Combined Use of Models and Measurements for Spatial Mapping of Concentrations and Deposition of Pollutants

by

Pamela Carisa Ambachtsheer

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Chemistry

Waterloo, Ontario, Canada, 2004

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

When modelling pollutants in the atmosphere, it is nearly impossible to get perfect results as the chemical and mechanical processes that govern pollutant concentrations are complex. Results are dependent on the quality of the meteorological input as well as the emissions inventory used to run the model. Also, models cannot currently take every process into consideration. Therefore, the model may get results that are close to, or show the general trend of the observed values, but are not perfect. However, due to the lack of observation stations, the resolution of the observational data is poor. Furthermore, the chemistry over large bodies of water is different from land chemistry, and in North America, there are no stations located over the great lakes or the ocean. Consequently, the observed values cannot accurately cover these regions. Therefore, we have combined model output and observational data when studying ozone concentrations in north eastern North America. We did this by correcting model output at observational sites with local data. We then interpolated those corrections across the model grid, using a Kriging procedure, to produce results that have the resolution of model results with the local accuracy of the observed values. Results showed that the corrected model output is much improved over either model results or observed values alone. This improvement was observed both for sites that were used in the correction process as well as sites that were omitted from the correction process.

Acknowledgements

I would like to thank Dr. J. Sloan for his guidance and support throughout my research. I would also like to thank Dr. R. Bloxam for his expert knowledge and advice.

Input data for the model were provided by Dr. A. Chtcherbakov, and Sunny Wong from the Ontario Ministry of the Environment. Technical support throughout my project was provided by Sunny Wong. Thanks go to both for their assistance.

I would also like to thank the other members of my committee, Dr. N. Reid, and Dr. K. T. Leung for their guidance.

Lastly I would like to thank my family, friends and Byung for their personal support these past three years. Dr. Jake Fisher also deserves special mention for his friendship and support.

Table of Contents

List of Tables

List of Figures

1 Introduction

In the last century, ozone concentrations in the troposphere have increased by a factor of three in the northern hemisphere due to human activities [Nielsen *et al.*, 1997]. Ozone, because of its ability to oxidize biological tissue, is a toxic air pollutant whose increasing concentrations in the troposphere are of great concern to atmospheric scientists. In 2002, according to Ontario's Ministry of the Environment, "Ontario's Ambient Air Quality Criteria (AAQC) for ozone were exceeded at 39 of 40 ambient air monitoring stations on at least one occasion." [Ontario Ministry of the Environment, 2002] This means that the ozone concentration throughout the summer of 2002 exceeded 80 parts per billion (ppb) at 39 monitoring stations — a sobering fact. At this ozone level, human health is notably affected since increased levels of ozone — the major component of photochemical smog — can cause or exacerbate diseases in the respiratory tract.

Yet, problems resulting from air pollutant emissions are not restricted to the atmosphere. The pollutants emitted to — and formed in — the troposphere also produce acids that deposit onto the earth's surface and eventually make their way into lakes, rivers and other waterways. These contaminants have a variety of effects, including the death of many water dependent organisms, as well as the tainting of water and animals ingested by the populace, which may cause further human health problems. Thus, reduction of the current levels of toxic emissions and ground level ozone is important not only to the health of fragile ecosystems, but to the human population as well. In order to protect the public from some of the harmful effects of ozone in the short term, accurate forecasts of high ozone levels will help those at risk to determine when they can safely be active outdoors in the summer.

While air quality forecasting will not directly solve pollution problems, this capability can play an important role in issuing timely health alerts in addition to providing important information to legislators who make decisions regarding emission controls. If models could predict accurately the effect on air quality of the various types and amounts of emissions and the influences of meteorology on these effects, then governments would gain the information they need to issue appropriate regulations, such as limiting the number of cars driven on days when

high vehicle emissions would result in especially poor air quality. This would limit the number of ozone episodes (periods of exceptionally high ozone concentrations) experienced by the regional population. Yet, in spite of the simplicity of the previous example, forecasting air quality and acid deposition is a complex problem.

Concentrations of ozone and other pollutants are predicted using regional chemical transport models that use meteorology, emissions, and geophysical data as inputs. These estimate the production, transport and destruction of primary and secondary air pollutants over the region. Primary pollutants are those that are emitted directly by anthropogenic or natural sources. Secondary pollutants are those that are chemically derived from the primary pollutants. Due to the complexity of the system, and the prohibitive computer resources needed, no model is accurate and efficient enough to produce data that are reliable enough for government agencies to depend upon them to warn the public of impending episodes or make meteorologically dependent emission legislation. As a result, scientists are currently studying a variety of ways to improve model output economically.

One approach to correcting model output is to use Four Dimensional Data Assimilation (FDDA). This method, which is used extensively in weather forecasting, corrects the model predictions using observed measurements. The observed data are assimilated with model output in the following four dimensions: the horizontal directions x and y, the vertical direction z, and lastly, time. The assimilation procedure "nudges" the model output to the correct values at the sites where measurements are taken, and the corrections are then interpolated over the regional domain, or grid. This system preserves the high resolution of the model output while also maintaining the absolute accuracy of the local measurements.

The assimilation can be used to correct final results or can be used to amend the initial – conditions file to be used in subsequent modeling. Accordingly, the model then uses the corrected output as the initial conditions for the next day's run. Moreover, because the initial

conditions have a significant impact on the quality of the model output, this correction, at least for weather forecasting, considerably improves the quality of the forecast.

The method was defined originally for meteorological modeling. It is not easily implemented for ozone modeling in the troposphere due to the lack of measured data for ozone concentrations. A comprehensive interpolation procedure that accurately estimates measurements between monitoring stations is used to overcome this difficulty. Another problem with the application of FDDA to ozone modelling is the absence of detailed measurements beyond the surface level (*i.e*. higher than about 15 m). Despite our focus on surface level ozone, vertical turbulence in the troposphere is significant and corrections made only in the surface layer would be lost within a small number of model iterations. In principle the model output for all vertical layers can be corrected by measuring vertical profiles or by vertical extrapolation of the surface measurements, but these are beyond the scope of this project, and therefore the interpolation of sparse surface data is the focus of this thesis.

Finding a suitable interpolation procedure for surface level ozone concentrations thus became the first step of this study. We chose a generally accepted procedure called Kriging, to approximate the ozone concentrations between monitoring stations. Next, we carried out the remaining steps in the assimilation of the observed data to improve the model predictions. The results were then statistically tested to determine their effectiveness in improving the surface level ozone profile for the Ontario region. To show statistically significant results, studying several months of data was necessary. Therefore, the spring and summer months — June through September — of 1996 were chosen for this research. The year 1996 was chosen simply because data were readily available for this year.

Our research entailed the computation of model output and compilation of observed measurements for the four months of the study. The differences between the model and observed results were then found and these were Kriged over the model domain. The Kriged difference file was then subtracted from the model results to produce corrected model output. This procedure not only maintained the local accuracy of the observed measurements, but also

retained the high resolution of the model output. This method was then statistically tested to measure the overall improvement of the profile in comparison to either the model output or measurements alone.

In addition to the important implications this project has with respect to efficient air quality forecasting, the interpolation procedure also has considerable value for Ontario Ministry of the Environment and Energy MOEE efforts to cut monitoring station costs. The results of this project prove that some monitoring stations can be eliminated without significantly degrading the quality of the measurement profile. Caution must be taken, however, when determining which stations will be removed, because some stations are important to the accuracy of the interpolation procedure. A sensitivity analysis must be completed in order to determine the effect each station has on the procedure. The method developed in this study has the potential to be easily modified to aid in the identification of key monitoring stations, thus identifying areas where money can be saved by station closures.

While the results of this study did not directly improve air quality forecasting capabilities, future work on the vertical extrapolation of surface level measurements will lead to a powerful four dimensional data assimilation procedure that could have the capability of drastically improving chemical transport model output, without significantly increasing requirements for computer resources. These advancements would then lead to a vast improvement in forecasting capabilities, thus improving the government's capability to issue health alerts due to poor air quality. Improved air quality forecasting also provides legislators the option of installing emission reduction regulations during times when the probability of ozone episodes is high, thereby maintaining high air quality. Both of these implications would lead to a decrease in health complications due to high ozone concentrations, thus improving Ontarians' overall quality of life.

2 Background

2.1 Previous Work

Four dimensional data assimilation has rarely been applied to atmospheric chemical transport models due to the lack of observed data for pollutants such as ozone. Surface level measurements are sparse, but in addition, measurements in the mid to upper troposphere are nonexistent. As a result, some satellite observations have been used; however, these measurements must be taken through the stratosphere. This affects the detail that can be attained for tropospheric ozone because of high concentrations of ozone in the stratosphere. The information obtained from satellite measurements is suitable for data assimilation with global model results [El Serafy et al. 2002; Lamarque et al., 2002], but it is not dense nor detailed enough to be assimilated with regional model outputs.

Numerous sources of error in regional models can occur, with imperfect input files being the most common. Consequently, scientists have started using data assimilation techniques to improve these inputs. In the following experiment the meteorology input was corrected using measurements. The models used in this experiment were MM5, a meteorology model developed by the National Center for Atmospheric Research (NCAR), and CALMET, a meteorology model that runs in conjunction with CALGRID, a photochemical grid model developed by the California Air Resources Board. MM5 data and observed meteorology were assimilated to produce accurate meteorological input for the chemical transport model CALMET/CALGRID [Barna and Lamb, 2000]. Twin experiments were done to show the benefit of data assimilation. The first run of the CALMET/CALGRID models, used as a reference used only MM5 data as meteorology input They then ran CALMET/CALGRID again, this time using as meteorology input the MM5 output assimilated with observed measurements. The model results were significantly improved after the assimilated (or in other words nudged) meteorological input was used. Therefore, by reducing the error in even one of the model inputs, the model output can be considerably enhanced.

Emission inventories are most widely believed to be the largest source of uncertainty in chemical transport model inputs. Thus, techniques to improve emissions inputs are beneficial to regional atmospheric modeling. To this end, a procedure by which existing emissions

inventories could be augmented with ambient measurements was designed by [Dominguez and Russell, 2001]. This study concluded that the model values were closer to measurements when ambient emissions measurements were used to correct emissions inventories.

Aside from nudging meteorology or emissions input, correcting initial concentrations of pollutants can also improve model results. In fact, modified initial concentrations of a secondary pollutant can be used to correct the concentrations of its precursors as well by completing some back calculation [Elbern and Schmidt, 2001]. Because initial concentrations of pollutants have influence on the model results for up to 40 hours of simulation, having good estimates for these values is important. In other words, if the initial concentrations are more accurate or more detailed, the model is more likely to produce good results. An experiment was conducted by Elbern and Schmidt whereby the full effect of initial conditions on model results were tested. It was important to choose a pollutant for which a great deal of data were available, in order to verify model results. The most convenient pollutant for such an experiment, therefore, was ozone. This experiment is very relevant to the work done in this thesis, so a detailed explanation is necessary.

An experiment was conducted whereby the first six hours of the simulation were each assimilated with corresponding observed ozone measurements. These were then used to nudge the concentrations of ozone precursors. Subsequently, the model was allowed to simulate 18 hours past the assimilation period. To test this technique, the model was first run with *detailed initial values* to produce "observed" or reference values for each model grid point. The model was then run with *general initial values* to get a "first guess run". Consequently, the first guess run followed the trend of the reference run for each pollutant of interest, but the values differed significantly, as expected [Elbern *et al.*, 1997]. After assimilating ozone alone for the six-hour assimilation period, the model produced values that were quite similar to the reference run for ozone, NO_2 and NO. In a later experiment, measured ozone was used for the assimilation period [Elbern and Schmidt, 2001] The technique continued to produce superior results for ozone as well as improved results for NO_2, HO, HO_2, HONO, HNO_4, PAN and HCHO.

Elbern's experiment was run in Europe with 400 observation sites. By using data from all sites, the six-hour assimilation period at the start of the model simulation was shown to improve results for up to 42 hours. This experiment also showed that removing the data from half of the available monitoring stations was not significantly detrimental to the improvements. Moreover, that the model results were improved for up to 42 hours of simulation is noteworthy, since the current goal for air quality forecasting models is to produce accurate forecasts for up to three days. This technique, then, could significantly improve the performance of these models. It was also noted, however, that with fewer stations, improvements would deteriorate quickly. For our project, there were 117 monitoring stations. In addition, the regional grid that was used for this project was significantly larger than the grid used in Elbern et. al.'s experiment. Therefore, more work needs to be done to improve this technique in order to obtain the results of Elbern *et al.* for the circumstances of our study.

While some related work has been done on improving model output, using observed data to improve model results is a relatively new area of study. Accurately nudging chemical transport model results is difficult, due to the need for extensive and detailed observational data. With the use of a good interpolation procedure, however, we hope to reap the benefits of data assimilation with only sparse observed measurements.

2.2 Modeling Process

Many different types of atmospheric models have been written to address a variety of scientific questions regarding physical and chemical processes in the troposphere. For example, one may study the fate of an air parcel, such as a plume from a smokestack, or instead focus on the general air quality over a specific region. Scale is also a question: studying global, regional, or local trends in pollutant concentration are all possible, but a different type of atmospheric model must be applied for each type of study.

A Lagrangian model is used to study the fate of air parcels because of its ability to follow the trajectory of the plume as well as its dispersal rate. Accordingly, weather patterns, wind speed, stack height, and emission rates are important inputs to these models. One can thus learn how far elevated concentrations of pollutants are transported before they are dispersed.

Furthermore, one can use the trajectory information to find the geographic locations where pollutants are transported. As a result, municipal and provincial governments can then study the plumes of major emission sources in areas within and surrounding their borders to determine the major sources of local pollution. These types of observations allow elected officials to make the most effective policies to control local pollution.

Concentration distribution represents another concern in air quality. Thus, instead of studying the fate of individual air parcels, one can study the concentration distribution of pollutants over a selected area. Eulerian models are used for this purpose. For these, the modeler first defines a grid over a region of interest, and next calculates the concentration for each grid square over a defined time period. Eulerian models must consider both mechanical transport and chemical processes. Thus, meteorological, emissions and geophysical data, as well as boundary and initial conditions, are important inputs to Eulerian models.

Within the category of Eulerian models, there are again several types, each defined by scale. Global models are one such example—these types study global trends in pollutant concentration distribution. Since the study region encompasses the entire planet, the resolution of such models is quite poor, with grid squares often as large as ten degrees latitude by ten degrees longitude. The poorer resolution necessitates longer time periods in order to record significant concentration changes in each grid cell. Consequently, these models are useful only in defining global transport of long-lived pollutants.

In order to refine the resolution, a smaller study region must be chosen. Regional Eulerian models have a grid square size ranging from 10 km x 10 km to 150 km x 150 km. The resolution of regional models is thus significantly better than that of global models. As a result, the time steps, (i.e. the time over which the model assumes conditions remain constant), can also be shortened, thus allowing for the study of shorter-lived pollutant species. Regional models are therefore most widely used to study pollution episodes. These are short time periods - three to seven days - during which elevated concentrations of pollutants are experienced. The emissions input file of the model can be modified to examine the result of deleting one or more emission sources, such as coal fired power plants. National or provincial

governments can also use regional models to monitor the pollution concentration distributions over their political domains.

The smaller grid size of regional models, however, introduces the need for horizontal boundary conditions. This problem is often overcome, by extracting approximate boundary values from the output of a global model. Models with nested grids have also been written to address this issue. The smallest grid of a model with nested grids would have the finest resolution, and its boundary conditions would be determined from the next larger grid.

Despite their improved resolution over global models, regional models are still limited in detail. They cannot, for example, be used to study the effects of rush hour traffic on the air quality at major intersections. As a result, microscale models have been developed for this type of study. For example, microscale models are used to study the pollution levels in areas next to roadways or intersections.

For the purposes of this project, a regional model named ADOM was applied. The model domain was centered over Windsor. It encompassed much of Canada and the USA. The grid squares were 36 km x 36 km, while the grid was 75 grid squares east/west, and 72 grid squares north/south. The grid also included 11 vertical levels where each level height was determined by air pressure values. This system is called the sigma coordinate system. There are two versions of this system, the sigma p system and the sigma z system. The sigma p system is defined directly by pressure and the levels are defined in the following way:

$$\sigma = \frac{p - p_{top}}{p_{s_0} - p_{top}} \qquad (1)$$

with p_{top} a specified constant "top of the atmosphere" pressure and p_{s_0} a reference state pressure that is constant in time but varies with terrain height. Because the heights are determined by pressure, the height of each layer varies with pressure. The sigma z system, on the other hand, is based on this system, but the physical heights are predefined. ADOM uses a sigma z coordinate system, which will be discussed in more detail in Chapter 4.

A Lambert Conformal Conic projection was used to fit the curved surface of the earth under a flat grid. Having an accurate map corresponding to the grid is important to analyze the model results.

The input data necessary to run the model for this project were provided from the Ontario Ministry of the Environment. Modelers at the MOEE modified MM5 output to provide meteorology input for ADOM [Chtcherbakov, 2002], while emission input was provided from a 1995 emissions inventory [USEPA, 2002a]. Detailed emissions data were, until recently, only compiled every five years. Therefore, given the lack of available emissions data, these values were deemed accurate enough for our purposes.

In summary, several types of atmospheric models exist: Lagrangian models study the fate of air parcels, such as smokestack plumes, while Eulerian models simulate pollution concentrations over a given domain for a series of grid cells. Moreover, several types of Eulerian models—global, regional, and microscale, to name a few—exist as well. For the purposes of this project, we used a regional model named ADOM, with input provided by the MOEE.

2.3 Ozone in the Troposphere

Ozone is highly toxic because of its ability to oxidize biological tissue. It is thus one of the most important gases to consider in the study of tropospheric air pollutants [Jacob, 1999]. Ozone is not directly emitted through anthropogenic sources but is a major component of smog – a phenomenon found mostly in urban areas. In fact, ozone is known as a secondary pollutant since it is the product of several photochemical reactions. Thus, ozone has been a favored topic of study among atmospheric scientists for many years. Accordingly, the chemistry of ozone in the troposphere is well known. This gas was chosen as the pollutant to study for this project because of its well-understood chemistry and its importance in air quality.

2.3.1 Sources of Ozone

Originally, the major source of ozone in the troposphere was thought to be transport from the stratosphere. Though this is indeed a source, the transport rate is only in the range of 1.5×10^{13}

moles per year — an amount that accounts for approximately 10 – 12% of ozone measured per year [Jacob, 1999]. Subsequently, it was discovered that NO_x was linked with O_3 production. Now it is known that ozone production is formed in several chain reactions that included the production of NO_2 from NO and $HO_2\cdot$. For example, the following chain results in the production of ozone.

$$CO + OH\cdot \xrightarrow{O_2} CO_2 + HO_2\cdot \qquad (C1)$$

$$HO_2\cdot + NO \rightarrow OH\cdot + NO_2 \qquad (C2)$$

$$NO_2 + h\nu \xrightarrow{O_2} NO + O_3 \qquad (C3)$$

Therefore, the net reaction from this mechanism is then,

$$CO + 2O_2 \rightarrow CO_2 + O_3. \qquad (C4)$$

Critical to this chain is the concentration of NO. With low concentrations of NO, ozone is destroyed in the following way:

$$NO + O_3 \rightarrow NO_2 + O_2 \qquad (C5)$$

Furthermore, lower concentrations of NO means that there is not enough NO to compete with the $HO_2\cdot$ radical which then also reacts with ozone leading to its destruction. At higher concentrations of NO, however, NO reacts with $HO_2\cdot$ leading to the production of ozone [Brasseur *et al.*, 1999]. Therefore, the efficiency of this chain depends on the abundance of NO_x. This mechanism accounts for approximately 70% of the chemically produced ozone in the troposphere [Jacob, 1999]. Also, the diurnal variation seen in the concentration of tropospheric ozone is partially due to the sunlight required for the photolysis of NO_x in the third step of this chain.

Since ozone is a secondary pollutant, it is relevant to identify the sources of the primary ozone precursors. NO_x, the most important, is emitted mostly by anthropogenic sources. Approximately 46% of NO_x emitted to the atmosphere comes from fossil fuel combustion, with the burning of biomass accounting for a further 26%. Additional sources include soil emissions, lightning, NH_3 oxidation, and, to a lesser extent, aircraft emissions and transport from the stratosphere [Jacob, 1999].

While NO_x is critical to the production of ozone, so too are the volatile organic compounds (VOCs) that react with NO to produce NO_2. One of these is methane. Methane emissions lead to the production of ozone by the following chain of reactions, starting with its oxidation by $OH\cdot$.

$$CH_4 + OH\cdot \rightarrow CH_3\cdot + H_2O \qquad (C6)$$

The methyl radical then reacts instantaneously with O_2 to yield the methyl peroxy radical $CH_3O_2\cdot$, as follows:

$$CH_3\cdot + O_2 + M \rightarrow CH_3O_2\cdot + M \qquad (C7)$$

where M is a third body . The methyl peroxy radical can then react with NO to produce NO_2 and a methoxy radical.

$$CH_3\cdot + NO \rightarrow CH_3O\cdot + NO_2 \qquad (C8)$$

The production of the methoxy radical leads to the production of $HO_2\cdot$ through a reaction with O_2 in the following way.

$$CH_3O\cdot + O_2 \rightarrow HCHO + HO_2\cdot \qquad (C9)$$

The resulting $HO_2\cdot$ then reacts with NO to produce further NO_2. NO_2 then photolyses in the presence of a third compound to produce ozone as in the first chain. Therefore, the net reaction from this chain is:

$$CH_4 + 4O_2 + 2\,h\nu \rightarrow HCHO + 2O_3 + H_2O \qquad (C10)$$

Since methane is the primary precursor to the formation of methyl peroxy radicals, it is important to discover the sources of methane. Though both natural and anthropogenic sources of methane exist, the major natural sources, including wetland and termite emissions, make up only 30% of methane emissions. In contrast, major anthropogenic sources, which include the burning of natural gas as well as emissions from livestock and rice paddies, far outweigh the natural sources, making up 70% of methane emissions.

Methane is one of the most important VOCs that are critical to the production of ozone, but other VOCs also produce peroxy radicals that react with NO_x in the presence of sunlight to produce ozone. These volatile organic compounds are emitted through solvent use, and by vehicles, residential buildings, and surface coatings. Moreover, Isoprene — a compound released by plants — is also a VOC. In some regions; its emissions can constitute a large percentage of the VOCs present.

Ozone, therefore, is produced in industrial and heavily populated regions as well as over busy roadways. From its sources, it is transported to more remote regions. Presumably one would find the highest tropospheric concentrations of ozone in urban or industrial regions since it is largely produced there, yet remote locations, such as Long Point, also experience high ozone concentrations. Furthermore, rural regions experience longer-lived ozone episodes than their urban counterparts. The reason for this can be found in how ozone is destroyed.

2.3.2 Sinks of Ozone

Ozone's removal from the troposphere is another area that needs explanation. Ozone is eliminated from the troposphere either chemically, or through dry deposition. Dry deposition is the settling of chemical compounds onto the earth's surface, be that surface dry or wet. This

is in contrast to wet deposition, which is the deposition of compounds during precipitation. Chemical reactions account for 88% of ozone loss [Jacob, 1999]. Therefore it is important to study the chemical mechanism for ozone loss.

Several reactions cause the chemical loss of ozone from the troposphere. The principle ozone sink is photolysis as in the following series of steps. It accounts for approximately 75% of tropospheric ozone loss by gas-phase routes [Seinfeld and Pandis, 1998].

$$O_3 + h\nu \rightarrow O(^1D) + O_2 \quad (C11)$$

$$O(^1D) + M \rightarrow O + M \quad (C12)$$

$$O(^1D) + H_2O \rightarrow 2OH\cdot \quad (C13)$$

Chemical loss of ozone also occurs through the following reaction:

$$OH + O_3 \rightarrow HO_2 + O_2. \quad (C14)$$

Removal of ozone in urban areas, however, is also greatly affected by high concentrations of NO_x. NO contributes significantly to the destruction of ozone in urban areas in a process called NO scavenging, which occurs by the following reaction.

$$NO + O_3 \rightarrow NO_2 + O_2 \quad (C15)$$

Note that sunlight is not required for ozone loss from the troposphere.
Ozone has a longer lifetime in rural regions due to the lack of NO. It is largely because of this that rural regions that are downwind from industrial centers can have longer-lived ozone episodes.

Slow ozone destruction in the spring and summer also occurs over large bodies of water. During these months, regions such as the Great Lakes experience high ozone concentrations with respect to the surrounding regions. The reasons for this are threefold. First, because ozone is not very soluble in water, ozone does not readily dry deposit into the Great Lakes or into the Atlantic Ocean. Meteorology also plays a major role in the higher ozone concentrations. Because the surface of the Lakes and the Ocean is cool in the spring and summer, the temperature profile over these bodies of water is stable with respect to thermal inversion. This leads to slow vertical mixing, which means that any ozone that is transported over the lakes does not readily react and so does not get destroyed [Bloxam *et al.*, 1993]. Also because of the slow vertical mixing NO_x tends to stay close to the surface. Because NO_x is not dispersed vertically, it exists at the surface in higher concentrations than would be found over land, causing much more ozone producing photolysis to occur [Bloxam *et al.*, 1993]. This phenomenon is important for our project since the model domain contains the Great Lakes and some of the Atlantic Ocean.

In summary, ozone is a secondary pollutant whose precursors have largely anthropogenic sources. In the presence of sunlight, NO_x photodissociate in the presence of VOCs to produce ozone. Also, ozone is lost from the troposphere through dry deposition, photolysis and reactions with HO_2 OH and NO.

2.4 Project Components

The necessary components of this project consist of the chemical transport model, including its input data, hourly observed measurements, and an interpolation procedure. Some introduction to these components is required, but much of the detail will be discussed in the following chapters.

From this project's inception, a great number of regional chemical transport models were available for use; however, many considerations must be taken when choosing an appropriate model. For instance the input data must be readily available. Additionally, the size of the input and output data files can limit model choices. Furthermore, installing and learning how to run each model takes time, since user-friendly packages are rare. There are therefore

distinct advantages to choosing a model that is accessible and well explained. For our project, there was a further constraint due to the many months of simulation required to produce relevant results, namely model speed and efficiency. As a result, for the purposes of this project, the model ADOM best fit all the requirements.

The assimilation process requires both model data and observed measurements; thus the next important component of the project was the collection of hourly measurment data. The required observed measurements could not be obtained from a single source, because our domain encompasses parts of both Canada and the U.S.A. Fortunately national networks exist in both nations that store historical hourly ozone measurements. For the Canadian data, the measurements from the NAPS (National Air Pollution Surveillance) database were used [Dann, 2002], while the CASTNET (Clean Air Status and Trends Network) database was used for the U.S. data [USEPA, 2002b]. Hourly measurements for the summer of 1996 were used to correct model output.

An interpolation procedure was essential to accomplishing this task because of the sparseness of the monitoring stations. For regional tropospheric ozone modeling, Kriging is a generally accepted interpolation procedure [Elbern and Schmidt, 1999]. It was chosen to interpolate observed data over the grid because it uses a sophisticated weighted average to find unknown concentrations. The weights are determined in a fashion that minimizes the estimation variance. First, semivariances of known concentrations are calculated and graphed with respect to the distance between the points. A function called the semivariogram is then fitted to the set of points and the weights are calculated using that function. The contour graphing software package Surfer contained a Kriging procedure, which facilitated this step of the project. Kriging was also used effectively to interpolate the differences between model and measured data over the grid.

To conclude, for this project, the model ADOM was chosen as the regional chemical transport model. While there were more accurate models to choose from, the speed and efficiency of ADOM were particularly required for a project that requires many months of simulation. In addition, hourly measurement data were obtained from the NAPS and CASTNET databases to

assimilate with the model output. The sparseness of the monitoring stations in these networks made necessary an effective interpolation procedure. Kriging was chosen for this purpose because of its advanced weighted average system of calculation. These three main components of the project will be discussed in more detail in the following chapters.

3 ADOM (Acid Deposition and Oxidant Model)

3.1 General Parameters of ADOM and the Model Domain

ADOM is an Eulerian long-range transport model developed in the 1980s for the Ontario Ministry of the Environment, The Canadian Atmospheric Environment Service and the Umweltbundesamt of the Federal Republic of Germany [Scire J.S. *et al.*, 1986]. It was designed to estimate concentrations and deposition of the major atmospheric pollutants, such as ozone and NO_x, over a medium scale grid. In our case, each grid square was 36 km by 36 km in size. The corresponding model domain was centered approximately over Windsor (at 40˚latitude and 100˚ longitude) and had 75 grid squares East/West, and 72 North/South. Additionally, there were twelve vertical layers to encompass the troposphere over the region, where the physical height of each layer was predefined using the sigma z coordinate system. The physical heights of the layers were 1.0m, 56.1938m, 135.8103m, 250.6563m, 416.3205m, 655.2896m, 1000m, 1497.2413m, 2214.5072m, 3249.1559m, 4741.6265m, 6894.5006m, and 10000m respectively.

The regional model ADOM was chosen for this project mainly because of its speed. Since we needed several months of model output, speed was of utmost importance, as long as accuracy was not significantly sacrificed. ADOM processed twenty-four hours of data in twenty minutes while Models-3 CMAQ, the most popular North American chemical transport model, processed the same amount of data in two hours on the same computer. While the newer models could have predicted ozone concentrations more accurately than ADOM, the project's success was not hindered since only the interpolation and assimilation procedures were being tested. Moreover, these could later be adapted to work with other models.

Once the model had been chosen, the next step was to acquire the necessary input file. These were provided by the MOEE. A lot of time and effort was spent on processing the meteorology, emissions and geophysical files, and this work deserves some attention.

3.2 Input Files for ADOM

The three largest input files of most chemical transport models are the meteorology, emissions and geophysical files. For these, raw data must be acquired and processed to produce files that can be used by the model. For our project, Dr. Andrei Chtcherbakov at the MOEE did much of this work with help from Dr. Robert Bloxam and Sunny Wong.

The raw meteorology files were acquired first. These are the most difficult files to obtain because of the meteorological detail that is required by ADOM. Data that were produced using MM5 were attained from the U.S. Environmental Protection Agency (USEPA). This grid for this data set was centered over 40˚latitude and 100˚ longitude. It was this data set that determined the grid that would be used in this project. Some processing was then required to produce files that were compatible with ADOM. First, the MM5 output fields were slightly different than those required for ADOM. To acquire all the necessary fields, the MM5 output was processed using a sub-procedure of CMAQ called MCIP (Meteorology-Chemistry Interface Processor), and the correct fields were then taken from this output. There was a fundamental difference, however, between the grid type used by MM5 and CMAQ, and that of ADOM. ADOM uses a sigma z coordinate system for the vertical layers, while the other two use sigma p levels. Therefore some interpolation of the vertical data was required to produce a file that was compatible with ADOM. This interpolation necessitated a mass conservation step to ensure that nothing was gained or lost during the conversion. Once this step was completed the resulting data file could be used as meteorology input to ADOM.

The emissions input files were the next to be produced. The national averages for individual anthropogenic point source emissions (such as factories), area source emissions (such as residential neighborhoods) and mobile sources (such as roadways), were acquired from the USEPA [USEPA, 2002a]. The US national averages were produced by the USEPA, and Environment Canada produced the Canadian averages. Next, twelve meteorologically representative days were chosen for the year, three from each season. For each season, one weekday, one Saturday and one Sunday were selected as representative days. It was important to choose meteorologically representative days because meteorology significantly affects the

concentrations of primary pollutants. The emissions, with the corresponding meteorology from the selected days were then processed using MEPPS (Models-3 Emissions Processor and Projection System), to produce daily emissions files. These were not yet compatible with ADOM. The VOCs in the emissions files produced by MEPPS were not speciated compatibly with ADOM requirements. MEPPS speciates VOCs in a way that is compatible with RADM (Regional Acid Deposition Model), as well as other models such as CMAQ. RADM's mechanism includes 15 VOC compounds (or groups of compounds) assembled by condensing the list of VOCs produced by the National Acid Precipitation Assessment Program (NAPAP). ADOM's VOC mechanism is also derived from NAPAP's list, but condenses the list further to only 11 compounds or compound groups. A mapping from RADM requirements to ADOM requirements was written by Dr. R. Bloxam at the MOEE to convert the data file to be compatible with ADOM. This mapping required a mass conservation test to be done, again to ensure that nothing was gained or lost during this procedure. This completed the anthropogenic emissions input. The biogenic emissions, which depend on meteorology and land use, were compiled by Dr. Chtcherbakov. A model called Biogenic Emissions Inventory System 2 (BEIS2) was run to produce hourly biogenic emissions data for the year. These emissions were then added to the anthropogenic emissions file to complete these input files. This was the last step in producing daily emissions files that were compatible with ADOM requirements.

The last major input file required by ADOM was the geophysical file. This file contains information such as land use for the model domain. For our project, the geophysical file was produced by modifying an existing file from a previous project. The existing domain, however, was centered over 40˚ latitude and 90˚ longitude. This map was converted using a procedure developed by the US Geological Survey. The map coordinates were converted back to latitude and longitude, and then re-projected to be centered over 40˚ latitude and 100˚ longitude. Lastly, because ADOM uses fewer land use categories than the model that was used for the prior project, the fields were modified to be compatible with ADOM requirements. This file was then visualized using a software package called PAVE (Package for Analysis and Visualization of Environmental Data), and checked visually to ensure that the land use and other geophysical features were correct. This fulfilled the last major input requirement for

ADOM. To understand how ozone is simulated in ADOM, the gas-phase chemistry module will have to be examined.

3.3 Gas-Phase Chemistry Mechanism of ADOM

In this module of ADOM, the complex photochemistry and related reactions were simulated to calculate the production of secondary pollutants such as ozone. The chemical mechanism included in ADOM was based on a mechanism developed in 1986 by Lurmann and Lloyd, which was condensed from a detailed mechanism including nearly 300 reactions among 100 species. In our study, however, ADOM's gas-phase chemistry mechanism included roughly 100 reactions among approximately 50 chemical species. The mechanism is listed below in Table 1, but some explanation is necessary.

Numerous techniques were employed in order to condense drastically the detailed chemical mechanism without losing important details [Lurmann and Lloyd, 1986]. First, organic species with similar chemistry were combined. Subsequently, unimportant reaction pathways were removed from the mechanism. Next, the concentrations of relatively inert species were given constant values, while irrelevant stable species were removed from the species list. In addition, variable stoichiometric coefficients were employed which means that terminally bonded alkenes represented by propene, the internally bonded alkenes represented by *trans*-2-butene, and biogenic alkenes such as isoprene could be treated as a single lumped $\geq C_3$ alkene [Lurmann and Lloyd, 1986]. Lastly, a steady state approximation was used to estimate concentrations for species with rapid production and destruction rates. These steps allowed the species list to be reduced by half, and the reaction set by two thirds. Accordingly, listed in the following table are the reactions relevant to ozone, along with their corresponding rate constants. The mechanism in its entirety can be found in Appendix A.

Table 3.1 Most important ozone related reactions in ADOM's gas-phase chemistry mechanism

	Reactions	Rate Constant (cm^3 molecule s units)
(R1)	$NO_2 + h\nu \rightarrow NO + O_3$	radiation dependent
(R2)	$NO + O_3 \rightarrow NO_2 + O_2$	$2.2 \times 10^{-12} e^{-1430/T}$
(R3)	$NO_2 + O_3 \rightarrow NO_3 + O_2$	$1.2 \times 10^{-13} e^{-2450/T}$
(R4)	$NO + NO_3 \rightarrow 2NO_2$	$8.0 \times 10^{-12} e^{250/T}$
(R5)	$NO_2 + NO_3 \rightarrow N_2O_5$	PT dependent
(R6)	$N_2O_5 \rightarrow NO_2 + NO_3$	special function
(R7)	$NO_2 + NO_3 \rightarrow NO + NO_2 + O_2$	$2.5 \times 10^{-14} e^{-1230/T}$
(R8)	$NO_3 + h\nu \rightarrow 0.15NO + 0.85NO_2 + 0.85O_3 + O_2$	$3.29k_1$
(R9)	$NO_3 + HO_2 \rightarrow HNO_3 + O_2$	2.5×10^{-12}
(R10)	$O_3 + h\nu \rightarrow 2OH$ (H_2O dependent)	special function
(R11)	$NO + OH \rightarrow HONO$	PT dependent
(R12)	$HONO + h\nu \rightarrow NO + OH$	$0.205k_1$
(R13)	$NO_2 + OH \rightarrow HNO_3$	PT dependent
(R14)	$HNO_3 + h\nu \rightarrow NO_2 + OH$	radiation dependent
(R15)	$HNO_3 + OH \rightarrow NO_3 + H_2O$	$9.4 \times 10^{-15} e^{778/T}$
(R16)	$N_2O_5 + H_2O \rightarrow 2HNO_3$	1.3×10^{-21}
(R17)	$CO + OH \rightarrow HO_2 + CO_2$	special function
(R18)	$O_3 + OH \rightarrow HO_2 + O_2$	$1.6 \times 10^{-12} e^{-940/T}$
(R19)	$NO + HO_2 \rightarrow NO_2 + OH$	$3.7 \times 10^{-12} e^{240/T}$
(R20)	$NO_2 + HO_2 \rightarrow HNO_4$	special function
(R21)	$HNO_4 \rightarrow NO_2 + HO_2$	special function
(R22)	$O_3 + HO_2 \rightarrow OH + 2O_2$	$1.4 \times 10^{-14} e^{-580/T}$
(R23)	$HO_2 + HO_2 \rightarrow H_2O_2 + O_2$ (H_2O dependent)	special function
(R24)	$H_2O_2 + h\nu \rightarrow 2OH$	radiation dependent
(R25)	$H_2O_2 + OH \rightarrow HO_2 + H_2O$	$3.1 \times 10^{-12} e^{-187/T}$
(R26)	$NO_2 + H_2O \rightarrow HONO + HNO_3 - NO_2$	4.0×10^{-24}
(R27)	$HNO_4 + h\nu \rightarrow NO_2 + HO_2$	$(1.0 \times 10^{-4})k_1$
(R28)	$HNO_4 + OH \rightarrow NO_2 + H_2O + O_2$	4.0×10^{-12}
(R29)	$SO_2 + OH \rightarrow SO_4 + HO_2$	PT dependent
(R30)	$HCHO + h\nu \rightarrow 2HO_2 + CO$	radiation dependent

4 Observed Data

Observed data recorded by the CASTNET (Clean Air Status and Trends Network) [USEPA, 2002a] and NAPS (National Air Pollution Surveillance) networks [Dann, 2002] was of the utmost importance to this project. Both networks report hourly measurements 24 hours per day, seven days per week. The regional grid for this project spanned Canadian territory as well as several central and eastern American states, therefore the measurement data could not be collected from one source. U.S. ozone data was obtained from the CASTNET database, while the NAPS database was used for the Canadian sector. From these networks, appropriate monitoring stations were selected, with the sites placed properly on the model domain. Next, appropriate averages were calculated to facilitate the comparison with model output.

The first step in compiling the observed measurements was to place the monitoring stations onto our model domain. Since station locations from both networks were given in latitudinal and longitudinal values, projecting these figures onto the flat model grid was necessary. A Lambert Conformal Conic projection was used for this purpose because the same projection was used to plot the map onto the grid. The mapping program Arc View was used to accomplish this.

Some analysis of the monitoring stations was required because each measurement was to be representative of the grid square on the domain in which it was located. This was necessary since the measurements were directly compared to the model results for its grid square. Stations that were not illustrative, then, of their respective grid squares were eliminated. These included stations in urban and industrial areas that measured highly variable ozone concentrations. For example a monitoring station a short distance away from a factory or busy highway would measure high values of ozone. Yet, while this value would be correct locally, it would not be representative of the larger area because ozone is destroyed quickly under these conditions as well. Thus, for this reason, measurements selected for this project were taken from stations located in rural, forested, unused, agricultural and residential areas. Ideally residential sites would have been removed as well, but the resulting network was far too sparse. This revision still left stations that recorded systematically unusual measurements with

respect to nearby measurements. These anomalies were therefore also removed from consideration.

Once the correct set of monitoring stations was chosen and plotted on the model grid, the measurements from each station were next compared to model data at the nearest grid point. (Since there is one model output value for each grid square, these values are assigned to the point at the center of each grid square. Therefore, it is frequently stated that the measurements were compared to the model result at the nearest grid point.) In some cases, however, the nearest grid point was located over a large body of water, such as one of the Great Lakes. Please note that over large water bodies, ozone concentrations are higher in spring and summer than nearby land based monitoring stations would measure for the reasons discussed in Chapter 2. Chemical transport models, however, take land use into consideration to account for these variations. In these regions, therefore, the model, rather than the interpolated measurements, will produce concentrations closer to the true concentrations. Accordingly, shoreline measurements were compared to a more distant model grid point located over land. Through this compromise, the data could still be used as a good estimate of terrestrial ozone concentrations a short distance away. The final set of monitoring stations could then be plotted on the model domain as seen in Figure 4.1.

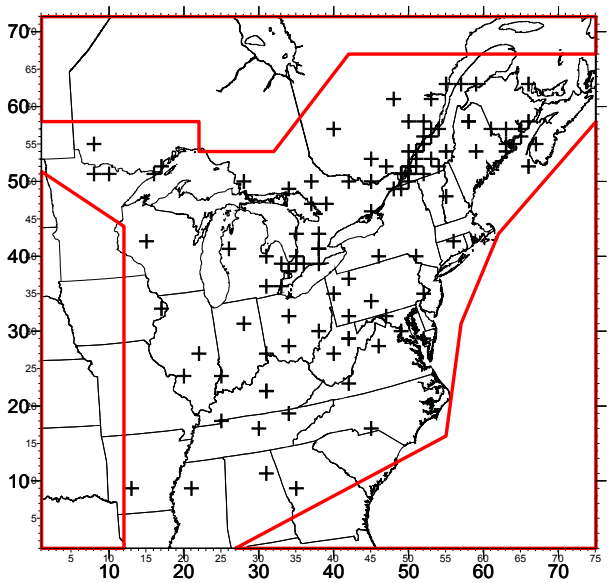

Figure 4.1: Valid ozone monitoring stations plotted on the ADOM model domain

As mentioned, the monitoring stations measured ozone hourly, 24 hours per day. Due to the fast chemistry associated with ozone, however, the hourly ozone concentrations fluctuated significantly throughout the day. Such fluctuations were impossible to model accurately and were also unimportant for the purposes of this project. Therefore, in order to reduce this background noise, calculating the daily eight-hour maximum ozone concentration was needed. The eight-hour maximum is defined as the eight hour time period in the day that has the highest average ozone concentration. The daily eight-hour maximums were then averaged over a study period, such as a week to further reduce noise.

Amalgamating many observed measurements, however, presents some difficulty. For example, measurements were not recorded for every hour of the day in some cases because of equipment malfunction. A threshold was therefore required to determine when data from a monitoring station should no longer be considered. For an average value to be valid, common practice requires 75% of the data points to be present. Thus for an eight hour time period to be considered, a minimum of six measurements were needed. Additionally, at least 18 of 24 eight hour time periods were required to calculate the daily eight-hour maximum. Problems may also arise if a station fails to record enough measurements to calculate a valid eight-hour maximum for each day in the study period. As a result, for a valid average to be calculated, an eight-hour maximum was required for 75% of the days in the study period. These eight-hour maxima averaged over the study period were then subtracted from corresponding eight-hour maxima from the model output to produce a difference file. If two stations were located in the same grid square, then their values were averaged before the comparison with model output.

In summary, monitoring stations were selected by how representative they were to their respective grid squares. Subsequently, they were plotted on the nearest grid point of the model domain, with those stations located over a large water body moved to the nearest terrestrial grid square. Additionally, in order to reduce background noise created by the small fluctuations in ozone concentration, the average eight-hour maximum for the study period was calculated for each site. Next, the amalgamated measurement data was ready to be compared with similarly compiled model outputs.

5 Modified Simple Kriging (MSK)

The measured data must be interpolated over the grid because there do not exist measurements for each grid point in the domain. For this project, Modified Simple Kriging was chosen, as it is an accepted interpolation scheme used in atmospheric sciences [Lefohn, A.S. *et al.*, 1987;] [Armstrong, M., 1998]. The estimation procedure known as Kriging was developed by the South African engineer D. G. Krige and was originally used in the field of geostatistics. It is a procedure by which unknown concentrations are estimated based on a weighted average of surrounding known values. The weights are selected to minimize the estimation variance such that the weights add up to one. This last criterion, called the unbiased constraint, essentially means that the weights are normalized. The estimation variance is minimized by associating the weights to the semivariances of known values with respect to distance through the use of a semivariogram (*vide infra*). Accordingly, before a proper description on the use of Kriging can begin, a more thorough explanation of the semivariogram is required.

5.1 Semivariogram

Semivariance is a measure of the difference between the values of samples and the physical distance between them. It is assumed, for example, that air samples taken one kilometer apart are more closely related than samples taken ten kilometers apart. Thus, semivariance increases with respect to distance between samples. In most cases, a maximum range exists beyond which samples are no longer related at all. At this range, a maximum variance is also reached and is represented by a plateau in the semivariances. The semivariogram (or simply the variogram) is a plot of semivariance as a function of distance [Lefohn A.S. *et al.*, 1987]. There are several types of variogram models that have been created to describe various distributions of related data. A few of these examples are the spherical, exponential, and Gaussian models for which the general equations respectively are,

$$\gamma(h)\begin{cases} = n + c\left[1.5\dfrac{h}{a} - 0.5\left(\dfrac{h}{a}\right)^3\right], & for\ h \le a \\ = n + c, & for\ h > a \end{cases} \qquad (2)$$

$$\gamma(h) = n + c\left[1 - \exp\left(-\frac{3h}{a}\right)\right] \quad (3)$$

$$\gamma(h) = n + c\left[1 - \exp\left(-\frac{3h^2}{a^2}\right)\right] \quad (4)$$

where n is the "nugget value" or the systematic error, c is the maximum variance less n, h is the distance from the known value and a is the maximum range of relevant data (i.e. the range where the plateau starts). Because of the shape of the semivariance graphs, the spherical model was determined to fit best the regional atmospheric distributions of ozone. This variogram was fitted to the semivariances using a least squares procedure. An example of this variogram can be seen in Figure 5.1. Here the plateau begins at a distance of 79 grid squares, which can be seen by a flattening of the variogram. This means that the range is 79 grid squares in this case.

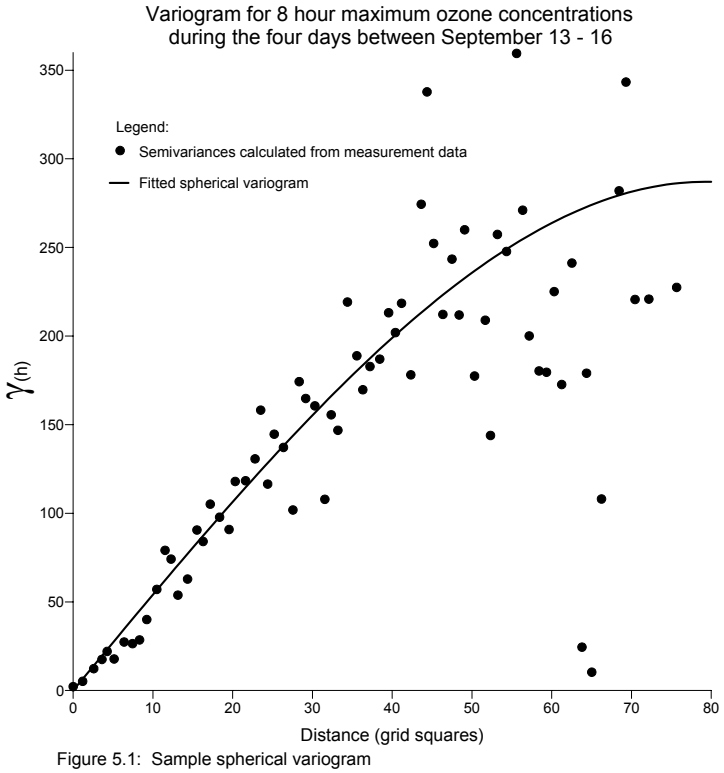

Figure 5.1: Sample spherical variogram

Mathematically, the exact variogram, γ(h) is defined as follows [Lefohn A.S. *et al.*, 1987]:

$$\gamma(h) = \frac{1}{2} Var[Z(x+h) - Z(x)] \qquad (5)$$

where $Z(x)$ is a measurement at an arbitrary location x; $Z(x+h)$ is the measurement at a grid point, and 'h' is the distance from location x to the grid point. Also, *Var* represents the variance between the measured concentration value at Z(x) and Z(x+h). Now, we know that $Var(Z(x)) = E[Z(x) - E(Z(x))]^2$, where $E(Z(x))$ is the expected value (or mean value) at x. For small values of h, $E(Z(x+h) - Z(x)) = 0$. Then

$$Var[Z(x+h) - Z(x)] = E[Z(x+h) - Z(x)]^2 \qquad (6)$$

and therefore,

$$\gamma(h) = \frac{1}{2} E[Z(x+h) - Z(x)]^2 \qquad (7)$$

Then, where there are $N(h)$ pairs of sample locations for a given distance h, the variance can be estimated by

$$\gamma(h) \approx \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i + h) - Z(x_i)]^2 \qquad (8)$$

From this point forward, Z(x) will be shortened to Z, with the point we are trying to solve for generalized as Z_k. Accordingly, k is an index where $0 \le k \le M$, where M is the number of unknown points, with all known observation points generalized by Z_i, and i is an index between 0 and the number of known points, P.

5.2 Modified Simple Kriging

The variogram is the means by which the Kriging method determines the weight of each measurement and minimizes the estimation error [Lefohn A.S. *et al.*, 1987]. Thus, to estimate

an unknown value using Kriging, a value is calculated based on a weighted average of its neighbors. For example, to estimate the unknown value Z_k^*, the following equation is solved,

$$Z_k^* = \sum_{i=1}^{n} \lambda_i Z_i \qquad (9)$$

where λ_i is the weight associated with the known value Z_i (*vide infra*). Moreover, when the distance between a known point and an unknown point is greater than or equal to the distance at which the plateau starts on the variogram, the weight is zero. In addition, the closer the observed point is to the unknown point, the greater the weighting, to a maximum value of one.

The next step in the procedure is to calculate the weights in an effort to minimize the estimation variance, subject to the unbiased constraint. This variance can be expressed in terms of the weights and γ by

$$Var\left[Z_k^* - Z_k\right] = -\sum_i \sum_j \lambda_j \gamma_{ij} + 2\sum_i \lambda_i \gamma_{ik} \qquad (10)$$

where Z_k is the exact unknown value, Z_k^* is the approximated unknown value, and i, j and k are the indices. γ_{ik} can be written as the difference between the value of the variogram for the distances separating the sample i and the sample being estimated, and the average value of the variogram within the area. γ_{ik} is the value of the variogram for the distance separating sample i and sample j. Minimizing this variance under the unbiased constraint (i.e. the weights add up to one) gives the following set of two equations to be solved simultaneously [Lefohn A.S. *et al.*, 1987].

$$\sum_{i=1}^{n} \lambda_i \gamma(i, j) - \mu = \bar{\gamma}(j, k) \qquad (11)$$

and $\sum_{i=1}^{n} \lambda_i = 1$. (12)

where $1 \le i, j \le n$, k is the index of the unknown value, and μ is a Lagrange multiplier. The Lagrange multiplier is simply a slack variable in the set of equations. A slack variable is one that is added to each inequality in a system of inequalities to transform it to a system of equalities. The slack variable is given a value equal to the difference between the right and left

hand sides of each inequality in the system [Chvatal, 1983]. Also $\gamma(i, j)$ is the value for the variogram for the distance separating sample *i* and sample *j*, and $\bar{\gamma}(j, k)$ is the average value of the variogram between point *j* and the sample being estimated.

Thus, to summarize the Kriging process, unknown values are estimated by first calculating the semivariogram using equation (8), then the weights are estimated using the set of equations (11) and (12), and then the value itself is calculated using equation (9). This method was used to interpolate differences between model and observed data across the grid.

5.3 How Kriging was Implemented

Because Kriging is commonly used for many types of modeling, writing a program to complete this step of the project was not necessary. Instead, a visualization software package called Surfer was used to do the Kriging to create contour graphs of ozone. Surfer was originally designed to visualize groundwater modeling results, and is therefore a sophisticated contouring package that includes a Kriging procedure.

The first step to applying Kriging is to calculate the variogram. To do this, Surfer has a variogram calculation procedure that employs user-supplied input data (in our case the measured data), to first calculate the associated semivariances. Next, the user chooses an appropriate variogram model and Surfer fits the model to the data using a Least Squares method to minimize error. This first step is imperative because the contours are later calculated using the fitted variogram model as described above. This method was then used to interpolate and contour the observed data collected from the monitoring stations, as well as the local differences between measured and modeled data. In contrast, no Kriging was necessary for the model output since the model already produced values for all grid points.

It must be stressed that Kriging is strictly an interpolation procedure. Within the model domain, however, there exist three regions around the borders of the grid in which there are no

monitoring stations. These regions are represented by the polygons outlined in red in Figure 4.1 in Chapter 4. The measured data would need to be extrapolated to estimate observed values in those regions. These regions were "blocked", i.e. taken out of the domain, to prevent Surfer from extrapolating measured data because the model results are a better estimate than extrapolated measured data. The model data should only be corrected within the regions where observational data are available.

In summary, Surfer first calculates semi-variances using equation (8), with a spherical variogram next fitted to the data. Accordingly, this variogram is used to interpolate the observed data across the central region of the model domain only. This interpolated data can now be compared with model output.

6 Procedure

Only recently has improving Chemical Transport Model output through data assimilation with observed measurements been explored. This process was complex and required several steps to accomplish. To produce an adequate test for this procedure, many weeks of model results were produced. In addition, corresponding observed measurements were collected, and compiled. These were then subtracted from the model output and the differences were interpolated over the model domain. The differences were then used to nudge the model output in an assimilation process. Moreover, some statistical analysis was done in order to test the effectiveness of the method. Lastly the method was also used for a sensitivity analysis, to determine key monitoring stations within the domain.

As previously stated, the first step of this project was to run the model ADOM in order to produce model output for four months — June 4th through September 30th, 1996, respectively — with the input required for the model collected from the MOEE. Once all the model results were produced, compiling them into a format that was compatible with the observed measurements was necessary.

For this project, model and observed data were compiled into 8-hour daily maxima that were averaged over a week, or episode. Initially, comparing the model and observed data directly was a possibility because the model outputted hourly concentrations of ozone and the monitoring stations also recorded hourly concentrations. While the model followed general trends of observed measurements, however, it failed to produce results that were temporally in sync with measurements. It also missed some of the minor fluctuations. For the purposes of this project, these inconsistencies were unnecessary noise. Thus, in order to fine-tune the data, eight-hour maximums were calculated for each day and averaged over a week. Consequently, seventeen week-long periods were studied between June 4 and September 30 and assimilated with observed measurements. In addition to the one-week periods, an ozone episode between June 27th and June 30th was also studied, as well as a period of particularly low ozone concentrations between September 13th and September 17th. Once the surface ozone data had been extracted from the model output and compiled into this format, they were ready to be

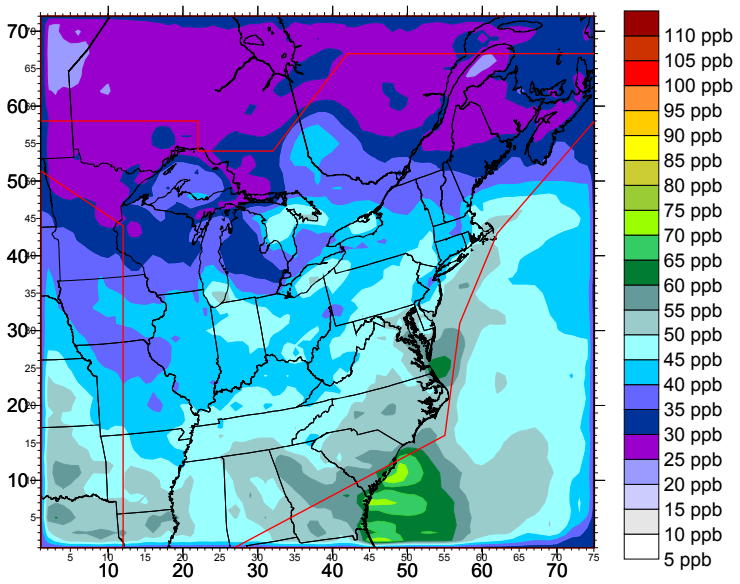

Figure 6.1: Model Output September 13 - 16

compared to similarly assembled measured data. Figures were also created in Surfer to visualize the model output in this format as in Figure 6.1.

Collecting the measured data from the CASTNET and NAPS databases was the next step. These data were also compiled into 8-hour maximum averages over the seventeen weeks between June 4th and September 30th. There were some missing measurements from the observed data, however, due to equipment failure. As a result, a threshold was needed to determine how many measurements were necessary to calculate an eight-hour average. Additional thresholds also had to be created to determine how many eight-hour averages were needed in order to determine an eight-hour maximum and how many eight-hour maximums were needed to calculate an eight-hour maximum average for a given week. Setting the threshold at 75% is common practice when working with these types of measurements; therefore at least 6 measurements were required to calculate an average for a given eight-hour period. Additionally, 18 eight- hour periods in a given day were needed to find a valid eight-hour maximum. Lastly, eight-hour maximums from at least five days out of seven were used to calculate a valid weekly average, while three eight-hour maximums were required to calculate the average for the four-day periods. The observed data could then be visualized as in Figure 6.2, and the observed data could also subtracted from corresponding model results.

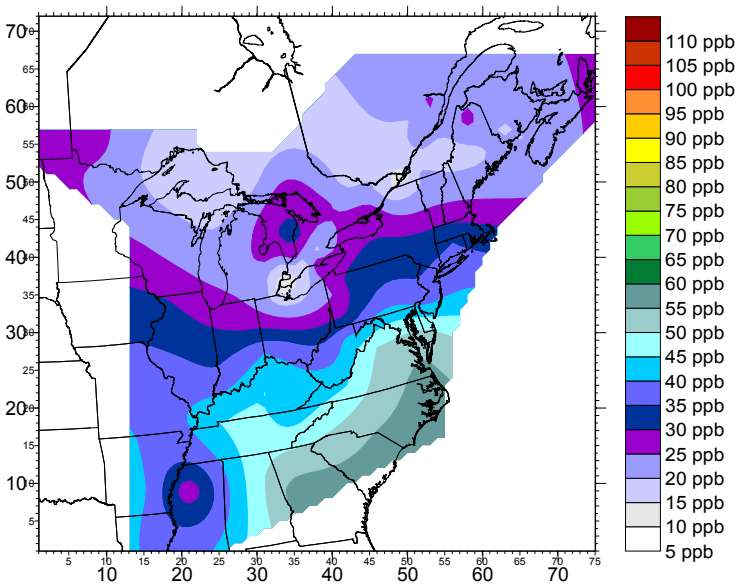

Figure 6.2: Observed Data September 13 - 16

The next step in the procedure was to calculate the differences between model output and observed measurements. Not every grid square contained a monitoring station. Consequently, model results for the grid squares that contained monitoring stations were extracted from the model output file. The observed data were then subtracted from those model results to produce a difference file. These differences were then interpolated over the central region of the grid.

The interpolation was not straightforward, because the interpolation did not occur over the entire grid. There are two ways to ensure that Surfer interpolates only over the central region of the grid. The most obvious method is to employ the blocking feature, which basically eliminates part of the domain. This could not be used, however, because once a grid contained a blocked region, subsequent calculations with that grid necessarily resulted in a final grid that also contained the blocked region. A blocked region could not be introduced at this point in the procedure because the differences were to be subtracted from the model results in the next step to produce a corrected picture that incorporated the entire grid. The second way to prevent extrapolation is to include zeros in the difference file for the grid points in the border regions. This is the method that was used. Figure 6.3 gives a visualization of a difference file with the zeros in the border region. This difference file was later blocked for display purposes only, as illustrated in Figure 6.4.

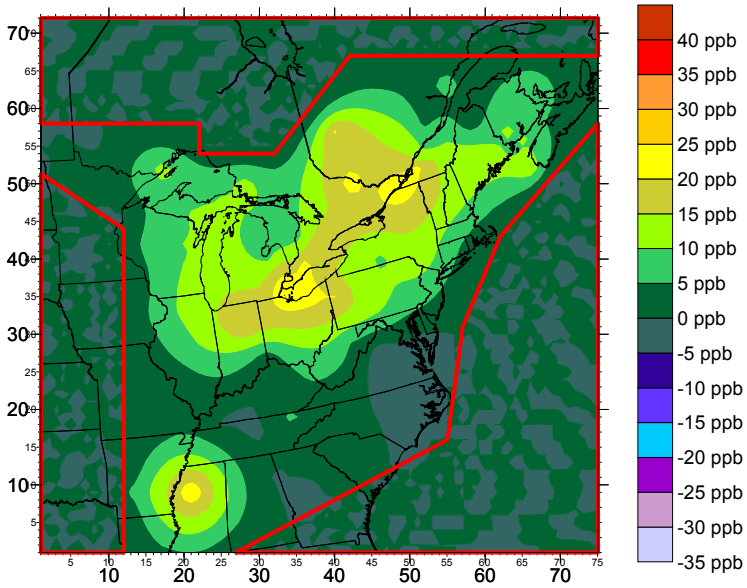

Figure 6.3: Unblocked Difference = Kriged (Model - Observed)
September 13 - 16

Once the interpolated difference file was obtained, the next step was to use it to correct the model output. To achieve this goal, the differences were subtracted from the model results. This preserved the resolution of the model output, while simultaneously improving the accuracy of the results. These enhancements can be seen in Figure 6.5. In the following steps, this corrected picture was statistically tested to quantify the improvement.

Measuring improvements was difficult because no perfect picture could be obtained to use as a

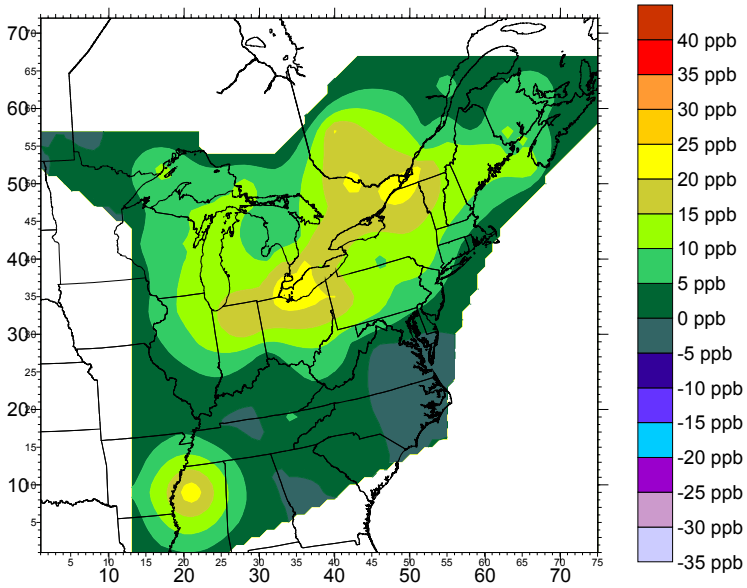

Figure 6.4: Difference = Kriged (Model - Observed)
September 13 - 16

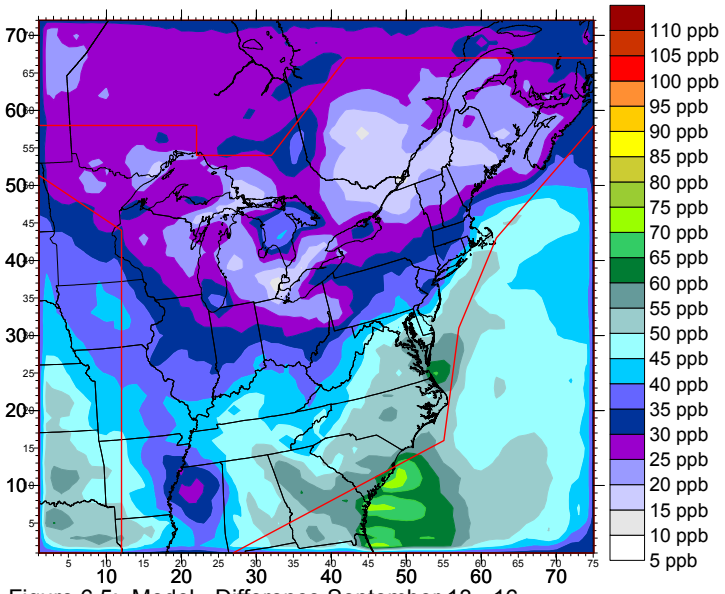

Figure 6.5: Model - Difference September 13 - 16

gauge. Consequently, the corrected picture was compared with the existing measurements in the central region, and with the model results in the border region in order to estimate error. To accomplish these calculations, the interpolated measurements were subtracted from the corrected picture. Next, the model values were subtracted from the borders to capture error at the junction between the central portion and the border region of the grid. Please note that both of these calculations were expected to produce values close to zero and, when all monitoring stations were used in our method, any variation from this supposition was due to the difference in resolution between the model output and observed data as shown in Figure 6.6. Nevertheless, the Normalized Chi Square, Normalized Bias, and Normalized Gross Error were calculated to provide a base case for further analysis. The equations for these are:

$$Normalized\ Chi\ Square\left(\frac{\chi^2}{N}\right)=\frac{1}{N}\sum_{i=1}^{N}\frac{\left(x_i-E_i\right)^2}{E_i}, \tag{13}$$

$$Normalized\ Bias=\frac{1}{N}\sum_{i=1}^{N}\frac{\left(x_i-E_i\right)}{E_i}, \tag{14}$$

$$Normalized\ Gross\ Error = \frac{1}{N}\sum_{i=1}^{N}\frac{|x_i - E_i|}{E_i}, \qquad (15)$$

where N is the number of points, x_i is the point being tested and E_i is the observed value.

In order to truly test the interpolation and assimilation method, we needed to find how much the error would increase if the monitoring network became more sparse. Therefore, some sites needed to be removed from the correction procedure and used only for error assessment. As a result, the set of monitoring stations was randomly divided into two groups, as illustrated in Figure 6.7. The method described above to attain a corrected model output file was then followed for each group separately. Each corrected picture was then compared against the entire collection of measurements to determine how well missing values could be estimated. Subsequently, the Normalized Chi Square, Normalized Bias, and Normalized Gross Error were again calculated in order to compare these values against the base case statistical calculations. This stage was necessary to discover the amount by which the error increased when fewer measurements were used in the procedure.

With this test completed, certain monitoring stations were found to be much more important to the interpolation process than others, due to their location. To identify all key sites would

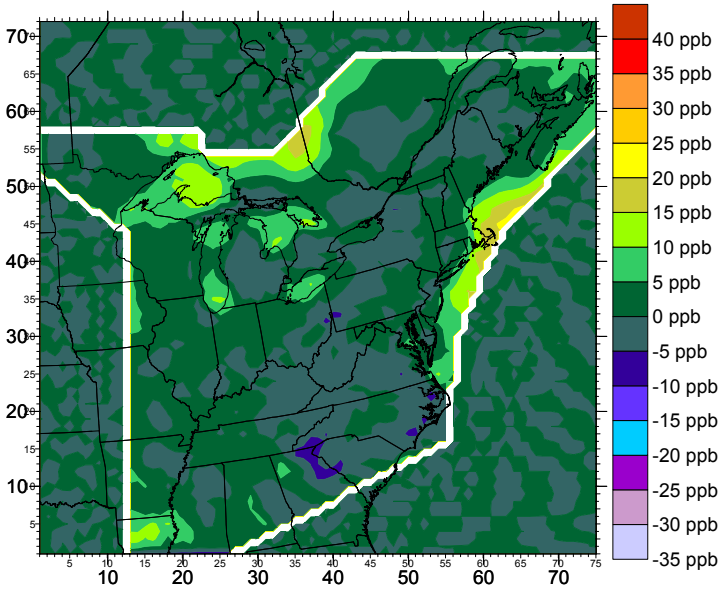

Figure 6.6: Check Corrected Values, September 13 - 16
Interior: Model - Difference - Observed
Exterior: Model - Difference - Model

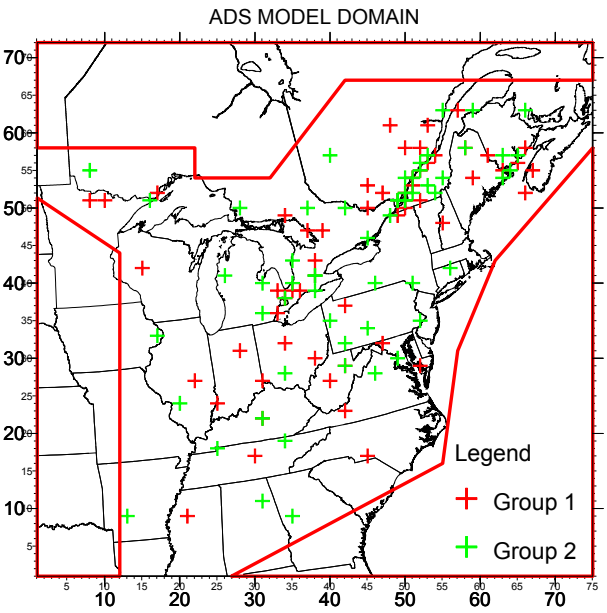

Figure 6.7: Monitoring station split

require an intensive sensitivity analysis. Our procedure would be able to complete much of this investigation. By removing a suspected key site from the correction procedure, we can then test how well the method predicts this known value. Moreover, the procedure's sensitivity to the site can be quantified by comparing the Chi Square, Bias and Gross Error values to the base case statistical calculations. Accordingly, this test was attempted for four sites as an example, but a lack of time prevented a full analysis from being done.

To summarize, several steps must be taken in order to enhance ADOM ozone output with observed measurements. Beginning with the model runs after the input data is collected, the ozone output for the surface layer must next be extracted and compiled into a format that can easily be corrected with observed measurements. Additionally, the observed data must also be collected and reformatted in order for comparison to the model output.

In our procedure, the difference between observed measurements and model output was found for each chosen time period with the differences then Kriged across the central region of the grid. Statistically testing the improvement made by the assimilation was then necessary.

Because of the sparseness of ozone measurements, a perfect picture of ozone concentration could not be obtained to gauge directly the improvements made by the procedure. Therefore, the sites were split into two groups, which were then tested individually against the entire collection of measurements. As a result, the statistical analysis was able to determine the amount of error that increased as the number of monitoring stations decreased. Additionally, a similar process was devised to perform a sensitivity analysis on the monitoring stations, with four stations analyzed to test this method. The analysis was not completed, however, and further study is required in order to find all key stations.

7 Results

Despite the sparse network of monitoring stations, the results produced by the procedure were encouraging since they showed that the assimilation procedure increased the local accuracy of the model results while preserving its resolution. This result could easily be seen in the contour graphs produced in Surfer, with the improvements then quantified in the statistical analysis. While visual representations of all results could not be displayed, the results for the four day episode in June will be explained in detail here, while the statistical results from the rest of the time periods will be summarized in Table 7.1 at the end of the chapter.

Figures 7.1 through 7.4 on the following page show the model, observed, difference and corrected model visualizations for all valid monitoring stations used in the assimilation procedure for the ozone episode from June 27 through 30, 1996. Figure 7.5 shows the difference between the corrected model output and the observed data in the central region of the domain, with the difference between the corrected data and the model output in the border regions of the grid. Improvements made by assimilating the model output with all the observed data is difficult to quantify. This is more easily accomplished when some monitoring stations are removed from the procedure and used only for testing purposes. This was done in the next step. Yet there remain several important points to note in this set of figures.

In Figures 7.1 and 7.2 — the model and observed results — we can see that while the overall pattern between the two is somewhat similar, the model under- predicts ozone quite significantly in the central region. This difference is further demonstrated in the *difference* slide in Figure 7.3. Under prediction of ozone concentrations during an ozone episode is a common problem in chemical transport models since they tend to flatten extreme peaks found in measured data (i.e. the model also over predicted ozone when there was an extreme low

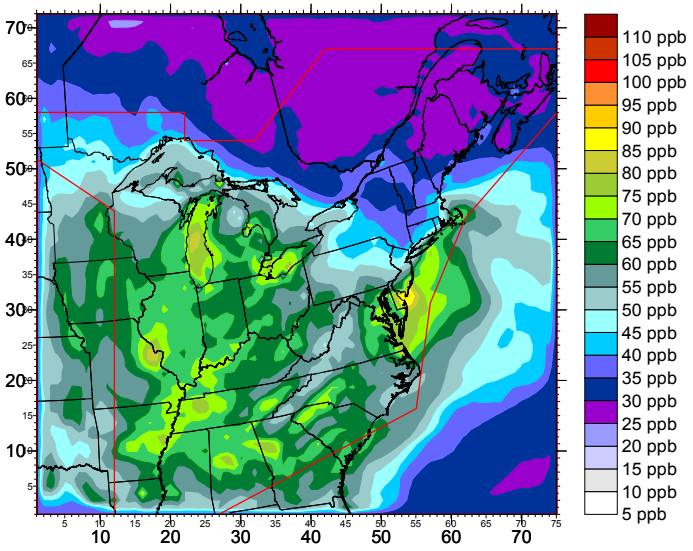

Figure 7.1: Model Output, June 27 - 30

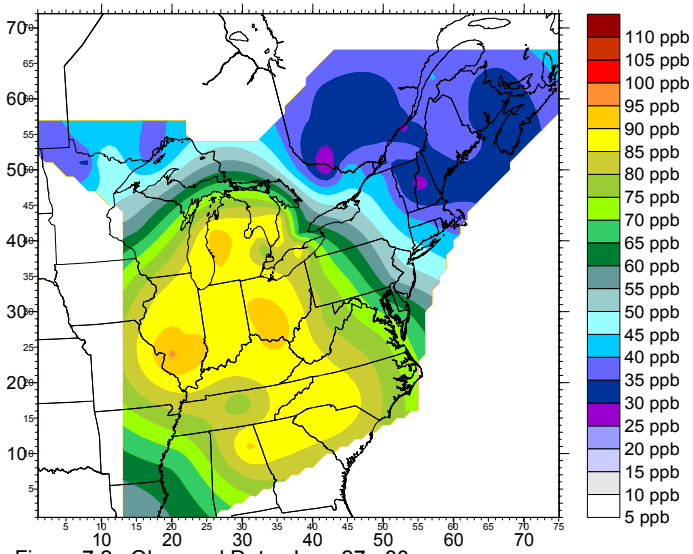

Figure 7.2: Observed Data, June 27 - 30

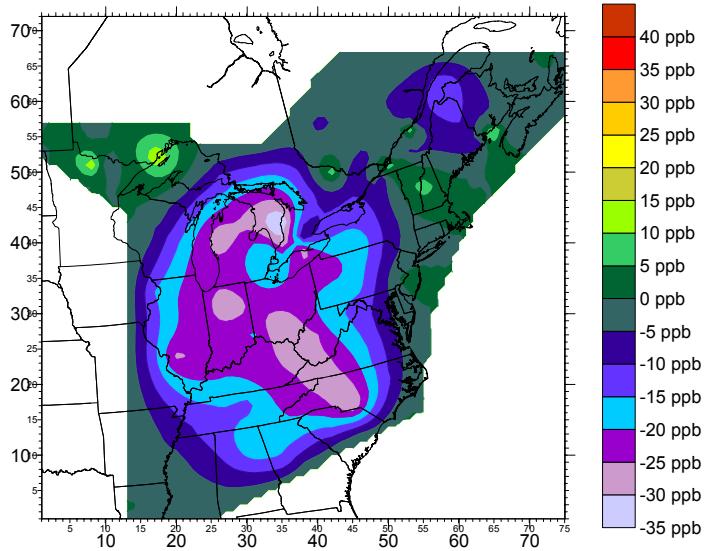

Figure 7.3: Difference = Kriged (Model - Observed)
June 27 - 30, all stations
Normalized Chi Square: 2.172 Normalized Bias: -0.124,
Normalized Error: 0.139

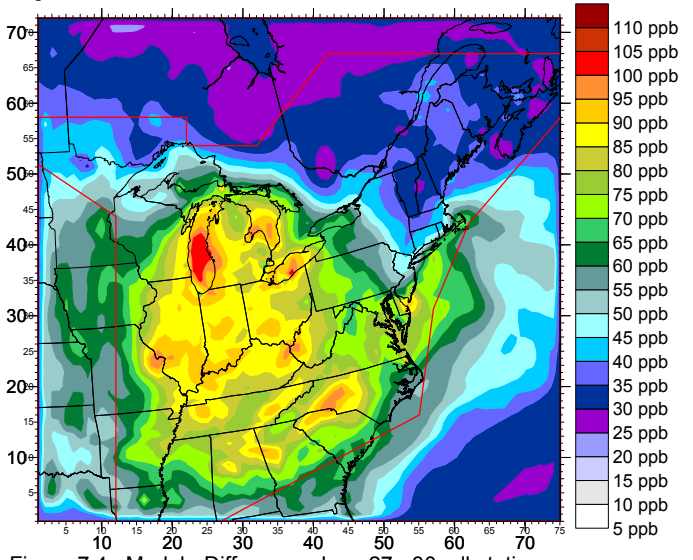

Figure 7.4: Model - Difference, June 27 - 30, all stations

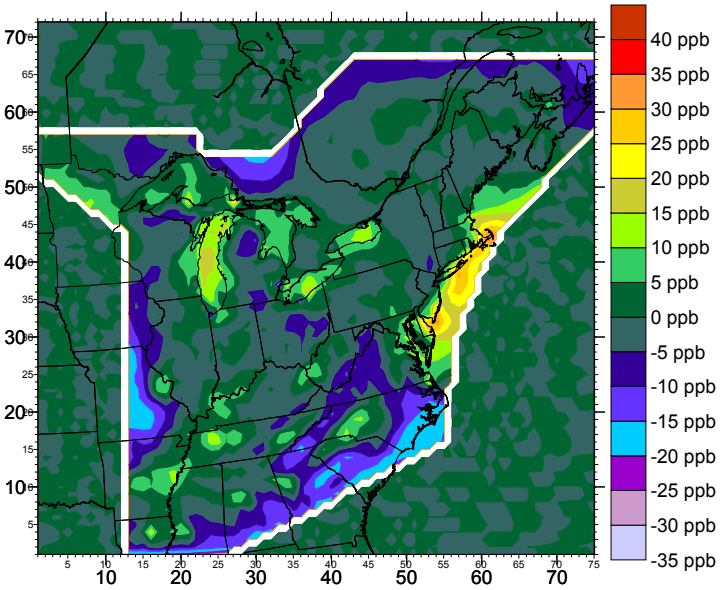

Figure 7.5: Check Corrected Values, June 27 - 30, all stations
Normalized Chi Square: 0.512, Normalized Bias: 0.009,
Normalized Error: 0.049

measured as well). This flattening is, however, a known drawback of chemical transport models and was anticipated. In fact, it was this error that the assimilation procedure attempted to correct.

Another feature to note in Figure 7.3 is the variance between the model output and the observed measurements over Southeast U.S., near the Atlantic coast and the Great Lakes. As mentioned in Chapter 4, the cause of this variance is largely the fault of interpolated measurements not being representative of the regions over large bodies of water. This differentiation was anticipated, however, and should in fact be preserved where possible. We see in Figures 7.4 and 7.5, the variation over the Atlantic coast and over the Great Lakes is maintained to a large extent. As a result, the corrected model output successfully incorporated the local accuracy of the measurements, without sacrificing areas of the grid where the model performed better than the interpolated measurements.

This initial view of the results of our assimilation procedure was positive, but the task remained to quantify the improvement of the corrected model output over the initial model output and the observed measurements. In order to accomplish this, the sites were split in half

with each resulting corrected model output data set tested against the whole set of measurements. Figure 6.7, in Chapter 6, shows how the monitoring stations were split between Group 1 and Group 2. The sites were randomly divided, with the only criteria being that it was necessary to have small areas in each set with no monitoring stations that were covered by the other set, since these holes were required to test the interpolation procedure. Accordingly, when the corrected picture was compared with the entire set of measurements, these areas were of the most interest when observing how well the interpolation procedure was able to estimate the values in those gaps. From these results, estimating how well the procedure would work on other existing gaps in both networks could then be accomplished. Once the analysis had been done with each group of sites, there were many points of interest to consider.

To illustrate in detail these features, the contour graphs of the data produced from the analysis of the ozone episode in June have been included. Figures 7.6 through 7.9 on the following page show the results for the first half of the stations, Group 1, while Figures 7.10 through 7.13 portray the results for the second half, Group 2. These contour graphs are illustrative of the results from the rest of the time periods which are summarized in Table 7.1.

In Figures 7.6 and 7.10 — the observed data from the respective groups — one should first observe how the data in these illustrations differs significantly from one another and from the contour graph produced when all monitoring stations were used. This variation is illustrated further in Figures 7.7 and 7.11 – the *difference* slides. Yet, despite these significant differences, Figures 7.8 and 7.12 show that the resulting corrected data from either group does not vary from Figure 7.4 as much as one would expect given the initial differences. This observation is further supported by the statistical analysis. Moreover, while the Normalized Chi Square values are worse when only half the sites are used, they are still much better than the observed or model values alone. The same observations may be stated for the Normalized Bias and Gross Error, thus indicating that the interpolation procedure estimated unknown values quite well.

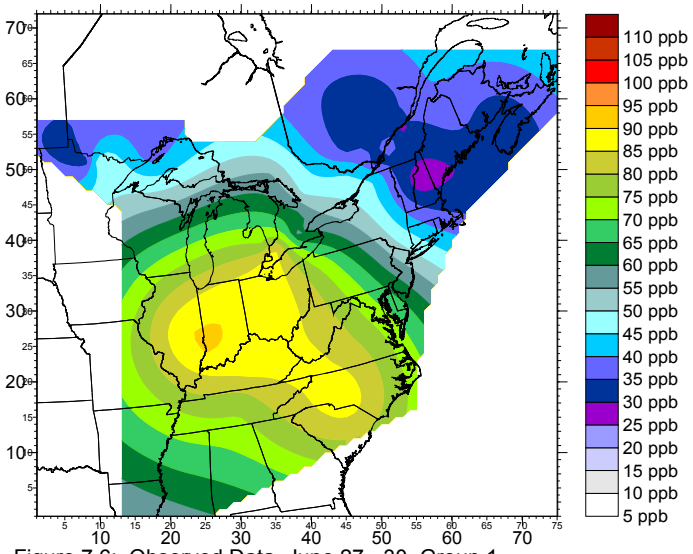

Figure 7.6: Observed Data, June 27 - 30, Group 1

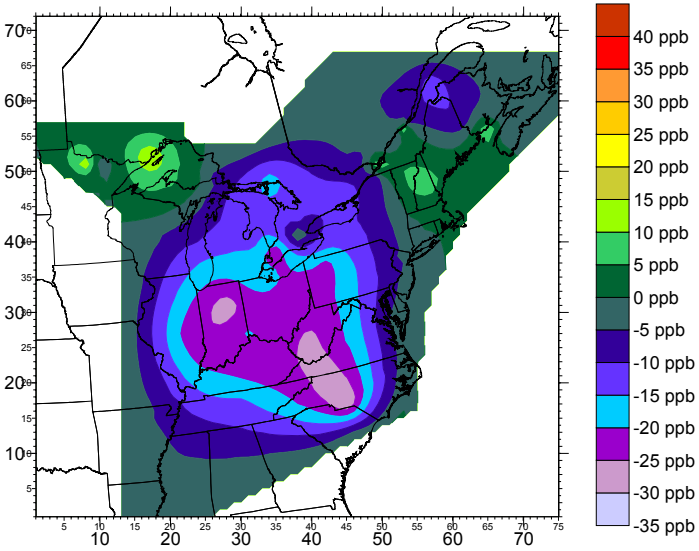

Figure 7.7: Difference = Kriged (Model - Observed)
June 27 - 30, Group 1

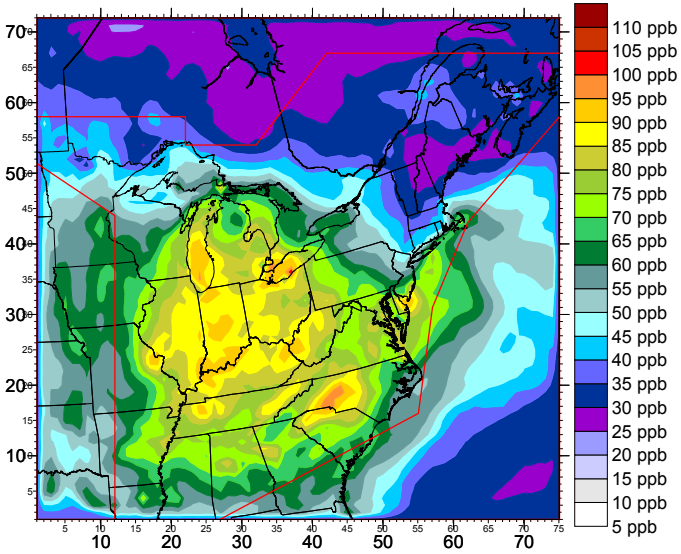

Figure 7.8: Model - Difference, June 27 - 30, Group 1

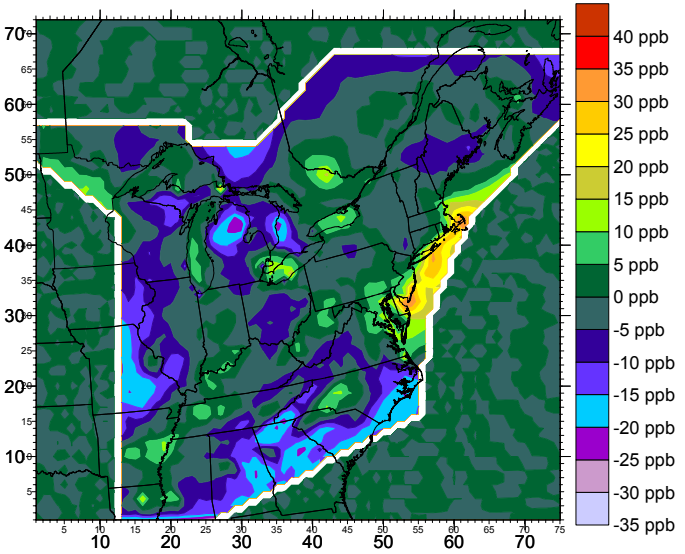

Figure 7.9: Check Corrected Values, June 27 - 30, Group 1
Normalized Chi Square: 0.630, Normalized Bias: -0.022,
Normalized Error: 0.057

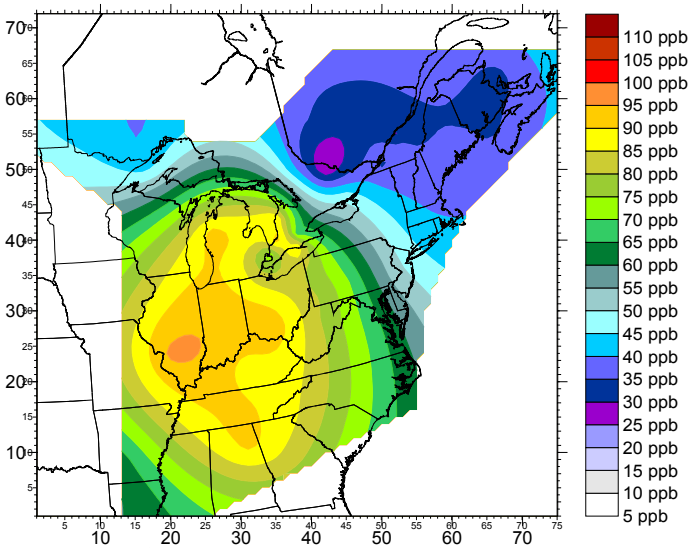

Figure 7.10: Observed Data, June 27 - 30, Group 2

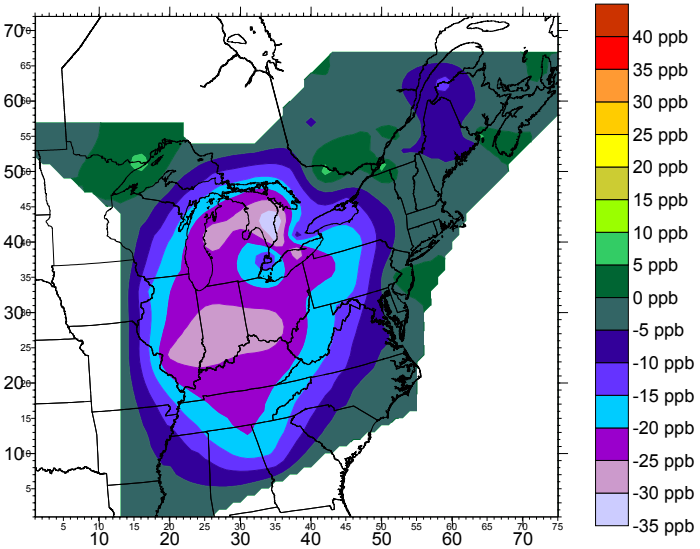

Figure 7.11: Difference = Kriged (Model - Observed)
June 27 - 30, Group 2

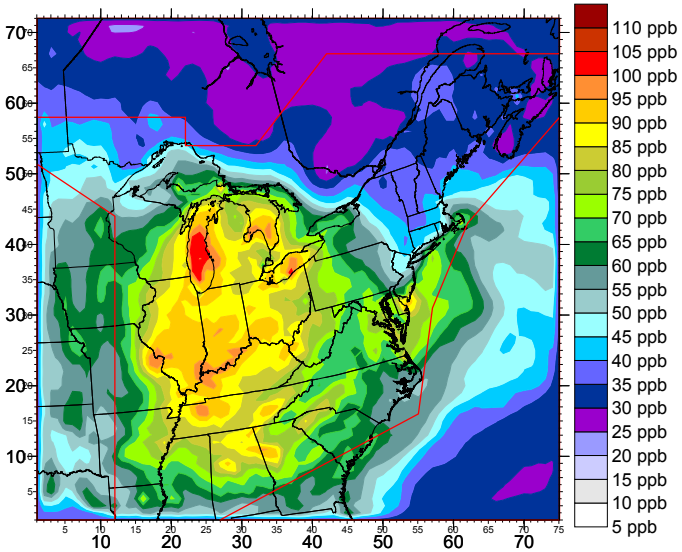

Figure 7.12: Model - Difference, June 27 - 30, Group 2

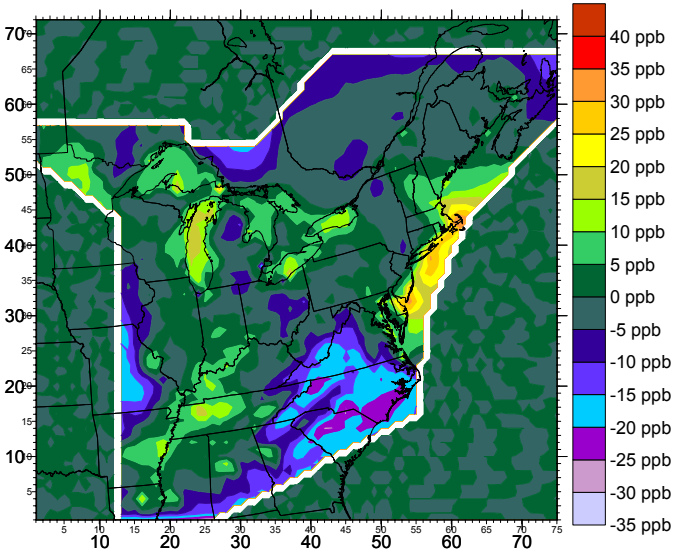

Figure 7.13: Check Corrected Values, June 27 - 30, Group 2
Normalized Chi Square: 0.678, Normalized Bias: -0.011,
Normalized Error: 0.060

Yet, while the interpolation procedure seems to work quite well when approximating unknown figures, caution must be taken in eliminating monitoring stations from the network. Many key sites exist that, if removed, would degrade the quality of the data quite significantly. For example, in the Great Lakes region of both Group 1 and Group 2, the error in the corrected picture is appreciably greater than that of the corrected picture using all stations. The error also increased along the Northwest edge of the central region and the central East Coast along the Atlantic Ocean. Thus, to combat this problem, completing a sensitivity analysis on potential stations to be eliminated is necessary to determine whether or not they represent key stations to the interpolation procedure.

An example of such an analysis was completed for four stations, with the results summarized in Table 7.2. Two of the stations, 060807 and CVL151, were labeled as possibly important while the other two, 063901 and ABT147, could be deleted with little consequence to the corrected model output, as can be seen from the statistical analysis. The results of the sensitivity analysis for the Ontario station 060807 from the NAPS network for the week from September 24 – 30, will be discussed in detail here while the statistical results from the rest of the weeks studied (i.e. June 25 – July 1, July 23 – 29, and August 20 – 26, 1996), is summarized in Table 7.2.

The following figures — Figures 7.14 – 7.22 — illustrate the results when 060807 was removed from the procedure as compared to the results when all the stations were used. The figures created using all of the monitoring stations in the assimilation procedure are on the left side, while the figures created without 060807 are on the right. A pink circle on the top left corner of each figure signifies the location of this station — one that deserves careful examination.

Without this station, the interpolation procedure overestimates ozone in the region to the North of Lake Superior from September 24 through September 30. This feature is emphasized further in the *difference* slide in Figures 7.17 and 7.18. Accordingly, the overestimation in the observed data results in an overestimation of ozone in the corrected model output as can be seen in Figures 7.19 and 7.20. Taking into consideration that only one station was removed from the procedure, most important to notice is the statistical analysis which shows a significant degradation in the quality of the corrected model output when 060807 is omitted. Moreover, for each week studied in the sensitivity analysis, the quality of the data decreased notably when this station was removed from the process. Therefore, regardless of meteorological conditions, the interpolation procedure was unable to correctly approximate the ozone concentration there. This implies that this station may be a key site, but further analysis must be done before a conclusion of this nature can be made. For the same reasons, CVL151 is a possible key station as well. Conversely, in the case of 063901 and ABT147, little or no change occurred in the Chi Square values, the Bias, or the Gross Error when these sites were removed from the process. The interpolation and assimilation procedure correctly estimated the ozone concentrations in these regions thus allowing them to be removed from the process with little consequence.

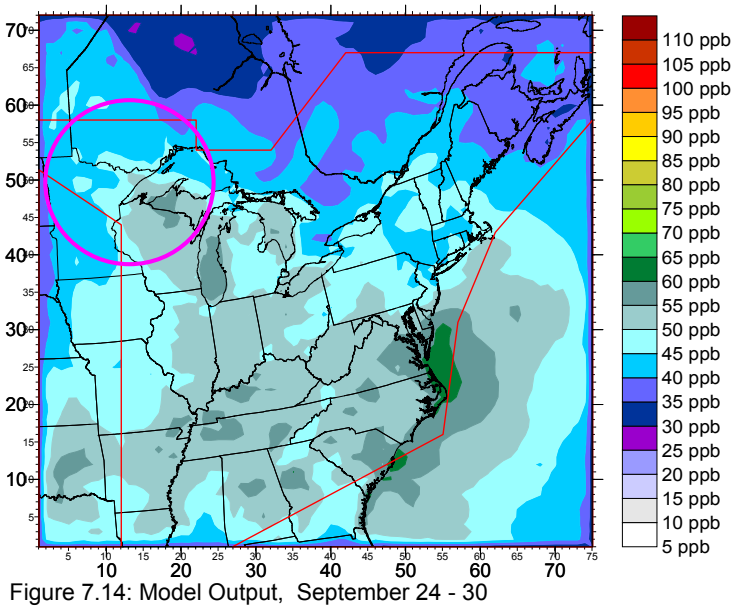

Figure 7.14: Model Output, September 24 - 30

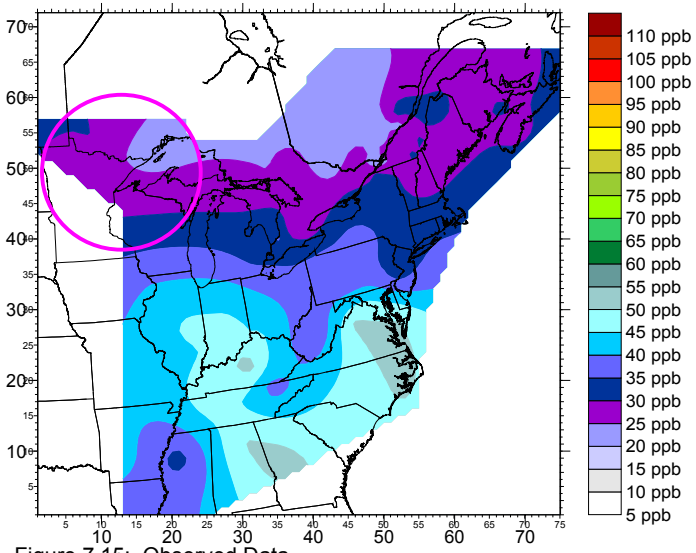

Figure 7.15: Observed Data,
September 24 - 30, all stations

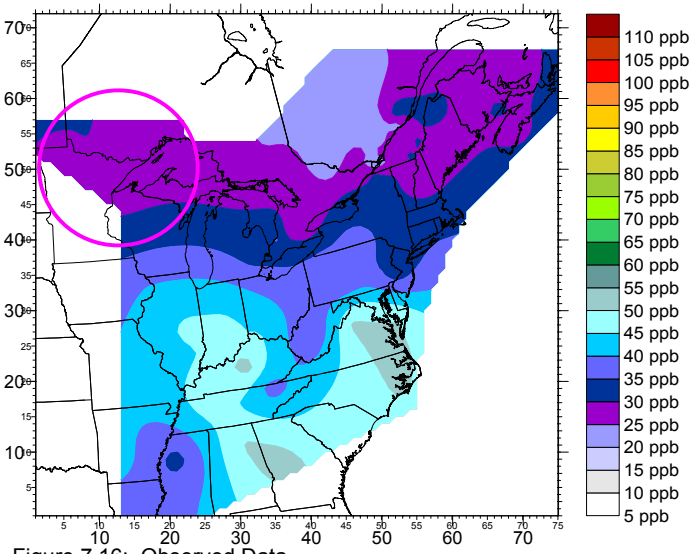

Figure 7.16: Observed Data,
September 24 - 30, all stations except 060807

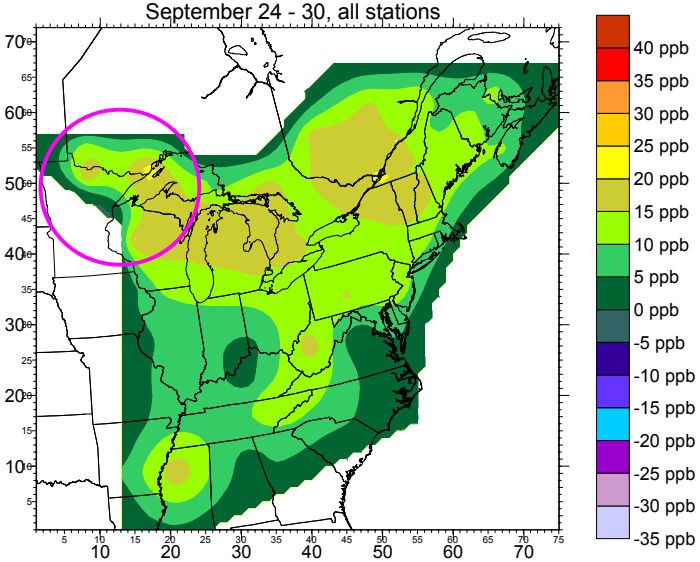

Figure 7.17: Difference = Kriged (Model - Observed),
September 24 - 30, all stations
Normalized Chi Square: 3.604 Normalized Bias: 0.285,
Normalized Error: 0.285

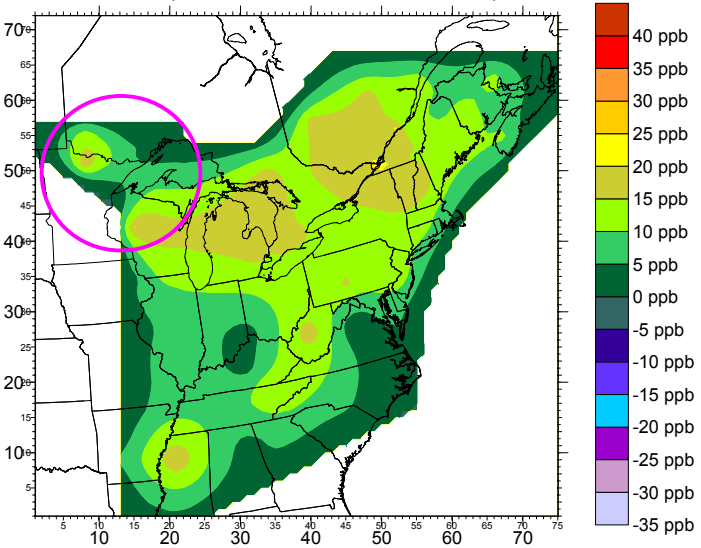

Figure 7.18: Difference = Kriged (Model - Observed),
September 24 - 30, all stations except 060807

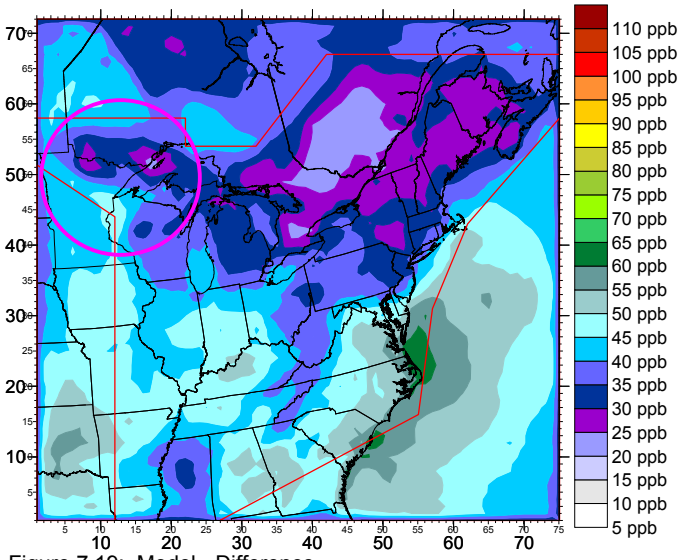

Figure 7.19: Model - Difference,
September 24 - 30, all stations

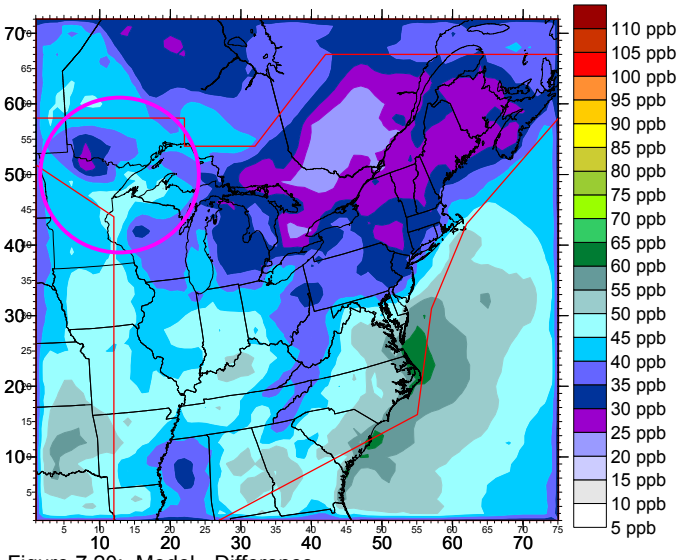

Figure 7.20: Model - Difference,
September 24 - 30, all stations except 060807

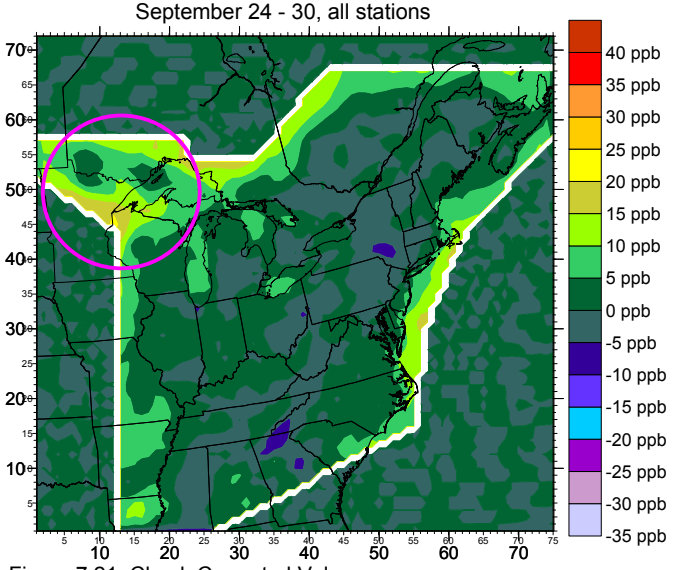

Figure 7.21: Check Corrected Values,
September 24 - 30, all stations
Normalized Chi Square: 0.553, Normalized Bias: 0.053,
Normalized Error: 0.066

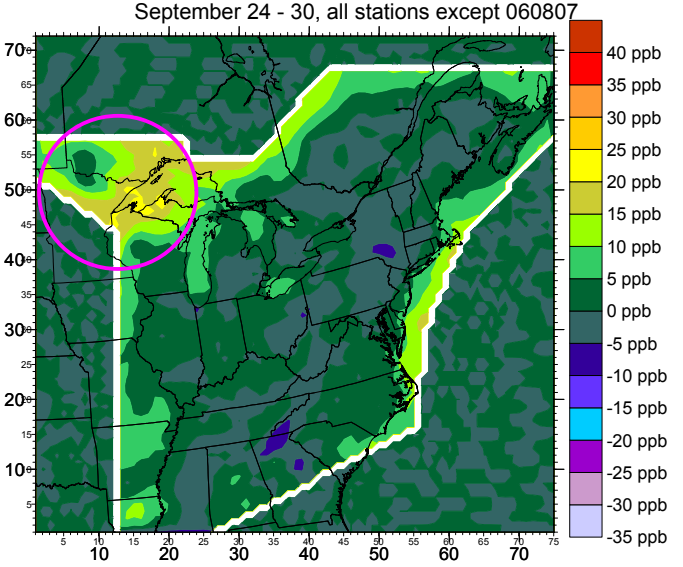

Figure 7.22: Check Corrected Values,
September 24 - 30, all stations except 060807
Normalized Chi Square: 0.748, Normalized Bias: 0.061,
Normalized Error: 0.075

The assimilation procedure developed in this project is therefore a versatile tool when modeling regional concentrations of ozone. The procedure itself produces a high-quality combination of model and measured data that maintains the high resolution of model output but includes the correct local accuracy of observed measurements. This procedure was tested by excluding half the monitoring stations from the assimilation procedure, and was found to satisfactorily approximate most of the concentrations from removed stations. Because there was significant variation in some areas of the grid, a sensitivity analysis was needed in order to find monitoring stations which were crucial to the success of the assimilation procedure. This analysis was started and two possible key stations were found. The sensitivity analysis was also done using the assimilation procedure in a way that isolated the site and monitored the decrease in quality of the corrected model output when the station in question was removed. Further sensitivity analysis, however, must be done to locate all key sites within the two networks to assess how much the quality of the data will be degraded by the shutting down of monitoring stations.

Table 7.1: Results Summary

	Normalized Chi Square				Normalized Bias				Normalized Gross Error			
	Before	After Assimilation Using			Before	After Assimilation Using			Before	After Assimilation Using		
	Assimilation	All Sites	Group 1	Group 2	Assimilation	All Sites	Group 1	Group 2	Assimilation	All Sites	Group 1	Group 2
Jun 04 - 10	0.9555115	0.2606435	0.3589583	0.0000003	0.0370339	-0.0181582	-0.0253716	0.0000012	0.0999769	0.0389062	0.0489413	0.0000012
Jun 11 - 17	0.5472079	0.3381160	0.4333304	0.3900643	-0.0307149	-0.0098196	-0.0185573	-0.0111365	0.0800369	0.0415622	0.0495522	0.0465519
Jun 18 - 24	0.7178381	0.3556158	0.5370666	0.4885466	0.0556809	-0.0080698	-0.0028189	-0.0097447	0.0888620	0.0470773	0.0598008	0.0603789
Jun 25 - Jul 01	0.6891619	0.3099188	0.4008231	0.4019864	-0.0523260	-0.0069415	-0.0173779	-0.0090365	0.0791165	0.0386110	0.0469639	0.0476529
Jul 02 - 08	0.7864081	0.2751474	0.3644793	0.4360625	-0.0770088	-0.0000010	-0.0110959	-0.0129731	0.1019527	0.0383397	0.0474012	0.0519338
Jul 09 - 15	0.3139657	0.2940696	0.3581530	0.3888980	0.0274293	-0.0051141	-0.0046644	-0.0136897	0.0610386	0.0405400	0.0461336	0.0502308
Jul 16 - 22	1.0728276	0.3727612	0.4047290	0.4706127	0.1131110	0.0201687	0.0206493	0.0205746	0.1166097	0.0455063	0.0491492	0.0543026
Jul 23 - 29	1.2193557	0.4025722	0.5950471	0.5584009	0.1484293	0.0378681	0.0513600	0.0506984	0.1491729	0.0540489	0.0695738	0.0679365
Jul 30 - Aug 05	0.6082223	0.4051629	0.6006539	0.5287970	0.0709519	0.0339297	0.0437723	0.0377597	0.0883888	0.0507112	0.0643497	0.0600135
Aug 06 - 12	1.0362188	0.3088016	0.3903907	0.7242486	0.0833307	0.0194376	0.0217925	0.0408169	0.1074799	0.0437170	0.0515251	0.0667665
Aug 13 - 19	0.5681850	0.2479251	0.3182037	0.4814928	0.0801023	0.0184684	0.0141717	0.0319809	0.0861935	0.0389618	0.0466397	0.0566712
Aug 20 - 26	1.4138656	0.5013657	0.5651230	0.8835734	0.1169393	0.0225844	0.0240685	0.0441626	0.1384568	0.0551521	0.0620476	0.0766026
Aug 27 - Sept 02	0.8121622	0.4289716	0.5349007	0.7019884	0.0389624	0.0334667	0.0353262	0.0498208	0.1067471	0.0524143	0.0615463	0.0705368
Sept 03 - 09	0.8260844	0.4500142	0.5087273	0.6565655	0.0856432	0.0351514	0.0334005	0.0518789	0.1151214	0.0563500	0.0643607	0.0715641
Sept 10 - 16	3.5499566	0.6351998	0.7234184	1.1107352	0.2842663	0.0529030	0.0567947	0.0799497	0.2845543	0.0698304	0.0797092	0.1014517
Sept 17 - 23	0.3821358	0.1697510	0.2399760	0.3943475	0.0315077	0.0095572	0.0085002	0.0265570	0.0706174	0.0350242	0.0450224	0.0529824
Sept 24 - 30	3.6042704	0.5525733	0.6730096	1.2261837	0.2851483	0.0525334	0.0673036	0.0861709	0.2851483	0.0661902	0.0788647	0.1034033
Episode: Jun 27 - 30	2.1717918	0.5157619	0.6297165	0.6783299	-0.1237422	-0.0092283	-0.0223519	-0.0112693	0.1386423	0.0485064	0.0568540	0.0598425
Low Period: Sept 13 - 16	3.5296403	0.4827736	0.6448229	0.8268253	0.2805292	0.0467259	0.0543545	0.0699712	0.2821059	0.0659998	0.0796863	0.0920601

Table 7.2: Sensitivity Analysis Summary

	Normalized Chi Square					Normalized Bias					Normalized Gross Error				
	After Assimilation Using All Sites					After Assimilation Using All Sites					After Assimilation Using All Sites				
		Except					Except					Except			
		060807	063901	ABT147	CVL151		060807	063901	ABT147	CVL151		060807	063901	ABT147	CVL151
Jun 25 - Jul 01	0.309919	0.313265	0.310548	0.307807	0.310475	-0.006941	-0.005634	-0.007490	-0.007375	-0.006489	0.038611	0.038887	0.038644	0.038579	0.038668
Jul 23 - 29	0.402572	0.444564	0.402572	0.416224	0.453056	0.037868	0.043023	0.037868	0.038795	0.043884	0.054049	0.057629	0.054049	0.054579	0.057652
Aug 20 - 26	0.501366	0.605611	0.501366	0.496966	0.612222	0.022584	0.028762	0.022584	0.021850	0.032222	0.055152	0.060252	0.055152	0.055187	0.060378
Sept 24 - 30	0.552573	0.748193	0.552573	0.554978	0.612153	0.052533	0.061182	0.052533	0.053343	0.058731	0.066190	0.074786	0.066190	0.066267	0.071037

8 Discussion

The success of the interpolation of observed measurements and the assimilation of model and observed data has many applications. The data set developed in this project can itself be used to start an archive of ozone concentrations. Also, the technique has many uses, including the development of background ozone concentrations for microscale modeling projects as well as model evaluation. Similarly, the interpolation and assimilation method can be used to perform sensitivity analysis on monitoring stations to evaluate their importance within the network.

The interpolation and assimilation techniques developed in this study were the focus of the work, but the assimilated model and observed data files can be used in future historical studies requiring 1996 ozone data. An archive of the data produced in a study as long as this one is important, because the time frame studied in modeling experiments will often be determined by the availability of data. Moreover, the archive can easily be expanded to produce accurate historical ozone concentration archives, which would be superior in most ways to the historical measurement data that are currently documented. The only drawback to the data files from this study is the loss of the hourly detail. For studies that require hourly data then, the interpolation and assimilation techniques could be used to improve the ozone profile.

For example, microscale modeling projects that require surface level regional background concentrations of pollutants would benefit from the use of the assimilation procedure. Currently, in order to calculate background concentrations, a modeler must use either the hourly measurements from often just one monitoring station, or the output from a regional chemical transport model. Neither of these options is ideal because microscale modeling projects require high-quality background concentrations in order to accurately calculate the pollution contribution of the emission source being tested. Assimilated model output and observed measurements, however, will provide detailed, accurate background concentrations. Also, for microscale modeling projects, the assimilation procedure needs no modification to extrapolate calculations vertically since only surface level concentrations are needed.

In addition to improving pollution data, the interpolation and assimilation procedures can aid in model evaluation and verification. Measurement data alone currently are used to evaluate a model's performance. This strategy can show large areas of discrepancy between model results and measurements, but isolating smaller regions that have great disagreement is often more useful. This allows a scientist to better determine the cause of disparity. Moreover, modeling over large regions often presents many discrepancies in results. In our case, for example, the model consistently under-predicted ozone concentrations — a common problem in regional CTMs. Under-prediction is caused by several contributing factors. First, uncertainty exists in the model inputs, and, in particular, the emissions inventory. If emissions are under predicted, the model will accordingly output lower concentrations of these pollutants. Furthermore, models assume even distributions of pollutants over each grid square while, in reality, pockets of high concentrations will exist within the grid square, thus leading to more chemical reactions than the model could predict. Therefore the model tends to flatten any extreme concentration peaks, an error that is fixed by the assimilation procedure.

Once this large error has been corrected using data assimilation, scientists can concentrate on problems in the model output that can be improved upon. When measurement data are subtracted from corrected model output, small regions with large discrepancies are highlighted. Scientists can also find regions within the domain that are consistently modeled incorrectly by analyzing several weeks of model output in this way. In this study, once the overall under prediction had been corrected by the assimilation procedure, we could see that ozone concentrations were high in comparison to measurements over the Great Lakes and along their shores. Over the lakes these values were most likely correct, but the shore values required improvement. Therefore, the integration between marine and terrestrial environments may be of particular interest to scientists looking to improve ADOM. Better models, however, do exist, making further work on this CTM unlikely, but the technique could be applied to other modeling projects.

Another benefit of this study is its ability to evaluate the importance of individual monitoring stations within the network. As discussed in Chapter 6, the station being tested is removed from the assimilation procedure, and the resulting corrected picture is tested against the full set of monitoring stations. In our study, four stations were tested, two from the CASTNET database, and two from the NAPS database. The two American stations were CVL151, located at the bottom left of the grid in Mississippi, and ABT147, located in Connecticut on the East Coast. The Canadian stations were both located in Ontario, just Northwest of Lake Superior and were called 060807 and 063901. From the statistics, we can see a significant difference between the corrected picture of the reference run and when either 060807 or CVL151 were removed. This indicates that the concentrations measured at these stations could not be predicted accurately thus making these stations indispensable to the procedure. Conversely, the concentrations at the other two stations were estimated accurately by the procedure, which is demonstrated by the statistics of the procedure run missing these stations being nearly identical to the reference run.

A detailed analysis, however, will need to be completed to determine precisely why the interpolation procedure could not accurately predict the ozone concentrations at each key site. One possible explanation for this result includes the distance between sites. In the case of CVL151, its closest neighbor is at least ten grid squares, or 360 km, away. Therefore, it is probable that the monitoring stations that surround it are too distant to predict accurately the ozone concentration there. The reasons behind 060807 being a key site are more difficult to comprehend.

The site 060807 is located just Northwest of Lake Superior, and Northeast of 063901. In fact, these two stations are at most three grid squares apart as shown in Figure 32. Moreover, 063901 is not a key station, meaning the concentrations at 063901 can be predicted accurately with the use of 060807 and other surrounding stations. Conversely,

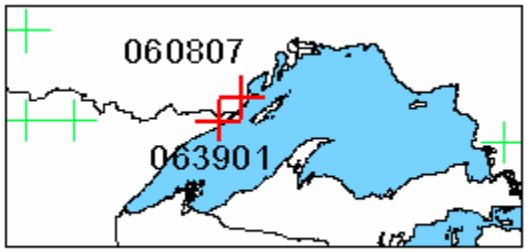

Figure 8.1: Location of 060807 and 063901

the concentrations at 060807 cannot be predicted using the concentrations at 063901 and neighboring stations. As shown in the figure, 063901 is more effectively surrounded by its neighbors than 060807, which could contribute to this phenomenon. Additionally, the closeness of 060807 to Lake Superior may also contribute to the uncertainty of the ozone concentrations measured there. As a result, an enhanced understanding of the geophysical nature of the area surrounding 060807 will need to be established to better determine the reasons for this station's relative importance. Furthermore, the possibility of this station being located downwind of an emission source, thus causing it to measure unusual concentrations, also remains. The likelihood of this placement will need to be explored in order to determine its validity as a key station.

While the two stations CVL151 and 060807 were determined to be essential to the interpolation and assimilation procedures, the 8-hour maximums at the other two stations, ABT147 and 063901, were predicted accurately by the procedure. Therefore, the network of monitoring stations would not be significantly affected were these stations to be shut down. As a result, using this procedure to test the importance of other stations may reveal a number of stations whose value to the network is minimal. Further analysis is needed to determine combinations of stations that can safely be removed. For example, removing one of two stations that are near to each other may be not greatly affect the interpolation procedure, but removing both might cause problems. Consequently, the MOE could save money by the removal of some monitoring stations, but much more analysis is needed to determine the optimal sites to shut down.

While the project overall was a success, there were limitations. For example, the border regions of the model domain could not be tested or corrected because of the lack of monitoring stations. Furthermore, measurements over the Great Lakes were missing, thus causing some uncertainty in the corrected values. As a result, these regions would benefit greatly from the addition of monitoring stations.

The project was also limited by the lack of measurements in the upper troposphere. Because of this, only the surface level of the model output could be corrected.

Consequently, the corrected model output could not be used as corrected initial conditions for subsequent model runs. This limitation, however, could perhaps be overcome by the vertical extrapolation of surface level measurements. Once a method for measurement extrapolation has been formulated, true four dimensional data assimilation for ozone modeling can be done.

While we were unable to accomplish four dimensional data assimilation for ozone in this project, the results of this work show that there could be significant improvement to model results using such a procedure. Moreover, this procedure can be used without modification to improve background concentrations for microscale modeling projects and to archive accurate ozone concentrations. Lastly, statistical tests of the corrected model output can be used to evaluate the importance of individual monitoring stations, as well as to locate geographical regions where a model consistently simulates concentrations incorrectly. Both of these tests are valuable because they can lead to the improvement of both the monitoring network and the model itself. Therefore, a major advancement towards accurate air quality forecasting has been successfully completed by this study.

9 Future Work

This project had two separate goals from its inception. The first was to progress towards providing accurate air quality forecasts, while the second was to develop a method by which the network of monitoring stations could be evaluated — one that would find which stations could be removed from the network without loss of quality to the measurement set. While major steps were taken towards achieving both goals, more work is required before they will be completely usable. In addition, there are other tests that can be done to further analyze the procedure developed here.

Chemical transport modeling results need vast improvement before they can be used for air quality forecasting. Model predictions can be improved upon by increasing the accuracy of the initial conditions for ozone. Yet, before the initial conditions of ozone can be corrected, surface level measurements need to be extrapolated vertically. In the troposphere, vertical mixing is a significant force, and therefore corrections made only in the surface layer would be lost within few model iterations. This problem would be solved, however, if the ozone concentrations could be corrected in the three spatial dimensions by extrapolating surface level measurements vertically. At that juncture, corrections could be made before each model run, as is done with meteorology, to complete a true four-dimensional data assimilation procedure. Given the improvements made to model output in this project, however, FDDA promises to make vast improvements to model predictions of ozone.

Using a four dimensional data assimilation procedure for ozone would correct the model indirectly. Another way to improve model results is to improve the quality of the model inputs. The quality of the emissions inventory is the most uncertain of model inputs. Since ozone is considered a secondary pollutant and not directly emitted to the atmosphere, the model will continue to produce inaccurate amounts of ozone if the concentrations of ozone precursors are inaccurate. Therefore, using an assimilation approach similar to Dominguez and Russell [Dominguez and Russell, 2001] to improve emissions input to the model would be beneficial. This does not negate the need for correcting ozone concentrations, as ozone is highly reactive and, thus, reactions that

destroy it would contribute significantly to NO_2 concentrations, for example. These products are important for both further ozone production and destruction. Furthermore, using corrected meteorology input has also been shown to improve chemical transport model results [El Serafy et al. 2002; Lamarque et al., 2002]. Therefore incorporating several of these techniques would provide a powerful ozone prediction capability.

In addition to aiding in more accurate air quality forecasts, this procedure could be used to provide government scientists with capabilities to determine which monitoring stations could be shut down without degrading the quality of the measurement network. Such an analysis was started in this project, but could not be completed. Thus, a full study to find each key site needs to be undertaken. This will not, however, complete the test to find dispensable sites since a group of sites may exist such that, if one site from the group is removed, the others can accurately predict the missing measurement. Conversely, if more than one is omitted, the remaining stations may not be able to predict the missing values accurately. Therefore a detailed analysis of the network will need to be completed in order to find groups of stations that can safely be removed from the network.

Further testing can also be done to better analyze the effectiveness of our procedure. For example the last step of the procedure, where the full complement of measured data is subtracted from the corrected model output, can be changed for the situation when half the monitoring stations were used. In this case, instead of subtracting the full complement, we could subtract just those values from the half of stations that were not used in the procedure. This tests only those known values which were estimated by the procedure. We can also test how sensitive the procedure is to errors in measurement by intentionally adding noise, or incorrect values, to the measurements. In this way, a sensitivity analysis would be performed on the procedure itself. Both these tests would provide further tests to the effectiveness and limitations of the interpolation and assimilation procedures.

By developing a data assimilation procedure for ozone to work with a chemical transport model, significant progress has been made towards accurate air quality forecasts and

monitoring network evaluation. Furthermore, while the progress made in this project was finite in both of its aims, much potential exists for its goals to be realized. For air quality forecasting, a four dimensional data assimilation procedure will need to be developed by first extrapolating the surface level measurements vertically, and using these three dimensional measurements to correct model output over time. Using techniques developed by other groups, in addition to ozone FDDA, will further enhance the quality of the model output and will likely lead to a high-quality air quality forecasting system.

Suggestions for further testing of the procedure were also given. These will aid in the finding of limitations of the procedure, as well as testing its effectiveness.

Lastly, evaluating which monitoring stations can be removed from the network without degrading the quality of the set of measurements it provides will require a full sensitivity analysis of the network using the procedures developed in this study. This method will include finding key sites as well as groups of sites that can be removed from the network, with saving government money by shutting down surplus monitoring stations being the end result of this work.

Appendix A: ADOM's Gas Phase Chemistry Mechanism

	Reactions	Rate Constant (cm^3 molecule s units)
(R1)	$NO_2 + h\nu \longrightarrow NO + O_3$	radiation dependent
(R2)	$NO + O_3 \longrightarrow NO_2 + O_2$	$2.2 \times 10^{-12} e^{-1430/T}$
(R3)	$NO_2 + O_3 \longrightarrow NO_3 + O_2$	$1.2 \times 10^{-13} e^{-2450/T}$
(R4)	$NO + NO_3 \longrightarrow 2NO_2$	$8.0 \times 10^{-12} e^{250/T}$
(R5)	$NO_2 + NO_3 \longrightarrow N_2O_5$	PT dependent
(R6)	$N_2O_5 \longrightarrow NO_2 + NO_3$	special function
(R7)	$NO_2 + NO_3 \longrightarrow NO + NO_2 + O_2$	$2.5 \times 10^{-14} e^{-1230/T}$
(R8)	$NO_3 + h\nu \longrightarrow 0.15NO + 0.85NO_2 + 0.85O_3 + O_2$	$3.29k_1$
(R9)	$NO_3 + HO_2 \longrightarrow HNO_3 + O_2$	2.5×10^{-12}
(R10)	$O_3 + h\nu \longrightarrow 2OH$ (H_2O dependent)	special function
(R11)	$NO + OH \longrightarrow HONO$	PT dependent
(R12)	$HONO + h\nu \longrightarrow NO + OH$	$0.205k_1$
(R13)	$NO_2 + OH \longrightarrow HNO_3$	PT dependent
(R14)	$HNO_3 + h\nu \longrightarrow NO_2 + OH$	radiation dependent
(R15)	$HNO_3 + OH \longrightarrow NO_3 + H_2O$	$9.4 \times 10^{-15} e^{778/T}$
(R16)	$N_2O_5 + H_2O \longrightarrow 2HNO_3$	1.3×10^{-21}
(R17)	$CO + OH \longrightarrow HO_2 + CO_2$	special function
(R18)	$O_3 + OH \longrightarrow HO_2 + O_2$	$1.6 \times 10^{-12} e^{-940/T}$
(R19)	$NO + HO_2 \longrightarrow NO_2 + OH$	$3.7 \times 10^{-12} e^{240/T}$
(R20)	$NO_2 + HO_2 \longrightarrow HNO_4$	special function
(R21)	$HNO_4 \longrightarrow NO_2 + HO_2$	special function
(R22)	$O_3 + HO_2 \longrightarrow OH + 2O_2$	$1.4 \times 10^{-14} e^{-580/T}$
(R23)	$HO_2 + HO_2 \longrightarrow H_2O_2 + O_2$ (H_2O dependent)	special function
(R24)	$H_2O_2 + h\nu \longrightarrow 2OH$	radiation dependent
(R25)	$H_2O_2 + OH \longrightarrow HO_2 + H_2O$	$3.1 \times 10^{-12} e^{-187/T}$
(R26)	$NO_2 + H_2O \longrightarrow HONO + HNO_3 - NO_2$	4.0×10^{-24}
(R27)	$HNO_4 + h\nu \longrightarrow NO_2 + HO_2$	$(1.0 \times 10^{-4})k_1$
(R28)	$HNO_4 + OH \longrightarrow NO_2 + H_2O + O_2$	4.0×10^{-12}
(R29)	$SO_2 + OH \longrightarrow SO_4 + HO_2$	PT dependent
(R30)	$HCHO + h\nu \longrightarrow 2HO_2 + CO$	radiation dependent
(R31)	$HCHO + h\nu \longrightarrow CO + H_2$	radiation dependent

	Reactions			Rate Constant (cm^3 molecule s units)
(R32)	$HCHO + OH$	$\longrightarrow$	$HO_2 + CO + H_2O$	1.0×10^{-11}
(R33)	$HCHO + HO_2$	$\longrightarrow$	AHO_2	1.0×10^{-14}
(R34)	$AHO_2 + NO$	$\longrightarrow$	$ACO_2 + HO_2 + NO_2$	$4.2 \times 10^{-12} e^{180/T}$
(R35)	$AHO_2 + HO_2$	$\longrightarrow$	$ACO_2 + H_2O + O_2$	2.0×10^{-12}
(R36)	$AHO_2 + AHO_2$	$\longrightarrow$	$2ACO_2 + 2HO_2 + 2O_2$	1.0×10^{-13}
(R37)	$ACO_2 + OH$	$\longrightarrow$	$HO_2 + H_2O + CO_2$	3.2×10^{-13}
(R38)	$NO_3 + HCHO$	$\longrightarrow$	$HNO_3 + HO_2 + CO$	3.2×10^{-16}
(R39)	$ALD_2 + OH$	$\longrightarrow$	$MCO_3 + H_2O$	$6.9 \times 10^{-12} e^{250/T}$
(R40)	$ALD_2 + NO_3$	$\longrightarrow$	$HNO_3 + MCO_3$	1.4×10^{-15}
(R41)	$ALD_2 + h\nu$	$\longrightarrow$	$MO_2 + HO_2 + CO$	radiation dependent
(R42)	$ALD_2 + h\nu$	$\longrightarrow$	$CH_4 + CO$	radiation dependent
(R43)	$MCO_3 + NO_2$	$\longrightarrow$	PAN	4.7×10^{-12}
(R44)	PAN	$\longrightarrow$	$MCO_3 + NO_2$	$1.9 \times 10^{16} e^{-13543/T}$
(R45)	$MCO_3 + NO$	$\longrightarrow$	$MO_2 + NO_2 + CO_2$	$4.2 \times 10^{-12} e^{180/T}$
(R46)	$MO_2 + NO$	$\longrightarrow$	$HCHO + NO_2 + HO_2$	$4.2 \times 10^{-12} e^{180/T}$
(R47)	$CH_4 + OH$	$\longrightarrow$	$MO_2 + H_2O$	$2.4 \times 10^{-12} e^{-1710/T}$
(R48)	$C_2H_6 + OH$	$\longrightarrow$	$ETO_2 + H_2O$	$1.7 \times 10^{-11} e^{-1232/T}$
(R49)	$ETO_2 + NO$	$\longrightarrow$	$ALD_2 + HO_2 + NO_2$	$4.2 \times 10^{-12} e^{180/T}$
(R50)	$C_3H_8 + OH$	$\longrightarrow$	R_3O_2	$1.18 \times 10^{-11} e^{-679/T}$
(R51)	$R_3O_2 + NO$	$\longrightarrow$	$0.03R_3N_2 + 0.46ALD_2 + 0.97NO_2 + 0.97HO_2 + 0.49KET$	$4.2 \times 10^{-12} e^{180/T}$
(R52)	$ALKA + OH$	$\longrightarrow$	RAO_2	$2.0 \times 10^{-11} e^{-500/T}$
(R53)	$RAO_2 + NO$	$\longrightarrow$	$\beta_1 NO_2 + \beta_2 NO + \beta_3 RAN_2 + \beta_4 ALD_2 + \beta_5 KET + \beta_6 ETO_2 + \beta_7 MO_2 + \beta_8 HO_2 + \beta_9 R_3O_2 + 0.06RAO_2$	$4.2 \times 10^{-12} e^{180/T}$
(R54)	$ALKA + NO_3$	$\longrightarrow$	$HNO_3 + RAO_2$	4.0×10^{-17}
(R55)	$RAN_2 + OH$	$\longrightarrow$	$RAN_1 + H_2O$	2.0×10^{-12}
(R56)	$RAN_1 + NO$	$\longrightarrow$	$2.5NO_2 - 0.5NO + 0.8HCHO + 2.1ALD_2$	$4.2 \times 10^{-12} e^{180/T}$
(R57)	$MO_2 + MO_2$	$\longrightarrow$	$1.4HCHO + 0.8HO_2 + O_2$	$1.5 \times 10^{-13} e^{220/T}$
(R58)	$ETO_2 + ETO_2$	$\longrightarrow$	$1.6ALD_2 + 1.2HO_2$	5.0×10^{-14}
(R59)	$R_3O_2 + R_3O_2$	$\longrightarrow$	$1.9ALD_2 + 0.28KET + 0.37HO_2$	5.0×10^{-14}
(R60)	$HO_2 + MO_2$	$\longrightarrow$	$ROOH + O_2$	5.0×10^{-12}
(R61)	$HO_2 + ETO_2$	$\longrightarrow$	$ROOH + O_2$	5.0×10^{-12}

	Reactions			Rate Constant (cm^3 molecule s units)
(R62)	$HO_2 + R_3O_2$	$\longrightarrow$	$ROOH + O_2$	5.0×10^{-12}
(R63)	$HO_2 + RAO_2$	$\longrightarrow$	$ROOH + O_2$	5.0×10^{-12}
(R64)	$HO_2 + MCO_3$	$\longrightarrow$	$ROOH + O_2$	5.0×10^{-12}
(R65)	$KET + OH$	$\longrightarrow$	KO_2	$1.2 \times 10^{-11} e^{-890/T}$
(R66)	$KO_2 + NO$	$\longrightarrow$	$0.05RAN_2 + 0.95NO_2 + 0.94ALD_2 + 0.94MCO_3$	$4.2 \times 10^{-12} e^{180/T}$
(R67)	$KET + h\nu$	$\longrightarrow$	$MCO_3 + ETO_2 + H_2O$	$2.6 \times 10^{-4} k_1$
(R68)	$KET + NO_3$	$\longrightarrow$	$HNO_3 + KO_2$	7.0×10^{-16}
(R69)	$KO_2 + HO_2$	$\longrightarrow$	$MGLY + MO_2 + H_2O$	3.0×10^{-12}
(R70)	$ETHE + OH$	$\longrightarrow$	EO_2	$1.66 \times 10^{-12} e^{474/T}$
(R71)	$EO_2 + NO$	$\longrightarrow$	$NO_2 + 2.0HCHO + HO_2$	$4.2 \times 10^{-12} e^{180/T}$
(R72)	$ALKE + OH$	$\longrightarrow$	PO_2	$4.1 \times 10^{-12} e^{537/T}$
(R73)	$PO_2 + NO$	$\longrightarrow$	$NO_2 + ALD2 + HCHO + HO_2$	$4.2 \times 10^{-12} e^{180/T}$
(R74)	$ETHE + O_3$	$\longrightarrow$	$HCHO + 0.4CHO_2 + 0.12HO_2 + 0.42CO + 0.06CH_4$	$1.2 \times 10^{-14} e^{-2633/T}$
(R75)	$ALKE + O_3$	$\longrightarrow$	$0.525HCHO + 0.5ALD2 + 0.2CHO_2 + 0.2CRO_2 + 0.23HO_2 + 0.215MO_2 + 0.095OH + 0.33CO$	$7.8 \times 10^{-14} e^{-2105/T}$
(R76)	$CHO_2 + NO$	$\longrightarrow$	$HCHO + NO_2$	7.0×10^{-12}
(R77)	$CHO_2 + NO_2$	$\longrightarrow$	$HCHO + NO_3$	7.0×10^{-13}
(R78)	$CHO_2 + H_2O$	$\longrightarrow$	ACO_2	4.0×10^{-18}
(R79)	$CRO_2 + NO$	$\longrightarrow$	$ALD2 + NO_2$	7.0×10^{-12}
(R80)	$CRO_2 + NO_2$	$\longrightarrow$	$ALD2 + NO_3$	7.0×10^{-13}
(R81)	$CRO_2 + H_2O$	$\longrightarrow$	$ACTA$	4.0×10^{-18}
(R82)	$EO_2 + EO_2$	$\longrightarrow$	$2.4HCHO + 1.2HO_2 + 0.4ALD2$	5.0×10^{-14}
(R83)	$PO_2 + PO_2$	$\longrightarrow$	$2.2ALD2 + 1.2HO_2$	5.0×10^{-14}
(R84)	$HO_2 + EO_2$	$\longrightarrow$	$ROOH + O_2$	3.0×10^{-12}
(R85)	$HO_2 + PO_2$	$\longrightarrow$	$ROOH + O_2$	3.0×10^{-12}
(R86)	$SO_2 + CHO_2$	$\longrightarrow$	$SO_4 + HCHO$	7.0×10^{-14}
(R87)	$SO_2 + CRO_2$	$\longrightarrow$	$SO_4 + ALD2$	7.0×10^{-14}
(R88)	$ALKE + NO_3$	$\longrightarrow$	PRN_1	1.26×10^{-13}
(R89)	$PRN_1 + NO_2$	$\longrightarrow$	PRN_2	6.8×10^{-12}
(R90)	$PRN_1 + HO_2$	$\longrightarrow$	$PRPN + O_2$	3.0×10^{-12}
(R91)	$PRN_1 + NO$	$\longrightarrow$	$2NO_2 + HCHO + ALD2$	$4.2 \times 10^{-12} e^{180/T}$

	Reactions			Rate Constant (cm^3 molecule s units)
(R92)	$CHO_2 + HCHO$	$\longrightarrow$	OZID	1.36×10^{-14}
(R93)	$CHO_2 + ALD2$	$\longrightarrow$	OZID	1.36×10^{-14}
(R94)	$CRO_2 + HCHO$	$\longrightarrow$	OZID	1.36×10^{-14}
(R95)	$CRO_2 + ALD2$	$\longrightarrow$	OZID	1.36×10^{-14}
(R96)	AROM + OH	$\longrightarrow$	$0.84TO_2 + 0.16CRES + 0.16HO_2$	1.52×10^{-11}
(R97)	$TO_2 + NO$	$\longrightarrow$	$NO_2 + HO_2 + 0.72MGLY + 0.18GLYX + DIAL$	$4.20 \times 10^{-12} e^{180/T}$
(R98)	GLYX + *hv*	$\longrightarrow$	PROD	$8.00 \times 10^{-3} k_1$
(R99)	GLYX + OH	$\longrightarrow$	$HO_2 + 2.0CO + H_2O$	1.15×10^{-11}
(R100)	MGLY + *hv*	$\longrightarrow$	$MCO_3 + HO_2 + CO$	$1.90 \times 10^{-2} k_1$
(R101)	MGLY + OH	$\longrightarrow$	$MCO_3 + CO + H_2O$	1.73×10^{-11}
(R102)	CRES + OH	$\longrightarrow$	$\beta_{12}HO_2 + 0.9ZO_2 + 0.9TCO_3 - 0.9OH + \beta_{13}NO_2$	4.25×10^{-11}
(R103)	NO_3 + CRES	$\longrightarrow$	$HNO_3 + \beta_{10}NO_2 + \beta_{11}OH$	1.00×10^{-11}
(R104)	OH + DIAL	$\longrightarrow$	$TCO_3 + H_2O$	2.80×10^{-11}
(R105)	TCO_3 + NO2	$\longrightarrow$	TPAN	4.70×10^{-12}
(R106)	TPAN	$\longrightarrow$	$TCO_3 + NO_2$	$1.95 \times 10^{16} e^{-13543/T}$
(R107)	TCO_3 + NO	$\longrightarrow$	$NO_2 + 0.92HO_2 + 0.89GLYX + 0.11MGLY + 0.05MCO_3 + 0.95CO + 0.79CO_2 + 2.0ZO_2$	$4.2 \times 10^{-12} e^{180/T}$
(R108)	ZO_2 + NO	$\longrightarrow$	NO_2	$4.2 \times 10^{-12} e^{180/T}$
(R109)	DIAL + *hv*	$\longrightarrow$	$0.98HO_2 + 0.02MCO_3 + TCO_3$	$1.00 \times 10^{-2} k_1$
(R110)	$HO_2 + TO_2$	$\longrightarrow$	$ROOH + O_2$	4.00×10^{-12}
(R111)	$HO_2 + TCO_3$	$\longrightarrow$	$ROOH + O_2$	4.00×10^{-12}
(R112)	$HO_2 + ZO_2$	$\longrightarrow$	$ROOH + O_2$	4.00×10^{-12}

10 Reference List

Armstrong, M., *Basic Linear Geostatistics*, Springer, Verlag, Berlin, Heidelberg, New York, 1998.

Barna, M. and B. Lamb, Improving ozone modeling in regions of complex terrain using observational nudging in a prognostic meteorological model, *Atmospheric Environment*, 34, 4889-4906, 2000.

Bloxam, R., Neiwiadomski, M, Wong, S., and Misra, P. K. Influence of Boundary Conditions and Surface Characteristics on Fine Mesh Urban and Regional Models. 1993. San Diego, California. Proceedings of A&WMA Conference on Regional Photochemical Measurments and Modeling Studies. 1993.

Brasseur, G., Orlando, J., Tyndall, G., *Atmospheric Chemistry and Global Change*, Oxford University Press, New York, Oxford, 1999.

Chtcherbakov, A., Meteorology Input for ADOM, (personal communication, 2002).

Chvatal, V., *Linear Programming*, W. H. Freeman and Company, New York, 1983.

Dann, T., NAPS network ozone data, 1996, (personal communication, 2002).

El Serafy, G., R. van der A, H. Eskes, and H. Kelder, Assimilation of 3D ozone field in global chemistry-transport models using Kalman filter, *Advances in Space Research*, 30, 2473-2478, 2002.

Elbern, H., H. Schmidt, and A. Ebel, Variational data assimilation for tropospheric chemistry modeling, *Journal of Geophysical Research, [Atmospheres]*, 102, 15967-15985, 1997.

Elbern, H. and H. Schmidt, Ozone episode analysis by four-dimensional variational chemistry data assimilation, *Journal of Geophysical Research, [Atmospheres]*, 106, 3569-3590, 2001.

Jacob, D. J., *Introduction to Atmospheric Chemistry*, Princeton University Press, Princeton, New Jersey, 1999.

Lamarque, J. F., B. V. Khattatov, and J. C. Gille, Constraining tropospheric ozone column through data assimilation, *Journal of Geophysical Research, [Atmospheres]*, 107, ACH9-1-ACH9/11, 2002.

Lefohn, A. S., H. P. Knudsen, J. A. Logan, J. Simpson, and C. Bhumralkar, An Evaluation of the Kriging Method to Prdict 7-H Seasonal Mean Ozone Concentrations for Estimating Crop Losses, *Journal of Air Pollution Control Association*, 37, 595-602, 1987.

Lurmann, F. W. and A. C. Lloyd, A Chemical Mechanism for Use in Long-Range Transport/Acid Deposition Computer Modeling, *Journal of Geophysical Research*, 91, 10905-10936, 1986.

Mendoza-Dominguez, A. and A. G. Russell, Emission strength validation using four-dimensional data assimilation: application to primary aerosol and precursors to ozone and secondary aerosol, *Journal of the Air and Waste Management Association*, 51, 1538-1550, 2001.

Nielsen, T., Hertel, O., Christensen, C. S., Egeloev, A. H., Granby, K., Hansen, A. B., Platz, J., and Skov, H., Comparison of measurements and modeling of ozone, other photochemical oxidants, precursors and atmospheric reaction products during a summer

high pressure episode in Denmark. 1. 1997. Proc. EUROTRAC Symp. '96: Transp. Transform. Pollut. Troposphere, 4th.

Ontario Ministry of the Environment, E. M. R. B. “Air Quality in Ontario, 2002”, Report, 32, Ontario Ministry of the Environment. Air Quality in Ontario, 2003.

Scire, J. S., F. W. Lurmann, P. Karamchandani, A. Venkatram, R. Yamartino, J. Young, and J. Pleim, ADOM*/TADAP Model Development Program, Volume 9, ADOM/TADAP User's Guide,*1986.

Seinfeld, J. H. and S. N. Pandis, *Atmospheric Chemistry and Physics, From Air Pollution to Climate Change*, John Wiley & Sons, Inc., 1998.

USEPA, Emissions Inventories, *ftp://ftp.epa.gov/EmisInventory/*, 2002a.

USEPA, CASTNET hourly ozone data, 1996, *http://www.epa.gov/castnet/ozone.html*, 2002b.