letter to the editor[1]

Comments on "Empirical Orthogonal Function-Analysis of Wind Vectors Over the Tropical Pacific Region"

Legler had to estimate n = 477-dimensional EOFs by means of m = 216 sample vectors $\mathbf{x}_1, \ldots \mathbf{x}_m$. The conventional way is to estimate the n*n 2nd moment matrix **X** by:

$$\hat{\mathbf{X}} = \frac{1}{m} \mathbf{WW}'$$

where **W** is the n*m matrix containing the sample x_i as i.th column and **W'** its complex conjugate transpose, and to calculate the EOFs $\hat{\mathbf{y}}_i$ as eigenvectors of $\hat{\mathbf{X}}$. The statements given below hold also if the considered data are real. In this case, **W'** is the transposed matrix.

If n is large, as it is in Legler's paper, it becomes difficult and time consuming to compute the eigenvalues and eigenvectors of the n*n matrix **WW'**. But if the number m of samples is less than n, the problem may be reduced by a simple trick, which is based on the statement:

A) If $r \neq 0$ is an eigenvalue of **W'W** with multiplicity s and eigenvectors $\mathbf{y}_1, \ldots \mathbf{y}_s$, then r is also an eigenvalue of **WW'** with the same multiplicity and eigenvectors $\mathbf{Wy}_1, \ldots \mathbf{Wy}_s$.

Proof: Let $r \neq 0$ be an eigenvalue of **W'W** with eigenvector $y \neq 0$, i.e. $\mathbf{W'Wy} = r\mathbf{y}$. The assumption $\mathbf{Wy} = 0$ leads to $r\mathbf{y} = \mathbf{W'Wy} = 0$, thus $r = 0$ contradicting $r \neq 0$. Thus, we have $\mathbf{Wy} \neq 0$.

$$(\mathbf{WW'})(\mathbf{Wy}) = \mathbf{W}(\mathbf{W'W})\mathbf{y} = r(\mathbf{Wy})$$

proves that **Wy** is eigenvector of **WW'** to eigenvalue r. r has the same multiplicity for **WW'** as for **W'W** if the vectors $\mathbf{Wy}_1, \ldots \mathbf{Wy}_s$ are linearly independent, which can easily be shown.

WW' is an n*n matrix and **W'W** a m*m one. **W'W** has m eigenvalues and **WW'** n. Each nonzero eigenvalue of **W'W** is one of **WW'** according to (A). It follows that maximally, min(n,m) useful EOFs can be estimated.

Summarized, the algorithm is as follows: If $n \leqslant m$, solve directly

$$\hat{r}_i \hat{y}_i = \hat{X} \hat{y}_i$$

If $n > m$, solve

$$\hat{r}_i z = \frac{1}{m} \mathbf{W'W\,z}$$

and define as EOF:

$$\hat{y}_i := \frac{\mathbf{Wz}}{\|\mathbf{Wz}\|}$$

where $\|\ \|$ denotes the Euclidean norm and $\hat{r}_i$ the eigenvalues.

In Legler's example, the matrix is reduced from 227 529 to 46 656 coefficients, which is a significant reduction.

Hans von Storch and Gerhard Hannoschöck
Meteorological Institute
University of Hamburg
Fed. Rep. Germany

[1] This section of the Bulletin is made available to members who wish to express opinions about problems of concern to the AMS. (For guidelines followed in accepting letters, see "Minutes of the Council," Bulletin of the AMS, **51**, p. 40, ¶4; **51**, p. 434, ¶7.) The opinions expressed in "Letters to the Editor" are those of the writers and do not represent the official position of the American Meteorological Society.

© 1984 American Meteorological Society

Response

I would like to thank the authors of the letter for indicating an interesting and useful trick for decreasing the size and difficulty in determining the eigenvectors and eigenvalues in the EOF problem. This technique will enable users to solve larger data sets when $M \leqq N$.

I became aware of this useful idea after completing the analysis. In addition, there is another way of reducing the size of the matrix. By simply inverting the $N \times M$ matrix to an $M \times N$ matrix, i.e., inverting the spatial and temporal data, the reduction in coefficients for this particular problem would be exactly the same as mentioned by Storch and Hannoschöck. In this way, the eigenvectors would be the temporal variability and the computed coefficients would contain the spatial patterns. Thus, these roles are reversed from my previous paper. Once again, this inversion is only advantageous when $M \leqq N$.

An interesting fact included in this discussion, is the cost of computing the eigenvectors. With the vector processors used today, (CRAY 1, CYBER 205 for example) the dollar cost of computing the eigenvectors is only a portion of the cost of using EOFs. Computing the coefficients, plotting the results, and handling the data require as much computer cost as computing the eigenvectors.

Using the recommended techniques will enable the user to reduce the central memory needed and provide an economical way to find EOFs of large data sets.

David M. Legler
Department of Meteorology
Florida State University
Tallahassee, FL 32306

© 1984 American Meteorological Society