Developments in Environmental Science, Volume 8
A. Bytnerowicz, M. Arbaugh, A. Riebau and C. Andersen (Editors)

ISSN: 1474-8177/DOI:10.1016/S1474-8177(08)00024-7

Chapter 24

A Statistical Model for Forecasting Hourly Ozone Levels During Fire Season

Haiganoush K. Preisler, Shiyuan (Sharon) Zhong, Annie Esperanza, Leland Tarnay and Julide Kahyaoglu-Koracin*

Abstract

Concerns about smoke from large high-intensity and managed low-intensity fires have been increasing during the past decade. Because smoke from large high-intensity fires are known to contain and generate secondary fine particles ($PM_{2.5}$) and ozone precursors, the effect of fires on air quality in the southern Sierra Nevada is a serious management issue. Various process-based models have been developed for forecasting PM and ozone levels in the presence and absence of fires. Although these models provide deterministic predictions, few of them give measures of uncertainties associated with these predictions. Estimates of uncertainties are essential for model evaluation and forecasting with known precision levels. In this chapter we present a statistical procedure for forecasting next-day ozone levels at given sites. The statistical model takes into account some of the known sources of ozone fluctuations, including changes in temperature, humidity, wind speed, wind direction and, during fire season, effects of smoke from fires. Other sources of variation not directly accounted for in the model—e.g., variability in daily amount of ozone produced by sources other than fire—are included in the uncertainty measure as random effect variables. The advantage of a model that is capable of estimating mean effects and uncertainties simultaneously is that evaluation of model performance is immediate and predictions are available with specific precision levels. The ability of the model in making accurate forecasting with specified precisions is demonstrated by applying it to real data set of observed ambient ozone and weather values at two sites in the Sierra Nevada for the

*Corresponding author: E-mail: hpreisler@fs.fed.us

period from 1 January to 31 July 2006. Forecasted $PM_{2.5}$ values from the BlueSky Smoke Dispersion Model are tested as a proxy for the amount of pollution precursors reaching a given site from specific fires. The forecasts from the statistical model may be useful as a tool for air quality managers to time-prescribed fire treatment.

24.1. Introduction

Concerns about smoke from large high-intensity and managed low-intensity fires have been increasing during the past decade. There is a growing awareness that smoke from some fires may cause exposure to hazardous air pollutants. Studies using satellite images and long-range atmospheric transport models have shown that smoke from large high-intensity fires can be transported across continents or even oceans, producing air quality impacts that can be detected on a continental scale or beyond. Managed low-intensity burns, although usually smaller and less intense, nevertheless may also have an impact on local and regional air quality.

Emissions from managed low-intensity and large high-intensity fires are of special concerns for landscape and air quality managers in the Sierra Nevada. Fire is a natural part of the landscape of the Sierra Nevada. Fire history studies show pre-European fires were widespread and both low-intensity and high-severity fires were frequent (Caprio & Lineback, 2000). Fire suppression beginning in the late 19th century has dramatically changed the historical fire regime, altering the ecological structure and function of Sierra forests. Current fuel conditions show substantial accumulation of live and dead fuels, and require that federal land and air quality managers develop strategies to reduce fire danger and maintain ecosystem integrity. As such, one of the strategies is to use managed low-intensity burns for air quality, safety and resource benefits.

Smoke generated by managed low-intensity burns can have a substantial impact on the air quality in the Sierra Nevada, especially on local scales. Particulate matter contained in fire smoke is one of the greatest concerns due to its impacts on public health and visibility (Billington et al., 2000). At larger scales, burning of forests can also generate substantial concentrations of O_3 downwind of the fire through reaction of nitrogen oxides (NO_x), volatile organic compounds (VOC) and sunlight (Finlayson-Pitts & Pitts, 1993; U.S. EPA, 2001), although only very large fires have been shown to contribute substantially to higher ozone levels at regional scales.

The Sierra Nevada is home for several National Parks and Wilderness Areas that have been, under the Clean Air Act, designated as Class-I airsheds and must maintain the highest standard for air quality. The Sierra Nevada and its National Parks are susceptible to poor air quality because of its close proximity to the San Joaquin Valley, CA that contains four of the top six most polluted cities in the United States. Pollutants emitted from sources in the San Joaquin Valley can be transported to the National Parks in the Sierra Nevada by terrain-induced regional and local circulations. In 2003 alone, Sequoia and Kings Canyon National Park recorded 72 days when the observed ozone level exceeded the federal 8-h ozone standard (NPS, 2003). Smoke from wildland and prescribed fires in the Sierra Nevada may lead to the exceedances of ozone and particulate air quality standards in the National Parks in the mountains, but also contribute to the already serious air pollution problem in the San Joaquin Valley through atmospheric transport. The Sierra Nevada federal land managers are, therefore, constantly challenged by the conflict between the use of managed low-intensity burns to reduce fire danger and maintain ecosystem integrity and the protection of air quality in the Class-I airsheds in the Sierra Nevada where the background ozone concentration is already high.

Land managers require a tool that may aid their decisions in planning burn operations in ways that create more efficient burn opportunities while minimizing their air quality impact. It has been suggested that a process-based atmospheric dispersion models may be a useful tool in this regard. However, although these models may provide deterministic predictions of amounts of pollutant being dispersed from particular sources (including fires) hardly any of them give measures of uncertainties associated with these predictions. Without measures of uncertainties, it is almost impossible to evaluate model performance. This chapter describes how a statistical model may be developed, using historical ozone and meteorological data from given sites together with output from an atmospheric dispersal model, to obtain estimates of next-day hourly ozone levels with estimates of uncertainties. The statistical methods are an extension of those used in Preisler et al. (2005). As an example of the procedures, the statistical model is used to produces next-day ozone estimates at two sites in the Sierra Nevada. The model uses meteorological variables at each site together with predicted fire-produced $PM_{2.5}$ (particulate matter with diameter less than or equal to 2.5 μm) values from the BlueSky Smoke Dispersion Modeling System (Larkin et al., 2008). The BlueSky $PM_{2.5}$ values are used as a proxy for potential contributions of smoke from fires in the region. Since atmospheric dispersion models such as BlueSky already utilize real-time meteorological variables to make predictions of PM concentrations, the question may be raised as to why meteorological variables are used again in the statistical model. One reason for including meteorological variables is to assess the output of the dynamic dispersal model. If the dynamic model is capturing the full effects of meteorological variables then these variables would ideally drop out of the statistical model (i.e., found not significant). The latter is particularly relevant when we are forecasting PM levels at a site using the PM values from a dynamic model as one of the predictors. When the goal is to forecast ozone levels, however, the weather variables at the site are likely to be the most important predictors of background diurnal ozone values. Since weather stations are commonly found at many sensitive locations in the Sierra Nevada and BlueSky real-time forecasts are also available for the region, the statistical model developed here for two sites in the region may be adapted and applied to other sensitive locations in the Sierra Nevada. The models are not limited to BlueSky. Forecasts from other dynamic transport models may be used in a similar fashion in order to appraise their ability to forecast next-day ozone or PM levels during fire season.

24.2. Methods

24.2.1. Data

Meteorological data were obtained from two weather stations in Sequoia National Park. The Lower Kaweah station is located at 1902 m above mean sea level (MSL) and Ash Mountain station is at 535 m MSL. The meteorological data included hourly values of temperature, relative humidity, wind speed, wind direction and solar radiation. Ozone concentrations (ppb) were also recorded at these stations (Fig. 24.1). Ozone concentrations were measured with a Thermo Environmental Model 49 UV absorption instrument operated by the National Park Service. The ozone monitor was calibrated at the beginning of the season and checked against a calibrator on a weekly basis. Air temperature and relative humidity were measured with a Vaisala temperature and humidity sensor mounted at approximately 2 m in a self-ventilated, louvered shelter. Wind speed and direction were monitored with a MetOne anemometer mounted on a 10-m tower.

Forecasted $PM_{2.5}$ values at the two locations were obtained from BlueSky output for the same period. BlueSky is a smoke dispersion modeling forecasting system that combines burn information with models of consumption, emissions, meteorology, and dispersion to yield a

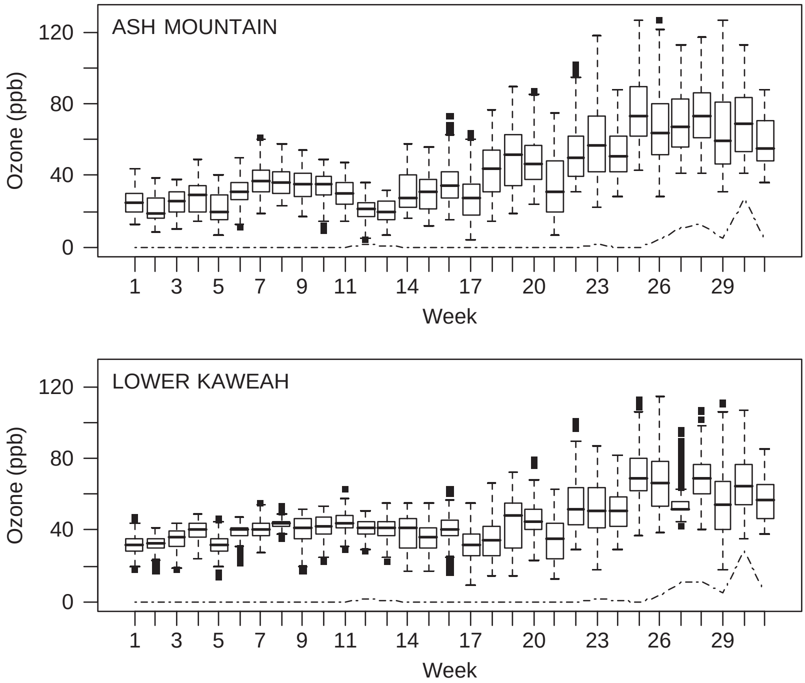

Figure 24.1. Boxplots of hourly ozone values at two sites in Sequoia National Park for the first 31 weeks (1 January to 31 July) in 2006. The dashed curves at the bottom of each panel are the weekly total $PM_{2.5}$ values from BlueSky simulations for that site.

prediction of trajectories and surface concentrations of particulate matter (both $PM_{2.5}$ and PM_{10}) from managed low-intensity fires, wildfires, and agricultural burn activities (Larkin et al., 2008). Currently BlueSky smoke predictions from wildfires are available daily for many locations in the United States. In California, BlueSky has been implemented by the California and Nevada Smoke and Air Committee (CANSAC, http://www.cefa.dri.edu/COFF/coffframe.php) (Brown et al., 2003) into its operational weather forecast system. The system employs the Fifth Generation Penn State University/National Center for Atmospheric Research (NCAR) Mesoscale Model (MM5, http://www.mmm.ucar.edu/mm5/) (Grell et al., 1994) in an operational mode. The initial and boundary condition inputs for the MM5 model are prepared using 6-hourly 40-km ETA forecasts and observations obtained from the NCAR's Unidata data stream. The BlueSky $PM_{2.5}$ forecasts for each day include smoke emissions from all fires reported by various agencies and individuals for the previous day. Simulations for each day start at 5 am local time. All fires from the previous day are assumed to have started at the starting time of the simulations.

We used the predicted $PM_{2.5}$ values from BlueSky forecast to characterize the amount of PM produced by smoke from both large high-intensity and managed low-intensity fires in the region. Outputs from other transport models, with the capability to produce spatially and temporally explicit values in real time, may also be used.

24.2.2. Statistical model

A statistical model is developed that links the observed ozone concentrations with the observed meteorological conditions and BlueSky-predicted $PM_{2.5}$ values at two sites in the Sequoia National Park. The goal here is to forecast next-day ozone concentrations at a given site. While meteorological conditions at the site are likely to be good predictors of background ozone concentrations they will not pick potential fluctuations due to a particular source of pollution such as fires. For that we use predicted $PM_{2.5}$ values from BlueSky output. Observed PM values at the site may be better predictors; however many sites, including the two in this study, do not have observed PM values. Even at sites that have PM observations, it is very difficult to separate the PM concentration produced by fire emissions from the contributions by other PM sources. The BlueSky-predicted PM values at a given location and a given time is a result of smoke plumes as they are transported and dispersed from the locations of all fires in a region.

Standard multiple regression models are not appropriate for the analysis of ozone data for a variety of reasons. First, the distribution of hourly ozone values is not well approximated by a Gaussian or other symmetric distribution (Fig. 24.2a). Second, the assumption of independent observations is not valid because hourly ozone values are serially correlated. Lastly, the relationships between ozone levels and the various predictors are nonlinear. For example, the relationship between hourly ozone and wind direction in the Sequoia National Park region is likely to be cyclical with ozone values being higher when wind directions are from the southwest potentially transport ozone precursors from the heavily polluted San Joaquin Valley.

In our study we used a multiple regression model with cube root of ozone as the dependent variable and with an autoregressive error term to account for the serial correlation. The cube transform of ozone values were more closely approximated by the Gaussian distribution

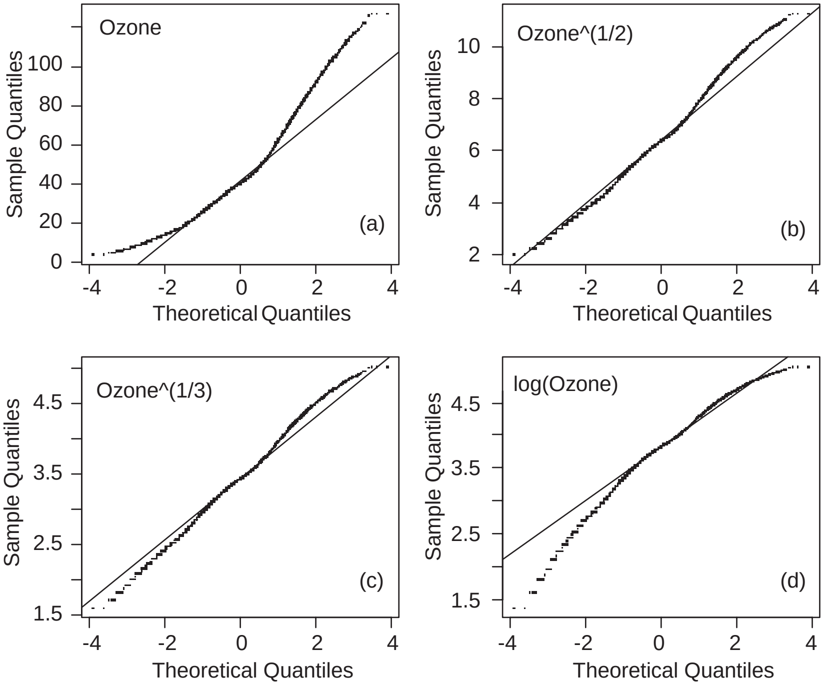

Figure 24.2. Normal probability plots for hourly ozone values (top left panel) and for three transformations. The cube root transformation of the observed ozone values (bottom left) appeared to be more closely approximated by the Gaussian distribution when compared to the square root (top right) or the logarithmic transformations (bottom right).

(Fig. 24.2c). The statement for the statistical model is

$$Y_t = \mu(\mathbf{X}_{t-24}) + \beta Y_{t-24} + \varepsilon_t$$

where Y_t is the cube root of the ozone value at t^{th} hour of the day; μ an additive function of the columns in **X**; $\mathbf{X}_{t-24}$ is a matrix of bases spline transforms of the predictor variable (e.g., weather and BlueSky variables) for the previous day; $\boldsymbol{Y}_{t-24}$ the cube root of the ozone value for the previous day; β a coefficient to be estimated and where $\varepsilon_t = \rho\varepsilon_{t-1}$ is the autoregressive error term. In ordinary regression it is often customary to use some parametric transformation of the predictors (e.g., polynomial or logarithmic function of X) to describe non-linear relationships between the predictors and the predictant. The regression equation described above uses non-parametric transforms of the predictors (e.g., splines), thus allowing the data to suggest the nonlinearities. The latter is achieved

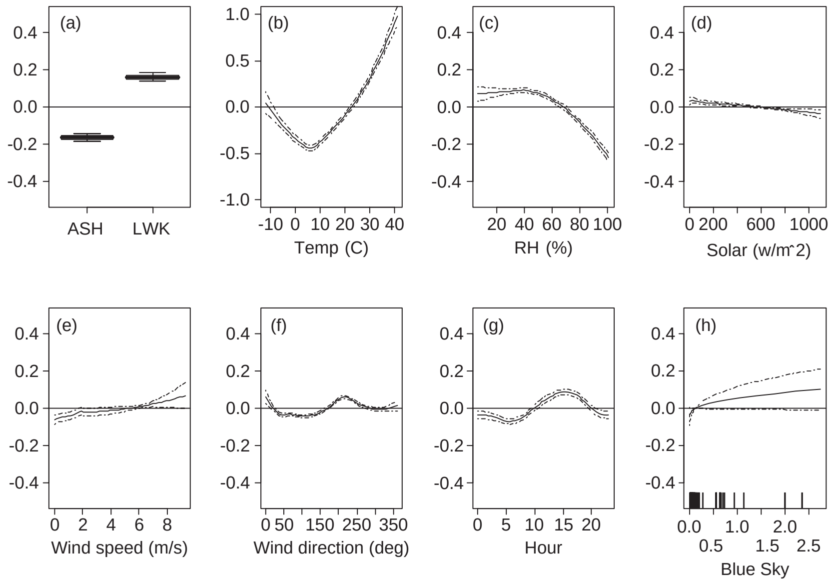

Figure 24.3. Estimated relationships (and 95% point-wise approximate confidence bounds) between predictors and hourly ozone levels. The effects are standardized to have mean zero. If the horizontal line at zero is completely within the confidence bounds then that predictor does not have a significant effect on diurnal ozone concentration. The BlueSky variable is the 80th percentile of the previous day's hourly $PM_{2.5}$ values produced by BlueSky. The hatch marks at the bottom of panel h are the distribution of the observed BlueSky variable.

by fitting polynomial regressions locally after dividing the range of the predictor variables into subsets at specified knots (Hastie et al., 2001). The plots in Fig. 24.3 are an example of fitted non-parametric relationships for the present data. Note that while in ordinary parametric regression the slopes and intercepts of the linear relationships for each predictor are estimated from the data, in non-parametric regression the whole non-linear curves are estimated from the data for all predictors simultaneously.

The above model allowed us to assess the ability of the previous day predictor variables in predicting next-day ozone levels. One may use a similar method to assess the skill of forecasted weather data on ozone levels.

The predictor variables used in the model were temperature, relative humidity, solar radiation, wind speed and wind direction as measured at the two sites. Other two predictors used were hour-of-day in addition to a predictor based on the hourly BlueSky-predicted $PM_{2.5}$ values for that location and time. The $PM_{2.5}$ predictor we used was the 80th percentile of the simulated hourly $PM_{2.5}$ values for the previous 24 h. It was anticipated that the meteorological variables in the model will account for most of the variability in ozone values due to local weather conditions at the time and the BlueSky variable will account for weather conditions at the regional scale that affect the transport of particulate matter and other possible ozone precursors from fires in the region in the previous day. The hour-in-day variable was included as a surrogate for unobserved factors (other than the measured local weather conditions and $PM_{2.5}$ values) that may be affecting ozone levels and that have a 24-hour daily cycle. All estimations were done with the R statistical package (R Development Core Team, 2006).

One outcome of interest to land and air quality managers is whether the ozone level at a particular sensitive site will exceed some critical value (e.g., 90 ppb). We used the forecasted hourly ozone values and their estimated standard errors to appraise the following decision rule:

If $\max_{24\,\mathrm{h}}(\hat{Y} + 2\,\hat{se}) \geq \sqrt[3]{90}$ prescribed burn not recommended on this day, where $\hat{Y}$ and $\hat{se}$ are the forecasted cube root of ozone and its corresponding estimated standard error. Using the above rule, prescribed burns are not recommended on days where the chance of ozone levels exceeding the critical 90 ppb value is greater than a small value ($\sim$2.5%).

We used data from 1 January to 30 July 2006 (the only time period that we had archived BlueSky forecasts) to build the model. We used the period 1 May to 30 July as a test period to calculate the error rate in our decision rule. Ozone concentrations for each day in the test period were forecasted from a model estimated using data from all the days except the day being forecasted.

24.3. Results

On average ozone values at the Ash Mountain site were significantly lower than the corresponding values at the Lower Kaweah site (Fig. 24.3a). All five local weather variables included in the model had significant effects on hourly ozone levels. The partial effect plots (Fig. 24.3b–f) are the estimated non-parametric functions describing the relationships between the predictors and the dependent variable (cube root of ozone). The significant relationship between ozone and hour-of-day (Fig. 24.3g) appears to indicate that there are other sources of variation with a 24-hour cycle that were not accounted for by the five meteorological variables in the model. The hour-of-day variable in the model may be viewed as a surrogate for these missing predictors. Finally, there is some evidence of increases in ozone levels with increasing values of the BlueSky-predicted $PM_{2.5}$ values (Fig. 24.3h). The effect of the BlueSky's $PM_{2.5}$ variable was significant (p-value = 0.0418), however, the standard errors around the estimated curve were large, due to the small sample sizes. The BlueSky-predicted $PM_{2.5}$ levels during the study period were very low. It remains to be seen if the relationship between the BlueSky predictor and ozone concentrations remains the same when a longer period is studied with a wider range of BlueSky values, namely, during a period with more fire activities including some intense fires.

The statistical model appeared to have considerable skill in forecasting ozone levels for the next 24 h using the local weather and ozone levels in the previous day (Figs. 24.4 and 24.5). The decision rule used to forecast whether ozone values will exceed a critical level seems to work well. The forecast missed only one day where the actual/observed ozone level was in excess of the critical level of 90 ppb while the forecasted maximum level (+2SE's) was lower than that. (Fig. 24.6). However, there were many days when the forecast was greater than 90 ppb while the actual observed value was below 90 ppb. One can decrease the number of false positives by making a decision rule based on a lower confidence bound. However, in doing so, the chance that the forecast will miss days with ozone in excess of 90 ppb will increase.

24.4. Discussions

Process-based models for forecasting PM levels during fire season may be a useful tool for land managers if their performance was evaluated and their accuracy of PM forecasts for a particular area or sensitive locations are documented. In order to evaluate model performance it is necessary to quantify the uncertainties of the model predictions. The statistical model developed here is an attempt to demonstrate how uncertainties may be quantified and precisions of forecasts estimated given observational data at a specific site. In this chapter we first demonstrated how one may quantify uncertainties using observational data by applying the statistical model to two meteorological stations in the Sierra Nevada. We then used the estimated statistical model to forecast next-day ozone values during the start of 2006 fire season using BlueSky forecasted $PM_{2.5}$ values as a surrogate for the level of fire activity on a given day. The model can be used as an aid to land managers in making a 'go' or 'no-go' decision with respect to managed low-intensity fires in the Sierra Nevada. To make a

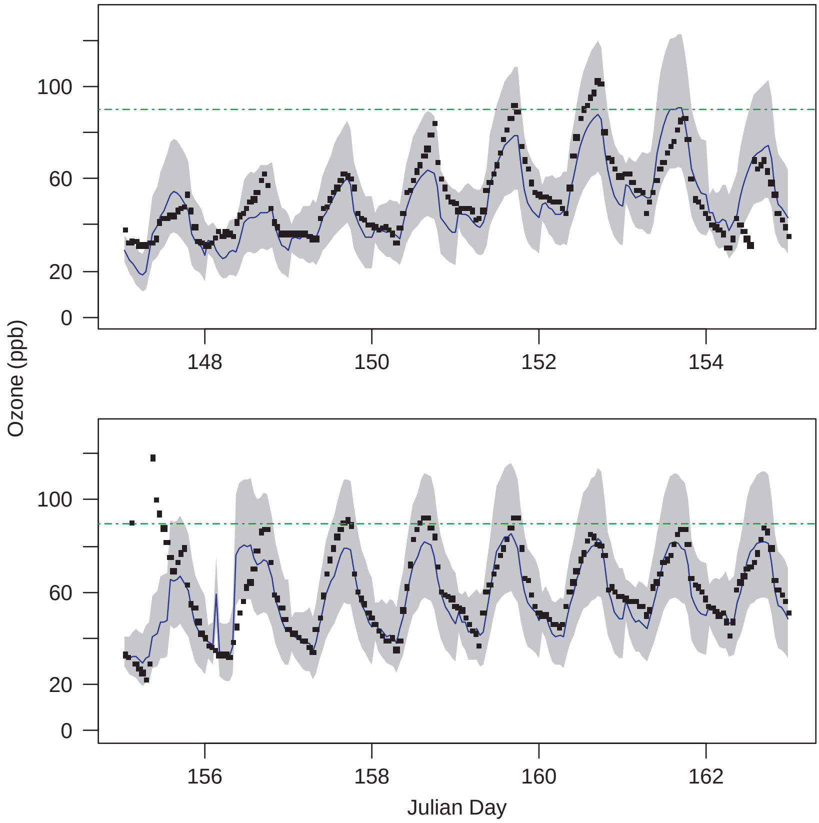

Figure 24.4. Observed (dots) and forecasted (blue curves) hourly ozone values at the Ash Mountain site for a period of 16 days (27 May to 11 June 2006). The gray regions indicate the forecasted point-wise approximate 90% bounds. The green dashed line is at 90 ppb.

forecast for given locations in the Sierra Nevada using the statistical model suggested in this study, the following steps need to be taken:

1. Install meteorological stations at sites near smoke sensitive areas in the Sierra Nevada (i.e., schools and hospitals). These meteorological stations need to have web-based accessibility to retrieve real-time hourly weather information. Co-location of ozone and $PM_{2.5}$ monitors would make these sites ideal for forecasting efforts and would complement efforts to accurately monitor air quality in this region.
2. Collect hourly weather and ozone data from these sites for at least one year. If the site also has the capability to collect $PM_{2.5}$ data, then the

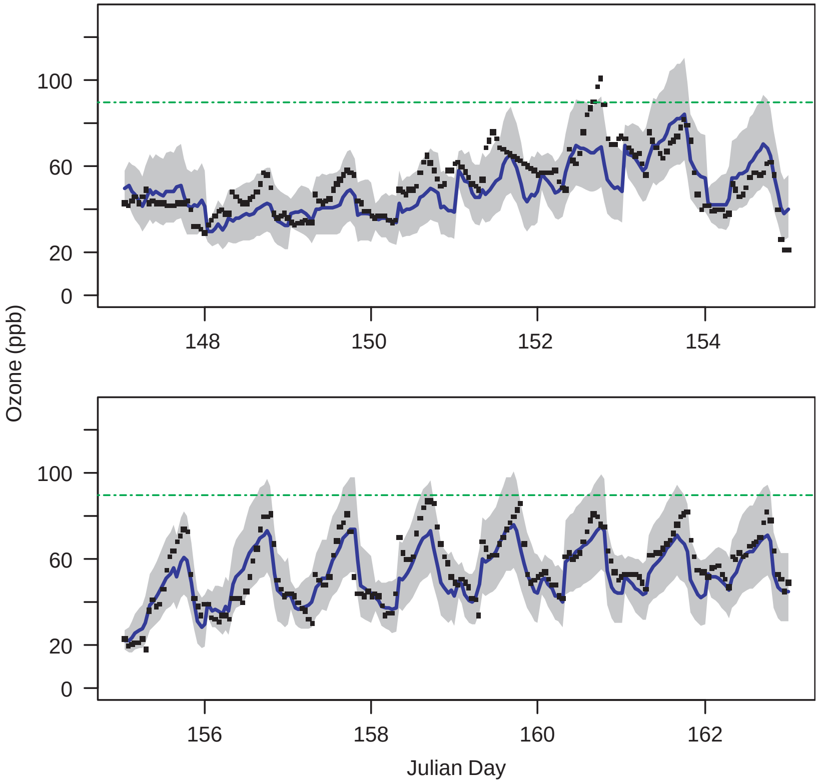

Figure 24.5. Observed (dots) and forecasted (blue curves) hourly ozone values at the Lower Kaweah site for a period of 16 days (27 May to 11 June 2006). The gray regions indicate the forecasted point-wise 90% bounds. The green dashed line is at 90 ppb.

data can be used directly in the statistical model to replace the BlueSky-predicted $PM_{2.5}$ values. Otherwise, BlueSky-predicted $PM_{2.5}$ values will also be needed for the same time period. The observed ozone and meteorological variables and the $PM_{2.5}$ values will then be used to evaluate the statistical model and determine the coefficients that best describe statistically the conditions for the specific location. Historical data at existing monitoring sites may also be used to train the statistical model.

3. During fire seasons, download previous-day data from the meteorological sites by 5:00 am present day.
4. Obtain BlueSky-predicted $PM_{2.5}$ values for that day, also by 5:00 am.
5. Produce forecasted hourly ozone values (and approximate confidence bounds) for the next 24 h starting at 5:00 am.

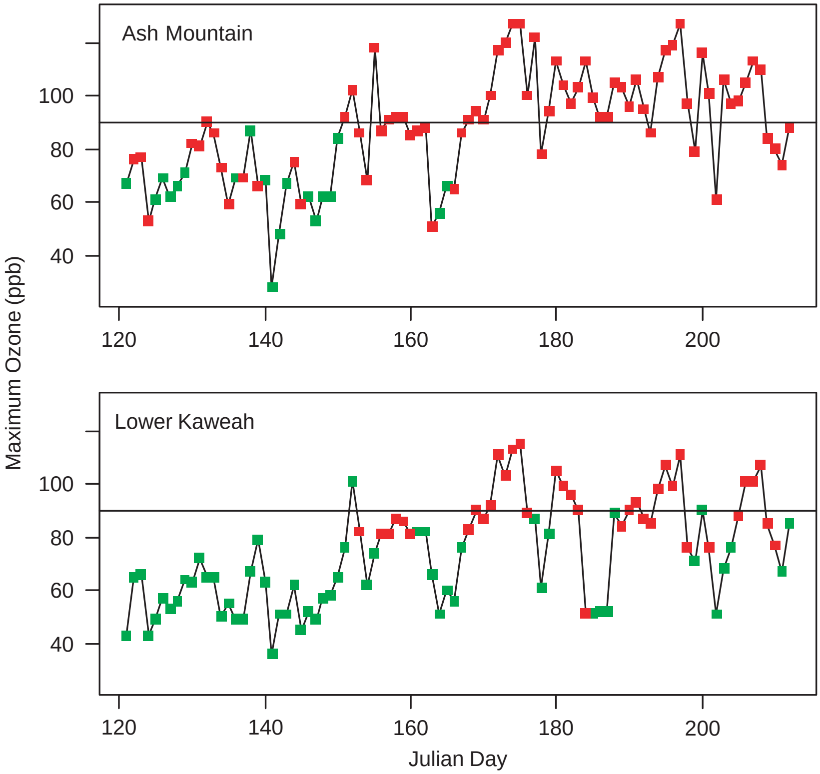

Figure 24.6. Observed 24-hour maximum ozone values at Ash Mountain and Lower Kaweah for the period 1 May to 30 July 2006. Red indicates days when the forecasted upper 95th percentile exceeded 90 ppb. Green indicates days when the forecasted upper 95th percentile was less than 90 ppb.

6. Produce a map indicating all the sites in the study area with red or green sites depending on whether the site is forecasted to exceed the critical level (e.g., 90 ppb hourly average) or not. The latter based on some criteria such as the one described above.
7. Repeat the process the next day.

The current statistical model may be improved by replacing the observed previous day weather conditions with the forecasted values from mesoscale weather forecast models for the current day. As mentioned earlier, mesoscale weather forecasts using the MM5 model, together with forecasts of PM concentrations due to large high-intensity and managed low-intensity fires using the BlueSky Smoke Dispersion Modeling System, are made available operationally by CANSAC for the State of California and Nevada at 4-km horizontal grid spacing.

One potential problem, however, with these mesoscale model forecasts is that for a given location in the Sierra Nevada Mountain Range that is subject to strong influence of local topography, relative large errors may occur in forecast of surface meteorology, especially wind speed and direction. At these locations, the statistical model based on actual on-site observations from previous day, as described in this study, may do a better job in predicting the likelihood of ozone exceedance for the next day at the site.

During this past fire season in 2006, more observational data have been collected at other locations in Sequoia and Yosemite National Parks and Sequoia National Forest in the Sierra Nevada. These data will be used to test the robustness of the statistical model for making ozone forecasting for different weather conditions and topographic settings.

ACKNOWLEDGMENT

We thank the reviewers for their constructive comments and Dr. Andrzej Bytnerowicz for the editorial work essential to the success of this book. We thank the National Park Service Air Resources Division for the hourly ozone and meteorological data from the Lower Kaweah and Ash Mountain stations. The research is funded by a grant from the USDA Forest Service under the Joint Fire Science Plan.

REFERENCES

Billington, J., Ommering, L., Haden, C., Linse, E., Ramalingam, R., and Black, K. 2000. Proposed Amendments of California's Agricultural Burning Guidelines. Staff Report. CALEPA, Air Resources Board, Planning and Technical Support Division, Sacramento, CA.

Brown, T.J., Fujioka, F., and Fontana, C. 2003. The California and Nevada smoke and air committee—an interagency partnership to meet decision-making needs. Proceedings American Meteorological Society Fifth Symposium on Fire and Forest Meteorology, 6 pp.

Caprio, A.C., and Lineback, P. 2000. Pre-twentieth century fire history of Sequoia and Kings Canyon National Parks: A review and evaluation of our knowledge. In Proceedings of the Conference on Fire in California Ecosystems: Integrating Science, Prevention, and Management. 17–20 November 1997, San Diego, CA.

Finlayson-Pitts, B., and Pitts, J. 1993. Atmospheric chemistry of tropospheric ozone formation: Scientific and regulatory implications. Air Waste 43, 1091–1100.

Grell, G.A., Dudhia, J., and Stauffer, D.R. 1994. A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5). NCAR Technical Note TN-398, June 1994, 138 pp.

Hastie, T., Tibshirani, R., and Friedman, J. 2001. The elements of statistical learning: Data mining, inference, and prediction. Springer-Verlag, New York.

Larkin, N.K., O'Neill, S.M., Solomon, R., Krull, C., Raffuse, S., Rorig, M., Peterson, J., and Ferguson, S.A. 2008. The bluesky smoke modeling framework. Atmos. Environ. (Accepted).

Preisler, H.K., Grulke, N.E., Bytnerowicz, A., and Esperanza, A. 2005. Analyzing effects of forest fires on diurnal patterns of ozone concentrations. Phyton (Austria) 45(Special Issue), 33–39.

R Development Core Team. 2006. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org.

U.S. EPA. 2001. Air quality criteria for particulate matter: Second external review draft. EPA 600/P-99/002aB, bB. http://www.epa.gov/nceawww1/pm_vol1.htm

This page intentionally left blank