Appendix C - Photochemical Modeling QA/QC Plan

Photochemical Modeling QA/QC Plan

In order to ensure that the inputs to the CAMx model are of the highest possible quality, the TCEQ Air Modeling Team performs a series of rigorous quality assurance procedures on the model input files. All data produced by external contractors are examined by a member of the air modeling staff.

Biogenic Emissions Estimates

Biogenic emissions estimates are produced using GloBEIS, which requires temperature data and photo synthetically active radiation data as inputs. Measured temperature data taken from the CAMS network, offshore buoys, an agricultural temperature network, and the National Weather Service are interpolated into hourly temperature fields. Contour plots are then made of the hourly fields to check for reasonableness. Desirable features of the data include:

Small diurnal temperature variations over the Gulf of Mexico;

Larger diurnal temperature variations over land;

Differences between temperatures in rural and urban areas. Urban areas should have higher overall temperatures; and

The effect of rain on the temperature. Any rain storms should decrease overall temperatures.

Photosynthetically active radiation data, the second type of input to the GloBEIS model, are extracted from satellite data. To qualitatively check the extracted data, they are compared to total solar radiation values measured at ground-based sensors. The time variation of the extracted data is compared to the time variation of the measured data. The two time variations should exhibit the same pattern.

In addition to the qualitative check above, a quantitative check is performed. Total solar radiation measured at a ground-based sensor is plotted against the derived photo synthetically active radiation for that same area. The slope of the line produced is usually close to 0.5, and the correlation coefficient is usually high, which indicates that the derived data is successfully picking up cloud fields.

Once the biogenic emissions estimates for every hour of the ozone episode have been produced by GloBEIS, qualitative checks are performed on them by creating tile plots and time series. Anomalies in the data are sought, including:

The timing of isoprene production, i.e., does isoprene production begin at sunrise and end at sunset?

Unexpected geographic allocation of emissions, e.g., are there emissions located over the bay or high isoprene production in downtown Houston?

To document quality assurance activities, a log file is created for each GloBEIS run. This log file contains the name and location of the input and output files, the date of the modeling run, the operator

name, and a brief description of the run. Further file traceability is provided by the tileplots produced, which are stamped with the name and date of the original file. This policy provides a second QA person the opportunity to trace file names from the tileplot back through the log file to the original input files.

Quantitative comparisons of the modeled data to measured data are documented by producing time series, scatterplots with regression statistics, and performance statistics. All graphs and statistics are stamped with the file name, date, location, and the date the plot was made.

Onroad Mobile Emissions Estimates

Estimates of onroad mobile emissions are produced by the Texas Transportation Institute (TTI). The data files received from TTI are processed in SAS to convert the times from daylight savings to standard time. Then the data are processed by three modules of EPS 2x: LBase, ChmSplt, and GRDEM. The resulting data files are then quality assured to ensure that the right amount of emissions are located in the right place (spatial allocation) at the right time (temporal allocation).
TTI provides summary files which have emissions totals for all pollutants for the entire nonattainment area as well as for each individual county. To ensure that no emissions are lost in the EPS 2x processing steps, the emissions totals for each pollutant in the message files of LBase, ChmSplit, and GRDEM are compared to the totals in the TTI summary files. First, the totals for the nonattainment area as a whole are compared to ensure that no gross emissions have been omitted. Then the totals for each county are compared.

To check the spatial and temporal allocation of the emissions, the data are fed into a tile plot program. The resulting plots are then checked to ensure that the emissions fall along the major roadways. In addition, the diurnal profile is checked to see if it follows the “Batman curve” (a curve with two peaks around 8 am and 5 p.m.) on Mondays through Thursdays.

Area Source Emissions Estimates

The processing of area source emissions is automated via UNIX scripts. First, the area source surrogates are tested for validity. Ten tons of test emissions are allocated to each county within the nonattainment area and viewed with tile plots to check that they are reasonably distributed throughout the counties. Second, the actual emissions are divided into various categories (e.g., logging, lawn and garden equipment, offshore shipping, etc.) by a PERL script. Third, a UNIX shell script is used to process the emissions through EPS 2x. Each category of emissions is processed separately. Once the emissions are gridded, each category is viewed using tileplots to check for reasonableness.

Quality assurance of the emissions is also automated. One program deletes all the intermediate EPS 2x files and confirms that EPS 2x generated no significant errors. Any error messages generated by EPS 2x are written to the program output. A second program tracks the emissions amounts through the EPS 2x processing steps. It summarizes the input and output emissions for every processing step, showing where any emissions were lost and why.

Point Source Emissions

Point source data are extracted from the TCEQ's Point Source Database (PSDB). Many QA/QC checks are performed on the PSDB data that are extracted. Initially, the annual emissions totals by pollutant, stack, SIC, and SCC are compared to the respective average of the previous five years. The annual emissions should generally be within three standard deviations of the five-year average. Any outliers that are not easily explained are reported to the Industrial Emissions Assessment Section (IEAS) team leaders for correction or explanation. Next, the stacks are checked for outlier stack parameters. The parameters of each stack are compared to the average of the previous five years to ensure that they are within three standard deviations of the five-year average for that type of facility (SIC and SCC). Again, any outliers that are not easily explained are reported to the IEAS team leaders for correction or explanation. Default stack parameters by source type are used to substitute for outlying stack parameters. Next, the data records are checked for incorrect/outlier stack coordinates (geographical locations). If a stack's coordinates are located outside of the county where the plant is located, they are changed to the coordinates of the center of the county. Default start times for selected pieces of equipment are substituted for equipment with faulty start times. There is also a check performed for missing SAROAD/AIRS pollutant codes, among many other small QA/QC checks performed directly on the extracted PSDB data.

Next, the overall performance of the modeling extract is checked against "Top 100" Paradox queries of the PSDB performed by the IEAS. The top 100 NOx-emitting stacks in the state, from the modeling extract, are compared to a similar list obtained from IEAS queries. This process is performed to ensure that no large emitters of NOx have been omitted during the extraction process or subsequent PSDB QA/QC checks. A similar check is performed for the top 100 VOC-emitting stacks, and any discrepancies between the two lists are reconciled prior to any further point source processing. A comparison of the "Top 100" lists with previous "Top 100" lists is performed, to ensure that large emitters in the state have not been accidentally dropped off the lists and to ensure that any new large emitters in the state are correctly on the lists.

The last step of the PSDB extract and QA/QC processing is the creation of AIRS Facility Subsystem (AFS) records, as required for EPS2x preprocessing. The AFS records for the state are then split into electrical generating utilities (EGU) and non-EGU (NEGU) AFS files. The primary reason for splitting the AFS file into EGU and NEGU AFS files is to ease facilitation of incorporating hourly CEM data from EPA's Acid Rain Program Database (ARPDB) into the EGU AFS file. The PSDB-to-ARPDB cross-reference file is updated each time a new modeling episode is chosen to ensure the best link between the PSDB stack identifiers and the ARPDB identifiers. The cross-reference is double-checked to ensure that every boiler in the ARPDB data matches a FIN/EPN (emission point/stack) in the PSDB data. This cross-reference must be used in the merge of the ozone-season daily PSDB data with the hourly ARPDB data. After the merge, a manual sampling of records is performed to make sure that there is a record for every pollutant for every hour of every day of the ozone episode, for each EGU in the ARPDB.

Once hourly records from the ARPDB records have replaced the corresponding PSDB records, at least two sources are chosen at random for quality assurance tracking. Two days of the episode are

chosen at random, and the emissions of NOx (and SO_2, if SO_2 is being modeled) for every hour are compared to their respective values in the original ARPDB data to ensure that no emissions were lost or gained during processing. The emissions of two potential problem (mis-allocation of emissions from ARPDB boilers to PSDB stacks) stacks, such as common stacks and multiple stacks, are also specifically checked. Hourly VOC and CO emissions for each ARPDB stack are allocated according to the hourly NOx profile of each day for each ARPDB stack. The calculated VOC and CO emissions are checked to ensure that the program has allocated them completely and accurately, and that the PSDB NOx/VOC and NOx/CO ratios are maintained for every hour of the day.

After the EGU data have been thoroughly checked, they are processed through EPS2x. "Tileplots" are created for the gridded low-level emissions and elevated emissions to facilitate additional QA/QC of spatial and temporal allocation. Consistent with electrical usage patterns, EGUs usually produce a diurnal profile with a broad peak near mid-afternoon. More details about tileplots as a QA/QC tool are given below.

If hourly Special Inventory (SI) data is available, incorporation of SI data follows a very similar QA/QC routine as the ARPDB data. SI data may be applied to both the EGU AFS file and the NEGU AFS file. AFS files are created for all states within the modeling domain, and ozone season or annual emissions records are given very nearly the same thoroughness of QA/QC as the state of Texas and the nonattainment area. When hourly data for other states is incorporated into those ozone season or annual records, the same level of QA/QC is applied as for Texas processing, to ensure that emissions are not lost or gained, and that the files are as accurate as the available raw data. A SAS program, sum_afs, is executed to compare the overall ozone season daily and hourly emissions for a day's worth of AFS records with that day's output from the EPS2x PREPNT module.

As with area sources, the processing of point source emissions through EPS2x is fairly automated via UNIX scripts. The point source emissions for the entire domain are typically divided into manageable pieces. Within each piece of the modeling domain, the data are completely related in that they have either come from the same data source, or the data will be expected to have similar growth and controls applied to them for the future case. For example, in recent modeling studies, the pieces included: Texas EGUs, Texas NEGUs, Louisiana EGUs, Louisiana NEGUs, Offshore, Mexico, Regional EGUs, and Regional NEGUs. The level of QA/QC drops off only slightly with distance from the nonattainment area.

QA/QC of the EPS2x processing steps is also fairly automated. Scripts are written to capture any error messages generated by EPS2x at each processing step.
A tracking summary program combines all of the message file input and output emissions values from each EPS2x processor program. This tracking program generates a single file for easy QA/QC of emissions gain and loss from a given portion of the inventory. As stated above, tileplots are generated for each low-level and elevated component of each portion of the inventory.
These steps are repeated for each piece of the modeling domain as it is processed through EPS2x. Low-level pieces of the point source inventory are merged together via EPS2x module MRGUAM. The input to MRGUAM is a list of the pieces to be included in each model run, so this list is QA-ed thoroughly. A summary output from MRGUAM of all of the emissions from each piece can be

compared to previous or similar merges to ensure that emissions are as expected. The low-level total emissions for each day is also provided as output from this step. As a final QA step for low-level sources, a tileplot of the total merged points is generated. The typical diurnal profile for low-level point sources is fairly flat across the day, with typical variations across hours being less that ten percent.

A similar process is performed for elevated point sources prior to input to the photochemical model. Since elevated sources generally only come from point sources, any change in elevated emissions is due to point sources. Tileplots are generated for every merged elevated file, the spatial distribution is compared to previous merges, the overall emissions total on the tileplot is verified, and the diurnal profile for each day is checked. The typical diurnal profile of elevated point sources peaks in mid afternoon because EGU emissions are the dominant source type in the elevated files.

MM5 QA/QC Procedures

Quality checking meteorological fields for MM5 requires several pre-processing steps. The output from each pre-processor provides a check for some of the variables and parameters which are part of MM5 input fields. Running MM5 itself requires verification through the job deck, mm5.deck, that switches and options have been correctly selected. The quality assurance of MM5 output is a central part of the technical work plan designed by TCEQ staff and Dr. Nielsen-Gammon of Texas A&M University and is intended to provide a rationale for the sensitivity analyses which followed the initial modeling.

A description of each of the pre-processor configurations was provided in the report *Initial Modeling of the August 2000 Houston-Galveston Ozone Episode, December 19, 2001*. Of particular importance was the evaluation and graphical inspection of surface characteristics and sea temperatures after running TERRAIN, PREGRID, and REGRIDDER. The “basic-state” of the Houston atmosphere, as characterized by the sea level temperature and pressure, lapse rate, and stratospheric isothermal temperature, was selected by Dr. Nielsen-Gammon to reflect the Houston summertime environment in the processor INTERPF.

Due to the importance of the land use and water characteristics on forcing the MM5 surface boundary, special attention was directed towards a realistic specification of key surface parameters. In particular, the adjustment of soil moisture availability was made by referencing climatological data, and in part by analyzing the Bowen ratio. These are referenced in more detail in the above referenced report.

The availability of global analysis fields and observational data is potentially very useful for model nudging. However, the use of data which has not already been checked for quality could destabilize MM5 or produce unrealistic responses which do not accurately reflect the true atmosphere. The global analysis fields which were used for model initialization and boundary conditions already have undergone review by NCEP modeling staff. Use of these fields for analysis nudging allowed for further inspection. Surface observations were collected and inspected by Dr. Nielsen-Gammon under a contract completed prior to MM5 modeling. Review and quality assurance of GPS sonde,

rawinsonde, and profiler data was performed by NOAA Environmental Technology Laboratory and Aeronomy Laboratory staff. The completed document will be available by August 31, 2002.

Quality Assurance

The use of generic input and output file names is part of somewhat complicated MM5 and pre-processor job decks, so separate output directories were created for each run to avoid over writing and to help manage post process analysis. Job decks were preserved in these directories, and a log of model runs was maintained at www.met.tamu.edu/results. A separate file was maintained on the TCEQ SGI Origin 3400 ("typhoon") labeled "state of the art model" to document the "best" current configuration of the MM5 model output to be used for SIP modeling.

Review of the job decks for MM5 and each of the pre-processors was accomplished by TCEQ staff. In addition to reports provided to TCEQ and associated WWW links, Dr. Nielsen-Gammon and TCEQ staff presented summary reports at Interim Science Coordinating Committee meetings and at Technical Review Committee meetings to members of the broader scientific community who participated in the TexAQS 2000 data intensive as well as stakeholders who have had long standing familiarity with meteorological modeling issues for the Houston-Galveston area.

Input File Tracking

TCEQ uses the air quality modeling log database to log and track the ozone photochemical modeling runs. The database contains the information of the air quality model used for a run, modeling input files and descriptions, job control or script file, and QA information. The database is implemented with the MySQL database system. The web browser based user interface was created to allow the modeling staff to make log record entry and query. The log database record fields are listed in the following table:

Air Quality Modeling Log Database

Item	Type	Description
id	int	unique ID number of a modeling run (used internally by the database)
log_date	date	date of the modeling run and logging
log_name	char	name of staff who makes the run and log entry
qa_date	date	date of QA is conducted
qa_name	char	name of staff who QAs the run
project	char	project name, e.g. HGMCR
episode	char	episode dates, e.g. 20000822-20000901
model_type	char	name of AQ model, e.g. CAMx
model_version	char	version of AQ model, e.g. 3.10
case	char	modeling case: base or future and version #, e.g. base1 (part of output file name)
scenario	char	sensitivity or control scenario, e.g. 070wind for 70% wind (part of output file name)
el_ei_type	char	case and scenario of elevated EI (part of input file name), e.g. base1.regular
lo_ei_type	char	case and scenario of low-level EI (part of input file name), e.g. base1.regular
met	text	description of meteorological inputs
ei_bio	text	description of biogenic EI inputs
ei_area	text	description of area and non-road mobile EI inputs
ei_point	text	description of popint source EI inputs
ei_mobile	text	description of on-road mobile EI inputs
other	text	description of other features
job	text	modeling job control files/script

This database is implemented with MySQL with the web-browser-based user interface for log entry and query.