LADCO IPM MODEL PARSED FILE POST-PROCESSING METHODOLOGY AND FILE PREPARATION (JANUARY 2005)

TECHNICAL MEMORANDUM

PECHAN

5528-B Hempstead Way
Springfield, VA 22151

703-813-6700 telephone
703-813-6729 facsimile

3622 Lyckan Parkway
Suite 2002
Durham, NC 27707

919-493-3144 telephone
919-493-3182 facsimile

P.O. Box 1345
El Dorado, CA 95623

530-295-2995 telephone
530-295-2999 facsimile

Prepared for:

Lake Michigan Air Directors Consortium

Prepared by:

Susy Rothschild
E.H. Pechan & Associates, Inc.
5528-B Hempstead Way
Springfield, VA 22151

February 8, 2005

I. INTRODUCTION

This technical memorandum describes the procedure for preparing air quality modeling input-formatted emission files based on Integrated Planning Model (IPM) output. These files will be used for Midwest Regional Planning Organization (MRPO) and Visibility Improvement-State and Tribal Association of the South (VISTAS) modeling for visibility, ozone, and fine particles.

A. POST-PROCESSING METHODOLOGY

The essence of the IPM model post-processing methodology is to take an initial IPM model output file and transform it into air quality model input files. ICF via VISTAS/MRPO provides an initial spreadsheet file containing unit-level records of both (1) "existing" units and (2) committed or new generic aggregates. All records have unit and fuel type data; existing, retrofit (for SO_2 and NO_x), and separate NO_x control information; annual SO_2 and NO_x emissions and heat input; summer season (May-September) NO_x and heat input; July day NO_x and heat input; coal heat input by coal type; nameplate capacity (MW), and State FIPS code. Existing units also have county FIPS code, a unique plant identifier (ORISPL) and unit ID (also called boiler ID) (BLRID); generic units do not have these data.

This section includes a description of the post-processing methodology.

The IPM data are processed using EPA-approved data files and methodology. The processing includes estimating various types of emissions and adding in control efficiencies, stack parameters, latitude-longitude coordinates, and State identifiers (plant ID, point ID, stack ID, process ID). Additionally, the generic units are sited in a county and given appropriate IDs. This processing is described in more detail below.

1. Generics

The data are prepared by transforming the generic aggregates into units similar to the existing units in terms of the available data. The generic aggregates are split into smaller generic units based on their unit types and capacity, are provided a dummy ORIS unique plant and boiler ID, and are given a county FIPS code based on an algorithm that sites each generic by assigning a sister plant that is in a county based on its attainment/nonattainment status. Within a state, plants (in county then ORIS plant code order) in attainment counties are used first as sister sites to generic units, followed by plants in PM nonattainment counties, followed by plants in 8-hour ozone nonattainment counties. Note that no LADCO or VISTAS States provided us with blackout counties that would not be considered when siting generics, so this process is identical to the one used for EPA IPM post-processing.

2. Adding Data for All Units

SCCs were assigned for all units; unit/fuel/firing/bottom type data were used for existing units' assignments, while only unit and fuel type were used for generic units' assignments.

Latitude-longitude coordinates were assigned, first using the EPA-provided data files, secondly using the September 17, 2004 Pechan in-house latitude-longitude file, and lastly using county centroids. These data were only used when the data were not provided in the 2002 NIF files.

Stack parameters were attached, first using the EPA-provided data files, secondly using a March 9, 2004 Pechan in-house stack parameter file based on previous EIA-767 data, and lastly using an EPA June 2003 SCC-based default stack parameter file. These data were only used when the data were not provided in the 2002 NIF files.

Plant ID (within State and county), point ID, process ID, and stack ID were then attached, first using the EPA-provided data files, or secondly using Pechan-generated defaults: the point ID is assigned the value of the given boiler ID preceded by '#', unless the boiler ID has a length of six [the length for the point ID], in which case the left-most character is replaced with '#'); and the default Pechan process ID is '01'. Default stack IDs within a plant are assigned for each unique stack height-diameter combination; the default Pechan stack ID is of the form '4N'. The process ID and stack ID default data were only used when the data were not provided.

3. Estimating Emissions

Additional data were required for estimating VOC, CO, filterable primary PM_{10} and $PM_{2.5}$, PM condensable, and NH_3 emissions for all units. Thus, ash and sulfur contents were assigned by first using 2002 EIA-767 values for existing units or SCC-based defaults; filterable PM_{10} and $PM_{2.5}$ efficiencies were obtained from the 2002 EGU NEI that were based on 2002 EIA-767 control data and the PM Calculator program (a default of 99.2% is used for coal units if necessary); fuel use was back calculated from the given heat input and a default SCC-based heat content; and emission factors were obtained from an EPA-approved October 7, 2004 Pechan emission factor file based on AP-42 emission factors. Note that this updated file is not the one used for estimating emissions for previous EPA post-processed IPM files.

Emissions for 28 temporal-pollutant combinations were estimated since there are seven pollutants (VOC, CO, primary PM_{10} and $PM_{2.5}$, NH_3, SO_2 and NO_x) and four temporal periods (annual, summer season, winter season, July day). Note that annual SO_2 and annual, summer season, and July day NO_x emission values are provided in the file.

- Annual emissions were first estimated by multiplying the fuel use, emission factor (which might include sulfur and/or ash content factor), removal control efficiency, and a units conversion factor.
- Summer emissions were estimated by multiplying the annual emissions by a ratio of the summer to annual heat input.
- Winter emissions were estimated by subtracting the summer emissions from the annual emissions.
- July day emissions were estimated by multiplying the annual emissions by a ratio of the July day to annual heat input.

The FoxPro program, and associated date files (in Excel) are included in the zip file, LADCOIPMdoc.zip, as is this document.

4. Crosswalk File

The Task 1 crosswalk file was used to obtain FIPS State and county, plant ID (within State and county), and point ID. If the FIPS State and county, plant ID and point ID are in the 2002 NIF tables, then the process ID and stack ID are obtained from the NIF; otherwise, defaults, described above, were used.

B. FILE PREPARATION

This section of the documentation describes the file preparation for these data; that is, how the post-processed IPM data were formatted as MRPO-specified NIF 3.0 tables. The tables were developed separately as two sets of tables for (1) IPM boilers that had a crosswalk match that were in the 2002 NIF files provided by LADCO and VISTAS, and (2) IPM boilers that were not

in the 2002 NIF files provided by LADCO and VISTAS. Note that records for a given plant can be included in both sets of tables because a boiler in the IPM scenario may have a match to a point in the 2002 NIF tables, while another boiler at the same plant may not.

NIF tables for each scenario were developed and based on the Task 1 crosswalk. Two special cases relating to the crosswalk match were handled as follows:

1. One-to many match: At a given plant, if one IPM boiler ID is matched to more than one point ID, the boiler data are put on the first point ID records; records from the other point IDs are deleted from the relevant tables.

2. Many-to one match: At a given plant, if more than one IPM boiler ID is matched to one point ID, all the boilers' emissions (tons), throughput (really heat input in MMBtu), and capacity (MW) are summed ("summed boiler") and put on that point ID's records in the relevant tables. The values for stack parameters and latitude-longitude values are those from the first record summed.

The two sets of tables that were developed are described below.

A. "**NIF**" – For IPM units that have a crosswalk match and are in the 2002 NIF tables, the following was done:

In the **SI** table:

Filled in the scenario ORIS plant code as a character variable.

All records without ORIS plant code values were deleted.

If two ORISPL match with one FIPS State+county - plant ID, both records are included in this table.

In the **EU** table:

Filled in the scenario boiler ID.

All records without boiler ID values were deleted.

Filled in maximum nameplate capacity (given in IPM scenario).

In the **EM** table:

Identified which process ID record to use for each emissions release point ID (and stack ID) by using the process ID that has the largest SO_2 + NO_x temporal emissions sum. If no SO_2 + NO_x for any process ID at that point, then used largest primary PM_{10} + primary $PM_{2.5}$ temporal emissions sum; if not that, used VOC + CO + NH_3 temporal emissions sum; if not that, used a default process ID associated with that point ID in EPA IPM runs; or as a last resort, used the Pechan default process ID.

Deleted all EM records if

(1) they are for point IDs without a match in the IPM file;

(2) they are for process IDs not chosen as the one for the given point ID;

(3) they are not for the 7 pollutants (VOC [changed from ROG], CO, NO_x, SO_2, NH_3, primary PM_{10}, or primary $PM_{2.5}$);

(4) they are for durations other than annual (20yy0101 – 20yy1231), summer season (20yy0501 – 20yy0930), winter season (20yy1001 – 20yy0430 [MRPO will change these start and end dates], and typical season day (20yy0721 – 20yy0721), where yy=09 or 18.

Added needed pollutant records (with identifiers) if they are not originally in the EM table.

Filled/added in pollutant code and start and end dates for each record.

Filled/added in emissions for the 7 pollutants for the 4 temporal periods.

Filled/added in emission unit numerator (TON) and emission type (30) for all kept and added records.

Filled/added in (emission) factor numeric values [which we will calculate], emission unit numerator (LB), and emission unit denominator (E6BTU) for annual SO_2, NO_x, primary PM_{10} and primary $PM_{2.5}$ and all pollutants' summer season, winter season and typical day emissions. Left blank if summed boiler data. Note that because of NIF 3.0 table format limitations, there may be some values that are not 0 but appear to be 0 because they have the value 0 when formatted to two decimal places for (emission) factor numeric values. Real 0 values are indicated with a value of -99 to distinguish them from very small values that display as zero in the NIF tables.

Filled/added in (emission) factor numeric values [the latest EPA-approved emission factor which was used to calculate the emissions], emission unit numerator (LB), and emission unit denominator (appropriate SCC unit for the SCC we used [which is not necessarily the SCC displayed in the NIF file]) for annual CO, VOC, and NH_3. Left blank if summed boiler data.

In the **PE** table:

Deleted all FIPS State+county - plant ID - point ID - process ID - start date - end date records that are not in the EM table; added those that are in the EM table.

Filled/added in actual throughput (heat input given in IPM scenario or else derived) and throughput numerator units (E6BTU) for the 4 temporal periods.

In the **CE** table:

Kept records for those EM table identified FIPS State+county - plant ID - point ID - process ID records, whether or not they have control devices in IPM file.

Controls for NO_x, SO_2, and filterable PM (not primary PM_{10} and $PM_{2.5}$) are included to "record" what information about controls was provided in the IPM file. Note that PM controls were not used to estimate PM_{10} and $PM_{2.5}$ emissions, and the PM pollutant code will match with no other NIF table's pollutant value. The fields used to determine the control devices (as best we could) are retrofit controls, existing controls, NO_x controls – in that order. Left blank if summed boiler data.

Filled/added in pollutant code, and primary, secondary, third control, fourth control device type codes. If there was a retrofit control, it was included first. We used the best possible acceptable NIF 3.0 codes for the IPM control devices.

Did not fill in any control efficiency information since none was provided in the IPM file.

In the **EP** table:

Only kept records for those EM table identified FIPS State+county - plant ID - point ID - process IDs.

Did not fill in any values.

Added in default process ID records when necessary.

Added in IPMSCC (character 8) at end of each NIF record to display the single SCC assigned to each IPM boiler. Left blank if summed boiler data.

In the **ER** table:

Only kept records for those EM table identified FIPS State+county - plant ID - stack ID records.

Added in default stack ID records when necessary.

Did not fill in any values.

B. "**NoNIF**" – For IPM units that are not in the NIF (which includes existing units with or without a crosswalk match as well as generic units), we did the following (besides filling in the record type for all tables):

In the **SI** table:

Added all records at the plant level.

Included the scenario ORIS plant code as a character variable, FIPS State+county code, plant ID (='ORIS'+ORISPL value if not matched in the crosswalk file), facility name, SIC primary (4911 for all), NAICS primary (22 for all), State abbreviation, and country (USA).

In the **EU** table:

Added all records at the plant-point ID level.

Included FIPS State+county code, plant ID (='ORIS'+ORISPL value if not matched in the crosswalk file), point ID (='#' + boiler ID value, unless the boiler ID has a length of six [the length for the point ID], in which case the left-most character is replaced with '#') – if there is no crosswalk match), boiler ID, SIC unit level (4911 for all), NAICS unit level (22 for all), and maximum nameplate capacity.

In the **EM** table:

Added all records at the plant-point-process ID-pollutant level.

Included the pollutant code, stack ID, start and end dates, emissions for the 7 pollutants for the 4 temporal periods, and emission unit numerator (TON) and emission type (30).

Added in (emission) factor numeric values [which we will calculate], emission unit numerator (LB), and emission unit denominator (E6BTU) for annual SO_2, NO_x, primary PM_{10} and primary $PM_{2.5}$ and all pollutants' summer season, winter season and typical day emissions. Left blank if summed boiler data. Note that because of NIF table format limitations, there may be some values that are not 0 but appear to be 0 because they have the value 0 when formatted to two decimal places for (emission) factor numeric values. Real 0 values are indicated with a value of -99 to distinguish them from very small values that display as zero in the NIF tables.

Added in (emission) factor numeric values [the latest EPA-approved emission factor which was used to calculate the emissions], emission unit numerator (LB), and emission unit denominator (appropriate SCC unit for the SCC we used [which is not necessarily the SCC displayed in the NIF file]) for annual CO, VOC, and NH_3. Left blank if summed boiler data.

In the **PE** table:

Added all records at the plant-point-process ID level.

Included start and end dates, actual throughput (heat input given in IPM scenario or else derived) and throughput numerator units (E6BTU) for the 4 temporal periods.

In the **CE** table:

Added records at the plant-point-process ID-pollutant level if they are in the EM file.

Controls for NO_x, SO_2, and filterable PM (not primary PM_{10} and $PM_{2.5}$) are included to "record" what information about controls was provided in the IPM file. The fields used to determine the control devices (as best we could) are retrofit controls, existing controls, NO_X controls – in that order. Left blank if summed boiler data.

Included pollutant code, and primary, secondary, third control, fourth control device type codes. If there was a retrofit control, it was included first. We used the best possible acceptable NIF 3.0 codes for the IPM control devices.

Did not provide any control efficiency information since none was provided in the IPM file.

In the **EP** table:

Added all records at the plant-point-process ID level.

Included stack ID and SCC (which is left blank if summed boiler data).

In the **ER** table:

Added all records at the plant-stack ID level.

Included stack ID, the five stack parameters (height, diameter, temperature, velocity, and flow), the latitude and longitude coordinates (may be county centroid), and the xy coordinate type (LATLON).

File layouts for the tables for the NIF and NoNIF records are included in the zip file, LADCOIPMdoc.zip, as is this document.

The list of supporting data files (besides the original IPM scenario file) used for this post-processing are listed below.

File Name	File Description
allgenerics	Lists all generics sited in the 4 scenarios.
centroid	Lists county centroids used.
controlchoicesall	Lists all control choice combinations in the 4 scenarios (used in condensable PM emission factor assignment).
f767bg02	2002 EIA-767 Boiler-Generator file used to relate into 2002 EGU inventory.
fipsatal	Lists State codes and names used.
inv02704	File with 2002 EIA-767-based data for PM control efficiencies, sulfur and ash content.
ipmtonifcontrols	Lists all the NOx, SO2, and PM NIF control device code assignments corresponding to the control devices in the 4 scenarios.
latlon091704	Lists power plant latitude-longitude coordinates used.
neiblr1	NEI boiler data file received from EPA used to fill in data if needed.
neiplt1	NEI plant data file received from EPA used to fill in data if needed.
neiplt2	NEI plant data file received from EPA used to fill in data if needed.
neiplt3	NEI plant data file received from EPA used to fill in data if needed.
nif2002_CE_completeset	Compilation of the three sets of 2002 NIF CE files received.
nif2002_EM_completeset_pt1	Compilation of the three sets of 2002 NIF EM files received (part 1).
nif2002_EM_completeset_pt2	Compilation of the three sets of 2002 NIF EM files received (part 2).
nif2002_EM_completeset_pt3	Compilation of the three sets of 2002 NIF EM files received (part 3).
nif2002_EP_completeset	Compilation of the three sets of 2002 NIF EP files received.
nif2002_ER_completeset	Compilation of the three sets of 2002 NIF ER files received.
nif2002_EU_completeset	Compilation of the three sets of 2002 NIF EU files received.
nif2002_PE_completeset	Compilation of the three sets of 2002 NIF PE files received.
nif2002_SI_completeset	Compilation of the three sets of 2002 NIF SI files received.
pmcdef	Latest EPA-approved condensable PM emission factor assignment (embedded within code).
sccdeflt	Lists default SCC-level stack parameters and default heat contents.
sccemfac100704	Latest EPA-approved uncontrolled emission factor file for CO, VOC, NH3, filterable PM-10, and filterable PM-2.5.
sccmatch	Lists all SCC combinations assigned in the 4 scenarios.
siteall	List of plants and their county's PM-10 (as of January 2004) and 8-hour ozone (as of April 2004) Nonattainment status used for siting generics.
stkp02or030904	Compilation of default stack parameters taken from the 1999 through the 2002 EGU inventory, with 2001 as the dominant year.
xwalkmrpoupdforprog	MRPO IPM-State ID crosswalk file for MRPO IPM post-processing. This file differs from the "true" crosswalk file (delivered December 9, 2004) in that it eliminates any one boiler-to-multiple points relationship, per LADCO agreement.

Additional file structures used solely for the FoxPro program that actually does the IPM post-processing.	
File Name	**File Description**
mrpoyylxxstat	File structure for mrpoYYLXXstat.xls, a file used to format and convert into a dbf the IPM scenario file delivered to Pechan.
mrpoyylxx	File structure for the adjusted version of mrpoYYlXXstat.dbf.
ladcostructure	File structure for a working .dbf file used throughout the FoxPro program.