IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol-6, Special Issue-4, May 2018 E-ISSN: 2347-2693

Implementing PCA on MST Radar data for Wind Analysis

M.Anitha[1*], J. Avanija [2]

[1*]Computer Science Engineering, Sri Vidyaniketan Engineering College, Tirupati, India
[2] Computer Science Engineering, Sri Vidyaniketan Engineering College, Tirupati, India

**Corresponding Author: anithanishi1026@gmail.com, Tel.: +9441601122*

Available online at: www.ijcseonline.org

Abstract**—**The data collected from MST radar uses traditional and statistical analysis for inferring wind components from the spectral data. There are several algorithms available for dimensionality reduction on big data using PCA. These algorithms are non -parametric and often implemented on high dimensional datasets. It will be quite interesting to use these analytical algorithms in the context of MST radar dataset. The existing algorithms are very week in estimation of Doppler at low SNR conditions at higher altitudes. Thus PCA algorithm has been applied on the MST Radar data to find Power Spectrum (PS) and from Power Spectrum Doppler Frequency components are estimated. The components are Zonal (U), Meridional (V), Windspeed (W) are estimated from Doppler Frequency. The PCA derived wind data has to be qualified with wind information from GPS radio-sonde thereafter.

Keywords**—** Principal Component Analysis, MST radar, GPS sonde, Wavelet-based denoising, cepstral, thresholding.

I. Introduction

THE Indian MST Radar is positioned at National Atmospheric Research Laboratory (NARL),Gadanki near Tirupati, Andhra Pradesh. The MST Radar is prominent atmospheric radar that operates at a frequency of 53MHz, provides information on atmospheric winds, thus enabling the computation of different parameters of the atmosphere. The Indian MST radar provides information on wind data in the Mesosphere, Stratosphere and Troposphere with a resolution of 150m starting above 3.5km. The signal collected from the radar is first subjected to spectrum estimation techniques, and from the estimated frequencies, the components such as zonal velocity U, meridional velocity V, and wind speed W are determined. The radar uses Doppler beam swinging method to determine the above said three wind components U, V, and W. the spectral data are collected by the radar using multiple beam positions (east, west, zenith-X, zenith-Y, north and south) with 16 µs Inter Pulse Period (IPP). A conventional method of frequency estimation used at the NARL is a software package named Atmospheric Data Processor (ADP). This processes the data in a sequence of steps, which begins by determining the Doppler profile of the radar echoes, thus generating the Doppler frequencies from these profiles. U, V, and W components can be brought out from the radial velocities which in turn came from the frequencies. Wind speed can be computed from the zonal and meridional velocities. The complex time series of the decoded and integrated signal samples are subjected to the process of Fast Fourier Transform (FFT) for on-line computation of the Doppler power spectra for each bin of the selected range window. This technique estimates the Doppler frequencies of the returned echoes accurately upto a certain height (≤18km), beyond this, estimation fails.

The existing algorithms such as Bispectral-based estimation which eliminates the noise using high computational algorithms, Multitaper spectral algorithm which will broaden spectral peak, Wavelet based denoising for estimating the Doppler frequencies and wind components, will all fail at the higher altitudes. In contradictory, PCA is a method where in it computes the wind components at the medium as well as at the higher altitudes with reduced variance and less complexity.

II. Related Work

The authors Amit Kumar Mishra and Bernard Mulgrew [1] has used PCA in SAR-ATR using the MSTAR data base, and comparison has been made with the conventional conditional Guassian model based Bayesian classifier. The results have been compared based on the percentage of the classification, receiver operating characteristics and performance with limited amount of training data. By all means, PCA based classifier was observed to outperform the conditional Guassian model based Bayesian classifier. The classifiers will be compared based on their performance over three criteria. The first criteria are over all the percentage of correct classification. The second criteria are the receiver operation characteristics of the two classifiers were compared. The last will be the reduced amount of training data.

The authors d. Uma Maheswara Rao, T. Sreenivasulu Reddy and G. Ramachandra Reddy [2] has analyzed the data collected from the indian MST Radar at Gadanki using PCA for various simulated signals like narrowband, wideband and exponential signals which may contain more than one frequency both in absence and presence of noise. They have observed that PCA works for low SNR, i.e., it successfully detects the frequency in the highly noise-corrupted signal also. By using PCA, the complexity reduces since a low-rank approximation. PCA is showing an SNR improvement of almost over 10dB over all heights. Standard deviation and Mean Square Error are also low when compared to the previous approaches.

The authors Neetha I. Eappen, T. Sreenivasulu Reddy, and G. Ramachandra Reddy [3] aims at spectral analysis of MST radar signals using a sparse spectrum estimated algorithm termed as SemiParametric/sparse Iterative Covariance-based Estimation. This algorithm was found to successfully estimate the spectrum for simulated data even in the scenario of low SNR. A semiparametric method for spectrum estimation has been applied on MST radar data and found to exhibit better results than the periodogram method even in the presence of noise. It has been analyzed by computing the MSE and output SNR for various SNR values. Since the estimation of power spectrum of the complex data turned out well, it was implemented on the radar data for the computation of the wind velocities.

III. METHODOLOGY

Principal Component Analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. There are two primary reasons for using PCA.

1.Data Reduction: PCA is most commonly used to reduce the information contained in a large number of original variables into a smaller set of new component dimensions, with a minimum loss of information.

2.Interpretation: PCA can be used to discover important features of a large data set. It often reveals relationships that were previously unsuspected, interpretations that would not ordinarily result.

PCA is typically used in data analysis when the number of input variables is too large for useful analysis. PCA should be used mainly for variables which are strongly correlated. The main advantage of PCA is dimensionality reduction. As shown in the Figure 1, the steps involved in the PCA for dimensionality reduction is explained below. Here, firstly we need to key in the data which has transformed using FFT method. Thus, by computing the mean and standard deviation from the data, we need to generate the correlation matrix. Eigen values and Eigen vectors are generated from the correlation matrix to analyze the final results.

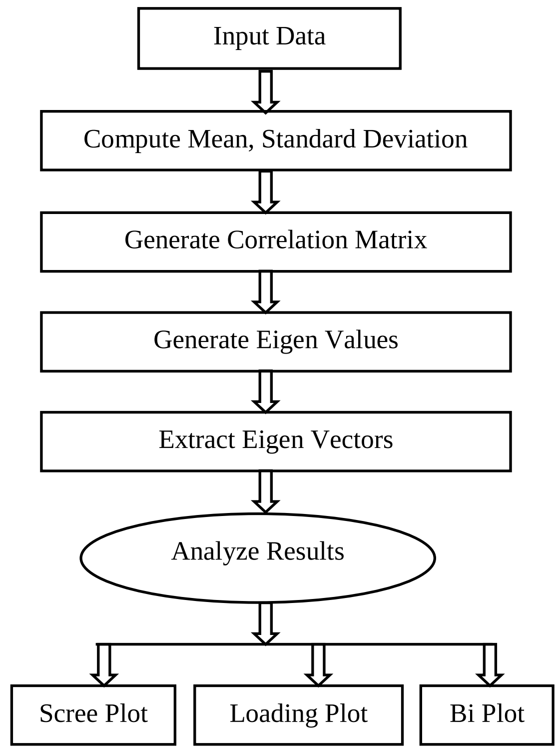

Fig 1 Flow chart steps involved in the PCA

The PCA involves in the computation of mean and standard deviation through the following mathematical formula.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x \rightarrow (1)$$

From (1), we will get the mean from the set of observations called range bin.

$$S = \sqrt{\frac{\sum\left(x-\bar{x}\right)^2}{n-1}} \rightarrow (2)$$

From (2), standard deviation can be brought out from the mean values. Each principal component in Principal Component Analysis is the linear combination of the variables and gives a maximized variance. The correlation matrix can be computed below (3).

$$C = \frac{1}{n-1}(x^T x) \rightarrow (3)$$

Let C be the matrix with X with each column’s mean subtracted from each variable and each scaled to get the covariance matrix. Let X be the matrix for n observations by p variables, and the covariance matrix is C. Let C be the matrix X with each column's mean subtracted from each variable and each column to fetch the correlation matrix.

Then for a linear combination of the variables, from (4) and (5), we need to find the Eigen values and Eigen vectors respectively

$$\det(C - \lambda I) \rightarrow (4)$$

Eigen values are sorted in descending order. The proportion of variance explained by the ith principal component is shown in (4).

$$(C - \lambda I)x \rightarrow (5)$$

Eigen vectors are also known as loadings or coefficients for principal components. Each column in C is the eigen vector corresponding to the eigen value of the principal component. Thus, the values in the descending order will be the principal components.

IV. Results and Discussion

The MST Radar data has been the key input to fetch the principal components.

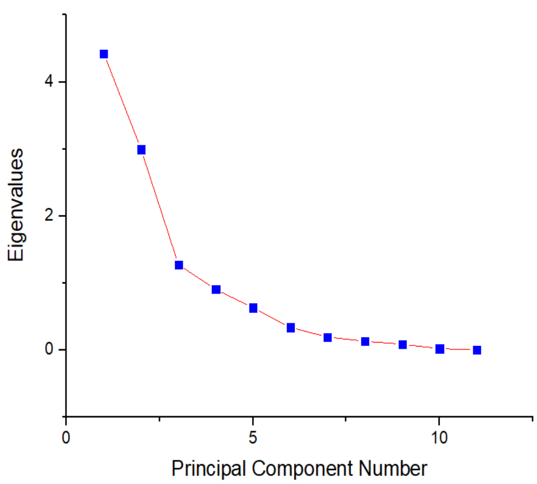

Fig 2. Showing the Eigen values.

A scree plot can be a useful visual aid for determining the appropriate number of principal components. The number of components depends on the "elbow" point at which the remaining eigenvalues are relatively small and all about the same size. This point is not very evident in the scree plot, but we can still say the fourth point is our "elbow" point.

inaccurately. But it is not the case with PCA. Standard

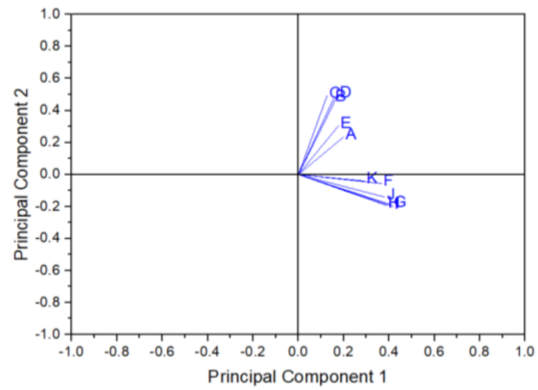

Fig 3. Showing the Principle Components

The **Loading Plot** reveals the relationships between variables in the space of the first two components. In the loading plot, we can see that A, B, C, D and E have similar heavy loadings for principal component 1. F, G, H,J, and K, however, have similar heavy loadings for principal component 2.

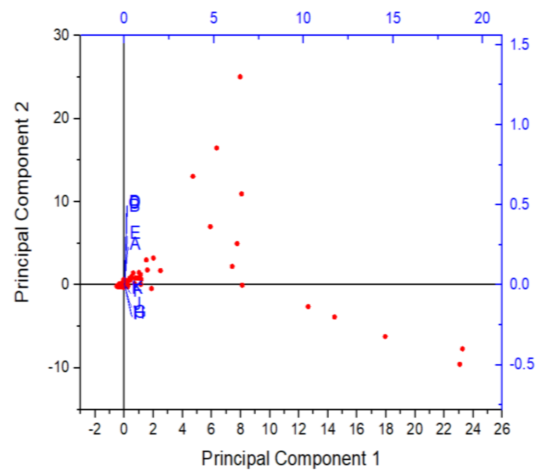

Fig 4. Showing the peak values

The biplot shows both the loadings and the score for two selected components in parallel. It can reveal the projection of an observation on the subspace with the score points. It can also find the ratio of observations and variables in the subspace of the first two components.

V. Conclusion and Future Scope

Thus by using the Principal Component Analysis, the complexity reduces in mathematical formulation and also in fetching the peak values at higher altitudes consistently from the all the sample MST Radar datasets. All the other algorithms fail at those altitudes and calculate the peaks

deviation and mean square error are also low when compared to the previous methods.

ACKNOWLEDGMENT

I would like to thank National Atmospheric Research Laboratory, Gadanki for providing the MST Radar Data to implement the PCA for wind analysis. I would like to thank my external project guide Mr. Manas for guiding and code implementation to generate the results. I would like to thank my internal guide Dr. J Avanija in support to complete this project.

REFERENCES

[1] V.K. Anandan, "*Spectral analysis of atmospheric signal using higher orders spectral estimation technique*", IEEE Transaction, Geosci. Remote Sens. 39 (9) (Sep.2001) 1890-1895

[2] T. Sreenivasulu Reddy, "*MST radar signal processing using cepstral thresholding*", IEEE Transaction Geosci. Remote Sens. 48 (6) (Jun.2010) 2704-2710

[3] Thatiparthi Sreenivasulu Reddy, "*MST radar signal processing using wavelet-based denoising*", IEEE Transaction Geosci. Remote Sens. Lett. 6 (4) (Oct.2009) 752-756

[4] P. Stoica, "*Smoothed non parametric spectral estimation via Cepstral thresholding*", IEEE Signal Process. Mag.23(6) (Nov.2006) 34-45

[5] D.A. Hooper, "*Signal and noise level estimation for narrow spectral width returns observed by the Indian MST radar"*, Radio Sci. 34(4)(1999) 859-870

