Data Analysis for the Lake Michigan Ozone Study

By:

Paul T. Roberts
Timothy S. Dye
Hilary H. Main
Sonoma Technology Inc.
5519 Skylane Blvd., Suite 101
Santa Rosa, CA 95403

INTRODUCTION

The Lake Michigan Air Quality Region (LMAQR), which encompasses portions of Wisconsin, Illinois, Indiana, and Michigan, continues to experience ozone levels in urban and rural areas that exceed the federal standard of 12 pphm. In recent years, hourly ozone concentrations within the region have exceeded 20 pphm. Most of the ozone-forming precursor emissions, volatile organic compounds (VOC) and oxides of nitrogen (NO_x), originate in the urban/industrial areas of northern Indiana, northeastern Illinois (the greater Chicago area), and the Milwaukee metropolitan area.

The unique meteorology of the LMAQR, namely the interaction of the lake/land breeze circulation with the predominant southerly synoptic flow, gives rise to a number of potential transport regimes. Precursor emissions and locally produced ozone from the urban/industrial areas have the potential to be transported northward along the lakeshore areas of eastern Illinois and Wisconsin, and eastward and northeastward toward the shoreline areas of Indiana and western Michigan. These transported pollutants may contribute to ozone concentrations that exceed 12 pphm far downwind (50-150 miles) of these areas. In addition to local emissions (those emitted within the LMAQR), precursors and ozone transported into the LMAQR from outside the region may also contribute to high ozone concentrations in the LMAQR.

The LMAQR states have planned a comprehensive field measurement and photochemical modeling program, the Lake Michigan Ozone Study (LMOS), as a basis for revised State Implementation Plans to be submitted to the USEPA. The LMOS includes a number of major elements: field measurements; emissions, meteorology, and photochemical modeling; information management; quality assurance; and data analysis.

This paper summarizes the data analysis effort which is currently underway using the LMOS field data taken in the summer of 1991; additional analysis results will be completed soon and will be presented at the meeting. Preliminary analyses were performed using data from selected 1990 and 1987 ozone episodes in the LMAQR[1,2]. Recommendations based on the preliminary analysis results were used to improve the design of the 1991 field study.

This paper presents an overview of the data analysis tasks. A systematic and complete data validation task is critical for successful data analysis and modeling. Case studies will be used to illustrate important phenomena influencing high ozone concentrations in the LMAQR. An understanding of these phenomena is important to an improved understanding of ozone formation and transport in the LMAQR and to an improved model development such as the LMOS. These phenomena include transport into the region along the southern and western boundaries, the lake/land breeze along the shores of Lake Michigan, and the transport of pollutants over the lake and eventually back to both the western and eastern shores of the lake. Detailed analysis results using the 1991 field study data will be completed soon and will be presented at the meeting.

The overall data analysis effort is designed to support the modeling and planning elements of the LMOS. The objectives of the overall data analysis effort are:

- To ensure that the LMOS modeling data base is consistent and quality assured;
- To obtain a better understanding of those processes that influence the formation and transport of ozone and ozone precursors in the LMAQR; and
- To recommend measurements systems for the 1991 field study and specifications for the LMOS modeling system which will adequately describe those processes that influence the formation and transport of ozone and ozone precursors in the LMAQR.

TECHNICAL APPROACH

The data analysis efforts required to achieve the data analysis objectives and to address the technical issues naturally fall into two phases, as defined by the data available for analysis and by the schedule. The first phase involved analysis of the 1990 field data and making recommendations for the 1991 field study and for the LMOS modeling system. Phase I was completed by the time of the 1991 field study[1]. The second phase involves analysis of the 1991 field data and is currently in progress.

We have prepared flow charts showing the data analysis tasks (see Figures 1 and 2). Connections between task boxes in these charts indicate the major flows of data and results (a few minor data flows are left out for clarity). The flow charts are useful in planning and managing the data analysis effort, since they illustrate how the tasks fit together, which tasks will provide results to which other tasks, and how the tasks will help meet the data analysis objectives. The time sequence of tasks is controlled by the availability of data and by the relationships between the tasks.

The major activities for Phase II are shown in Figure 1. In general, data and task results flow from left to right: from data validation, to describe and display tasks, on to various interpretative analysis tasks, and finally to tasks which integrate the results of previous tasks. Figure 2 is similar to Figure 1, except it shows more details for Task 4. Note that we have divided Task 4 into a number of subtasks, based on the large number of activities involved and the size of the efforts.

The first task of Phase II (Task 2) involves evaluating the 1991 field data set and preparing it for subsequent analyses; the major purpose of this task is to assess the accuracy, precision, and validity of the data set. The next tasks (Tasks 4a-4g) involve 'describing and displaying' various portions of the data set; these tasks are designed to summarize the data, make the data useful to other tasks, and illustrate specific transport and transformation mechanisms at work in the LMAQR. Task 4e includes delivering validated sets of data to the emissions modeling and photochemical modeling contractors. The following tasks (Tasks 4h-4k and Tasks 5-9) involve interpretative data analyses; these tasks integrate large amounts of information to describe the evolution of pollutant concentrations in the LMAQR and the relevant physical and chemical processes which influence pollutant concentrations. The last three tasks bring together the results of earlier tasks into a cohesive picture. Task 10 provides a summary of the intensive field monitoring periods; Task 11 revises the conceptual model of ozone formation and transport in the LMAQR; and Task 12 provides a final report documenting the data analysis results.

DATA VALIDATION

Field data sets need to be evaluated before they can be used in data analyses and modeling. Serious errors in data analysis and modeling results can be caused by individual data values. Some of these data values are erroneous while others may deviate from the true measurement value due to measurement error. Potential problems might involve any number of causes, from simple instrument malfunction or mis-applied calibration information to the use of an measurement system inappropriate to measure that parameter. In order to minimize unnecessary errors and uncertainties, all values in the data set should be reviewed, evaluated, and flagged.

A rigorous evaluation of the data set involves analyses to:

- evaluate the precision and accuracy estimates of the field measurements;
- compare measurements of the same parameter from different measurement systems (aloft temperature and dew point via aircraft and rawinsonde; upper air winds via doppler acoustic sounder, rawinsonde, radar profiler, and aircraft, for example);

- compare surface meteorology measurements with upper air meteorology measurements at altitudes near the surface;
- compare surface air quality measurements with air quality aircraft measurements at altitudes near the surface;
- test the internal consistency of the various air quality and meteorology measurements; and
- evaluate the adequacy of specific measurement systems to provide data for planned analyses and modeling, identify when specific data are not adequate for the planned analyses and modeling.

Data validation is an integral part of all data analysis tasks. Data validation is defined as the process which identifies deviations from measurement assumptions and flags individual measurements as valid, valid but suspect, or invalid, based on pre-defined criteria. The data validation process is a continuing process throughout the data analysis and model development and evaluation efforts. Generally, there are three levels of data validation: Level I, Level II, and Level III.

Level I data validation takes place in the field or in the laboratory and consists of:

- flagging values when significant deviations from measurement assumptions have occurred;
- verifying computer file entries against data sheets;
- eliminating values for measurements which are known to be invalid because of instrument malfunctions;
- replacing data from a backup data acquisition system in the event of failure of the primary system; and
- adjusting measurement values for quantifiable calibration or interference biases.

Each measurement investigator performs Level I validation on his measurements.

Level II data validation is performed after data from various measurement methods have been assembled in the master data base. Level II applies consistency tests based on known physical relationships between variables in the assembled data. Examples of these tests are:

- wind speeds and directions should be similar for surface measurements and nearby upper air measurements;
- upper air wind speeds and directions should be similar for nearby measurements (Doppler acoustic sounder versus rawinsonde, versus radar profiler, etc.);
- dew point should always be less than temperature;
- temperature and dew point measured by rawinsonde and by aircraft should be similar;
- routine and AIRTRAK measurements for NO and for NO_x should agree; and
- NO_2 via luminol should be the same or larger than NO_2 by difference between chemiluminescence NO_x and NO.

Data adjustments for quantifiable biases are made in Level II validation if they are discovered after assembly of the master data base.

Level III validation is really a part of the data interpretation process. The first assumption upon finding a measurement which is inconsistent with physical expectations is that the unusual value is due to a measurement error. If, upon tracing the path of the measurement, nothing unusual is found, the value can be assumed to be a valid result of an environmental cause. Unusual values are usually identified during the data interpretation process as:

- extreme values;
- values which would otherwise normally track the values of other variables in a time series; or
- values for observables which would otherwise normally follow a qualitatively predictable spatial or temporal pattern.

All data validation actions at each level should be recorded in a data validation summary which accompanies the data volumes. Data base records should contain flags to identify the level of validation which they have received at any point in their existence.

AIR QUALITY AND METEOROLOGICAL CHARACTERISTICS OF OZONE EPISODES IN THE LMAQR

Basic characteristics of the ozone air quality problem in the LMAQR include the following observations:

- The typical ozone season in the LMAQR lasts from April through September; however, most of the ozone exceedance days occur in June, July, and August;
- On a regional scale, local stagnation and low wind speeds, high temperatures, intense solar radiation, and shallow mixing depths are associated with high ozone concentrations;
- A typical ozone episode might include a period of light winds and strong subsidence (producing a 'local' ozone exceedance) followed by a period of southerly or southwesterly winds which probably transport ozone and ozone precursors to downwind receptor sites;
- Monitoring sites in Wisconsin record ozone concentrations over the Federal Standard most often, followed by sites in Illinois and Michigan, and then Indiana;
- Most of the high ozone concentrations occur at monitoring sites within a few miles of the shoreline;
- The year-to-year variation in peak ozone concentration and in the number of days exceeding the ozone standard is large, up to a factor of four;
- For the LMAQR as a whole, about one-third of the ozone exceedances are single-day events and about one-third are two-day events. Single-day events are most common in Michigan and Indiana, whereas multi-day events are more common in Wisconsin and Illinois; and
- The greatest number of ozone exceedances are measured at monitoring sites in Illinois from 1400-1600 LST (local standard time), in Wisconsin from 1500-1700 LST, in Indiana from 1400-1700 LST, and in Michigan from 1500-1900 LST. Varying ozone peak times (on the average and on specific episode days) indicate variations in relative contributions of local and transported contributions.

The attribution of source contributions to high ozone concentrations in the LMAQR is difficult, since the formation of ozone can occur during complicated mixing and transport processes. Most ozone exceedance days are synoptically similar, with a surface anticyclone located over the LMAQR or to the

southwest of the LMAQR and a 500 mb ridge centered over the central or eastern United States. These conditions generally produce a subsidence inversion, winds become light or southerly, temperatures increase, and air quality deteriorates. Details will be discussed in a companion paper[3]. Favorable conditions exist for local pollution production and stagnation as well as long-distance pollution transport into and within the LMAQR by the prevailing southerly flow. Long-distance transport may occur from upwind sources as far away as St. Louis.

Several mechanisms, in addition to prevailing synoptic flow, can modify pollutant transport within the LMAQR. The low-level jet and the interaction of synoptic flow with the lake/land breeze are intermediate transport mechanisms which link synoptic-scale and local-scale flow. During the talk, we will discuss case-study examples using the 1991 field data to illustrate the various transport mechanisms, including transport into the region along the southern and western boundaries, the lake/land breeze along the shores of Lake Michigan, and the transport of pollutants over the lake and eventually back to both the western and eastern shores of the lake.

A low-level jet may rapidly and effectively transport pollutants several hundred kilometers. A 10 ms^{-1} low-level jet can transport pollutants anywhere in the LMAQR within 12 hours (e.g., a 5-10 ms^{-1} low-level jet can transport pollutants from Chicago to Milwaukee in 4 to 7 hours). Once transported, the pollutants can be mixed down to the surface by morning convection.

Interaction of the prevailing synoptic-scale flow with the lake/land breeze creates a second mechanism for intraregional pollutant transport. The lake/land breeze is driven by differential heating between the land and lake. During the day air rises over the land, flows lakeward, subsides, and completes the circulation by flowing landward. The opposite circulation occurs at night. The interaction of the prevailing southerly regional flow with the lake/land breeze flow may result in northward pollution transport and at the same time recirculation of pollution on and offshore.

On a local scale, the lake/land breeze transports pollutants by onshore and offshore flow and by interaction with the Thermal Internal Boundary Layer (TIBL). A TIBL results when surface roughness and heat fluxes differ between two air masses (over land and water). TIBL depth increases with downwind fetch. Early morning transport of precursors (NO_x and hydrocarbons) in the remnants of the land breeze can result in ozone formation over the lake. When the flow reverses from offshore to onshore, the ozone burden is transported shoreward for interaction with the TIBL and distribution to inland receptors. The influences of the land breeze on the ozone and precursor transport, ozone formation, and distribution are not fully understood.

To address intra-region pollutant transport mechanisms, the 1991 LMOS field study was designed with surface and aloft measurement systems concentrated along two two-dimensional data planes (2DDPs). These 2DDPs are shown in Figure 3. The primary 2DDP was located along the Wisconsin-Illinois border, out to the center of the lake, and then onshore in southwestern Michigan. The secondary 2DDP crossed the lake from Wisconsin north of Milwaukee to central Michigan. Upper air meteorological and air quality measurements were made at specific locations along these 2DDPs so that we could estimate the pollutant flux crossing these planes and to characterize the detailed structure of the land-lake breeze and investigate its influence on high ozone concentrations. Examples will be shown of pollutant concentrations, atmospheric structure, and pollutant flux estimates along these 2DDPs.

Over 800 hydrocarbon and carbonyl samples were collected at the surface and aloft during LMOS. At the surface, two-hour integrated samples were collected at several locations in the study area two to four times per day. The surface samples were assembled into the following groups for initial analyses: two urban/industrial sites (Northwestern University, IL and Gary, IN - CITY), five Wisconsin sites (Sheboygan to Bayview - WI), two Michigan sites (Borculo and Benton Harbor - MI), three northern Illinois sites (Rockford, Zion, and a boat offshore - IL), two upwind sites (Kankakee and Jasper County - UP), and two

other sites (OTH). Figure 4 shows the flight paths of the aircraft during the study. Grab samples were collected during up to three flights per day at various altitudes and locations along the flight paths.

To facilitate comparisons, we have determined the 25 most abundant organic species in the data set. Also many of the species in a sample may be below detection. Hydrocarbons and carbonyl species in each sample were ranked by concentration and a frequency distribution was determined which showed the number of times each nonmethane organic compound (NMOC) was one of the 30 most abundant. This frequency distribution was then ranked to identify the 25 most abundant species aloft and at the surface. The 25 most abundant hydrocarbons in LMOS accounted for 37% and 61% of the NMOC carbon aloft and at the surface. The carbonyl species accounted for an average of 15% of the surface NMOC, which is about the same as in the Los Angeles surface NMOC (16%, Lurmann and Main, 1992). The carbonyl species were over 50% of the NMOC aloft in LMOS, which is significantly higher than the 35% observed in Los Angeles. Validation of the carbonyl data has not yet been completed, therefore, the following analyses are limited to the hydrocarbons.

We compared the NMHC composition of the LMOS data to NMHC data collected in 41 cities by EPA in the summer of 1984-1986 (Jeffries et al., 1988). The EPA data set contained 773 samples and was carefully screened for outliers. Figure 5 compares the abundant NMHC species in LMOS, listed in roughly chromatographic order, to the 41-city average. For this comparison, the surface and aloft LMOS data includes samples collected throughout the day, and the EPA samples were from 0600-0900. The LMOS aloft and surface compositions differed significantly for several species. The aloft samples contained less acetylene, propene, isoprene, 1,2,4-trimethylbenzene than the LMOS surface samples but contained proportionately more ethane and toluene. NMOC concentrations aloft were a third to a half of the surface concentrations. The EPA data differed significantly from the LMOS data for acetylene, ethane, ethene, propane, and i- and n-butanes. The surface average LMOS data contains both rural and urban samples. When the LMOS urban samples are compared separately to the EPA average, the profiles match very closely.

The NMHC composition of samples from each flight path and surface group area shown in Figure 6. The composition of samples collected during different flight paths varied significantly for several species, including ethene, propane, benzene, toluene, and total xylenes. The boundary flight path (FP01) data had the highest weight percent (wt%) of C2-C3 species and the lowest wt% of C4+ species. In contrast, the FP02 data had the lowest wt% of C2-C6 species and the highest percent of C7+ species. The surface sample compositions also varied significantly. The largest difference was the propane and propene composition at the upwind sites. There were several samples from both sites which had high proportions of these species and thus, at this point in the validation process, we have chosen to retain these data in the data set. Additional analyses are underway to investigate the compositional differences.

ACKNOWLEDGEMENTS

The work described in this paper was funded by the Lake Michigan Air Directors Consortium (LADCO) as part of the Lake Michigan Ozone Study (LMOS). The work was directed by the LADCO Technical Director, Mr. Mike Koerber. STI 90210-1209.

References:

1. Roberts, P.T., T.S. Dye, S.R. Hanna, and C.M. Haga (1991) *Preliminary Data Analysis for the Lake Michigan Ozone Study.* Report prepared for the Lake Michigan Air Directors Consortium by Sonoma Technology Inc. and Sigma Research Corporation, STI-90212-1157-DR, June.

2. Roberts, P.T., T.S. Dye, S.R. Hanna, and C.M. Haga (1991) *Preliminary Data Analysis for the Lake Michigan Ozone Study.* Presented at Tropospheric Ozone and the Environment II: Effects, Modeling, and Control; Air and Waste Management Association International Speciality conference; November 4-7, 1991; Atlanta, Georgia. Also presented at the Air, Water, and Waste Technologies conference; Session III A; November 11-14, 1991; Cobo Center; Detroit, Michigan. STI-90212-113.

3. Hanna, S.R. and C. M. Haga (1992) *Meteorological Patterns Associated with High Ozone Concentrations During the Lake Michigan Ozone Study* To be presented at the Air and Waste Management Association 85th Annual Meeting, June 21-26, 1992, Kansas City, Missouri, paper 92-87.06.

4. Jeffries, H.E., K.G. Sexton, and J.R. Arnold (1988) *Validation Testing of New Mechanisms With Outdoor Chamber Data, Vol. 2: Analysis of VOC Data for the CB4 and CAL Photochemical Mechanisms.* Prepared for USEPA, Agreement CR-813107.

5. Lurmann, F.W.. H.H. Main (1992) *Analysis of the Ambient VOC Data Collected in the Southern California Air Quality Study.* Prepared for California Air Resources Board, STI-99120-1161-FR, November.

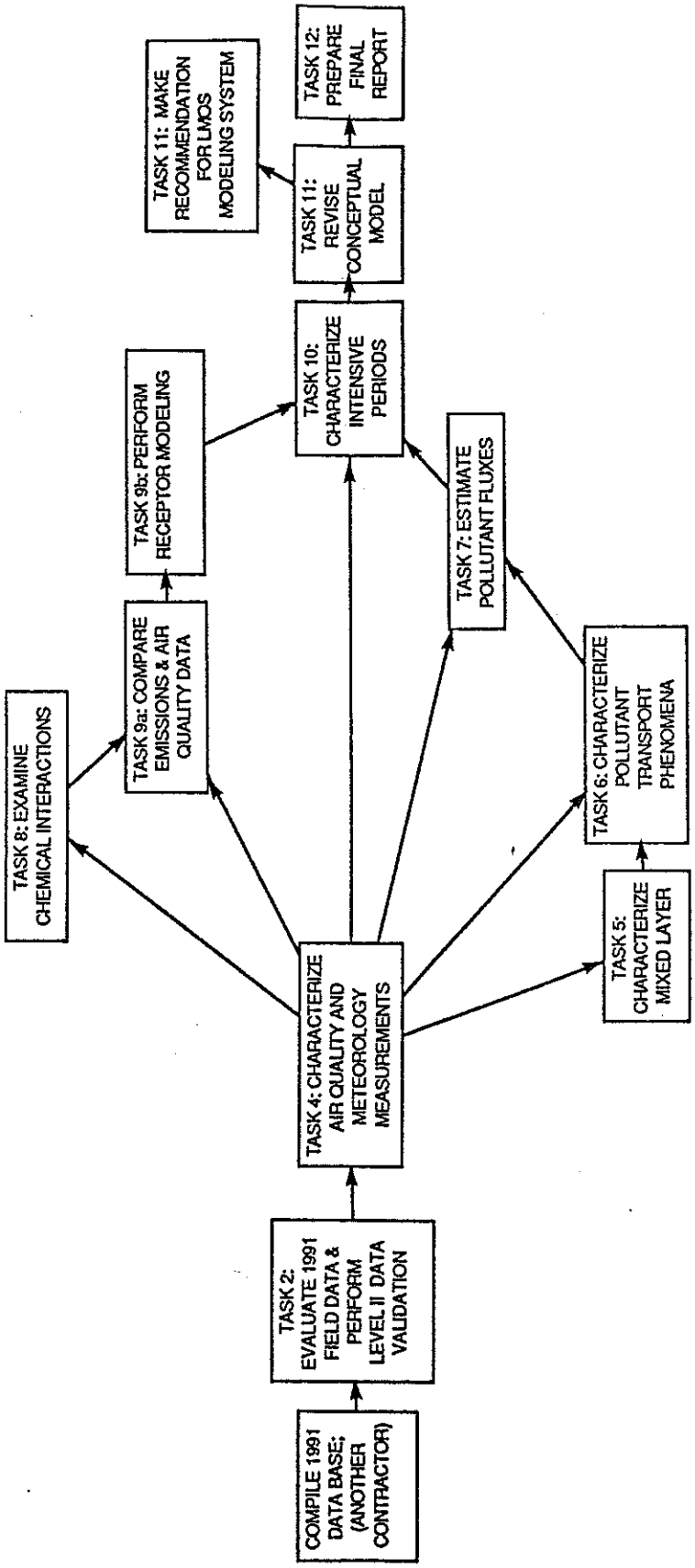

FIGURE 1. Flow Chart for Phase II LMOS Data Analysis tasks.

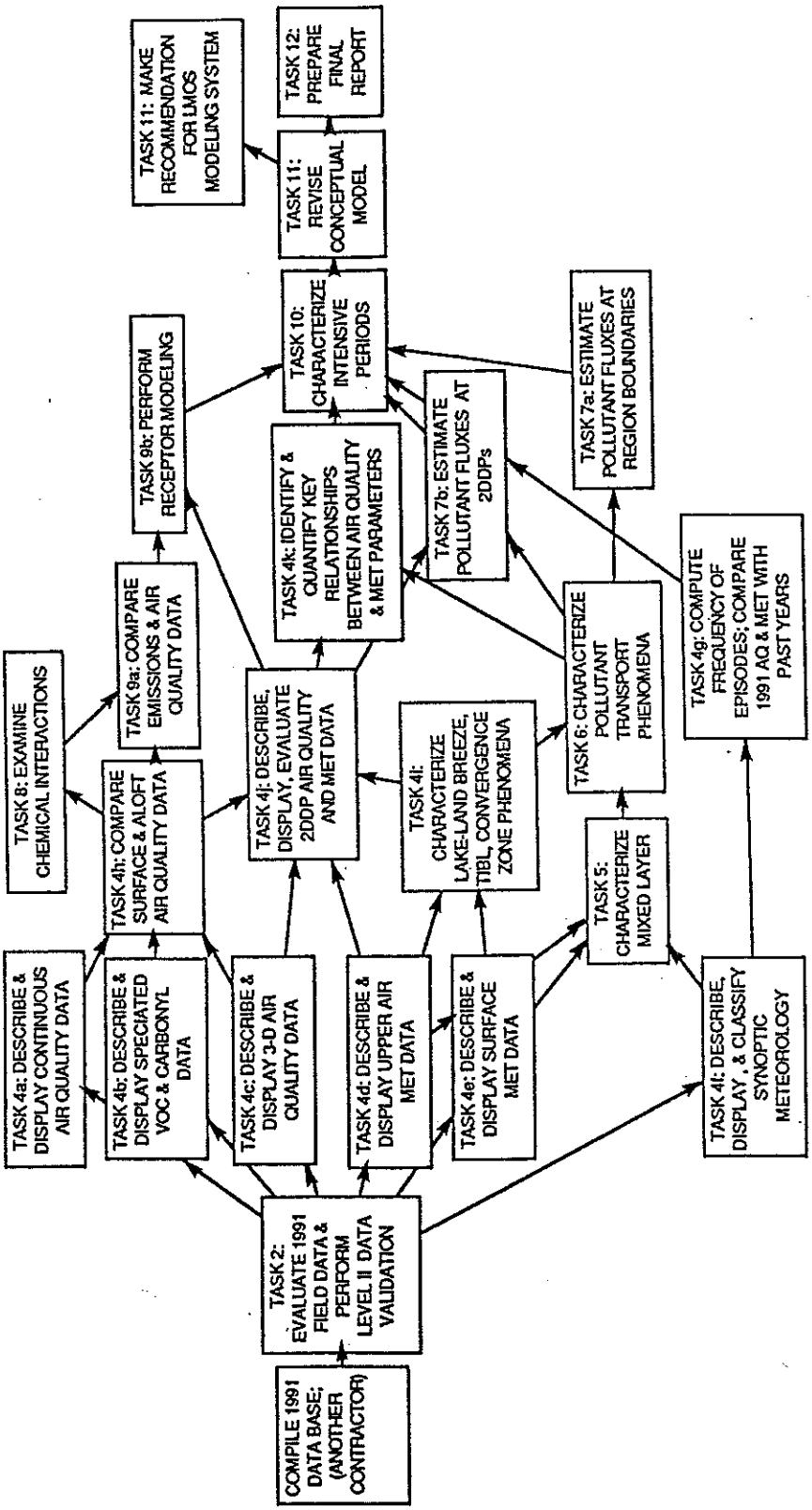

FIGURE 2. Flow chart for Phase II LMOS data analysis tasks (more detail).

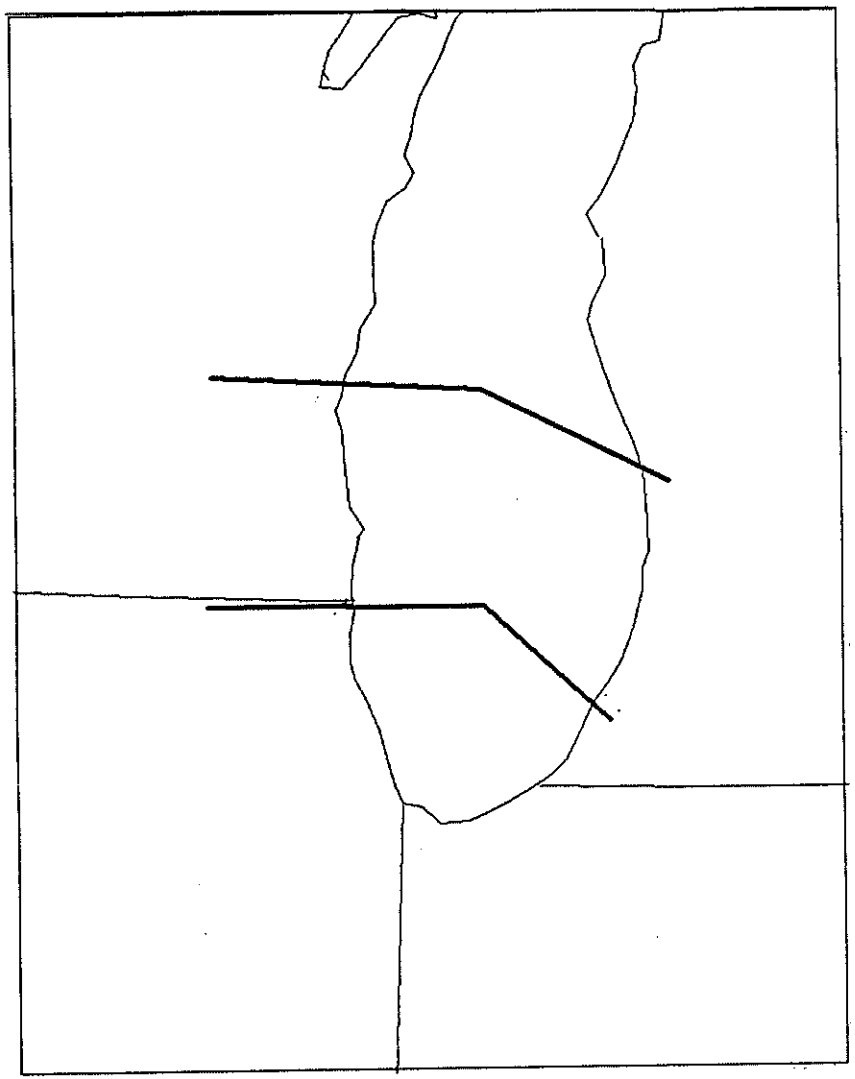

FIGURE 3. Map of the LMAQR showing the location of the 2-dimensional data planes (heavy lines).

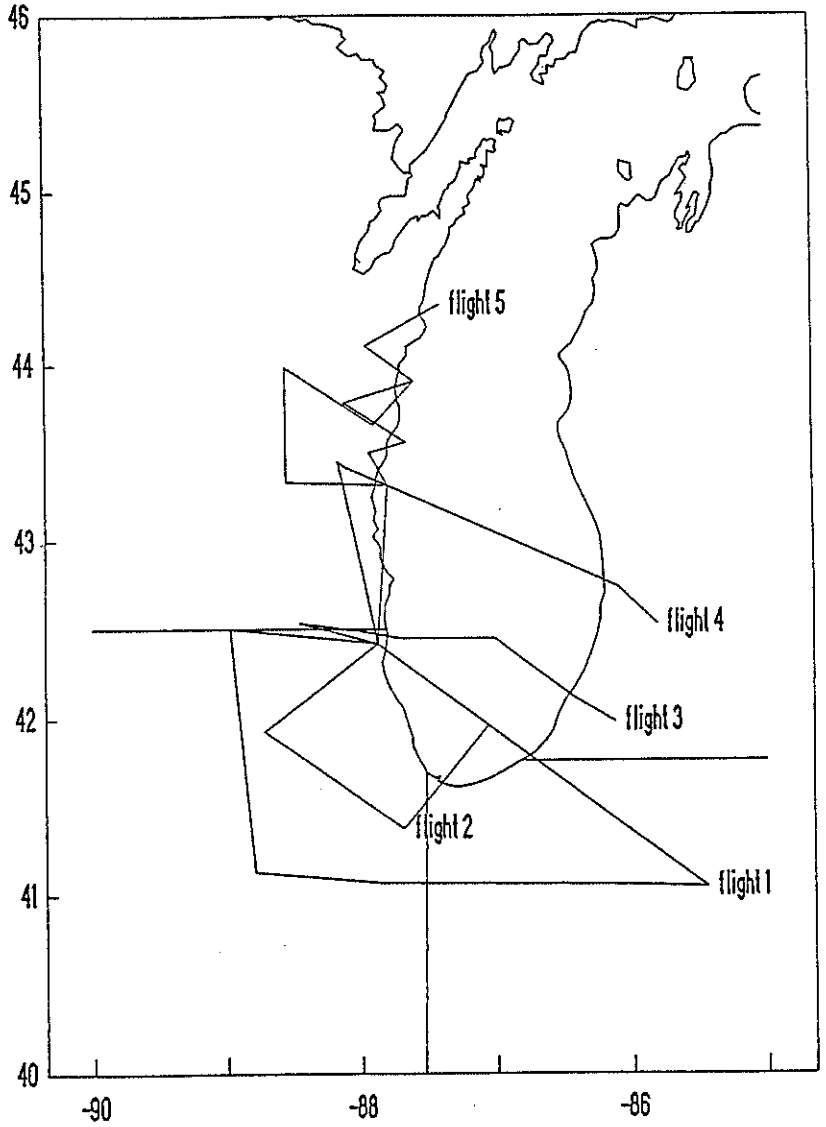

FIGURE 4. Sampling Aircraft Flight Patterns during LMOS.

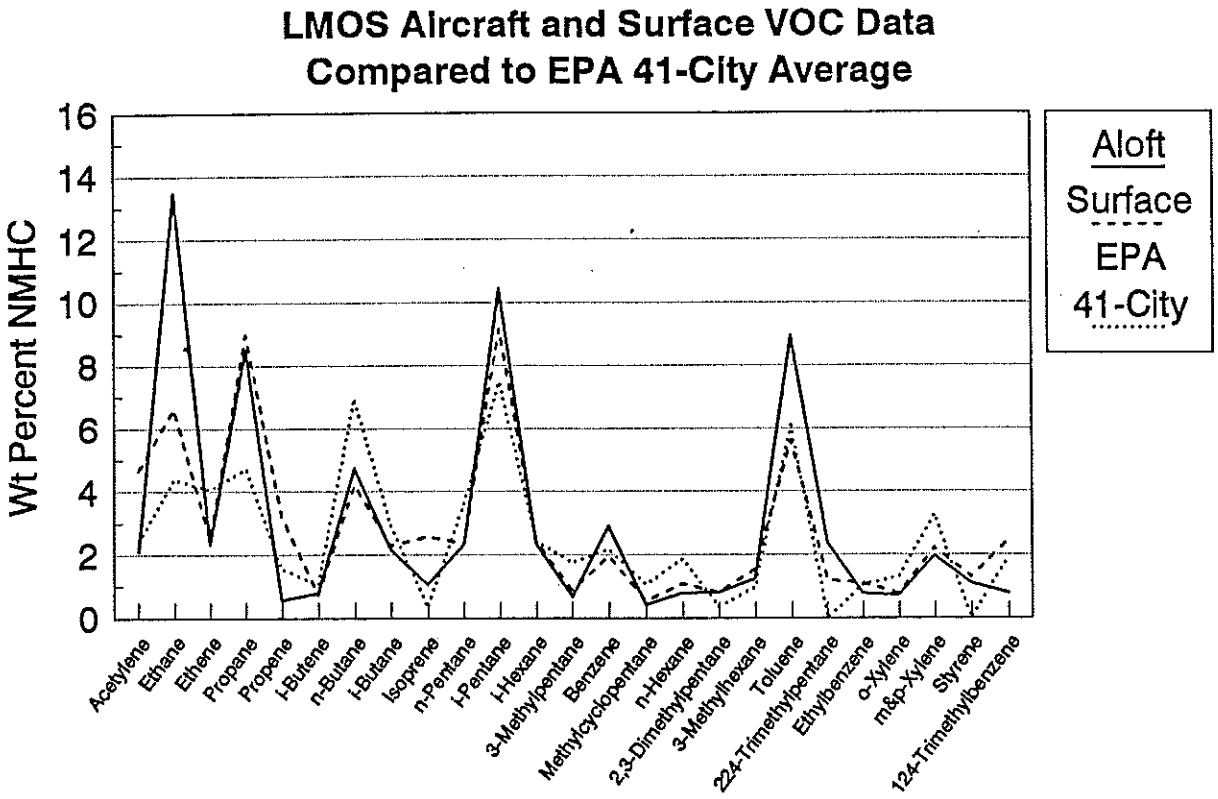

FIGURE 5. Mean percent NMHC composition of EPA 41-City 6-9 a.m. and LMOS aircraft and surface (all sampling times) data.

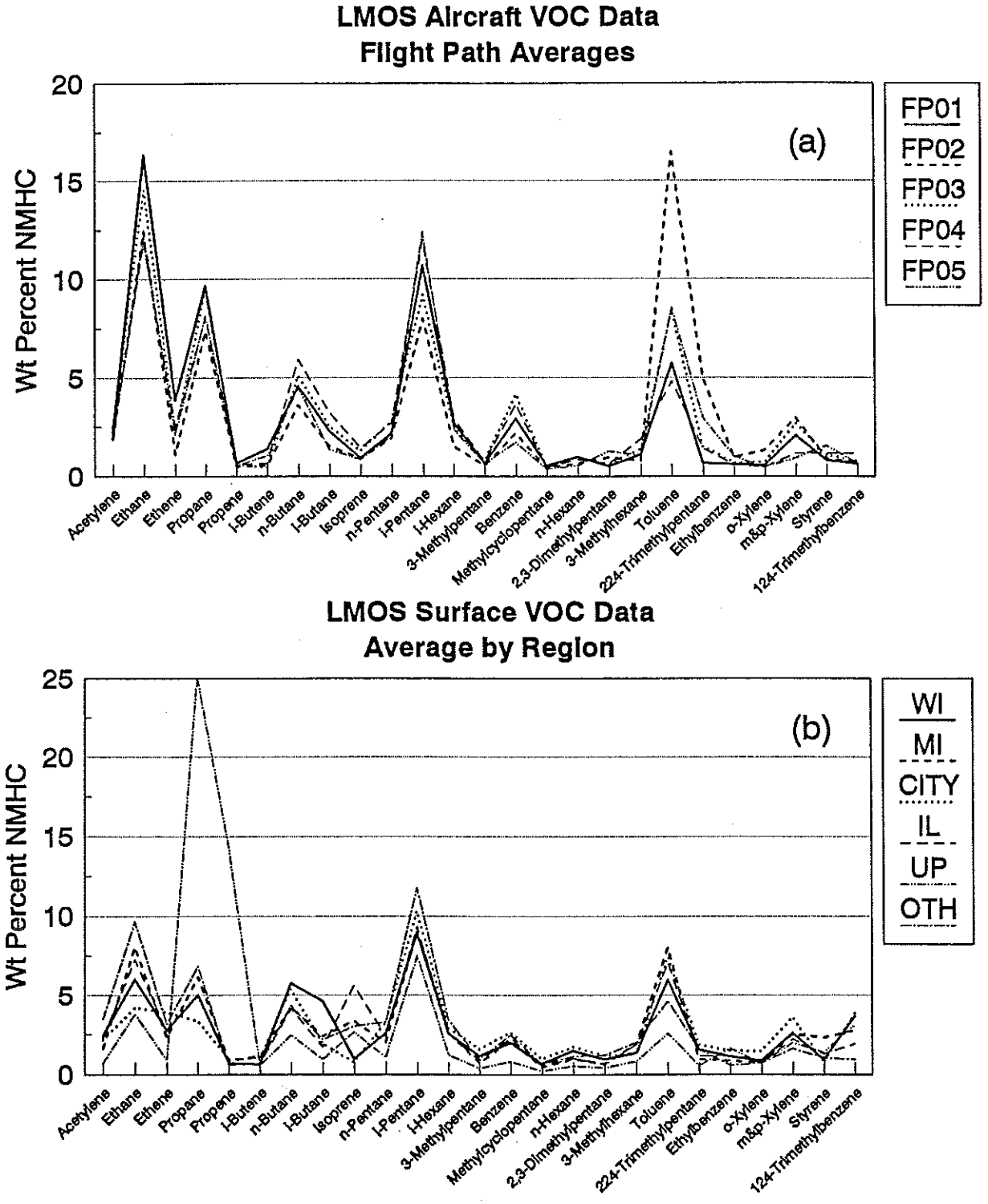

FIGURE 6. Mean percent NMHC composition of a) samples from each aircraft flight path and b) surface samples grouped by region.