5.5A OZONE FORECASTING TOOL DEVELOPMENT TO SUPPORT FORECASTING FOR THE EPA AIRNow PROGRAM

Dianne S. Miller*, Clinton P. MacDonald, Timothy S. Dye
Sonoma Technology, Inc., Petaluma, CA

1. INTRODUCTION

As understanding of the effects of poor air quality on human health expands and as the public becomes more aware of these effects, providing forecasts of unhealthy air quality to the public becomes increasingly important. Air quality forecasts are generally used by the public to avoid unnecessary exposure to pollutants and by state and local agencies to determine when to declare "Ozone Action" or "Spare the Air" days when the public is asked to reduce emissions (e.g., carpooling, telecommuting, and refueling after dark). Timely forecasts of unhealthy air quality are critical to allow the public ample time to reduce exposure and enable them to plan ways to reduce emissions. These forecasts are currently collected across the United States by the U.S. Environmental Protection Agency's (EPA) AIRNow program.

AIRNow is a public outreach program that provides real-time and forecasted air quality information to the public via the Internet (http://www.epa.gov/airnow) and media outlets (television and newspapers). State and local air quality agencies collect real-time ozone and particulate matter (PM) data and send the data to EPA for processing. In addition, these agencies forecast air quality for the current and next day and send the forecasts to EPA each day. AIRNow acts as a clearinghouse for all this information, distributes it to the public, and provides the information (observations and forecasts) to media companies. The Weather Channel and *USA Today* distribute the air quality forecasts on their web sites (e.g., http://www.weather.com) and on the weather page of *USA Today*. Information is also sent to Weather Service Providers (WSPs), who provide data to television, web, and newspaper clients throughout the United States.

In summer 2000, *USA Today* began including Air Quality Index (AQI) forecasts for 32 of the 36 cities for which it provides detailed weather forecasts. AQI forecasts for 4 of the 36 cities—Columbus, Ohio; Memphis and Nashville, Tennessee; and Minneapolis, Minnesota—were not included because these cities did not have air quality forecasting programs. To help these cities build forecasting programs to meet *USA Today*'s need for the 2001 season and to help develop state and local agency staffs' forecasting capabilities, EPA and the Lake Michigan Air Directors Consortium (LADCO) funded Sonoma Technology, Inc. (STI) to provide daily ozone forecasts from May 1 through September 30, 2001, develop forecasting tools, and transfer the tools and knowledge gained to state and local agencies to aid in forecasting during future ozone seasons.

*Corresponding author address: Dianne S. Miller, Sonoma Technology, Inc., 1360 Redwood Way, Suite C Petaluma, CA 94954; e-mail Dianne@sonomatech.com

The types of forecasting tools developed by STI include conceptual models of the important processes that influence air quality, forecast guidelines, and regression equations. The conceptual models were developed from historical ozone data, synoptic weather maps, and upper-air soundings. Regression equations were developed using historical ozone data, surface meteorological observations, and Eta weather model data. Regression equations were chosen as a tool because they are easy to use, are objective, and have performed well forecasting ozone in other areas. Because accurate regression equations could not be developed for Minneapolis, forecast guidelines were developed for this city, which are not discussed in this paper.

This report summarizes the activities and results of the development of forecasting tools for use by Columbus, Ohio and Memphis and Nashville, Tennessee. The report covers data sources (Section 2), case study and statistical analyses (Section 3), and tool performance (Section 4).

2. DATA

Developing forecasting tools requires analysis of historic air quality and meteorological data to understand the meteorological processes that influence ozone concentrations. Thus, when the meteorological processes are forecasted to occur, this understanding can be used to predict air quality. In general, the understanding of these processes can be achieved through both objective and statistical analysis of the data.

Historical ozone data were acquired for each forecasting city and nearby sites to characterize the ozone episodes and to develop forecasting tools. Columbus had 8 monitors evenly distributed across the city. Nashville had 7 monitors, all located south or east of the city. Memphis only had 2 monitors (see Figure 2-1) which, due to their lack of spatial coverage, cannot capture high ozone under all meteorological conditions. Hourly 1-hr ozone concentrations for each monitor in the forecast region were obtained from EPA's Aerometric Information Retrieval System (AIRS). A five-year data set (May through October, 1996-2000) was sufficient to capture enough episodes for analysis and provide a robust data set for developing forecasting tools. From the hourly data, the daily maximum 8-hr average ozone concentration was determined for each monitor, as well as the regional maximum 8-hr average ozone concentration.

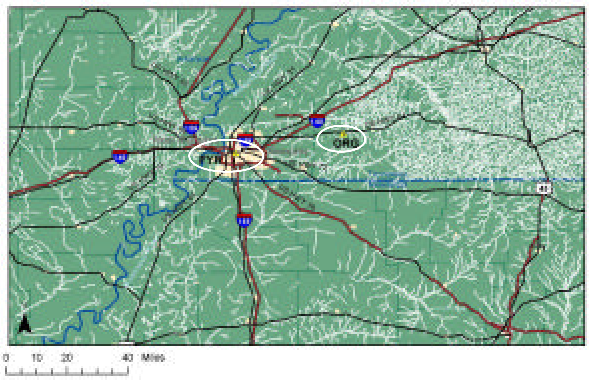


Figure 2-1. Memphis site map.

Historical meteorological data were acquired for each city for May through October, 1996 to 2000, from regional climate centers. Hourly surface observations from the primary National Weather Service (NWS) monitoring site in each city included temperature, dew point temperature, relative humidity, pressure, cloud cover, precipitation, and wind speed and direction. Upper-air rawinsonde data from Wilmington, Ohio and Nashville, Tennessee included aloft height, temperature, dew point temperature, wind speed, and wind direction, at 00Z (1900 EST) and 12Z (0700 EST). Additional upper-air data were extracted from the Eta model initializations at 00Z (1900 EST) and 12Z (0700 EST). General large-scale weather patterns at the surface and aloft (~500 m) were obtained from the NWS Daily Weather Maps.

All data were checked for quality prior to use. Extreme values were eliminated, and values were checked for consistency with nearby data points. All data were converted to Greenwich Mean Time (GMT) begin time. Units were standardized to m/s for all wind speed variables, and degrees centigrade or Kelvin for all temperature variables. The hourly surface meteorological data were used to compute additional variables, such as maximum surface temperature, daily temperature range, and average resultant wind speeds for various time periods such as 0900 to 1300 EST and 1300 to 1800 EST. Some additional variables were calculated from the Eta model data as well, including resultant wind speed and direction and relative humidity at all levels. The Eta model data were merged with surface weather data and ozone data into a database with one record set per day, resulting in 920 summer days of data for each city over the five-year period.

3. TOOL DEVELOPMENT

Several steps were performed to develop the forecast tools for each city. After collecting all the historical data, an ozone climatology was developed to study long-term trends. Next, a conceptual model was developed to identify and understand how meteorological processes affect ozone; the model results help develop forecast guidelines and support variable selection for regression equations. Finally, statistical analysis was performed using scatter plots, correlations, and factor analysis to determine the final set of variables to be used for developing regression equations.

3.1 *Ozone Climatology*

An ozone climatology was developed for each forecast region to better understand the nature of the air quality in each city. Annual, monthly, weekly, and day-of-week frequencies of exceedances of the federal 8-hr ozone standard were analyzed. An exceedance is any day when the 8-hr average ozone concentration is at least 85 ppb. In addition, diurnal cycles of ozone for exceedance days at each site were reviewed. As an example, Figure 3-1 shows the historical monthly exceedances for Columbus; most of the ozone exceedances in Columbus occurred in June and July. May and September show fewer exceedances, and October shows no exceedances at all from 1996 to 2000. Another example, Figure 3-2, shows the diurnal cycle of ozone for Columbus on exceedance days. This graph illustrates that the highest ozone concentrations occur between 1300 and 1700 EST, and the 8-hr exceedances are due to many hours with concentrations modestly higher than 85 ppb in contrast to a few hours of very high concentrations. This type of information was used to help guide STI forecasters: for example, an ozone exceedance is more likely to occur in June and not in October.

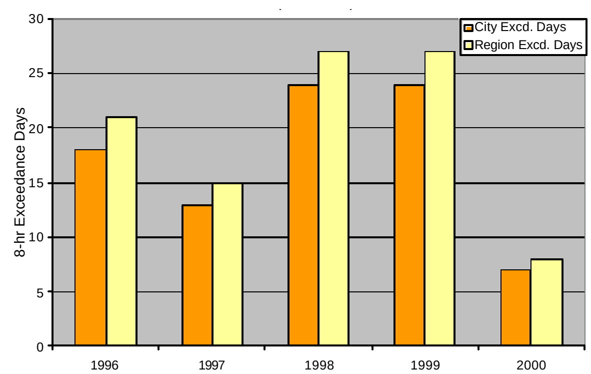


Figure 3-1. Average annual 8-hr ozone exceedances by month in Columbus based on data from 1996-2000.

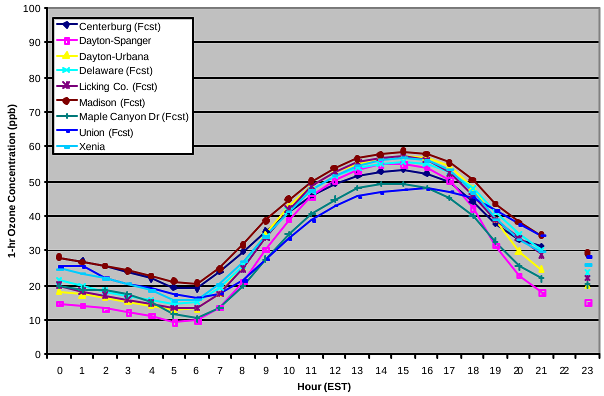


Figure 3-2. Average diurnal ozone concentrations at monitoring sites in Columbus, Ohio, from 1996 through 2000, on days with 8-hr exceedances.

A summary of the climatology for each city is shown in Table 3-1. The first portion of the table shows that Memphis had the most annual exceedances between 1996 and 2000 compared to the other cities, but Columbus averaged more days above 110 ppb than any other city. Columbus had fewer overall exceedances than Memphis and Nashville but had the highest ozone recorded. The middle part of the table shows that, on a monthly average, both Memphis and Nashville experienced the most days when the maximum 8-hr ozone concentration was greater than 100 ppb in August whereas Columbus experienced the most days with a maximum 8-hr ozone concentration greater than 100 ppb in May and June.

Table 3-1. Ozone climatology summary for Columbus, Memphis, and Nashville.

Average days per year			
Ozone	Columbus	Memphis	Nashville
>80	25.6	40.6	35.8
>84	20	30.8	26.4
>90	11.4	19.6	16.6
>100	3.8	6.4	4.2
>110	1	0.4	0.4
Days per month when ozone > 100 ppb			
Month	Columbus	Memphis	Nashville
May	1.2	1	0.6
June	1	1	0.8
July	0.8	1.2	1
August	0.2	1.6	1.2
September	0.6	1.6	0.6
October	0	0	0
Maximum 8-hr ozone concentration from 1996-2000			
	Columbus	Memphis	Nashville
Max O3	130	123	119

3.2 *Meteorological Processes and Conditions That Influence Ozone*

After examining climatological factors, physical relationships between meteorological processes and ozone were studied to develop a conceptual understanding. This conceptual understanding of how meteorology affects ozone concentrations is important for two reasons: (1) forecasters can use it to subjectively forecast air quality and (2) understanding the important physical processes guides the selection of potential predictor variables for developing an objective forecast tool.

3.2.1 General Weather Processes That Influence Ozone Concentrations

There are four basic meteorological processes that affect ozone concentrations: transport, horizontal dispersion of pollution by wind, solar radiation, and vertical mixing.

- Transport. Regionally, ozone or ozone precursors can be transported into an area if there is a pollution source upwind of the forecast area, such as another city. The transported pollutants can combine with local ozone and ozone precursors to produce the observed ozone concentrations.
- Horizontal dispersion. Variation in the wind spreads pollutants horizontally within a region, which acts to lower the pollutant concentrations. In addition, winds can continually move pollution away from the source, which also acts to lower pollutant concentrations in a city.
- Vertical mixing. The depth of vertical mixing from the surface to the top of the boundary layer strongly influences ozone or ozone precursor concentrations. A deeper mixed layer typically results in lower pollutant concentrations whereas shallow mixing allows pollutant concentrations to increase near the surface. The depth of the mixed layer strongly depends on the existence, strength, and altitude of an aloft temperature inversion. If the inversion is strong and shallow, vertical mixing will be confined.
- Solar radiation. Solar radiation triggers the reactions that form ozone; thus, if there is more incoming radiation and enough ozone precursors are present, there will be more ozone.

The variability of these processes, which in turn affects the variability in pollution, is governed by other processes and conditions including diurnal heating and cooling cycles, the movement of large-scale high and low pressure systems, and local and regional topography. In general, high ozone occurs when an aloft ridge axis is overhead or just upstream, which causes subsidence, or sinking air, in the mid- to low-levels of the atmosphere. As the air sinks, it adiabatically warms; this warming stabilizes and dries the atmosphere, which may, in turn, form a temperature inversion. These conditions limit vertical mixing and cloud production, which, in turn, increase ozone photochemistry. On the other hand, aloft low-pressure systems are associated with rising motion; no inversion; clouds or rain; strong vertical mixing; and low ozone concentrations.

In addition to influencing clouds and stability, aloft pressure systems influence the strength and location of large-scale surface high and low pressure systems. The large-scale surface pressure patterns interact with local forcing to produce the observed local flows. The local forcing is generally driven by diurnal temperature changes and topography. Often, just downstream of an aloft high-pressure system, a surface high develops. For the area under the influence of the surface high, local flows often dominate. Since local flows are partly controlled by diurnal heating, these flows often recirculate air and are often light, allowing pollution to build up. Also of importance is the contribution of the transport of material from upwind regions. The transported material combines with local emissions to produce the observed pollution concentrations. The transport patterns are controlled by both regional and local flow patterns.

3.2.2 Case Study Examples

High- and low-ozone episodes for each area were analyzed to understand these processes and conditions. Example case studies are presented for a high-ozone and a low-ozone day in Columbus. Several similar case studies were completed to help develop a conceptual model for each city.

High ozone case: August 6, 2001

The maximum 8-hr average ozone concentration was 93 ppb in Columbus, Ohio, on August 6, 2001. The following meteorological conditions associated with high ozone concentrations on August 6 are illustrated in Figure 3-3:

- A broad high-pressure ridge extended across the continental United States with maximum heights of 600 decameters (dm) centered over western Nebraska and heights of 594 dm over Columbus (Figure 3-3a). These high heights indicate a strong high pressure system.
- A moderate temperature inversion was situated at the surface with a weaker inversion near 700 mb; the weaker inversion was likely caused by subsidence from the high-pressure ridge aloft (Figure 3-3b). Another sign of subsidence is the very dry air aloft, shown by the low dew point temperatures above 700 mb. This resulted in generally clear skies throughout the Columbus region, which provided sufficient solar radiation to trigger photochemical reactions. The moderate temperature inversion near the surface limited the depth of vertical mixing and trapped the ozone near the surface.
- High pressure at the surface extending over the entire eastern United States (Figure 3-3c) and weak pressure gradients caused light surface winds. The dashed line on the map in Figure 3-3d shows the 24-hr backward surface trajectory. The short length of the trajectory and its origin point near Columbus show that there was limited transport at the surface from other source regions into Columbus and little horizontal dispersion which allowed ozone to build up in Columbus. Aloft backward trajectories were also short although these are not shown in the figure.

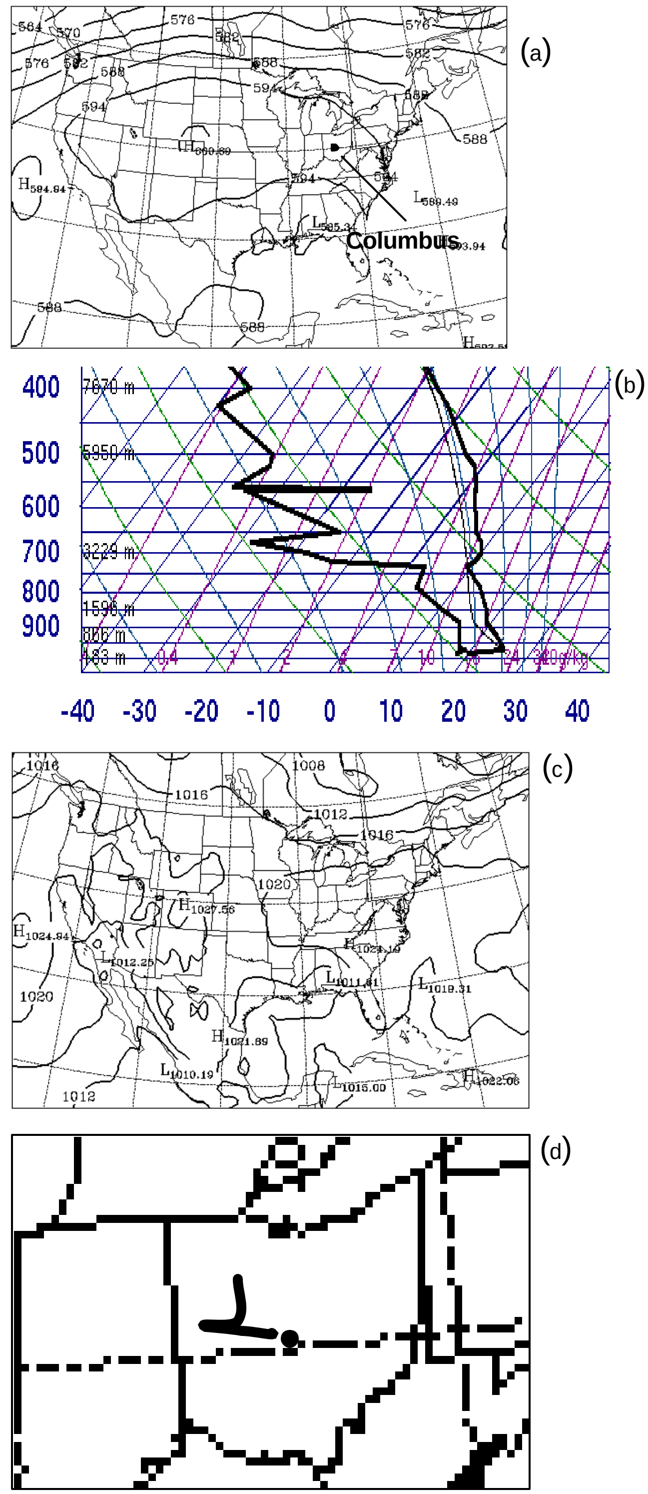


Figure 3-3. Meteorological conditions associated with high ozone concentrations in Columbus, Ohio, on August 6, 2001: (a) 500-mb heights in dm at 12Z (0700 EST). Contours are every 6 dm. (b) Temperature sounding at 12Z (0700 EST) from Wilmington, Ohio. The bold line on the right shows temperature in °C, and the bold line on the left shows the dew-point temperature in °C. Vertical units are

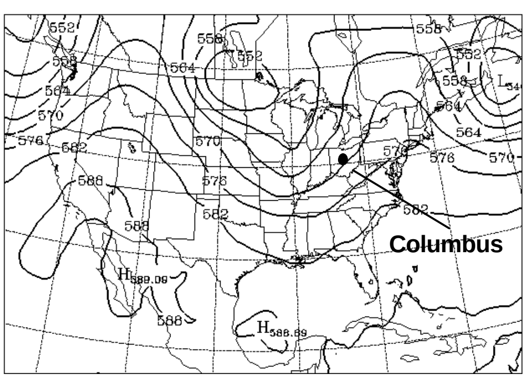
shown in m and in mb. (c) Surface pressure pattern at 12Z (0700 EST). Isobars are every 4 mb. (d) Close-up showing the 24-hr backward surface trajectory.

Low ozone case: June 1, 2001

The maximum 8-hr average ozone concentration was 41 ppb in Columbus, Ohio, on June 1, 2001. The following meteorological conditions associated with the low ozone concentrations on June 1 are shown in Figure 3-4:

- A large, deep trough of low pressure was centered over Illinois, with minimum heights of 552 dm centered over North Dakota and Minnesota and heights of about 570 dm over Columbus (Figure 3-4a).
- There was no inversion in the lower atmosphere. This suggests good vertical mixing throughout the atmosphere, resulting in no build-up of ozone at the surface (Figure 3-4b). The relative humidity was 100%, as shown by the overlapping temperature and dew point temperature lines, indicating cloud cover. There was also precipitation on this day which was confirmed by radar imagery during the morning hours. The conclusion is that there was little or no solar radiation at the surface in the Columbus area; consequently, little ozone formation through photochemistry occurred.
- A well-developed surface low-pressure system was situated over Illinois at 12Z (0700 EST), which indicated strong surface winds (Figure 3-4c). These strong surface winds resulted in the dispersion of continuous emissions in Columbus and transport from the St. Louis, Missouri, area (Figure 3-4d). However, transport was not an issue because the source region of the trajectory was influenced by the same synoptic weather conditions which produced low ozone concentrations in Missouri.

(a)



(b)

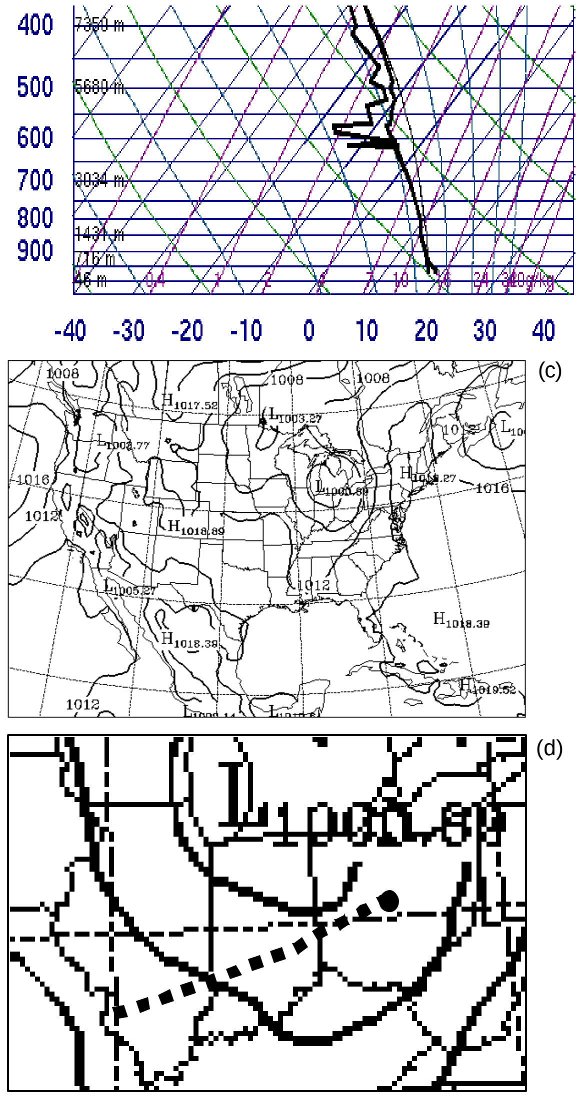


Figure 3-4. Meteorological conditions associated with low ozone concentrations in Columbus, Ohio, on June 1, 2001: (a) 500-mb heights in dm at 12Z (0700 EST). Contours are every 6 dm. (b) Temperature sounding at 12Z (0700 EST) from Wilmington, Ohio. The bold line on the right is temperature in °C, and the bold line on the left is dew-point temperature in °C. Vertical units are shown in both m and mb. (c) Sea-level pressure at 12Z (0700 EST) and 24-hr backward surface trajectory ending at 23Z (1800 EST) in Columbus, Ohio. (d) Close-up showing the 24-hr backward surface trajectory.

In summary, these and other case studies showed that high ozone concentrations occurred in Columbus when there was a strong high-pressure ridge and dry air aloft, a moderate or strong temperature inversion near the surface, clear skies, and light surface winds. Additional case studies showed that high ozone concentrations also occurred in Columbus when there was transport of pollutants from regions to the south into the Columbus area, provided that the other meteorological conditions for high ozone also occurred. The case studies showed that low ozone concentrations occurred when there was a strong low-pressure trough aloft, no temperature inversion near the surface, cloudy skies and/or rain, and moderate to strong surface winds with no transported pollutants.

3.3 *Statistical Analysis*

The conceptual models developed for each city explain the meteorological processes that are important to ozone formation and helped guide the selection of specific variables from the extensive predictor variable list. Scatter plots, correlations, and factor analyses were used to limit the list of predictor variables and to develop objective forecast tools. Data, described by the limited list of variables, were used with statistical software to develop several regression equations, which are discussed in detail in Section 4.1.

The first step of the statistical analysis was to make scatter plots of the potential predictor variables versus ozone. The scatter plots showed two things: the correlation and the type of relationship between the variable and ozone, either linear or logarithmic. Low correlations ($r^2 < 0.2$) provided a quick way of eliminating variables.

Of all the variables, maximum surface temperature had the best correlation with ozone for Columbus, Memphis, and Nashville. Figure 3-5 shows a scatter plot of temperature versus ozone for Nashville. Figure 3-5a shows the linear relationship and Figure 3-5b shows the logarithmic relationship with correlations (r^2) of 0.45 and 0.46, respectively. Because ozone has a log normal distribution, a logarithmic regression equation provided slightly better results.

After the scatter plot analysis, factor analysis was run to eliminate co-linear variables. Factor analysis takes all possible predictor variables and groups them together based on their correlation with each other. If several variables have a correlation coefficient greater than 0.71 (r^2>0.5), they are considered co-linear, meaning the variables change together. Thus, using one variable will have the same effect on statistical equations as using all the variables. In addition, using more than one co-linear variable may over-fit the data, resulting in an equation that may not be representative of all the processes that influence ozone concentrations. As an example, a list of co-linear variables for Memphis included 925-mb relative humidity at 12Z (0700 EST), 925 mb relative humidity at 00Z (1900 EST), surface relative humidity at 12Z (0700 EST), surface relative humidity at 00Z (1900 EST), average daytime surface relative humidity from 14Z to 23Z (0900 to 1800 EST), diurnal temperature range, and 850 mb relative humidity at 12Z (0700 EST). Only one variable from this list, surface relative humidity at 00Z (1900 EST), was used in the equation.

Once the groups of co-linear variables were identified, different variable combinations were used with statistical software to develop several regression equations to predict current- or next-day ozone for each city. The goal was to develop equations with an r^2 of at least 0.65 for the current-day equation and 0.60 for the next-day equation; equations failing to meet this criterion were not considered.

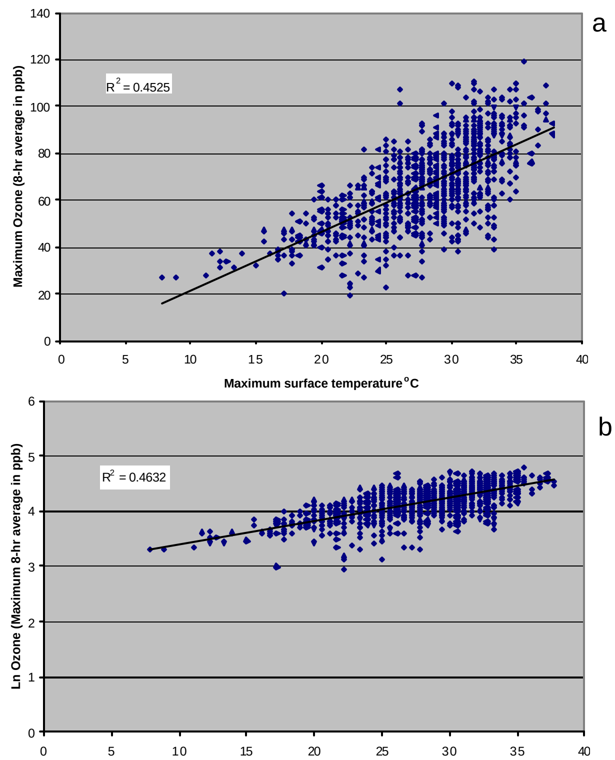


Figure 3-5. Scatter plot of (a) temperature vs. ozone and (b) temperature vs. natural log of ozone for Nashville, Tennessee.

The physical reasonableness of the equations was evaluated using the conceptual models discussed in Section 3.1. Every predictor variable was checked for proper correlation and physical contribution to the equation. For example, the conceptual model showed that high ozone occurs when surface temperatures are hot; thus, temperature should be positively correlated with ozone and have a high r^2 as shown in Figure 3-5. As another example, 500-mb wind direction had very low correlation with ozone (-0.04 for Memphis). This variable was eliminated from the equations because it was statistically insignificant. The standard coefficient was the final factor considered in the equations; it measures the relative weight of each variable towards the predicted ozone. If a variable contributed less than 0.03 towards the final r2, it was eliminated, because it had so little influence on the ozone forecast.

4. REGRESSION EQUATIONS AND THEIR PERFORMANCE

For Columbus, Memphis, and Nashville, two final equations were selected for each city, one for the current-day and one for the next-day ozone forecast. The two equations have all the same variables, with the addition of the natural log of yesterday's maximum 8-hr ozone concentration, shown as lnO3Y, to the current-day equation. The meteorological variables are defined in Table 4-1. The equations are all exponential and predict maximum 8-hr ozone concentrations (ppb) for the current day (O3D1) and next day (O3D2).

Table 4-1. Table of equation variable names and descriptions.

Variable	Description
AMClouds	Average cloud cover from 6 a.m. to 12 p.m. EST
DailyClouds	Average cloud cover from 6 a.m. to 6 p.m. EST
MaxT	Maximum daily surface temperature
MSLPI00	Mean sea-level pressure at 00Z
RH1000I00	Relative humidity at 1000 mb at 00Z
RWS500I00	Resultant wind speed at 500 mb at 00Z
RWS850I12	Resultant wind speed at 850 mb at 12Z
Trange	Difference between the maximum and minimum surface temperatures
U850I00	U component of the wind at 850 mb at 00Z
U925I12	U component of the wind at 925 mb at 12Z
V1000I12	V component of the wind at 1000 mb at 12Z
V500I00	V component of the wind at 500 mb at 00Z
V500I12	V component of the wind at 500 mb at 12Z
V925I00	V component of the wind at 925 mb at 00Z
WS12to9	Average resultant wind speed from 12 a.m. to 9 a.m. EST
WS1to6	Average resultant wind speed from 1 p.m. to 6 p.m. EST
WS9to6	Average resultant wind speed from 9 a.m. to 6 p.m. EST

Current day (Columbus):

O3D1 = exp(1.67968 + 0.01624 * MaxT – 0.00259 * RWS500I00 + 0.00718 * Trange – 0.00720 * WS1to6 + 0.00676 * V925I00 – 0.00341 * RH100I00 + 0.30766 * lnO3Y) (4-1)

Next day (Columbus):

O3D2 = exp(2.42110 + 0.02405 * MaxT – 0.00206 * RWS500I00 + 0.00321 * Trange – 0.00592 * WS1to6 + 0.00692 * V925I00 – 0.00406*RH1000I00) (4-2)

Current day (Memphis):

O3D1 = exp(12.3537 + 0.01953 * MaxT – 0.02566 * WS9to6 – 0.04890 * AMClouds – 0.00505 * U925I12 – 0.00430 * V500I12 + 0.00785 * V1000I12 – 0.01153 * RWS850I12 – 0.00682 * RH1000I00 – 0.00009 * MSLPI00 + 0.31770 * lnO3Y) (4-3)

Next day (Memphis):

O3D2 = exp(16.10887 + 0.02922 * MaxT – 0.02317 * WS9to6 – 0.04620 * AMClouds – 0.00621 * U925I12 – 0.00438 * V500I12 + 0.01118 * V1000I12 – 0.01502 * RWS850I12 – 0.00773 * RH1000I00 – 0.00012 * SLP00) (4-4)

Current day (Nashville):

O3D1 = exp (7.182282 + 0.020936 * MaxT – 0.039516 * WS12to9 – 0.030987 * DailyClouds + 0.005541 * V925I12 – 0.003547 * V500I00 – 0.005971 * RH1000I00 – 0.000045 * MSLPI00 – 0.007204 * U850I00 + 0.346756 * lnO3Y) (4-5)

Next day (Nashville):

O3D2 = exp (10.093939 + 0.033022 * MaxT – 0.041539 * WS12to9 – 0.020072 * DailyClouds + 0.007093 * V925I12 – 0.003701 * V500I00 – 0.007061 * RH1000I00 – 0.000062 * MSLPI00 – 0.008655 * U850I00) (4-6)

To evaluate forecast performance, the equations for Columbus, Memphis, and Nashville were tested on 16 to 41 days from the summer of 2001 and compared with independent forecasts, made operationally by humans, and the observed ozone concentrations. The Columbus equation was run from July 21, 2001, to August 6, 2001, the Memphis equation was run from July 20, 2001, to August 30, 2001, and the Nashville equation was run from June 14, 2001, to July 6, 2001, and from August 10, 2001, to August 30, 2001. During these times, AQI levels ranged from Good to Unhealthy for Sensitive Groups in each city. This ensured that the equations were tested on a wide range of conditions. Three different statistics were used to measure forecast performance:

1. Percent correct, the ability to predict exceedances and non-exceedances, was computed using Equation 4-7:

$$PercentCorrect = \frac{(a+d)}{N} * 100 \qquad (4\text{-}7)$$

where:

a is the number of non-exceedances forecast that verified as non-exceedances

d is the number of exceedances forecast that verified as exceedances

N is the total number of forecasts

2. Bias, the average difference from the observed value, was calculated using Equation 4-8:

$$Bias = \frac{1}{N}\left(\sum_{1}^{N}(f-o)\right) \qquad (4\text{-}8)$$

where:

f is the forecasted ozone value

o is the observed ozone value

N is the total number of forecasts

3. Accuracy, the absolute average difference from the observed value, was computed using Equation 4-9:

$$Accuracy = \frac{1}{N}\left(\sum_{1}^{N}|f-o|\right) \qquad (4\text{-}9)$$

where:

f is the forecasted ozone concentration

o is the observed ozone concentration

N is the total number of forecasts

The performance measures for each regression equation are shown in Table 4-2. For Columbus, human forecasts performed better than the equation at predicting exceedances and non-exceedances (percent correct), correctly predicting all 17 current-day forecasts, and 13 of 17 next-day forecasts. The equation correctly predicted 12 of 16 current-day forecasts, and 10 of 16 next-day forecasts. Human forecasts were more accurate than the equation for both current-day and next-day forecasts, but the equation showed less bias in both cases. Figure 4-1 shows plots of the observed maximum 8-hr average ozone concentration, the next-day human forecast, and the next-day equation forecast. Human forecasters performed better than the equations at forecasting the exceedance days, but they did not accurately predict them all, such as August 1, 2001. Note that the equation did predict the general increasing and decreasing trends in ozone concentrations.

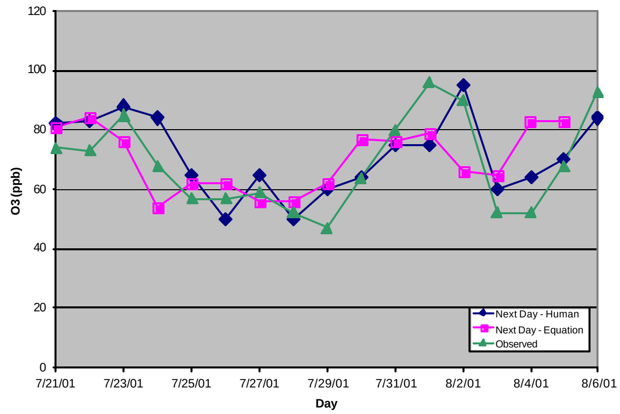


Figure 4-1. Time series of observed, human forecasted, and equation forecasted 8-hr average maximum ozone concentrations in Columbus from July 21, 2001 to August 6, 2001.

The Memphis equation performed slightly better than humans at predicting the AQI category. Both forecast methods were correct for 38 of 42 current-day forecasts while humans succeeded with 33 of 41 next-day forecasts and the equation correctly predicted 36 of 41 next-day forecasts, as shown in Table 4-2. Like Columbus, the Memphis equation showed less bias than humans for both current- and next-day forecasts, although the equation was biased low for current-day forecasts while humans were biased high. Humans were more accurate for current day forecasts while the equation was more accurate for next-day forecasts. The plots in Figure 4-2 show that the trends for both the equation and human forecasts for Memphis were not captured as well as those for Columbus. Neither humans nor the equations predicted three of the four exceedance days (July 26, August 16, and August 25, 2001) or the extremely low day (August 8, 2001) shown in Figure 4-2.

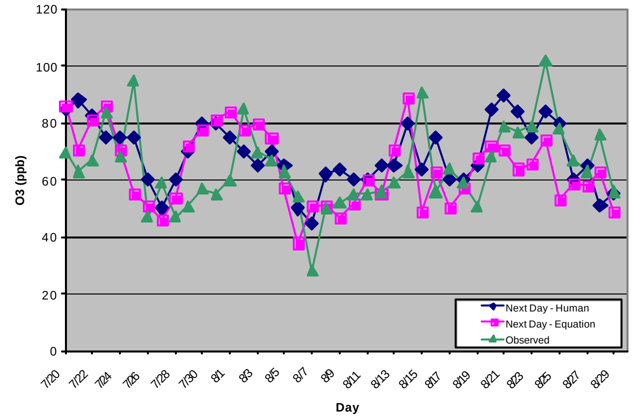


Figure 4-2. Time series of observed, human-forecasted, and equation-forecasted maximum 8-hr average ozone concentrations in Memphis from July 20, 2001, to August 30, 2001.

The Nashville equations performed better than humans for most metrics. The equations were better than humans at predicting the current- and next-day ozone category, were less biased for current- and next-day forecasts, and were more accurate for next-day forecasts. Humans were more accurate for current-day forecasts. Like Columbus and Memphis, the equation did not predict the exceedance days (June 20 and August 15, 2001) or the low days, as shown in Figure 4-3. However, on one exceedance day, August 15, 2001, the equation missed by only 1 ppb, forecasting a maximum 8-hr ozone concentration of 84 ppb. Humans correctly forecasted the two exceedance days but missed the two extremely low days (June 15 and July 6, 2001).

Table 4-2. Statistics comparing the Columbus, Memphis, and Nashville equations to human forecasts.

		Percent Correct		Bias (ppb)		Accuracy (ppb)	
		Current Day	Next Day	Current Day	Next Day	Current Day	Next Day
Columbus	Human	100% (17/17)	76%(13/17)	1.76	2.76	5.65	7.94
	Equation	75% (12/16)	63% (10/16)	0.56	1.81	6.69	8.06
Memphis	Human	90% (38/42)	80% (33/41)	3.3	5.3	6.3	10.1
	Equation	90% (38/42)	88% (36/41)	-2.3	0.6	9.4	9.2
Nashville	Human	91% (29/32)	88% (28/32)	6.1	7.5	8.0	11.4
	Equation	94% (30/32)	91% (29/32)	0.1	4.0	9.6	10.7

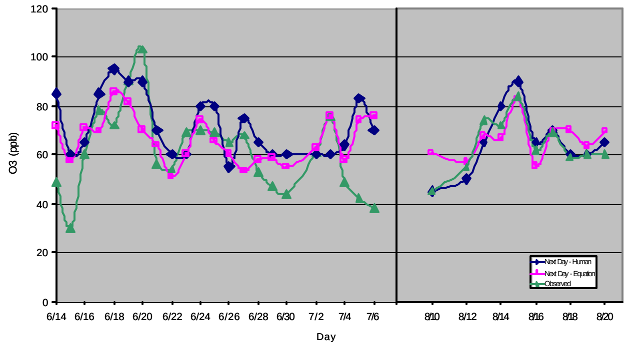


Figure 4-3. Time series plots of human and equation forecasts and observed ozone for Nashville from June 14 to July 6, 2001, and from August 10 to August 20, 2001.

5. RECOMMENDATIONS

The regression equations developed for Columbus, Memphis, and Nashville performed relatively well but under-predicted the high ozone days, a typical problem with regression. The equations are best used to complement other forecasting methods, such as subjective human forecasts. The regression equations might be used to give the human forecaster a quick, "ball-park" prediction, which the forecaster could then refine using experience and a conceptual model. When the regression tool predicts low ozone, the forecaster might use this prediction alone to save time.

We suggest three ways of potentially improving the forecast tools:

- Developing separate regression equations specifically for higher ozone concentrations. These equations could be run in conjunction with the equations discussed in this document when the output from the equations is above a certain concentration.
- Developing guidelines or criteria for Columbus, Memphis, and Nashville similar to those developed for Minneapolis that go beyond the basic conceptual model. These rules could be useful in identifying days when the equations are under-predicting ozone concentrations.
- Testing the equations on a larger data set covering a wider range of ozone concentrations.

ACKNOWLEDGMENTS

This report was prepared in conjunction with work sponsored by the U.S. Environmental Protection Agency (EPA) and Lake Michigan Air Directors Consortium (LADCO). The authors wish to thank the following individuals for their assistance in this project: Mr. Michael Koerber (LADCO), Mr. Rick Strassman (Minnesota Pollution Control Agency), Mr. Robert Brawner (Tennessee Air Pollution Control Division), Mr. Jonathan Howell and Mr. George King (Memphis/Shelby County Department of Health, Air Quality Division), and Ms. Laura Koprowski (Mid-Ohio Regional Planning Commission). We also thank our forecaster colleagues at Sonoma Technology, Inc.: Mr. Charley Knoderer, Mr. Craig Anderson, Ms. Beverly Thompson, Mr. Shane Motley, and Mr. Duc Nguyen.