ON THE PHYSICAL INTERPRETATION OF EMPIRICAL ORTHOGONAL FUNCTIONS

C. Eugene Buell

Science Applications, Inc.
Colorado Springs, Colorado

1. INTRODUCTION

It has been shown (Buell, 1975) that the topography of the empirical orthogonal functions associated with a two-dimensional meteorological field is strongly dependent on the shape of the region concerned and only slightly dependent on the nature of the covariance function and its scale parameter. A heuristic treatment of the reasons for this is carried out. First the basic equations are set up together with an iterative solution algorithm. Next a simple exercise in a one dimensional situation is carried through that illustrates the importance of the end points of the line interval over which the empirical orthogonal functions are defined. These ideas are then extended to two or more dimensions.

2. THE EQUATIONS

It is customary to consider the empirical orthogonal function (EOF's hereafter) as the solutions of the standard matrix proper value/proper function problem.

$$C\Phi = \Phi\Lambda \quad (1)$$

where C is a positive definite covariance matrix (though a correlation coefficient matrix is sometimes used), Φ is a matrix with columns the EOF's and Λ is a diagonal matrix the elements of which are the proper values. The proper functions (EOF's) are orthonormal in that

$$\sum_i \phi_{ij}\,\phi_{ik} = \delta_{jk}$$

where δ_{jk} is the Kronecker delta and the j 'the proper function ϕ_{ij}, i = 1, ---, n is associated with the j'th proper value, λ_j, the proper values being in descending order $\lambda_1 \geq \lambda_2 \geq --- \geq \lambda_u > 0$.

When a field of property on a continuum is concerned (the element c_{ij} of C being the two-point covariance of the property at the point P_i and at the point P_j) the formulation in (1) fails to account for the nature of the continuum concerned. A better approach is to formulate the problem as an integral equation

$$\lambda_i\,\phi_i(x) = \int_A K(x,y)\,\phi_i(y)dy \quad (2)$$

where the kernel function K(x,y) corresponds to the covariance matrix C in (1). The kernel is known only at the data point pairs so that $K(x_i, x_j)$ corresponds exactly to the covariance matrix element c_{ij}. (To reduce the notation, the coordinates x, y, x_i, etc are understood to to be multidimensional as required by the context.) Two important elements are made explicit in the integral equation formulation (2) which are not explicit in the matrix formulation (1); the role of the domain of integration (indicated by A) and the role of the "area element" dy. The solution process for (2) reduces to equations similar to (1) since the kernel is known only at the data points. Thus one has

$$\lambda\,\phi(x_j) = \sum_k K(x_j,x_k)\,\phi(x_k)A_k \quad (3)$$

where the subscript i is dropped from λ and ϕ. If the data points are located on an equally spaced grid, the area elements associated with each point A_k are all the same and their common value absorbed into the proper value λ. In this case (3) and (1) are identical in form. (The importance of the area elements A_k for non-uniform data points is discussed in Buell, 1978.) We will concern ourselves here with the way the shape of the domain of integration enters into this restricted problem.

3. THE ALGORITHM

There are many methods of finding the proper values and proper functions of (1). The method of Jacobi rotations and the QR-algorithm (and its variations) are efficient of computer time, but do not give any insight into how the shape of the domain of integration in (2) affects the solution (even when simplified so that the algebra of (1) may be used). An old iterative technique that was used before high-speed computers (and is still of some utility) can provide an answer (Anderson, 1958). Let C be the covariance matrix, λ_1 the largest proper value, and ϕ_1 the corresponding proper function (a column vector over the data points, ϕ_1 = col (ϕ_{11}, ϕ_{21}, ---, ϕ_{n1})). Now let $x_{(o)}$ be an initial column vector (an estimate of ϕ_1, it must not be orthogonal to ϕ_1) and let $x_{(i)}$ represent the i'th iteration of this vector starting with $x_{(o)}$.

Step 1. Normalize

$$y_{(i)} = x_{(i)}/\sqrt{x'_{(i)}x_{(i)}} \tag{4}$$

where $x'_{(i)}$ is the transpose of $x_{(i)}$.

Step 2. Iterate.

$$x_{(i+1)} = Cy_{(i)} \tag{5}$$

and then go to step 1.
The result of this algorithm is that

$$\lim_{i\to\infty} y_{(i)} = \pm \phi_1 \tag{6}$$

$$\lim_{i\to\infty}\sqrt{x'_{(i)}x_{(i)}} = \lambda_1 \tag{7}$$

To obtain the second proper value/function the first proper value/function must be "removed" from C. Construct $C^{(2)}$ using

$$C^{(2)} = C - \lambda_1\phi_1\phi_1' \tag{8}$$

or

$$c^{(2)}_{ij} = c_{ij} - \lambda_1\phi_{i_1}\phi_{j_1} \tag{9}$$

The process above is then repeated to get λ_2 ϕ_2. For the third proper value/function, λ_2, ϕ_2 are "removed" from $C^{(2)}$ to obtain $C^{(3)}$, etc.. Further, this technique is applicable to the integral equation (2) directly. Consequently, though we confine ourselves to a particular case of (2) (all "area elements" A_k equal) the process being used is quite general.

4. A ONE DIMENSION EXAMPLE

To illustrate the strong influence of the end points for a one dimensional domain of integration consider the covariance function $c_{ij} = \exp\,[-(i-j)^2/2L^2]$, i,j = 1, ---, 8, L=2. The elements of the first and fifth rows are shown in the upper part of Fig. 1. To obtain the first proper value/function one may choose $x_{(o)}$ = col (1, 1, 1, 1, 1, 1, 1, 1). The normalized $y_{(o)}$ is shown as the line of x's in the lower part of Fig. 1. The construction of $x_{(i+1)}$ from $y_{(i)}$ in Step 2 may be written in full as

$$x_j^{(i+1)} = \sum_k c_{jk}\, y_k^{(i)},\ j = 1, ---, n, \tag{10}$$

where the iteration count is the superscript and the vector component index is the subscript. The j'th component of $x_{(i+1)}$ is then the sum of products of the j'th row of C with the components

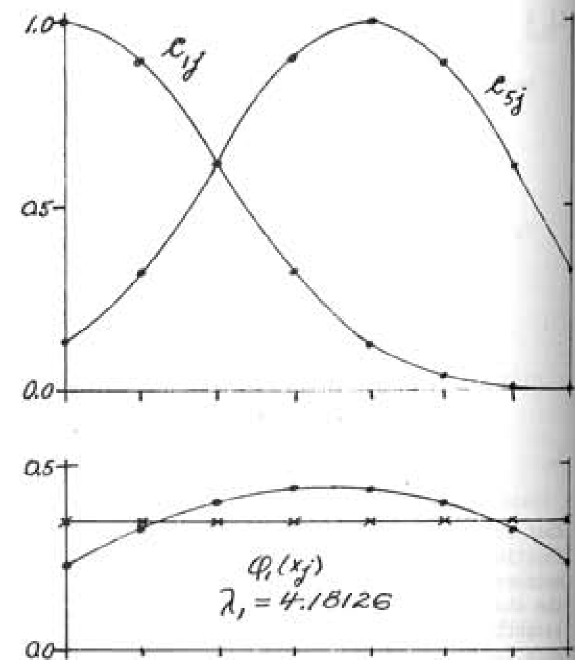

Fig. 1. Upper: The covariance functions for row 1 and row 5 as a function of the column index, j. Lower: The normalized initial estimate $x_{(o)}$ (the line) and the final proper function $\phi_1(x_j)$.

of the normalized vector $y_{(i)}$. It is also the integral with fixed index j over the interval concerned multiplied by a weight function $y_{(i)}$. Starting with a uniform weight function $x_{(o)}$ the values of $x_{(i)}$ are the areas under the covariance function for fixed row index. It is easily seen from the upper part of Fig. 1. that for rows 1 and 8 only "half" of the covariance function lies in the interval while for rows 4 and 5 the covariance function is nearly centered at the middle of the interval. The area computed is then smallest for rows at the end of the interval and largest in the middle. On the next step of the iteration the weight function is no longer uniform and its low values at the end-points (high values at the mid-points) serves to reinforce this tendency on successive iterations. The final value of the vector ϕ_1 and the corresponding proper value λ_1 are shown in the lower part of Fig. 1.

The elements of the first four rows of the modified $C^{(2)}$ are shown in the top of Fig. 2. The elements of the last four rows are obtained from symmetry about the mid-point of the interval. These are obtained by subtracting from each row of C, top of Fig. 1., a curve of the same "shape" as that of ϕ_1, bottom of Fig. 1, multiplied by the factor $\lambda_1\phi_1$, which changes for each row. Since ϕ_{i_1} is smallest for i = 1,8 the least is removed while ϕ_{i_1} is largest for i = 4, 5 and the most removed. This is an end point effect and virtually independent of the details of the shape of the initial covariance function. To find the second proper value/

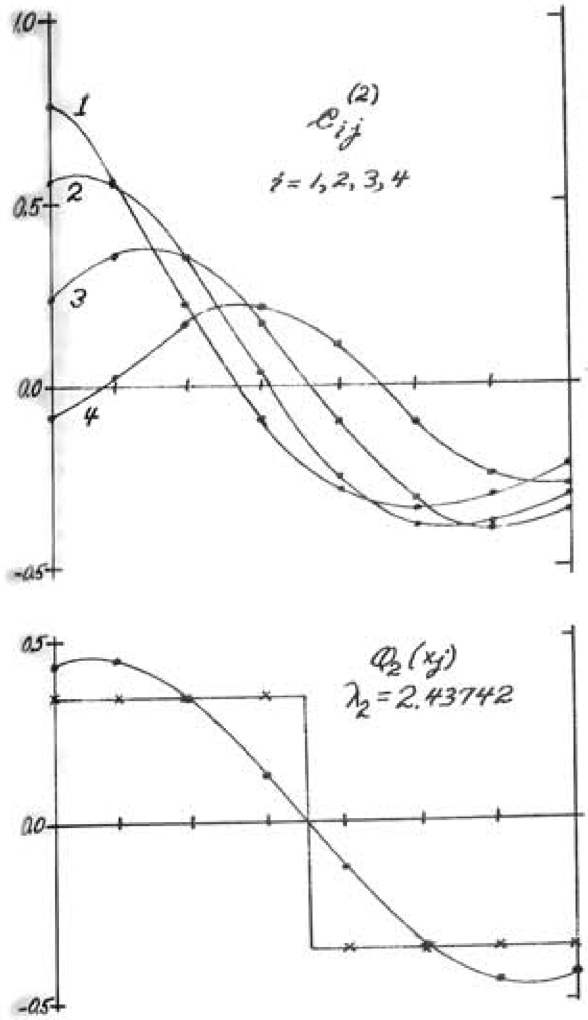

Fig. 2. *Upper: The elements of the modified matrix* $C^{(2)}$, *rows 1,2,3,4, as a function of column index, j. Lower: The normalized initial estimate* $x_{(o)}$ *(step function) and the final proper function* $\phi_2(x_j)$.

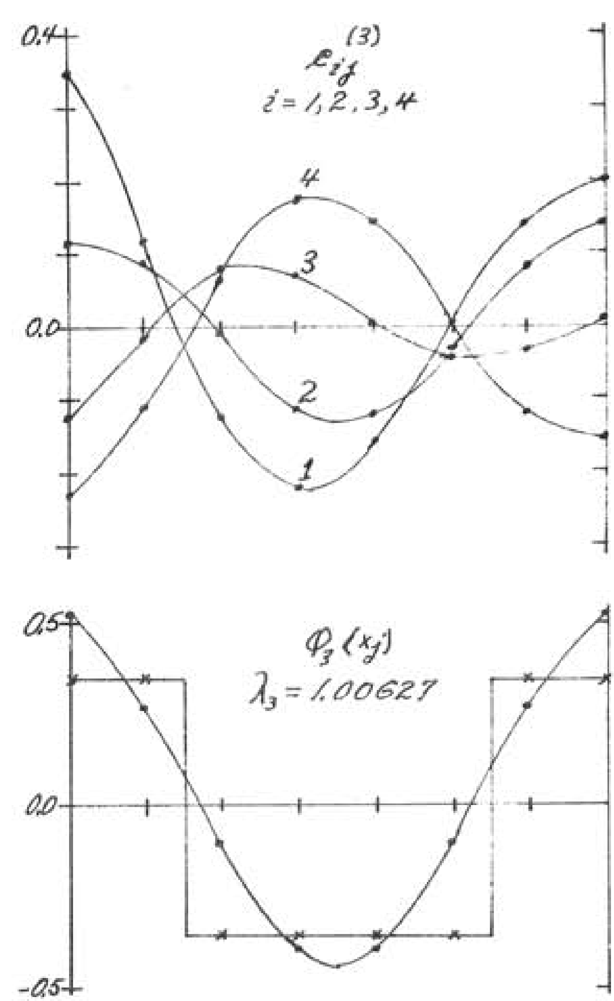

Fig. 3. *Upper: The elements of the modified matrix* $C^{(3)}$, *rows 1,2,3,4, as a function of the column index, j. Lower: The normalized initial estimate* $x_{(o)}$ *(step function) and the final proper function* $\phi_3(x_j)$.

function, an old rule of thumb is to make the initial guess look something like a typical row of the reduced matrix ($C^{(2)}$ here). Taking into account the signs only, a reasonable guess is $x_{(o)} = (1,1,1,1,1,-1,-1,-1,-1)$ shown normalized in the lower part of Fig. 2. Carrying out the iterations, the final result for ϕ_2 and the corresponding λ_2 are shown in the bottom of Fig. 2.

Figs. 3 and 4 continue the process two more steps to λ_3, ϕ_3 and λ_4, ϕ_4.

4.1 Remarks

The covariances c_{ij} in this example are all positive. When this is not the case, the negative values are usually for a large difference $|i-j|$ (or $|x_i - x_j|$). The end effect is to make ϕ_1 even more humped in the middle and may even introduce negative values of $\phi_1(x_i)$ at the end of the interval.

The shape of the proper functions is to a large extent, but not completely, independent of the details of the covariance function. For example, the covariance function $c_{ij} = 1 - |i-j|/h$, $|i-j|<h$, $c_{ij} = 0$, $|i-j| \geq h$, $(h = 5)$ was used with the result that the first two proper functions agreed with those obtained above to two significant figures while the third checked to one significant figure.

The example is strictly symmetric about the center of the interval. This need not be the case. For asymmetric covariance functions, the resulting proper functions would also be asymmetric.

The above argument is strictly heuristic and applies only for sufficiently well behaved covariance functions (as are encountered in the atmosphere).

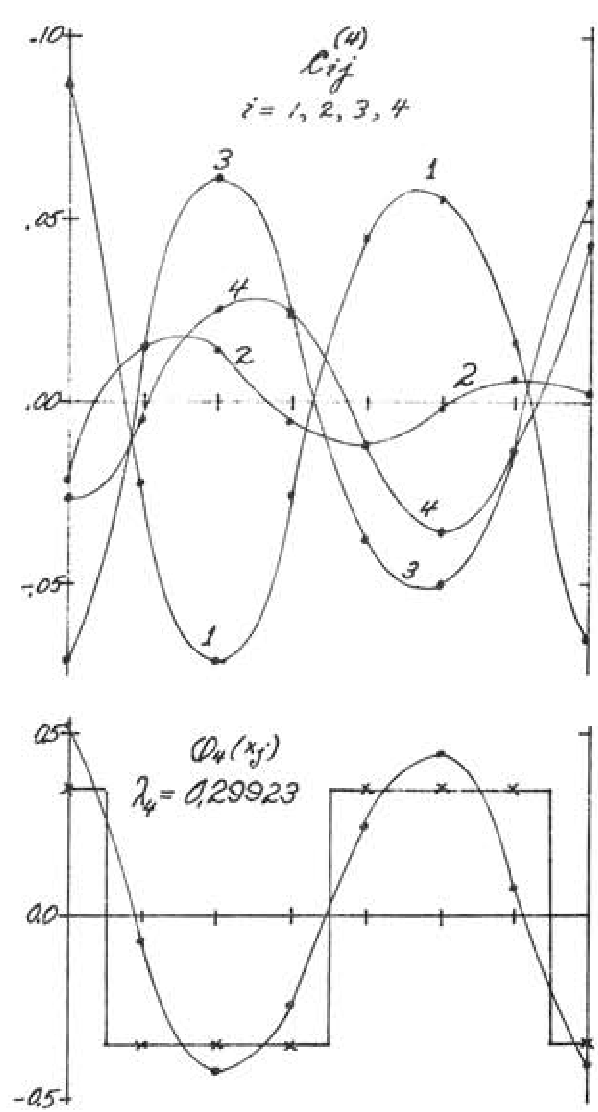

Fig. 4. Upper: The elements of the modified matrix $C^{(4)}$, rows 1,2,3,4, as a function of the column index j. Lower: The initial estimate $x_{(o)}$ (the step function) ad the final proper function $\phi_4(x_j)$.

5. TWO OR MORE DIMENSIONS

In a one dimensional continuum it is convenient to order the rows/columns of C in the order of the data points of the interval concerned. This is no longer possible in two or more dimensions. The elements of C are two-point functions of the data $c_{ij} = c(x_i, x_j)$ where x_i represents the coordinates of P_i, say (ξ_i, η_i) in two dimensions. For fixed row index i, the elements of the columns j range over the points of the data network that "cover" the domain A.

The first step of the iteration process remains much the same as in one dimension but requires a two dimensional representation of the results. A reasonable estimate for the initial vector $x_{(o)}$ is col (1, ---, 1) while the first proper function $\phi_1(x_i) = \phi_1(\xi_1, \eta_1)$, i = 1, ---, n will be represented by a dome-shaped surface over the domain A. Note that when the row index corresponds to a point P_i that lies near the edge of A about half of the large positive part of the covariance function $c_{ij} = c(x_i, y_j)$ as a function of P_j(P_i fixed) is outside A and so is not involved (as at an end point of a one dimensional interval). If the domain A has corner points even more of the large positive part is outside A so that the values for $\phi_1(x_i)$ will be still smaller. An additional effect in two dimensions is that the summation (integration in the continuum case) is over a two dimensional region so that the number of terms for which P_j is widely separated from P_i(fixed) is much larger than one dimension. Thus, negative values of $\phi_1(x_i)$ are more easily introduced for P_i near the border when c_{ij} is negative and P_i, P_j widely separated. The situation is illustrated schematically in Fig. 5.

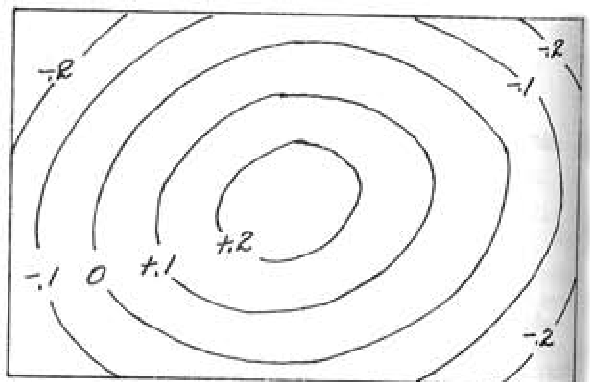

Fig. 5. Schematic representation of the first proper $\phi_1(x_j)$ for the rectangular region shown in the plane.

The algorithm next requires that λ_1, ϕ_1 be "removed" from C to form $C^{(2)}$ as in (8) or (9) resulting in the two dimensional analog of the upper part of Fig. 2. This is much more difficult to represent graphically. Several cases are shown schematically in Fig. 6. It is obvious that a good initial choice for $x_{(o)}$ consists of 1's at the points of half the area and -1's at the points of the other half. Whether the -1's lie in the lower or right hand half would depend on a good guess as to the predominant contour pattern. If patterns like Fig. 6a predominate then the right/left halves may be chosen; if like Fig. 6b, then the top/bottom halves. In either case, the resulting second proper function will be zero near the middle of the region with a positive hill in one half and a negative "hill" in the other half. (For a square region it is quite possible that the pattern similar to Fig. 6c would predominate so that the division would be for NW/SE halves or NE/SW halves.)

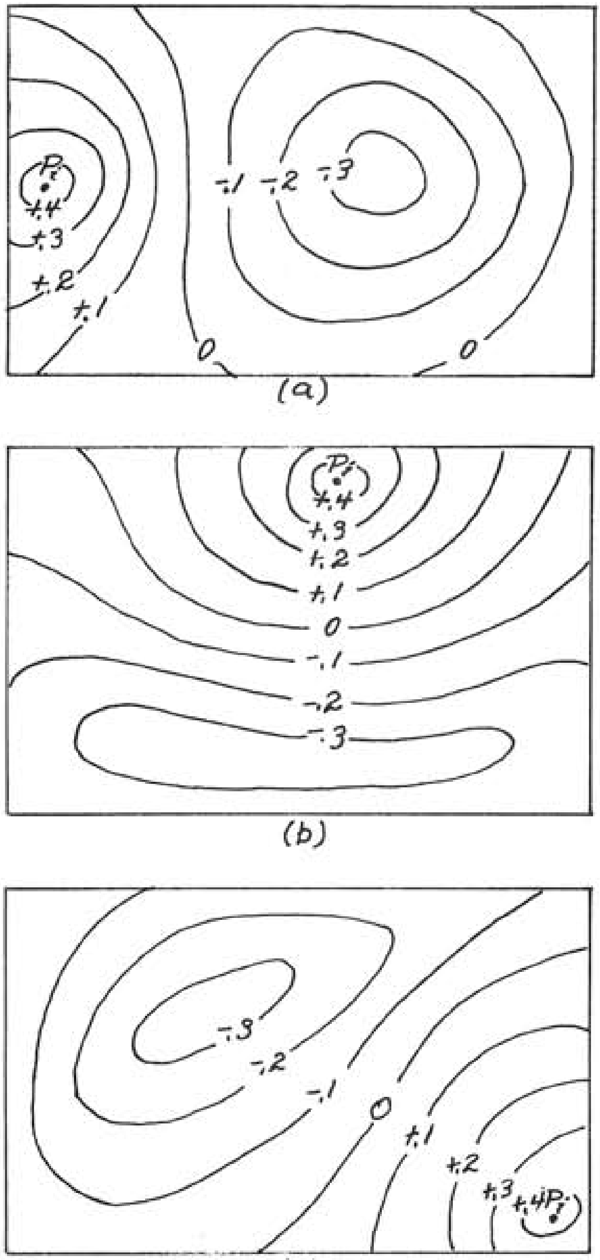

Fig. 6. Schematic representation of the elements of the modified matrix $C^{(2)}$ as a function of location x_j corresponding to column j for rows in which (a) x_i is near the left side, (b) x_i is near the top side, and (c) x_i is near the lower right corner.

The removal of λ_2, ϕ_2 from $C^{(3)}$ has no analog in one dimension. Assume that the proper function $\phi_2(x_i)$ is positive in the left half of A, negative in the right half, and the curve $\phi_2(x_i) = 0$ lies near the middle. Then for rows of $C^{(3)}$ for which P_i lies in the left and right quarters of A are expected to have small elements. On the other hand the rows corresponding to P_i lying centrally in the top and bottom quarters of A will be nearly the same as they were in $C^{(2)}$ since $\phi_2(x_i)$ is small. For a reasonably symmetric covariance function, one would then be lead to expect that a reasonable choice of $x_{(o)}$ to obtain λ_3, $\phi_3(x_i)$ would be the one passed over in obtaining λ_2, $\phi_2(x_i)$. Assuming this to be the case, then $\phi_3(x_i)$ would resemble $\phi_2(x_i)$ rotated through 90°.

Beyond this point, the complexities of the two (or more) dimensional case becomes too complex to describe simply. It is easily recognized that the same influences that were seen in the one-dimensional case are playing an even more important role; namely that for the rows of C corresponding to points near the boundary of the domain considered nearly half of the "global" covariance function lies outside the domain and is consequently ignored. This strongly influences the shape of the first proper function, $\phi_1(x_i)$, and thence strongly influences the next step, the construction of $C^{(2)}$. The effect is passed on step by step until the terms of $C^{(k)}$ are dominated either by rounding in the computer or the irregularities of sampling.

In two dimensional regions that have rotational symmetry one encounters pairs of identical proper values. In such cases, the proper functions are related to each other by the same rotation. Further, which proper function comes first seems to be due to rounding. Even the topography of the function selected is not well determined. For example, if $\lambda_2 \equiv \lambda_3$ and $\phi_2(x_i)$, $\phi_3(x_i)$ are the corresponding proper functions, then $[a\ \phi_2(x_i) + b\phi_3(x_i)]$ and $[b\phi_2(x_i) - a\phi_3(x_i)]$ are also proper functions corresponding to the proper value provided $a^2 + b^2 = 1$. Though multiple proper values rarely occur except in mathematical exercises, pairs of proper values that are relatively close together are not uncommon and are associated with a certain amount of ambiguity as to which proper function is associated with which proper value (Buell, 1979). They also occur in connection with domains that are rotationally symmetric (See Törnevik, 1977 where $\lambda_1 = 38.2$, $\lambda_2 = 33.2$ while $\phi_1(x_i)$ and $\phi_2(x_i)$ are nearly the same if one is rotated 90° and $\lambda_6 = 1.6$, $\lambda_7 = 1.4$ and the same is true for $\phi_6(x_i)$ and $\phi_7(x_i)$; the domain a 7 x 7 square; the property concerned is sea level pressure.)

6. CONCLUSIONS

The two-point covariance functions of atmospheric properties such as geopotential, temperature, wind components, etc., show a strong tendency to be very similar from place to place (with the possible exception of the tropics). When a region with well defined boundaries is concerned, the EOF's computed over this region are expected to be very strongly influenced by the geometrical shape of the region and to a large extent independent of where the region is located. As a consequence, the interpretation of the topography of the EOF's in terms of geographical area and associated meteorological phenomena should be looked on with suspicion unless the influence of the effect of the shape of the region has been completely accounted for. Otherwise, such interpretations may well be on a scientific level with the observations of children who see castles in the clouds.

7. REFERENCES

Anderson, T.W., 1958; *Introduction to Multivceriate Statistical Analysis* John Wiley and Sons, Inc. N.Y., xii + 374 pp (See Ch. 11, pp. 272-287).

Buell, C. Eugene, 1975; The topography of empirical orthogonal functions. *4th Conf. on Prob. and Statis, in Atmos. Sci., Nov. 18-21, 1975,* Amer. Meteor. Soc., Boston, Mass, pp. 188-193.

Buell, C.E., 1978; The number of significant proper functions of two dimensional fields. *J. Appl. Meteor.,* Vol. 17, 717-722.

Buell, C.E., 1979; Reply (to comment by Glahn on Buell, 1978), *J. Appl. Meteor.*, Vol. 18, pp. 714-717.

Törnevik, H., 1977; Application of EOF to sea level forecasting. *Workshop on the use of Empirical Orthogonal Functions in Meteorology, 2-4 Nov, 1977,* European Center for Medium Range Forecasts, Fitzwilliam House, Skimped Hill, Bracknell, Berkshire, England, pp. 112-133.