August 8, 2003

REQUEST FOR PROPOSAL

Development of a New (Open) Emissions Model

The Lake Michigan Air Directors Consortium is seeking contractor assistance to develop a new (open) emissions model. Emission models, such as SMOKE and EMS-2003, are used to process emissions inventory data to provide the necessary input files for air quality models. A new construct for emissions modeling is proposed, which relies on the strength of the existing models, takes advantage of a public domain relational database management system, and meets the needs of air quality modelers. This new model will be used to support the development of State Implementation Plans for fine particles, regional haze, and ozone.

You are invited to submit a proposal to develop the open emissions model. Proposals must be received no later than 5 p.m. CDT on Friday, September 5, 2003. Ten (10) paper copies and one electronic copy (WordPerfect or Word) of the proposal are required and should be sent to:

Mr. Michael Koerber
Executive Director
Lake Michigan Air Directors Consortium
2250 East Devon Avenue, Suite 250
Des Plaines, IL 60018

e-mail: ladco@ladco.org

No late proposals will be accepted, and the offer shall remain effective for a period of 60 days from the date of the mailing.

Your response to this Request for Proposal (RFP) should include a complete technical proposal to describe your approach to accomplishing the tasks outlined below in the Statement of Work. The response should include a summary of your organization's capability and experience in the field of work. Please limit the technical proposal to 15 pages. In addition, please provide a complete cost estimate that sets forth a detailed breakdown of projected expenditures by task, including person hours by labor category, travel, and other direct charges.

All contracts will be issued by the Lake Michigan Air Directors Consortium and managed by the Consortium's Executive Director. It is anticipated that the Consortium will award a cost plus fixed fee type contract as a result of this solicitation. The Consortium may consider awarding another type of contract, if consistent with the objectives of this project and the interests of the Consortium

Funds available for this contract are federal funds from the U.S. Environmental Protection Agency and contractors must meet requirements associated with the use of federal funds. All information, software, documentation, and data delivered under this contract will be in the public domain.

All inquiries regarding this RFP should be directed to Michael Koerber either in writing at the above address or via e-mail at ladco@ladco.org. Written responses to inquiries will be sent to all interested bidders.

Statement of Work

The purpose of this project is to develop a new (open) emissions model that will be used by state and local agencies to support the development of State Implementation Plans for fine particles, regional haze, and ozone. A "white paper" on the new model is available at http://64.27.125.175/tech/emis/misc/OPEM.pdf. The white paper discusses how to assess the usefulness of emissions modeling systems by highlighting the following characteristics that every system should have:

> (1) *Quality Contro*l: Inventory quality is highly variable and often of low quality for important parts of the modeling process. The model must be able to identify critical and non-critical errors in the data inputs and report them in an effective way.
>
> (2) *Transparency:* In the near future, there may be models for point, area, on-road mobile, off-road mobile, biogenic, electric utility, livestock, soil, and fires. It will be important for users to easily understand the processors and data sets, so that model operation can be easily deconstructed when problems occur.
>
> (3) *Performance:* Emissions models are one part of a larger modeling construct, and should not require significant additional hardware and software.
>
> (4) *Public Domain*: A good emissions model should be written in such a way that community enhancement and improvement should be straightforward, and not encumbered by any licensing or other proprietary limitations.

In terms of quality control, the most important feature for an emissions model is to quickly and efficiently diagnose problems with the inventories and report them in a concise way. In the past, emissions models identified thousands of potential errors, far more than inventory developers were capable of fixing or even reading. Often the implication of these errors were unclear and unimportant to the overall modeling process. Moreover, important problems were often overlooked, buried in thousands of lines of output. Quality assurance tools must, therefore, be built to identify significant problems clearly including the implication of found errors. Priority should be given to those issues that are expected to affect the modeling process. Your proposal should include some hierarchical error identification and correction method, such as the four-tiered error reports in EMS-2003

Transparency refers to the ability to trace the sources of data and what happens to them during the emission modeling process. This is critical to the success of any emissions model, because it allows the end user to understand the reliability and consistency of the data given the many different calculations and variety of data types. The language used to code an emissions model must be readable and easily learned so that the user can understand how individual pieces of data are handled. Additionally, it must be easy for a user to see the data that the model is using at any step in the process. In order to evaluate output that seems erroneous, users need to be able to run basic queries on data sets to trace the exact equations used for a specific value. For example, the user should be able to reference a data set and request only records for a given state, county, and SCC code. A critical component of your proposal is to include a program similar to SAS fsview that will allow the user to open any SQL data set and do simple queries on the data set.

Data transparency is further improved by a system that makes complex ad-hoc report writing possible by the lay emissions modeler after a few months of experience with the model and its source code. This allows investigations of new data issues as they arise. Processors should also be developed that report emissions estimates (in both tabular and graphical formats) through the different stages of emissions modeling such as input processing, spatial processing, temporal processing, and, finally, speciation processing. These reports should look at state and county totals by major sector of the inventory. Additionally, some method should be devised to store summaries of past modeling runs for easy comparison with current values.

Model performance has two distinct components: software and hardware. Satisfactory software performance requires the model to run on similar platforms as the photochemical models with minimum modifications. The current standard operating system for photochemical transport models is Red Hat LINUX along with the Portland Group Fortran compiler. Your proposal should use only software compatible with the Red Hat LINUX platform and with no additional costs beyond the Fortran compiler

Hardware performance requires that only modest adaptations to the hardware used to run the photochemical transport models need to be made to run the emissions model. Currently, the hardware recommendations for running a photochemical transport model include a dual 2.8+ Ghertz Intel Processors, 2 Gbyte of RAM, 800 Gbytes of IDE hard drive, and 200 Gbytes of SCSI hard drive running REDHAT LINUX version 8.0 with the Portland Group Fortran compiler. Often disk input/output (I/O) is the most limiting factor in emissions modeling. The best way to resolve disk I/O problems now is to use striped RAID SCSI drives. A four-drive striped SCSI disk array costs less than $1,500 and should be the only hardware improvements needed for emissions modeling. Your proposal should specify that the model will be two- to five-times faster than the photochemical model for a given domain. (This might not be possible for the MOBILE6 processors.)

Finally, it is important that your proposal clearly state your commitment that all products of this contract are open source, the model contains no proprietary components, and no part of the final products shall be proprietary. Your proposal should explain your organization’s commitment to delivering a product suitable for continued, contributed improvements by the modeling community following the open source development model.

Model Fundamentals: This project should cover the development of the four base emissions models: point sources, area sources, on-road mobile sources, and biogenics. Additional supporting models include: gridding surrogates model, chemical speciation model, and point and area source growth/control model. In developing a new emissions model, the best attributes of SMOKE, EMS-2003, and EPS should be used. The model should be built with a modular design, so that common activities, like temporalization, speciation, or spatial allocation, can use the same tools, whenever possible, and new modules may be added with relative ease in the future.

The model should be written primarily in Postgres SQL. We expect where pre-existing Fortran code exists, like MOBILE6, Fortran will be used. Other programs/tools can be used as long as those tools are Gnu Public License (GPL) code. Additionally the contractor should try and use code where Linux RPM's are available. If the contractor believes that non-GPL code will be used, then prior written consent should be obtained from the Consortium and included in your proposal. It is important to create a model that is not difficult to install. Your proposal should clearly explain your use of non-SQL programming to accomplish any of the tasks.

The emissions model should follow the RPO data exchange protocol (http://64.27.125.175/tech/emis/protocol/protocol.html) for all data sets input by the model. For datasets not covered by the RPO data exchange protocol, the model developers should consult with the Consortium before setting formats.

The point source model should use NIF3.0 as its native input format. It should use the EMS-2003 temporal processors as a blueprint for how to deal with temporalization of NIF files and the selection of the "best" NIF EM record to reflect emissions for a given day. Additionally, the model should be built so that the temporalization components of the RPO data exchange protocol are used and that the full options provided by that file type are allowed. This includes application of hourly profiles by day-of-week, by month-of-year, or season-of-year temporal profiles. The model should retain three temporal data sets that would be applied hierarchically: a national default set, a national SCC default set, and a state/tribe/RPO-supplied set. Spatial processors should be capable of projecting the following coordinate systems: Lambert, Lat/Long, Polar Stereographic, and UTM. The model should be capable of reading in the day- and hour-specific CEM data types and applying day specific emissions estimates for both point and area sources. This also includes the output of hour specific stack parameters for plume-in-grid and plume rise equations. The model should be capable of outputting to CAMX and CMAQ elevated source files. For CMAQ this would include the ability to run the plume rise calculations and generate 3 dimensional input files. If a user does not require CMAQ output, then IO/API should not be required for model operation.

The area source model should also use NIF3.0 as its native input format and have all the temporal capabilities of the point source model. Area sources will differ from point sources in that a robust spatial processing tool will be required to grid national surrogates for the local grid. An additional tool should be provided that will create sub grids within a larger spatial surrogate set. For example, Subgrid.sas in EMS-2003 can read in a continental scale spatial surrogate file for a 4 km grid cell size and create an eastern US subgrid at 36 km grid cell size based on user inputs. The model should also include spatial processors capable of generating raw spatial surrogates from available ARC/INFO shape files, such as TIGER line files, census coverages, landuse shape files, and state/tribe supplied shape files.

The mobile source model should contain an imbedded version of MOBILE6 (of, if available, USEPA's next generation mobile model). The mobile source model must be capable of managing and applying the temporal aspects of mobile sources at the highest level. This means that the model should be capable of reading in the RPO data exchange protocol file formats and apply all MOBILE6 input variable by road type, link, month, and hour for speed, VMT, vehicle mix, and vehicle starts. The model must be capable of running MOBILE6 in slow full-resolution mode, and in an abbreviated mode (i.e., by averaging certain sub-county effects, as in SMOKE). The spatial processors of the model should read the same surrogate file as the area sources. They should also be capable of reading link end point coordinates and calculate the fraction of link length in each grid cell. The spatial processors should read Lambert, Lat/Long, and UTM coordinate systems, but also State Plane coordinates.

The biogenics model should resemble the BIOME model in EMS-2003, which is a rewrite of BEIS3 in the programming language SAS. We would like BEIS3 rewritten in SQL. While the biogenics model will be written in SQL, it is important that the input and output data sets are transparent, easily accessed, and in the same format as BEIS3. The model should be able to report not only emissions estimates but intermediate values, such as emissions at standard conditions, emission factors, and land use and plant type summaries by grid cell and county. There should be no Fortran code used in the biogenics model.

Spatial surrogate processing for point, area, and mobile sources will need to follow the guidelines set out in the draft document "Using Tribal Inventories" (http://64.27.125.175/tech/emis/protocol/Using Tribal Inventories_v4.pdf) (It is our understanding that the biogenics model will not change.)

Temporal processing should assume the modeling of global scale inventories. This means that the model should be capable of adjusting all emissions estimates for the primary time zone of the modeling system. This temporal allocation will require the programs to calculate emissions for different days and even seasons in much the same way as EMS-2003 handles the NIF files in point and area sources. An example of this is EMS-2003's capability of reaching across different seasonal EM records to locate the most appropriate emissions estimates.

The model should be capable of having multiple chemical mechanisms, including current versions of CB-IV and SAPRC. The model should be capable of building lumped chemical mechanism speciation profiles from the raw speciation profiles by chemical species The speciation model should be capable of matching the different naming requirements for species and provide data sets for output to photochemical models, such as CAMx, CMAQ, and if the user supplies the raw speciation values, REMSAD. It should contain the most current version of USEPA's SPECIATE database and read speciation files from the RPO data exchange protocol. The model should also be written to allow the use of toxics inventories, if users develop appropriate inputs to the speciation model. This will require that no pollutants are "hard-coded" into the model, and that it is designed to allow the user to easily add new pollutants and species.

The growth and control model should be capable of applying growth and control factors supplied in files defined by the RPO data exchange protocol. Unlike the current EMS-2003 and SMOKE growth and control models, we would like the growth and control model to be able to interpolate growth and control between the years when explicit control factors are available. This requires intermediate year control records, for those controls that are phased-in over time. Clarification of the exact interpolation methods can be made in the planning steps of the model development.

A time consuming element of current emissions processing is quality assurance (QA) of new data. The model should contain tunable QA tools. That is, the user should be able to specify the number of tests, quantity of intermediate data retention, and variety of emissions estimates. Once it has been established that the new data is of acceptable quality and enhanced QA and error checking is not necessary, then the user should have the option of turning-off much of the QA to speed up processing and decrease disk space requirements. We propose a scale of 1 to 10, where 1 would have the lowest level of QA and intermediate data retention and 10 would have the highest level of QA and data retention.

Another consideration is the link between the meteorological models, such as MM5, and the emissions model. Meteorological models provide input to both emissions and photochemical models. A strong link between the meteorological and emissions models is desired. For Models3/CMAQ and EMS-2003, the connection is accomplished with external processors. The linking processor here should be transparent, capable of reading the meteorological model's binary output format, and written completely in SQL.

We believe that user interfaces are over-rated as a practical tool for modeling episodes that span several days. The best emissions models have a robust batch processing methodology that allows users to easily configure the model, run the episodes, and control the numerous options available. The new model should have this capability.

There will be several (i.e., 5-10) beta testers available for the model and the contractor should provide some minimal beta tester support for the model components when they are released and should be capable of responding to reasonable comments from the testers.

A comprehensive users guide for the model with a technical formulation section will be needed. The users guide should contain: (1) references to hierarchical selection criterion for default values in the various spatial, speciation, and temporalzation steps; (2) step-by-step instructs on how to set up a new domain, including grid selection, chemistry selection, raw data input, surrogate creation, option selection, and model execution; (3) identification of data sets needed to run the model, and (4) guidelines for quality assurance of emissions data. You should also provide a blank template, so that future members of the user community can add-on to the users guide as new modules and technical documents are added.

Your proposal should include a one-time, five-day training session on how to use the model, with a focus on explaining model operation and basic functions. Training should assume that attendees have moderate experience with emissions modeling. Training should include a comprehensive review of temporal, spatial, and speciation data sets; hands-on installation; and local modeling grid definition and construction. All training materials should also be made freely available, so that sponsors may host the presentation materials on-line. The cost estimates should include costs for developing the training materials and example data sets, and separate cost estimates for holding the training, so that we may ascertain the cost of additional training. Assume three attendees from each RPO and two attendees from USEPA in your cost estimates.

Optional Tasks

In addition to the core modeling system outlined in the Statement of Work, you may wish to include the following additional (optional) tasks in your proposal.

Optional task 1 should include additional tools in the growth and control model to calculate costs for control strategies and report those costs for point and area sources. The cost estimates should be based on user inputs, such as cost per emission unit and cost per ton values. The model should have a wide variety of reports showing costs by state, county, strategy, scenario, and year.

Optional task 2 should include incorporation of USEPA's NONROAD source model in much the same way that the MOBILE6 code is included. The model would read in date and meteorological parameters to calculate hour-specific emissions estimates for the nonroad sector. The model should be programmed to report significant information about vehicle populations, fuel use, and activity factors.

Optional task 3 should be a tool to help users build modeling domains by selecting points on a map graphically and drawing a map of this new domain with various background maps, including counties, states, tribal lands, and countries. The tool should select a node point on an existing grid, if a subgrid is being built. It should also be capable of understanding and compensating for buffer cells needed for the photochemical model. These new grid definitions should then be output in a form that the model can read. Operation over the web is desirable.

Optional task 4 is a graphical tool to help users identify problems with the existing SCC-to-spatial surrogate cross-reference. This tool would show which surrogates are used for each SCC code. This tool would also show where default surrogates are not available or where there is no cross reference for new SCC's.

Optional task 5 is an external graphical QA tool, similar to PAVE, that allows the user to view the spatial and temporal aspects of a select part of the input, intermediate, or output data sets. This visualization tool should, at a minimum, provide the following options:

- a tool for plotting the hourly, daily, or monthly distribution of emissions, with line and bar charts;
- a tool to show simple spatial plots of gridded or county emissions, with background maps with states and counties;
- a tool for tracking a single source through the entire emissions modeling process, including spatial, speciation, and spatial allocation decisions made by the model; and
- having platform independence (i.e., works on Windows) is preferred, but not required, if it will compromise LINUX performance or significantly effect cost.

Optional task 6 is a tool for the interpretation of travel demand model outputs directly into the motor vehicle model. This would be in lieu of requiring groups to convert the output of those models into RPO Data exchange protocol outputs. The proposed travel demand models we would prefer are UTPS, TP+, Viper, EMME/2, and TRANPLAN. Your proposal should show your expertise in understanding advanced travel demand modeling issues.

Level of Effort

The amount of funding for the development of the new emissions model, as outlined in the Statement of Work is $250,000. Any additional funding considered necessary to develop a more complete model (see, for example, the Optional Tasks identified above) should be included as options to the primary work effort.

Schedule

The period of performance is approximately 18 months. The work is expected to start in October 2003 and continue through April 2005. The model, along with appropriate documentation, is expected to be delivered in phases at approximately 6-month intervals:

Point and area source models, including speciation	May 1, 2004
Biogenics model	September 1, 2004
Mobile sources model	January 1, 2005
Optional elements	April 1, 2005

The Consortium will provide a small group of beta testers for draft versions of the models to work with the contractor to ensure model portability and functionality.

Deliverables
A kick-off conference call will be held after the contract is awarded to review the scope of work and schedule for this project. A brief project work plan will be due two weeks after the kick-off call. One electronic copy (WordPerfect or Word) of the work plan shall be delivered to the Consortium's Executive Director.

In lieu of written progress reports, monthly conference calls will be held to review the status of the work and discuss any outstanding issues.

Twenty (20) paper copies and one electronic copy (WordPerfect or Word) of the draft and final documentation, including a users guide and other relevant technical reports shall be delivered to the Consortium's Executive Director.

A draft and final model source code, including all data necessary to run the model, shall be delivered to the Consortium's Executive Director.

Two meetings to discuss the project results will be held in the upper Midwest (e.g., Chicago).

Key Personnel
The contractor shall identify key personnel for performance of the work under this RFP. The contractor must specify the amount of time that the key personnel will dedicate to the project. Any change in key personnel associated with the project shall be made only with prior written approval of the Consortium's Executive Director. These conditions pertain to both prime contractor and subcontractor personnel.

Evaluation Criteria
Evaluation and rating of proposals will be based on the following criteria:

(1) The contractor's understanding of the overall objectives of the emissions modeling project.

(2) The proposed technical approach, completeness of coverage with respect to the statement of work, and responsiveness to the available project resources.

(3) The experience, expertise, and other qualifications of the principal investigator and other personnel assigned to the project, and the level of effort proposed for the principal investigator.

Of particular note is the experience of key personnel in the following areas:

- writing SQL based programs, including SQL versions, platforms, complexity
- emissions model application, including models used and inventory complexity
- inventory experience, especially with travel demand models, complex temporalization, and biogenics inventories