P1.3

MONITORING CLIMATE VARIABILITY IN RELATION TO AIR QUALITY: A REGIONAL TEMPERATURE-BASED INDEX FOR GROUND-LEVEL OZONE

Daniel Y. Graybeal and Arthur T. DeGaetano
Northeast Regional Climate Center, Cornell University, Ithaca, New York

1. INTRODUCTION

High levels of tropospheric ozone (a.k.a. ground-level ozone, GLO hereafter) are well known to pose a risk to human health. The GLO concentration at any given location is a product of two factors: the availability of constituents needed for its local photochemical production or destruction, and the stagnant or transient meteorological conditions conducive to its local buildup or transport elsewhere. GLO results from a photochemical reaction in the presence of nitrous oxides and volatile hydrocarbons. Thus, its production is positively correlated with solar radiation amount and duration, and the reaction rate increases with temperature. As emissions from power plants and internal combustion engines contribute to GLO formation, high concentrations typically originate and often remain in urban areas, though not necessarily, owing to mechanisms of transport and boundary layer decoupling (e.g., Solomon et al. 2000).

Much recent work has focused on identifying the relationships between meteorological factors and high GLO days, so that the effects may be removed to study secular trends in ozone over the decadal scale (Smith et al. 2002). However, little research (Cox and Chu 1996) has focused on the long-term variability of meteorological conditions conducive to excessive GLO, particularly on the multi-decadal scale. As interest in climate monitoring and seasonal forecasting continues to grow, links of these conditions to large-scale ocean-atmospheric circulation patterns and teleconnections would be valuable. In support of its applied monitoring goals, the National Climate Data Center has posted three long-term climate impact indices (Karl et al. 1996): the Residential Energy Demand Temperature Index (REDTI), the Crop Moisture Stress Index (CMSI), and the Air Stagnation Index (ASI) (NCDC 2003). ASI is available for the period beginning 1973, while the REDTI and CMSI extend back to 1895 and 1900, respectively. All three indices are nationwide in scope.

Meteorological conditions responsible for high GLO exposure at a place on a given day are a combination of regional and local components. A warm-season, slow-moving anticyclone provides the high insolation and temperatures and limited mixing that allows for buildup, while the changing wind field to its fore and aft influence local advection of GLO among adjacent urban areas (Aneja et al. 1999). To produce an index of conditions favorable for high GLO production over a spatial scale larger than a medium-sized city and over a multidecadal time scale requires an upscaling approach. This study considers meteorological components integrated over the temporal scale of the ozone season, as well as the fairly climatically homogeneous spatial scale of the region or subregion. In snow hydrology, for example, the roughly analogous upscaling is between hourly energy-balance modeling at a point and daily temperature index modeling of snow melt over an area; while the daily model is usually not able to capture extreme melt variability like the hourly model, it is usually quite capable of modeling snow processes over the broader spatial and temporal scales, as temperature is integrative of most important processes at those scales (Rango and Martinec 1995).

2. DATA AND METHODS

2.1 *Meteorological Data*

The only meteorological element known to exhibit a strong relationship with ozone concentration and for which the available record extends back at least 100 years is temperature. Over this time span, minima and maxima at monthly and daily resolution are available from the USHCN network (Karl et al. 1990) for more than 1000 stations, and monthly means are available for 343 U.S. Climate Divisions (USCDs). Both the REDTI and CMSI are based on USCD data, which are equally weighted averages of conditions over a broad area, as opposed to sampling just rural areas, as the USHCN does. High GLO is usually distributed spatially near cities or downwind from them, so the greater spatial information present in the USCD average is helpful, for the purpose of developing a climatological index to potential GLO occurrence in metropolitan areas taken together in a region. Thus, USCD monthly mean temperatures were obtained from NCDC (http://www.ncdc.noaa.gov) for the period 1895 through 2002 and for 23 divisions in the Northeast U.S. classified as metropolitan.

Classification of USCDs as metropolitan was based on land area and 2000 Census population, both available from NCDC. Simple univariate stem-leaf plots were constructed for each of the variables total population and population density, and the largest breaks were noted, suggesting samples from different populations. For total population, the largest break occurred at about 1.7 million persons, whereas for population density, a less prominent but probably more meaningful break of around 92.7 persons km^{-2} was

---

**Corresponding author address*: Dr. Daniel Y. Graybeal, 1123 Bradfield Hall, Cornell University, Ithaca, NY 14853; e-mail: dyg2@cornell.edu.

Fig. 1. Metropolitan USCDs, with the Northeast region highlighted.

chosen. That threshold allowed the North Carolina Piedmont to be included, for example. An intersection of the two criteria proved too restrictive (excluding St. Louis, for example), so the union was employed. Fifty-seven USCDs (Fig. 1) ranking above either of the thresholds were classified as metropolitan, containing nearly 63% of the total 2000 U.S. population. For comparison, the top 50 metropolitan areas in the 1990 Census contained 57% of the total U.S. population.

2.2 *Ozone Data*

Ozone data were obtained from the AIRS data base compiled by the U.S. Environmental Protection Agency (EPA) and made available on-line by the Center for Air pollution Impact and Trend Analysis (CAPITA) at Washington University in St. Louis, Missouri (http://capita.wustl.edu). Data were available from ten seasons (1986 through 1995, defined as 15 April through 15 September) and at hourly resolution, from 1500 stations around the U.S. and more than 400 in the Northeast. Hourly concentrations were transformed into annual counts of days exceeding the U.S. EPA threshold of 80 ppbv daily maximum, eight-hour average, GLO concentration. To bridge the gap in scale between the GLO monitoring station and the metropolitan USCD, an annual count for a division was incremented if any GLO station within 0.5° longitude or latitude of the USCD geographic center reported a daily maximum, eight-hour average concentration over the threshold. The size of the geographic window was adjusted to approximate the dimensions of the USCD, if significantly larger than 0.5° on a side. In the Northeast, a median of 6 GLO stations reported hourly in each search box (range 1 to 15). All ozone data were passed through temporal and spatial quality assurance; details are available from the authors.

2.3 *Index development*

For each metropolitan USCD, relationships between annual ozone exceedance counts and temperature were explored through correlation and simple linear regression, following the development of the REDTI and CMSI. Prior to including temperature in the analysis, the ozone counts time series were examined for evidence of a trend; a decrease from 1986 to 1995 was anticipated, due to the effects of the 1990 Clean Air Act and as noted in other studies (e.g., Smith et al. 2002). While previous studies have removed the effects of year-to-year weather variability to highlight more clearly the true ozone trends, the goal in this paper is to remove the effects of regulation over the short time span and isolate the variation due to weather and climate. Trend detection was affected by the outlying high ozone counts in 1988, so a resistant regression of counts against year was fit, in addition to an ordinary least squares (OLS) fit. Many resistant regressions are available; the method chosen estimates the median, rather than the mean, and minimizes the least absolute value of residuals. Regressions were fit using the STATA software. Ozone time series were OLS-detrended only if the slopes of both regressions were more than one standard error away from zero and at least one of them was significant at the 90% level. Only three time series were detrended.

Following detrending, 21 aggregate monthly mean temperatures were obtained for each ozone season (April through September of a given year). The means for each month individually, along with the composite for April-May, May-June, etc., on up to the six-month composite were calculated. The aggregate having the best Pearson correlation with GLO exceedance days was selected as most representative of mean temperature for the ozone season. As a double-check, the counts of exceedance days were summed by month, and the months with the highest counts were compared with the months found best-correlated. As the focus of this index is gauging potential for high GLO exposure, the next step in creating the index was to weight the standardized temperature-based predictions of seasonal counts of high-GLO days by the ratio of the population in the individual USCD to that over all metropolitan Northeast USCDs. A population-weighted sum was then obtained for the Northeast region as a whole (for its metropolitan USCDs), and this sum was rescaled 0 to 100 based on the historic range. This procedure is akin to those used in developing the REDTI and CMSI. Operationally, this index may be updated and rescaled each GLO season.

3. RESULTS WITH DISCUSSION

3.1 *General Results for the Northeast U.S.*

By Northeast USCD, Pearson correlation between seasonal counts of high-GLO days and best aggregate mean temperature exhibited a median (mean) of 0.776 (0.747), ranging from 0.541 to 0.952. Median (mean) explained variance is thus 60.2% (57.0%). These correlations are similar in magnitude to those supporting the REDTI and CMSI. The middle 50% of the rescaled, jackknifed (Studentized) residuals are within five days of the observed seasonal exceed-

Table 1. Statistics summarizing temperature-ozone exceedance count relationships found over the Northeast U.S. region; the last column is Studentized residuals (jackknifed and standardized).

Stat.	r	r^2 (%)	Adj. r^2 (%)	Resid.	Student.
Min.	0.541	29.3	20.5	–21.0	–29.5
25%	0.668	44.6	37.7	–4.8	–4.8
Median	0.776	60.2	55.3	–0.2	–0.1
75%	0.830	68.8	64.9	3.8	4.0
Max.	0.952	90.5	89.3	20.7	34.1
Mean	0.747	57.0	51.6	0.0	0.4
S. Dev.	0.115	16.8	18.9	6.2	7.9

ance counts; the range is within about a month's worth (Table 1). For nine of the 23 USCDs, best correlation with temperature was found using the May-August aggregate, while for another nine USCDs, either July or August or both of those months in aggregate were best correlated with ozone exceedance day counts. Of exceedance days in the Northeast in the 10-year period, 31.7% fell in July, 26.2% in June, and 23.0% in August. The weighted mean month of the temperature relationships (e.g., a May-August aggregate scores 6.5, a July-August aggregate 7.5) was 6.89, a (late) bias of +0.14 months relative to the weighted mean month based on the exceedance day counts themselves.

The index developed here (Fig. 2) exhibits very good agreement with a similar, annual index to the potential for high GLO exposure in the Northeast U.S., based on daily climatology and probabilistic modeling using an extreme-value distribution (Cox and Chu 1996). As their index is based on relative rankings, the present index was ranked accordingly, and a Pearson correlation measure between the indices of 0.646 was obtained (coefficient of determination 41.7%), increasing to 0.732 for the top 20 and with virtually no relationship in the bottom half of the paired rankings. Cox and Chu (1996) rank 1988 highest, whereas 1955 ranks highest in this index. Years in the remainder of the top five in their series are 1953, 1955 and 1983 at rank 3.5, and 1991 and 1993 at rank 5.5. In this study, ranks second through fifth belong to 1991, 1988, 1959, and 1993, in descending order. This study ranks 1983 and 1953 ninth and eleventh, whereas 1959 is ranked 15th in Cox and Chu (1996). They base their index on a threshold, the seasonal 99th percentile, daily maximum, one-hour GLO concentration, as opposed to days above a daily maximum, eight-hour threshold.

3.2 *Applying a Spatial Synoptic Classification*

Several approaches were undertaken to explore practical ways in which the index may be improved and expanded to other regions in the U.S.

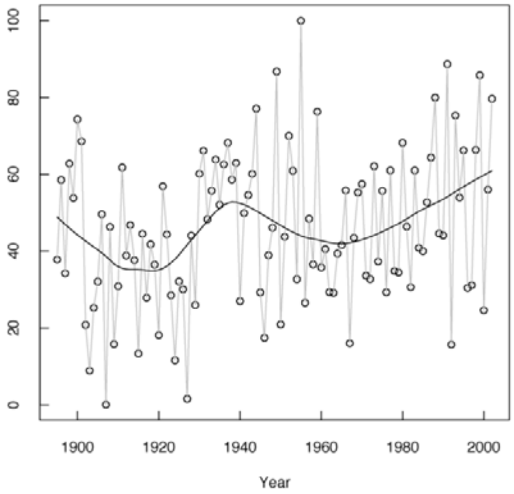

Fig. 2. Population-weighted, climatological index to high GLO exposure potential, for metropolitan Northeast USCDs (with *f* =0.3 lowess smooth).

The Spatial Synoptic Classification (SSC2) of Sheridan (2002) is hybrid scheme beginning with expert manual identification, using standard surface synoptic charts, of "seed days" typifying six air mass types commonly found across North America. These seed days are specified for each surface reporting station but allow geographic variability in the mean and range of the typical elements. The procedure continues by discriminant analysis to classify the remaining days in the station's period of record into one of the six main types or a transitional type. The multivariate treatment of weather elements offers the advantage of simplifying interpretation while accounting for synergistic interrelationships among weather elements. GLO production is known to result from a combination of several related elements that can be tracked efficiently in this manner.

Daily SSC2 calendars were obtained for the Philadelphia airport station for a case study of conditions in years with prominent local residuals in exceedance counts from the seasonal aggregate mean temperature relationship. Of the six main SSC2 types regularly occurring at Philadelphia, Moist Moderate (MM) most favorably balances strong correlation with exceedance counts but weak correlation with mean temperature, potentially providing valuable independent information. Summer frequency of cool, cloudy MM air is negatively correlated overall (–0.457) with the seasonal GLO exceedance count.

3.3 *Counts of Days above 32°C*

A simple approach that focuses on the more extreme conditions conducive to production of high GLO levels is a count of a season's days above 32°C. Counts were obtained overall, on days not recording precipitation, and for consecutive-day (temperature) events, again at Philadelphia. Maximum degree-day sums (base 32°C) were also obtained, for all consecutive-day events. Mean temperature bested all of these in Pearson correlation with annual counts of exceedance days at Philadelphia. Correlation measures were 0.759, 0.687, 0.628, and 0.552, for mean temperature, days above 32°C with no precipitation, days above 32°C, and maximum degree-day sum, respectively.

3.4 *Monthly Mean Diurnal Temperature Range*

This variable (DTR hereafter) was estimated by subtracting monthly mean minimum from monthly mean maximum temperatures available from the USHCN, also for the Philadelphia-area USCD. At this USCD, July-August composite mean temperature was most highly correlated with exceedance counts. Whereas July DTR was highly correlated with both mean temperature and exceedance counts, August DTR correlated well only with exceedance counts (Pearson measure 0.598), potentially providing independent information. This positive correlation supports the link between DTR and less cloud cover and higher daytime radiation receipts, elements in addition to temperature that are important to GLO formation. The importance of this element in August, contrasted with the July-August aggregate period, may be due to advection becoming more important, relative to radiation, in determining local mean temperatures.

3.5 *Monthly Incoming Solar Radiation*

Finally, monthly total insolation was modeled (DeGaetano et al. 1997), using representative surface airways data. At New York and Philadelphia, July-August insolation and exceedance counts were negatively correlated, and to a lesser absolute extent than temperature (Pearson measure of –0.359, compared to +0.689 for temperature at New York's USCD). Counts in 1988 largely influence this result, but for a different reason than expected. That season was characterized by relatively frequent frontal overrunning while also exhibiting relatively frequent GLO episodes. For May-August, insolation and temperature were correlated to a Pearson (Spearman) measure of 0.374 (0.541), suggesting insolation affords only limited additional information. This conclusion is further supported by Houston's results, where insolation and exceedance counts are virtually uncorrelated in August-September (time of maximum insolation correlation).

4. OTHER U.S. REGIONS AND CONCLUSIONS

Construction of climatological indices of the potential for GLO exposure was attempted for two other regions, Texas and California (Fig. 1). Temperature-exceedance count relationships were not as strong as in the Northeast (mean Pearson correlation of 0.555 in Texas and 0.538 in California), likely owing to bimodal within-season ozone dynamics and to climatic and physiographic heterogeneity, respectively. In Texas, exceedance days peak in frequency in August-September and secondarily in May, but the best temperature relationship was generally found for August. In California, unimodal exceedance frequency peaks in July-September but the best temperature regressions are probably not related; they are from May. These results suggest seasonal aggregate temperature may best model seasonal exceedance counts in regions where it best serves as a proxy for the truly active climatic factors. Moving from regional to national indexing may require incorporation of other factors and time frames, such as insolation or vertical structure, summarized over days exhibiting an ozone exceedance.

5. REFERENCES

Aneja, V. P., R. G. Oommen, A. J. Riordan, S. P. Arya, R. J. Wayland, and G. C. Murray, 1999: Ozone patterns for three metropolitan statistical areas in North Carolina, USA. *Atmos. Environ.*, **33**, 5081-5093.

Cox, W. M. and S.-H. Chu, 1996: Assessment of interannual ozone variation in urban areas from a climatological perspective. *Atmos. Environ.*, **30**, 2615-2625.

DeGaetano, A. T., K. L. Eggleston, and W. W. Knapp, 1997: Redevelopment of a solar radiation estimation model based on ASOS and satellite-derived cloudiness data. Preprints, *10th Conf. Appl. Climatol.*, Reno, NV, Amer. Meteor. Soc., 167–170.

Karl, T. R., C. N. Williams, Jr., F. T. Quinlan, and T. A. Boden, 1990: *United States Historical Climatology Network (HCN) Serial Temperature and Precipitation Data*. Environmental Science Division, Pub. No. 3404, Carbon Dioxide Information and Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN, 389 pp.

Karl, T. R., R. W. Knight, D. R. Easterling, and R. G. Quayle, 1996: Indices of climate change for the United States. *Bull. Amer. Meteor. Soc.*, **77**, 279-292.

National Climatic Data Center, 2003: *National Climate Impact Indicators*. [Available on-line at http://lwf.ncdc.noaa.gov/oa/climate/research/cie/cie.html. Accessed 29 October 2003.]

Rango, A. and J. Martinec, 1995: Revisiting the degree-day method for snowmelt computations. *Water Resour. Bull.*, **31**, 657–669.

Sheridan, S. C., 2002: The redevelopment of a weather-type classification scheme for North America. *Int. J. Climatol.*, **22**, 51–68.

Smith, M., P. Yau, T. Shivley, and R. Kohn, 2002: Estimating long-term trends in tropospheric ozone levels. *Int. Stat. Rev.*, **70**, 99–124.

Solomon, P., E. Cowling, G. Hidy, and C. Furiness, 2000: Comparison of scientific findings from major ozone field studies in North America and Europe. *Atmos. Environ.*, **34**, 1885–1920.