Using Ordinary Kriging to Estimate the Seasonal W126, and N100 24-h Concentrations for the Year 2000 and 2003

by

Allen S. Lefohn, Ph.D.
A.S.L. & Associates
111 North Last Chance Gulch
Suite 4A
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, NC 27514

October 25, 2005

1.0 Introduction

This report describes how Ordinary Kriging was used to estimate the 6-month W126 and N100 24-h ozone exposure values for the year 2000 and 2003 for the United States. A.S.L. & Associates has published previously its kriging results in peer-review papers and reports (Knudsen and Lefohn, 1988; Lefohn et al., 1988; Lefohn et al., 1992; Lefohn et al., 1997).

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 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.

A brief discussion follows on why ordinary kriging was chosen for this study rather than another method.

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 is used to estimate spatial means when the data have a strong trend and the trend can be modeled by simple functions. Trend is scale dependent. For example Montana Tech sits on the south side of a hill high above the valley Butte, Montana. A model of the elevations around Montana Tech would show that a trend in the values exists when you look north. If you want to model the elevation to the north of Montana Tech, it can be accurately done with a simple straight line. At the scale of 1/4 mile the local data has a trend. This trend doesn't exist for far, however. If you continue north for 60 miles, you encounter Helena, Montana. Along the way the elevation rises and falls many times as you cross mountains and valleys. On the scale of 60 miles, there is no trend in the elevation. 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 of this fact, Universal Kriging was not chosen for this study.

Co-kriging is an extension of kriging used when estimating a 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 chosen for this study because the ozone indices N100 and W126 are sampled at each location. Also, there has been no study has yet been done 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. Elevation may be a promising variable but there is not a sufficient number ozone monitors across a range of elevation to develop the covariance models.

Ordinary kriging was selected for this study based on how well it has performed on prior years data and because the statistical characteristics of the data in 2000 and 2003 make Ordinary Kriging the appropriate choice of estimator. The data displayed no trend at the scale of the modeling; thus universal kriging was not appropriate. The covariance models (variogram) exhibited local stationary and thus, Ordinary Kriging was the appropriate technique to use.

The authors have used ordinary kriging to make ground-level ozone models for the W126 for the years from 1982 to 2003. 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.

Ordinary Kriging is a spatial estimation method where the error variance is minimized. This error variance is called the kriging variance. It is based on the configuration of the data and on the variogram, hence is is homoescedastic (Yamamoto, 2005). It is not dependent on the data used to make the estimate. Recently, Yamamoto derived a error variance for ordinary kriging that is conditional to the data values. He referred to this variance as the Ordinary Interpolation variance. Yamamoto has shown that the ordinary interpolation variance is a better measure of accuracy of the kriging estimate. The ordinary kriging programs used for this study were modified to calculate the new error variance, named the Ordinary kriging interpolation variance (NKVAR) and output it along with the traditional kriging variance. The 95% error bound based on the new variance was reported also. It is believed that the new method used in this study 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.

1.1 Approach

A.S.L. & Associates provided the 2000 and 2003 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, data capture, the second highest daily maximum concentration, and the 4th highest 8-hour daily maximum average concentration were calculated for the EPA’s designated ozone season. Following receipt of the 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 in 2000 and 2003. The computer files provided contained summarized air quality data, a 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, there is not a valid estimate of the corrected seasonal cumulative index.

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 EPA-defined ozone season for each monitoring site that experienced sufficient data capture was provided by A.S.L. & Associates to the U.S. Forest Service project manager.

2.0 Scope of Work

Estimate the seasonal 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 seasonal 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, as mentioned above, another estimate of the error variance, which is referred to as the Ordinary kriging interpolation variance (NKVAR) was determined. The 95% error bound based on the new variance was reported also.

3.0 Steps in Modeling

In its 1982 Kriging study, NCLAN investigators were concerned about the selection of stations to be included in the air quality analysis (Heck *et al*., 1984). In urban settings, the ozone concentrations were thought to be lower at city-center than at rural locations because of nitric oxide titration in the city. Therefore, NCLAN investigators hypothesized that by using city-center monitoring stations to predict rural ozone levels, the resultant estimations might be biased low. Because of this concern, specific monitoring stations located in large metropolitan areas were not included in the 1982 NCLAN analysis.

Because significant changes have occurred to all metropolitan areas in the last 20 years, the method of filtering city-center sites used in the 1982 NCLAN study was re-examined. Using 2003 data, Table 1 shows statistics for several of the large metropolitan areas of the US. In Table 1, the first line in each case is the mean variance of the N100 values for the city-center sites included in the 1982 study. The second line is a summary of all the sites within the metropolitan area.

With the exception of Los Angeles, the 2003 set of sites had essentially the same N100 values as the entire metropolitan area. In Los Angeles the filtered city center sites are higher than the complete set of sites. Table 2, shows the same information for the W126 index.

Based on this test, it was decided to not use the filtering protocol developed in the 1982 NCLAN study, and to instead use all the monitoring data.

Table 1. City-Center Site Comparison for N100.

N100 City Center Analysis			
	N	Mean	Variance
Cincinnati			
City Center	8	9.1	61.1
City plus surrounding area	19	9.4	65.5
St. Louis			
City Center	14	12.5	15.4
City plus surrounding area	20	10.3	27.3
Chicago			
City Center	15	6.3	53.7
City plus surrounding area	19	9.4	65.5
Los Angeles			
City Center	24	61.7	6591.4
City plus surrounding area	42	84.2	11289.3

Table 2 City-Center Site Comparison for W126.

W126 City Center			
	N	Mean	Variance
Cincinnati			
City Center	8	24.4	21.3
City plus surrounding area	19	23.9	45.7
St. Louis			
City Center	14	24.6	44.7
City plus surrounding area	20	24.7	34.8
Chicago			
City Center	15	17.6	36.5
City plus surrounding area	39	18.0	30.6
Los Angeles			
City Center	24	40.6	681.2
City plus surrounding area	42	50.2	1114.7

Ozone exposure indices for California are known to be significantly different (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, longitude of all the monitoring sites from the AIRS database were compared and updated to coordinates supplied by ASL & Associates. In addition, many of the AIRS sites have incorrect elevations, missing elevations, or elevations listed in feet rather than meters. The elevations of all monitoring sites were checked and update 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 rest 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 of the N100 values. 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 for each 1/2° by 1/2° degree cell were determined. The N100 exposure for each 1/2° by 1/2° cell in California and the rest 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 Table 3. The search parameters are shown below.

Search Radius		California	US
Maximum search radius	=	500 km	1100 km
Max.number of sites used to estimate a cell	=	12	12
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 3.

Table 3. Example Output from Ordinary Kriging Program.

Latitude	Longitude	N100	Ordinary Kriging Var.	95%EB	OK Interpolation Variance	OKIV 95%EB	No. of Samples
44.5	-66.5	0.12	97.20	19.72	0.20	0.90	15
45.0	-66.5	0.12	98.41	19.84	0.19	0.87	15
44.0	-67.0	0.15	93.95	19.39	0.24	0.98	15
44.5	-67.0	0.12	86.52	18.60	0.20	0.89	15
45.0	-67.0	0.12	93.04	19.29	0.18	0.86	15

2000 year data

N100 Exposure Index

Table 4 shows summary statistics for the N100 index. Table 5 shows the variogram parameters for N100.

Table 4. Year 2000 N100 Statistics.

AREA	MEAN	VARIANCE	STD.DEV.	MIN	MAX	NUMBER
California	27.5	2431.2	49.3	0.0	230.6	150
Rest of US	7.8	168.8	13.0	0.0	126.9	889

Table 5. Year 2000 N100 Variogram Parameters.

AREA	Co	C1	Range1	C2	Range2	Angle	AF ratio
CAL	400	1900	200	400	1000	135	2.0
US	40	90	160	40	900	0	1.5

W126 Exposure Index – Basic Statistics

Table 6 shows summary statistics for the W126 index. Table 7 shows the variogram parameters for W126

Table 6. Year 2000 W126 Statistics.

AREA	MEAN	VARIANCE	STD.DEV.	MIN	MAX	NUMBER
California	33.6	976.5	31.2	0.0	121.0	150
Rest of US	24.9	185.8	13.6	0.0	96.7	889

Table 7. Year 2000 W126 Variogram Parameters.

AREA	Co	C1	Range1	C2	Range2	Angle	AF ratio
CAL	200	500	300	300	1000	-45	2.0
US	45	140	2250	0	0	0	3.0

2003 year data

N100 Exposure Index

Table 8 shows summary statistics for the N100 index. Table 9 shows the variogram parameters for N100

Table 8. Year 2003 N100 Statistics.

AREA	MEAN	VARIANCE	STD.DEV.	MIN	MAX	NUMBER
California	42.5	6569.7	81.1	0.0	420.1	149
Rest of US	6.2	81.4	9.02	0.0	58.2	982

Table 9. Year 2003 N100 Variogram Parameters.

AREA	Co	C1	Range1	C2	Range2	Angle	AF ratio
CAL	500	5000	200	900	1000	135	2.0
US	10	45	200	30	666	0	1.3

W126 Exposure Index

Table 10 shows summary statistics for the W126 index. Table 11 shows the variogram parameters for W126

Table 10. Year 2003 W126 Statistics.

AREA	MEAN	VARIANCE	STD.DEV.	MIN	MAX	NUMBER
California	38.8	1252.4	35.4	0.0	148.3	149
Rest of US	23.1	112.3	10.5	0.3	88.5	982

Table 11. Year 2003 W126 Variogram Parameters.

AREA	Co	C1	Range1	C2	Range2	Angle	AF ratio
CAL	400	950	225	200	1125	135	2.0
US	40	50	225	25	1125	0	1.0

The kriged results were prepared in 4 digital files with the following information: latitude, longitude, estimated kriged value (e.g., W126 or N100), kriged variance, 95% error bound, NKVAR, N95% error bound, and number of samples used to determine the kriged estimate.

4. 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.; 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.

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.

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.