An EOF study of the Earth's ozone field

D.Manin, R. Everson, and L. Sirovich

The Rockefeller University, New York City, 10021

Abstract

In a previous paper we discussed the spectral properties of the Earth's ozone layer, obtained using Empirical Orthogonal Function decomposition of the Total Ozone Mapping Spectrometer (TOMS) database. Here we present other aspects of the analysis, including the EOF method adapted for incomplete datasets, analysis of spatial structure and temporal variation of first several eigenfunctions, and an extended study of small-scale properties of ozone concentration fields.

Geophysical datasets collected by satellite-based devices currently provide a wealth of information which needs to be assimilated and interpreted. In this paper we analyze the global ozone concentration fields measured by Total Mapping Ozone Spectrometer [1]. The interest in atmospheric ozone is due to both its environmental role (ozone hole phenomenon [2, 3]), and the fact that it can be regarded as a passive tracer, thus providing an additional insight in the properties of atmospheric circulation. This is an extended version of a previous paper [4], where the principal attention was paid to the spectral properties of the ozone concentration fields.

The TOMS device was installed on a satellite, Nimbus, which completed sixteen rotations in a 24-hour period, on a south-north synchronous orbit. Thus, every global field (see example in Fig. 1) consists of 14 separate records taken roughly 1.5 hours apart. Every field is represented on a 288×180 grid, so the resolution is order 100 km. The measurement method is based on the absorption by ozone of light reflected from the surface, so it fails in the polar night regions (a black circle around North Pole in Fig. 1). Occasional technical failures also lead to missing pixels or whole records, as also can be seen in Fig. 1. The TOMS database contains daily ozone fields for the period of 13 years, which amounts to 270 Gigabytes of data. Due to computational requirements, we analyzed the subset consisting of 4 snapshots per month (approximately weekly).

The Karhunen-Loeve (KL) procedure [5, 6, 7] is ideally suited for analyzing massive spatiotemporal datasets, due to its ability to extract persistent features. In particular, the snapshot method [8] considerably reduces the needed computation effort. However, the incompleteness of data required modification of the methodology. We discuss the numerical procedure in Section 1. Section 2 is devoted to the large-scale temporal and spatial properties of the ozone field. In Section 3 we report the results on statistical properties on the smaller scale.

543

1 KL procedure for gappy data

The Karhunen–Loeve procedure, applied to the array of data $c(x,t)$, is aimed at finding the best sum-of-products representation

$$c(x,t) = \sum_k \phi_k(t)\psi_k(x) \tag{1}$$

The variational problem of minimizing the net error

$$\Delta = \left| c - \sum_k \phi_k(t)\psi_k(x) \right|^2 \to \min \tag{2}$$

reduces to the linear eigenproblem

$$\hat{K}_x\psi_k = \lambda_k\psi_k, \quad \phi_k = (\psi_k, c)_x \tag{3}$$

where round brackets denote averaging (inner product) in x: $(f,g)_x = \int fg\,dx$; and $\hat{K}_x$ is the spatial correlation operator:

$$K_x(x,y) = (c(x,t), c(y,t))_t, \quad \hat{K}_x\psi = (K_x(x,y), \psi(y))_y \tag{4}$$

The idea of the snapshot method [8] is that because of the symmetry between x and t in this formulation, one can instead solve the eigenproblem for ϕ_k

$$\hat{K}_t\phi_k = \lambda_k\phi_k, \quad \psi_k = (\phi_k, c)_t \tag{5}$$

where $\hat{K}_t$ is the temporal correlation operator. For practical applications, where both x and t are discrete variables, advantage of choosing the appropriate alternative can be essential. Typically with image data (when t enumerates images, and x indexes pixels), the number of images can be much smaller than the number of pixels in each image, and consequently a large gain in performance can be achieved by solving the eigenproblem for the $\hat{K}_t$ matrix, rather than for the much larger $\hat{K}_x$.

For discrete problems, the KL procedure produces the *complete basis* (ϕ_k, ψ_k), so that the net error (2) is zero. The total number of eigenfunctions required is the smaller of the number of snapshots and the number of pixels. Moreover, if the eigenfunctions are numbered in the order of decreasing energy, $\lambda_0^2 < \lambda_1^2 < \ldots < \lambda_k^2 < \ldots$, then every sub-basis comprising the first n eigenfunctions is the best of all n-dimensional bases.

Unfortunately, the problem is much more complicated, and the solution is much less elegant, when the data is partially missing. Denote by $D(x,t)$ the characteristic function (with values from $\{0,1\}$) describing the domain where the data is known. To generalize the optimization problem (2), we should consider integrating over the domain where the data is known:

$$\Delta = \int D(x,t)\left(c(x,t) - \sum_k \phi_k(t)\psi_k(x) \right)^2 dx\,dt \to \min \tag{6}$$

It can be easily seen that this problem has infinite number of solutions with $\Delta = 0$. Indeed, if we fill in the gaps in data with arbitrary values and apply the classical KL procedure,

we will obtain the basis which represents the filled-in snapshots exactly everywhere, and in particular in regions with authentic data. This means that the KL procedure does not provide a preferable way to fill in the missing data points in the generic situation.

The approach we choose instead is based on the fact that KL eigenfunctions can be actually found independently of each other. Namely, the pair (ϕ_0, ψ_0) provides the best approximation $c(x, t) \approx \phi(t)\psi(x)$, the pair (ϕ_1, ψ_1) optimally approximates the *reduced* dataset $c_1(x, t) = c(x, t) - \phi_0(t)\psi_0(x)$, etc. In this way, we seek the k-th eigenfunction for gappy data by solving

$$\Delta = \int D(x, t)\Big(c_{k-1}(x, t) - \phi_k(t)\psi_k(x)\Big)^2 dx\, dt \rightarrow \min \tag{7}$$

with

$$c_k(x, t) = c(x, t) - \sum_0^{k-1} \phi_k(t)\psi_k(x) \tag{8}$$

Variational differentiation of (7) over ϕ and ψ gives a pair of nonlinear equations

$$\phi(t) = \frac{\int D(x', t)c(x', t)\psi(x')\, dx'}{\int D(x', t)\psi(x')^2\, dx'} \tag{9}$$

$$\psi(t) = \frac{\int D(x, t')c(x, t')\phi(t')\, dt'}{\int D(x, t')\phi(t')^2\, dt'} \tag{10}$$

Note that when all data is known, $D(x, t) \equiv 1$, and the denominators are constants (rather than functions of x or t), which reduces the task to solving the classical linear eigenvalue problem. Although the nonlinear problem (9,10) cannot be solved directly, it is well suited for solution by iterations (the next approximation being computed by substituting the previous one in the RHS). We cannot prove the convergence, but since the problem is in a sense close to the linear one, for which iterations do converge to the eigenvector with the greatest real part of eigenvalue, we can expect the similar behavior in (9,10). Indeed, the process converged well. However the amount of computation needed to iterate (9,10) for every eigenfunction was still very demanding (several hours per eigenfunction, on the average, for about 600 snapshots by 51840 pixels).

The resulting basis is not necessarily orthogonal. However, on practice the angle between different eigenfunctions,

$$\alpha_{nm} = \arccos \frac{(\psi_n, \psi_m)}{\sqrt{|\psi_n|^2 |\psi_m|^2}}$$

did not exceed 1–3°. Thus, in what follows we consider ψ_n as a good approximation to the "true" eigenfunctions.

To evaluate the effect of this method, we compared the result with the naive technique of filling the missing data by time-average values and applying the classical KL procedure. The spectra of eigenvalues resulting from the two methods are very close, but there is essential difference in eigenfunctions in the polar regions. Fig. 2 shows the 6th eigenfunction for the southern hemisphere, obtained by both methods. The "naive" procedure creates a huge artifact in the polar night region, where the lack of data is most drastic. Beyond the polar circle, where gaps in the data are rare and random, the results agree well.

545

2   First eigenfunctions

The ozone content in the Earth's stratosphere is maintained due to the balance between complicated photochemical processes of its generation and destruction. The photochemical lifetime of ozone can range from days to months [9], depending on latitude and altitude, and thus the ozone layer exhibits seasonal variations. It turns out that the first three eigenfunctions capture the entire seasonal variation, and the entire temporal regularity, while the time courses of other eigenfunctions have an appearance of random functions.

The temporal course $\phi_0(t)$ of the zeroth eigenfunction (Fig. 3) is almost constant in time and, consequently, ψ_0 reflects the temporally averaged ozone concentration. The "Fourier coefficient" $\phi_0(t)$ demonstrates weak annual variation of total ozone content (with distinct annual minima in northern winter and secondary minima in southern winter), as well as a weak secular trend of 2.5% per decade (average for the period 1980–92), more pronounced between 1980 and 1986. (For exhaustive investigation of trends of atmospheric ozone, based on TOMS data, see [10]) No other eigenfunction possesses manifest secular trend.

The first and second eigenfunctions (Fig. 4) describe almost pure annual oscillation. The temporal coefficient ϕ_1 has extrema in Spring and Fall, while ϕ_2 is extreme in Winter and Summer. Especially interesting is the south-hemisphere part of ψ_1, with its negative extremum over Antarctica and positive extremum over moderate latitudes. Taking into account that the corresponding temporal coefficient reaches its annual negative extremum in September–October, we conclude that ψ_1 describes Antarctic ozone depletion in Fall, while sunlight generates ozone in mid-latitudes. In other words, ψ_1 is in a way related to the ozone hole phenomenon [11].

ψ_2 is probably related to two different mechanisms. In South Hemisphere, it is non-zero only over Antarctica, and positive in winter, corresponding perhaps to the late-Fall breakdown of Antarctic polar vortex, and consequent fast rise of ozone concentration. In Northern Hemisphere, ψ_1 is positive in Summer in tropics, when and where ozone is generated by sunlight. The peculiar pattern of ψ_2 over NH resembles generic mixing of a passive scalar blob (see, e.g. [12]), however it is not clear how significant this similarity is.

The general structure of following several eigenfunctions reflects the fact that the spatial spectrum of a typical ozone concentration field is a decreasing function of wavenumber (see below). EEFs from 3 to 8, shown in Fig. 5, appear as large-scale structures around the poles, where variation is much greater than in the equatorial region. Neither of them reveals any specific physical mechanism, though some qualitative observations can be made. Thus the temporal variation of EEFs 6 and 7 (Fig. 6) shows sharp peaks of activity in December–January, more pronounced in some years than in others. Spatially, both functions are mostly located around the North Pole, with little variation in the SH. Temporal courses of EEFs 4 and 8 suggest the presence of quasi-biennial oscillations, although the total observation time of 12 years is probably too short for a reliable spectral analysis.

Finally, note that processes in the two hemispheres seem to be largely decoupled, as demonstrated by hemispherical asymmetry of EEFs 3, 5, 6, and 7. Since the atmospheric transport across the equator is indeed small, and the patterns of atmospheric circulation are different in South and North Hemispheres (in part, because of greater influence of orography in the NH), it is natural to perform the EOF analysis for the two hemispheres separately.

This will be presented elsewhere.

3 Small-scale eigenfunctions and spectrum

The spectrum of eigenvalues λ_n, is very important because it shows for how much variance every eigenfunction is responsible [13, 4]. Indeed, the magnitude of projection of $c(x, t)$ onto $\psi_n(x)$ is

$$
\begin{aligned}
E_n &= \int \left(\int c(x, t)\psi_n(x)\, dx \right)^2 dt \\
&= \int \psi_n(x) \left(\int \psi_n(x') K(x, x')\, dx' \right) dx \\
&= \int \psi_n(x)\lambda_n\psi_n(x)\, dx \\
&= \lambda_n
\end{aligned}
\tag{11}
$$

Hence eigenvalues are measures of average variance per eigenfunction, and the total variance is given by their sum total:

$$
\left\langle \int c^2(\mathbf{x})d\mathbf{x} \right\rangle = \operatorname{Tr} K = \sum_n \lambda_n.
\tag{12}
$$

In Fig. 7 we display in doubly logarithmic form λ_n versus index n. As is seen the variance spectrum falls, to good approximation, on two different power laws

$$
\lambda_n \propto
\begin{cases}
n^{\alpha_i}; & n < 35, \quad \alpha_i = -0.85 \pm .035 \\
n^{\alpha_o}; & n > 50, \quad \alpha_o = -1.56 \pm .022
\end{cases}
\tag{13}
$$

The error bounds appearing in (13) are based only on the least squares fit to the data, and not on the methods used in arriving at the spectrum which appears in Fig. 7.

It is possible to transform this spectrum to the traditional wavenumber spectrum [4]. The key observation here is that higher-order eigenfunctions (see samples in Fig. 5), i.e. those with smaller variance, consist of positive and negative patches of roughly the same length scale. This is natural, in view of the well-known fact that for processes possessing spatial homogeneity, empirical eigenfunctions are nothing else but sinusoids. Indeed, consider for simplicity a one-dimensional process $f(x, t)$, such that the spatial correlation function between points x and y depends only on the distance between them:

$$
\langle f(x, t)f(y, t) \rangle_t = K(x - y)
\tag{14}
$$

Then Fourier transform (denoted by tilde) of the eigenproblem (5) reduces to the equation

$$
\tilde{K}(k)\tilde{\psi}(k) = \lambda\tilde{\psi}(k)
\tag{15}
$$

which is in the generic case solved by sinusoids $\tilde{\psi}(k) = \delta(k - k_0)$, $\lambda = \tilde{K}(k_0)$.

On the global scale, the ozone distribution is nonhomogeneous, owing to latitudinal dependencies of the ozone source/sink and atmospheric circulation, influence of the underlying surface, etc. However, on smaller scales we can expect local homogeneity, and the appearance of higher-order eigenfunction supports this notion. To quantitively verify that small-scale eigenfunctions are characterized by a distinct length scale, and to determine it, we consider

547

the correlation length of each eigenfunction as the position of the first minimum of the spatial correlation function

$$B_\psi(r) = \langle \psi_n(\mathbf{x})\psi_n(\mathbf{y})\rangle_{|\mathbf{x}-\mathbf{y}|=r}$$

where $|\mathbf{x} - \mathbf{y}|$ stands for the distance between the two points on the sphere. In Fig. 9 a sample correlation function is presented for the eigenfunction ψ_{44}, shown in Fig. 8. These functions B_ψ decay with oscillations, and possess a distinct first minimum corresponding to the average distance L_n between adjacent maxima and minima of ψ_n. This distance can be interpreted as half wavelength.

Wavelength determined in this way can be related to the eigenfunction index by the following argument. Assume that the eigenfunctions are homogeneously distributed in the two-dimensional Fourier space. Then the number of those with wavenumber smaller than some k, is order

$$N \propto k^2, \text{ whence } k \propto N^{1/2}. \tag{16}$$

Assuming, in addition, that the spectral density is decreasing with wavenumber increasing, we can interpret N above as being proportional to the eigenfunction index. Finally, the correlation length, being in the inverse relation with wavenumber, should satisfy

$$L_n \propto n^{-1/2} \tag{17}$$

The correlation length is plotted in Fig. 10 versus the eigenfunction index in log-log coordinates. It is clear from this figure that (17) provides an excellent fit to the data in the two asymptotic regimes. The region of the *knee*, transitory between the two spectral asymptotics, is the only anomaly and it appears as a plateau in the figure and corresponds to just one scale.

$$2\pi/k_* = L_* \approx 4000 \text{ km.} \tag{18}$$

We refer the reader to ref. [4] for the discussion of the knee region.

Now that the relation (17) between eigenfunction index and its wavenumber is established, we can translate the power laws for λ_n, (13) to the wavenumber form. In keeping with customary practice we consider the variance per wavenumber $E_c(k) = k\mathcal{E}_c(k)$. It follows from (13) and (16) that

$$E_c(k) \propto \begin{cases} k^{-2/3}, & k < k_* \\ k^{-2}, & k > k_* \end{cases} \tag{19}$$

The more precise exponents are entered for suggestive reasons. Surprisingly enough, neither value has been predicted by existing theories of passive scalar advection, which propose either $k^{-5/3}$ (Obukhov and Corrsin [14, 15, 16]), or k^{-1} (Batchelor [17], see also Chertkov *et al.* [18]). Leaving the theoretical discussion for elsewhere, here we will present an additional empirical analysis in order to establish the exponent -2 in (19) in a more solid manner.

4   Confirmation of the k^{-2} spectrum

The k^{-2} spectrum of a scalar field can be interpreted as an evidence that the scalar variance is dominated by (curvi)linear discontinuities. In fact, it is this reasoning that lead Saffman [19]

548

to propose the k^{-4} energy spectrum for the two-dimensional turbulence, by arguing that the vorticity field should develop jumps, and concluding that k^{-2} vorticity spectrum means k^{-4} spectrum of kinetic energy.

Fig. 11 shows the field of squared gradient of the concentration (for the arbitrarily picked snapshot of March 1, 1995). It is clearly seen that large gradient values are mostly concentrated in nearly one-dimensional filaments. Consider the structure function of such a field, $D_c(r) = \langle (c(\mathbf{x}) - c(\mathbf{y}))^2 \rangle_{|\mathbf{x},\mathbf{y}|=r}$. For the distances $l < r < L$, where l is the characteristic thickness of a jump, and L is the characteristic distance between jumps, the average difference between two points is $\langle (c(\mathbf{x}) - c(\mathbf{y}))^2 \rangle = P \delta c^2$, where P is the probability to have a jump between the points $\mathbf{x}$ and $\mathbf{y}$, and δc is the typical amplitude of a jump. Since $P \propto r$, the structure function, should be proportional to r, and correlation should behave as $B_c(r) = \langle c^2 \rangle - D_c(r)/2 \sim \langle c^2 \rangle - O(r)$. This also implies the power spectrum $E_c(k) = O(k^{-2})$. The actual structure function is shown in Fig. 12. It is practically linear for distances between about 300 to 1500 km.

The correlation function of a scalar field determines the correlation function of its gradient [20] by

$$B_{ll}(r) = \langle g_l(\mathbf{x}) g_l(\mathbf{y}) \rangle_{|\mathbf{x},\mathbf{y}|=r} = -B_c''(r) \tag{20}$$

$$B_{nn}(r) = \langle g_n(\mathbf{x}) g_n(\mathbf{y}) \rangle_{|\mathbf{x},\mathbf{y}|=r} = -\frac{B_c'(r)}{r} \tag{21}$$

where primes denote derivatives over r, and subscripts l and n refer to longitudinal and normal components of the gradient, that is, its projections onto the straight line joining the two points, and onto its normal, respectively. For the linear $B_c(r)$, this gives $B_{ll} = 0$, which is easily understandable, since the difference in concentration between two points can be written as

$$c(\mathbf{y}) - c(\mathbf{y}) = \int_0^r g_l(s)\, ds \tag{22}$$

where s is the coordinate along the line joining the two points. That this difference grows on the average as $r^{1/2}$ requires the integrand be delta-correlated on the distances order r. In Fig. 13, B_{ll} and B_{nn} are shown for a particular snapshot. Although B_{ll} is non-zero, it is essentially smaller than B_{nn} at small scales, which is a reasonable agreement, making the allowance for the amplification of errors in numerical differentiation of experimental data. Actually, it is easy to see from (21), (20) that $B_{ll} \ll B_{nn}$ implies that $B_c(r)$ is linear in r.

Thus, the scalar power spectrum $E_c(k) = O(k^{-2})$ seems to pass several tests. We will attempt to provide a theoretical explanation and discuss the failure of available theories to provide such.

5 Conclusions

Global atmospheric ozone concentration fields from TOMS device were analyzed in terms of Karhunen–Loeve method (Empirical Orthogonal Functions). This analysis reveals characteristic spatio-temporal patterns, including the Antarctic ozone hole phenomenon. A new variant of the KL procedure was developed to deal with data containing gaps resulting from both measurement method limitations (polar nights), and occasional device malfunction. By

549

analyzing spatial structure of eigenfunctions, it is shown that the eigenvalue spectrum can be reliably related to the wavenumber spectrum. This wavenumber spectrum contains two power-law domains with a knee between them. An explanation is proposed for the locations and extent of the knee region. Neither of the exponents in the power-law domains is predicted by available theories of passive scalar advection by 2D fluid flow. For the small-scale power-law domain, an additional empirical evidence is obtained in favor of the hypothesis that the corresponding k^{-2} asymptotic is due to the spatial distribution of ozone, typically consisting of regions with almost-constant concentration, separated by one-dimensional near-discontinuities.

References

[1] P. Grimares and R. McPeters, editors. *Total Ozone Mapping Spectrometer (TOMS) Data, 1978–1993*. NASA Goddard Space Flight Center, 1993.

[2] M.R.Schoeberl, R.S.Stolarski, and A.Kreuger. *International Journal of Supercomputer Applications*, 16(377), 1989.

[3] M.H. Proffitt et al. *Nature*, 347:31, 1990.

[4] L. Sirovich, R. Everson, and D. Manin. Turbulent spectrum of the earth's ozone field. *Phys. Rev. Letters*, 74(13):2611–2614, 1995.

[5] L. Sirovich and R. Everson. *International Journal of Supercomputer Applications*, 6(1):50–68, 1992.

[6] G. W. Stewart. *SIAM Rev.*, 35, 1993.

[7] G. Berkooz et al. *Ann. Rev. Flu. Mech.*, 25, 1993.

[8] L. Sirovich. Turbulence and dynamics of coherent structures. *Quarterly of Applied Mathematics*, 155(3):561–590, 1987.

[9] R. Garcia and S. Solomon. *J. Geophys. Res.*, 90(D2):3850, 1985.

[10] J.R.Herman and R.McPeters. *J. Geophys. Res.*, 98(D7):12783–12793, 1993.

[11] R.P. Wayne. *Chemistry of Atmospheres*. Oxford, 1991.

[12] M. R. Schoeberl and J. T. Bacmeister. *NATO ASI Series*, 18:135, 1992.

[13] B. Knight and L. Sirovich. *Phys. Rev. Lett.*, 65:1356, 1990.

[14] A.M. Obukhov. *Izv. Acad. Nauk SSSR Geogr. Geofiz*, 13:58–69, 1949.

[15] S. Corrsin. *Appl. Phys.*, 22:469, 1954.

[16] A.S. Monin and A. Yaglom. *Statistical Fluid Mechanics*, volume 2. MIT, 1971.

[17] G.K. Batchelor. *J. Flu. Mech.*, 5:113, 1958.

[18] M. Chertkov, G. Falkovich, I. Kolokolov, and V. Lebedev. Statistics of a passive scalar advected by a large-scale velocity field: analytic solution. (preprint wis/94/t/jan-ph). *Phys. Rev. E (to appear)*.

[19] P.G. Saffman. *Stud. Appl. Math.*, 50:377–383, 1971.

[20] S. Panchev. *Random Functions and Turbulence*. Pergamon Press, 1971.
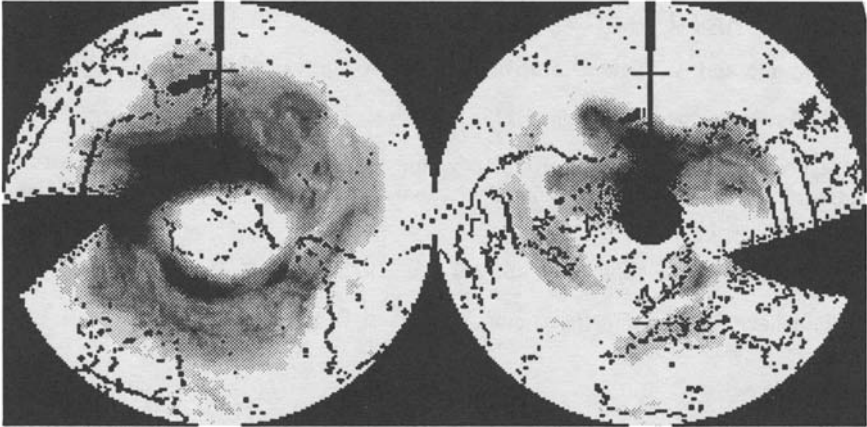
Data of 10/22/80

Figure 1: A typical snapshot of the global ozone field. Outlines of continents are shown.

a   b
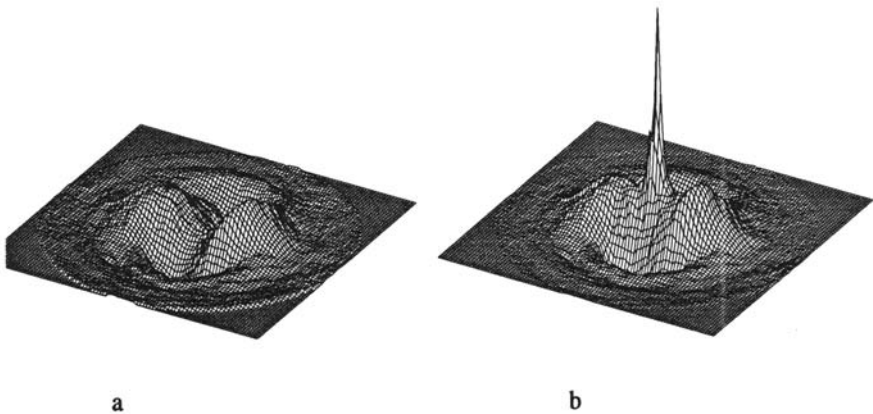
Figure 2: Southern-hemisphere topographic view of ψ_6, obtained by iterations (a) and by the naive method (b).

552
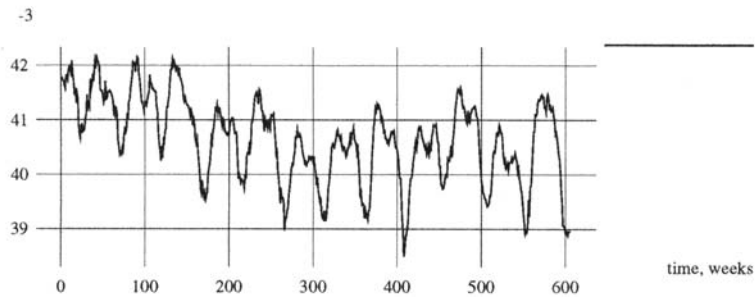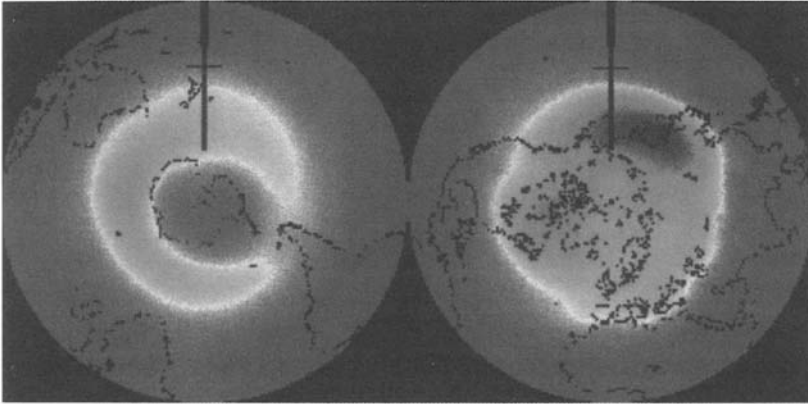
Eigenfunction 0

Figure 3: Eigenfunction 0, ψ_0, and its time history.
(See color Plate 3.)
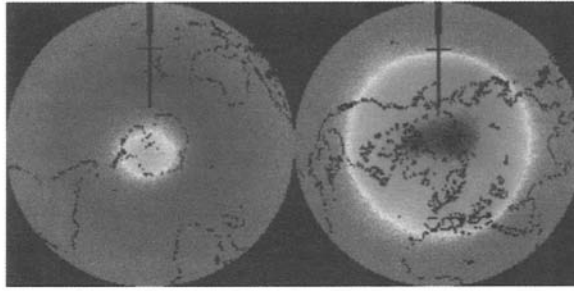
Eigenfunction 1
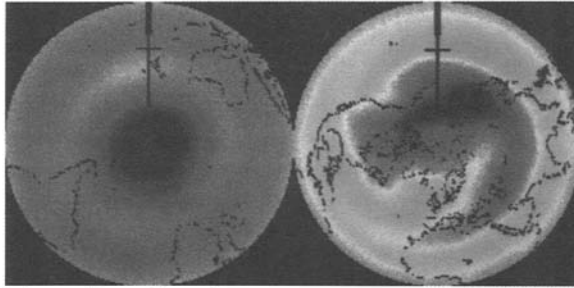
Eigenfunction 2

Temporal coefficients
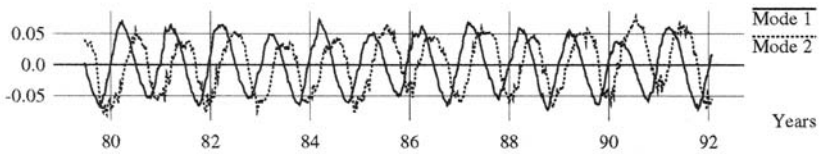
Figure 4: Eigenfunctions 1 and 2, and their time histories.
(See Color Plate 4.)

554
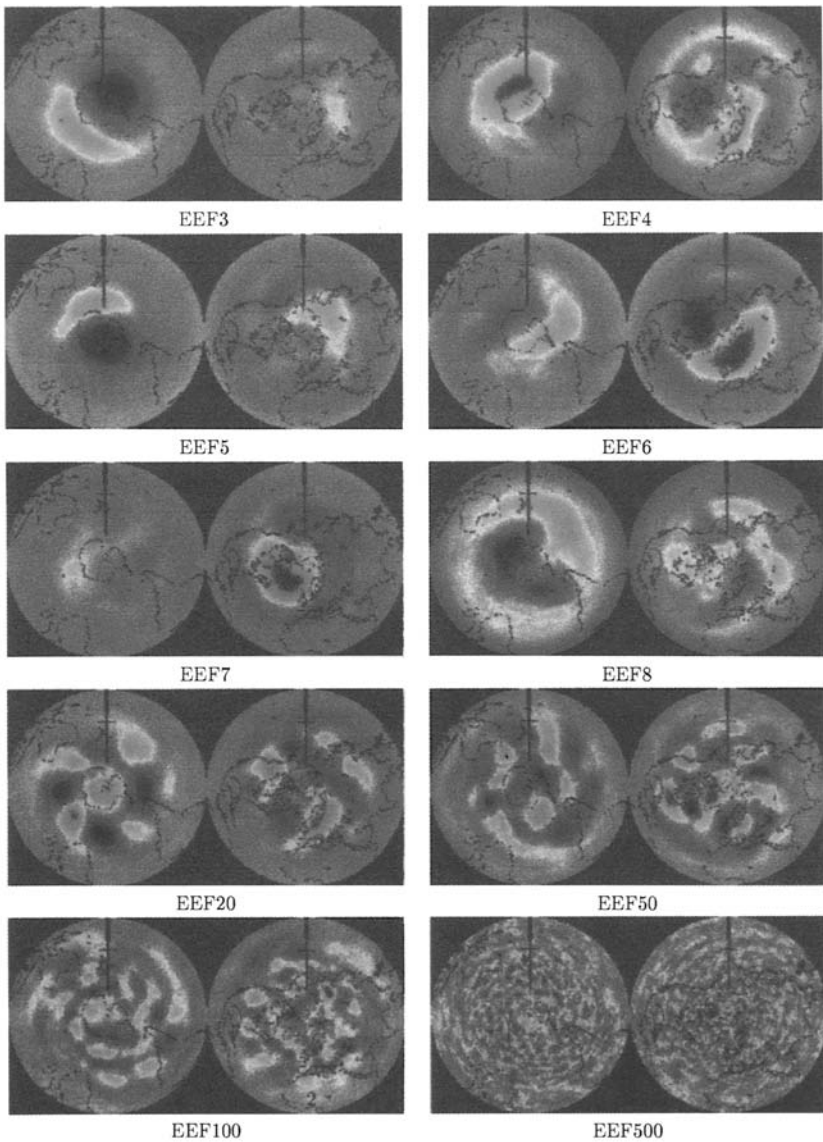
Figure 5: A selection of eigenfunctions.
(See Color Plate 5.)

555
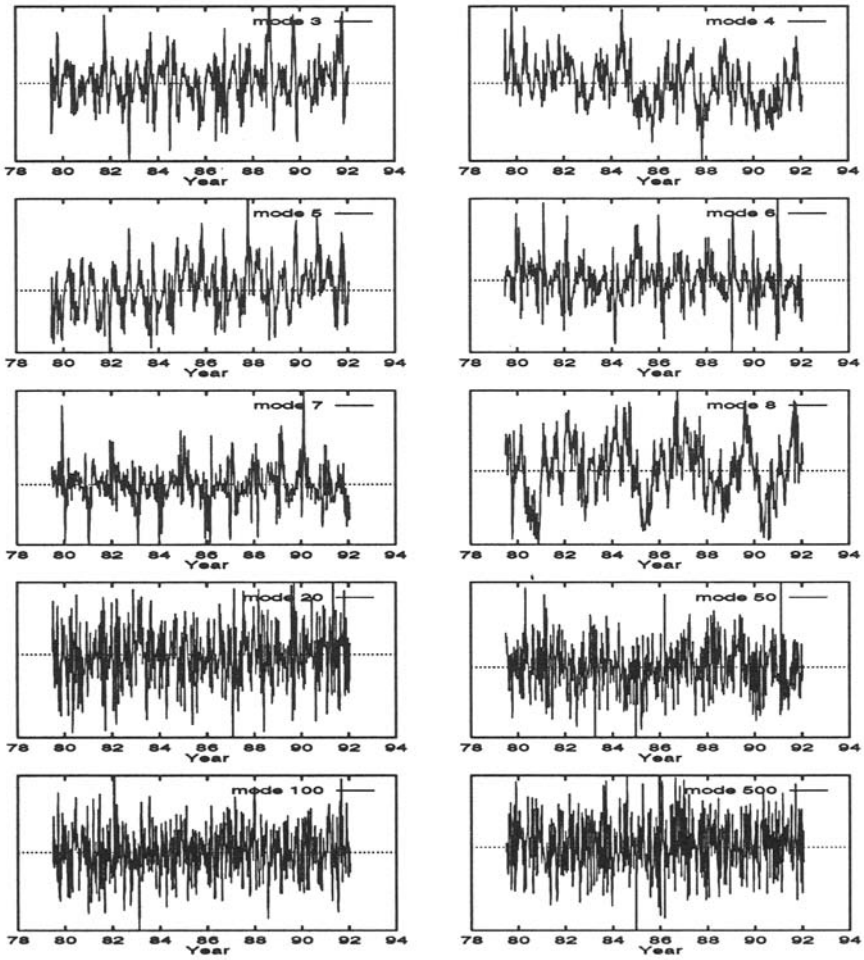
Figure 6: Time histories for eigenfunctions shown in Fig. 5

556

Eigenspectrum: mode energy vs. mode number
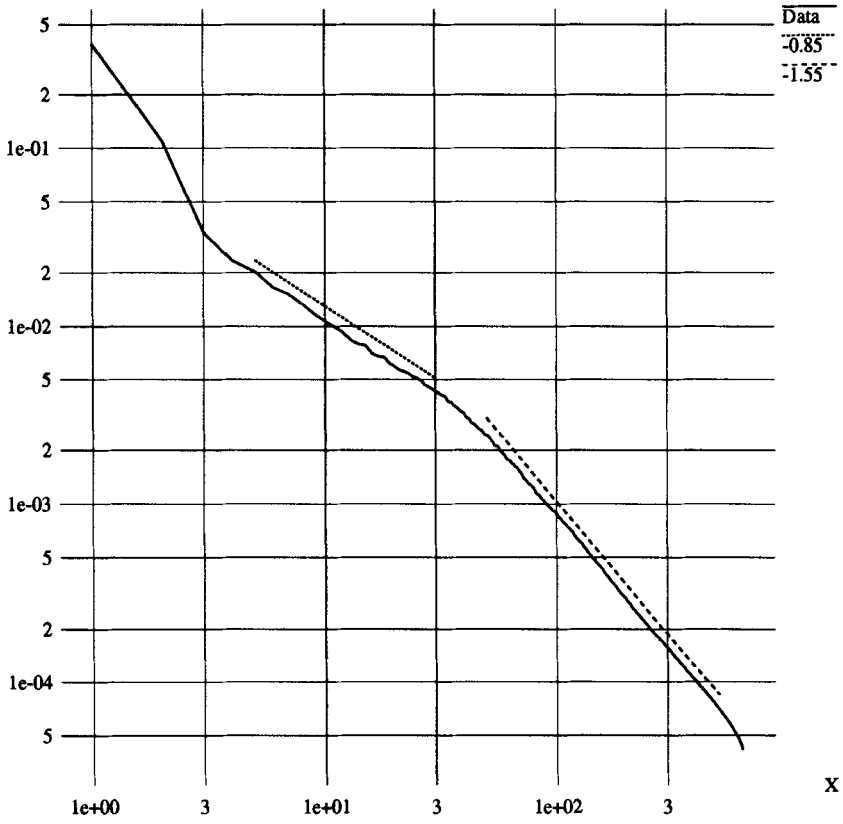
Figure 7: Log–log plot of the global eigenvalue spectrum.
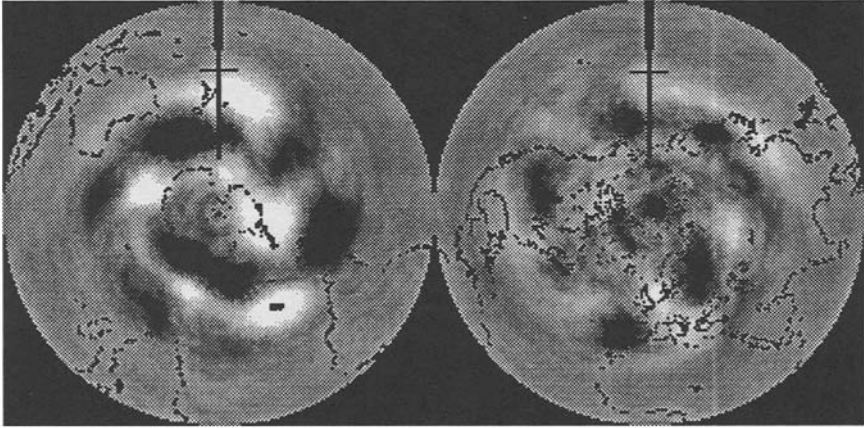
Eigenfunction 44

Figure 8: Eigenfunction 44.
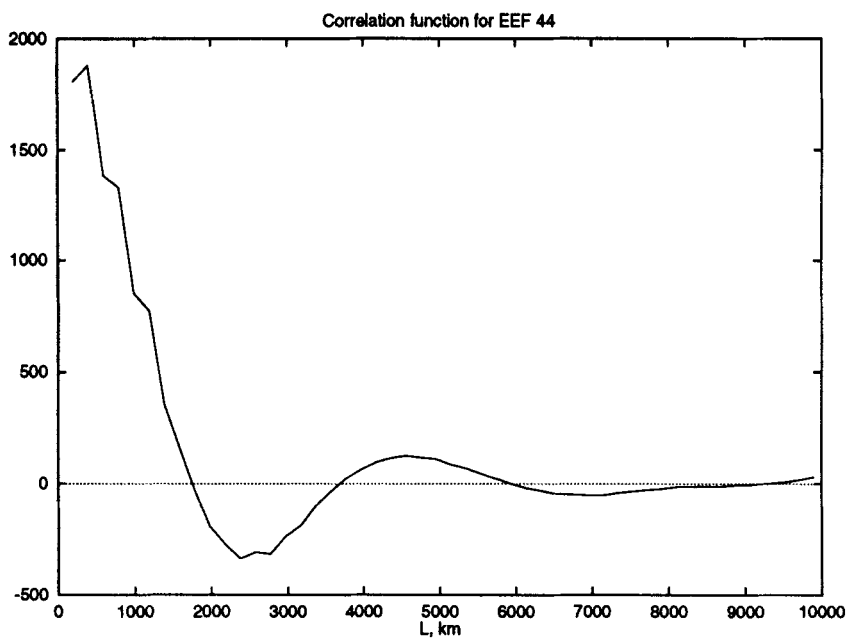
Figure 9: Spatial correlation function of the eigenfunction ψ_{44}, shown in Fig. 8
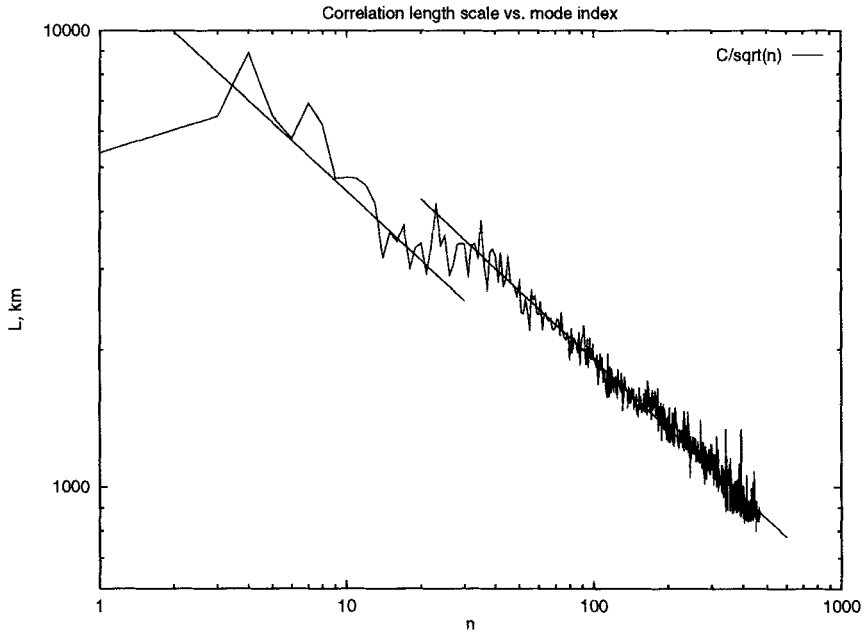
Figure 10: Correlation length scale of eigenfunction ψ_n vs. its index n.
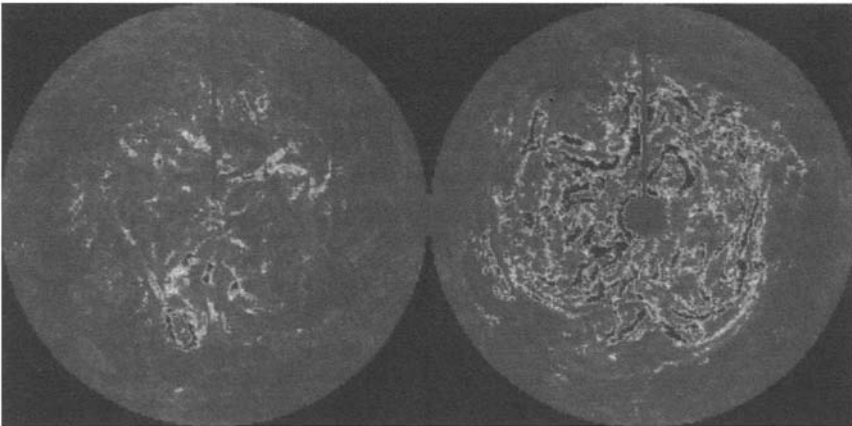
Squared gradient of concentration as of 03/01/85

Figure 11: Squared gradient of the ozone concentration field.
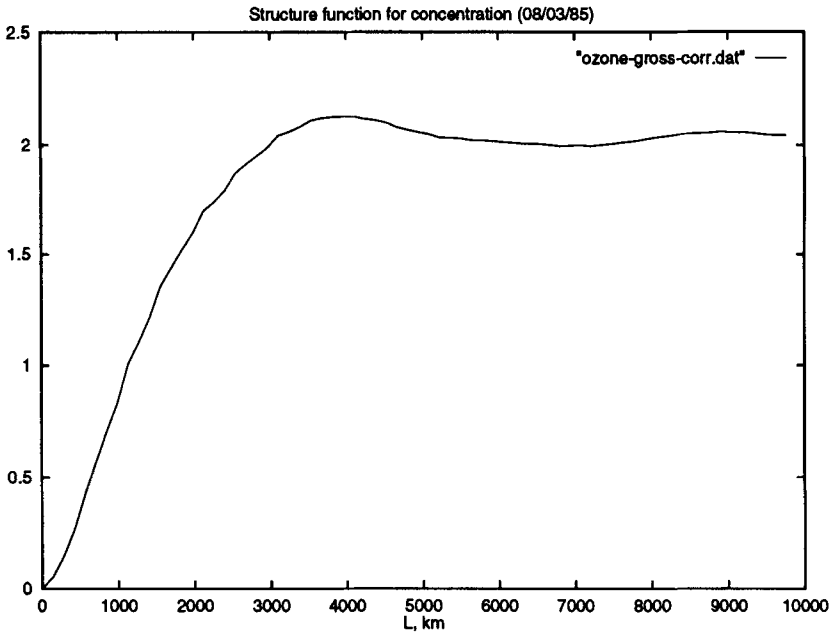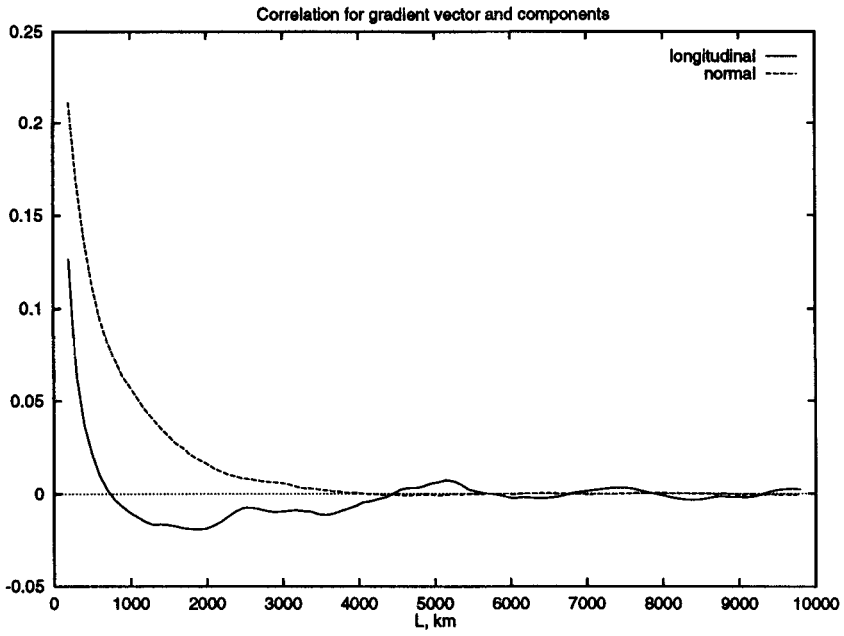*(See Color Plate 6.)*

560

Figure 12: A concentration structure function.

Figure 13: Correlation functions for the gradient components.