Bull. Iraq nat. Hist. Mus.
(2015) 13 (3): 9-21

GEOSTATISTICAL ANALYSIS AND MAPPING OF OZONE OVER IRAQ

Hussain Zaydan Ali
Ministry of Science and Technology-Baghdad-Iraq
E-Mail: hussainzayali53@yahoo.com

ABSTRACT

The Ozone Monitoring Instrument (OMI) measures the reflected solar radiation in the ultraviolet and visible part in the spectral range that is between 270 and 500 nm, using two channels with a spectral resolution of about 0.5 nm. Ground-level tropospheric ozone is one of the air pollutants of most concern. In the troposphere, near the Earth's surface, human activities lead to ozone concentrations several times higher than the natural background level. To evaluate the ozone distribution over Iraq, the ozone data from OMI were analyzed using geostatistical techniques. Theoretical spherical models provided the best fit for all monthly experimental variograms. The parameters of these variograms (sill, range and nugget) were evaluated. Finally, predictive ozone maps were derived for all points of the study area, by use of geostatistical algorithms (kriging). High prediction accuracy was obtained in all cases as cross-validation showed.

Keywords: Pollution, Geostatistical Analysis, Kriging, tropospheric ozone mapping.

INTRODUCTION

Data from the Ozone Monitoring Instrument (OMI) onboard the Aura satellite (from 2005 to 2012) all processed using geoststistical analyst. The Ozone Monitoring Instrument (OMI) flies on the National Aeronautics and Space Administration's Earth Observing System Aura satellite launched in July 2004 (Figure (1)). OMI is a ultraviolet/visible (UV/VIS) nadir solar backscatter spectrometer, which provides nearly global coverage in one day with a spatial resolution of 13 km 24 km. Trace gases measured include O3, NO2, SO2, etc. OMI's unique capabilities for measuring important trace gases with a small footprint and daily global coverage will be a major contribution to our understanding of stratospheric and tropospheric chemistry and climate change. OMI's high spatial resolution is unprecedented and will enable detection of air pollution on urban scale resolution. OMI measures the reflected solar radiation in the ultraviolet and visible part in the spectral range that is between 270 and 500 nm, using two channels with a spectral resolution of about 0.5 nm. The light entering the telescope is spatially pseudo-depolarized using a scrambler and then split in two channels: the UV channel (full performance range 270–365 nm) and the VIS channel (full performance range 365–500 nm) [1, 2 and 5].

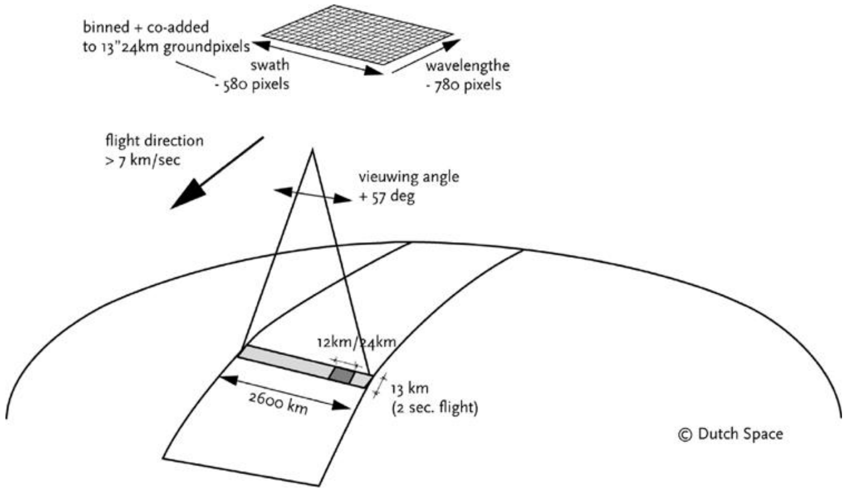

Figure (1): OMI measurement principle.

The UV channel consists of two sub channels with the following full performance ranges: the UV-1, ranging from 270–310 nm, and the UV-2 ranging from 310-365 nm. The spectral and spatial sampling of the UV-1 are reduced by a factor of two compared to the UV-2. The full performance range of the VIS-channel ranges from 365–500 nm. Ground-level tropospheric ozone is one of the air pollutants of most concern. It is mainly produced by photochemical processes involving nitrogen oxides and volatile organic compounds in the lower parts of the atmosphere. Ozone levels become particularly high in regions close to high ozone precursor emissions and during summer, when stagnant meteorological conditions with high insolation and high temperatures are common. Ground ozone levels is a topic of considerable environmental concern, since excessive level of ozone are taken as indicative of high pollution. Human activities have led to much higher ground-level ozone concentrations.

In the troposphere, near the Earth's surface, human activities lead to ozone concentrations several times higher than the natural background level. Too much of this ground-level ozone is “bad” as it is harmful to breathe and also damages the environment. When ozone mixes with other air pollutants, especially nitrogen oxides and particulate matter, it can form harmful smog. This smog occasionally takes place in polluted city areas. In addition, ozone is a greenhouse gas which may have important global climatic consequences. Ozone is a natural component of the troposphere, produced by photochemical reactions of nitrogen oxides and volatile organic compounds (VOC), and collectively called ozone precursors, enhanced by temperature and sunlight. Emissions from car exhausts, power plants and industrial facilities are the major sources of nitrogen oxides and VOC [1, 12, 13 and 14]. In this study the ozone data for Iraq were analyzed using Geostatistical techniques in ArcGIS. Thus, first, during the exploratory analysis of data, it was revealed that they were distributed normally, which is a desirable property for the subsequent stages of the geostatistical study. Secondly, during the structural analysis of data, theoretical spherical models provided the best fit for all monthly experimental variograms. The parameters of these variograms (sill, range and nugget) were calculated. Finally, predictive ozone maps were derived for all points of the study area (Iraq), by use of geostatistical algorithms (kriging). High prediction accuracy was obtained in all cases as cross-validation showed. Many studies about the temporal and spatial variability of ground level ozone were conducted [1, 17 and 18]. This study aimed to determine the best method (lowest cross validation error) for interpolating the spatial distribution of ozone data over Iraq.

METHODOLOGY

The presence of a spatial structure where observations close to each other are more alike than those that are far apart (spatial autocorrelation) is a prerequisite to the application of geostatistics. The experimental variogram measures the average degree of dissimilarity between unsampled values and a nearby data value, and thus can depict autocorrelation at various distances. The value of the experimental variogram for a separation distance of h (referred to as the lag) is half the average squared difference between the value at z (xi) and the value at z (xi + h):

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \qquad (1)$$

Where N (h) is the number of data pairs within a given class of distance and direction. If the values at z (xi) and z (xi + h) are auto correlated the result will be small, relative to an uncorrelated pair of points. From analysis of the experimental variogram, a suitable model (e.g. spherical, exponential) is then fitted, usually by weighted least squares, and the parameters (e.g. range, nugget and sill) are then used in the kriging procedure [7, 11 and 16].
Fitting a variogram model.

The variogram must be expressed as a mathematical function before being used for kriging. This is typically achieved by fitting a suitable function to the experimental variogram. Each function is defined in terms of a small number of parameters that are selected to best-fit the function to the experimental variogram. In this study we use two functions, namely spherical and circular. Below the spherical function adopted, which is defined by:

$$\gamma(h) = \begin{cases} c_0 & \text{when } h = \varepsilon \text{ (a very small lag)} \\ c_0 + \mathrm{c}\left(\frac{3h}{2a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right) & \text{when } 0 < h \le a \\ c_0 + \mathrm{c} & \text{when } h > a \end{cases} \qquad (2)$$

Where c0 is the nugget variance, c+c0 is sill, h is the lag and a is the range. All variograms computed in this study are all fitted with spherical model. The spherical model is the most commonly used model for experimental data. This function is expressed in terms of three parameters, namely; a range of spatial correlation, c0 the nugget effect and c1 the sill value. When a variogram is plotted using discrete experimental data points, it is called an experimental or sample variogram. A theoretical model can be fitted through the experimental data points to quantify spatial patterns. The shape and description of a “classic” variogram is shown in Figure (2).

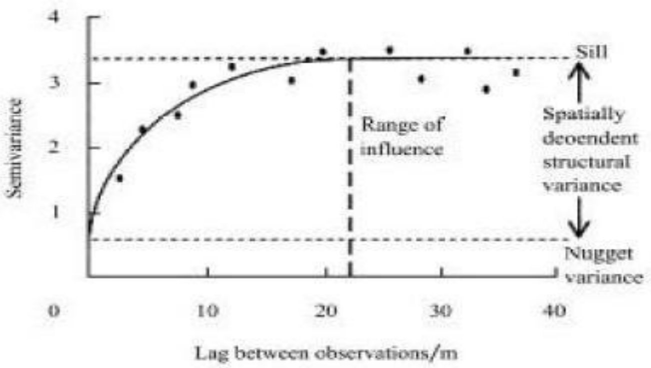

Figure (2): The variogram.

There are three key terms in each model, the sill, the range, and nugget variance. The sill corresponds to the overall variance in the dataset and the range is the maximum distance of spatial autocorrelation. The nugget variance is the positive intercept of the variogram and can be caused by measurement errors or spatial sources of variation at distances smaller than the sampling interval or both [9, 15 and 19].

With the wide and increasing applications of the spatial interpolation methods, there is also a growing concern about their accuracy and precision. As any other statistical modeling techniques, the spatial interpolation methods also produce a certain degree of errors associated with the estimation. The statistics of the differences (absolute and squared) between the measured and predicted values at sampled points are often used as an indicator of the performance of an inexact method. Several error measurements have been proposed. Commonly used error measurements include: mean error (ME), mean absolute error (MAE), mean squared error (MSE) and root mean squared error (RMSE). I am used for determining the degree of bias in the estimates, often referred to as the bias but it should be used cautiously as an indicator of accuracy because negative and positive estimates counteract each other and resultant ME tends to be lower than actual error. RMSE provides a measure of the error size, but it is sensitive to outliers as it places a lot of weight on large errors. MSE suffers the same drawbacks as RMSE. Whereas MAE is less sensitive to extreme values and indicates the extent to which the estimate can be in error. MAE and RMSE are argued to be similar measures, and they give estimates of the average error, but they do not provide information about the relative size of the average difference and the nature of differences comprising them. Of course, we can also use cross-validation in together with these measurements to assess the performance of both exact and inexact methods. RMSE and MAE are argued to be among the best overall measures of model performance as they summarize the mean difference in the units of observed and predicted values [4, 9, 10 and 17]. The purpose of Geographic Information System (GIS) is to provide a spatial framework to support decisions for the intelligent use of earth's resources and to manage the man-made environment. Biologists, botanists, planners, petroleum engineers, foresters, and corporate executives are increasingly relying on GIS to help them make critical decisions. By putting spatial data in an integrated system where it can be organized, analyzed, and mapped, patterns and relationships that were previously unrecognized may emerge. The air pollution continues to be a significant environmental problem in many countries. Geostatistical analysis methods have been available for several decades, but were not integrated into any GIS modeling environments. A newly released software package, Geostatistical Analyst links GIS and geostatistical analysis methods. With Geostatistical Analyst a continuous surface (a map or a distribution model) can be created from measured sample points. Data collection usually can only be conducted at a limited number of point locations due to logistical and financial limitations; however scientists and managers are increasingly interested in continuous surface estimates. In order to generate such a surface some type of interpolation method must be used to estimate data values for those locations where no samples were taken. Geostatistical methods, such as kriging apply regionalized variables and describe spatial dependencies between the instances of random variables by using variograms. A variogram is a graphical display of a variance of measurements over the distance between the measurement sites. If there are spatial dependencies the variance between the observations on two points normally increases with increasing distance until at a specific range a maximum value is reached. Kriging is considered to be the most sophisticated geostatistical method as it can potentially provide the most accurate results [3, 6, 8 and 16]. In this work an application of the Geostatistical Analyst for development of O3 distribution models will be discussed.

RESULTS AND DISCUSSION

Many studies have employed distance weighting methods, but kriging is the only one which incorporates the spatial correlation into its estimation algorithm. Kriging has been used more widely due to its many advantages. Although kriging requires an abundance of sample points to be an accurate spatial interpolation method, even when relative small data sets and not exhaustive samplings are available it is a reliable technique for investigating the distribution of tropospheric ozone. Maps were constructed using the Geographic Information System (GIS) software and the Geostatistical analyst extension (ESRI). The Kriging method for surface interpolation method was used. In this method, a value is estimated by averaging the values of sample data points in the vicinity around it. The closer a sample point is to the point being estimated, the more influence, or weight, it has in the process. The produced ozone maps are shown in figures (3) to (8).

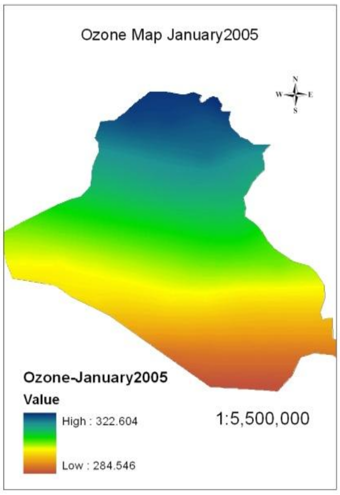

Fig (3): Ozone Map January 2005

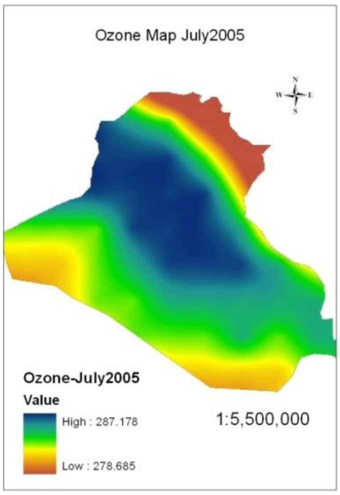

Fig (4): Ozone Map July 2005.

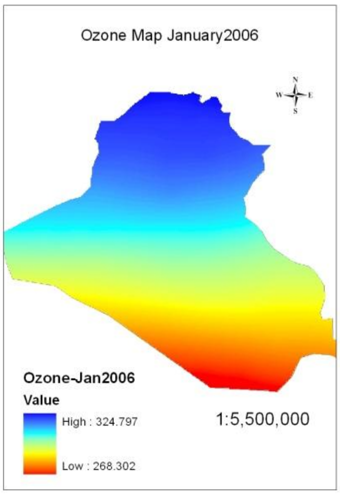

Fig (5): Ozone Map January 2006.

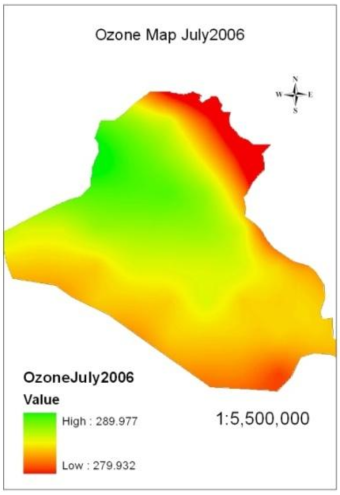

Fig (6): Ozone Map July 2006.

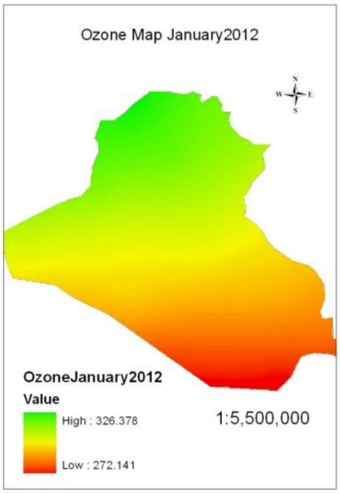

Fig (7): Ozone Map January 2012.

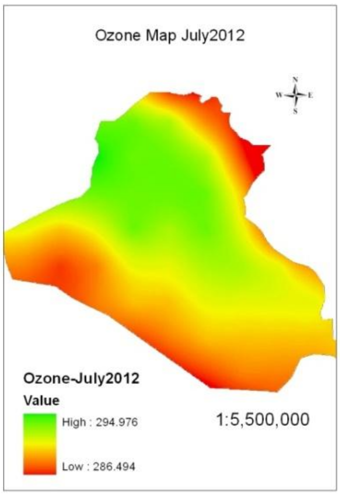

Fig (8): Ozone Map July 2012.

Cross validation is used to compare measured values with interpolated values using only the information available in the sample data set. A cross validation study can help to choose between different weighting procedures, between different search strategies, or between different estimation methods. The sample value at a particular location is temporarily discarded from the sample data set; the value at the same location is then estimated using the remaining samples. Once the estimate is calculated, the calculated value can be compared with the true value that was initially removed from the sample data set. This procedure is repeated for all available sample values. In this study all parameters of methods were optimized for minimum cross validation error. Table (1) below shows the cross validation errors for months: January and July, for the years: 2005, 2006, and 2012.

Table (1): Cross Validation Results

Criteria	January 2005	January 2006	January 2012	July 2005	July 2006	July 2012
Mean	-.0057	-.00285	-.0126	-.00575	-.00827	-.00984
Root-Mean-Square	.7913	.7436	1.204	.6503	.6226	.6723
Average Standard Error	5.733	8.775	7.812	1.295	1.293	1.389
Mean Standardized	-.00053	-.00029	-.00087	-.003283	-.0040	-.004229
Root-Mean-Square Standardized	.1295	.0797	.1415	0.4875	.4711	.4773

Since a strong spatial dependence between ozone data is observed, the geostatistical algorithms, particularly the ordinary kriging, provide accurate estimates, as cross validation confirmed. Predictions errors are fair indicators of the variability in predictions, as cross validation revealed, and normality is apparent, probability maps were finally generated. They are very useful tools for hazard assessment and decision support, as presented in figures (9) to (13).

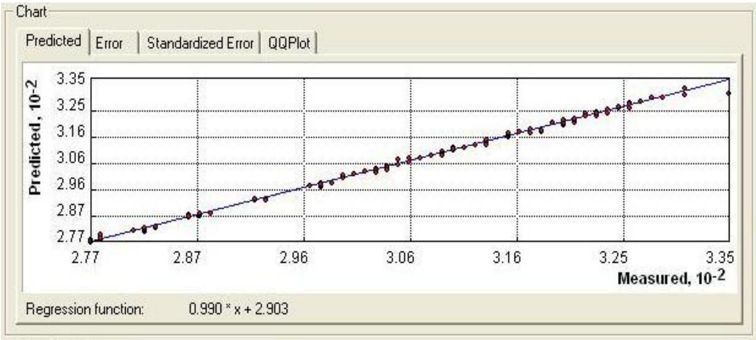

Fig (9): Measured vs. Predicted Ozone Data In January 2005.

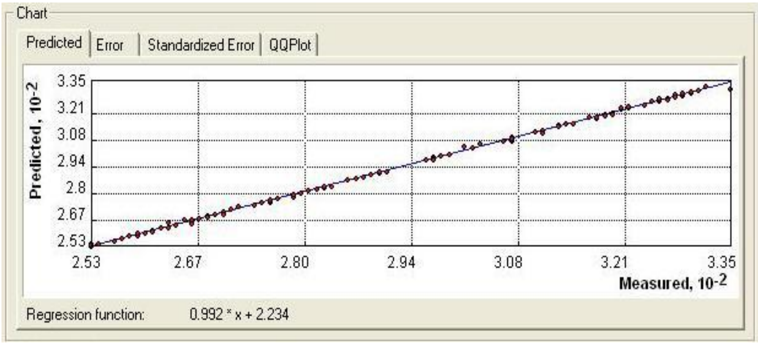

Fig (10): Measured vs. Predicted Ozone Data In January 2006.

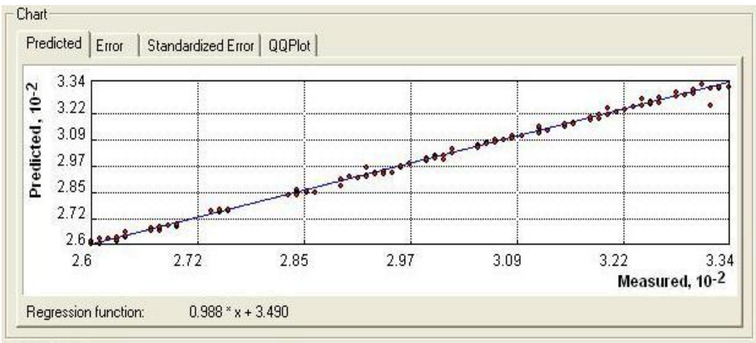

Fig (11): Measured vs. Predicted Ozone Data In January 2012.

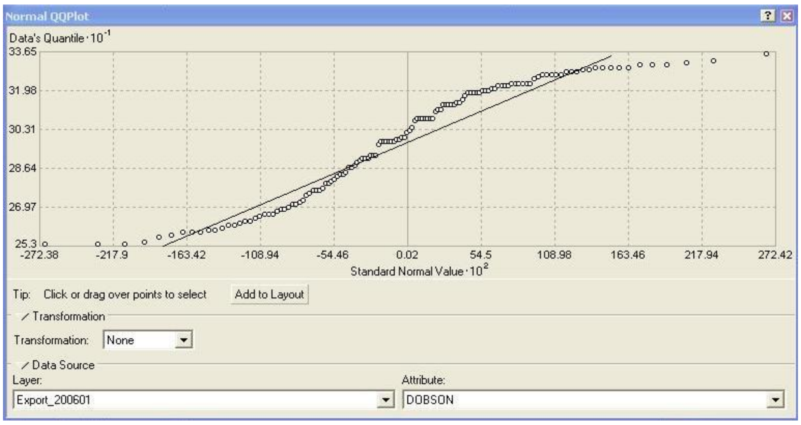

Fig (12): QQPlot of Ozone Data in January 2006.

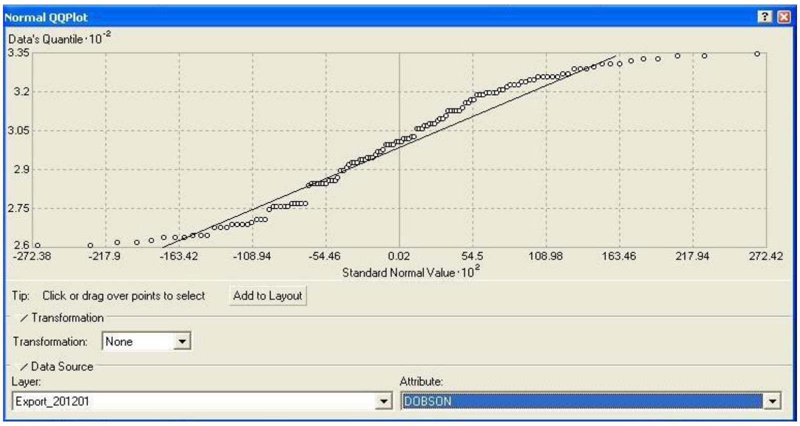

Fig (13): QQPlot of Ozone Data in January 2012.

CONCLUSION

Global monitoring of ozone is essential as it plays an important role in the chemical processes occurring in the atmosphere and has a major impact on the climate. Tropospheric and stratospheric ozone are highly variable in both space and time and thus in order to correctly quantify its effect on stratospheric chemistry, air quality and radiative forcing it is necessary to develop accurate global measurements. The main health concern of exposure to ambient ground-level ozone is its effect on the respiratory system, especially on lung function. Several factors influence these health impacts, including the concentrations of ground-level ozone in the atmosphere, the duration of exposure, average volume of air breathed per minute (ventilation rate), and the length of intervals between short-term exposures. The main goal of interpolation is to discern the spatial patterns of ozone data by estimating values at unsampled locations based on measurements at sample points. Geostatistics provides an advanced methodology to quantify the spatial features of the studied variables and enables spatial interpolation, kriging. In addition, geographical information systems (GIS) and geostatistics have opened up new ways to study and analyze spatial distributions of regionalized variables, i.e. distributed continuously on space. Moreover, they have become useful tools for the study of hazard assessment and spatial uncertainty. Without a GIS, analysis and management of large spatial data bases may not be possible. The ground-level ozone in an urban environment must be studied by means of high-resolution ozone maps, which are essential tools to properly diagnose and propose control measures with the aim of minimizing its effects. In this work, geostatistical techniques are considered to model the ambient air ozone distribution over the experimental area. Although the real spatial complexity of ozone surfaces can not be captured, the proposed techniques provide some reliable surfaces at enough spatial resolution to correctly visualize the spatial patterns of this pollutant. Polluted areas in Iraq have to be delimited. Future actions against ozone should be particularly aimed at reducing the high levels in these zones. Consequently, the ozone maps can influence decisions concerning air-quality policy, which, in turn, affect the attitudes and behaviors of the general public.

ACKNOWLEDGMENT

The author is grateful to Dr. Ali M. Al-Salihi and Zahraa M. Hassan, Atmospheric Sciences Department, College of Science, Al-Mustansiriyah University for providing the data.

LITERATURE CITED

Al-Salihi, Ali M. and Hassan, Zahraa M. (2014). “Analysis of Temporal and Spatial Patterns of Ozone Over Iraq”, M.Sc. Thesis, Atmospheric Sciences Department, College of Science, Al-Mustansiriyah University.

Bodeker, GE.; Scott, JC.; Kreher, K. and McKenzie, RL. (2001). Global ozone trends in potential vorticity coordinates using TOMS and GOME intercompared against the Dobson network: 1978-1998. J Geophysical Res; 106, 23029-42.

Burrough, P. and McDonnell, R. (Ed.). (1998). Principles of Geographical Information Systems. Oxford University Press.

Cressie, N., (1985). Fitting variogram models by weighted least squares, Mathematical Geology, 17(5), 563-586.

Crutzen PJ. (1971). Ozone production rates in an oxygen-hydrogen nitrogen oxide atmosphere. J Geophysical Res; 76: 7311-27.

ESRI (Ed.). (2001). Using ArcGIS Geostatistical Analyst. ESRI Press.

Goovaerts P., (2001). “Geostatistical modeling of uncertainty in soil science,” Geoderma, vol. 103, pp. 3-26.

Isaaks, E. and Srivastava, R. (Ed.). (1989). An Introduction to Applied Geostatistics. Oxford University Press.

Johnston, K.; Ver Hoef, J. M.; Krivoruchko, K. and Lukas, N. (2003). Using ArcGis Geostatistical Analyst, ESRI, 300 pp.

Kravchenko, A. and Bullock, D. (1999). A Comparative Study of Interpolation Methods for Mapping Soil Properties. Agronomy Journal, 91, 393-400.

Lark, R.M., (2002). Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood, Geoderma, 105 (1-2): 49-80.

Lelieveld, J. and Denterner F.J. (2000). What controls tropospheric ozone? Journal of Geophysical Research, vol. 105, pp. 3531-3551.

Logan J.A., (1985). Tropospheric ozone: seasonal behavior, trends, and anthropogenic influence, Journal of Geophysical Research, vol. 90, pp. 10463-10482, 1985.

Monks, P.S., (2000). A review of the observations and origins of the spring ozone maximum, Atmospheric Environment, vol. 34, pp. 3545-3561.

Wahba, G. and Wendelberger, J. (1980). Some new mathematical methods for variational objective analysis using splines and cross-validation. Monthly Weather Review 108: 1122–1145.

Warrick, A.W. and Myers, D.E. (1987). Optimization of sampling locations for variogram calculations, Water Resources Research, 23 (3), 496-500.

Webster, R. and Oliver, M.A. (2001). Geostatistics for Environmental Scientists. Brisbane, Australia: John Wiley & Sons Ltd.

Zeigler, M. (1999). Modeling Our World. The ESRI Guide to Geodatabase Design. ESRI Press, Redlands, CA.

Zimmerman, D.; Pavlik, C.; Ruggles, A. and Armstrong, M. (1999). An Experimental Comparison of Ordinary and UK and Inverse Distance Weighting. Mathematical Geology, 31, 370-395.

Bull. Iraq nat. Hist. Mus.
(2015) 13 (3): 9-21

التحليل الجيواحصائي وإنتاج خرائط الأوزون فوق العراق

حسين زيدان علي
وزارة العلوم والتكنولوجيا - بغداد - العراق

الخلاصة

يقيس جهاز مراقبة الأوزون (OMI) الإشعاع الشمسي المنعكس من الجزء الطيفي الفوق البنفسجي والمرئي بين ٢٧٠ و ٥٠٠ نانومتر ، باستخدام قناتين بقابلية تحليل طبقية حوالي ٠,٥ نانومتر. يعتبر الأوزون التروبوسفيري على مستوى سطح الأرض من ملوثات الهواء الأكثر اهتماما . في طبقة التروبوسفير وقرب سطح الأرض تقود النشاطات البشرية إلى تركيز للأوزون عدة مرات أكثر من مستوى الخلفية الطبيعي. لتقدير توزيع الأوزون فوق العراق ، فان بيانات الأوزون من OMI تم تحليلها باستخدام التقانات الجيواحصائية. لقد وفرت الموديلات الكروية النظرية أفضل تطابق لكل الـ variograms الشهرية العملية. إن معاملات هذه الـ variograms والتي تتضمن sill, range و nugget قد تم تقديرها أخيراً، وتم اشتقاق الخرائط المتوقعة للأوزون لكل نقاط منطقة الدراسة، باستخدام الخوارزميات الجيواحصائية (كرجنج). تم الحصول على توقعات دقيقة لكل الحالات كما بينت نتائج التدقيق المتقاطع.

الكلمات الدالة: تلوث ،تحليل جيواحصائي ،كرجنج ،إنتاج خرائط الأوزون التروبوسفيري.