

Ozone

Ozone gas is produced by photochemical dissociation of atmospheric oxygen as a result of shortwave solar ultraviolet radiation. It is highly concentrated in the stratosphere, mainly between 15 and 40 km. Stratospheric ozone is necessary for life on earth, since it absorbs solar ultraviolet radiation, which should have mutagenetic effects on living tissues of plants and animals (*see* **Mutagenesis, environmental**). In recent years it has been discovered that the ozone layer is thinned in the ozonosphere above Antarctica. This has been caused mainly by chlorofluorocarbons (CFCs) from human activities, which are emitted into the stratosphere. This phenomenon, commonly known as the 'ozone hole', is different from the increase of ozone in the troposphere, which is the lowest part of the atmosphere near the earth's surface, i.e. a thin stratum of atmosphere which on average occurs at 10 km altitude. The chemical and biological processes that regulate life on earth, and most of the meteorological processes, occur in the troposphere.

Ozone occurs in the troposphere for different reasons: the transportation downwards of stratospheric air in special meteorological situations, the production of ozone in the troposphere itself by chemical reaction of precursor compounds, and by natural emissions. Ozone production by chemical reaction is increased by emissions of anthropogenic origin in the atmosphere. A consequence of the increase of tropospheric ozone concentration is the contribution to planetary warming. Ozone is produced as a secondary pollutant by a chemical reaction with NO_x and volatile organic compounds (VOCs), an extended class of gas compounds, mainly non-methane hydrocarbons, which are emitted by oil and its by-products and by biological materials, with concentrations higher than that of the natural background, commonly known as photochemical smog. In the highly polluted areas of industrialized countries, the production of ozone is therefore high. Vegetation causes the natural emission of many nonmethane compounds.

The photochemical reactions that produce ozone depend on meteorological conditions such as solar radiation, wind speed, temperature and pressure. Solar radiation enters directly into the main reactions producing ozone (in the high atmosphere $\text{O}_2 + h\nu \rightarrow \text{O} + \text{O}$; in the low troposphere $\text{NO}_2 + h\nu \rightarrow \text{NO} + \text{O}$,

where $h\nu$ is a quantum of solar radiation). Wind speed determines transport and accumulation of primary pollutants. The temperature influences directly the kinetics of reactions producing ozone and determines the mixing height, which affects the accumulation of primary pollutants. Photochemical smog episodes are usually associated with high-pressure situations. Owing to the strong dependence on meteorological factors, ozone concentrations are highly seasonal. The solar radiation cycle (*see* **Radiation and radiative transfer**) produces minimum values in winter and maximum values in summer. During the warm season, the solar radiation that causes ozone production is high. Moreover, in that period the exchange of air between the lowest troposphere levels and the free atmosphere is scarce, and the wind at soil is usually light. A spring peak in surface ozone also occurs due to vertical transport of stratospheric ozone down into the troposphere. Seasonal variations are different according to each region, because of climatic peculiarities. A daily cycle is also present, since a peak in concentration occurs in the early afternoon. In urban areas ozone concentrations diminish during the night. In fact, the absence of solar radiation associated with the emission of NO_x causes the conversion reaction $\text{O}_3 + \text{NO} \rightarrow \text{NO}_2 + \text{O}_2$. Conversely, in rural areas concentrations are stationary due to the absence of NO_x sources.

Ozone, in high concentrations, is a toxic gas that can damage pulmonary tissues (*see* **Inhalation toxicology**). In the presence of high ozone concentration, breathing problems may affect children and people suffering from **asthma** or undergoing intensive physical effort. Negative effects of tropospheric ozone in plants and in agriculture can reduce crop productivity. It is therefore necessary to control ozone production. This can be done by controlling the emission of nitrogen oxides, mainly due to transport. In the large scale, the reduction of VOC emissions is enough to reduce ozone production, in the small scale both VOCs and NO_x must be reduced. The control and evaluation of pollution levels are linked to the definition of **air quality** standards. Standards and guidelines are stated according to the concentration levels that are considered not to be dangerous for health, and are proposed as quality goals (*see* **Clean Air Acts**). Ozone pollution is monitored at single monitoring sites, often organized in spatial networks. At each site ozone is reported as average hourly concentration, typically in parts per billion. National environmental protection

authorities establish air quality regulations in the form of air quality standards and associated compliance criteria (*see* **Standards, environmental**). The most used standards are based on the daily 1-h maximum among the stations of a given area, or on an 8-h maximum. Compliance criteria are based on the number of exceedances of this measure over a fixed threshold in a period (*see* **Exceedance over threshold**). Standards evolve differently in different countries as a function of exposure conditions and the socioeconomic situation. In the last few decades the phenomenon of ozone pollution has been analyzed extensively. Researchers proposed, besides chemical and meteorological deterministic models, statistical models for the analysis of ozone data. Statistical analysis of ozone data is motivated by the need to summarize large amounts of data collected in time and space, forecast, discover trends and evaluate the effects of ozone pollution on health.

Pollution due to ozone is analyzed not only as a metropolitan scale phenomenon, but also on regional and global scales. Analyses have been performed on numerous datasets. The Chicago area dataset is one of the most intensively analyzed [6], together with the Houston area dataset [7]. The meteorological conditions of the two areas are very different. Many other areas, not only in the US, have been studied, although not with the same intensity. The spatial and temporal synthesis of ozone data is a choice that influences subsequent analysis. The most frequent temporal synthesis is a summary like the mean, maximum, and/or median of hourly averages and number of exceedances over a time interval. The time interval considered depends on the aim of the analysis. Daily syntheses are frequently used. When the object of investigation is a trend, longer intervals, like weeks, months and years, are used. Syntheses on intervals shorter than one day are typical of short-term forecasting. Analyses are performed mainly for ozone season, running between April and October. In order to avoid complications due to the presence of a spring nonanthropogenic peak, the period considered can be also June–September.

Various probability models have been proposed for fitting air quality data: the two- and three-parameter **lognormal distribution**, the two- and three-parameter gamma distribution, and the **Weibull distribution**. In particular, Holland and Fitz-Simons [13] find that the best fit for ozone is the three-parameter lognormal distribution. Important issues include modeling distributions of maxima

over fixed time periods, or modeling exceedances over critical thresholds. In fact, extreme values are useful for assessing the impact of high air pollution. Moreover, air quality standards are formulated in terms of the highest level of permitted emissions. The traditional theory of extreme values is widely used in such contexts (*see* **Extreme value analysis**). Smith [23] illustrates the classical extreme value theory and emphasizes the point-process viewpoint of high-level exceedances. This approach includes the main methods for modeling extreme values as special cases, and suggests a general strategy to handle the complicated features of real data due to short-range correlation, seasonal variation, and the need to test for long-term trend. The theory is used for trend detection in ground-level ozone in the Houston area, without taking meteorology into account. No trend in the overall ozone levels is found, while a marked downward trend is detected in the extreme values. Also, a nonhomogeneous **Poisson process** model has been proposed for extreme values, where all the parameters of the exceedance model are allowed to depend on meteorological variables. A study by Smith and Shively [24] indicates clearly a decrease in ozone exceedances in the Houston area when meteorological effects are removed. This article, by means of a two-dimensional nonhomogeneous Poisson process, models jointly the continuous size of exceedances given that an exceedance has occurred, and the discrete frequency of exceedances over a threshold level as functions of time and meteorological covariates. When ozone data are collected in a spatial network, techniques can be used that either model explicitly the spatial data structure or model univariate spatial summaries of the network. These can be summaries such as means, medians, maxima or results of reducing dimensionality techniques looking for a unique value to be modeled. One part of the studies on ozone pollution models the relationship among ozone values in time and space. The most recent studies tend to include also information on meteorology. The reason is due to both the search for more accurate predictive models and the need to take **confounding** effects into account when investigating ozone trends and effects on health. Some studies investigate spatio-temporal variability on a planetary scale, trying to detect global trends. In an early work, Bloomfield et al. [1] studied global trends in total ozone by means of a frequency domain approach and by

considering spatial and temporal variability, fitting a model via **maximum likelihood**. They found no strong evidence of a trend in global ozone in the 1970s. Niu and Tiao [17] used data from the total ozone mapping spectrometer (TOMS) aboard the NUMBUS-7 satellite (*see Remote sensing*), which has been collecting data since 1978. They reduced the data to monthly averages for the period 1979–1989, and took advantage of the regular spatial grid of sample points to model the average ozone observation at each latitude and longitude. The model is additive, with a component for the time and a component for the seasonal cycle. The error term is modeled as a space–time autoregressive moving average (STARMA) process (*see Time series, periodic*). The results show negligible trends in equatorial latitudes, but with increasingly negative and statistically significant trends in total ozone depletion moving towards the poles.

Other studies concern the regional scale, mainly in order to investigate whether, in rural areas, ozone pollution can damage vegetation. Logan [16] studied the spatial coherence by means of cross-correlation coefficients, Lefohn et al. [15] used **kriging**, Cox and Clark [5] used **factor analysis** for the spatial relationships between urban sites. Eder et al. [9] used rotated principal components analysis for ozone data collected both in space and time to find situations with homogeneous characteristics, making a data reduction by means of independent transformed variables. A rotation of the original components is performed in order to better identify similar areas. The subregions were found to correspond well with the path and frequency of anticyclones. Guttorp et al. [12] developed models for the space–time correlation structure that enable us to spatially interpolate data in a moderately homogeneous region. They first proposed an AR(2) model for a temporal prewhitening of the **time series** at multiple monitoring sites. Spatial and space–time covariances between the prewhitened series at different sites were then computed. Standard **variogram estimation** techniques were applied in a new space obtained by transforming the geographical coordinates, using multidimensional scaling, into coordinates defined in a space where distances correspond to spatial dispersion and the correlation structure is isotropic and homogeneous. Carroll et al. [3] proposed a model where, for the first time, meteorology and spatio-temporal dependence of ozone data are considered together. They modeled the square root

of ozone concentrations as a sum of a deterministic trend depending on time and meteorology, and of a stochastic component modeled as a Gaussian random field (*see Random field, Gaussian*). The predictions for ozone obtained from the model were used in conjunction with population counts obtained from the census to construct exposure indices.

Some proposals to model with high-frequency data for short-term forecasting are based on autoregressive moving average (ARIMA) models (*see Time series*). Since the goal of such analyses is mainly to predict peaks, ARIMA modeling was revealed to be not very effective without concomitant variables: this is a common problem in forecasting high pollution levels. Simpson and Layton [22] explored dynamic regression noise models, where ozone data of another site is used for prediction, exploiting the idea that peaks at different locations are highly correlated. This kind of result can be useful when only one monitoring station needs to operate full time. Robeson and Steyn [19] compared different statistical models for forecasting daily maximum ozone concentrations: a univariate deterministic–**stochastic model**, a univariate ARIMA model, and a bivariate temperature and persistence-based regression model. The ARIMA model showed nearly the same predictive performance as the regression model, while the third model did not perform well. Graf-Jaccottet [11] analyzed ozone mean daily concentration considering as a covariate the number of hours between sunrise and sunset in a day. Instead of giving a parametric form to the error distribution, she used a model that transforms both sides of the regression equation via a **Box–Cox transformation**. The model is constructed with an AR(1) structure for errors; environmental factors are not introduced explicitly. This is useful for prediction, since often meteorological conditions cannot be controlled. Estimation of long-term trend is important for assessing the effectiveness of environmental policy. In doing so, it is important to stress that variations in time of meteorology have an important effect on ozone behavior. Situations might be detected where the introduction of guidelines has reduced the trend of ozone concentration, with exceptions due to peculiar meteorological conditions. For this reason, many proposals exist to remove meteorological effects on the observed ozone trend. Trends in ozone concentrations are usually estimated by fitting models to the distribution of ozone levels, with meteorological variables as independent variables.

Temperature, wind speed and direction are among the most commonly used variables, together with mixing height, pressure and opaque cloud cover. Both surface and upper air measurements are included in the models. Two main approaches have been proposed for evaluating the effect of meteorology on ozone: one is based on **classification** methods, used to isolate days with the same meteorological conditions, and the other on regression methods. **Linear models** have been found to be inadequate for modeling the complex relationship between ozone and meteorology. Cox and Chu [4] used in fact **generalized linear models** to describe the relationship between surface ozone concentration measurements and meteorological variables in 43 urban areas in the US, obtaining adjusted trend estimates for those areas. They modeled the probability of observing daily maxima of ozone according to a Weibull model, in which the scale parameter depends on a linear combination of daily values of meteorological variables and of an annual trend. Bloomfield et al. [2], for the Chicago area data, related the median value among the daily maxima of ozone to a nonlinear combination of meteorological variables, seasonality and yearly trend, specifying a parametric model that produces nearly normal errors. The dataset is divided into subsets, in order to test the model consistency via **cross-validation**. Gao et al. [10], using the same data, modeled the dependence of ozone concentration on meteorology and day in a nonparametric way. Their model contains also an additive parametric constant for the year effect. Davis et al. [8] compared a single-stage clustering technique, based on an average linkage, with a two-stage clustering technique, based on an average linkage in conjunction with k means, in order to find a classification scheme to investigate ozone dependence on meteorology in the Houston area. Prior to clustering, **collinearity** among meteorological variables is eliminated using a singular value decomposition to prevent variables that are highly correlated from influencing the results. The two-stage approach provides a better segregation of ozone concentration compared with the single-stage approach. Seven distinct meteorological regimes are identified in which daily 1-h maximum concentrations are significantly different. **Generalized additive models** are then used in each cluster in order to identify the meteorological variables most closely associated with ozone concentrations. In almost all clusters, the daily maximum surface temperature,

the daily mean v component of surface wind and the total daily global radiation are most important. Although the dependence of ozone on meteorology seems to be unique within individual clusters, cluster-specific regression models fit the data better than models fitted using the whole dataset. A nonlinear treatment by means of a regression tree was developed by Huang and Smith [14]. After selecting the daily maxima of the Chicago area network, a tree is grown and pruned using meteorological variables as covariates, until identifying clusters of homogeneous meteorological conditions at different ozone levels. The variation of the trend among clusters is interpreted by looking for the most suitable **analysis of variance** (ANOVA) model. In order to reduce the **standard errors**, an adjustment is performed by means of an empirical Bayes analysis (*see Empirical Bayes methods*). Results enlighten a downward trend and show that the diminution is stronger at higher ozone levels. In [18], daily maximum 1- and 8-h averages from a network in western Washington and northwest Oregon are used to construct a univariate regional daily synthesis derived from a canonical covariance analysis. This analysis determines the dominant pattern of association between the ozone network and the meteorology and emissions spatial fields. The synthesis is then adjusted for both regional surface and mesoscale meteorology and, subsequently, for regional estimated emission rates. A transformation for each meteorological predictor is utilized to achieve an approximately linear relationship with each monitor response. Both daily 1- and 8-h maximum surface ozone regional syntheses display moderate increasing long-term trends after meteorological adjustment. Since ozone data are collected over time, they may be serially correlated, as happens when meteorological persistence occurs. The use of techniques that rely on the hypothesis of independent observations may be inappropriate, since the asymptotic standard errors of the parameters are underestimated. In such cases, corrective methods are employed or resampling schemes, such as **jackknife** or **bootstrap resampling** [2], are used to obtain meaningful confidence intervals. The problem of **missing data** is typical of ozone analysis. The most recent developments try to model explicitly this feature, as in [24] or [2]. Reynolds et al. [18] estimated the **spatial covariance** matrix across an ozone network, in the presence of missing data, using the **EM algorithm**.

In recent years, concern has increased about the effect of air pollution on human health. Measuring the effects of ozone on health is difficult due to the presence of confounding factors such as the effect of autocorrelation, the influence of long-term trends, and the possible existence of a threshold level of ozone below which there is no observable effect. Confounding is also due to weather and other forms of air pollution. Owing to the second confounding factor set, the effects of pollutants on health are in general considered jointly. Most studies explore the relationships between daily death counts, or admissions to hospital, and covariates representing the long-term trend, meteorology and air pollution. The simplest way of analyzing such relationships is the classical linear model [25] that considers logarithmic or square-root transformations of daily values (*see Logarithmic regression*). Another approach is provided by nonlinear models such as Poisson regression models [21, 26] (*see Categorical data*). A European project has proposed standardized procedures for data management and analysis of the effect of pollution on health. A *meta-analysis* [28] shows that exposure to ozone seems to be associated with daily hospital admissions due to chronic obstructive diseases and, more generally, all respiratory diseases.

References

- [1] Bloomfield, P., Oehlert, G., Thompson, M.L. & Zeger, S. (1983). A frequency domain analysis of trends in Dobson total ozone record, *Journal of Geophysical Research* **88**, 8512–8522.
- [2] Bloomfield, P., Royle, J.A., Steinberg, L.J. & Yang, Q. (1996). Accounting for meteorological effects in measuring urban ozone levels and trends, *Atmospheric Environment* **30**, 3067–3077.
- [3] Carroll, R.J., Chen, R., George, E.I., Li, T.H., Newton, H.J., Schmiediche, H. & Wuang, N. (1997). Ozone exposure and population density in Harris County, Texas (with discussion), *Journal of the American Statistical Association* **92**, 392–415.
- [4] Cox, W.M. & Chu, S. (1993). Meteorologically adjusted ozone trends in urban areas: a probabilistic approach, *Atmospheric Environment* **27B**, 425–434.
- [5] Cox, W.M. & Clark, J. (1981). Ambient ozone concentration patterns among Eastern US urban areas using factor analysis, *Journal of the Air Pollution Control Association* **31**, 762–766.
- [6] Davis, J.M., Eder, B.K. & Bloomfield, P. (1998). Modeling ozone in the Chicago urban area, in *Case Studies in Environmental Statistics*, Lecture Notes in Statistics, Vol. 132, D. Nychka, W.W. Piegorsch & L.H. Cox, eds, Springer-Verlag, New York, pp. 5–26.
- [7] Davis, J.M., Eder, B.K. & Bloomfield, P. (1998). Regional and temporal models for ozone along the Gulf Coast, in *Case Studies in Environmental Statistics*, Lecture Notes in Statistics, Vol. 132, D. Nychka, W.W. Piegorsch & L.H. Cox, eds, Springer-Verlag, New York, pp. 27–50.
- [8] Davis, J.M., Eder, B.K., Nychka, D. & Yang, Q. (1998). Modeling the effects of meteorology on ozone in Houston using cluster analysis and generalized additive models, *Atmospheric Environment* **32**, 2505–2520.
- [9] Eder, B.K., Davis, J.M. & Bloomfield, P. (1993). A characterization of the spatiotemporal variability of non-urban ozone concentrations over the eastern United States, *Atmospheric Environment* **27A**, 2645–2668.
- [10] Gao, F., Sacks, J. & Welch, W.J. (1996). Predicting urban ozone levels and trends with semiparametric modeling, *Journal of Agricultural, Biological, and Environmental Statistics* **1**, 404–425.
- [11] Graf-Jaccottet, M. (1993). A flexible model for ground ozone concentration, *Environmetrics* **4**, 23–37.
- [12] Guttorp, P., Meiring, W. & Sampson, P.D. (1994). A space–time analysis of ground-level ozone, *Environmetrics* **5**, 241–254.
- [13] Holland, D.M. & Fitz-Simons, T. (1982). Fitting statistical distributions to air quality data by the maximum likelihood method, *Atmospheric Environment* **16**, 1071–1076.
- [14] Huang, L. & Smith, R.L. (1999). Meteorologically-dependent trend in urban ozone, *Environmetrics* **10**, 103–118.
- [15] Lefohn, A.S., Knudsen, H.P. & McEvoy, L.R. Jr (1988). The use of kriging to estimate monthly ozone exposure parameters for the southeastern United States, *Environmental Pollution* **53**, 189–224.
- [16] Logan, J.A. (1989). Ozone in rural areas of the United States, *Journal of Geophysical Research* **94**, 8611–8632.
- [17] Niu, X. & Tiao, G.C. (1995). Modelling satellite ozone data, *Journal of the American Statistical Association* **90**, 969–983.
- [18] Reynolds, J.H., Das, B., Sampson, P.D. & Guttorp, P. (1998). Meteorological adjustment of Western Washington and Northwestern Oregon surface ozone observations with investigation of trend. NRCSE Technical Report Series, no. 15.
- [19] Robeson, S.M. & Steyn, D.G. (1990). Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations, *Atmospheric Environment* **24B**, 303–312.
- [20] Rushton, L. (1997). Overview and critique of the air pollution and health: a European approach (APHEA) project, Concawe, Report no. 99/54.
- [21] Schwartz, J. (1993). Air pollution and daily mortality in Birmingham, Alabama, *American Journal of Epidemiology* **137**, 1136–1147.
- [22] Simpson, R.W. & Layton, A.P. (1983). Forecasting peak ozone levels, *Atmospheric Environment* **17**, 1649–1654.

6 Ozone

- [23] Smith, R.L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone, *Statistical Science* **4**, 367–383.
- [24] Smith, R.L. & Shively, T.S. (1995). Point process approach to modelling trends in tropospheric ozone based on exceedances of a high threshold, *Atmospheric Environment* **29**, 3489–3499.
- [25] Smith, R.L., Davis, J.M. & Speckman, P. (1999). Human health effects of environmental pollution in the atmosphere, in *Statistics for the Environment*, Statistical Aspects of Health and the Environment, Vol. 4, V. Barnett, K.F. Turkman & A. Stein, eds, Wiley, Chichester, pp. 91–115.
- [26] Zidek, J.V. (1997). Interpolating air pollution for health impact assessment, in *Statistics for the Environment*, Pollution Assessment and Control, Vol. 3, V. Barnett, K.F. Turkman & A. Stein, eds, Wiley, Chichester, pp. 251–268.

(See also **Atmospheric Dispersion; Chemistry; Global circulation; Global environmental change; Global warming; Simulation-optimization modeling**)

DANIELA COCCHI & CARLO TRIVISANO