12-1-2007

Ground-Level Ozone Across Kentucky: Modeling and a Synoptic Analysis of High Concentrations

John Walker
Western Kentucky University

Follow this and additional works at: http://digitalcommons.wku.edu/theses

Part of the Geology Commons

Recommended Citation
http://digitalcommons.wku.edu/theses/385

This Thesis is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Masters Theses & Specialist Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact topscholar@wku.edu.
GROUND-LEVEL OZONE ACROSS KENTUCKY: MODELING AND A SYNOPTIC
ANALYSIS OF HIGH CONCENTRATIONS

A Thesis
Presented to
The Faculty of the Department of Geography and Geology
Western Kentucky University
Bowling Green, Kentucky

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science

John Miller Walker, Jr.

December 2007
GROUND-LEVEL OZONE ACROSS KENTUCKY: MODELING AND SYNOPTIC ANALYSIS OF HIGH CONCENTRATIONS

Date Recommended ______________________

Dr. Gregory B. Goodrich, Director of Thesis

Dr. Stuart A. Foster

Dr. Jun Yan

Dr. L. Michael Trapasso

Dean, Graduate Studies and Research  Date
12/11/07
# Table of Contents

Table of Contents ................................................................. i

List of Figures ........................................................................... ii

List of Tables ............................................................................. v

Abstract ...................................................................................... vi

## Chapter

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction ........................................................................ 3</td>
<td></td>
</tr>
<tr>
<td>2. Literature Review .................................................................. 10</td>
<td></td>
</tr>
<tr>
<td>3. Data and Methodology ................................................................ 27</td>
<td></td>
</tr>
<tr>
<td>3.1 Data .................................................................................. 27</td>
<td></td>
</tr>
<tr>
<td>3.1.1 Model Development ....................................................... 27</td>
<td></td>
</tr>
<tr>
<td>3.2 Methodology ....................................................................... 29</td>
<td></td>
</tr>
<tr>
<td>3.2.1 Assessing Model Performance ......................................... 31</td>
<td></td>
</tr>
<tr>
<td>3.2.2 Synoptic Patterns Associated with High O₃ Concentrations .... 32</td>
<td></td>
</tr>
<tr>
<td>3.2.3 Impact of Potential Policy Changes .................................. 32</td>
<td></td>
</tr>
<tr>
<td>4. Results and Discussion ................................................................ 34</td>
<td></td>
</tr>
<tr>
<td>4.1 Development of the Standard Model ....................................... 34</td>
<td></td>
</tr>
<tr>
<td>4.2 Development of the Hi Model ............................................... 39</td>
<td></td>
</tr>
<tr>
<td>4.3 Operational use of the Hybrid Model for Kentucky during June 2007</td>
<td>40</td>
</tr>
<tr>
<td>4.4 Synoptic Analysis of High O₃ Concentrations ......................... 47</td>
<td></td>
</tr>
<tr>
<td>4.5 Environmental Protection Agency’s Exceedance Policies .......... 54</td>
<td></td>
</tr>
<tr>
<td>4.5.1 Changes in the 1-hour and 8-hour Exceedance Policy .......... 55</td>
<td></td>
</tr>
<tr>
<td>4.5.2 Proposed 8-hour Exceedance Policy .................................... 58</td>
<td></td>
</tr>
<tr>
<td>5. Conclusions and Future Research .......................................... 61</td>
<td></td>
</tr>
</tbody>
</table>

References ................................................................................. 65
List of Figures

Figure 1: Map of AQS and CASNET monitors across Kentucky as of 2005. Source: AIRS (2007), USDA (2007), annotated by author .......................................................... 5

Figure 2: Location of the AQS monitor at MCNP. Source: NAUS (2006), AIRS (2007), annotated by author .......................................................... 6

Figure 3: Metropolitan Statistical Areas that have a daily $O_3$ forecast. Source: Cobourn (2006), Cobourn and Lin (2004), US Census (2006), annotated by author ............................................. 7

Figure 4: This figure shows the concentration of $O_3$ in the Earth’s atmosphere measured in milli-Pascals of pressure. Source: Gleanson (2006) .......................................................... 10

Figure 5: The Air Quality Index was developed by the EPA to help the public understand the health affects of $O_3$. 8-Hour $O_3$ concentration and Air Quality Index values are matched with a specific color. This index makes it easier for the general public to know what precautions they need to take in order to avoid adverse health effects. Source: ENHS (2007), USEPA (2007a), annotated by author .......................................................... 14

Figure 6: The diagram shows the vertical structure of the boundary layer throughout the day. Source: Atkins (2001), annotated by author .................................................. 17

Figure 7: This graph represents the cycle in $O_3$ concentration during a typical day for a low-elevated location downwind of an urban center. Source: McKendry and Ludgren (2000) .................................................. 18

Figure 8: Tong et al. (2006) studied the spatiotemporal differences between $O_3$ concentrations at different elevations for GSM and MCNP. Source: AIRS (2007), NAUS (2006), annotated by author .................................................. 19

Figure 9: This figure shows the diurnal pattern of 8-hour $O_3$ exceedance for MCNP and GSM during 1996-1998. A count of one was used every time an hour of the day reached exceedance. Those counts were summed and then graphed by hour. Source: Tong et al. (2006), annotated by author .................................................. 20

Figure 10: Spatial interpolation methods of Inverse Distance Weighting (a), Spline (b), and Kriging (c) give different results. This figure shows the result of the three different methods when elevation is interpolated from a grid of sample points. Source: Childs (2004), annotated by author .................................................. 23
Figure 11: This map shows MOS stations used by the NWS in the Continental United States. Source: NOAA (2006b, 2007), annotated by author.

Figure 12: This map shows the 50 MOS stations that was used as sample points for the O₃ model in this research. Source: NOAA (2006b, 2007), annotated by author.

Figure 13: A 30 day moving average of O₃ subtracted by Pᵢ for MCNP during the 2004 and 2005 O₃ seasons. The negative (positive) values in the graph are days when the model over- (under-) predicted O₃. Source: NPSARD.

Figure 14: Graphical representation of the difference in Oᵢ at MCNP and Pᵢ for KBWG during the June 2007 model run. June 17th is missing due to incomplete data. Source: NPSARD, NOAA (2006a).

Figure 15: Kentucky O₃ forecast for June 24, 2007. Source: NOAA (2006b, 2007), annotated by author.

Figure 16: Kentucky O₃ forecast for June 11, 2007. Source: NOAA (2006b, 2007), annotated by author.

Figure 17: Standard Error Map of the Kentucky O₃ forecast for June 11, 2007.

Figure 18: Standard Error Map of the Kentucky O₃ forecast for June 24, 2007.

Figure 19: Distribution of O₃ in the 95th percentile of MCNP 98-03 dataset. Source: NPSARD.

Figure 20: Temporal distribution of Dry Tropical days that were within the 95th percentile. Source: NPSARD.

Figure 21: Temporal distribution of Dry Moderate days that were within the 95th percentile. Source: NPSARD.

Figure 22: Temporal distribution of Moist Tropical days that were within the 95th percentile. Source: NPSARD.

Figure 23: The National Drought Mitigation Centers outlook on drought conditions for August 31, 1999. Source: NDMC (1999).
Figure 24: A moving average of 45 days was applied to 1-hour O₃ observations during the summer months from 1998 to 2005. The exceedance limit of 125 ppb is indicated by the red line. Source: NPSARD.

Figure 25: A moving average of 45 days was applied to 8-hour O₃ observations during the summer months from 1998 to 2005. The exceedance limit of 85 ppb is indicated by the red line. Source: NPSARD.

Figure 26: A moving average of 45 days was applied to 8-hour O₃ observations during the summer months from 1998 to 2005. The exceedance limit is within a range of 70 to 75 ppb indicated by the red line. Source: NPSARD.

Figure 27: Counties that would obtain non-attainment status for the proposed 8-hour average ambient O₃ policy. Source: USEPA (2007b).
List of Tables

Table 1: Results of correlation testing between O₃ and meteorological variables for the Standard model. The model evaluation statistics of the model are also shown..............35

Table 2: Summary of coefficients from the correlation testing in Table 1.......................36

Table 3: The monthly PDSI values for the central climate division of Kentucky for the 1998 to 2005 O₃ seasons. Source: ESRL (2007).................................................38

Table 4: Summary of coefficients from the Hi model correlation testing.......................39

Table 5: Results of correlation testing between O₃ and meteorological variables for the Hi model.........................................................................................40

Table 6: MOS forecast for KBWG and observations made at MCNP for June 2007. Days that the Hybrid model selected the Standard (Hi) model are denoted by S (H). The difference between the O₃ at MCNP and the Pₛ at KBWG are also shown. The 17th of June is missing because of incomplete data on that day. Source: NPSARD.........................42

Table 7: The 95th percentile of O₃ stratified by air mass. Time period for the 95th percentile is 1998 to 2003. Source: NPSARD.........................................................49

Table 8: PDSI values and the corresponding classifications that define the index values. Source: Palmer (1965), annotated by author.....................................................54

Table 9: PDSI values for Kentucky’s Central Climate Division each month during 1998 to 2003. Source: ESLR (2007).................................................................54

Table 10: Number of days that exceeded the 1- and 8-hour average ambient O₃ policies. Source: NPSARD.................................................................57

Table 11: Number of days that would exceed the proposed 8-hour average ambient O₃ policy. Source: NPSARD.................................................................58
Abstract

Rural areas are often more susceptible to high concentrations of ground-level ozone (O₃) than urban areas. However, rural populations are, for the most part, unaware of this problem. Currently the rural areas of Kentucky have no daily forecast for O₃.

This research addresses the issue by using methodologies from previous Kentucky O₃ modeling research to develop a daily forecast model within Geographic Information Systems. The rural O₃ model developed by Kendrick (2005) will be used in this research, as a Standard model, along with an application of the model introduced by Cobourn and Hubbard (1999), as the Hi model, to be used on days that O₃ concentrations are expected. When the forecasted maximum temperature is less (greater) than 87°F, the diurnal temperature range is less (greater) than 27, and the probability for precipitation is less (greater) than 50 percent then the Hybrid will choose the Standard (Hi) model. Data for the both models came from the Model Output Statistic by the National Weather Service. The Standard model proves to be successful in forecasting O₃ while the Hi model is less accurate.

Synoptic meteorology conditions were analyzed to find patterns that are associated with high O₃ concentrations for rural areas. Data collected at Mammoth Cave National
Park (MCNP) (30 miles north of Bowling Green, KY) during 1998 to 2005 was used in this analysis. The methodology presented by Sheridan (2002) was used to define the overall synoptic patterns that were present during 1998 to 2003. It was found that Dry Moderate and Dry Tropical air masses frequently had high O₃ associated with them during late August and early September.

The Environmental Protection Agency (EPA) is responsible for policies that fulfill the primary and secondary goals of the National Ambient Air Quality Standards (NAAQS), which includes O₃. Prior to 1998 it was determined that an 1-hour average ambient O₃ measurement less than 125 parts per billion (ppb) would be sufficient in achieving both the NAAQS primary and secondary goals. However, it was found that health problems could still occur at levels less than 125 ppb, so the policy was changed to an 8-hour average of 85 ppb. Many researchers explained that this new policy might cause rural areas to break exceedance more often (Baumgardner and Edgerton, 1998; Cobourn and Hubbard, 1999; Barna et al., 2001; Sistla et al., 2001; Reynolds et al., 2003). In this research it was found that the number of exceedance days at MCNP increased by 43 days from the 1-hour to the 8-hour policy.

The number of exceedance days has the potential to increase at MCNP if the EPA accepts a proposed 8-hour policy in early 2008. The reason for this proposal is because recently it has been discovered that O₃ at levels lower than 85 ppb for 8-hours can affect human health. The proposal requires that the current 8-hour average be adjusted from 85 ppb to a range within 70 - 75 ppb. If accepted the proposed policy would not take effect until 2013.
Introduction

Tropospheric or ground-level ozone (O$_3$) is an air pollutant that is found near the Earth’s surface. Ozone at the ground-level is a secondary pollutant formed when nitrogen oxide (NO$_x$) molecules, volatile organic compounds (VOCs), and sunlight combine in a photochemical reaction. A volatile organic compound, also known as isoprene, is a gas released naturally from vegetation and artificially from products such as cleaning chemicals. High concentrations of ground-level O$_3$ are a concern to agriculture, ecosystems, and human health (Diem and Comrie, 2001). More than half of the world’s population lives in what are considered urbanized areas, where the combustion of oil byproducts by vehicles and industry creates a large amount of air pollution (Ellis et al., 2000).

In North America, where most of the world’s urban sprawl is taking place, 80 percent of the population lives in urbanized areas. Ozone precursor emissions, of NO$_x$ and VOCs, from these urban areas are transported across long distances by meteorological processes and eventually turn into O$_3$ by the photochemical reaction already described. The problems of high concentrations in one city can become the burden of another city. This is known as geographic spillover. In the eastern United States geographic spillover is a considerably large problem (Garcia, 2006).

Previous studies have found that the Midwest is susceptible to high concentrations of O$_3$. Modeled data from the Ozone Transportation Assessment Group, which represents the 37 eastern states that are affected by O$_3$, show that much of the O$_3$ affecting the Lake Michigan area is generated from the Ohio River Valley (Jeffery, 1997). High-pressure
systems moving through the Midwest interact with increased levels of precursor emissions (mainly NO\textsubscript{X}) from the Ohio River Valley (Aneja et al., 2000). Several power plants and industries located in the Ohio River Valley are responsible for the high precursor emissions (Miller, 1999) of NO\textsubscript{x} (Meagher, 2006).

The problem of high O\textsubscript{3} is not limited to urbanized areas. In rural areas precursor emissions from these urban sources can produce high concentrations of O\textsubscript{3}. Sillman (1999) stated that rural areas are more sensitive to NO\textsubscript{X} than urban areas, because of the VOCs production from vegetation in rural areas (Meagher, 2006). The result is a larger concentration of O\textsubscript{3} for the Ohio River Valley than in metropolitan cities along the east coast such as Boston (Miller, 1999). In fact, on a global scale rural areas actually receive more of the burden (i.e. adverse heath effects) from air pollution than urban areas (Smith, 2006).

There have been extensive studies done on O\textsubscript{3} in urban environments (Jeffery, 1997; Kang et al., 2003) but very little research has been done on rural environments. This is due in large part to O\textsubscript{3} recently being recognized as a regional problem and not a local problem (Jeffery, 1997). The result is that rural populations are unaware of the potentially high O\textsubscript{3} concentrations that can exist and the potential health problems that may occur.

As of 2005 Kentucky had 34 O\textsubscript{3} monitors in operation composed of two different networks operated by the Environmental Protection Agency (EPA) (Figure 1). Of the 34 monitors 29 are part of the Air Quality System (AQS) and four are a part of the Clean Air Status Trends Network (CASNET). AQS encompasses 1100 O\textsubscript{3} monitors across the United States generally in urban areas or population centers. CASNET is composed of 56 stations located in rural areas mainly in the eastern United States (USEPA, 2006c).
Ground-Level Ozone Monitors across Kentucky

![Map of AQS and CASNET monitors across Kentucky as of 2005. Source: AIRS (2007), USDA (2007), annotated by author](image)

Since 1998 O₃ and meteorological observations have been recorded by an AQS monitor at Mammoth Cave National Park (MCNP) in Edmonson County Kentucky (Figure 2). MCNP land usage is described as rural and forested (AIRS, 2007) and has had a history of pollution problems. In 2001, MCNP ranked as one of the top five national parks polluted by O₃ (NPCA, 2006; Pringle, 2004). The data collected at MCNP has helped provide researchers (Kang et al., 2003; Kendrick, 2005; Tong et al., 2006) with a sample of O₃ pollution in the predominately rural state of Kentucky.
While valuable data is being collected from $O_3$ monitors throughout the state, such as the one at MCNP, the rural populations in Kentucky are still unaware of days when potentially high $O_3$ concentration will occur. As of today seven urban areas (one micropolitan and six metropolitan statistical areas) have a daily $O_3$ forecast available during the months from March to October, which are considered $O_3$ season (Cobourn, 2006) (Figure 3).
Ground-Level Ozone Forecast for Kentucky Metropolitan Areas

Metropolitan Statistical Areas

- Paducah
- Clarksville
- Owensboro
- Bowling Green
- Louisville/Jefferson
- Lexington-Fayette
- Cincinnati-Middletown
- Huntington-Ashland

*Paducah is a Metropolitan Statistical Area

Counties highlighted in this map are only ones that fall within Kentucky. The Paducah, Clarksville, Louisville-Jefferson, Cincinnati-Middletown, and Huntington-Ashland Statistical Areas include counties from surrounding states. A complete list of Metropolitan Statistical Areas can be found at: http://www.census.gov/popest/metro/popest_2006_st.txt

During the Ozone Season (March to October) forecasts are available at: http://lavellespd.louisville.edu/ozone/forecast_forestKY.html

Of Kentucky’s 4 million citizens roughly 2 million live in a metropolitan statistical area (US Census, 2007). These areas are mainly found in the western and northern portions of the state along the Ohio River. Many of Kentucky’s rural areas are in the central and eastern part of the state. These areas, which are the most susceptible to high concentrations of O₃, are un-served by a daily O₃ forecast model. An O₃ forecast model for these un-served areas would benefit those Kentuckians living in rural areas. Such a model would bring more attention to the regional implications of high O₃ concentrations across Kentucky.

The primary goal of this research is to create a model that predicts O₃ for rural areas of Kentucky, such as MCNP. The model for this research will apply known relationships between O₃ and meteorological variables at MCNP (Kendrick, 2005) to Kentucky. It has
been found that the relationship between meteorological conditions and concentrations of
O\textsubscript{3} exhibit a strong correlation (Spellman, 1999). The variables that will be used in this
model were found to explain 49 percent of the variance for the maximum 1-hour and 8-
hour O\textsubscript{3} average at MCNP (Goodrich, 2006). The working hypothesis in this research is
that the model will over-estimate days of low O\textsubscript{3} and under-estimate days of high O\textsubscript{3}
based on previous findings by Kendrick (2005).

Two more topics will be examined in this research. These topics include an analysis of
the synoptic meteorology patterns associated with high O\textsubscript{3} concentrations. The 95\textsuperscript{th}
percentile of ranked O\textsubscript{3} recorded at MCNP between 1998 and 2003 will be used to
represent high concentrations. For Kentucky, the synoptic patterns related to high O\textsubscript{3}
should be high-pressure systems with warm/dry air during mid July and August (Aneja et
al., 2000).

The final topic covered in this research will be an examination of the Environmental
Protection Agencies (EPA) O\textsubscript{3} exceedance policy changes within the past decade
(USEPA, 2006a). The previous policy stated that an area is in exceedance if the 1-hour
average contains more than 125 parts per billion (ppb) of O\textsubscript{3}. The current policy states
that an area is in exceedance if the 8-hour average contains more than 85 ppb. The
number of days that exceed the current 8-hour policy will be more than the previous 1-
hour policy for MCNP (Goodrich, 2006).

Also discussed in the exceedance chapter will be the impact of a proposed revision to
the current 8-hour policy that would designate an area in exceedance if the 8-hour
average of O\textsubscript{3} is within a range of 70 to 75 ppb (USEPA, 2007d). The days in exceedance
of the proposed policy will be even more than the current policy. This proposed policy is
awaiting EPA approval which is expected to come some time in March 2008. If approved the policy would not take effect until 2013.
Chapter 2

Literature Review

The atmosphere contains two types of O$_3$; stratospheric and tropospheric. Stratospheric O$_3$ is a natural occurring gas located in the stratosphere. The stratosphere ranges from 10 to 50 km above the Earth's surface. Concentrations of O$_3$ are present throughout the stratosphere, but the highest are between 20 and 30 km above the Earth's surface (Lutgens and Tarbuck, 2001) (Figure 4). This layer of O$_3$ absorbs a large portion of ultraviolet light from the sun (Gleanson, 2006). The mass absorption of ultraviolet light makes it possible to sustain life on Earth. Tropospheric O$_3$, the focus of this research, is found below 3 km. This O$_3$ is a pollutant that is harmful to plant life and to human health.

![Figure 4](image)

Figure 4: This figure shows the concentration of O$_3$ in the Earth's atmosphere measured in mili-Pascals of pressure. Source: Gleanson (2006)

Tropospheric O$_3$ is a secondary pollutant that is formed when a number of chemicals and sunlight (hv) combine in a photochemical reaction. The process of O$_3$ creation begins
with a reaction between VOC and a hydroxide radical (OH) (Equation 1). Nitrogen oxide (NO$_x$) is then converted to nitrogen dioxide (NO$_2$) through a reaction with hydroperoxyl radical (HO$_2$) (Equation 2) or RO$_2$ that generates OH (Equation 3) (Sillman, 2007).

Finally hv removes one oxygen (O) atom from NO$_x$ (Equation 4). The O then collides with an oxygen (O$_2$) molecule creating O$_3$ (Equation 5). The collision of O and O$_2$ requires energy to be removed by an atom or molecule which is referred to as M (Brown et al., 2003). M is a catalyst for O$_3$ creation.

$$\text{VOC} + \text{OH} \rightarrow \text{RO}_2 + \text{H}_2\text{O} \quad \text{(Equation 1)}$$

$$\text{HO}_2 + \text{NO}_x \rightarrow \text{OH} + \text{NO}_2 \quad \text{(Equation 2)}$$

$$\text{RO}_2 + \text{NO}_x \rightarrow \text{VOC} + \text{HO}_2 + \text{NO}_2 \quad \text{(Equation 3)}$$

$$\text{NO}_2 + \text{hv} \rightarrow \text{NO} + \text{O} \quad \text{(Equation 4)}$$

$$\text{O} + \text{O}_2 + \text{M} \rightarrow \text{O}_3 + \text{M} \quad \text{(Equation 5)}$$

In 1990, the Clean Air Act (CAA) amended the National Ambient Air Quality Standard (NAAQS). NAAQS has two main goals for protecting the U.S. population from air pollution. The primary goal of the NAAQS is to protect public health of children, the elderly, and people with respiratory problems. The second goal is to protect the public welfare, including vegetation, animals, and buildings (USEPA, 2007c). Six of the air pollutants monitored by the NAAQS are O$_3$, carbon dioxide (CO$_2$), lead (Pb), NO$_2$, particulate matter, and sulfuric acids (H$_2$SO$_4$). Ozone is the only pollutant that will be addressed in this research.

Prior to 1997 the Environmental Protection Agency (EPA) required an average measurement for ambient O$_3$ not exceed 125 ppb for a 1-hour period. However, studies found that adverse health effects still occur at levels less than 125 ppb and for periods
longer than 1-hour (DEP, 2006). Humans that are overexposed to unhealthy levels of O₃ can experience respiratory problems such as chest pains and coughing. When overexposure to O₃ occurs over a number of days, more problems occur such as lung infections and inflammation (Ellis et al., 2000). Health problems associated with overexposure to O₃ are not limited to the respiratory system. For example, headaches and dizziness are also associated with O₃ (Lennartson and Schwartz, 1999).

In 1997, the EPA changed the requirement for average ambient O₃ measurements to be less than 85 ppb for 8-hours to address the effects of long-term exposure. Two classifications are used to describe an area’s compliance with the new requirement. If a measurement is less than 85 ppb for the 8-hour average, it is declared in attainment. The other classification, non-attainment occurs when an area has an average measurement that is more than 85 ppb for 8-hours. Areas not in compliance can be upgraded to attainment status by having measurements over a “3-year average of the annual 4ᵗʰ-highest daily maximum 8-hour concentrations is less than or equal to 0.08 ppm (85 ppb)” (USEPA, 2006a: 4). In other words, an area’s fourth highest concentration of O₃ over a three-year period must be less than 85 ppb. This requirement is fulfilled by the area in violation having a State Implemented Plan (SIP), developed by the state officials. SIPs discuss how the state will reduce the emissions so that the problem can be eliminated in both the short and long terms (Georgopoulos, 1995).

In 1996, two years before the new 8-hour policy, 39 million Americans lived in a non-attainment area, with the potential for more people to be in non-attainment after the policy change (Cobourn and Hubbard, 1999). Two years later in 1998 the number of American living in non-attainment areas reached 51 million (Lin et al., 2000). Recently,
researchers documented that over 100 million Americans, or one-third of the U.S. population, live in non-attainment areas (Maxwell-Meier and Chang, 2005).

In June of 2007 the EPA proposed an alternative policy that would strengthen the NAAQS in regard to O\textsubscript{3}. The alternative policy would require the average ambient O\textsubscript{3} measurement not exceed a range of 70 to 75 ppb. The exact value for exceedance is still to be determined. This proposed policy is in response to scientific data that shows humans have adverse health effects occurring at levels less than 85 ppb for an 8-hours average. If this policy is approved in March of 2008, it will take effect in 2013 (USEPA, 2007d).

The 1997 change in the NAAQS has placed five metropolitan (Owensboro, Bowling Green, Louisville/Jefferson, Lexington-Fayette, and Huntington-Ashland) and one micropolitan (Paducah) statistical area in Kentucky at risk for non-attainment status (Cobourn and Lin, 2004). All six of these areas, including the Hopkinsville (part of the Clarksville MSA) and Covington (part of the Cincinnati-Middletown MSA) metropolitan area, have an O\textsubscript{3} forecast available online at http://www.louisville.edu/speed/mechanical/ozone/fcst/oz_fcst_KY.html (Cobourn, 2006) (Figure 3).

Louisville, KY is one metropolitan area that responded to its potential non-attainment status by developing the Kentuckiana Ozone Prevention Coalition, which is now called Kentuckiana Air Education (KAIRED). The KAIRED program’s primary objective is to educate the general public on how individual actions can impact local air quality (KAIRED, 2006). This education encourages the public to limit behavior that would add to increased O\textsubscript{3} production. To accomplish this the public is asked to use alternate modes of transportation, such as public transportation systems or car pooling, that are practical
alternatives potentially reducing the number of precursor emissions emitted (Kaire, 2006; Hubbard and Cobourn, 1998).

Kaire is also responsible for issuing warnings in Louisville when days of high O3 are expected to occur (Hubbard and Cobourn, 1998). The threshold for issuing warnings are derived from the Air Quality Index that has been developed by the EPA to explain the health risk associated with different O3 concentrations (USEPA, 2007a) (Figure 5).

<table>
<thead>
<tr>
<th>8-hour Average Ozone Concentration (ppb)</th>
<th>Air Quality Index Values</th>
<th>Air Quality Descriptor</th>
<th>Health Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to 64</td>
<td>0 to 50</td>
<td>Good</td>
<td>No health effects are expected.</td>
</tr>
<tr>
<td>65 to 84</td>
<td>51 to 100</td>
<td>Moderate</td>
<td>Sensitive individuals participating in outdoor activity may experience health effects from long term exposure.</td>
</tr>
<tr>
<td>85 to 104</td>
<td>101 to 150</td>
<td>Unhealthy for Sensitive Groups</td>
<td>Members of the sensitive group may experience respiratory symptoms and reduced lung function.</td>
</tr>
<tr>
<td>105 to 124</td>
<td>151 to 200</td>
<td>Unhealthy</td>
<td>Members of the sensitive group have a higher chance of experiencing respiratory symptoms and reduced lung function. Everyone else should reduce prolonged or heavy exertion outdoor.</td>
</tr>
<tr>
<td>125 to 374</td>
<td>201 to 300</td>
<td>Very Unhealthy</td>
<td>Members of the sensitive group should avoid outdoor exertion. Everyone else should avoid prolonged or heavy exertion outdoor.</td>
</tr>
<tr>
<td>375 and Greater</td>
<td>301 to 500</td>
<td>Hazardous</td>
<td>Everyone should avoid physical activity outdoors.</td>
</tr>
</tbody>
</table>

* Members of the sensitive group include: children who are active outdoors, adults involved in moderate or strenuous outdoor activities, individuals with respiratory disease such as asthma, and individuals with visual susceptibility to ozone.

Figure 5: The Air Quality Index was developed by the EPA to help the public understand the health affects of O3. 8-Hour O3 concentration and Air Quality Index values are matched with a specific color. This index makes it easier for the general public to know what precautions they need to take in order to avoid adverse health effects. Source: ENHS (2007), USEPA (2007a), annotated by author.

Kaire and programs like it are successful because of the ability to forecast high O3 concentrations with a 24-hour warning. This is useful in planning purposes for both the primary and secondary goals of the NAAQS. The warning allows people time to plan activities when they will be at less risk of experiencing O3 related respiratory problems (Kang et al., 2005). Ozone monitoring stations are a valuable asset when establishing a
forecast model. The EPA requires that monitors measure meteorological conditions and 
O₃ chemistry situations (Abraham and Comrie, 2004).

Meteorological conditions modulate the level of O₃ in an area (Milanchus et al., 
1998). In fact, O₃ concentrations are affected by meteorological conditions two to three 
times more than changes in emission levels (Garcia, 2006). Synoptic meteorological 
conditions associated with high O₃ concentrations consist of stagnate summer days with 
long periods of sunlight (Dueñas et al., 2005). Generally, high-pressure systems have 
these characteristics. A high-pressure system has to remain stationary, or stagnate, over 
an area for several days in order to form high concentrations of O₃. The characteristics of 
a stagnate high pressure system are a four day period of sea level pressure greater than 
1014 mb, surface wind speed less than 4 m/s, broken cloud cover, and rain events that 
last less than 2 hours (O'Connor et al., 2005).

The moisture content in the atmosphere is an important variable in predicting O₃ 
concentration. O'Connor et al. (2005) explained that dry atmospheric conditions are 
strongly correlated to high O₃ concentrations. On the other hand, when the atmosphere is 
moist with cloud cover, O₃ is reduced (Milanchus et al., 1998). The reason is water vapor 
in the atmosphere acts as an absorber of radiation (Peavy et al., 1985), which reduces the 
availability of sunlight for a photochemical reaction. Also, if water vapor condenses and 
precipitation follows then the precursor emissions are “washed” out of the atmosphere.

As already explained meteorological conditions that contribute to high O₃ most often 
occur during the summer months (USEPA, 2006a). Cascadia in the Pacific Northwest had 
an outbreak of high O₃ from July 11 to 14, 1996. It should be noted that during this time 
the 1-hour ambient O₃ policy was in place, but of the seven monitors recording O₃, six
recorded a violation of the 8-hour standard (Barna et al., 2001). A case study by Barna et al. (2001) found that synoptic features present included an upper level high-pressure system over the west coast United States and a developing surface thermal low-pressure over California.

Synoptic meteorological patterns are important for determining long-term O\textsubscript{3} concentrations, but diurnal changes are important to study as well. Peaks and dips in O\textsubscript{3} concentrations are experienced in cycles during the day as the angle of the sun changes (Ryan et al., 2000). The increase and decrease in O\textsubscript{3} are classified in four categories linked to the height of the nocturnal boundary layer that develops at night and cools the air at the surface. The four categories during a given day are nighttime (0000-0400), morning (0600-1000), afternoon (1200-1600), and evening (1800-2200) (Rao et al., 2003).

During morning hours the boundary layer (BL) (an inversion layer that forms over the night-time hours due to radiational cooling) begins to break down (Sistla et al., 2001). At this time of day the BL is at its lowest height. Also, winds are generally calm. As sunlight reaches the Earth, photochemical reactions take place producing O\textsubscript{3}. The morning hours are a critical time of day because these hours determine the amount of O\textsubscript{3} that will form during the day (Rao et al., 2003). During the mid-morning hours (1000-1200) there is an increase in O\textsubscript{3} production when the Earth’s surface begins to receive more sunlight than in the morning hours. This brings about an increase in precursor emissions (NO\textsubscript{x} and VOCs) being converted to O\textsubscript{3}. In addition, vertical mixing, due to convection of the sun’s energy, is more pronounced in the mid-morning hours. Vertical mixing allows entrainment, or one wind flow being captured by another, to take O\textsubscript{3} from the residual
layer to the Earth’s surface. The residual layer contains the previous day’s O₃ and/or O₃ that has been advected to the area. (Doran et al., 2006) (Figure 6).

Figure 6: The diagram shows the vertical structure of the boundary layer throughout the day. Source: Atkins (2001), annotated by author

O₃ reaches a peak during the afternoon hours (Figure 7). Also, the BL, which has been mixed for approximately five hours, reaches its highest point of the day. In an urban area the peak in O₃ is usually found in a 50 to 100 km radius around a city center (Sillman, 1999). Solar radiation and the BL height decreases in the late afternoon and evening hours. Photochemical reactions slow down and O₃ begins to be removed from the atmosphere (Sistla et al., 2001). A reduction in precursor emissions and a lack of solar radiation during these hours interrupts the photochemical reaction process that created O₃ earlier in the day.
Diurnal patterns in $O_3$ vary between rural monitors at sea-level and those well above sea-level (Sistla et al., 2001). Baumgardner and Edgerton (1998) studied differences in $O_3$ between elevations. The difference between $O_3$ concentrations on mountaintops and low elevations has to do with the temperature inversion that develops at night or the BL. Lower elevations influenced by the BL are "blanketed" from the air above it more easily destroying $O_3$ (Baumgardner and Edgerton, 1998). When the BL is well mixed, low elevation $O_3$ concentration is less than high elevation $O_3$ (Aneja et al., 2000).

Sites at high elevations are above the inversion layer. Ozone concentrations at this elevation are influenced by radiational cooling that causes mixing with the troposphere. Ozone is then transported from the troposphere by vertical mixing. For example, a site with an elevation greater than 900 m can experience long exposure to $O_3$ (Baumgardner and Edgerton, 1998).

The Greater Smokey Mountains (GSM) and MCNP are rural locations affected by $O_3$. Both national parks have a monitor to record $O_3$, but the elevations of the two monitors are different. The MCNP $O_3$ monitor is at 230 m above sea level and the GSM monitor is...
1243 m above sea level (Figure 8). Tong et al. (2006) used O₃ data collected at GSM and MCNP during 1990 to 2002 to study the spatiotemporal (related to time and space) differences between MCNP and GSM (Figure 8). It was found that MCNP did not experience exceedance after sunset because of the BL reducing O₃ concentrations. The BL prevented air masses with high O₃ concentrations from reaching the Earth’s surface. MCNP most often had exceedance during midday and sunset hours (Figure 9).

Figure 8: Tong et al. (2006) studied the spatiotemporal differences between O₃ concentrations at different elevations for GSM and MCNP. Source: AIRS (2007), NAUS (2006), annotated by author

GSM had O₃ exceedance during all hours of the day “regardless of the availability of sunlight” (Tong et al., 2006: 180). As opposed to MCNP, GSM had the most exceedance during the hours between sunset and late evening. GSM is above the BL making GSM more susceptible to high concentrations of O₃.
Figure 9: This figure shows the diurnal pattern of 8-hour O\textsubscript{3} exceedance for MCNP and GSM during 1996-1998. A count of one was used every time an hour of the day reached exceedance. Those counts were summed and then graphed by hour. Source: Tong et al. (2006), annotated by author

For urban areas the decay of O\textsubscript{3} is aided by an increase of NO\textsubscript{x} from vehicle emissions (Baumgardner and Edgerton, 1998) during the evening hours. Generally this decay is a result of oxidation involving one oxide of nitrogen (NO) and an O\textsubscript{3} particle (Spellman, 1999) (Equation 6). Nitrous dioxide (NO\textsubscript{2}) and O\textsubscript{2} are the outcomes of oxidation. This process is known as titration.

\[
\text{NO} + \text{O}_3 \rightarrow \text{NO}_2 + \text{O}_2
\]  

(Equation 6)

In rural areas the titration process usually goes unnoticed because there are less sources of NO\textsubscript{x} emissions. The result for rural areas are pattern that remain consistent throughout the day (BIOSTRESS, 2007).

Maxwell-Meier and Chang (2005) suggested that a different approach needs to be taken when managing air pollution for urban and rural areas. Time series data of O\textsubscript{3} and meteorological conditions need to be analyzed. For this to be accomplished a method
needs to be developed where short-term or high frequency changes were removed from the data (Milanchus, et al., 1998). Separating the short-term and long-term changes provides better conclusions that could be drawn from the resulting data (Wise and Comrie, 2005). Policy makers can design better O₃ reduction procedures when there is an understanding of long-term climate effects (Diem and Comrie, 2001).

To investigate the trends in time-series data from urban and rural areas Maxwell-Meier and Chang (2005) used the Kolmogorov-Zuebenko filter. The filter examines data for long-term yearly differences including intra-day, diurnal, synoptic, seasonal, and long term changes over time between urban and rural areas. Time scales of the filter data are intra-day less than 10 hours, diurnal lasting 24 hours, synoptic lasting between 2 and 21 days, seasonal lasting over a year, and long-term is the change over several years (Rao et al., 2003).

The filter shows that seasonal and long-term changes are more dominant in rural areas. For example, the short-term changes in Georgia for high O₃ concentrations are greater in urban areas when compared to rural areas. Additionally, urban areas commonly experience a peak in high O₃ during July, while rural regions experience a higher peak in June (Dueñas et al., 2005). Urban and rural areas should then be treated separately in policy decisions (Maxwell-Meier and Chang, 2005).

More often than not there are a limited number of O₃ monitors serving an area. Cities and regions have used different techniques in developing O₃ models to compensate the sparse number of O₃ monitors. There are two ways of doing this the first is using a spatial interpolation method of inverse distance weighting (IDW), spline, or kriging. The second way is using a trend surface analysis or linear regression.
Spatial interpolation is the “prediction of exact values of attributes at un-sampled locations from measurements made at [sample] points with in the same area” (O’Sullivan and Unwin, 2003: 220). Sample points for the model are O₃ monitors. To begin the interpolation process a grid of points are laid over the study area. Values are then interpolated for each point in the grid and contour lines are drawn to represent the surface of the study area (O’Sullivan and Unwin, 2003).

There are uncertainties in making estimations for the area surrounding the O₃ monitor. The uncertainties are local spatial variability, which include terrain changes and other data representing physical structures (Abraham and Comrie, 2004), are not factored into spatial interpolation so the area is not represented correctly in the model. The use of IDW (Figure 10a), which is one of the simplest techniques in spatial interpolation, skews a forecast outside of the sample area because local spatial variability is not taken into account. In IDW more credibility is given to sample interpolation points near each other, such as a nearest neighbor approach, and not to points with a greater distance from one another. Other methods of spatial interpolation are spline and kriging.

The spline method is a piecewise polynomial interpolation (Figure 10b). This is accomplished by the surface passing thought the sample points and the surface has a minimum curvature (ESRI, 2005). There are two different variations of spline called regularized and tension. Regularized spline minimizes the surface curvature by integrating the first, second, and third derivative. Tension spline uses more point data and only the first and second derivative. Tension spline has a smoother surface (Childs, 2004).
While both IDW and spline are useful spatial interpolation methods, Childs (2004) explained that kriging is the most powerful interpolation method to use in pollution modeling. Kriging is similar to IDW in that kriging takes a nearest neighbor approach, however kriging uses a search radius around control points that are weighted to produce a value at un-sampled locations (Figure 10c). The weighting is based on the distance between sample points, un-sampled locations, and the distribution of sample points (ESRI, 2007). When kriging and IDW are compared the error in estimating values for un-sampled locations is lower in kriging than in IDW (Liebhold et al., 1993). For this reason kriging is regard as one of the best interpolation methods for geo-statistics (Moore and Carpenter, 1999).

Figure 10: Spatial interpolation methods of Inverse Distance Weighting (a), Spline (b), and Kriging (c) give different results. This figure shows the result of the three different methods when elevation is interpolated from a grid of sample points. Source: Childs (2004), annotated by author

An accurate high-resolution model can be created using a linear regression-based equation. Simple linear regression is used to find if two datasets are linearly correlated. For O₃ estimations simple linear regression treats metrological conditions as independent variables so that the dependent variable of O₃ can be estimated. The independent and dependent variables are then fitted to a line to find the relationship between the two (O’Sullivan and Unwin, 2003). When variables show a statistically high relationship with one another, the variables are used in a regression model. In has been found that a strong
correlation exists between meteorological conditions and concentrations of O₃ (Spellman, 1999).

Both kriging and linear regression have been applied to the O₃ problem in Tucson, AZ. This was done to find the best way of mapping O₃ in a city that has several monitors in a concentrated area. Using linear regression over kriging is the preferred method of O₃ mapping due to linear regressions ability to provide a measurement of error (PDEQ, 2007) and take into account factors of spatiotemporal variation (Abraham and Comrie, 2004). In this application spatiotemporal variations are the “prevailing wind patterns, topography, and the spatial distribution of sources of air pollution across the metropolitan area (e.g. roadways)” for the city of Tucson (PDEQ, 2006: 2). The Pima County Department of Environmental Quality published the results of this study online, which can be viewed at http://www.airinfonow.org/html/Ozone_Compare.html (PDEQ, 2006).

There are many different types of linear regression than just simple linear regression have been applied to O₃ mapping.

Nonlinear regression model has been used to analyze O₃ in Louisville. The model previously was a multiple regression model that was refined to a nonlinear equation in 1998. Multi-linear regression uses multi-temporal data from the available stations taking observations (Diem and Comrie, 2002). Nonlinear regression gives better accuracy when compared to linear regression and neural networks (Cobourn and Lin, 2004). This is due to “real-world systems” being nonlinear in nature (Spellman, 1999), such as the curvature of the Earth’s surface.

Some O₃ models have been developed using neural networks. A neural network is an algorithm that is based on processing elements (Spellman, 1999). All of the processing
elements are designed to operate in the same way or with the same topology (Narasimhan et al., 2000). This allows the network to react to bias from the input data and adjust to compensate for the bias (Spellman, 1999). The advantage of the network is that there are no large data gaps or spatiotemporal errors (Narasimhan et al., 2000). A neural network is proficient in nonlinear effects between variables that are affected by mathematical linkages in the network (Cobourn and Lin, 2004). The input is retrieved from either single or numerous sources that ultimately give a single output that is in line with a non-linear function (Cobourn et al., 2000). Since the neural network is nonlinear there is no need to be concerned with correlations between variables (Spellman, 1999).

These three operations of neural networks improve modeling but do not make neural networks overwhelmingly better than multiple regression models as documented by Comrie (1997). Comrie (1997) explained that models generally over-and under-estimate days of low and high O₃ concentrations. The neural networks can reduce the estimation error and can be superior to multi-linear regression models if lagged O₃ data are used. Lagged O₃ data are classified as data that are not dependant on real-time information (Comrie, 1997). Also, increasing the number of processing elements used in neural networks could potentially make this technique more valuable (Soja and Soja, 1999).

Soja and Soja (1999) compared and contrasted different linear and non-linear regressions along with neural networks. A shift from simple linear regression to multi- and non-linear regression took place when temperature was found to highly correlate with O₃ concentration. Both of these sophisticated models (multi- and non-linear) were shown to be superior in the explanation of meteorological patterns that influence O₃. However, limitations in the different models include the inability to accurately predict low and,
more importantly, high $O_3$ days. Multi-linear regression and neural networks generally explain 60 to 80 percent of the correlated variables with an error of 10 to 20 ppb (Cobourn and Hubbard, 1999). The overestimation in the models occurs on days when there are windy conditions and few hours of sunlight. Conversely, the models underestimate $O_3$ concentration on weekdays when there is more traffic as opposed to the weekends (Soja and Soja, 1999).

Model data is inevitably going to have under- and over-estimations. However, Cobourn and Hubbard (1999) suggest that errors can be reduced by using a Hybrid regression model. The model is composed of two separate equations that treat the relationships used in the regression of meteorological conditions as non-linear. The two equations are called the Standard and Hi-Lo model. When used together the model is referred to as the Hybrid model. The Standard model makes predictions of $O_3$ when conditions are sunny, hot, and stagnant (Cobourn et al., 2000). The second model or the Hi-Lo model is used for an improved detection of high $O_3$ concentrations. This model is used when the temperatures are greater than 87°F, wind speeds are less than 6 mph, and less than a quarter of the sky has cloud cover. Another element of the Hybrid model is the use of an air mass trajectory term. This is useful when forecasting for the long range transport of $O_3$. 
Chapter 3

Data and Methodology

3.1 Data

The data used in this research came from the O₃ monitor at MCNP. The monitor is located at the southern edge of the park (Figure 2) in a 100 by 100 foot area. The monitor has four instruments that record meteorological variables and a Model 49 UV Photometric O₃ analyzer, which is used to measure O₃. All instruments operate daily taking observations every hour. For a detailed description of how the UV Photometric O₃ analyzer and the meteorological instruments work see Kendrick (2005).

The National Park Service Air Resource Division (NPSARD) is responsible for the stations maintenance and collecting observations. At the conclusion of each month the NPSARD send all hourly observations for the past month to Air Resource Specialist, Inc. in Fort Collins, CO for validation. This validation process takes 45 days to complete and at the conclusion the data is considered accurate.

3.1.1 Model Development

The first step in this research was to recalibrate the model used by Kendrick (2005). This was done because errors were found in the dataset used by Kendrick (2005) to develop the O₃ model on which his conclusions were reached. The errors were discovered when Kendrick’s (2005) dataset was cross-referenced with a validated dataset provided from NPSARD from March 1st, 1998 through October 30th, 2003 or MCNP 98-03. Since this was the case the MCNP 98-03 dataset was used so that the model developed in this research is accurate.
The meteorological predictor variables used in the recalibrated model consisted of MaxT, DTR, and Pd (Kendrick, 2005) (Equation 7). The variable of MaxT is the forecasted daily maximum temperature for a given day. DTR, or diurnal temperature range, is the forecasted maximum temperature subtracted by the forecasted minimum temperature for a given day. Pd, is a precipitation binary component, represents the probability of precipitation. A Pd of 0 was used on days that the probability of precipitation is less than 50 percent. On days that the probability of precipitation was greater than 50 percent then 1 was used for Pd.

\[ \hat{Y} = (\beta_1 \cdot 1:0 \cdot Pd) + (\beta_2 \cdot \text{MaxT}) + (\beta_3 \cdot \text{DTR}) + \beta_0 \] (Equation 7)

Kendrick (2005) tested the ability of 14 different equations that could be used to predict or forecast O3. For this research Kendrick's (2005) Equation 12 (Equation 7 in this research) was used since all variables for the equation are easily accessible. Equation 7 in this research will be referred to as the Standard model from this point forward.

Improvements were made to the Standard model in this research. As explained by Cobourn and Hubbard (1999) a Hi-Lo model can improve the accuracy in O3 forecast on days of high and low O3 concentrations. The same concept was used in this research to develop a Hi model to improve O3 forecast on days with expected high concentrations of O3. The Hi model was used when a there is a forecast of a MaxT greater than 87°F, a DTR greater than 27, and Pd of 0. These variables were determined from the 95th percentile of MCNP 98-03 dataset and variables used by Cobourn and Hubbard (1999). If the criteria for the Hi model was not met then the Standard model was used. Both the Standard and Hi models were incorporated into this research by using the Hybrid model algorithm as explained in the previous chapter.
Predictor variables for the Hybrid model will come from the Model Output Statistic (MOS). MOS is used by the National Weather Service to produce a detailed forecast for stations across the United States, Puerto Rico, and the Virgin Islands. These stations and forecasts for the stations can be viewed at http://www.nws.noaa.gov/mdl/synop/products/bullform.mex.htm (NOAA, 2006a).

3.2 Methodology

While MOS encompasses approximately more than 1000 stations across the Continental United States (Figure 11) only 50 stations were used as sample points in this research. Of the 50 sample points 13 were from Kentucky with the remaining 38 points within in a 70-mile radius of Kentucky’s state border (Figure 12).

Figure 11: This map shows MOS stations used by the NWS in the Continental United States. Source: NOAA (2006b, 2007), annotated by author
Sample Points For the Kentucky Ground-Level Ozone Model

The model used for this forecast is based on meteorological relationships and ozone at Mammoth Cave National Park (MCNP). The model was developed by Kenny (2005). Kenny, D. 2005. The Relationship Between Meteorological Patterns and Rural Ground Ozone Concentration. Master of Geoscience Thesis. Department of Geography and Geology. Western Kentucky University.

Figure 12: This map shows the 50 MOS stations that were used as sample points for the O₃ model in this research. Source: NOAA (2006b, 2007), annotated by author

The reason for including the 37 sample points neighboring Kentucky was to avoid what is known as edge effect. Edge effect is when an artificial boundary is placed around a study area so that the study area is more manageable. This problem results in points at the edge of the study area only neighboring sites in the center of the study area (O’Sullivan and Unwin, 2003). In other words sites just outside of a study area can provide useful information to the patterns being observed in the study area.

Since the available data points were distributed sporadically, spatial interpolation was used (Lee and Pielke, 1996). As explained in the previous section, spatial interpolation is the “prediction of exact values of attributes at un-sampled locations from measurements made at [sample] points with in the same area (O’Sullivan and Unwin, 2003: 220).” The 50 sample points from MOS were transformed to grids along with the rest of the study area.
Spatial interpolation filled the empty grid cells. Kriging from the Geostatistical Analyst in GIS was used for spatial interpolation.

All of the steps listed above were completed in ArcGIS using ArcObjects. ArcObjects are computer objects designed for application development within the ArcGIS Desktop. The objects have behavior and properties such as buttons and tools. To develop applications an object-oriented program needs to be developed (Burke, 2003). The programming language for this object-oriented approach is Visual Basic Applications.

In this research the application programming consisted of designing a user interface that first prompted the user to choose a text file containing the forecast variables for all 50 MOS stations. Based on those variables the program then selected either the Standard model or Hi model for each station. A field was added a shapefile containing the MOS stations and populated with the Pi for each station. The program then performed kriging of the ozone between points which resulted in a raster file. Then the user manually reclassified the raster to represent the Air Quality Index used by the EPA (Figure 5).

3.2.1 Assessing Model Performance

The models performance was determined by data collected at MCNP from the NPSARD during March 1st, 2004 to October 30th, 2005. This is a cross validation technique, which uses data that is independent of the data that was used in the model development. The models performance was reported in root mean square error (RMSE), mean average error (MAE), mean statistical error (MSE), and d-index as suggested by Willmott (1982). Both RMSE and MAE are the top measurement of a models overall performance by summing the mean of the observed O3 (O3) and the models predicted O3 (P0) (Willmott, 1982). The difference between the RMSE and MAE is that RMSE is
sensitive to outliers while MAE is not sensitive to outliers (Kendrick, 2005). RMSE and MAE are defined as:

\[
\text{RMSE} = \sqrt{\left( \frac{1}{N} \sum_{i=1}^{N} (O_i - P_i) \right)^2} \tag{Equation 7}
\]

\[
\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |O_i - P_i| \tag{Equation 8}
\]

In RMSE and MAE as well as the other two performance test MSE and d-index, N is the number of observation used in test dataset. MSE is defined as:

\[
\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (O_i - P_i)^2 \tag{Equation 9}
\]

Finally, d-index is an indication of the agreement between \(P_i\) and \(O_i\). This agreement is given by a value that ranges between 0 and 1. The closer the d-index value is to 1 then the better the agreement is between \(P_i\) and \(O_i\). The equation for d-index is:

\[
\text{d-index} = 1.0 - \frac{\text{MSE}}{\text{PE}} \tag{Equation 10}
\]

3.2.2 Synoptic Patterns Associated with High O\(_3\) Concentrations

The second part of this research was an explanation of the synoptic meteorology patterns associated with high O\(_3\) days for MCNP. Data from the MCNP 98-03 dataset was ranked from the highest to lowest O\(_3\) concentrations. The 95\(^{th}\) percentile of the ranked dataset was used to examine the synoptic patterns associated with those days. The air masses types for these patterns came from the spatial synoptic classification scheme developed by Sheridan (2002).

3.2.3 Impact of Potential Policy Changes

The final part of this research was focused on the policy change that took place in 1998 when the standards changed for ambient O\(_3\) measurements. Data collected at MCNP
from 1998 to 2005 was used to determine exceedance under the current (previous) 8-hour (1-hour) policy. It has been noted that the change in policy would put many areas specifically rural in non-attainment (Baumgardner and Edgerton, 1998; Cobourn and Hubbard, 1999; Barna et al., 2001; Sistla et al., 2001; Reynolds et al., 2003). The recent proposal by the EPA to raise exceedance standards within a range of 70 to 75 ppb (USEPA, 2007d) was also examined.
Chapter 4

Results and Discussion

4.1 Development of the Standard Model

To build the Standard model a stepwise linear regression analysis was conducted on the MCNP 98-03 dataset. For the regression the dependent variable was average ambient 8-hour \( O_3 \) and the independent variables were meteorological conditions of \( \text{MaxT}, \text{DTR}, \) and \( \text{Pd} \). The stepwise linear regression eliminates the variables that are not significant and adds variables when they show significance. Miles and Shevlin strongly advise researchers to proceed with caution when using stepwise linear regression because the computer is relied on to determine the level of significance that a variable has, “… when the computer has no idea about the theory that may determine which variables are important (2001: 38).” However, the overall significance of dependent and independent variables in Kendrick’s (2005) linear regression is similar to the stepwise linear regression in this research. As it has been noted before, Kendrick’s (2005) model is based on data that has discrepancies with observations taken by NPSARD at MCNP, but other researchers have found similar variables to be significant in predicting \( O_3 \) such as Cobourn and Hubbard (1999). Thus, the results of this stepwise analysis are accepted as an accurate relationship between the dependent and independent variables.

The \( R^2 \) value in Table 1 represents the amount of variance explained by the independent variables in the Standard model. Variance is a proportion that should be viewed as a percent by simply moving the decimal place over two digits to the right of the \( R^2 \) value (Miles and Shevlin, 2001). For this regression the independent variables of \( \text{MaxT}, \text{DTR}, \) and \( \text{Pd} \) explain 55 percent of the dependent variable \( O_3 \). This \( R^2 \) value is
0.07 higher than the $R^2$ that Kendrick (2005) received for the same equation, meaning that this model explains 15 percent more variance than that of Kendrick. Again the difference is because the dataset used by Kendrick (2005), which had discrepancies with the observations taken by NPSARD. The value for $R$ in Table 1 represents the amount of correlation between the dependent variables, in this research $O_3$, and the independent variables, of MaxT, DTR, and Pd (Miles and Shevlin, 2001: 32).

Table 1: Results of correlation testing between $O_3$ and meteorological variables for the Standard model. The model evaluation statistics of the model are also shown.

<table>
<thead>
<tr>
<th>Standard Model Summary</th>
<th>R</th>
<th>$R^2$</th>
<th>RMSE</th>
<th>MAE</th>
<th>d-index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.746</td>
<td>0.556</td>
<td>9.62</td>
<td>7.70</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Predictors: (Constant, MaxT, DTR, Pd)

Other variables such as average wind direction, average wind speed, the noon average of sea level pressure, and days since the last recorded precipitation were added to the Standard model to search for improved variance. However, these additions did not add to the explained variance so they were removed. Also considered is the fact that MOS does not include forecast for noon sea level pressure or days since the last recorded precipitation. In other words these variables could not be used even if they were found to be significant.

Table 2 shows the resulting coefficients that are used in the Standard model. Of the three independent variables DTR has the most importance in explaining $O_3$ followed by MaxT and Pd. An explanation of why DTR is weighted so heavily is because high values in DTR equating to warm days with little moisture or cloud cover in the atmosphere. While MaxT and Pd have less importance in the Standard model than DTR both variables are critical to determining DTR. With a high MaxT and no precipitation it is conceivable
that the DTR for a given day will be high, thus the meteorological conditions for high \( \text{O}_3 \) concentrations are established.

Table 2: Summary of coefficients from the correlation testing in Table 1.

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Constant)</td>
<td>8.485</td>
<td>0.528</td>
</tr>
<tr>
<td>MaxT</td>
<td>0.651</td>
<td>0.317</td>
</tr>
<tr>
<td>DTR</td>
<td>0.765</td>
<td>-0.155</td>
</tr>
<tr>
<td>Pd</td>
<td>-5.161</td>
<td>-7.928</td>
</tr>
</tbody>
</table>

Dependant Variable: \( \text{O}_3 \)

The result of the model's error, or the subtraction of the observed \( \text{O}_3 \) (\( \text{O}_j \)) from the predicted \( \text{O}_3 \) (\( \text{P}_i \)), was averaged over a period of 30 days to create a smooth graph (Figure 13). The values that are over- (under-) predictions are associated with negative (positive) values in Figure 13. This smoothing essentially removed sharp changes or spikes. It should be noted that none of the days in the test dataset of 2004 to 2005 meet the criteria to use the Hi model. As a result, performance testing on the Hi model could not be included in this research. The reason for none of the days reaching the criteria is somewhat explained by the unusually cool period in 2004 when the average temperature was 4 to 5 degrees cooler in south-central Kentucky than during the years that the Hi model were developed (the MCNP 98-03 dataset).
Figure 13: A 30 day moving average of \( O_3 \) subtracted by \( P_1 \) for MCNP during the 2004 and 2005 \( O_3 \) seasons. The negative (positive) values in the graph are days when the model over- (under-) predicted \( O_3 \). Source: NPSARD

Overall the model was generally an over-predictor of \( O_3 \) for both 2004 and 2005. In 2004, the model over-predicted \( O_3 \) the most during early summer and fall. This under-performance is partially attributed to the abnormally wet year that central Kentucky had during 2004. The days that are associated with this under-performance are Moist Tropical (MT) air masses with recorded precipitation. In a topic to be discussed later, the Palmer Drought Severity Index (PSDI) values showed that central Kentucky was in a severe wet spell (Table 3). This weather pattern added to the under-prediction of the Standard model.
Table 3: The monthly PDSI values for the central climate division of Kentucky for the 1998 to 2005 O₃ seasons. Source: ESRL (2007)

<table>
<thead>
<tr>
<th>Year</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>-2.02</td>
<td>1.17</td>
<td>0.87</td>
<td>2.4</td>
<td>2.21</td>
<td>-0.68</td>
<td>-1.47</td>
<td>-1.51</td>
</tr>
<tr>
<td>1999</td>
<td>-1.64</td>
<td>-1.78</td>
<td>-1.94</td>
<td>-1.39</td>
<td>-2.29</td>
<td>-3.07</td>
<td>-3.48</td>
<td>-3.31</td>
</tr>
<tr>
<td>2000</td>
<td>-3.53</td>
<td>-2.95</td>
<td>-2.59</td>
<td>-2.49</td>
<td>-2.11</td>
<td>-1.78</td>
<td>-1.64</td>
<td>-2.31</td>
</tr>
<tr>
<td>2001</td>
<td>-2.31</td>
<td>-2.92</td>
<td>-2.71</td>
<td>0</td>
<td>-0.01</td>
<td>0.22</td>
<td>0.04</td>
<td>0.87</td>
</tr>
<tr>
<td>2002</td>
<td>0.72</td>
<td>0.83</td>
<td>1.28</td>
<td>-0.28</td>
<td>-0.65</td>
<td>-0.92</td>
<td>1.16</td>
<td>2.29</td>
</tr>
<tr>
<td>2003</td>
<td>1.47</td>
<td>1.69</td>
<td>1.96</td>
<td>2.6</td>
<td>2.56</td>
<td>2.95</td>
<td>3.99</td>
<td>3.71</td>
</tr>
<tr>
<td>2004</td>
<td>2.1</td>
<td>2.33</td>
<td>3.39</td>
<td>3.02</td>
<td>3.28</td>
<td>3.36</td>
<td>2.74</td>
<td>3.14</td>
</tr>
<tr>
<td>2005</td>
<td>-0.45</td>
<td>-0.13</td>
<td>-0.4</td>
<td>-0.91</td>
<td>-1.02</td>
<td>1.33</td>
<td>-0.77</td>
<td>-1.44</td>
</tr>
</tbody>
</table>

In 2005 the climate patterns exhibited typical conditions found in summer months. It is interesting that patterns seen in 2004 during later summer and early fall are the opposite for 2005. In other words during late August of 2004 the model has less over-prediction than in 2005 when the over-prediction dips. In late September of 2005 the model has more over-prediction than 2004 when the over-prediction peaks (Figure 13). This test serves as a brief explanation of the over-predictions in the model from 2004 to 2005 however model evaluation statistics are a better judge of the models over all accuracy and performance.

The statistics consisting of RMSE, MAE and d-index were conducted on the 2004 to 2005, to examine the Standard model performance. RMSE was 9.62 ppb, MAE was 7.70 ppb, and d-index was 0.82 (Table 1). Excluding the d-index value these statistics are distinct improvements over Kendrick’s (2005) top two models. Kendrick’s (2005) top model (Equation 13 in Kendrick (2005)) explained 53 percent of variance between independent and dependent variables.

Kendrick (2005) reported the model error statistics in the top model as an RMSE of 11.04, a MAE of 8.70, and 0.84 for d-index. Direct comparisons of the Standard model in this research and its counterpart in Kendrick’s (2005) research can not be done since
Kendrick (2005) only reported the statistical error in the top two models. However, this comparison is not important since this research suggests that a model Kendrick (2005) classified beneath the top two, or the Standard model, is actually better than the two chosen as top performing in that research.

4.2 Development of the Hi Model

The dataset for this model is built on the 95th percentile of the data used in the Standard model. Using the 95th percentile of ranked O₃ data includes those conditions that are associated with high O₃ concentrations. In fact 30 percent of the days in the 95th percentile meet the meteorological criteria for the Hi model, as explained in the previous chapter.

This Hi model uses independent variables of DTR and MaxT. Both of these variables have similar importance to the model with the coefficient values (Table 4). With only two days in the 95th percentile recording precipitation the Pd variable was removed as a variable since it did not significantly contribute to the explanation of variance in O₃.

Table 4: Summary of coefficients from the Hi model correlation testing.

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>B 37.274</td>
<td>Std. Error 11.648</td>
</tr>
<tr>
<td>MaxT</td>
<td>.442</td>
<td>.135 Beta .338</td>
</tr>
<tr>
<td>DTR</td>
<td>.408</td>
<td>.132 Beta .32</td>
</tr>
</tbody>
</table>

Dependant Variable: O₃

The overall variance of both DTR and MaxT describes 27 percent of the variance in the 95th percentile (Table 5). This low R² value suggests other independent variables need to be added to the model for a better explanation. As explained previously the addition of other independent variables currently available, such as wind direction, did not improve
the variance explained. At this time the Hi model is somewhat limited in accurately forecasting O₃, but was used in this research to begin exploring the possibilities and potential of having such a model available to policy makers and the general public.

Table 5: Results of correlation testing between O₃ and meteorological variables for the Hi model.

<table>
<thead>
<tr>
<th>Hi Model Summary</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>.523</td>
<td>.274</td>
</tr>
</tbody>
</table>

Predictors: (Constant, MaxT, DTR)

4.3 Operational use of the Hybrid Model for Kentucky during June 2007

The Hybrid model was used during June 2007 to demonstrate the model’s ability to forecast O₃ across Kentucky. Variables for the model were collected daily (except for June 17th) from MOS for each of the 50 sample points as discussed in the previous chapter. To validate the model, O₃ at MCNP and the Pᵢ for Bowling Green, KY (KBWG) were compared. This was done since the model is based on relationships between O₃ and meteorological conditions at MCNP and KBWG is the nearest MOS stations to MCNP.

The difference between the Oᵢ and Pᵢ were drastically different on everyday from the 1st to the 30th of June (excluding the 17th) (Figure 14). As with the analysis in Section 4.1 many of the days were an under-prediction (indicated by negative values) for Oᵢ. These under-predictions appear to be more dramatic in Figure 14 because of the smaller temporal scale. The testing in this section is different from Section 4.1 because Pᵢ was based on observed data while Pᵢ in this section was based on MOS forecast data. This implies that the model runs in this section are an operational use to forecast O₃ and can not rely on observed data for forecast.
Figure 14: Graphical representation of the difference in $O_i$ at MCNP and $P_i$ for KBWG during the June 2007 model run. June 17th is missing due to incomplete data. Source: NPSARD, NOAA (2006a)

Table 6 lists the MOS forecast for each day, excluding the 17th, for the month of June. This table also has the model type (either Standard or Hi) that the Hybrid selected and the difference between $O_i$ and $P_i$ for each day. These observations were compared to the meteorological observations taken by the NPSARD at MCNP in Table 6.
Table 6: MOS forecast for KBWG and observations made at MCNP for June 2007. Days that the Hybrid model selected the Standard (Hi) model are denoted by S (H). The difference between the Oj at MCNP and the Pi at KBWG are also shown. The 17th of June is missing because of incomplete data on that day. Source: NPSARD

<table>
<thead>
<tr>
<th>June</th>
<th>MOS forecast for KBWG</th>
<th>Observations at MCNP</th>
<th>Model</th>
<th>Difference between Oj and Pi</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MaxT</td>
<td>MaxT</td>
<td>DTR</td>
<td>Pd</td>
</tr>
<tr>
<td>1</td>
<td>86 23 0</td>
<td>87 21 0</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>2</td>
<td>85 21 1</td>
<td>83 17 1</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>3</td>
<td>84 20 1</td>
<td>82 16 0</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>4</td>
<td>83 21 0</td>
<td>78 15 0</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>5</td>
<td>79 19 0</td>
<td>79 17 1</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>6</td>
<td>85 27 0</td>
<td>84 26 0</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>7</td>
<td>94 32 0</td>
<td>89 20 0</td>
<td></td>
<td>H</td>
</tr>
<tr>
<td>8</td>
<td>87 15 0</td>
<td>84 15 1</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>9</td>
<td>85 22 0</td>
<td>80 15 0</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>10</td>
<td>84 27 0</td>
<td>82 23 0</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>11</td>
<td>81 24 0</td>
<td>81 17 0</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>12</td>
<td>86 29 0</td>
<td>83 22 0</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>13</td>
<td>88 28 0</td>
<td>86 23 0</td>
<td></td>
<td>H</td>
</tr>
<tr>
<td>14</td>
<td>90 28 0</td>
<td>86 21 0</td>
<td></td>
<td>H</td>
</tr>
<tr>
<td>15</td>
<td>88 26 0</td>
<td>86 20 0</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>16</td>
<td>88 28 0</td>
<td>86 26 0</td>
<td></td>
<td>H</td>
</tr>
<tr>
<td>17</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>18</td>
<td>91 33 0</td>
<td>83 12 0</td>
<td></td>
<td>H</td>
</tr>
<tr>
<td>19</td>
<td>86 17 1</td>
<td>79 11 0</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>20</td>
<td>83 21 0</td>
<td>83 19 0</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>21</td>
<td>88 32 0</td>
<td>86 28 0</td>
<td></td>
<td>H</td>
</tr>
<tr>
<td>22</td>
<td>85 23 0</td>
<td>82 17 1</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>23</td>
<td>88 23 0</td>
<td>87 20 1</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>24</td>
<td>91 28 0</td>
<td>80 13 1</td>
<td></td>
<td>H</td>
</tr>
<tr>
<td>25</td>
<td>90 25 0</td>
<td>82 14 1</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>26</td>
<td>93 26 0</td>
<td>89 19 1</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>27</td>
<td>92 24 0</td>
<td>84 14 1</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>28</td>
<td>90 21 1</td>
<td>87 17 1</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>29</td>
<td>88 21 1</td>
<td>85 17 1</td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>30</td>
<td>86 22 0</td>
<td>83 12 1</td>
<td></td>
<td>S</td>
</tr>
</tbody>
</table>

Many of the days that had the highest error happen to be days when Hybrid model selected the Hi model based on the MOS forecast. In hindsight the Hybrid model should have selected the Standard model on everyday during June 2007 since the criteria for the Hi model was never met based on observations at MCNP. In other words the issue is with over- or under-estimations made by MOS, which caused the Hybrid model to fail. One of
those under-estimations is found in the last week of June when MOS forecasted there would be less than a 50 percent chance of precipitation and the observed data showed that precipitation was recorded at MCNP.

The most glaring problem with using MOS as a source for sample data can be found on June 24th (Figure 15). This day had the highest difference between \( O_1 \) and \( P_1 \) (Table 6). The MOS forecasted that DTR would reach 28, a MaxT of 91, and a 0 for Pd. These three forecast conditions triggered the Hybrid model to select the Hi model. It can be seen in Table 6 that the observed variables were below the criteria of the Hi model and instead the Standard model should have been selected. For June 24th using the observed variables at MCNP and the Standard model the \( P_1 \) was improved to 48 ppb, which is a 41 ppb improvement over the forecasted \( P_1 \) for that day. On average the \( P_1 \) for days when the Hi model was selected improved by 16 ppb when the Standard model and observed variables were used.
Even with days of extreme error MOS provided reliable data on a handful of days including on June 11th. On this day the difference between $O_3$ and $P_1$ was perfect with the zero. From this data the individual point forecast at same locations should be considered accurate since the model is based off of relationships at MCNP and the difference between MCNP and KBWG is negligible (Figure 16).
The forecast in Figure 16 is not without flaws however, a standard error map was generated to show the amount of uncertainty in the interpolated values (Figure 17). These values show there is a +/- 4 to 5 ppb of uncertainty in O₃ across Kentucky in the O₃ values between sample point locations. This day is better than on June 24th which had an uncertainty of +/- 1 to 12 ppb of O₃ (Figure 18). The difficulty in examining these standard errors are that no true thresholds can be set that separates the reliable forecast from the unreliable since the forecast depends on many dynamic factors that change with each model run (Fraczek and Bynerowicz, 2007). While these standard error maps are useful in determining error in the interpolated forecast they do not explain the Hybrid models failure in accurately predict O₃ at each sample point or in this case at KBWG.
The model used for this forecast is based on meteorological relationship and ozone at Mammoth Cave National Park (MCP). The model was developed by Kendrick (2005).


Figure 17: Standard Error Map of the Kentucky O3 forecast for June 11, 2007.

Figure 18: Standard Error Map of the Kentucky O3 forecast for June 24, 2007.
The challenge highlighted on June 24\textsuperscript{th} and other days of extreme over-estimation, shown in Table 6 and again in Figure 14, is that this model is a prediction of $O_3$ based on a prediction of meteorological variables. Currently MOS offers the only format that all required model variables can be obtained. This challenge of using MOS is one that must be accepted since the Hybrid model is for operational use in rural areas without $O_3$ monitors or forecasts.

4.4 Synoptic Analysis of High $O_3$ Concentration

On average the 95\textsuperscript{th} percentile is characterized by dry summer days with high DTR and MaxT values. Many of these days are dominated by Dry Moderate (DM) and Moist Tropical (MT) air masses (Table 7). Sheridan (2002) defined both DM and MT air masses as the following. DM air masses are originally Pacific air masses (Maritime Polar) that move east passing over the Rocky Mountains and becoming dry and warm through adiabatic processes (Sheridan, 2002). The MT air mass is typically warm and humid and is often associated with partly cloudy skies and convective thunderstorms partially during the summer months. The source region for this air mass for Kentucky is the Gulf of Mexico. A MT air mass is in the warm sector of low pressure or west of a surface level high pressure system (Sheridan, 2002).

The temporal pattern of the 95\textsuperscript{th} percentile shows a minor peak in $O_3$ during the early summer followed by an anomalously low period in mid summer. After the mid-summer low period $O_3$ drastically increases in the late summer and early fall months (Figure 19). The range of $O_3$ from August to September is from 80 to 105 ppb and includes 28 days of exceedance. Climatologically the upper level jet begins to shift southward at this time. It
is more common to see dry continental air masses making their way south from Canada being advected over tropical air from the summer to fall transition.

Figure 19: Distribution of $O_3$ in the 95th percentile of MCNP 98-03 dataset. Source: NPSARD

The days in Figure 19 were stratified by air mass to search for which synoptic patterns in the 95th percentile are associated with the high $O_3$ concentrations. As shown in Table 7 the most frequent air masses are tied between MT and DM air masses. Dry Tropical (DT) is the third most frequent air mass associated with high $O_3$ (Table 7). The DT air mass is commonly the hottest and arid conditions advected from Texas and Oklahoma. The data in Table 7 is graphed by air mass for DM, DT, and MT as in Figure 19 to examine the temporal distribution of high $O_3$ concentrations.
Table 7: The 95th percentile of O₃ stratified by air mass. Time period for the 95th percentile is 1998 to 2003. Source: NPSARD

<table>
<thead>
<tr>
<th>Air Mass Type</th>
<th>Occurrence</th>
<th>Average O₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moist Tropical</td>
<td>26 days</td>
<td>85</td>
</tr>
<tr>
<td>Dry Moderate</td>
<td>26 days</td>
<td>87</td>
</tr>
<tr>
<td>Dry Tropical</td>
<td>22 days</td>
<td>89</td>
</tr>
<tr>
<td>Transition</td>
<td>4 days</td>
<td>94</td>
</tr>
<tr>
<td>Moist Moderate</td>
<td>1 day</td>
<td>83</td>
</tr>
</tbody>
</table>

From Table 7 the third highest O₃ concentrations are associated with DT air masses. Many of the DT days are in the month of September (Figure 20). During 1999 a period of 5 days in a row starting on September 1st had observed O₃ above 80 ppb. As defined by Sheridan’s (2002) the DT air mass is associated with hot and arid conditions. These days had DTR’s that were greater than 30 with no recorded precipitation. In other words, the conditions were ideal for high O₃ concentrations.
One of the most frequent air mass types in Table 7 is DM. The DM air mass has a temporal pattern similar to that of DT (Figure 21). Of the 26 DM days in Figure 19, 18 of them were during 1999 with almost all of them in the month of August. While DM is one of the most frequent air masses the O₃ concentration associated DM is lower in comparison to the DT air masses.
Finally the MT air mass has some of the lowest O₃ concentrations when compared to DT (Figure 20) and DM (Figure 21). These days are more sporadic between 1998 and 2003 with no sharp peak in O₃ as in DM and DT. While a majority of the days in Table 7 are MT this air mass is not the most efficient at producing high O₃ days (Figure 22). With MT the likelihood of high O₃ is significantly reduced by cloud cover (Ryan et al., 2000), which reduces the amount of solar radiation and precipitation that washes out precursor emissions (Peavy et al., 1985). As a result the O₃ concentrations associated with MT are not as high when compared to the DT and DM concentrations in Figures 20 and 21.
Figure 22: Temporal distribution of Moist Tropical days that were within the 95th percentile. Source: NPSARD

As previously discussed many of the days in Table 7 with DT and DM occurred in 1999. During the late summer months of that year the southeastern portion of the United States was well into a severe drought (Figure 23). This drought could have enhanced the conditions necessary for high O₃ concentrations in DT and especially in the DM air mass.
As described by Palmer (1965) a drought is an extended amount of time without moisture. The Palmer Drought Severity Index (PDSI) is a water balance model that is used to examine precipitation and temperature over time (Heim, 2002). The values for PDSI are a classification that changes during dry or wet weather conditions (Hayes, 2007) (Table 8).
Table 8: PDSI values and the corresponding classifications that define the index values. Source: Palmer (1965), annotated by author

<table>
<thead>
<tr>
<th>Index Value</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥4.00</td>
<td>Extremely Wet</td>
</tr>
<tr>
<td>3.00 to 3.99</td>
<td>Very Wet</td>
</tr>
<tr>
<td>2.00 to 2.99</td>
<td>Moderately Wet</td>
</tr>
<tr>
<td>1.00 to 1.99</td>
<td>Slightly Wet</td>
</tr>
<tr>
<td>0.50 to 0.99</td>
<td>Incipient Wet Spell</td>
</tr>
<tr>
<td>0.49 to 0.49</td>
<td>Near Normal</td>
</tr>
<tr>
<td>-0.50 to -0.99</td>
<td>Incipient Drought</td>
</tr>
<tr>
<td>-1.00 to -1.99</td>
<td>Mild Drought</td>
</tr>
<tr>
<td>-2.00 to -2.99</td>
<td>Moderate Drought</td>
</tr>
<tr>
<td>-3.00 to -3.99</td>
<td>Severe Drought</td>
</tr>
<tr>
<td>≤-4.00</td>
<td>Extreme Drought</td>
</tr>
</tbody>
</table>

The central climate division for Kentucky, where MCNP is located, experienced abnormally dry conditions as a result of the drought during 1999 (Figure 23). The PDSI values in the central climate division are classified as severe to extreme beginning in the mid summer and continuing into that fall (Table 9). It is difficult to distinguish the level of impact the drought had on the synoptic patterns, but it appears that the two are related since many high O₃ days occurred during 1999.

Table 9: PDSI values for Kentucky’s Central Climate Division each month during 1998 to 2003. Source: ESLR (2007)

<table>
<thead>
<tr>
<th></th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sept</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>-1.63</td>
<td>-1.82</td>
<td>-2.02</td>
<td>1.17</td>
<td>0.87</td>
<td>2.4</td>
<td>2.21</td>
<td>-0.68</td>
<td>-1.47</td>
<td>-1.51</td>
<td>-2.21</td>
<td>-2.18</td>
</tr>
<tr>
<td>1999</td>
<td>-1.25</td>
<td>-1.47</td>
<td>-1.64</td>
<td>-1.78</td>
<td>-1.94</td>
<td>-1.39</td>
<td>-2.29</td>
<td>-3.07</td>
<td>-3.48</td>
<td>-3.31</td>
<td>-3.96</td>
<td>-4.12</td>
</tr>
<tr>
<td>2000</td>
<td>-3.84</td>
<td>-3.41</td>
<td>-3.53</td>
<td>-2.95</td>
<td>-2.59</td>
<td>-2.49</td>
<td>-2.11</td>
<td>-1.78</td>
<td>-1.64</td>
<td>-2.31</td>
<td>-2.26</td>
<td>-1.98</td>
</tr>
<tr>
<td>2001</td>
<td>-2.35</td>
<td>-1.97</td>
<td>-2.31</td>
<td>-2.92</td>
<td>-2.71</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.22</td>
<td>0.04</td>
<td>0.87</td>
<td>1.24</td>
<td>1.22</td>
</tr>
<tr>
<td>2002</td>
<td>-0.14</td>
<td>-0.79</td>
<td>0.72</td>
<td>0.83</td>
<td>1.28</td>
<td>-0.28</td>
<td>-0.65</td>
<td>-0.92</td>
<td>1.16</td>
<td>2.29</td>
<td>2.39</td>
<td>2.67</td>
</tr>
<tr>
<td>2003</td>
<td>1.83</td>
<td>2.54</td>
<td>1.47</td>
<td>1.69</td>
<td>1.96</td>
<td>2.6</td>
<td>2.56</td>
<td>2.95</td>
<td>3.99</td>
<td>3.71</td>
<td>3.99</td>
<td>3.67</td>
</tr>
</tbody>
</table>

4.5 Environmental Protection Agency’s Exceedance Policies

Prior to 1998 the EPA required that average ambient O₃ not exceed a level of 125 ppb for 1-hour. Scientific data revealed that effects to human health occurred at levels that were lower than 125 ppb for 1-hour, which violated the NAAQS primary goal of protecting public health (USEPA, 2006a). Beginning in 1998 the EPA changed its
exceedance policy requiring that levels of the average ambient \(O_3\) not be greater than 85 ppb for 8-hours. This change in policy increased the number of exceedance days in the data collected at MCNP from 1998 to 2005. Also examined will be the impact of a proposed exceedance policy by the EPA. The proposed policy would remain an 8-hour average but within a range of 70 to 75 ppb instead of the current 85 ppb. The proposed policy is to ensure better protection for members of the “sensitive group” which include children or adults that are involved in outdoor activity, individuals with respiratory disease, and individuals that are susceptible to \(O_3\) (Figure 5). If approved in March of 2008 in the policy would not go into effect until 2013.

4.5.1 Changes in the 1-hour and 8-hour Exceedance Policy

During the six year study period at MCNP only one day exceeded the 1-hour policy for \(O_3\). It is difficult to understand the impact of the 1-hour policy had since just one day was in exceedance. To further investigate the 1-hour policy a trend analysis was used. To find trends in the 1-hour \(O_3\) data, a 45-day moving average was applied on the summer months of 1998 to 2005 (Figure 24). This moving average produces a smooth trend showing the pattern of nearly 2000 days of observation.

During the study period there were two noticeable peaks in \(O_3\). The first peak occurred during mid August of 1998 and the second during early September of 1999. As explained in the previous section, this time period was characterized by periods of little precipitation with DM and DT air masses allowing precursor emissions to become concentrated in the atmosphere. After 1999 the trend in \(O_3\) immediately decreased and remained below an average of 75 ppb for the rest of the dataset. At no time does the trend come close to the 1-hour exceedance limit.
Figure 24: A moving average of 45 days was applied to 1-hour $O_3$ observations during the summer months from 1998 to 2005. The exceedance limit of 125 ppb is indicated by the red line. Source: NPSARD

Using the 8-hour policy the number of exceedance days increased by 43 (Table 10). Nearly 74 percent of the 8-hour exceedance days occurred late in the summer months of 1998 and 1999. The same application of the 1-hour trend was applied to the 8-hour trend (Figure 25). The peaks and dips in the 8-hour trend graph are consistent with that of the 1-hour, but the trend is closer to exceedance in the 8-hour trend.
Table 10: Number of days that exceeded the 1- and 8-hour average ambient O₃ policies. Source: NPSARD

<table>
<thead>
<tr>
<th>Average Ambient O₃</th>
<th>Exceedance Days*</th>
</tr>
</thead>
<tbody>
<tr>
<td>125 ppb - 1 hour</td>
<td>1</td>
</tr>
<tr>
<td>85 ppb - 8 hours</td>
<td>44</td>
</tr>
</tbody>
</table>

*Note: Exceedance days are from 1998 to 2005 O₃ seasons at MCNP

Figure 25: A moving average of 45 days was applied to 8-hour O₃ observations during the summer months from 1998 to 2005. The exceedance limit of 85 ppb is indicated by the red line. Source: NPSARD

Meteorological conditions, such as maximum temperature and DTR, were the same for 1- and 8-hour datasets. However, the reason for the dramatic difference in the two policies is the level of aggregation applied to the data. By averaging the ambient O₃ measurements from 1- to 8-hours the values become more homogenous. Thus, 8-hour data leads to more exceedance because the values are similar. For example, on July 26, 1999, the only 1-hour exceedance day, the maximum 1-hour measurement was 127 ppb while the maximum 8-hour measurement was 105 ppb.
4.5.2 Proposed 8-hour Exceedance Policy

To examine the impacts of the proposed exceedance policy would have on rural areas, data collected from the 1998 to 2005 O\textsubscript{3} seasons at MCNP were used in the analysis. Since the EPA proposes the new 8-hour average ambient O\textsubscript{3} exceedance criteria to be within a range of 70 to 75 ppb, both minimum and maximum of the range were examined.

The difference between the current 85 ppb limit and the 75 ppb limit is 107 exceedance days (Table 11). The number of exceedance days increases even more for the 70 ppb limit when a total of 258 days would be considered in exceedance. If the proposed policy is approved in 2008, then the number of exceedance days for 2013 could increase by five times the amount for the 85 ppb limit at MCNP.

Table 11: Number of days that would exceed the proposed 8-hour average ambient O\textsubscript{3} policy. Source: NPSARD

<table>
<thead>
<tr>
<th>Average Ambient O\textsubscript{3}</th>
<th>Exceedance Days*</th>
</tr>
</thead>
<tbody>
<tr>
<td>75 ppb - 8 hours</td>
<td>151</td>
</tr>
<tr>
<td>70 ppb - 8 hours</td>
<td>258</td>
</tr>
</tbody>
</table>

*Note: Exceedance days are from 1998 to 2005 O\textsubscript{3} seasons at MCNP

The trend in O\textsubscript{3} for the proposed policy breaks the limit of exceedance in late summer of 1998 and 1999 (Figure 26). In contrast the trend in current 8-hour policy was near 85 ppb, but it never went over that limit. As it can be seen from Figure 26 the potential to surpass exceedance is higher for the proposed policy. However, this does not represent the true potential since this is a range and not a set number for exceedance.
Figure 26: A moving average of 45 days was applied to 8-hour O₃ observations during the summer months from 1998 to 2005. The exceedance limit is within a range of 70 to 75 ppb indicated by the red line. Source: NPSARD

The difference between the 1-hour and current 8-hour policy was aggregation of data. However, the difference between the current 8-hour policy and the proposed policy is aggregation but lowering the exceedance limit. The increase in exceedance days of the proposed policy is not the only problem that rural areas, such as MCNP, or even urban areas would face. From projections recently released by the EPA, the proposed policy would significantly affect non-attainment status of counties across the state.

In Kentucky, eight counties would face non-attainment status for violating the attainment policy of a “3-year average of the annual 4ᵗʰ-highest daily maximum 8-hour concentration...” (USEPA, 2006a: 4) remaining less than 75 ppb instead of the current 85 ppb. The projection of this non-attainment status is based on 2003 to 2005 data collected
by AIRS and CASNET programs (Figure 27). The same datasets used to project non-attainment for a 75 ppb was used for 70 ppb.

In the 70 ppb analysis, 24 counties in Kentucky would be classified in non-attainment. That is, 82 percent of the counties with at least one O₃ monitor would be in violation. In all, there are 29 counties in Kentucky with at least one O₃ monitor.

![Counties with Monitors Violating Alternate 8-hour Ozone Standards (70 and 75 parts per billion)](image)

Figure 27: Counties that would obtain non-attainment status for the proposed 8-hour average ambient O₃ policy. Source: USEPA (2007b)
Chapter 5

Conclusions and Future Research

The main goal of this research was to design a model that could be used to forecast ground-level O$_3$ for the rural areas across the state of Kentucky. A Hybrid model approach similar to Cobourn and Hubbard (1999) was used to design the most robust O$_3$ forecast possible for rural areas. This model selects either a Standard or Hi model based on the MOS forecast for a given day. If the MOS forecasted a MaxT was greater than 87°F, a DTR greater than 27, and a Pd of 0 then the Hi model would be used. If all three of these criteria were not met then the Standard model would be chosen.

The Standard component of the Hybrid model explained more than half of the variance associated with O$_3$ and meteorological variables (MaxT, DTR, and Pd). The difference between observed O$_3$ and predicted O$_3$ during 2004 to 2005 showed that the Standard model on average is an over-prediction of O$_3$. Most of the over-predicting (under-predicting) was in the late (early) months of the O$_3$ seasons. The results of model evaluation statistics proved not only that this model is an accurate predictor of O$_3$ but is also an improvement over the same model that originated out of Kendrick’s (2005) research. In fact, this model is a better performer over the model that Kendrick (2005) considered the best. As already explained the dataset that Kendrick (2005) used for 1998 to 2003 did match the same dataset that was retrieved from the NPSARD. The Hi model that was developed in this research did not explain more than 27 percent of the variance between the O$_3$ and meteorological variables of MaxT and DTR. A different approach needs to be explored with the Hi model so it can be used with confidence.
Future research using the Hybrid model should include using the Kentucky Mesonet (KYMN) as a source for input variables. Data from the KYMN will be available from over 100 stations in nearly every county across the Commonwealth in the next few years. The KYMN will provide data from each station every 15 minutes 24 hours a day (KYMN, 2007). This would be an improvement over the current MOS forecast that only provides a static daily forecast. By using real-time data for input it is conceivable the Hybrid model could be a dynamic forecast model updating every 15 minutes. Data provided from the KYMN will be air temperature, relative humidity, wind speed/direction, soil moisture/temperature, solar radiation, and precipitation. This real-time data eliminates having a prediction of $O_3$ based on a prediction of meteorological variables. This will take some time to implement however.

KYMN stations would need to be in operation for at least 5 to 10 years so that a climatological record could be established at the stations. Once this is done an assumption would have to be made that the same relationships between $O_3$ and meteorological variables at MCNP would apply to all KYMN stations. This assumption is used in the current Hybrid model that the same relationships at MCNP exist at each MOS sample point.

Also in future research the top model in Kendrick (2005) should be used as the Standard model. It is obvious that the model would need to be recalibrated, but once this is done it is possible that a more accurate Standard model and even Hi model could be used in place of the ones presented in this research. The reason for that is the addition of a solar radiation coefficient to the model. As explained by Kendrick (2005) solar radiation is important component of $O_3$ concentrations. The importance of solar radiation
is that in the summer months days have long hours of radiation and the zenith angle of the sun is low (Ryan et al., 2000).

Finally the data from KYMN could also allow researchers to add an elevation component to the model. As examined by Tong et al. (2006) there is a significant change in O₃ concentration at different elevations. This is an important factor that needs to be further examined across the state of Kentucky. Across Kentucky there is roughly 1180 m of elevation difference between its western and eastern regions. This would greatly benefit the rural areas in the eastern portion of the state that are at a much higher elevation then the west. However, this component would not improve the current Hybrid model since elevation is a non factor with the model being built just from the data at MCNP.

In this research high concentrations of O₃ were days that the maximum 8-hour ambient average was greater than or equal to 80 ppb for 8-hours. It was also found that high O₃ concentrations are associated with DM air masses that advect over Kentucky during late summer and early fall. These air masses are likely the result of dry continental air masses migrating from Canada that are modified by moist summer conditions. However, DT air masses, which represent air that originated over the desert Southwest, have some of highest concentrations of O₃ associated with them. There were 48 days that were either DM or DT. A longer study period of 10 years or greater should be examined for MCNP. This length of time would capture both climatic anomalies and normal conditions for MCNP providing a better climatic record to examine high O₃ concentrations.
The final topic examined was the previous 1-hour, current 8-hour, and proposed 8-hour O₃ policies. Rural areas have and potentially will be the most affected areas from policy changes. As shown in this research the change from the 1-hour 125 ppb policy to the current 8-hour 85 ppb policy drastically increased the number of exceedance days from 1 to 44 at MCNP. Also, with the proposed 8-hour 70 – 75 ppb policy the number of exceedance days will increase an exorbitant amount from the current policy for rural areas.

The southern central and eastern portions of the Commonwealth are areas that have the most to lose by the changes in exceedance policy. However, at this time only a small sample of O₃ data is available for rural communities (Figure 1). With no data being retrieved from some of the most rural areas it is impossible to figure out the type of impact, either negative or positive, that the proposed policy may have or for that matter the impact the current policy is having.

In this research it has been shown that high O₃ concentrations are not bound to just urban areas. There needs to be more support from government organizations such as the National Forest Service and the EPA on monitoring O₃ in rural areas. More academic studies are need on O₃ in rural areas to add to the limited research on this topic.
References


65


Smith, K. 2006. Rural Air Pollution: A Major But Often Ignored Development Concern. Presented at the Commission on Sustainable Development Thematic Session on Integrated Approaches to Addressing Air Pollution and Atmospheric Problems on May 4, New York, NY


