

Nonlinear complex principal component analysis of the tropical Pacific interannual wind variability

Sanjay S. P. Rattan and William W. Hsieh

Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, British Columbia, Canada

Received 6 May 2004; accepted 7 October 2004; published 2 November 2004.

[1] Complex principal component analysis (CPCA) is a linear multivariate technique commonly applied to complex variables or 2-dimensional vector fields such as winds or currents. A new nonlinear CPCA (NLCPCA) method has been developed via complex-valued neural networks. NLCPCA is applied to the tropical Pacific wind field to study the interannual variability. Compared to the CPCA mode 1, the NLCPCA mode 1 is found to explain more variance and reveal the asymmetry in the wind anomalies between El Niño and La Niña states. *INDEX TERMS*: 4215 Oceanography: General: Climate and interannual variability (3309); 3339 Meteorology and Atmospheric Dynamics: Ocean/atmosphere interactions (0312, 4504); 3309 Meteorology and Atmospheric Dynamics: Climatology (1620); 4522 Oceanography: Physical: El Niño; 4504 Oceanography: Physical: Air/sea interactions (0312). **Citation**: Rattan, S. S. P., and W. W. Hsieh (2004), Nonlinear complex principal component analysis of the tropical Pacific interannual wind variability, *Geophys. Res. Lett.*, *31*, L21201, doi:10.1029/2004GL020446.

1. Introduction

[2] Principal component analysis (PCA) also known as empirical orthogonal function (EOF) analysis [von Storch and Zwiers, 1999; Jolliffe, 2002] is a multivariate statistical method widely used to compress datasets and to extract features. Complex PCA (CPCA) is PCA generalized to complex variables. It has been used to analyze 2-dimensional vector fields such as winds [Legler, 1983] and currents [Stacey *et al.*, 1986], where the 2-D vectors are expressed as complex variables. CPCA has also been used to analyze real data complexified first by the Hilbert transform [Horel, 1984].

[3] Linear methods such as PCA has a tendency to scatter the energy of a single oscillatory phenomenon into numerous unphysical modes [Hsieh, 2004]. Nonlinear PCA (NLPCA) via a neural network (NN) approach [Kramer, 1991] has been applied to meteorological/oceanographic datasets, where it has largely alleviated the scattering problem associated with PCA and has revealed the underlying nonlinear structure of the data (see the review by Hsieh [2004]).

[4] For nonlinear feature extraction in the complex domain, the nonlinear CPCA (NLCPCA) method has recently been proposed using a complex-valued NN and applied to the tropical Pacific sea surface temperatures [Rattan and Hsieh, 2004]. This research letter will be the first application of the NLCPCA to a 2-D vector field, the monthly tropical Pacific wind data.

2. Method and Data

2.1. Method

[5] Let $\mathbf{Z} = \mathbf{X} + i\mathbf{Y}$ be a complex matrix with dimension $m \times n$. We take n to be the number of time points and m the number of spatial points, with zero mean in time. A CPCA of \mathbf{Z} seeks a solution that contains r ($r \leq m, n$) linearly independent complex unitary vectors or eigenvectors in the columns of \mathbf{Q} ($m \times r$) such that [Strang, 1988]:

$$\mathbf{Z} = \mathbf{Q}\mathbf{A}, \quad (1)$$

where the rows of \mathbf{A} ($r \times n$) contain the r complex principal component (CPC) time series. The first l CPC can serve as input to the NN for NLCPCA.

[6] The Kramer [1991] auto-associative NN for NLPCA can be adapted to the complex domain (Figure 1) to nonlinearly generalize CPCA. After the layer of input neurons came 3 “hidden” layers of neurons, with the first layer called the encoding layer, followed by the bottleneck layer (with a single complex neuron), then by the decoding layer. A nonlinear transfer function f_1 maps from \mathbf{a} , the input column vector of length l , to the first hidden layer, $\mathbf{h}^{(a)}$, a column vector of length q with elements

$$h_k^{(a)} = f_1 \left[\left(\mathbf{W}^{(a)} \mathbf{a} + \mathbf{b}^{(a)} \right)_k \right], \quad (2)$$

where $\mathbf{W}^{(a)}$ is a $q \times l$ weight matrix, $\mathbf{b}^{(a)}$ is a column vector of length q containing the bias parameters, and $k = 1, \dots, q$. The neurons at the bottleneck, the decoding layer and the output layer are given respectively by

$$u = f_2 \left(\mathbf{w}^{(a)} \cdot \mathbf{h}^{(a)} + \bar{b}^{(a)} \right), \quad (3)$$

$$h_k^{(u)} = f_3 \left[\left(\mathbf{w}^{(u)} u + \mathbf{b}^{(u)} \right)_k \right], \quad (4)$$

$$a_j' = f_4 \left[\left(\mathbf{W}^{(u)} \mathbf{h}^{(u)} + \bar{\mathbf{b}}^{(u)} \right)_j \right], \quad j = 1, \dots, l, \quad (5)$$

(see Rattan and Hsieh [2004] for details of the NLCPCA method). It is well known that a feed-forward NN only needs one hidden layer of neurons for it to model any nonlinear continuous function [Bishop, 1995]. For the forward mapping $u = f(\mathbf{a})$, where u is the nonlinear CPC (NLCPC), this hidden layer is provided by the encoding layer, while for the inverse mapping $\mathbf{a}' = \mathbf{g}(u)$, with \mathbf{a}' the NLCPCA model output, it is provided by the decoding layer. For the typical 1-hidden layer feed-forward NN, the transfer function from the input to the hidden layer is

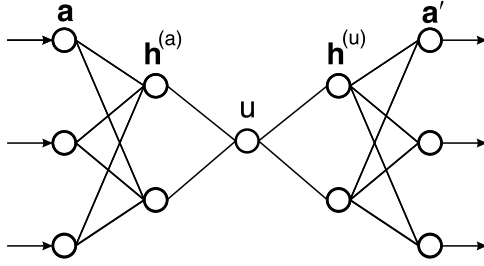


Figure 1. The complex-valued NN model for nonlinear complex PCA (NLCPCA) is an auto-associative feed-forward multi-layer perceptron model. There are l input and output neurons or nodes corresponding to the l CPCs or the number of rows of \mathbf{A} used as input. Sandwiched between the input and output layers are 3 hidden layers (starting with the encoding layer, then the bottleneck layer and finally the decoding layer) containing q , 1 and q neurons respectively. The network is composed of two parts: The first part from the input to the bottleneck maps the input \mathbf{a} to the single nonlinear complex principal component (NLCPC) u by the functions f_1 and f_2 . The second part from the bottleneck to the output \mathbf{a}' is the inverse mapping by the functions f_3 and f_4 . For auto-associative networks, the target for the output neurons are simply the input data. Increasing the number of neurons in the encoding and decoding layers increases the nonlinear modelling capability of the network.

nonlinear, while the transfer function from the hidden layer to the output is usually linear [Bishop, 1995]. Hence the transfer functions f_1, f_2, f_3, f_4 are respectively nonlinear, linear, nonlinear and linear, where the linear function is simply the identity function.

[7] The nonlinear complex transfer function that is used is the hyperbolic tangent ($\tanh(z)$), with certain constraints on z . In the complex plane $\tanh(z)$ has singularities at $(\frac{1}{2} + p)\pi i$, $p \in \mathbb{N}$ and these have to be removed to achieve convergence [Kim and Adali, 2002]. If the magnitude of z is constrained within a circle of radius $\frac{\pi}{2}$ then the singularities do not pose any problem and the transfer function is bounded. This requires a restriction on the magnitudes of the input data and the (weight and bias) parameters: Each element of the r th row of \mathbf{Z} was divided by the maximum magnitude of an element in that row, so each element of \mathbf{Z} has magnitude ≤ 1 . The parameters were randomly initialized with magnitude ≤ 0.1 , and a weight penalty term was added to the objective function J , i.e.,

$$J = \frac{1}{n} \sum_{j=1}^n \| \mathbf{a}_j - \mathbf{a}'_j \|^2 + p \left(\sum_{k=1}^q \| \mathbf{w}_k^{(1)} \|^2 + \| \mathbf{w}^{(2)} \|^2 + \sum_{k=1}^q \| \mathbf{w}_k^{(3)} \|^2 \right), \quad (6)$$

where the first term on the right hand side is the mean square error between \mathbf{a}' and \mathbf{a} , and the second term is the weight penalty term, with $\mathbf{w}_k^{(1)}$, $\mathbf{w}^{(2)}$ and $\mathbf{w}_k^{(3)}$ denoting respectively the vectors containing all the weight and bias parameters from the hidden layers 1, 2 and 3, and the weight penalty parameter p having typical values from 0.01 to 0.1. During the optimization of J , the real and the imaginary

components of the weight and bias parameters were separated and kept in a single real vector while optimization was done by the MATLAB function “fminunc”. After optimization, the predicted CPC \mathbf{a}' from the model output can be multiplied by the spatial eigenvectors from \mathbf{Q} to give the predicted values.

2.2. Data

[8] The monthly ship and buoy wind data from the Florida State University (FSU) pseudo-stress analysis [Stricherz et al., 1997] were used. Consider a wind field $\mathbf{Z} = \mathbf{X} + i\mathbf{Y}$ where \mathbf{X} and \mathbf{Y} are $m \times n$ matrices of the zonal and meridional components of the wind respectively. These components are calculated from the zonal and meridional wind stress data (\mathbf{L}_x and \mathbf{L}_y): $\mathbf{X} = \mathbf{L}_x / (\mathbf{L}_x^2 + \mathbf{L}_y^2)^{1/4}$, $\mathbf{Y} = \mathbf{L}_y / (\mathbf{L}_x^2 + \mathbf{L}_y^2)^{1/4}$ [Wang and Weisberg, 2000]. The data period is January 1961 through December 1999, covering the whole tropical Pacific from 124°E to 70°W, 29°S to 29°N with a grid of 2° by 2°. After the climatological monthly mean was removed, the data were smoothed by a 3-month running mean.

3. Results

[9] Prior to NLCPCA, traditional CPCA (i.e., complex EOF analysis) was first performed to reduce the dimensions of the data. The first two CPCs accounted for 15.3% and 10.7% of the total variance. The first and the second CPCs were also rotated in the complex plane by 13° and 64° respectively so that the mean value of the argument of the rotated CPCs were nearly 0°, i.e., the variance is mainly along the real axis [Hardy and Walton, 1978]. The spatial anomalies associated with the first 2 CPCA modes are shown in Figure 2, with Figure 2a showing the wind anomalies during maximum El Niño.

[10] The six leading CPCs (with 46% of the total variance) were used as the inputs to the NN model (Figure 1). These input variables were first normalized by removing their mean and the real components were divided by the

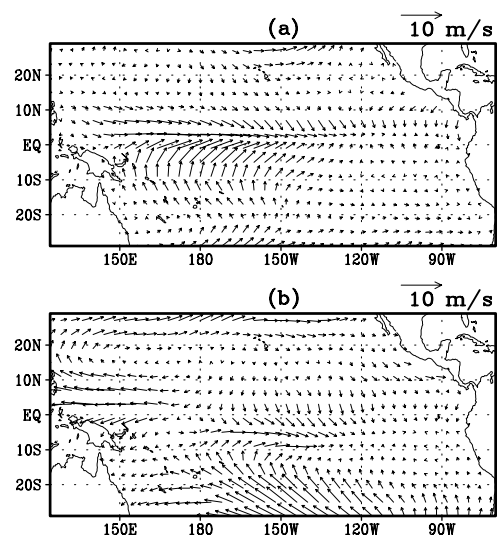


Figure 2. The spatial patterns of the CPCA (a) mode 1 and (b) mode 2 (plotted when the real component of the corresponding CPC is maximum).

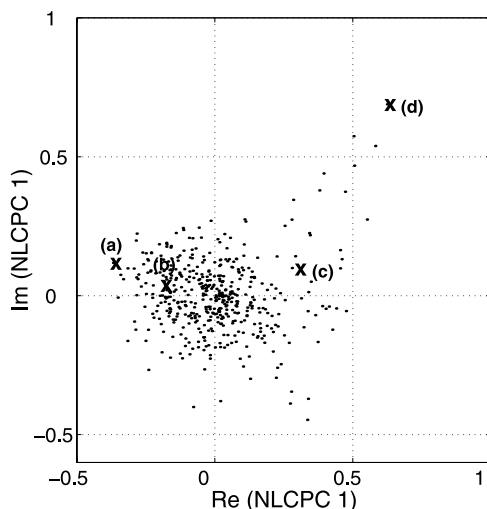


Figure 3. The first NLCPC u shown in the complex plane as dots, with crosses indicating the (a) minimum $\text{Re}(u)$ (strongest La Niña), (b) half minimum $\text{Re}(u)$ (weak La Niña), (c) half maximum $\text{Re}(u)$ (weak El Niño) and (d) maximum $\text{Re}(u)$ (strongest El Niño). The four corresponding spatial anomaly patterns are shown in Figure 4.

largest standard deviation among the 6 real CPCs while the imaginary components were divided by the largest standard deviation among the 6 imaginary CPCs. Division by the individual CPC's standard deviation was not done in order to avoid exaggerating the importance of the higher modes.

[11] The number q of hidden neurons used in the encoding/decoding layer of the NN model was varied between 2 and 10. While a relatively large q tends to give smaller mean square error during the NN training, it also tends to give overfitted solutions due to the relatively large number of network parameters. Based on a general principle of parsimony, $q = 6$ was chosen in this study. Values of the penalty parameter p used ranged from 0.01 to 0.1. For each p ,

25 randomly initialized runs were made. Also, 20% of the data was randomly selected as test data and withheld from the training of the NN model. Runs where the mean square error was larger for the test data set than for the training data set were rejected to avoid overfitted solutions. Among the remaining NN runs, the one with the smallest mean square error was selected as the solution.

[12] The first NLCPC shown in Figure 3 had been rotated by -90° in the complex plane (while the weights in the third hidden layer had also been rotated by 90°). The NLCPCA mode 1 explained 17.4% of the total variance compared to 15.3% explained by the CPCA mode 1. As the NLPC varies, the NLCPCA mode 1 yields nonstationary spatial anomaly patterns, in contrast to the CPCA mode 1 which yields a standing oscillation pattern with the amplitude varying according to the CPC.

[13] Four spatial patterns of NLCPCA mode 1 corresponding to points near the minimum $\text{Re}(u)$, half minimum $\text{Re}(u)$, half maximum $\text{Re}(u)$, and maximum $\text{Re}(u)$ are shown in Figure 4. In Figure 4a (strongest La Niña conditions) the equatorial Pacific displays anomalous easterly winds, with the strongest winds in the equatorial western Pacific. As the negative real component of NLCPC 1 decreases to about half its minimum, the easterly wind anomalies weaken over the equatorial Pacific as shown in Figure 4b to about half the maximum La Niña wind magnitude.

[14] Under El Niño conditions, the tropical Pacific wind field has reversed in direction (Figures 4c and 4d). In Figure 4d, during maximum El Niño, an easterly wind anomaly is observed in the far western equatorial Pacific together with strong westerly anomalies in the central equatorial Pacific. In contrast to the two La Niña pictures (Figures 4a and 4b) which look quite similar except for the magnitude, the weak El Niño state (Figure 4a) is quite different when compared to the strong El Niño state (Figure 4d), e.g., Figure 4c shows westerly anomalies located further west with much less than half the magnitude of Figure 4d, as well as missing the easterly anomalies at the far western equatorial Pacific and the off-equatorial anomalies.

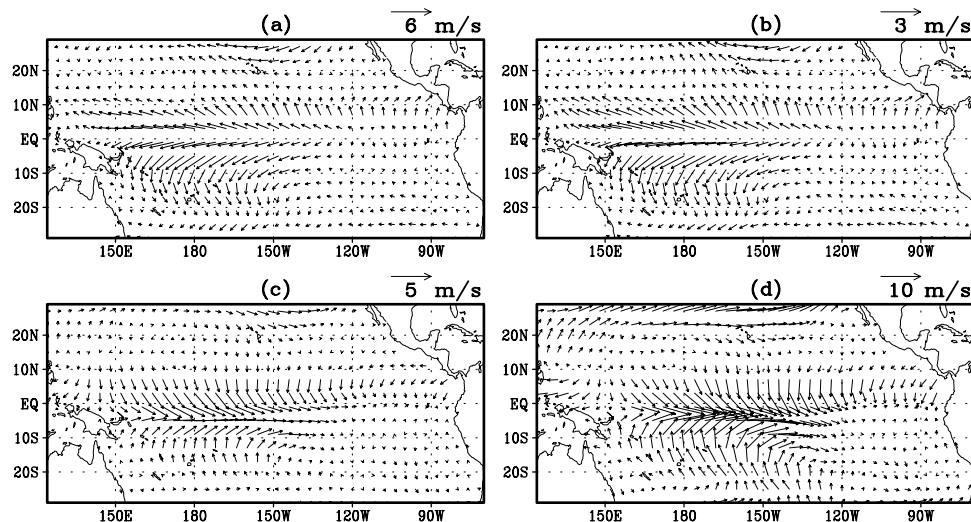


Figure 4. The spatial patterns of the NLCPCA mode 1 showing spatial patterns near the (a) minimum $\text{Re}(u)$ (strongest La Niña), (b) half minimum $\text{Re}(u)$ (weak La Niña), (c) half maximum $\text{Re}(u)$ (weak El Niño) and (d) maximum $\text{Re}(u)$ (strongest El Niño). Different scalings are used, as indicated at the top right corner of each panel.

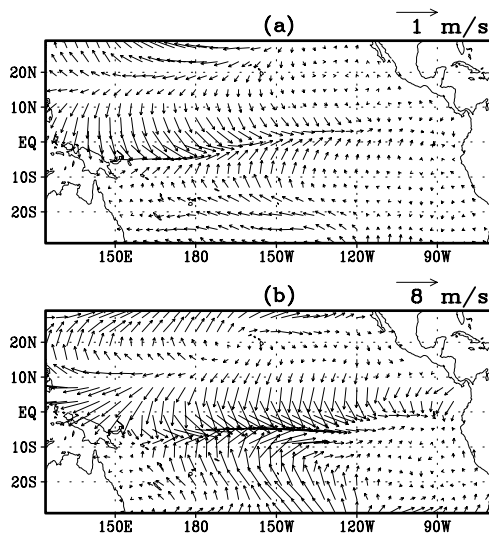


Figure 5. The NLCPCA mode 1 spatial pattern minus the CPCA mode 1 pattern during the (a) strongest La Niña, and (b) strongest El Niño. Different scalings are used in Figures 5a and 5b.

[15] The asymmetry between strong El Niño and strong La Niña is evident from Figure 4a (with anomaly center near 0°N 175°E) and Figure 4d (with center near 5°S 160°W). In contrast, the CPCA mode 1 yields anti-symmetrical stationary patterns for El Niño and La Niña. During maximum real CPC 1 (Figure 2a) the patterns for strong El Niño are captured whereas the minimum real CPC 1 represents the maximum La Niña features. The La Niña spatial patterns when plotted involves a 180° rotation of the El Niño wind directions and look similar to Figure 4a. Hence the CPCA centres for both strong El Niño and La Niña are near 0°N 175°E , i.e., the CPCA mode 1 completely failed to characterize the asymmetry between El Niño and La Niña which results in the asymmetric El Niño-Southern Oscillation (ENSO) features being scattered into CPCA mode 2 (Figure 2b) and higher modes. Compared to NLCPCA mode 1, CPCA mode 1 also substantially underestimated the magnitude of the maximum El Niño (Figure 2a), as well as missing the easterly anomalies in the far western equatorial Pacific and the off-equatorial anomalies found in Figure 4d. Figure 5 shows the difference between the NLCPCA mode 1 and CPCA mode 1 during the strongest La Niña and the strongest El Niño, revealing the difference during the latter to be much greater than during the former, i.e., the CPCA mode 1 does not accurately describe the wind anomalies during strong El Niño conditions.

[16] To test whether the El Niño and La Niña asymmetry have been biased by outliers, we removed the two strongest El Niño episodes and the two strongest La Niña episodes (i.e., a total of 4×12 monthly values) from the input data before NLCPCA was again performed. The resultant spatial patterns again exhibited the asymmetry between El Niño and La Niña.

[17] The NLCPCA mode 2 was extracted from the residual. Again with the NLCPC and CPC rotated so their variance is mainly along the real axis, we found that the correlation between the Southern Oscillation Index (SOI) and $\text{Re}(\text{NLCPC})$ is 0.80 for mode 1 and 0.22 for mode 2. In

contrast, the correlation between SOI and $\text{Re}(\text{CPC})$ is 0.77 for mode 1 and 0.52 for mode 2. In other words, for CPCA, the second mode also contains significant ENSO signal, as the nonlinear ENSO mode cannot be described by a linear mode and is scattered into higher modes, but the NLCPCA mode 1 has been much more effective in extracting the ENSO signal, so the second nonlinear mode is less correlated with the SOI than the second linear mode is.

4. Conclusions

[18] Linear statistical methods such as PCA are often too simplistic to describe real-world systems, with a tendency to scatter a single oscillatory phenomenon into numerous unphysical modes [Hsieh, 2004]. Two-dimensional vector fields like the horizontal wind and ocean currents have commonly used the linear CPCA method for feature extraction. By using a neural network approach, the new NLCPCA method allows a nonlinear generalization of the CPCA. Applied to the tropical Pacific horizontal wind anomaly data, the NLCPCA mode 1 explained 17.4% of the total variance (versus 15.3% for the CPCA mode 1), and gave an accurate description of the ENSO oscillation from strong La Niña to strong El Niño, revealing the considerable asymmetry in the oscillation. The NLCPCA code (written in MATLAB) is downloadable from <http://www.ocgy.ubc.ca/projects/clim.pred/download.html>.

[19] **Acknowledgments.** Dr. Aiming Wu kindly assisted on the GRADS plotting software. This work was supported by the Natural Sciences and Engineering Research Council of Canada and the Canadian Foundation for Climate and Atmospheric Sciences.

References

- Bishop, C. M. (1995), *Neural Networks for Pattern Recognition*, Oxford Univ. Press, New York.
- Hardy, D. M., and J. J. Walton (1978), Principal component analysis of vector wind observations, *J. Appl. Meteorol.*, *17*, 1153–1162.
- Horel, J. D. (1984), Complex principal component analysis: Theory and examples, *J. Clim. Appl. Meteorol.*, *23*, 1660–1673.
- Hsieh, W. W. (2004), Nonlinear multivariate and time series analysis by neural network methods, *Rev. Geophys.*, *42*, RG1003, doi:10.1029/2002RG000112.
- Jolliffe, I. T. (2002), *Principal Component Analysis*, Springer-Verlag, New York.
- Kim, T., and T. Adali (2002), Fully complex multi-layer perceptron network for nonlinear signal processing, *J. VLSI Signal Process.*, *32*, 29–43.
- Kramer, M. A. (1991), Nonlinear principal component analysis using autoassociative neural networks, *AIChE J.*, *37*, 233–243.
- Legler, D. M. (1983), Empirical orthogonal function analysis of wind vectors over the tropical Pacific region, *Bull. Am. Meteorol. Soc.*, *64*(3), 234–241.
- Rattan, S. S. P., and W. W. Hsieh (2004), Complex-valued neural networks for nonlinear complex principal component analysis, *Neural Networks*, in press.
- Stacey, M. W., S. Pond, and P. H. LeBlond (1986), A wind-forced Ekman spiral as a good statistical fit to low-frequency currents in a coastal strait, *Science*, *233*, 470–472.
- Strang, G. (1988), *Linear Algebra and its Applications*, Jovanovich, San Diego, Calif.
- Stricherz, J., D. M. Legler, and J. J. O'Brien (1997), TOGA pseudostress atlas 1985–1994, vol. 2, Pacific Ocean, Florida State Univ., Tallahassee.
- von Storch, H., and F. W. Zwiers (1999), *Statistical Analysis in Climate Research*, Cambridge Univ. Press, New York.
- Wang, C., and R. H. Weisberg (2000), The 1997–98 El Niño and evolution relative to previous El Niño events, *J. Clim.*, *13*, 488–501.

W. W. Hsieh and S. S. P. Rattan, Department of Earth and Ocean Sciences, University of British Columbia, 6339 Stores Road, Vancouver, British Columbia, Canada V6T 1Z4. (whsieh@eos.ubc.ca)