# VIRTUAL SENSOR TECHNOLOGY FOR MARS EXPLORATION.

**Howard Houben,** *Bay Area Environmental Research Institute, Sonoma, CA 95476 (houben@humbabe.arc.nasa.gov)*.

With the proliferation of atmospheric observing instruments on and around Mars, some consideration should be given to techniques that facilitate intercomparison of their measurements and results. A typical challenge might be to use the measurements of one instrument to infer what the other would observe. In a pilot study of a terrestrial problem of this type, Srivastava et al. [2005] attempted to infer MODIS channel 6 (1.6$\mu$m) radiances — which are useful in discriminating clouds from snow-covered terrain — from 5 other MODIS channels (1, 2, 20, 30, and 31), which correspond well to AVHRR channels (with much longer time coverage). Three machine learning methods (essentially nonlinear regression methods) were used. The most successful was a multi-layer perceptron, a type of neural network. Neural networks have been previously used for nonlinear principal component analyses of meteorological datasets [Hsieh, 2001]. However, for many purposes a straightforward linear principal component analysis is adequate.

A principal component analysis determines the eigenvalues and eigenvectors of the correlation matrix of a system. Typically, only a few of the eigenvalues are large and the corresponding eigenvectors — usually referred to as empirical orthogonal functions or EOFs in the meteorological context — represent the most important modes of variation of the system. Such EOFs have been used by meteorologists for some 50 years [Lorenz, 1956] to reduce the dimensionality of the problem for analysis. Important climatic oscillations like ENSO (El Niño - Southern Oscillation) and the North Atlantic Oscillation are EOFs of the Earth's climate system. Recently, Martinez-Alvarado et al. [2005] have begun to apply such an analysis to Mars GCM output.

Extended empirical orthogonal functions (EEOFs) include not only spatial correlations, but correlations across the time dimension as well [Weare and Nasstrom, 1982] and are therefore keys to the predictability of the system. There have thus been a number of efforts to predict the evolution of terrestrial SST anomalies in general [Smith et al., 1996; Aires and Chedin, 2000] and ENSO in particular [Roulston and Neelin, 2003] with EOFs. As a rule, these have been disappointing. Since the prediction of future meteorological variables based on EEOFs is mathematically equivalent to the virtual sensor problem of inferring unobserved variables using EOFs — and both procedures may be valuable in martian studies — it is worth examining the computational details of the problem for clues as to when success or failure is to be expected.

Correlation Matrix for MODIS Channels

| | 1 | 2 | 6 | 20 | 30 | 31 |
|---|---|---|---|---|---|---|
| 1 | 1. | .9980 | .6287 | .8778 | .8785 | .8784 |
| 2 | .9980 | 1. | .6564 | .8786 | .8774 | .8773 |
| 6 | .6287 | .6564 | 1. | .7369 | .6979 | .6984 |
| 20 | .8778 | .8786 | .7369 | 1. | .9977 | .9977 |
| 30 | .8785 | .8774 | .6979 | .9977 | 1. | 1. |
| 31 | .8784 | .8773 | .6984 | .9977 | 1. | 1. |

We make use of the MODIS correlation matrix of Srivastava et al. [2005]. Only 3 eigenvalues of the system contribute any significant variance (Figure 1). The corresponding eigenspectra are shown in Figure 2. The implications of the principal component analysis are that all MODIS spectra can be decomposed into a linear combination of these EOFs. The virtual sensor problem is to determine that decomposition with incomplete information (i.e., when the radiance in channel 6 is unknown). This problem is mathematically straightforward. The correlation matrix $C$ is diagonalized as

$$UDU^{-1}, \tag{1}$$

where $D$ is a diagonal matrix of eigenvalues and $U$ is the orthonormal matrix whose columns are the eigenvectors. Any given spectrum is characterized by a state vector $y$ whose elements are the coefficients of the corresponding eigenvectors, and the spectrum $x$ is given by

$$x = Uy. \tag{2}$$

To determine the radiance in an unobserved channel, $x_n$, we determine $y$ by inverting the relation

$$x = \tilde{U}y, \tag{3}$$

where $\tilde{U}$ is the (non-square) matrix obtained by removing the $n$th row ($U_n$) of $U$. Finally, we find $x_n$ from the relation

$$x_n = U_n y. \tag{4}$$

Using the generalized matrix inverse, we obtain a one-line expression

$$x_n = U_n(\tilde{U}^T\tilde{U})^{-1}\tilde{U}^T x, \tag{5}$$

which is a simple linear transformation. However, the usability of this formula depends on the conditioning of the matrix inverse. It is easy to show that the matrix

$$\tilde{U}^T\tilde{U} = I - U_n^T U_n, \tag{6}$$

has an eigenvector $U_n^T$ with eigenvalue $1 - U_n U_n^T$. The condition number of the matrix is inversely proportional

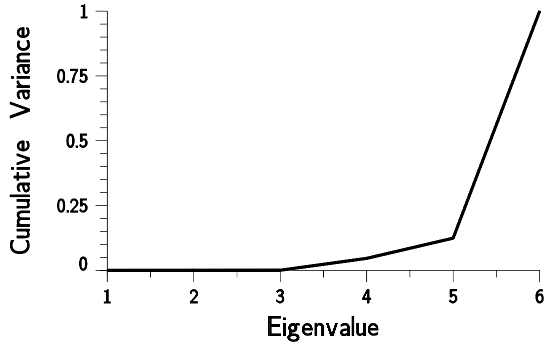


Figure 1: Cumulative fraction of variance explained by eigenfunctions of the MODIS correlation matrix. Only 3 modes are required to reproduce this matrix.

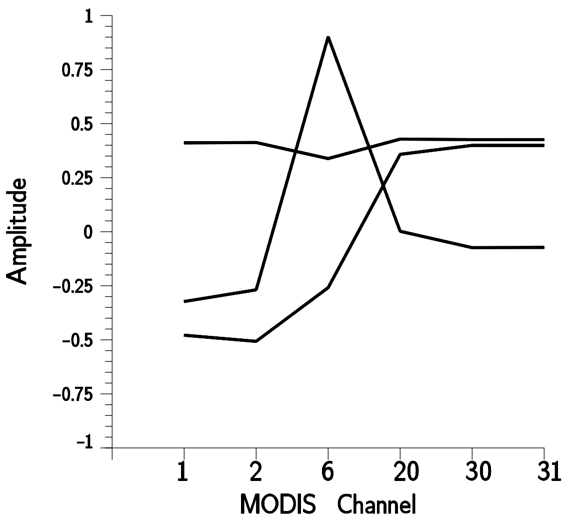


Figure 2: Leading eigenspectra of the MODIS correlation matrix. All these modes load signifi cantly on channel 6, explaining all of its variation.

to this eigenvalue. That means that when $U_n U_n^T \approx 1$ and the variability of $x_n$ is well represented by the EOF model, it is impossible to predict $x_n$. On the other hand, if $x_n$ loads heavily on an eigenvector that is not in the model (i.e., that does not represent any significant variation in the observations), it should be easy to predict from the other channels in the spectrum. This conclusion gives a mathematical basis to the tradeoff between analysis and prediction.

In the Srivastava et al. [2005] attempt to predict MODIS channel 6, it is indeed the case that $U_n U_n^T \approx 1$. More complex nonlinear methods are thus the only hope for solving this problem. On the other hand, the other channels are easily predicted (as could be inferred directly from the correlation matrix).

An analysis of this type would be useful in distributing a meteorology network, say, on Mars. If redundancy is desired, for example, locations should be chosen that are highly correlated (at least in model simulations). If, on the other hand, the intention of the network is to span all of the atmospheric EOFs, uncorrelated locations will be required.

## References

Aires, F., and A. Chedin, 2000. Independent Component Analysis of multivariate time series: Application to the tropical SST variability, *J. Geophys. Res.,* **105,** 17437–17455.

Hsieh, W. W., 2001. Nonlinear principal component analysis by neural networks, *Tellus,* **A53,** 599–615.

Lorenz, E. N., 1956. Empirical orthogonal functions and statistical weather prediction. Technical Report: Statistical Forecast Project Report I, Dept. Meteorology, MIT, 49 pp.

Martinez-Alvarado, O., I. M. Moroz, P. L. Read, and S. R. Lewis, 2005. Reduced-order models of the martian atmospheric dynamics, ENOC-2005, Eindhoven, Netherlands.

Roulston, M. S., and J. D. Neelin, 2003. Non-linear coupling between modes in a low-dimensional model of ENSO, *Atmosphere-Ocean,* **41,** 217–231.

Smith, T. M., R. W. Reynolds, R. E. Livezey, and D. C. Stokes, 1996. Reconstruction of historical sea surface temperature using empirical orthogonal functions, *J. Climate,* **9,** 1403–1420.

Srivastava, A. N., N. C. Oza, and J. Stroeve, 2005. Virtual sensors: Using data mining techniques to efficiently estimate remote sensing spectra, *IEEE Trans. Geoscience and Remote Sensing,* **43,** 590–600.

Weare, B. C., and J. S. Nasstrom, 1982. Examples of extended empirical orthogonal function analysis, *Mon. Wea. Rev.,* **110,** 481–485.