METDAT METEOROLOGICAL DATABASE SYSTEM: DATA ACQUISITION, PROCESSING, AND ACCESSIBILITY THROUGH THE AIRNOW PORTAL

Garnet B. Erdakos*, Kenneth J. Craig, Jennifer L. DeWinter, and Stephen B. Reid
Sonoma Technology, Inc., Petaluma, California, USA

1. INTRODUCTION

On behalf of the U.S. Environmental Protection Agency (EPA), we developed the Omnibus Meteorological Database, or MetDat meteorological database system, to support adjustment of air quality trends to account for year-to-year meteorological variability. The MetDat database includes a synthesis of raw and computed meteorological parameters from surface and upper-air observations, model reanalyses, and trajectory calculations. It includes a wide variety of more than 100 quality-controlled parameters for 1995 through 2011 for more than 700 sites across the continental United States (Fig. 1).

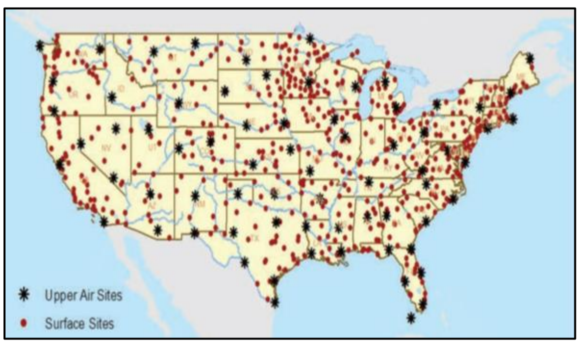

Fig. 1. Locations of surface and upper-air meteorological monitoring sites included in the MetDat database.

MetDat is useful for developing statistical regression equations for ozone forecasting and for other applications. The regression models describe the relationship between ozone concentrations and meteorological variables, and one or two meteorological variables often explain most of the variance in ozone concentrations. For example, Fig. 2 shows, for a sample case, that the daily maximum and minimum hourly temperatures (Tmax and Tmin) explain more than 60% of the variance in ozone concentrations.

**Corresponding author:* Garnet B. Erdakos, Sonoma Technology, Inc., 1455 N. McDowell Blvd., Suite D, Petaluma, CA 94954; email: gerdakos@sonomatech.com

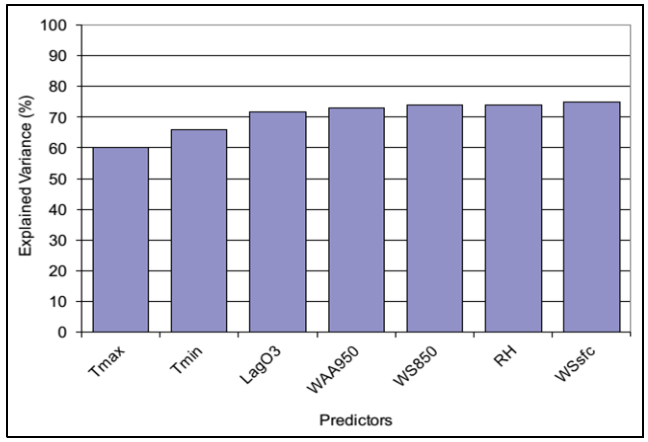

Fig. 2. Chart of accumulated explained variance in daily ozone concentrations provided by seven variables incorporated into a regression equation. Tmax = daily maximum hourly temperature; Tmin = daily minimum hourly temperature; LagO3 = the preceding day's maximum 8-hr average ozone concentration; WAA950 = warm air advection at 950 mb; WS850 = wind speed at 850 mb; RH = relative humidity; WSsfc = average surface wind speed.

One application of statistical regression models for forecasting ozone is to estimate fire impacts on ozone concentrations. If an air agency demonstrates to the EPA's satisfaction that monitored ozone concentrations were influenced by exceptional events, such as large wildfires, the corresponding ambient data can be excluded from use in regulatory decision-making.

The EPA has made a subset of MetDat data, including 64 meteorological parameters, available through the AirNow-Tech data portal (https://www.airnowtech.org/metdat/metdathome.cfm). AirNow-Tech is a password-protected website for air quality data management analysis and decision support. It contains a suite of tools connected to the EPA's national data portal. It includes a data acquisition tool for the subset of MetDat parameters, which are useful for developing regression equations for ozone forecasting. The data sites are paired with ozone monitoring sites through EPA-established links. MetDat data from the portal are available for 2007 through 2011 in both unpivoted and pivoted format; the pivoted format is ready for direct use in a statistical software package.

2. THE METDAT DATABASE SYSTEM

The MetDat database system consists of streamlined computer scripts and programs for acquiring and processing historical observed and predicted meteorological data to produce a database of more than 100 meteorological parameters. Input data for MetDat include National Climatic Data Center (NCDC) Integrated Surface Hourly (ISH) and Integrated Global Radiosonde Archive (IGRA) quality-controlled surface and upper-air meteorological observation data sets, and the regional, long-term (1979 to present), dynamically consistent, high-resolution, three-dimensional National Center for Environmental Prediction (NCEP) North American Regional Reanalysis (NARR) data set. Using extraction algorithms, parameterizations, and the Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) model, MetDat produces observed and derived meteorological parameter outputs for either short- (1, 3, 4, 6, and 12-hour) or long-term (daily and monthly) time periods at selected monitoring sites.

3. METDAT IN THE AIRNOW-TECH PORTAL

To facilitate the development of regression equations for ozone forecasting, a web-based data acquisition tool for MetDat was integrated into the AirNow-Tech software system (Fig. 3). We identified a subset of 64 MetDat parameters that are most useful for developing regression equations for ozone forecasting. We used existing site links established by the EPA to pair ozone monitoring sites with MetDat sites so that users can easily identify the more representative meteorological data for their area of interest.

Federal, state, local, and tribal air quality agencies can access the MetDat data portal by logging in to www.airnowtech.org. (Users can register for an AirNow-Tech account on this page.) MetDat is linked under the Tools menu (circled in red in Fig. 3). Once on the MetDat page, users select the surface meteorological monitoring site for which to download data, the meteorological parameters to download for the chosen site, and the desired range of years and months.

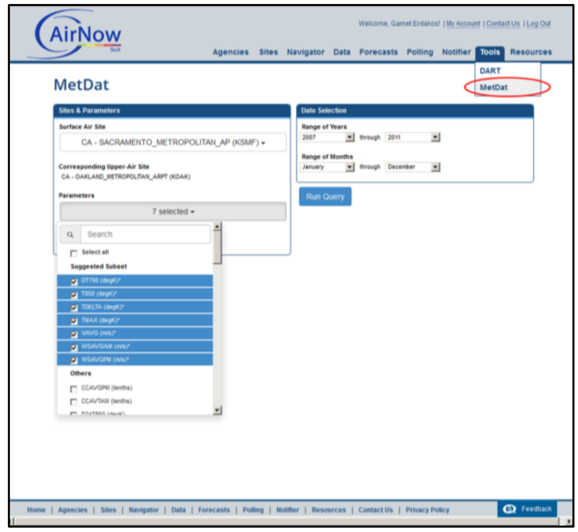

Fig. 3. The MetDat data acquisition tool in the AirNow-Tech data portal. Selections are shown for a download of MetDat data for Sacramento International Airport, the subset of seven suggested parameters, and all available dates.

Users can hover over any meteorological parameter in the parameter selection list to view a full description. Parameters marked with an asterisk are a suggested subset to develop ozone forecasting regression equations. Resource pages are linked on the entry page to the MetDat portal (https://www.airnowtech.org/metdat/metdathome.cfm). Resources include parameter definition tables, a data inventory table, and more information about the suggested parameter subset.

Many of the meteorological parameters are calculated from hourly surface observations. Occasionally, less than 75% of the potential hours of data are available for calculating the metrics. These incomplete data can be included in the data set query by checking the box next to "Include incomplete data."

The data request is submitted by choosing Run Query. File links will appear in the Results pane when the data set is ready to be downloaded. Then, a link for an unpivoted data file will appear; this file includes columns for date, parameter, value, and completeness status. If the "Include incomplete data" box was not checked, a link to a pivoted file, which displays each parameter in a separate column, will also appear.

4. CASE STUDY: USING METDAT DATA TO EVALUATE THE IMPACT OF SMOKE ON OZONE

From June 20 through June 22, 2008, lightning strikes ignited a series of wildfires in northern California. Over the next several weeks, more than one million acres burned. Much of the burned acreage was within 200 miles of Sacramento, and the Sacramento region was covered in thick smoke from June 23 through most of July. During this period, area air quality monitors recorded extremely high ozone concentrations (Sacramento Metropolitan Air Quality Management District, 2011).

In support of an exceptional events demonstration package submitted by the Sacramento Metropolitan Air Quality Management District (SMAQMD) for these fire-related ozone exceedances, STI applied a statistical regression method to describe the relationship between ozone concentrations and multiple meteorological variables. Regression analysis is a well-documented technique that has been successfully used to forecast pollutant concentrations in many parts of the United States (U.S. Environmental Protection Agency, 2003).

We had previously developed a regression equation for the Sacramento region based on six years (1997 to 2003) of monitored ozone data for Sacramento County, as well as surface and upper-air meteorological data for the Sacramento area for the same six-year period. These meteorological data are available in the MetDat database, but they are not available through the AirNow portal; only MetDat data for 2007 through 2011 are available through the AirNow portal. We developed the regression equation in 2004 to assist with daily ozone forecasting in the Sacramento area; during evaluation of fire impacts on ozone for 2007 and 2008, we updated the equation to account for reductions in ozone precursor emissions that occurred between the development (1997 to 2003) and application (2007 to 2008) periods.

In addition, we estimated the error associated with this prediction method by evaluating differences between daily predictions and observations for the 2006, 2007, and 2008 ozone seasons. After calculating these differences, we plotted the distribution of error and determined that the errors were normally distributed around the mean, indicating that the errors were random with no directional bias. We then calculated the 95th percentile values (approximately two standard deviations) for the error distribution to provide extremely conservative upper and lower bounds for the predictions. For the upper bound, we found that 95% of the daily differences did not exceed 27.6 ppb of ozone, as shown in Fig. 4. We used this value as a statistical threshold, or "regular upper limit," for predicted ozone concentrations.

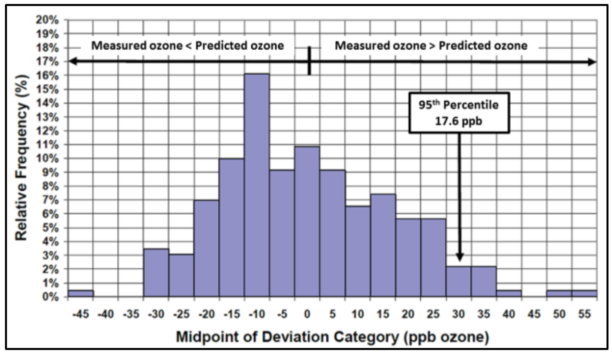

Fig. 4. Distribution of differences in observed and predicted ozone concentrations for Sacramento County for the summers of 2007 and 2008.

Results from the statistical techniques described above were used to estimate fire impacts on Sacramento ozone concentrations during selected days in June and July 2008 by comparing predicted values for no-fire conditions against observations. These analyses showed that estimated fire impacts on ozone concentrations in Sacramento ranged from 53 to 84 ppb on June 23, June 27, and July 10, 2008 (see Fig. 5), and demonstrated that "but for" the wildfires, there would have been no exceedances of the federal 1-hr ozone standard that was in effect for the Sacramento Metropolitan area on those three dates.

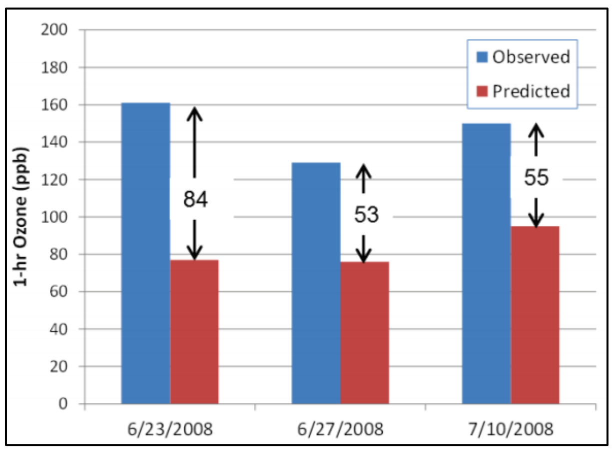

Fig. 5. Observed vs. predicted ozone concentrations for Sacramento for three dates in the summer of 2008.

To account for uncertainties associated with the regression equation, the 95th percentile difference of 27.6 ppb between observed and predicted ozone values was added to ozone predictions to establish the regular upper limit described above, and the observed concentrations for the three days in question still exceeded this upper bound (Sacramento Metropolitan Air Quality Management District, 2011).

5. CONCLUSION

The MetDat database includes more than 100 quality-controlled meteorological parameters for 1995 through 2011 and more than 700 sites across the continental United States. A subset of MetDat parameters is accessible through the password-protected AirNow-Tech data portal. As shown in the case study here, MetDat data can be used in combination with monitored ozone data to develop a statistical regression equation for ozone forecasting, and the regression equation can then be used to estimate wildfire impacts on ozone concentrations by comparing predicted values for no-fire conditions against observations. If an air agency demonstrates to the EPA's satisfaction that measured ozone concentrations were influenced by wildfires, the ambient data can be excluded from use in regulatory decision-making.

In the case study here, very conservative methods were used to establish acceptable uncertainty bounds, as the fire impacts were large, and measured concentrations were well above the federal 1-hr ozone standard of 120 ppb. In most cases, the federal 8-hr ozone standard of 75 ppb will be of concern, and measured concentrations may exceed that standard by more modest amounts. As a result, uncertainty considerations will play an important role in the application of this method to fire-influenced exceptional events.

6. REFERENCES

Sacramento Metropolitan Air Quality Management District (2011) Exceptional events demonstration for 1-hour ozone exceedances in the Sacramento regional nonattainment area due to 2008 wildfires. Report to the U.S. Environmental Protection Agency, March 30.

U.S. Environmental Protection Agency (2003) Guidelines for developing an air quality (ozone and $PM_{2.5}$) forecasting program. Prepared by Dye T.S., MacDonald C.P., Anderson C.B., Hafner H.R., Wheeler N.J.M., and Chan A.C. at Sonoma Technology, Inc., Petaluma, CA, STI-902461-2295-FR, EPA-456/R-03-002, June. Available at http://www.epa.gov/airnow/aq_forecasting_guidance-1016.pdf.