note that Season is intentionally available here but not in the Time Span panel. The UI will examine all of the selected time spans and find the shortest duration. That duration and those shorter will be shown in this panel. Thus, if the user selected two time spans both with a monthly duration, then the list shown would be Month, Day, and Hour. If the user selected two time spans, one with a yearly duration and one with a daily duration, the list shown would only be Day and Hour.

There is no default Output Timestep.

Various factors and values are available for each pollutant. The user can elect to have emission rates as a function of time output for each pollutant. The user can also elect to output emission rates as a function of distance (i.e. VMT related) output for each pollutant. Finally, the user can elect to simply get the total mass of pollutant output within each Output Timestep.

Choosing any of the Output factors or values requires the user to select appropriate units for their output. The units lists will be disabled if not required by the user’s selections. Both common English and Metric units are available, thus making output in Tons per Kilometer possible (but not advisable) along with the more common Grams per Mile. Since it is only a rate, the Time Units list will include Year, Month, Week, Day, Hour, and Second regardless of the Output Timestep or time span durations.

The user must select at least one of the Output factors or values, with the Emissions (mass/time step) option being selected by default.

The user must make appropriate selections in all areas of this panel, as the model cannot run without this information.

19. Additional Outputs



In addition to outputs for pollutants, MOVES can provide internal data used in its calculations. This panel is used to select which of those internal data sets should be provided. A wide variety of additional outputs could be included on this panel, for example, VMT, source populations, future year source use type distributions, etc.

20. Other UI Notes

Throughout the UI, list boxes are used extensively. These list boxes are to support a feature that allows the user to give focus to the list box then start typing text with the list box automatically moving/scrolling the displayed items to that which matches the user's typed text. The textual matching will not be case sensitive and the internally accumulated text will be reset upon a non-typing action (i.e. mouse usage or loss of input focus). This feature is used in lieu of providing an explicit "Search" feature on the various panels.

Buttons are used along with list boxes on most panels of the UI, typified by the "Select All", "Delete", and "Add" buttons. It is expected that these buttons will be enabled/disabled based upon selection in their associated list box. For example, a "Delete" button will be disabled until a selection is made in its list box.

Prior to starting a model run that involves uncertainty, the user will be presented with a popup window that again warns the user of the potentially extremely long execution time. This will be a restatement of the warning on the Uncertainty panel. The user will have the option of continuing and starting the run or to abort the run. Should they abort the run, the UI will navigate the user to the Uncertainty panel.

Though not shown here, it is expected that during a model run, the user will be shown a progress indicator and be allowed to pause, resume, and cancel the run. It is also likely the user will require the ability to halt a run at a checkpoint then resume it from there at a later date and even with a different mix of distributed computers contributing to the execution.

21. Revision History

Changes from Revision 2 to Revision 3:

- In the "RunSpec UI Overview" section, updated the main screen image
 - Added the "Additional Outputs" item under "Output"
 - Added the scroll bar to the navigation panel
 - Replaced the default screen with the EPA logo and the progress bar
- In the "Macroscale Geographic Bounds" section, replaced "50 States", "49 States", and "47 States" with "Nation"
- In the "Pollutants and Processes" section, updated the image with a transposed grid and revised wording to reflect this
- Added the "Additional Outputs" section

Changes from initial version to Revision 2:

- Added descriptions to every screen.
- Changes throughout the user interface, impacting nearly every screen, per MOVES Design Team meetings.
- Added section "19. Other UI Notes"

**Appendix B to Draft Design and Implementation
Plan for EPA's Multi-Scale Motor Vehicle and
Equipment Emission System (MOVES)**

**MOVES Output
Database Design**

August 29, 2002
ASD, EPA

Table of Contents

1. 'Entity' section	3
Entity Table	3
2. 'Column' section	4
Column Table	4
Table(s) of "distanceUnits" Column	7
Table(s) of "emissionMass" Column	7
Table(s) of "emissionMassPerDistance" Column	7
Table(s) of "emissionMassPerTime" Column	7
Table(s) of "fuelTypeID" Column	7
Table(s) of "HCExpressAs" Column	8
Table(s) of "HPClass" Column	8
Table(s) of "linkID" Column	8
Table(s) of "locationID" Column	8
Table(s) of "massUnits" Column	8
Table(s) of "modelYear" Column	8
Table(s) of "monteCarloRunNumber" Column	8
Table(s) of "monteCarloSimulationNumber" Column	8
Table(s) of "MOVESOutputRowID" Column	9
Table(s) of "MOVESRunID" Column	9
Table(s) of "MOVESRunID" Column	9
Table(s) of "onOrOffRoad" Column	9
Table(s) of "outputTime" Column	9
Table(s) of "pollutantID" Column	9
Table(s) of "processID" Column	9
Table(s) of "roadTypeID" Column	9
Table(s) of "runDateTime" Column	9
Table(s) of "runSpecFileName" Column	10
Table(s) of "scale" Column	10
Table(s) of "SCC" Column	10
Table(s) of "segment" Column	10
Table(s) of "sourceCategoryID" Column	10
Table(s) of "sourceTypeID" Column	10
Table(s) of "stateID" Column	10
Table(s) of "timeStepHours" Column	10
Table(s) of "timeUnits" Column	11
Table(s) of "zoneID" Column	11

Name	Datatype	Null Option	Comment	Is PK	Is FK
			in the case where the user selected "Nation" as the geographic granularity for the output. locationID will be default otherwise if "State" level granularity was selected.		
massUnits	VARCHAR(5)	NULL	massUnits is the standard textual abbreviation for the max unit. Ex: "g", "kg", "ton"	No	No
modelYear	SMALLINT(4)	NOT NULL		No	No
monteCarloRunNumber	SMALLINT(4)	NOT NULL		No	No
monteCarloSimulationNumber	SMALLINT(4)	NOT NULL		No	No
MOVESOutputRowID	INTEGER	NOT NULL	Unique number automatically sequentially assigned to each MOVESOutput instance as it is entered into this database. Number has no other significance.	Yes	No
MOVESRunID	INTEGER	NOT NULL	Run id, implemented as auto increment integer in MySQL.	Yes	No
MOVESRunID	INTEGER	NOT NULL	Run id, implemented as auto increment integer in MySQL.	No	Yes
onOrOffRoad	CHAR(1)	NOT NULL	onOrOffRoad can be "N" for NONRoad, or "O" ("oh") for OnRoad, and " " (space) for both. The default value is the first one in the enumeration (index = 1)	No	No
outputTime	DATETIME	NOT NULL		No	No
pollutantID	INTEGER	NOT NULL		No	No
processID	INTEGER	NOT NULL		No	No
roadTypeID	INTEGER	NOT NULL	roadTypeID is not redundant with linkID in the cases where	No	No

Name	Datatype	Null Option	Comment	Is PK	Is FK
			the user wants road type as a dimension but does not want geographic detail to the link/zone (or perhaps even to the County) level.		
runDateTime	DATETIME	NOT NULL		No	No
runSpecFileName	VARCHAR(255)	NULL	runSpecFileName can be null if the user has not saved their runspec prior to launching the simulation. scale ENUM("MACRO", "MESO", "MICRO") NOT NULL	No	No
scale	CHAR(5)	NOT NULL		No	No
SCC	CHAR(10)	NOT NULL	SCC holds both OnRoad and OffRoad SCC codes and may be all 0s (zeroes) to represent "all" SCC codes at once.	No	No
segment	TINYINT(2)	NOT NULL		No	No
sourceCategoryID	INTEGER	NOT NULL		No	No
sourceTypeID	INTEGER	NOT NULL		No	No
stateID	INTEGER	NOT NULL	stateID, locationID, zoneID, and linkID can all be default in the case where the user selected "Nation" as the geographic granularity for the output.	No	No
timeStepHours	REAL	NOT NULL	timeStepHours is the number of hours in the selected time step. For example, if the user wanted seasonal timesteps, the hours would be $365.25\text{days/year} * 24\text{hours/day} / 4\text{seasons/year}$.	No	No
timeUnits	VARCHAR(5)	NULL	timeUnits is the standard textual abbreviation for the time unit. Ex: "sec", "hour", "min",	No	No

Table(s) of "HCEXpressAs" ColumnTable(s) of "HCEXpressAs" ColumnTable(s) of "HCEXpressAs" ColumnTable(s) of "HCEXpressAs" ColumnTable(s) of "HCEXpressAs" ColumnTable(s) of "HCEXpressAs" ColumnTable(s) of "HCEXpressAs" Column
Name
MOVESRun

Table(s) of "HPClass" ColumnTable(s) of "HPClass" ColumnTable(s) of "HPClass" ColumnTable(s) of "HPClass" ColumnTable(s) of "HPClass" ColumnTable(s) of "HPClass" ColumnTable(s) of "HPClass" Column
Name
MOVESOutput

Table(s) of "linkID" ColumnTable(s) of "linkID" ColumnTable(s) of "linkID" ColumnTable(s) of "linkID" ColumnTable(s) of "linkID" ColumnTable(s) of "linkID" ColumnTable(s) of "linkID" Column
Name
MOVESOutput

Table(s) of "locationID" ColumnTable(s) of "locationID" ColumnTable(s) of "locationID" ColumnTable(s) of "locationID" ColumnTable(s) of "locationID" ColumnTable(s) of "locationID" ColumnTable(s) of "locationID" Column
Name
MOVESOutput

Table(s) of "massUnits" ColumnTable(s) of "massUnits" ColumnTable(s) of "massUnits" ColumnTable(s) of "massUnits" ColumnTable(s) of "massUnits" ColumnTable(s) of "massUnits" ColumnTable(s) of "massUnits" Column
Name
MOVESRun

Table(s) of "modelYear" ColumnTable(s) of "modelYear" ColumnTable(s) of "modelYear" ColumnTable(s) of "modelYear" ColumnTable(s) of "modelYear" ColumnTable(s) of "modelYear" ColumnTable(s) of "modelYear" Column
Name
MOVESOutput

Table(s) of "monteCarloRunNumber" ColumnTable(s) of "monteCarloRunNumber" ColumnTable(s) of "monteCarloRunNumber" ColumnTable(s) of "monteCarloRunNumber" ColumnTable(s) of "monteCarloRunNumber" ColumnTable(s) of "monteCarloRunNumber" Column
Name
MOVESOutput

Table(s) of "monteCarloSimulationNumber" ColumnTable(s) of "monteCarloSimulationNumber" ColumnTable(s) of "monteCarloSimulationNumber" ColumnTable(s) of "monteCarloSimulationNumber" ColumnTable(s) of "monteCarloSimulationNumber" ColumnTable(s) of "monteCarloSimulationNumber" Column
Name
MOVESOutput

Table(s) of "MOVESOutputRowID" ColumnTable(s) of "MOVESOutputRowID" Column2Table(s) of "MOVESOutputRowID" ColumnTable(s) of "MOVESOutputRowID" ColumnTable(s) of "MOVESOutputRowID" ColumnTable(s) of "MOVESOutputRowID" ColumnTable(s) of "MOVESOutputRowID" Column

Name

MOVESOutput

Table(s) of "MOVESRunID" ColumnTable(s) of "MOVESRunID" ColumnTable(s) of "MOVESRunID" ColumnTable(s) of "MOVESRunID" ColumnTable(s) of "MOVESRunID" ColumnTable(s) of "MOVESRunID" ColumnTable(s) of "MOVESRunID" Column

Name

MOVESRun

Table(s) of "MOVESRunID" ColumnTable(s) of "MOVESRunID" ColumnTable(s) of "MOVESRunID" ColumnTable(s) of "MOVESRunID" ColumnTable(s) of "MOVESRunID" ColumnTable(s) of "MOVESRunID" ColumnTable(s) of "MOVESRunID" Column

Name

MOVESOutput

Table(s) of "onOrOffRoad" ColumnTable(s) of "onOrOffRoad" ColumnTable(s) of "onOrOffRoad" ColumnTable(s) of "onOrOffRoad" ColumnTable(s) of "onOrOffRoad" ColumnTable(s) of "onOrOffRoad" Column

Name

MOVESOutput

Table(s) of "outputTime" ColumnTable(s) of "outputTime" ColumnTable(s) of "outputTime" ColumnTable(s) of "outputTime" ColumnTable(s) of "outputTime" ColumnTable(s) of "outputTime" ColumnTable(s) of "outputTime" Column

Name

MOVESOutput

Table(s) of "pollutantID" ColumnTable(s) of "pollutantID" ColumnTable(s) of "pollutantID" ColumnTable(s) of "pollutantID" ColumnTable(s) of "pollutantID" ColumnTable(s) of "pollutantID" ColumnTable(s) of "pollutantID" Column

Name

MOVESOutput

Table(s) of "processID" ColumnTable(s) of "processID" ColumnTable(s) of "processID" ColumnTable(s) of "processID" ColumnTable(s) of "processID" ColumnTable(s) of "processID" ColumnTable(s) of "processID" Column

Name

MOVESOutput

Table(s) of "roadTypeID" ColumnTable(s) of "roadTypeID" ColumnTable(s) of "roadTypeID" ColumnTable(s) of "roadTypeID" ColumnTable(s) of "roadTypeID" ColumnTable(s) of "roadTypeID" ColumnTable(s) of "roadTypeID" Column

Name

MOVESOutput

Table(s) of "runDateTime" ColumnTable(s) of "runDateTime" ColumnTable(s) of "runDateTime" ColumnTable(s) of "runDateTime" ColumnTable(s) of "runDateTime" ColumnTable(s) of "runDateTime" ColumnTable(s) of "runDateTime" Column

Table(s) of "timeUnits" ColumnTable(s) of "timeUnits" ColumnTable(s) of "timeUnits" ColumnTable(s) of "timeUnits" ColumnTable(s) of "timeUnits" ColumnTable(s) of "timeUnits" ColumnTable(s) of "timeUnits" ColumnTable(s) of "timeUnits" Column
Name
MOVESRun

Table(s) of "zoneID" ColumnTable(s) of "zoneID" ColumnTable(s) of "zoneID" ColumnTable(s) of "zoneID" ColumnTable(s) of "zoneID" ColumnTable(s) of "zoneID" ColumnTable(s) of "zoneID" ColumnTable(s) of "zoneID" ColumnTable(s) of "zoneID" Column
Name
MOVESOutput