

Statistical interpolation of ozone measurements from satellite data (TOMS, SBUV and SAGE II) using the kriging method

B. J. S. Tranchant, A. P. Vincent

▶ To cite this version:

B. J. S. Tranchant, A. P. Vincent. Statistical interpolation of ozone measurements from satellite data (TOMS, SBUV and SAGE II) using the kriging method. Annales Geophysicae, 2000, 18 (6), pp.666-678. hal-00316677

HAL Id: hal-00316677 https://hal.science/hal-00316677

Submitted on 18 Jun 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Statistical interpolation of ozone measurements from satellite data (TOMS, SBUV and SAGE II) using the kriging method

B. J. S. Tranchant^{1,2,*}, A. P. Vincent^{1,2}

¹ Département de physique, Université de Montréal, C.P. 6128, Succ. centre-ville, Montréal, Québec H3C 3J7, Canada ² Centre de Recherche en Calcul Appliqué, 5160 Boul. Décarie, Bureau 400, Montréal, Québec, H3X 2H9, Canada

Received: 9 July 1999 / Revised: 18 February 2000 / Accepted: 22 March 2000

Abstract. This study demonstrates that ordinary kriging in spherical coordinates using experimental semi-variograms provides highly usable results, especially near the pole in winter and/or where there could be data missing over large areas. In addition, kriging allows display of the spatial variability of daily ozone measurements at different pressure levels. Three satellite data sets were used: Total Ozone Mapping Spectrometer (TOMS) data, Solar Backscattered UltraViolet (SBUV), and the Stratospheric Aerosol and Gas Experiment (SAGE II) ozone profiles. Since SBUV is a nadir-viewing instrument, measurements are only taken along the sunsynchronous polar orbits of the satellite. SAGE II is a limb-viewing solar occultation instrument, and measurements have high vertical resolution but poor daily coverage. TOMS has wider coverage with equidistant distribution of data (resolution $1^{\circ} \times 1.25^{\circ}$) but provides no vertical information. Comparisons of the resulting SBUV-interpolated (column-integrated) ozone field with TOMS data are strongly in agreement, with a global correlation of close to 98%. Comparisons of SBUVinterpolated ozone profiles with daily SAGE II profiles are relatively good, and comparable to those found in the literature. The interpolated ozone layers at different pressure levels are shown.

Key words: Atmospheric composition and structure (middle atmosphere – composition and chemistry) – Meteorology and atmospheric dynamics (middle atmosphere dynamics)

1 Introduction

A thorough knowledge of the tri-dimensional structure of ozone near the polar vortex is a key requirement for a better understanding of ozone depletion occurring in the spring hemisphere.

Most ozone measurements so far have been taken in integrated-column form, either by satellite, such as TOMS (Total Ozone Mapping Spectrometer) and more recently GOME (Global Ozone Monitoring Experiment) (Munro et al., 1998), or by the ground observation Dobson network. The combination of these two types of observation provides both relatively good planetary coverage of the total ozone and a relatively good trend over time (Bojkov and Fioletov, 1995). However, vertical data are far less abundant. The few ground-based stations available for vertical ozone profile measurement are located mostly in the Northern Hemisphere (Mégie, 1992). Ground measurements are taken using either ozonesondes or the "Umkehr" method. Only a few satellite measurements (retrievals) provide the vertical distribution of the ozone field, such as SAGE (Stratospheric Aerosol and Gas Experiment), SBUV (Solar Backscatter UltraViolet), HALOE (HALogene Occultation Experiment) or more recently GOME (Eichmann et al., 1997) Generally, however, all of these techniques have difficulty in obtaining simultaneously both good resolution and good coverage. This study deals with three kinds of asynoptic data. The first is the SBUV dataset, for which data are non-equidistantly distributed. An example is shown in Fig. 1, which displays satellite tracks (symbolized by crosses) on October 5, 1988. We see that the main advantage of Sun-synchronous orbits is that these orbits can reach high latitudes (near-polar orbits), that is, they can reach the region where the ozone depletion occurs. SAGE II is the second dataset, which we used to validate the interpolated ozone profiles from the SBUV data. The TOMS dataset was the third source of asynoptic data. It differs from the other two in that the (total ozone) data

Correspondence to: B. J. S. Tranchant e-mail: Benoit.tranchant@cerfacs.fr

^{*}Present address:

Centre Européan de Recherche et de Formation Avancée en Calcul Scientifique, 42 av. Gaspard Coriolis 31057 France



Fig. 1. Satellite tracks (crosses) correspond to SBUV measurements and circles correspond to SAGE II measurements on October 5, 1988

are distributed two-dimensionally. Although the former data sources provide good resolution, some events show data missing over large areas. The total ozone field displayed in Fig. 2, taken on the same date as Fig. 1, is an example of the large gaps found in the data.

Of course, there are many techniques for filling in missing data in satellite-derived fields, or for generating synoptic maps from asynoptically gathered data. These range from mathematical to statistical interpolation and data assimilation methods such as the variationnal methods (3DVAR-4DVAR) of Pellerin et al. (1999). Methods such as trajectory analysis, which use analyzed meteorological data, have also been developed (e.g. Morris et al., 1995). We adopted a statistical approach using the kriging method, which was originally developed to process mining geostatistics. Kriging has been used for other spatial-estimation applications (e.g., Lefohn and Runeckles, 1987), and more recently to predict O_3 exposures in certain areas (Lefohn *et al.*, 1997). Kriging has been also shown to be highly efficient as an interpolating scheme for semi-Lagrangian advection on unstructured meshes for ocean modelling (LeRoux et al., 1997). Recently, Bergeron et al. (1999) have also used 2D kriging seismic fields for wavelet

applications. The advantage of the kriging technique is that we use covariance functions or semi-variograms that depend only on the displacement vector, and not on the location in space (intrinsic hypothesis). Consequently, this method can handle spatial auto-correlations. Here, we apply the kriging technique both for retrieving the horizontal structure of ozone at different altitudes (3D) with sufficient resolution, and to fill in large gaps where data are missing (zero values) in 2D ozone fields.

2 Datasets

2.1 Ozone profile dataset: SBUV and SAGE II

SBUV data were chosen for two reasons. Firstly, many comparisons have been made with ground measurements (De Luisi *et al.*, 1994; McPeters *et al.*, 1994) and these report that the SBUV dataset provides good results. Secondly, SBUV provides the best horizontal resolution among the available datasets, and also provides vertical profiles.

The Solar Backscatter UltraViolet radiometer (SBUV) was on board the Nimbus 7 satellite. Its purpose was to estimate global ozone distribution by measuring backscattered solar radiation in the ultraviolet Hartley-Huggins bands (Kidder and Vonder Haar, 1995). SBUV infers the vertical ozone profile from backscattered albedo, the ratio of the radiance backscattered from the terrestrial atmosphere to the extraterrestrial solar irradiance, at 12 discrete wavelengths from 255.5 to 339.8 nm (Chandra and McPeters, 1994). The profile inversion is performed in 12 Umkehr layers using an optical statistical method, see e.g., Fleig et al., (1990) and Bhartia et al. (1996). The approximate vertical resolution of SBUV measurement data is 8 km McPeters et al., 1994). The archived data we used are provided in ozone column amounts (Dobson Units) for 12 distinct layers that are similar to those used by the ground-based Umkehr technique (Mateer and DeLuisi, 1992). The corresponding pressure and approximate altitude (at the layer center) are shown in Table 1. The layers are approximately 5 km thick. Despite this relatively poor





Table 1. Umkehr layers used for ozone profile retrievals with SBUV

Layer number	Pressure range (mbar)	Pressure at layer centre (mbar)	~Height at layer centre (km)
1 ^a	253.3-1013	507	5.5
2	127-253	179	12.5
3	63.3-127	89.6	17.0
4	31.7-63.3	44.8	21.3
5	15.8-31.7	22.4	25.8
6	7.92-15.8	11.2	30.4
7	3.96-7.92	5.6	35.2
8	1.98-3.96	2.8	40.2
9	0.990-1.98	1.4	45.5
10	0.495-0.990	0.700	51.0
11	0.22467-0.495	0.350	56.5
12	0.0-02467	_	-

^a Layer 1 is a double layer and has no geophysical information [Bhartia *et al.*, 1996]

vertical resolution, comparisons have been made over a few years with other data from different measurement instruments. The quality of the reprocessed Version 6, which we use for comparison with external data, has been assessed by Gleason and McPeters, (1995). These data were reprocessed to correct chopper synchronization error. Comparisons of ozone profiles between SBUV and SAGE II (Stratospheric Aerosol and Gas Experiments), which has better vertical resolution but poorer horizontal resolution, have been conducted by McPeters et al. (1994), and they attest to the high quality of SBUV ozone profiles. Good agreement between SBUV and the Umkehr mean profiles from five stations in the Northern Hemisphere has been found by De Luisi et al. (1994). Recently, it has been shown that the SBUV instrument can capture large-scale variability in ozone over a broad range of atmospheric pressures (0.3-100 mbar) (Bhartia et al., 1996). Unfortunately, SBUV cannot scan spatially; it simply produces a 188 km nadir view, so horizontal resolution is affected (see Fig. 1). Indeed, even if measurements cover the entire world every 24 h (\sim 14 tracks), resolution is only limited along the orbital tracks. In addition, measurements are taken along tracks from 80° south to 80° north due to the inclined orbit, and no data are available during the polar night since SBUV measures scattered sunlight (Kidder and Vonder Haar, 1995).

We also used archived data provided as volumetric mixing ratios (ppmv) for 17 pressure levels to compare with the SAGE II ozone profiles. The SAGE II instrument is aboard the Earth Radiation Budget Satellite (ERBS) (Kidder and Vonder Haar, 1995). It uses a limb-viewing solar occultation measurement technique that has approximately 1-km vertical resolution (McPeters *et al.*, 1994). The retrieved ozone profiles are provided as density concentrations (or as volumetric mixing ratios) at 70 altitudes, with a resolution of between 0.5 and 1 km. This is a high vertical resolution compared to SBUV, but ozone measurements are taken at a given latitude only once every month. For example, there are only 30 measurement points on October 5, 1988 (see circles on Fig. 1).

2.2 Total column ozone dataset: TOMS and SBUV

SBUV archived data also include the total amount of ozone, calculated by taking the four longest absorbed wavelengths we used for comparison. The best agreement (error $\approx 1\%$) between SBUV and SAGE II was found when comparing integrated column measurements over a period of 6 years (McPeters *et al.*, 1994).

Unlike the SBUV measurements, the TOMS (Total Ozone Mapping Spectrometer) measurements were taken by scanning across the track of the satellite. TOMS measures the Earth's backscattered ultraviolet on six discrete wavelength channels. The TOMS field of view is a 49-km scan view. Thus the archived daily data we used are mapped on a grid of 1° latitude by 1.25° longitude. The total ozone retrieval used by TOMS is quite similar to that of SBUV, except that TOMS uses larger wavelengths than those of the SBUV scale. In this study, we use archived Nimbus 7 TOMS data reprocessed with the Version 7 algorithm, which provides accurate results for the Northern hemisphere. TOMS data agree strongly with the average of 30 Northern Hemisphere Dobson/Brewer stations up to a solar zenith angle of 80° (McPeters and Labow, 1996; McPeters *et al.*, 1996).

3 The kriging approach

Kriging is a local estimation technique that provides the best linear unbiased estimator (BLUE) of the unknown characteristic studied (Journel and Huijbregts, 1989). Kriging is a minimum variance method which is closely connected with statistical interpolation methods (Daley, 1991). The term "kriging" is derived from the name of D.G. Krige, who first worked on moving averages (Krige, 1951). Originally developed for mining applications, kriging is an interpolation technique quite similar to the Wiener-Kolmogorov theory prediction technique (Ripley, 1981; Matheron, 1965). In this study, both ordinary kriging (OK) and universal kriging (UK) were used. Hereafter, we speak of UK; OK should be viewed as a specific case of UK.

The kriging approach assumes that a physical phenomenon [here the ozone measurement z(X)] can be considered as a random function Z(X) that consists of several random variables over a domain D with $X \in D$.

One of the principal assumptions is that the random function is the sum of a mean value, the "drift" m(X), and a fluctuation e(X) (error or residual term).

$$Z(X) = m(X) + e(X) \tag{1}$$

The drift quantifies the regular and continuous variation of Z(X) according to the scale of the observational data (Journel and Huijbregts, 1989). Fluctuations can been regarded as a small-scale random process and the mean value as a large-scale variation (Ripley, 1981). By definition, the drift is the non-stationary expectation of the random function Z(X), and the residual term has a zero expectation, also known as the non-bias condition.

$$E\{Z(X)\} = m(X) \quad \text{with} \quad E\{e(X)\} = 0 \quad \forall X \tag{2}$$

When the drift is supposed to be a constant value (zero order), which means that the random function Z(X) is considered stationary or quasi-stationary, we deal with ordinary kriging (OK). In our case, global daily data representation, it seems reasonable to suppose that constant drift provides a good approximation. On the other hand, when there is a gradual trend in the data, UK is more appropriate. Then, the drift is no longer constant, and we assume that we know the shape of the trend, it can be linear, quadratic or cubic depending on the data variation in time over the domain. Consequently, the drift can be spanned by the *k* basis function f'(X):

$$E\{Z(X)\} = m(X) = \sum_{l=0}^{k} a_l f^l(X)$$
(3)

where a_l represents the unknown polynomial coefficients.

The kriging approach concerns a linear predictor $\hat{Z}_K(X)$, which is in our case a linear combination of ozone data values $Z(X_i)$.

$$\hat{Z}_{K}(X) = \sum_{i=1}^{n} \lambda_{i} Z_{i} \quad \text{with} \quad Z_{i} = Z(X_{i})$$
(4)

where λ_i are the weights calculated to ensure that the estimator is unbiased. The non-bias condition entails

$$E\{Z(X) - \hat{Z}_{K}(X)\} = E\{Z(X)\} - E\{\hat{Z}_{K}(X)\} = 0$$
(5)

It is equivalent to a zero error in expectation, and from Eqs. (3) and (4) we obtain:

$$m(X) = \sum_{i=1}^{n} \lambda_i m(X_i) \tag{6}$$

and thus

$$\sum_{l=0}^{k} a_{l} \left[\sum_{i=1}^{n} \lambda_{i} f^{l}(X_{i}) - f^{l}(X) \right] = 0$$
(7)

We want Eq. (7) to be true for any values of the unknown a_l , so a constraint on the estimator weights is:

$$\sum_{i=1}^{n} \lambda_i f^l(X_i) = f^l(X) \quad \forall l = 0 \text{ to } k$$
(8)

In the case of constant drift (OK), the sum of the estimator weights Eq. (8) is equal to one. Here, we see that this constraint is closely related to the minimum variance estimation (Daley, 1991). The kriging estimate is defined by the values of the weights λ_i that minimize estimation variance, or that minimize the mean-square error:

$$E\{[Z(X) - \hat{Z}_{K}(X)]^{2}\}$$

= $E\{Z(X)\}^{2} - 2\sum_{i=1}^{n} \lambda_{i}E\{Z(X)Z(X_{i})\}$
+ $\sum_{i,j=1}^{n} \lambda_{i}\lambda_{j}E\{Z(X_{i})Z(X_{j})\}$ (9)

This estimation variance appears as a quadratic form in $\lambda_i \ \lambda_j$, subject to the k non-bias constraints of Eq. (8). Using the Lagrange method of undetermined multipliers, a system of (n + k) linear equations and (n + k) unknowns (the n weights λ_i and the k Lagrange multipliers μ_l) is obtained. This system Eq. (9) is called the "universal kriging system" (Journel and Huijbregts, 1989; Delfiner and Delhomme, 1973).

.

$$\begin{cases} E\{Z(X)Z(X_i)\}\\ =\sum_{j=1}^n \lambda_j E\{Z(X_i)Z(X_j)\} - \sum_{l=1}^k \mu_l f^l(X_i) \quad \forall i = 1 \text{ to } n\\ \sum_{i=1}^n \lambda_i f^l(X_i) = f^l(X) \quad \forall l = 0 \text{ to } k \end{cases}$$
(10)

Then, we see that the kriging system relates the covariance between sample points $C_{ij} = E\{Z(X_i)Z(X_j)\}$, the covariance between each sample *i* and estimated points $C_i = E\{Z(X_i)Z(X)\}$ and the unknown weights λ_i . Written in matrix form the system Eq. (10) is:

$$\begin{pmatrix} C_{ij} & f^{l}(X_{i}) \\ & & \\ f^{l}(X_{j}) & 0 \end{pmatrix} \begin{pmatrix} \lambda_{1} \\ \vdots \\ \lambda_{n} \\ \mu_{0} \\ \vdots \\ \mu_{k} \end{pmatrix} = \begin{pmatrix} C_{1} \\ \vdots \\ C_{n} \\ f^{1}(X) \\ \vdots \\ f^{k}(X) \end{pmatrix}$$
(11)

This system Eqs. (10/11) is known as the primal kriging system. In our study (large data sets), we want to use kriging as a local interpolation method. Consequently, it is more appropriate to solve the dual kriging system Eqs. (12/13), which is strictly equivalent to the former but it allows us to use directly Z_i values instead C_i ; see for example LeRroux *et al.* (1997). So, Eq. (4) can be rewritten with $\hat{Z}_K(X)$ as the sum of two terms (drift + fluctuation):

$$\hat{Z}_{K}(X_{i}) = \sum_{l=0}^{k} a_{l} f^{l}(X_{i}) + \sum_{j=1}^{n} \theta_{j} C_{ij}$$
(12)

or in matrix form:

$$\begin{pmatrix} C_{ij} & f^{I}(X_{i}) \\ & & \\ f^{I}(X_{j}) & 0 \end{pmatrix} \begin{pmatrix} \theta_{1} \\ \vdots \\ \theta_{n} \\ a_{0} \\ \vdots \\ a_{k} \end{pmatrix} = \begin{pmatrix} \hat{Z}_{1} \\ \vdots \\ \hat{Z}_{n} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$
(13)

where the unknowns a_l and θ_j replace μ_l and λ_j . Thus, at every data location X_i , we have the value of the interpolated ozone data which is the true ozone data

value, i.e., $\hat{Z}_K(X_i) = Z(X_i)$. To summarize, in this approach, the dual kriging system Eqs. (12–13) is solved, i.e., the covariance matrix is inverted to solve for the weight θ_j and the coefficient a_i . Consequently, the interpolation at every location X (on a regular grid) can be further realized with the same system.

4 The covariance function: generalized covariance and variogram

To solve the dual kriging system Eq. (12) we need to know the covariance matrix C_{ij} . At this stage, we have to introduce the "intrinsic hypothesis" that the covariance between the sample points $Z(X_i)$ and $Z(X_j)$ depends only on the distance *h* between the two points X_i , X_j . In this study, we calculate the geodesic distance *h* in degrees.

$$h = |X_i - X_j| \tag{14}$$

Our main approach is to assume that no trend can be detected in our daily satellite ozone data which is not true at a seasonal scale, see e.g. DeLuisi *et al.* (1994) and (Niu *et al.* 1992). Therefore, we mainly deal with ordinary kriging which involves a constant drift and then use "semi-variogram" $\gamma(h)$ models to fill the covariance matrix C_{ij} . A variogram is used to model the way two values in space or time are correlated. It is related to the covariance C(h) by the following equality (Christakos, 1984):

$$\gamma(h) = C(0) - C(h) \tag{15}$$

where C(0) is the covariance at zero lag (h = 0) or the variance (σ^2) , also called "sill" in geostatistics. In fact, we calculate the "sample semi-variogram" $\hat{\gamma}(h)$ which is an estimator of the semi-variogram $\gamma(h)$ defined as follows:

$$\gamma(h) = \frac{1}{2} E\{ [Z(X) - Z(X+h)]^2 \} \quad . \tag{16}$$

The sample semi-variogram is calculated from the mean of the squared differences between two measurements $z(X_i)$ and $z(X_i + h)$:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(X_i) - z(X_i + h)]^2$$
(17)

where N(h) is the number of point-pairs separated by h. When sample isotropic semi-variograms are calculated, they are fitted from theoretical 1D semi-variogram models that become a prerequisite for the covariance matrix and then for kriging interpolation. Some theoretical models have been proposed in the literature. The most currently used (Journel and Huijbregts, 1989) are listed in Table 2. All these models are usually described using three parameters.

1. C_0 is the error variance at every data point, it is called the "nugget" effect. In our case, it could be interpreted as the introduction of a sampling error at the point measurement. By definition, these noises are without biases (white noise) and uncorrelated. At zero

Table 2. Semi-variogram models

1D theoretical semi-variogram models Spherical C_0 $C_0 + C_1[1.5(h/R) - 0.5(h/R)^3]$ $C_0 + C_1$	if $h = 0$ if $0 < h < R$ if $h \ge R$
Exponential C_0 $C_0 + C_1[1 - \exp(-3h/R)]$ $C_0 + C_1$	if $h = 0$ if $0 < h < R$ if $h \ge R$
Gaussian C_0 $C_0 + C_1[1 - \exp(-(h/R)^2)]$ $C_0 + C_1$	if $h = 0$ if $0 < h < R$ if $h \ge R$

 C_0 is the nugget effect, $C_0 + C_1$ corresponds to the asymptotic variance (σ^2), *h* is the separation distance and *R* the range of influence

lag (h = 0), no nugget effect leads to $\gamma(0) = 0$ or C(0) = 1 which is assumed throughout this study.

- 2. *R* is called the "range", it is the lag distance at which all successive values are independent of each other. Beyond this range, points are uncorrelated.
- 3. $C_0 + C_1$ corresponds to the asymptotic variance (σ^2) or the "sill" (maximum semi-variogram value).

Figure 3a is an example of a sample isotropic semivariogram (crosses) calculated from SBUV data (total ozone). This semi-variogram calculation can be made with the different lag increment h. We define the increment as the minimum distance found in the dataset in Fig. 3a. Results are similar with larger increments but they are smoother as in Fig. 3b, c. We can distinguish several peaks, which correspond to nested structures. Here, by analogy with structural analysis in geostatistics, the existence of different scales can reasonably be established. Moreover, these nested structures can be conveniently represented as the sum of a number of variograms, instead of using only one (Journel and Huijbregts, 1989). If we consider the smallest separation distances, such as for the event of October 19, 1988, (i.e., from $h = 1.85^{\circ}$ to $h \approx 20^{\circ}$), the point measurements are correlated since they exhibit less than the experimental dispersion variance σ^2 , which is equal here to 3253. If we use a gaussian model as in Fig. 3a, with $C_0 = 0$, $C_1 = 3350$ and R = 13, we find that it is a good fit. Consequently, for the total ozone field, a correlation length of 12° – 15° (\approx 1300 km), which is of the order of magnitude of the synoptic scale (≈ 1000 km), is relevant. Furthermore, for h between approximately 60° and 120°, point measurements are also correlated, since $\gamma(h)$ is also below the sill (σ^2). If we look at the region between approximately 30°S and 60°S, it corresponds to the region of ozone maxima along the polar vortex during Austral spring (see Fig. 2). The two latitudes circles 30° and 60° correspond to the two previous separation distances $h = 60^{\circ}$ and 120° . Then, not only did we find correlated measurements at the synoptic scale without privileged direction, but we also found zonal correlated measurements on the equatorial side of the polar Austral vortex. This last point shows that one



Fig. 3. a Isotropic semi-variogram calculated from the total ozone field provided by the SBUV dataset and fitted with the gaussian model (*solid line*) ($C_0 = 0$, $C_1 = 3350$ and R = 13), October 5, 1988. **b** Isotropic semi-variogram calculated from ozone concentration of layer 5 provided by the SBUV dataset, October 5, 1988. **c** Isotropic

of the main characteristic of the total ozone distribution at this period, i.e., ozone maxima located on the equatorial side of the vortex can be proven to be correlated with the variogram analysis. Fig. 3b displays the semi-variogram calculated from layer 5. The shape of this variogram is different from the one shown in Fig. 3a. The same peaks exist, but are less marked and particularly so at 40°. As for the previous case, a periodicity of 25° shows the influence of satellite tracks. Whereas in Fig. 3a the dynamic influence of the vortex showed up, this is not the case for layer 5 since ozone seems more homogeneously distributed (gaussian shape). We can also fit this semi-variogram with a single polynomial of the third order.

Semi-variograms have been computed for all layers, and four classes can be distinguished. Semi-variograms of layer 3 and 4 have the same shape as the total ozone field. Those of layers 2, 6 and 7 have the same shape as those of layer 5. Except for layer 8, where zonal correlated measurements are present with two marked peaks centered at 40° and 140° (Fig. 3c), the other uppermost layers (9 to 12) that are not shown here have



semi-variogram calculated from ozone concentration of layer 8 provided by the SBUV dataset, October 5, 1988. **d** East–west and north–south semi-variograms calculated from total ozone field by the SBUV dataset, October 5, 1988

a marked continuous sill from 40° to 140° . Apparently, this last shape of variogram indicates no particularly correlated measurements for separation distance over 20° . Nevertheless, these layers are of lesser importance due to the fact that the ozone concentration at these levels is much smaller compared to lower levels.

We have chosen to compute isotropic semi-variograms even though we are aware of this limitation. Indeed, atmospheric ozone data do not have homogeneous and isotropic second-moment statistics. So, it is possible to distinguish two main directions, such as north-south and east-west, or to split our domain into sub-domains (latitude-dependent) to account for the non-homogeneity of the statistics. Figure 3d presents the east-west and north-south semi-variograms of the previously illustrated total ozone field. These variograms have been calculated by using four north-south and east-west quadrants of 45° for every data points. The resulting shape of the east-west semi-variogram in Fig. 3d is close to the isotropic one (Fig. 3a) whereas the north-south one has a different shape with is a slight shift to smaller separation distances. This means that

SBUV data have higher variability in the north-south direction (roughly along satellite tracks), which is generally true for the total ozone field. As for the isotropic semi-variograms of layers 3, 4 and the integrated column, there is a strong peak at 40° for the east-west semi-variogram. This corresponds to a length scale of approximately 4000 km. Sirovich et al. (1995) have also observed this length scale using Karhunen-Loève empirical eigenfunctions. From physical arguments based on the analogy with the Von Karman vortex trail, they showed that this length scale is related to the baroclinic instability, which is characterized by a pattern of waves numbering about seven. In our case, this could be associated with a zonal wave (wave number = 7), but it could also be viewed as a planetary wave. They claim, from enstrophy cascade considerations, that dynamic activity leads to this length scale at which mechanical energy is supplied. The high variations in the north-south variogram are due to the proximity of the satellite tracks at the pole, and there is no further influence from the vortex. Thus, both isotropic and anisotropic semi-variograms can provide information about the variability of the total ozone field. In terms of interpolation by kriging, this does not change the results significantly. Indeed, in fitting eastwest or north-south semi-variograms with our theoretical 1D semi-variogram models, the changes are not large. Despite the limitations of those models, we show that in our case this previous approach had some advantages over the covariance function used with universal kriging.

Indeed, in the case of universal kriging, we use the covariance function or "generalized covariance" K(h) = C(h) of the random function Z(X) of order k, which was developed for geostatistical inference in kriging applications (Matheron, 1973). We shall take the first approach, following the work of LeRoux *et al.* (1997), using simple polynomial models of generalized covariance K(h), namely h, h^3 , h^5 and $h^2\ln(h)$. These covariances were mainly developed for Euclidean geometry. In the next section, we shall use different degrees of drift, linear, quadratic and cubic degrees of fixed drift (Eq. 3) to show that is not really appropriate for our global daily data.

5 Validation

5.1 Cross-validation

We have deliberately omitted a consecutive group of SBUV measurements (October 5, 1988) in order to further interpolate and so fill in these missing data points (\approx 13). The size of the gaps was chosen to be approximately equal to the spacing between orbit tracks (\approx 25°). These gaps were repeated for every satellite track at all levels between 45°S and 70°S. Only the results of layer 5 and the total column are presented, but these can be considered representative of most layers over different days. Figure 4a, b displays the error between SBUV data and those found by interpolation,

Fig. 4. a Biases between SBUV data and those found by interpolation with kriging and by bilinear interpolation from Delaunay triangulation versus satellite tracks. Biases are normalized by the original standard deviation. Results correspond to areas of gaps in the 14 satellite tracks for the total ozone field on October 19, 1988. b Biases between SBUV data and those found by interpolation with kriging and by bilinear interpolation from Delaunay triangulation versus satellite tracks. Biases are normalized by the original standard deviation. Results correspond to areas of gaps in the 14 satellite tracks. Biases are normalized by the original standard deviation. Results correspond to areas of gaps in the 14 satellite tracks for layer 5 on October 19, 1988

normalized by the standard deviation of the SBUV measurements, versus the 14 satellite tracks. These were found using ordinary kriging by applying the exponential model (see Fig. 3b), which provides better results than the others (3rd polynomial model included), and using a bilinear interpolation from a Delaunay triangulation of the data. In both cases, the two curves have the same shape, which means that the position of satellite tracks influences both interpolation methods. The interpolation with kriging gives significantly lower errors in more than 75% of cases, even when using a 1D isotropic semi-variogram. There are more results smaller than ± 1 standard deviation (between the two dotted lines) for kriging than for bilinear interpolation for both cases, and 70% of errors are smaller than ± 1 standard deviation, which can be regarded as accurate results. Figure 5 displays the relative mean error (%) versus the mean separation distance \bar{h} for layer 5 and for total ozone. The relative errors are quite close for the



Kriging

Bilinear

Satellite Tracks

4

3

2

3.5

2.5

1.5

1

0.5

-0.5

-1

-1.5

0

error/σ₀



Fig. 5. Relative biases between SBUV data and those found by interpolation with kriging and by bilinear interpolation from Delaunay triangulation versus the mean separation distance h. Results are averaged over gaps of the 14 satellite tracks for the total ozone field and for layer 5 on October 19, 1988

two interpolation methods. The error is higher for the total ozone field (5–13%) than for layer 5 (5–10%) for $3 \le h \le 12$, and lesser ($\approx 4\%$) for $h \le 3$. In both cases, kriging leads to a smaller error.

5.2 Validation with SAGE II ozone profiles

For October 5, 1988, the SAGE II ozone profiles dataset contains only 30 ozone profiles: 15 in the Southern Hemisphere at the same latitude (between 71.7°S and 71.9°S) and 15 in the Northern Hemisphere (between 21.5°N and 26.9°N) (see Fig. 1). Universal Kriging (*linear drift and* $K(h) = h^2 \ln(h)$) and ordinary kriging (exponential model) were used to interpolate SBUV ozone concentrations in volumetric mixing ratios at 17 pressure levels (100 mbar–0.1 mbar) in order to match the data point measurements of SAGE II. A detailed comparison of average SBUV and SAGE II ozone profiles has already been performed by McPeters et al. (1994) during the period from October 1984 to June 1990. These comparisons were performed based on Umkehr layer ozone profiles in Dobson units to determine whether SBUV and SAGE II yielded consistent ozone trends, so our comparison is noticeably different. Ours is a local daily comparison, and for simplicity, volumetric mixing ratios provided by both datasets were chosen instead of number density concentrations. Figure 6a, b displays ozone profiles for, respectively, the nearest (0.51°) and the farthest (2.16°) of the SBUV point measurements in the Southern Hemisphere (presence of the polar vortex). The two kriging methods are in agreement except (Fig. 6b) in the region where the ozone mixing ratio is at a maximum (10 mbar-4 mbar), and (Fig. 6a) where the values found with universal kriging are closest to those of SAGE II. Moreover, the main discrepancies between SBUV and SAGE II are in this same region, and between 70 mbar and 40 mbar. These two figures show that the fact that one SBUV

point measurement is closer does not influence the accuracy of the interpolation. At other pressure levels, the interpolated mixing ratios match those of SAGE II. Equivalent ozone profiles for the Northern Hemisphere (tropics) are shown in Fig. 6c, d, respectively the closest (0.8°) and one of the farthest (5.2°) SBUV point measurements. The interpolated SBUV data points match those of SAGE II for the farthest distance to the real SBUV data point, whereas there is a larger error in the region of ozone mixing ratio maxima (20 mbar to 4 mbar) toward the nearest distance. Due to the lack of resolution, SBUV cannot reproduce variations near the maximum of the ozone-mixing ratio. In fact, the biases between the two types of profile can be attributed to measurement errors and/or profile-derivation errors as much as to the interpolation itself. These results corroborate what we had found previously. Indeed, the relative error due to the interpolation is smaller than 6%. which is on the order of, but even less than, the error between SAGE II and Umkehr ozone profiles found by Newchurch et al. (1995).

5.3 Validation with the TOMS ozone field

Our method consisted of interpolating every Umkehr layer which has 1098 data points (October 19, 1988) on the TOMS equidistant grid (resolution 288 $(1^{\circ}) \times 180 (1.25^{\circ})$) using both the universal and ordinary kriging interpolation methods described previously. Next, we integrated over all layers in order to compare with TOMS measurements. In addition, we also directly interpolated total ozone amounts provided by SBUV. Table 3 displays the statistical results of our computation for a fixed linear drift with different covariance models (universal kriging) and two variogram models (ordinary kriging). The coefficient of correlation and the standard deviation of the error between the two 2D ozone fields were measured, i.e., between TOMS data and integrated kriged SBUV data.

When universal kriging is used, the two generalized covariance models are very close (h and $h^2 \ln(h)$), whereas h^3 and h^5 are not well adapted to our dataset. This result is consistent with the shape of $\gamma(h)$ where its variations from the origin to 25° were closer to being linear than polynomial (see Fig. 3a–d). Different degrees of drift have also been tested, but this does not alter the results. This means that the data have no drift in mean. It is also a sign to use ordinary kriging (quasi-stationary or stationary random function) in this case, since no drift (constant mean) seems required. Moreover, the isotropic variogram of Fig. 3a does not present parabolic variation for a large separation distance h, which usually indicates the absence of drift (Journel and Huijbregts, 1989). Finally, a global correlation close to 98% was found between the two total ozone fields. These very good statistical results were found using ordinary kriging with a larger number of points since the pole areas are covered. A theoretical exponential model and a polynomial (third order) fit were used and give same results.





SAGE II

SBUV (OK) SBUV (UK)

Fig. 6. a Mixing ratio ozone profiles of SAGE II (*line*) in the Southern Hemisphere compared to those obtained with interpolation of SBUV, ordinary kriging (+) and universal kriging (*). The nearest data point of SBUV is at $h = 0.51^{\circ}$. Latitude $= -71.7^{\circ}$, longitude $= 60.8^{\circ}$, October 5, 1988. **b** Mixing ratio ozone profiles of SAGE II (x) in the Southern Hemisphere compared to those obtained with interpolation of SBUV, ordinary kriging (+) and universal kriging (*). The nearest data point of SBUV is at $h = 2.16^{\circ}$. Latitude $= -71.9^{\circ}$, longitude $= 156.4^{\circ}$, October 5, 1988. **c** Mixing ratio ozone

profiles of SAGE II (x) in the Northern Hemisphere compared to those obtained with interpolation of SBUV, ordinary kriging (+) and universal kriging (*). The nearest data point of SBUV is at $h = 0.8^{\circ}$. Latitude = 22.7°, longitude = 2.9°, October 5, 1988. **d** Mixing ratio ozone profiles of SAGE II (x) in the Northern Hemisphere compared to those obtained with interpolation of SBUV, ordinary kriging (+) and universal kriging (*). The nearest data point of SBUV is at $h = 5.2^{\circ}$. Latitude = 21.5°, longitude = 75.4°, October 5, 1988

Table 3. Statistical results obtained by kriging of SBUV data using both ordinary (two different variogram models) and universal kriging (fixed drift: linear with different generalized covariances). Comparison is done with TOMS data (October 19, 1988)

	Mean (O3)	Maximum (O3)	Standard deviation	Coefficient of correlation
TOMS	299.8 ^a	517.0 ^a		
Universal kriging Covariances models				
h	298.7	509.	52.8	0.967
h^3	300.9	515.	62.5	0.842
$h^2 \ln(h)$	298.8.	513.	54.	0.966
h^5	305.8	1855.	380.	0.161
Ordinary kriging Variogram models				
Exponential	296.8	507.0	49.9	0.972
Polynomial (3rd order)	295.1	507.7	53.3	0.973

^a In Dobson Units

6 Results and comparisons

6.1 Two-dimensional total ozone fields

For October 5, 1988 (Fig. 2), a large area of TOMS data is missing. We have previously demonstrated that kriging is an accurate interpolator, so we can reasonably interpolate the SBUV data measurements on the TOMS grid. The corresponding kriged SBUV ozone field found using the exponential model is shown in Fig. 7. The two ozone fields are very similar, and the kriging has been able to fill in holes where there was missing data in the TOMS ozone field, even at the poles. Maxima and minima are located at the same place in both hemispheres. The main difference is at the edge of the polar vortex (\sim between 40°S and 65°S) where kriging has not been able to reproduce filamentation. Consequently, the interpolated ozone field is relatively smooth compared to the original TOMS ozone field. The best agreement is found in tropical regions where horizontal ozone gradients are weaker.

6.2 Two-dimensional ozone fields at different pressure levels

Figure 8a, b represents the 2D ozone field obtained using ordinary kriging with the exponential model at two different pressure levels, 89.6 mbar (layer 3) and 11.2 mbar (layer 6). These two ozone fields are quite different, like their semi-variogram. Maxima are not located at the same latitude. At 89.6 mbar, relative maxima of ozone concentration are located along the vortex edge at middle latitudes in the Southern Hemisphere, and over the pole (high latitudes) in the Northern Hemisphere. There is a large zone over the tropics where ozone concentration is minimal. The ozone distribution of this layer is very similar to that of the total integrated column with maxima along the polar vortex.

At 11.2 mb, relative maxima are found on each side of the equator, extending to the tropical regions (\sim between 30°S and 30°N). Relative minima are located over the poles, with a large ozone minimum in the Northern Hemisphere. The vortex is not as well defined as in the previous figure.

Finally, the vertical cross-section of ozone from the North Pole to the South Pole at 0° longitude from the kriged SBUV dataset is shown in Fig. 9. This is representative of the longitudinal mean ozone concentration shown by Holton (1979). We observe a relative maximum in tropical latitudes where convection is strong and ozone production is maximal. A maximum at high latitudes is observed, but for a lower altitude. This is due to the strong poleward and downward mass flow (Holton, 1979). The ozone is produced at the tropics and transported to middle and high latitudes by this meridional flow.

7 Discussion

We have shown that it is possible to reproduce a daily three-dimensional ozone field using only the SBUV nonequally spaced data. Despite the fact that kriging does not fully capture the smallest structures, such as filaments, the final picture still appears realistic in terms of total ozone concentration maps. Even if the daily











Fig. 9. Latitude-altitude cross section of the total amount of ozone in Dobson Units at 0° longitude on October 5, 1988, obtained using ordinary kriging. Contour interval is 5 DU

comparison of ozone profiles with SAGE II is not exhaustive, it provides an indication of the quality of our kriging interpolation. Our approach shows that kriging used as an interpolation method is quite efficient, with a mean relative error of less than 10% for a separation distance *h* between two points ranging from 1.5° to 12° . Results with kriging show that the interpolation error is less than ± 1 standard deviation in more than 70% of cases.

Obviously, this statistical approach was a first step and it cannot replace methods such as trajectory mapping or data assimilation. Indeed, SBUV has only 14 nadir measurements around a fixed latitude circle, so any daily disturbance with a zonal wave greater than 7 will always be aliased. This is important for midlatitudes where baroclinic disturbances of synoptic scale occur. Nevertheless, the statistical inference using isotropic and anisotropic 1D semi-variograms provides information about spatial variability, and showed a signal at $\approx 40^{\circ}$, which has already been found from TOMS ozone fields as significant scale to explain energetic characteristics see (Sirovich et al., 1995). In addition, atmospheric ozone data do not have homogeneous and isotropic second-moment statistics, as the 1D variogram approach assumes. Thus, we could improve on this point by introducing the 2D notion of the semivariogram, which reveals the degree and direction of anisotropy (Nielsen, 1993). We know that a data assimilation method would give more informations. Although the kriging interpolation presents some similarities with data assimilation methods such as optimal interpolation (OI) or more generally Kalman filter, it is a simpler technique. Data assimilation is more complex since it requires an a priori state estimate or background field. This background field can be obtained either with a theoretical linear model or with climatology fields. Moreover, to further improve the kriging interpolation

we could also use such climatologies to infer more appropriate trends. It so happens that measurements are biased. Thus, a systematic bias could be introduced in the observation measurements by including the nugget effect ($C_0 \neq 0$). Co-kriging could also be implemented to interpolate multivariate measurements. In our case this would correspond to interpolating the same variable (SBUV measurement) at different times, in both forward and reverse time, as in the spatial statistical interpolation scheme of Rogers (1977) which has been used by Randel and Wu (1995) to provide a useful ozone climatology.

This attempt to retrieve a correct ozone distribution was motivated by the need for better information on the vertical structure of ozone. This kriging interpolation method provides good knowledge of the behaviour of the parameter being interpolated, and thus can be considered as a first step toward being the method closest to data assimilation. Not only could we improve our understanding of stratospheric and tropospheric ozone in regions where measurements from the ground are missing, but also the resulting kriged ozone field can be used as the most realistic initialization point for further data assimilation or for more physical studies concerning the behaviour of the ozone transport in the stratosphere (Vincent and Tranchant, 1999).

Acknowledgements. SBUV data were obtained from NOAA/NES-DIS, with support from the NOAA Climate and Global Change Program Atmospheric Chemistry Element. We would like to thank Daniel LeRoux for his help in implementing the Kriging algorithm. Delaunay triangulation and linear interpolation were performed using MATLAB 5.2. Thanks also to the two anonymous reviewers for helpful suggestions.

Topical Editor D. Murtagh thanks D.W. Tarasick and another referee for their help in evaluating this paper.

References

- Bergeron, S. Y., A. V. Vincent, D. A. Yuen, B. J. S. Tranchant, and C. Tchong, Viewing seismic velocity anomalies with 3-D continuous gaussian wavelets, *Geophys. Res. Lett.*, 26(15), 2311, 1999.
- Bhartia, P. K., R. D. Peters, C. L. Mateer, L. E. Flynn, and C. Wellemeyer, Algorithm for the estimation of vertical ozone profile from the backscattered ultraviolet (BUV) technique, J. Geophys. Res., 101, 18 793–18 806, 1996.
- Bojkov, R. D., and V. E. Fioletov, Estimating the global ozone characteristics during the last 30 years, *J. Geophys. Res.*, 100, 16 537–16 551, 1995.
- Chandra, S., and R. D. McPeters, The solar cycle variation of ozone in the stratosphere inferred from Nimbus 7 and NOAA 11 satellites, J. Geophys. Res., 99, 20 665–29 671, 1994.
- Christakos, G., On the problem of permissible covariance and variogram models, *Water Resources Res.*, **20**, 251–265, 1984.
- Daley, R., Atmospheric data analysis, Cambridge University Press, New York, 1991.
- Delfiner, P., and J. P. Delhomme, Optimum interpolation by kriging, *Proc. NATO ASI, Display and analysis of spatial data*, Eds. J. C. Davis and M. J. McCullagh, 96–114, John Wiley, New York, 1973.
- De Luisi, J. J., C. L. Mateer, D. Theisen, P. K. Bhartia, D. Longenecker, and B. Chu, Northern middle-latitude ozone

profiles features and trends observed by SBUV and Umkehr, 1979-1990, J. Geophys. Res., 99, 18 901–18 908, 1994.

- Eichmann, K.-U., K. Bramstedt, M. Weber, V. Rozanov, R. deBeek, R. Hoogen, and J. P. Burrows, ozone profile retrieval from GOME satellite data: validation and applications, *Proc. 3rd ERS Symposium, Space at the service of our environment*, Florence, Italy, 17–21 March 1997.
- Fleig, et al., Nimbus 7 solar backscatter ultraviolet (SBUV) ozone product user's guide, NASA Ref. Publ., 1234, 117 pp, 1990.
- Gleason, J. F., and R. D. McPeters, Corrections to the Nimbus 7 solar backscatter ultraviolet data in the "non-sync" period (February 1987 to June 1990), J. Geophys. Res., 100, 16 873– 16 877, 1995.
- Holton, J. R., An introduction to dynamic meteorology, vol 23 of Internal geophysics series, Academic Press, New York, 1979.
- Journel, A. G., and J. Huijbregt, *Mining geostatistics*, 600 pp, Academic Press, London, 1989.
- Kidder, S. Q., and T. H. Vonder Haar, Satellite meteorology, an introduction, 466 pp, Academic Press, San Diego, 1995.
- Krige, D. G., A statistical approach to some basic mine valuation problems on the Witwatersrand, J. Chem. Metall. Min. Soc. S. Afr., 52, 119–139, 1951.
- Lefohn, A. S., and V. C. Runeckles, Establishing a standard to protect vegetation-ozone exposure/dose considerations, *Atmos. Environ.*, 21, 561–568, 1987.
- Lefohn, A. S. W. Jackson, D. S. Shadwick, and H. P. Knudsen, Effect of surface ozone exposures on vegetation grown in the southern Appalachian Mountains: identification of possible areas of concern, *Atmos. Environ.*, 31(11), 1695–1708, 1997.
- LeRoux, D. Y., C. A. Lin, and A. Staniforth, An accurate interpolating scheme for semi-Lagrangian advection on an unstructured mesh for ocean modelling, *Tellus*, **49A**, 117–138, 1997.
- Mateer, C. L., and J. J. DeLuisi, A new Umkehr inversion algorithm, J. Atmos. Terr. Phys., 54, 537–556, 1992.
- Matheron, G., Les Variables régionalisées et leur estimation, Masson, Paris, 1965.
- Matheron, G., The intrinsic random functions and their applications, Adv. Appl. Prob., 5, 439–468, 1973.
- McPeters, R. D., and G. J. Labow, Assessment of the accuracy of 14.5 years of Nimbus 7 TOMS version 7 ozone data by comparison with the Dobson network, *Geophys. Res. Lett.*, 23, 3695–3698, 1996.
- McPeters, R. D., T. Miles, L. E. Flynn, C. G. Wellemeyer, and J. M. Zawodny, Comparison of SBUV and SAGE II ozone profiles: implications for ozone trends, J. Geophys. Res., 99, 20 513–20 524, 1994.
- McPeters, R. D., P. K. Bhartia, A. J. Krueger, J. R. Herman, B. M. Schesinger, C. G. Wellenmeyer, C. J. Seftor, G. Jaross, S. L. Taylor, T. Swissler, O. Torres, G. Labow, W. Byerly, and R. P. Cebula, Nimbus-7 total Ozone Mapping Spectrometer (TOMS) data products user's guide, *NASA Ref. Publ.*, 1384, April, 1996.
- Mégie, G., Stratosphère et couche d'ozone, 108 pp, Masson, Paris, 1992.
- Morris G. A., et al., Trajectory mapping and applications to data from the Upper Atmosphere Research Satellite, J. Geophys. Res., 100, 16 491–16 505, 1995.
- Munro, R., R. Siddans, W. J. Reburn, and B. J. Kerridge, Direct measurement of tropospheric ozone distributions from space, *Nature*, 392, 168–171, 1998.
- Newchurch, M. J., D. M. Cunnold, and H. J. Wang, Stratospheric aerosol and gas experiment II-Umkehr ozone profile comparisons, J. Geophys. Res., 100, 14 029–14 042, 1995.
- Nielsen, A. A., 2D semivariograms in Ed. Cilliers, P. Proc. Fourth South African Workshop on Pattern Recognition, 25–35. Simon's Town, Africa, 1993.
- Niu, X., et al., Trends in column ozone based on TOMS data: dependence on month, latitude, and longitude, J. Geophys. Res., 97, 14 661–14 669, 1992.

- Pellerin S., P. A. Michelangeli, B. Tranchant, P. Gauthier, and G. Brunet, Assimilation of TOVS total ozone data: impact on the forecasts and analyses, *Proc. Third WMO Symposium on data assimilation*, Québec City, 7–11 June 1999.
- Randel, W. J., and F. Wu, Climatology of stratospheric ozone based on SBUV and SBUV/2 Data:1978–1994. NCAR Technical Note, NCAR/TN-412+STR, National Center For Atmospheric Research, Boulder, Colorado, USA, April, 1995.
- Ripley, B. D., Spatial statistics, 252 pp, John Wiley, New York, 1981.
- Rogers, C. D., Statistical principles of inversion theory, in *Inversion methods in atmospheric remote sounding*, Ed. A. Deepak, pp 117–138, Academic Press New York, 1977.
- Sirovich, L., R. Everson, and D. Manin, Turbulent spectrum of the Earth's ozone field, *Phys. Rev. Lett.*, **74**, 2611–2614, 1995.
- Vincent A. P., and B. J. S. Tranchant, Anisotropic turbulent diffusion for ozone transport at 520 K, J. Geophys. Res., 104(D22), 27 203, 1999.