Using Ordinary Kriging to Estimate the April - September 24-Hour W126 and N100 Ozone Exposure Metrics for 2010 for the United States

by

Allen S. Lefohn, Ph.D.
A.S.L. & Associates
302 North Last Chance Gulch
Suite 410
Helena, Montana 59601

H. Peter Knudsen, Ph.D.
Professor
Montana Tech of the
University of Montana
Butte, Montana 59707

Douglas S. Shadwick
320 Eastwood Road
Chapel Hill, North Carolina 27514

September 21, 2011

1. Introduction

This report describes how Ordinary Kriging was used to estimate the 6-month W126 and N100 24-h ozone exposure metrics for 2010 for the United States. A.S.L. & Associates has published previously its kriging results in peer-review papers and reports (e.g., Knudsen and Lefohn, 1988; Lefohn *et al*., 1988; Lefohn *et al*., 1992; Lefohn *et al*., 1997).

Kriging is a method of interpolation which predicts unknown values from data observed at known locations. This method uses variograms to express the spatial variation, and it minimizes the error of predicted values, which are estimated by spatial distribution of the predicted values.

Kriging is a family of estimators used to interpolate spatial data. This family includes Ordinary Kriging, Universal Kriging, Indicator Kriging, Co-kriging, and others. The choice of which method of kriging to use depends on the characteristics of the data and the type of spatial model desired. The most commonly used method is Ordinary Kriging, which was selected for this study.

Indicator kriging is used when it is desired to estimate a distribution of values within an area rather than just the mean value of an area. As the purpose of the study was to estimate the mean values of the N100 and W126 exposure indices within an area rather than the distribution of values, Indicator Kriging was not selected.

Universal Kriging was not selected for this study for several reasons. Universal kriging is used to estimate spatial means when the data have a strong trend and the trend can be modeled by simple functions. Ozone data may display trends over small geographic areas but at the scale of the entire United States, there is no trend that can be modeled by simple functions. Because the data display no trend at the scale of the modeling, Universal Kriging was not appropriate.

Co-kriging is an extension of kriging used when estimating one variable from two variables. The two co-variables must have a strong relationship and this relationship must be defined. Use of Co-Kriging requires the spatial covariance model of each variable and the cross-covariance model of the variables. The method can be quite difficult to do because developing the cross-covariance model is quite complicated. Developing the relationship between the two variables can also be complicated. Practice in the mining industry limits Co-Kriging to the case when the variable being estimated is under sampled with respect to the second variable. If all samples have both variables, industry has found no benefit gained from the use of Co-Kriging. Co-Kriging was not selected for this study because the O_3 indices N100 and W126 are sampled at each location. Also, there has been no study that has identified a secondary variable from more sampling sites that is highly correlated to these exposure indices that can be used to predict these indices.

Ordinary Kriging was selected for this study based on how well it has performed using prior years' data. The covariance models (i.e., variogram) exhibit local stationary and thus Ordinary Kriging was the appropriate method to use.

The authors have used Ordinary Kriging to make surface-level ozone models for the W126 exposure index for the years from 1982 through 2010. While the ozone values vary from year to year, the statistical character of the data remains remarkably constant from year to year. The covariance models are similar in each year and the spatial anisotropy exhibited by the co-variance models is similar in each year.

2. Approach

A.S.L. & Associates provided the 2010 ozone hourly data to Mr. Douglas Shadwick for characterizing the 24-hour W126 and N100 monthly values and then summarizing the information into 6-month (April – September) values. In addition, the second highest daily maximum concentration and the 4th highest 8-hour daily maximum average concentration were calculated for the same time period. The data capture was determined based on the April – September period. Following receipt of the characterized data from Mr. Shadwick and checking of the results, A.S.L. & Associates provided Dr. Knudsen with the April – September (6 month) W126 and N100 24-h exposure indices for monitoring sites that met the data capture criteria in 2010. The computer files provided contained summarized air quality data, monitoring site identification codes, site latitude and longitude, and site characterization code information (i.e., urban, suburban, rural, etc.).

Mr. Shadwick corrected the characterized data for missing values. The estimate of cumulative indices from hourly average data (e.g., the W126 cumulative and N100 ozone indices) will be biased low if a part of the hourly average data is missing. A correction scheme has been adopted to estimate, in particular, a cumulative index for seasonal values of the indices. The correction scheme has two components.

1. A monthly value of each index is calculated. If at least 75% of the hourly data are available for the month, a corrected monthly cumulative index is calculated as the uncorrected monthly cumulative index divided by the data capture (as a fraction).

2. If there are any months with less that 75% data capture and the two chronologically adjacent months each have at least 75% data capture, then a corrected monthly cumulative index for the month with less than 75% data capture is calculated as the arithmetic average of the corrected monthly cumulative indices for the two adjacent months.

If all of the months contained within a season have valid estimates (in the sense described above) of the corrected monthly cumulative index, the corrected seasonal cumulative index is calculated as the sum of the corrected monthly cumulative indices. Otherwise, a valid estimate of the corrected seasonal cumulative index is not reported.

In addition to the data provided to Dr. Knudsen, the second highest daily maximum 1-hour concentration and the fourth highest 8-hour average daily maximum concentration that occurred over the April – September period for each monitoring site that experienced sufficient data capture was provided by A.S.L. & Associates to the U.S. Forest Service project manager.

3. Scope of Work

Estimate the April - September W126 and N100 exposure index value for each 1/2^o by 1/2^o cell in the United States excluding Alaska and Hawaii.

Specific tasks performed included:

1. Check and verify the latitude, longitude and elevation of each site.
2. Calculate and model variograms for each exposure index values for each year.
3. Krig the W126 and N100 values for each year.
4. Prepare files that contain information describing the kriged values, the coordinates, variance, and the 95% error bound for each 1/2 ^{o} by 1/2^o cell.

The traditional kriging variance and the 95% Error bound were reported as in similar years. In addition, another estimate of the error variance, which is referred to as the Ordinary Kriging interpolation variance (OKIV), was determined (Yamamoto, 2005). Yamamoto derived an error variance for Ordinary kriging that is conditional to the data values. He has shown that the Ordinary interpolation variance is a better measure of accuracy of the kriging estimate. We believe that the method used by Yamamoto to determine the interpolation variance is a better estimate of the error variance than the kriging variance. In particular, for skewed data, it is believed that the new variance is a much better estimate of the error variance. The 95% error bound based on the new variance was reported also.

4. Steps in Modeling

Ozone exposure indices for California are known to be significantly different (i.e., higher) from other areas of the United States. Therefore California was analyzed separately in this study.

The following steps were performed for each exposure index.

1. Data Checking

   a. The latitude and longitude of all the monitoring sites from the AQS database were compared and updated to coordinates supplied by ASL & Associates. In addition, many of the AQS sites have incorrect elevations, missing elevations, or elevations listed in feet rather than meters. The elevations of all monitoring sites were checked and updated with elevations supplied by Bill Jackson, U.S. Forest Service, using a 90-meter digital elevation model.

b. Maps showing monitoring sites and data values were plotted.

2. Calculation of data statistics.

   a. Basic statistics histograms were calculated independently for California and for the remainder of the U.S.

   b. Histograms were prepared.

3. Variograms for each exposure index were calculated. Experience gained in prior studies of ozone indices was used in determining the parameters for calculating and modeling the variograms. Particular care was used to determine the presence and likely directions of anisotropy. After calculation of the variograms, a theoretical model was fitted to them.

4. Kriged values of W126 and N100 for each 1/2^{o} by 1/2^{o} degree cell were determined. The N100 and W126 exposure for each 1/2^{o} by 1/2^{o} cell in California and the remainder of US were estimated using Ordinary Kriging. The main input to a kriging program is the variogram parameters and the search parameters. The variogram parameters are listed in Tables 3 and 5. The search parameters are shown below.

Search Radius		California	US
Maximum search radius	=	750 km	900 km
Max. number of sites used to estimate a cell	=	12	15
Min. number of sites used to estimate a cell	=	1	1

5. A file with kriged values was prepared. An example is shown in Table 1.

Table 1. Example Output from Ordinary Kriging Program.

Latitude	Longitude	W126	Ordinary Kriging Var.	95%EB	OK Interpolation Variance	OKIV 95%EB	No. of Samples
44.5	-66.5	6.08	18.1	8.51	2.93	3.42	15
45	-66.5	5.16	11.81	6.87	2.64	3.25	15
44	-67	6.55	20.8	9.12	5.22	4.57	15
44.5	-67	6.28	11.36	6.74	2.3	3.03	15
45	-67	5.2	3.94	3.97	4.8	4.38	15

5. 2010 Data

5.1 N100 Exposure Index – Basic Statistics

Table 2 shows summary statistics for the N100 index. Table 3 shows the variogram parameters for N100.

Table 2. 2010 N100 Statistics.

AREA	MEAN	VARIANCE	STD.DEV.	MIN	MAX	NUMBER
California	9.28	384.	19.6	0.0	137.1	159
Rest of US	0.81	5.1	2.3	0.0	23.4	1028

Table 3. 2010 N100 Variogram Parameters.

AREA	Co	C1	Range1	C2	Range2	Angle	AF ratio
CAL	75	225	100	100	800	135	4
US	0.75	0.75	100	2.5	700	45	42.0

5.2 W126 Exposure Index – Basic Statistics

Table 4 shows summary statistics for the W126 index. Table 5 shows the variogram parameters for W126

Table 4. 2010 W126 Statistics

AREA	MEAN	VARIANCE	STD.DEV.	MIN	MAX	NUMBER
California	27.4	687	26.2	0.0	116.4	159
Rest of US	17.1	72	8.5	0.0	54.0	1028

Table 5. 2010 W126 Variogram Parameters.

AREA	Co	C1	Range1	C2	Range2	Angle	AF ratio
CAL	100	200	350	390	550	135	3.00
US	20	20	250	40	1800	90	1.5

The kriged results were prepared in two digital files with the following information: latitude, longitude, estimated kriged value (e.g., W126 or N100), kriged variance, 95% error bound, OKIV, N95% error bound, and number of samples used to determine the kriged estimate.

6. References

Heck, W.W., O.C. Taylor, R. M. Adams, G.E. Bingham, J.E. Miller, E.M. Preston and L.H.Weinstein. 1984. National Crop Loss Assessment Network (NCLAN) 1982 Annual Report. EPA-600/3-84-049. U.S. Environmental Protection Agency, Corvallis, Oregon. pp. 198-219.

Knudsen, H.P. and A.S. Lefohn. 1988. The Use of Spatial Statistics to Characterize Regional Ozone Exposures, pp 91-105. In: Assessment of Crop Loss from Air Pollutants. W.W. Heck, O.C. Taylor, D.T. Tingey (eds). Elsevier Applied Science Publishing, London, U.K.

Lefohn, A.S., H.P. Knudsen, and L.R. McEvoy, Jr.1988. The Use of Kriging to Estimate Monthly Ozone Exposure Parameters for the Southeastern United States. *Environmental Pollution*. 53:27-42.

Lefohn, A.S., H.P. Knudsen, D.S. Shadwick, and K.A. Hermann. 1992. Surface Ozone Exposures in the Eastern United States (1985-1989). In:Transactions of the Response of Southern Commercial Forests to Air Pollution Specialty Conference (R.B. Flagler, editor). Air & Waste Management Association, Pittsburgh, PA. pp. 81-93.

Lefohn, A.S.; Jackson, W.; Shadwick, D.S.; Knudsen, H.P. 1997. Effect of surface ozone exposures on vegetation grown in the Southern Appalachian Mountains: Identification of possible areas of concern. *Atmos. Environ*. 31: 695-1708.

Yamamoto, J.K. 2005. Comparing ordinary kriging interpolation variance and indicator kriging conditional variance for assessing uncertainties at unsampled locations, In:Application of Computers and Operations Research in the Mineral Industry – Dessureault, Ganguli, Kecojevic,& Dwyer editors, Balkema.