A discussion of empirical orthogonal functions and their application to vertical temperature profiles

By J. C. ALISHOUSE, L. J. CRONE, H. E. FLEMING, F. L. VAN CLEEF and D. Q. WARK,
Environmental Science Services Administration, National Environmental Satellite Center, Washington, D.C.

(Manuscript received May 16, 1966)

ABSTRACT

It has been found that empirical orthogonal functions have a special application in the deduction of temperature soundings from radiometric measurements made by artificial satellites. The methods of Lorenz, Obukhov, and Holmström for the generation of empirical orthogonal functions are discussed and compared. These methods all reduce to that of solving eigenvalue problems. It is shown that Holmström's approach is identical to that of E. Schmidt and that his results are identical to those of Obukhov. Furthermore, the computational procedure of Holmström is found to be equivalent to O. D. Kellogg's method for finding eigenfunctions, which in the discrete case reduces to the well-known power method with deflation for matrices. A measure of variance is given to select the number of empirical functions to be used relative to the errors in physical measurements.

1. Introduction

As described by Wark & Fleming (1966), radiometric observations from satellites in several discrete intervals of the 15-micron carbon dioxide band can be used to determine the temperature profile of the atmosphere through application of the integral form of the radiative transfer equation. Closed-form approximations of the temperature profile were found to have great weaknesses. Following suggestions by J. Charney and B. Bolin, the authors have employed empirical orthogonal functions in the solution of the inverse problem and have, thereby, largely overcome these weaknesses.

The linearized integral form of the radiative transfer equation for terrestrial conditions is given by Wark & Fleming (1966) as

$$\frac{I(\nu,\,0)-\beta(\nu)}{\alpha(\nu)} = -\int_0^{t_0} B[\nu_r,\,T(t)]\,\frac{d\tau(\nu,\,t)}{dt}\,dt, \quad (1)$$

where I $(\nu,0)$ is the radiance at the frequency ν and in the vertical direction (zenith angle $\theta=0$); $\alpha(\nu)$ and $\beta(\nu)$ are linearization factors relative to some reference frequency ν_r; $B[\nu_r, T(t)]$ is the Planck radiance as a function of the independent variable t (t is some function of pressure and t_0 corresponds to the level at which all the $\tau(\nu,t)$ have gone to zero); and $\tau(\nu,t)$ is the fractional transmittance in the spectral interval between level t and the top of the atmosphere.

To obtain the temperature profile, one must solve equation (1) for the Planck radiance $B[\nu_r, T(t)]$ as a function of the independent variable t; $B[\nu_r, T(t)]$ is transformed to temperature $T(t)$ only as a final step. Because the radiance $I(\nu,0)$ is measured by a satellite in only a finite number N of narrow spectral intervals, the Planck radiance is approximated by a linear combination of N empirical orthogonal functions. Furthermore, the empirical orthogonal functions are computed from a sample set of M measured temperature profiles $\{T_m(t)\}_{m=1}^{M}$; if $\overline{B[\nu_r, T_m(t)]}$ represents the average Planck radiance of this measured set, it is more efficient to approximate the difference $b(t)$ between the Planck radiance $B[\nu_r, T(t)]$ in (1) and the average Planck radiance of the sample set of profiles, $\overline{B[\nu_r, T_m(t)]}$, than it is to approximate the Planck radiance itself. That is,

$$b(t) = B[\nu_r,\,T(t)] - \overline{B[\nu_r,\,T_m(t)]}, \quad (2)$$

and then, instead of solving equation (1), solve the equation

$$g(\nu) = \frac{I(\nu, 0) - \beta(\nu)}{\alpha(\nu)} - J(\nu, 0)$$

$$= -\int_0^{t_0} b(t) \frac{d\tau(\nu, t)}{dt} dt, \qquad (3)$$

where

$$J(\nu, 0) = -\int_0^{t_0} \overline{B[\nu_r, T_m(t)]} \frac{d\tau(\nu, t)}{dt} dt.$$

In this case the empirical orthogonal functions are computed from the set of differences $\{b_m(t)\}_{m=1}^{M}$ where

$$b_m(t) = B[\nu_r, T_m(t)] - \overline{B[\nu_r, T_m(t)]}, \qquad (4)$$

in which the profiles $T_m(t)$ are the members of the sample set of M measured temperature profiles. (Note that (2) and (4) differ in that the former is the solution of (3) whereas the latter defines the individual elements of a sample set from which the empirical orthogonal functions are computed.)

2. The generation of the empirical orthogonal functions

Lorenz (1956), Sellers (1957) and Shorr (1957) have successfully used empirical orthogonal functions in connection with other meteorological problems. They reduced the problem of finding the empirical functions (vectors, actually) to that of calculating the eigenvectors of a symmetric matrix of the form $\mathbf{S} = \mathbf{A}^T\mathbf{A}$. For our problem, the elements of the mth row of $\mathbf{A}$ are the mth differences of (4) evaluated at a finite number of pressure levels of the mth profile of the sample; the number of rows is equal to M, the number of profiles in the sample. The matrix $\mathbf{A}^T$ is the transpose of the matrix $\mathbf{A}$. The eigenvectors associated with $\mathbf{S}$ are the required empirical functions and, of course, are orthogonal by virtue of the symmetry of $\mathbf{S}$.

A more rigorous approach is given by Obukhov (1960), in which the derivation is for continuous functions rather than for vectors. The empirical functions in his paper are eigenfunctions associated with an integral operator instead of a matrix operator. This operator is also symmetric, and the related empirical functions are likewise orthogonal.

An approach based on variational methods is given by Holmström (1963) in which the set of empirical functions is the solution to an extremal problem. A pair of sequences of N continuous functions $\{\phi_k(t)\}$, $\{\psi_k(m)\}$ is required such that the total variance

$$\sigma_N^2 = \int_0^{t_0}\int_0^M \left[b(t, m) - \sum_{k=1}^{N} \psi_k(m)\phi_k(t)\right]^2 dm\, dt \qquad (5)$$

is a minimum. The index m is now treated as a continuous variable (hence M no longer need be an integer), and the difference $b(t,m)$ is the continuous analog of that given by (4). The solution to this problem is given by Holmström (1963) in terms of the following set of equations:

$$\int_0^{t_0} \left[b(t, m) - \sum_{k=1}^{n-1} \psi_k(m)\phi_k(t)\right] \phi_n(t)\, dt = \psi_n(m),$$

$$n = 1, \ldots, N, \qquad (6\text{a})$$

$$\int_0^{M} \left[b(t, m) - \sum_{k=1}^{n-1} \psi_k(m)\phi_k(t)\right] \psi_n(m)\, dm = \lambda_n\phi_n(t),$$

$$n = 1, \ldots, N, \qquad (6\text{b})$$

where $$\lambda_n = \int_0^M \psi_n^2(m)\, dm. \qquad (6\text{c})$$

The continuous functions $\phi_k(t)$ become the required N empirical orthogonal functions and the coefficients ψ_k are functions of the continuous variable m. In practice the independent variable t is made discrete and m is an integer, so these equations can be solved numerically.

3. Equivalence of the empirical functions of Holmström and Obukhov

Although relations (6a) and (6b) do not appear in Obukhov (1960), the empirical functions generated in each case are the same because Holmström's equations lead to an eigenvalue problem for the same symmetric integral operator as that given by Obukhov.

To see this, first apply the separate orthogonal properties of the functions ϕ_k and ψ_k—which are proved in Holmström (1963)—to (6a) and (6b) in order to contract these equations into the form

$$\int_0^{t_0} b(t, m)\phi_n(t)\,dt = \psi_n(m), \quad n = 1, \dots, N, \tag{7a}$$

and

$$\int_0^{M} b(t, m)\psi_n(m)\,dm = \lambda_n\phi_n(t), \quad n = 1, \dots, N, \tag{7b}$$

which follows from the fact that $k < n$ in each case. When $n = 1$, the sums in (6a) and (6b) vanish of themselves.

Now eliminate $\psi_n(m)$ from these pairs of equations for $n = 1, \dots, N$, interchange the order of the integrations, and drop the subscripts to arrive at the eigenvalue problem

$$\int_0^{t_0}\left[\int_0^M b(t, m)b(t', m)\,dm\right]\phi(t')\,dt' = \lambda\phi(t), \tag{8a}$$

where the eigenvalue

$$\lambda = \int_0^M \left[\int_0^{t_0} b(t, m)\phi(t)\,dt\right]^2 dm \tag{8b}$$

follows by substitution of (7a) into (6c).

The bracketed symmetric kernel in (8a) is the kernel $B(t,t')$ given by Obukhov (1960); hence, the corresponding eigenfunctions of (8a) are those of Obukhov and are the required empirical orthogonal functions. Therefore, not only is the expansion of Holmström (1963) identical in form to that of Obukhov (1960) [as is noted by Holmström (1963)], but it is indeed the identical expansion.

Similar results hold for the $\psi_n(m)$. These are derived by using (7b) to eliminate the $\phi_n(t)$ from (7a), and then by using the procedure described earlier to obtain the dual of equation (8a),

$$\int_0^M \left[\int_0^{t_0} b(t, m)b(t, m')\,dt\right]\psi(m')\,dm' = \lambda\psi(m), \tag{9}$$

where the eigenvalue is given by (6c).

The solution of this problem in terms of the adjoint orthogonal systems, defined by equations (7a), (7b), (8a), and (9), was originally given by E. Schmidt (1907). [For a summary of results see Courant & Hilbert (1953), pages 159 and 161]. However, Holmström's derivation of the solution to the extremal problem (5) appears to be novel.

4. Computational procedures

The procedure suggested by Holmström for explicitly obtaining empirical orthogonal functions by means of equations (6a), (6b), and (6c) is first to set $n = 1$, which reduces (6a) and (6b) to (7a) and (7b), respectively, for $n = 1$. The process is begun by inserting an initial guess $\phi_1^{(0)}(t)$ in place of $\phi_1(t)$ in (7a) and computing $\psi_1^{(0)}(m)$, which in turn is used in (7b) and (6c) to find the next approximation $\phi_1^{(1)}(t)$ to $\phi_1(t)$. This iteration is continued until some previously determined convergence criterion is satisfied.

The identical process may also be treated from the following point of view. Use the steps described in the previous section to obtain a pair of equations analogous to (8a) and (9),

$$\int_0^{t_0}\left[\int_0^M b(t, m)b(t', m)\,dm\right]\phi_1^{(i)}(t')\,dt' = \lambda_1^{(i+1)}\phi_1^{(i+1)}(t) \tag{10a}$$

and

$$\int_0^M \left[\int_0^{t_0} b(t, m)b(t, m')\,dt\right]\psi_1^{(i)}(m')\,dm' = \lambda_1^{(i+1)}\psi_1^{(i+1)}(m), \tag{10b}$$

where the factor

$$\lambda_1^{(i+1)} = \int_0^M \left[\int_0^{t_0} b(t, m)\phi_1^{(i)}(t)\,dt\right]^2 dm = \int_0^M [\psi_1^{(i)}(m)]^2\,dm \tag{10c}$$

follows from (6c) and (8b). In the limit, as i becomes infinite, equations (10a) and (10b) become (8a) and (9), respectively, with the subscript 1.

For the cases $n = 2, \dots, N$, (6a), (6b) and (6c) can be treated similarly, except that now the kernel is deflated; that is, linear combinations of the previously found eigenfunctions are subtracted off the original kernel, so that the resulting kernel is orthogonal to all the previous eigenfunctions. In principle this procedure enables one to find all the eigenfunctions of the integral operator in question, and was devised by Kellogg (1922).

Thus the iteration process described by (10a), (10b) and (10c), coupled with the deflation process, is equivalent to the computational procedure of Holmström (1963). The discrete analog of this method, which can be applied to the formulation by Lorenz (1956), reduces to the power method with deflation for finding the eigenvalues and corresponding eigenvectors of a matrix operator (see e.g., Wilkinson (1965), pp. 571–572 and pp. 584–585).

5. Selection of the number empirical functions

In (3) the solution is obtained by expressing the indicial function as

$$b(t) = \sum_{j=1}^{N} c_j \phi_j(t), \tag{11}$$

where $\phi_j(t)$ is the empirical orthogonal function associated with the j^{th} largest eigenvalue, and the c_j are the coefficients which are determined in the inversion process given by Wark & Fleming (1966) [see also Malkevitch & Tatarskii (1965)]. The number of empirical functions available is limited only by the size M of the sample set. However, the number of terms N in (11) is severely limited in two respects by the physical errors of measurement.

As shown by Twomey (1966), the components of the kernel $d\tau(\nu,t)/dt$ must be linearly independent in the sense that no one component is expressable as a linear combination of the others to within the experimental error of $g(\nu)$ of (3). It is for this reason that the instrument described by Hilleary *et al.* (1966) has been designed for only six spectral intervals, and, hence, in this case $N \leq 6$.

A second consideration is that the deduction of a temperature profile by indirect means to an overall accuracy greater than that achieved by the radiosonde data used in the sample is not justified. According to Hodge & Harmantas (1965), the average error of temperature measurement with a radiosonde is about 1°C. This is equivalent to an error of about 1 erg/(cm² sec strdn cm⁻¹) in the Planck radiance at $\nu_r = 691$ cm^{-1}. Thus, the index N in (11) must be such that the residue σ_N^2 in (5), when averaged over t and m, exceeds this error. The mean residue is given by

$$\overline{\sigma_N^2} = \frac{1}{t_0 M}\int_0^{t_0}\int_0^{M}\left[b(t,m) - \sum_{n=1}^{N}\phi_n(t)\psi_n(m)\right]^2 dm\,dt$$
$$= \frac{1}{t_0 M}\left\{\int_0^{t_0}\int_0^{M} b^2(t,m)\,dm\,dt - \sum_{n=1}^{N}\lambda_n\right\}, \tag{12}$$

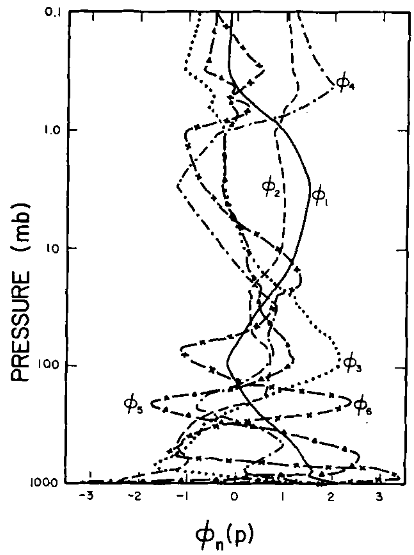

Fig. 1. The first six empirical orthogonal functions $\phi_n(p)$ of the North American sample set of 106 temperature soundings versus the logarithm of pressure p.

where the second line results from applying (7b), the orthonormal properties of the ϕ_n, the orthogonal properties of the ψ_n, and (6c) to the first line.

In studying the temperature inversion problem, sample sets of radiosonde data were selected for use in two separate studies. The first, the heterogeneous sample, was the synoptic situation over North America, the Caribbean, and the adjacent Atlantic and Pacific areas at 1200 G.M.T., January 14, 1961; this was applied to theoretical studies (Wark *et al.*, 1965; Wark & Fleming, 1966), and the first six related empirical functions are shown in Figure 1. The second, the homogeneous sample, was a set of 90 radiosondes taken at Ft. Worth, Texas during August and September; it was applied to a high-altitude (10 mb) balloon flight of a spectrometer, testing the temperature inversion method experimentally (Hilleary *et al.*, 1966; Wark *et al.*, 1966).

Table 1 lists the values of $\overline{\sigma_N}$ of the heterogeneous and homogeneous samples, using, successively, the first through the sixth terms.

It is obvious from Table 1 that, in the application of expansion (11) to the solution of (3), more empirical functions may be used from the heterogeneous sample than from the homogeneous one. If the number of spectral intervals for $g(\nu)$ exceeds the index N in (11), the excess

Table 1. *Values of $\overline{\sigma_N}$ for successive values of N*

N	$\overline{\sigma_N}$ [erg/(cm^2 sec strdn cm^{-1})] Heterogeneous sample	Homogeneous sample
1	4.51	1.92
2	3.14	1.65
3	2.38	1.39
4	1.90	1.26
5	1.60	1.12
6	1.39	1.01
0	9.71	2.52

number may be used in a least squares sense to improve the accuracy of the solution. Conversely, if the sample set is so heterogeneous that the largest permissible N exceeds the number of spectral intervals, then the sample set is not optimal. Procedures and criteria for selecting the "best" sample set are currently under investigation.

Explicit details of the application of empirical orthogonal functions to the solution of the temperature sounding problem (3) are given in Wark and Fleming (1966).

The authors are indebted to Mrs. Jean Hyatt and Mrs. Mary Marini for their assistance with the computations.

REFERENCES

1. Courant, R. & Hilbert, D. 1953. *Methods of Mathematical Physics, Vol. I.* New York, Interscience Publishers, Inc., 561 pp.
2. Hilleary, D. T., Heacock, E. L., Morgan, W. A., Moore, R. H., Mangold, E. C. & Soules, S. D. 1966. Indirect measurements of atmospheric temperature profiles from satellites: III. The spectrometers and experiments. *Monthly Weather Review 94,* 367–377.
3. Hodge, M. & Harmantas, C. 1965. Compatibility of United States radiosondes. *Monthly Weather Review 93,* 253–266.
4. Holmström, I. 1963. On a method for parametric representation of the state of the atmosphere. *Tellus 15,* 127–149.
5. Kellogg, O. D. 1922. On the existence and closure of sets of characteristic functions. *Mathematische Annalen 86,* 14–17.
6. Lorenz, E. 1956. Empirical orthogonal functions and statistical weather prediction. *Scientific Report No. 1, Statistical Forecasting Project,* Massachusetts Institute of Technology, Department of Meteorology, Cambridge, Mass., 49 pp.
7. Malkevitch, M. S. & Tatarskii, V. I. 1965. Determination of vertical atmospheric temperature profiles by means of outgoing radiation in a CO_2 absorption band. *Kosmicheskie Issledovaniya 3,* 444–456.
8. Obukhov, A. M. 1960. The statistically orthogonal expansion of empirical functions. Akademiya Nauk, SSSR. *Izvestiya Seriya Geofizicheskaya 1960,* pp. 432–439. (English Translation by the American Geophysical Union, Nov. 1960, pp. 288–291.)
9. Schmidt, E. 1907. Zur Theorie der linearen und nichtlinearen Integralgleichungen. I. Entwicklung willkürlicher Funktionen nach Systemen vorgeschriebener. *Mathematische Annalen 63,* 433–476.
10. Sellers, W. 1957. A statistical-dynamic approach to numerical weather prediction. *Scientific Report No. 2, Statistical Forecasting Project,* Massachusetts Institute of Technology, Department of Meteorology, Cambridge, Mass., 151 pp.
11. Shorr, B. 1957. Empirical orthogonal functions applied to prediction of the sea-level pressure field. *Scientific Report No. 3, Statistical Forecasting Project,* Massachusetts Institute of Technology, Department of Meteorology, Cambridge, Mass., 88 pp.
12. Twomey, S. 1966. Indirect measurements of atmospheric temperature profiles from satellites: II. Mathematical aspects of the inversion problem. *Monthly Weather Review 94,* 363–366.
13. Wark, D. Q. & Fleming, H. E. 1966. Indirect measurements of atmospheric temperature profiles from satellites: I. Introduction. *Monthly Weather Review 94,* 351–362.
14. Wark, D. Q., Fleming, H. E., Lienesch, J. H. & Van Cleef, F. 1965. The determination of the atmospheric temperature field from simulated satellite measurements. *Symposium on Radiation Processes in the Atmosphere,* Leningrad, August 1964, International Union of Geodesy and Geophysics, Monograph No. 28, p. 56.
15. Wark, D. Q., Saiedy, F. & James, D. G. 1966. Indirect measurements of atmospheric temperature profiles from satellites: VI. High altitude balloon testing. Submitted to *Monthly Weather Review.*
16. Wilkinson, J. H. 1965. *The Algebraic Eigenvalue Problem.* Oxford, Clarendon Press, 662 pp.

ОБСУЖДЕНИЕ ЭМПИРИЧЕСКИХ ОРТОГОНАЛЬНЫХ ФУНКЦИЙ И ИХ ПРИМЕНЕНИЯ ДЛЯ ПРЕДСТАВЛЕНИЯ ВЕРТИКАЛЬНЫХ ПРОФИЛЕЙ ТЕМПЕРАТУРЫ

Было найдено, что эмпирические ортогональные функции имеют специальное применение в получении вертикальных профилей температуры из радиометрических измерерий, проделанных с искусственных спутников. Обсуждаются и сравниваются методы получения эмпирических ортогональных функций Лоренца, Обухова и Хольмстрёма. Все эти методы сходятся к решению задачи на собственные функции и собственные значения, поскольку показано, что метод Хольмстрёма идентичен методы Е. Шмидта и результаты идентичны результатам Обухова. Кроме того, найдено что вычислительная методика Хольмстрёма эквивалентна методу Келлога для отыскания собственных функций, которая в дискретном случае сводится к хорошо известному степенному методу с понижением порядка матриц. Дается оценка дисперсии для определения оптимального числа используемых эмпирических функций по отношению к шуму в физических измерениях.