JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES
www.journalimcms.org

ISSN (Online) : 2454 -7190 Vol.-17, No.-4, April (2022) pp 21-31 ISSN (Print) 0973-8975

ROTATED EMPIRICAL ORTHOGONAL FUNCTION ANALYSIS FOR SPATIO-TEMPORAL DATA ANALYSIS

Shreyasi Debnath[1], Mourani Sinha[2]

[1]MSc student, Department of Mathematics, Techno India University, West Bengal, EM 4/1, Salt Lake, Sector V, Kolkata 700091, India

[2]Professor, Department of Mathematics, Techno India University, West Bengal, EM 4/1, Salt Lake, Sector V, Kolkata 700091, India

Email: [1]shreyasidebnathnaturalist@gmail.com, [2]mou510@gmail.com

Corresponding Author: **Dr Mourani Sinha**

https://doi.org/10.26782/jmcms.2022.04.00003

(Received: December 3, 2021; Accepted: March 23, 2022)

Abstract

Given any space-time field, Empirical orthogonal function (EOF) analysis finds a set of orthogonal spatial patterns along with a set of associated uncorrelated time series or principal components (PCs). Spatial orthogonality and temporal uncorrelation of EOFs and PCs respectively impose limits on the physical interpretability of EOF patterns. This is because physical processes are not independent, and therefore physical modes are expected in general to be non-orthogonal. Rotated empirical orthogonal functions (REOF) were introduced to generate general localized structures by compromising some of the EOF properties such as orthogonality. EOF and REOF analysis are carried out for the significant wave height (SWH) data for the Bay of Bengal (BOB) region for the period 1958 to 2001. Separate experiments were conducted for all the months together and also for July and December representing the southwest and northeast monsoon periods. The first eigenmodes account for 84%, 68% and 59% of the total variability for the above three cases respectively. The REOF proved to be more effective than EOF for the above region.

Keywords: Rotated empirical orthogonal functions, Principal components, Data analysis, Significant wave height, Bay of Bengal.

I. Introduction

The Empirical orthogonal function (EOF) analysis is a powerful tool for data compression and dimensionality reduction. The EOF technique decomposes the space-time distributed data into spatial modes ranked by their temporal variances. Typically, the EOFs are found by computing the eigenvalues and eigenvectors of a spatially weighted anomaly covariance matrix of a field. The technique compresses the spatial variability of the data into a few eigenmodes. As a result, a set of spatial modes and the associated temporal amplitude functions are obtained. The spatial modes provide information about spatial patterns, while the temporal amplitude functions describe the dynamics. EOF techniques go back to Hotelling (1933) who introduced principal component analysis (PCA), another name for EOFs. A review of PCA or EOFs can be found in Kutzbach (1967). EOFs, however, are not restricted to multivariate statistics or atmospheric sciences. They extend to the analysis of stochastic fields in the mathematical literature where they are known under the name Karhunen-Lo`eve basis functions (Lo`eve 1978). Detailed analyses of EOFs can be found for example in Wilks (1995), von Storch and Zwiers (1999), and Jolliffe (2002).

The Empirical Orthogonal Function technique aims at finding a new set of variables that capture most of the observed variance from the data through a linear combination of the original variables. Since EOFs have been introduced in atmospheric science by Lorenz (1956), it has become a statistical tool of fundamental importance in the atmosphere, ocean, and climate science for exploratory data analysis and dynamical mode reduction. Craddock (1973) discusses problems in analyzing multivariate data in meteorology. The EOF terminology is due to Lorenz (1956) who applied it in a forecasting project at the Massachusetts Institute of Technology.

The simplicity and the analytic derivation of EOFs are the main reasons behind its popularity in atmospheric science. The physical interpretability of the obtained patterns is, however, a matter of controversy because of the strong constraints satisfied by EOFs, namely orthogonality in both space and time. Physical modes such as normal modes (Simmons et al. 1983) are not in general orthogonal. This shortcoming has led to the development of rotated empirical orthogonal functions (REOFs), Richman (1986). REOFs yield in general localized structures by compromising some of the EOFs properties such as orthogonality. EOFs and REOFs are mainly based on using the spatial correlation of the field, an important feature of climate data. Extended EOFs (Weare and Nasstrom, 1982) is a technique that attempts to incorporate both the spatial and the temporal correlation. The method has since become a useful tool to extract dynamic structure, trends and oscillations, and filtering data.

By construction EOFs constitute directions of variability with no particular amplitude and are stationary structures, i.e. they do not evolve in time. The principal component attached to the corresponding EOF provides the sign and the overall amplitude of the EOF as a function of time. This provides a simplified representation of the state of the field at that time along that EOF. In other words, EOFs do not change the structure in time, they only change signs and overall amplitude to represent the state of the atmosphere. When EOFs are nondegenerate they can be studied individually. When they are degenerate, however, the separation between them becomes problematic despite being orthogonal and their PCs uncorrelated.

Jolliffe (2002) stated that the central idea of principal component analysis (PCA) or EOF analysis is to reduce the dimensionality of a data set consisting of a large number of interrelated variables while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the principal components, which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables.

The review of Hannachi et al. 2007 focused on five different methods based on EOFs to analyse various climate data. The methods considered here are conventional EOFs, REOFs, simplified EOFs, extended EOFs, and complex/Hilbert EOFs. They began by reviewing the conventional EOFs method. In particular, they have highlighted its benefits, such as easy computation, efficient data reduction, and use geometric properties. They have also presented its major drawbacks, such as predictable relations between EOFs, and physical interpretability. Next we have reviewed REOFs, which have been introduced to overcome some of the previous drawbacks related to orthogonality or uncorrelatedness of EOFs or PCs respectively and also interpretation. The method attempts to rotate a fixed number of EOF patterns using either an orthogonal or oblique rotation matrix subject to maximizing a simplicity criterion. The EOFs can be either unscaled or scaled by the square root of the associated eigenvalues. Various criteria exist in the literature, but the overall result is that all rotations can be classified into four classes: (i) orthogonal rotation of EOFs, (ii) orthogonal rotation of EOFs scaled by the square root of the associated eigenvalues, (iii) oblique rotation of EOFs, and (iv) oblique rotation of scaled EOFs.

Simplified EOFs, make use of some useful properties of EOFs and REOFs simultaneously. It attempts to achieve simultaneously successive variance maximisation, spatial orthogonality of EOFs, and simplicity of REOFs. This is achieved by solving the same eigenvalue problem of EOFs but with an extra constraint of simplicity that depends on a threshold parameter. The obtained optimisation is non-quadratic and involves using advanced numerical methods based on numerical solutions of ODEs. A threshold parameter of the order of $\frac{\sqrt{p}}{3}$, where p is the number of variables, is found to provide a reasonable balance between variance maximization and simplicity of the patterns.

Extended EOFs (or EEOF) are presented as a way to overcome some of the shortcomings of EOFs, namely the use of only spatial correlation. EEOF makes use of spatial as well as temporal correlation by extending the familiar state vector by including explicitly the time information after choosing the delay parameter. The method can be used as a tool to filter the data, isolate a trend, or even separate an oscillatory component buried in the noisy data. The complex EOFs method constitutes another approach used to identify propagating disturbances. Unlike EEOFs, complex EOFs use the complex formulation of propagating waves and involve the Hilbert transform of the field to form the complexified field with no parameter to fix.

Significant wave height (SWH) is a very important wave parameter in ocean engineering studies. Spatial variation of SWH is crucial for studying wave energetics, while analysis of its temporal variation contains elements of wave dynamics. Empirical orthogonal function (EOF) is a widely recognized technique for analyzing such spatio temporally distributed data as two dimensional fields of SWH simulated by a numerical model. The spatial modes provide information about spatial patterns, while the temporal amplitude functions describe the dynamics. The first eigenmode representing maximum variability of the total variance suggests that the spatial structure of the ocean basin is more or less regular in a hydro-dynamic sense. At the same time, a low value for the first eigenmode is representative of the fact that the ocean basin is most probably multi-connected (with the presence of islands or complex geometry and bathymetry).

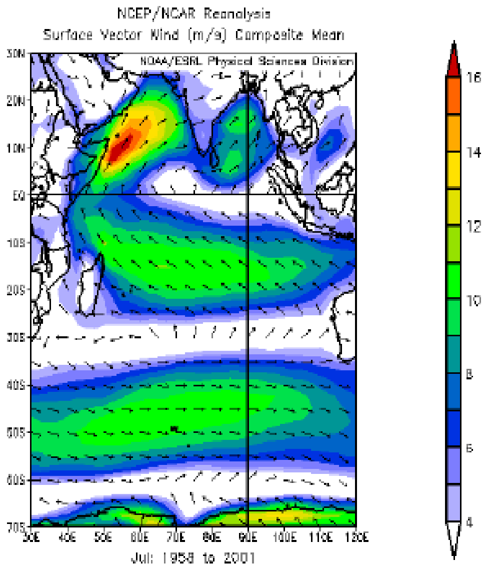
In this work, we have applied EOFs and REOFs to the SWH data of the Bay of Bengal (BOB) region for the period 1958 to 2001 to understand the Spatio-temporal pattern of the parameter. There are very few studies in which the leading eigenmodes of SWH variability are deduced from REOF analysis.

II. Data

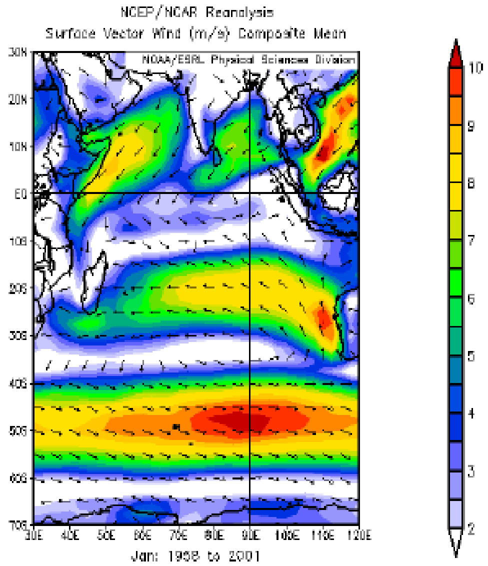
Six hourly SWH data from 1958 to 2001 (44 years) is downloaded from ERA-40 analyzed ocean wave dataset of the European Centre for Medium-Range Weather Forecasts (ECMWF). The spatial resolution of the dataset is one degree by one degree. EOF analysis has been carried out over the BOB region (78E to 98E and 25N to 5N) for the above 44 years. For the BOB region, EOF analysis is performed first for all the months and then separately for July and December. To obtain more localized or nonorthogonal features REOF were used on the initially obtained EOFs.

III. Results and Discussions

Winds over the BOB reverse twice annually and as a result, the region is forced by seasonally reversing monsoon winds. During winter, dry northeasterly winds blow over the BOB resulting in the northeast monsoon, while during summer, there is a wind reversal, with the moist southwesterly wind blowing over the region, resulting in the southwest monsoon. Forced by these winds, circulation in the BOB has a generally eastward direction during summer and westward during winter. Thus, it follows that understanding the variability of wave heights in the BOB requires prior knowledge of the wind variability. Figures 1a and 1b show the monthly means of the surface wind vector during the southwest and northeast seasons from NCEP/NCAR reanalysis dataset for the period 1958-2001 over the Indian Ocean.



(a)



(b)

Figure 1a: Averaged surface wind vector over the Indian Ocean for July

Figure 1b: Averaged surface wind vector over the Indian Ocean for January

In the first experiment, the SWH data for all the months from January to December for all the 44 years (1958-2001) are subjected to EOF analysis. The first and second eigenmodes account for 84% and 6.5% of the total variability of the SWH data for the BOB region. The cumulative first three eigenvectors account for 93.3% of the total variability. Figures 2a and 2b give the first spatial mode for the EOF and REOF respectively. Similarly figures 3a and 3b represent the second spatial mode. Figure 4 shows the time series (44 years) for PC1 corresponding to the first EOF, in which the annual signal is clearly visible.

From the first spatial mode (EOF1) depicted in Figure 2a, one can see that the maximum amplitude is observed in the northeast BOB. The minimum amplitude is seen along the southeast coast of India. The amplitude is increasing towards the north along the coast. The year to year variation of the time-series values for PC1 corresponding to the first EOF is depicted in Figure 4. The plot demonstrates that the temporal variations are of annual periodicity with a maximum occurring during the southwest monsoon period.

The possible explanation of the largest loading in the northeast part of the central Bay may be the following. The maximum wind speed occurs during the southwest monsoon. Accordingly wave height should in general be maximum in these months as reflected in the plot for the PC1 in Figure 4. During the southwest monsoon, the dominant direction of propagation of the waves should thus be north-eastward. Considering these facts, it is natural to expect that the maximum loading should occur in the northeast corner of the Bay. Another possible reason could be the north-eastward propagation of southern ocean swells.

Figure 2b showing the first spatial mode of the rotated EOFs has a similar maximum loading towards the head of the Bay. The area of the minimum loading is much more expanded. The second spatial mode (EOF2) in figure 3a gives a spatial pattern of maximum loading at the head of the Bay but the rotated EOF2 is more localized and shows the maximum loadings at the south-east coast of India.

Since rest of the eigenvectors contribute insignificant variability, the corresponding plots are not shown in this work.

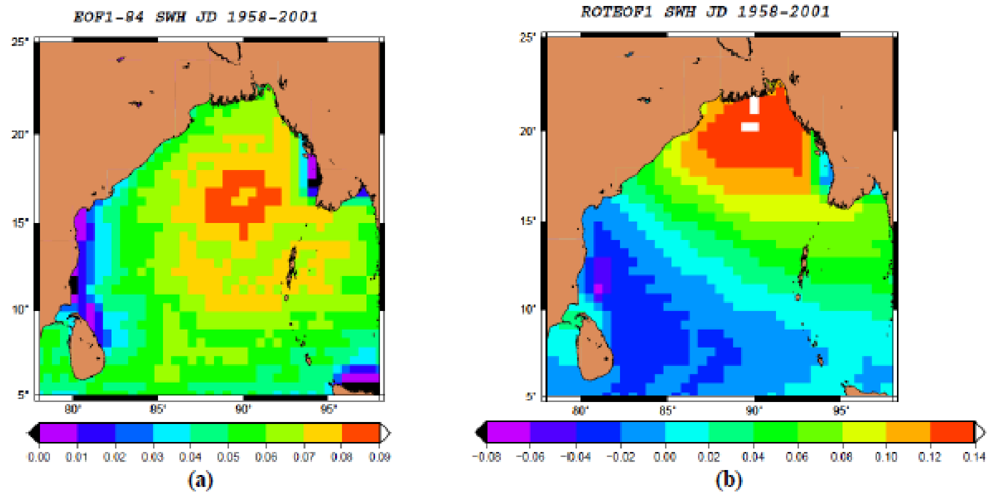


Figure 2a: The first spatial EOF mode for SWH analysis
Figure 2b: The first spatial REOF mode for SWH analysis

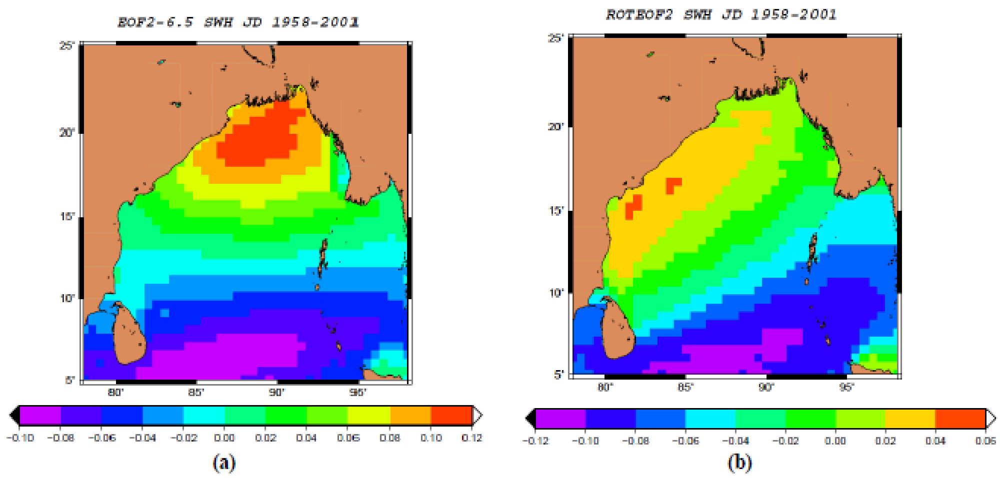


Figure 3a: The second spatial EOF mode for SWH analysis
Figure 3b: The second spatial REOF mode for SWH analysis

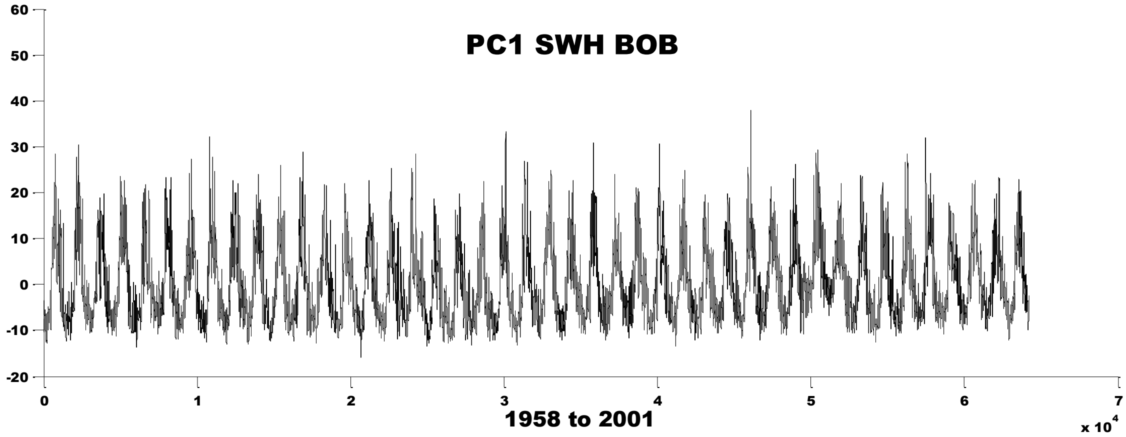
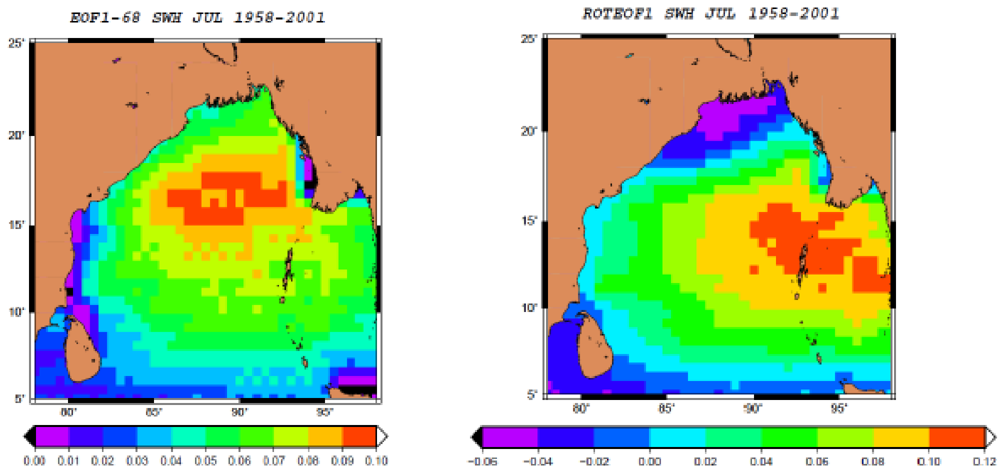


Figure 4: The first principal component for the SWH analysis (44 years)

Next experiments were conducted separately for the July and December months. EOF and REOF were applied to SWH data for the period 1958 to 2001 to extract the dominant modes of variability.

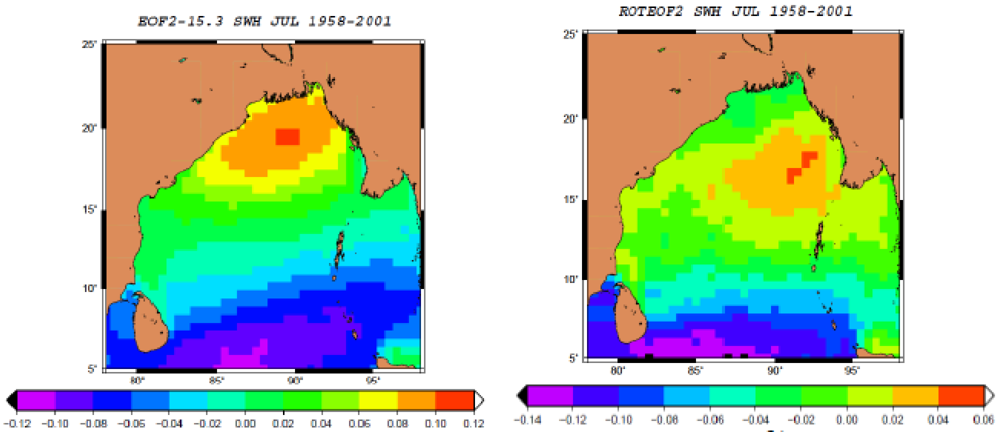
For July the first and second eigenmodes account for 68% and 15.3% of the total variability of the SWH data. In December the second eigenmodes account for 59.3% and 12.3% of the total variability. The cumulative first three eigenvectors account for 88% in July and 81.5% in December of the total variability. The month of July with stronger wind speeds exhibits larger variability as expected.

Figures 5a and 5b give the first spatial mode for the EOF and REOF for July respectively. Spatial patterns in figures 6a and 6b represent the second spatial mode. If we again compare EOF and REOF the patterns are more precise and localized. In December compared to July the EOF1 and REOF1 (figures 7a and 7b) show maximum loadings at lower latitudes. The northeastward propagation of the waves is maximum during the southwest monsoon. The second mode (figures 8a and 8b) gives a distinct pattern of maximum loading on the southeast coast of India in December.



(a) (b)

Figure 5a: The first spatial EOF mode for SWH analysis during July
Figure 5b: The first spatial REOF mode for SWH analysis during July



(a) (b)

Figure 6a: The second spatial EOF mode for SWH analysis during July
Figure 6b: The second spatial REOF mode for SWH analysis during July

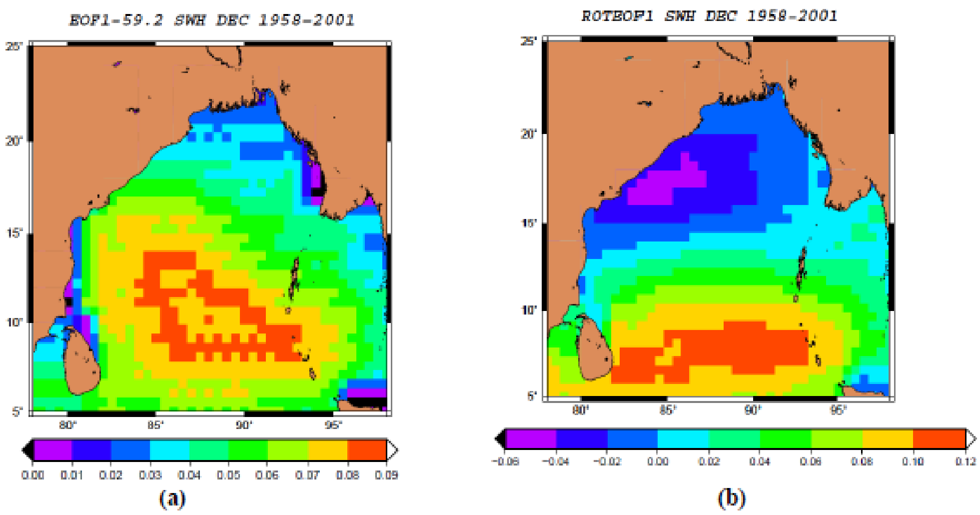


Figure 7a: The first spatial EOF mode for SWH analysis during December
Figure 7b: The first spatial REOF mode for SWH analysis during December

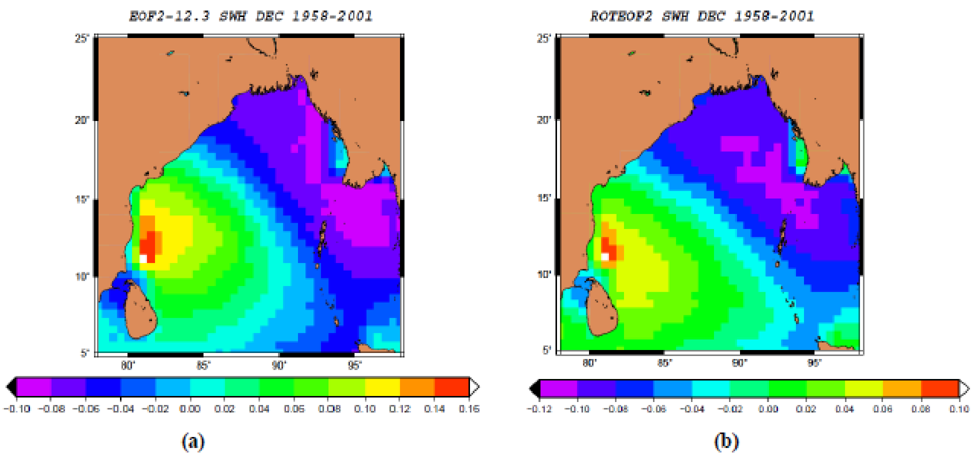


Figure 8a: The second spatial EOF mode for SWH analysis during December
Figure 8b: The second spatial REOF mode for SWH analysis during December

IV. Conclusions

SWH data of the BOB region have been subjected to EOF and REOF analysis for identifying the dominant modes of variability. Analysis has been done for all the months together for 44 years and also separately for July and December for the same period (1958-2001). The first and second eigenmodes account for 84% and 6.5% of the total variability for all the months together. The variability reduces to 68% and 15.3% for July only and 59.3% and 12.3% for December only. In winter the sea-states are calmer having less variability. The first EOF has its strongest impact on the regular pattern of southwest monsoon winds in the BOB. The spatial pattern of the first EOF for the BOB region shows the highest loading in the northeast part of the central bay. This is due to the northeastward movement of the waves during the southwest monsoon with its accompanying strong southwesterly winds. Another factor favouring this loading is the northeastward propagation of swells originating in the Southern Indian Ocean. The first principal component corresponding to the first EOF exhibits annual periodicity. The rotated EOF patterns are more precise and localized. Results indicate REOF to be more powerful than EOF. Thus EOF and REOF serve as an effective means of physical interpretability of spatial and temporal patterns within a data.

Conflicts of interest

The authors declare that they have no conflicts of interest to report regarding the present study.

References

I. Craddock, J. M., 1973: Problems and prospects for eigenvector analysis in meteorology. The statistician, 22, 133-145.

II. Crommelin, D. T., and A. J. Majda, 2004: Strategies for model reduction: Comparing different optimal bases. J. Atmos. Sci., 61, 2206–2217.

III. Farrell, B. F., and P. J. Ioannou, 1996: Generalized stability theory. Part I: Autonomous operators. J. Atmos. Sci., 53, 2025–2040.

IV. Hannachi, A., I. Joliffe, and D. Stephenson, 2007: Empirical orthogonal functions and related techniques in atmospheric science: Areview. Int. J. Climatol., 27, 1119–1152, doi:10.1002/joc.1499.

V. Hotelling, H., 1933: Analysis of a complex of statistical variables into principal components. J. Educ. Psych,, 24, 417-520.

VI. Jolliffe, I. T., 2002: Principal Component Analysis. Springer-Verlag, 2nd Edition, New York.

VII. Kleeman, R., 2008: Stochastic theories for the irregularity of ENSO. Philos. Trans. Roy. Soc., 366A, 2509–2524, doi:10.1098/rsta.2008.0048.

VIII. Kutzbach, J. E., 1967: Empirical eigenvectors of sea-level pressure, surface temperature and precipitation complexes over North America. J. Appl.Meteor., 6, 791-802.

IX. Lo`eve, M., 1978: Probability theory, Vol II, 4’th ed., Springer-Verlag, 413pp.

X. Lorenz, E. N., 1956: Empirical orthogonal functions and statistical weather prediction. Technical report, Statistical Forecast Project Report 1, Dept. of Meteor., MIT, 1956. 49pp.

XI. Monahan, A. H., and A. Dai, 2004: The spatial and temporal structure of ENSO nonlinearity. J. Climate, 17, 3026–3036.

XII. North, G. R., T. L., Bell, R. F. Cahalan, and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. Mon.Weather Rev., 110, 699-706.

XIII. North, G. R., 1984: Empirical orthogonal functions and normal modes. J. Atmos. Sci., 41, 879–887.

XIV. Penland, C., 1996: A stochastic model of Indo-Pacific sea surface temperature anomalies. Physica D, 98, 534–558.

XV. von Storch, H., and F. W. Zwiers, 1999: Statistical Analysis in Climate research, Cambridge University Press, Cambridge.

XVI. Weare, B. C., and J. S. Nasstrom, 1982: Examples of extended empirical orthogonal function analysis. Mon. Weath. Rev., 110, 481-485.

XVII. Wilks, D. S., 1995: Statistical Methods in the Atmospheric Sciences. Academic Press, San Diego.