PAPER • OPEN ACCESS

DINEOF reconstruction for creating long-term cloud-free sea surface temperature data records: A Case study in Lombok Strait, Indonesia

View the article online for updates and enhancements.

DINEOF reconstruction for creating long-term cloud-free sea surface temperature data records: A Case study in Lombok Strait, Indonesia

Yochi Okta Andrawina[1], Ratu Almira Kismawardhani[2], Hasti Amrih Rejeki[3]

[1]Erasmus Mundus Joint Master Degree Programme Marine Environment, Département d'astrophys., Géophysique et Océanographie (AGO) Université de Liège - ULiège, Belgium

[2]Graduate School of Science, Department Geophysics, Physical Oceanography Laboratory, Tohoku University, Japan

[3]Indonesian School of Meteorology Climatology and Geophysics, Indonesia

[1]yochiokta@gmail.com

Abstract. A long-term reliable sea surface temperature (SST) satellite data record is requisite resources for monitoring to understand climate variability. Creating a long-term data record especially for climate variability requires a combination of multiple satellite products. Consequently, missing data issues are inevitable. Hence, DINEOF (Data Interpolating Empirical Orthogonal Functions) has been applied to attain a complete and coherent multi-sensor SST data record with EOF-based technique by reconstructing the missing data. Unfortunately, the technique can lead to large discontinuities in the data reconstruction due to images depiction within long time series data. For that reason, filtering the temporal covariance matrix had been applied to reduce the spurious variability and more realistic reconstructions are obtained. However, this approach has not yet tested in tropical region with higher evaporation which cause incomplete satellite image coverage. Therefore, the objective of this research is to reconstruct SST of Lombok strait with data gaps up to 58.16% in one year. It is successfully reconstructed until the last iteration of 42 optimal EOF modes with the convergence achieved up to 0.9806 x10^{-3}, including previous set-aside data for internal cross-validation. The results highlight that the DINEOF method can effectively reconstruct SST data in Lombok Strait.

1. Introduction

Satellite-determined ocean sea surface temperature (SST) is one of the geophysical informational indexes from infrared perceptions that have been broadly utilized in oceanography because of their broad inclusion in reality. These geophysical informational collections, explicitly those acquired from satellites, regularly contain holes (missing qualities) due to the presence of cloud, mists, downpour, or essentially because of deficient track inclusion. Nonetheless, these SST information are frequently affected by mists' quality in the air, breakdowns in the satellite, or pictures commotions, which can cause missing information. Moreover, the deficiency of information may arrive at a high rate in certain periods. A total informational index is alluring or even fundamental for some, applications utilizing satellite-determined information. Satellited-based data is not always producing cloud coverage problem.

Radiometer satellite sensor problem may subject to the other source of limitation such as coastal mask or rain contaminations. Despite, cloud limitation brings many disadvantages to the satellite coverage.

The need is building up to complete informational collections worldwide, regional and local scale especially in the geophysical data sets which located in a tropical area such as Lombok Strait, Indonesia can encounter some problem while obtaining the data from satellites, often contain gaps (missing values) due to this location remain has several presences of clouds, rain, or simply due to incomplete track down coverage may affecting the raw data from the satellite. Several analysis methods, input for hydrodynamic model and data visualization are mostly applied where complete data sets are preferred or even important. Hence, Data Interpolating Empirical Orthogonal Functions (DINEOF) is introduced to help the lost coverage of satellite data and has been applied to SST data, in particular [3,9]. There are presently various strategies, like spline interpolation [10] and optimal interpolation (OI) [11] that have been developed to deal with the recovery of missing data. Unfortunately, both of them, especially OI, need *a priori* information about the error statistics of the data [10,11].

2. Method

2.1. The Data and Study Area
We use level three (L3) global ocean with near real-time observation with high-resolution SST data of $0.1°\times0.1°$ global grid. The data is multiple satellite sensors daily mean SST from January 2019 until November 2020. A bias correction has been applied based on multi-sensor median reference in the intercalibrated technique from CMEMS[5]. It is correcting the large-scale (global) cross-sensor biases for both bias correction and validation. Our area study is defined by ($115.05°$ E-$119.95°$ E) and ($10.05°$ S-$5.05°$N) covering the area with monsoon dynamics [12]. This research aims to reconstruct the sea surface temperature data in Lombok Strait to prove the possibility of DINEOF method can be used in the largest of cloudy area in the tropical sea which condensation may often occur in order to reproduce the clouds and several phenomena such as internal wave, monsoon, and El Niño–Southern Oscillation [12].

2.2. The Data Analysis
The DINEOF as *Data* that has *INterpolating* in an *Emphirical* and *Orthogonal Function* [3,4,9]) can be used to reconstruct a monovariate or multivariate data sets. Regarding [1,2,6], this technique is using a parameter-free technique based on Singular Value Decomposition (SVD) with an iterative EOF decomposition process to calculate missing data in the satellite data sets due to several factors and get the optimal number of modes by cross-validation. In this study, a DINEOF approach was examined to overcome the barriers for multisensory data combinations with considering the relevant information between variables. This extended EOF is used to consider the interrelationships between related variables and infer data at missing spatial and temporal data over the location, due to this multisensory combined record data sets in Lombok Strait.

Basically, this method is established in four steps which consist of an iterative method to calculate the field values at missing positions [1,2]. Firstly, this calculation begins with a spatial and temporal mean is removed from the analysed data, and the missing data are initialized to zero. At that point one EOF is determined from this field, and the missing information is supplanted with the qualities got by the EOF decay. This method is rehashed until a union measure for the missing information esteems indicated by the client is reached. The iterative repetition method is done based on a decomposition into principal modes of variations hereinafter, the missing pixels are reconstructed using a truncated Singular Value Decomposition. In this term, if X is a matrix containing time series of images:

$$X = USV^T \qquad\qquad (1)$$

We note that this U is the spatial pattern of EOFs since V is the temporal EOFs and S remains as the singular values. This temporal covariance matrix is filtered in order to enhance the coherence between successive images [4]. Then the missing values were replaced by truncated EOFs reconstruction:

$$X_r = \sum U_p S_p V_p^T (p = 1 \dots k)$$

(2)

Where Up and Vp are the p the column of the spatial and temporal EOF U and V, respectively and $p= 1...k$. With the new qualities for the missing information, the SVD (*Sp as solitary worth over the p segment*) deterioration is performed once more. The two last advances are rehashed until intermingling is acquired for the missing qualities. At that point, the entire iterative strategy is performed for $k = 2, 3,..., kmax$ EOFs, where *kmax* is a predefined number that ought to change as per the underlying lattice attributes. For every k, an assessment for the missing qualities is acquired. The ideal number of EOFs held for the remaking is dictated by cross-approval: various focuses (regularly 1% of the underlying information) is saved and considered as absent. At each EOF assessment, the mistake between those underlying focuses and their remaking is determined, so the ideal number of EOFs limiting this blunder can be resolved. Alongside the remade information, neighborhood blunder fields mirroring the exactness of the data reconstruction can be likewise produced [2].

Regarding the data from different sensor satellites, the multivariate DINEOF reconstruction of both SST IFREMER and CMEMS global Sea data were organized in two pseudo-2-D matrices Xs (IFREMER SST matrices) and Xm (CMEMS SST matrices), respectively. Particularly, each matrix represents the series of spatial pixels, and the combined matrix (X0) is calculated as follows:

$$X_0 \begin{bmatrix} X_s \\ X_m \end{bmatrix}$$

(3)

These two EOFs are determined, and the entire system is rehashed three EOFs with the ideal number of held modes EOFs to be determined will be dictated by cross-approval: a couple of important information (normally 1% of the complete information) are dismantled toward the start of the technique and hailed as absent [2]. At each EOF cycle, the worth determined from the EOF arrangement is contrasted with the genuine estimation of these hailed data. The number of EOFs that minimizes this difference is retained as the optimal number to reconstruct the data set [4].

3. Results and Discussion

### 3.1. DINEOF Preprocessing of data

The initial data of Lombok Strait is depicted below (Figure 1). The remaining white coverage retrieved from raw satellite data shows the missing values as a grid map. The number of mask land points of this data is 235 over the dimension of the file of 50 x 51x 579. The mean value of the SST initial data is 28.4719 ℃ with a standard deviation of 1.3760.
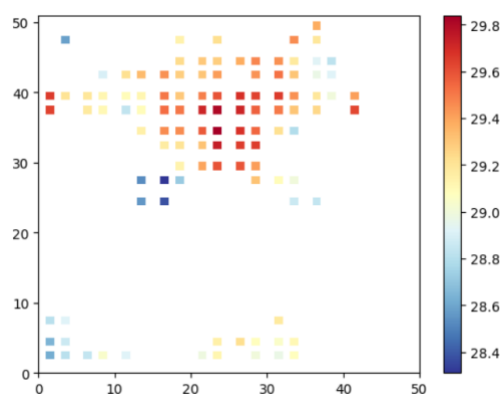
Figure 1. Initial data of SST in Lombok Strait

To create a land-sea mask, DINEOF had been used 2315 x579 size matrix datasets. Figure 2.a. represents the range of missing values over a time scale in the y-axis. In the right part is the grid maps of missing values data in space (Figure 2.b). This initial data as a first DINEOF analysis after three tests had been applied to classify pixels by suspecting the departure from the DINEOF truncated EOF basis, the departure from a local median, and the proximity to clouds, land, or missing values. After diffusion of covariance matrix activated with parameters, alpha is 0.10 and number of iterations is three, 58.16% of the missing data had been taking out, to the next step of cross-validation over 6257 points. Hereafter, in the data is weighted by summing of these three tests to obtain this Figure 3. as initial data of given SST research in the Lombok Strait area.

| (a) | (b) |
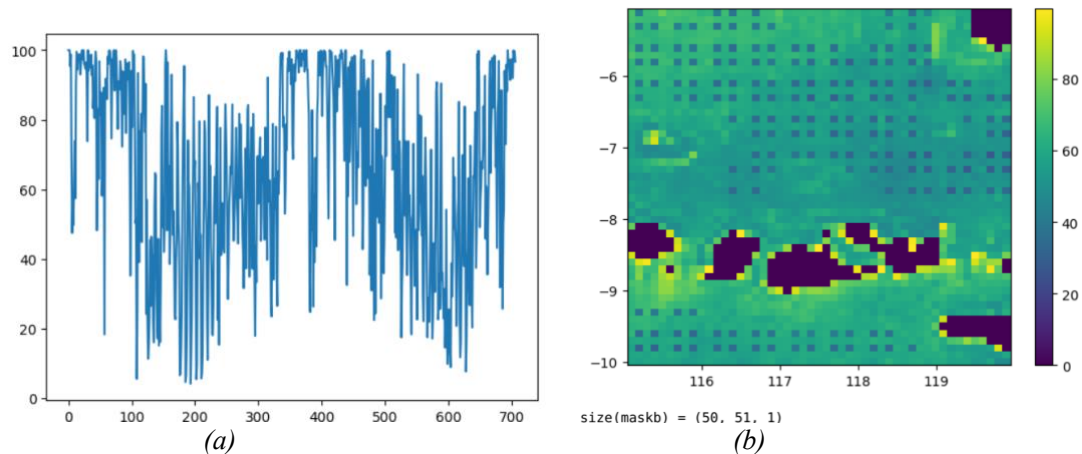
Figure 2. (a) Graph of missing data in time and (b) The percentage of missing data in space from the land-sea mask data

The pixel of the data could set the missing value, which the covariance values are 94.645 with an expected error of cross-validation had been calculated is 0.2590, and produced this land-sea mask of average SST values over the region as it is shown below (Figure 3.).

| a. | b. |
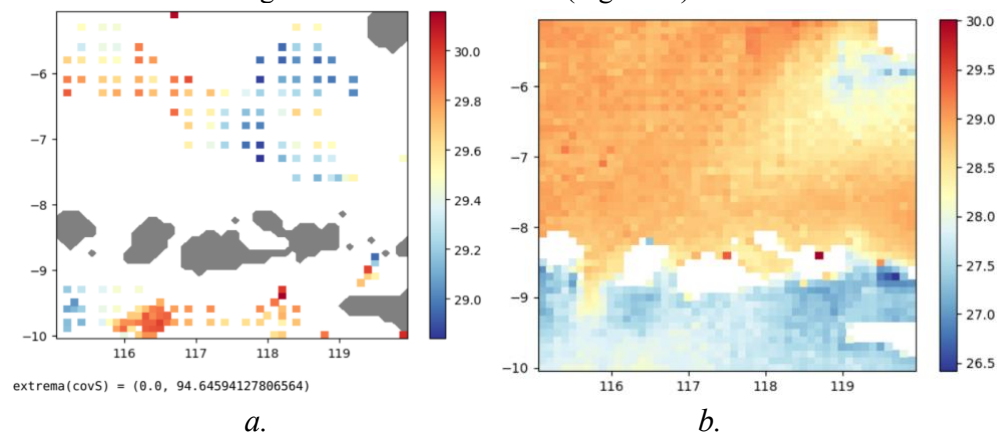
Figure 3. (a) Grid maps of land-sea mask and (b) average values over a study period of Lombok Strait

3.2. Optimal EOF Number Determination

The optimal number of EOF mode in this study reaches 42 repetitions corresponded by the shown data below. The interpolation method was applied, and we get the convergence of about 0.9806×10^{-3}. The estimation of error gained from the processes is shown in **Figure 4**. The second EOF will be started according to the previous single converged EOF file. Hence, the second calculation of error estimation is started. From here, the procedure will be stopped if the error is increasing. Otherwise, continue to the next EOF mode iteration. **Figure 4** shows the error value drops on the second mode of EOF so that the

iteration can be continued. We use a cross-validation technique to find the optimum number of EOF modes that was done from the iteration of error estimation calculation.
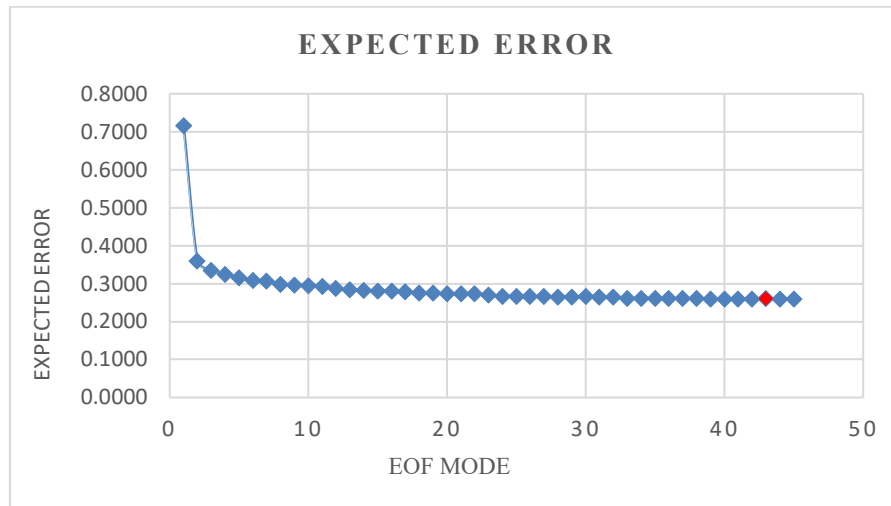

Figure 4. The Error Calculation of the Data

For EOFs algorithm, we adjust the alpha and number iteration in order to filter the data and set the numerical variable for the computation of the required singular values and associated modes to gain the minimum and maximum of EOFs to be calculated. After 42 EOF modes, obtains as the optimal EOFs, the data has finished to be reconstructed by 18 iterations on each mode, and the convergence value is.

Visualization of the DINEOF reconstruction in the Lombok Strait (Figure 5.) with the dimension of the data over 42 x 51 plotted into the array of 50 x 51 x 579 fruitfully reconstructed the Lombok Strait from the missing data before, in here by applying a filter to the temporal covariance matrix and spiking the removal in temporal EOFs.

No	DINEOF	IO
1		
2		
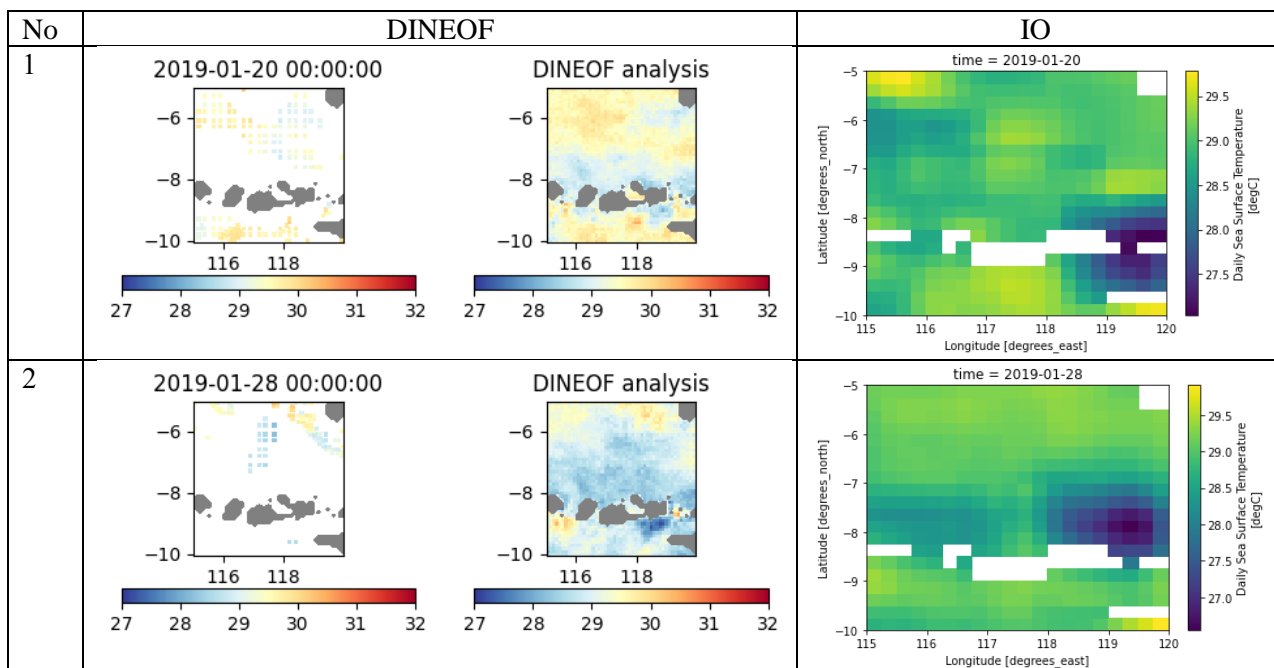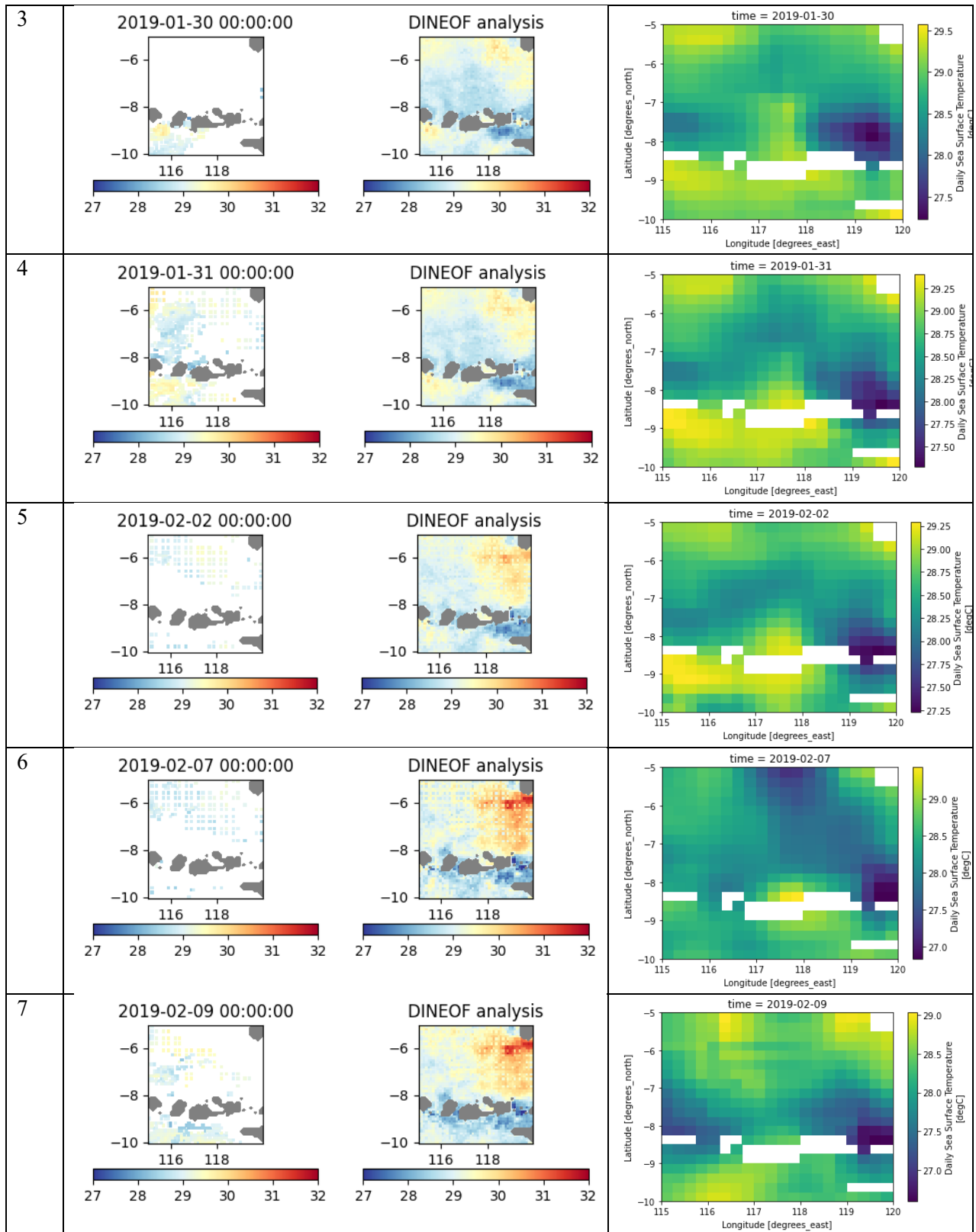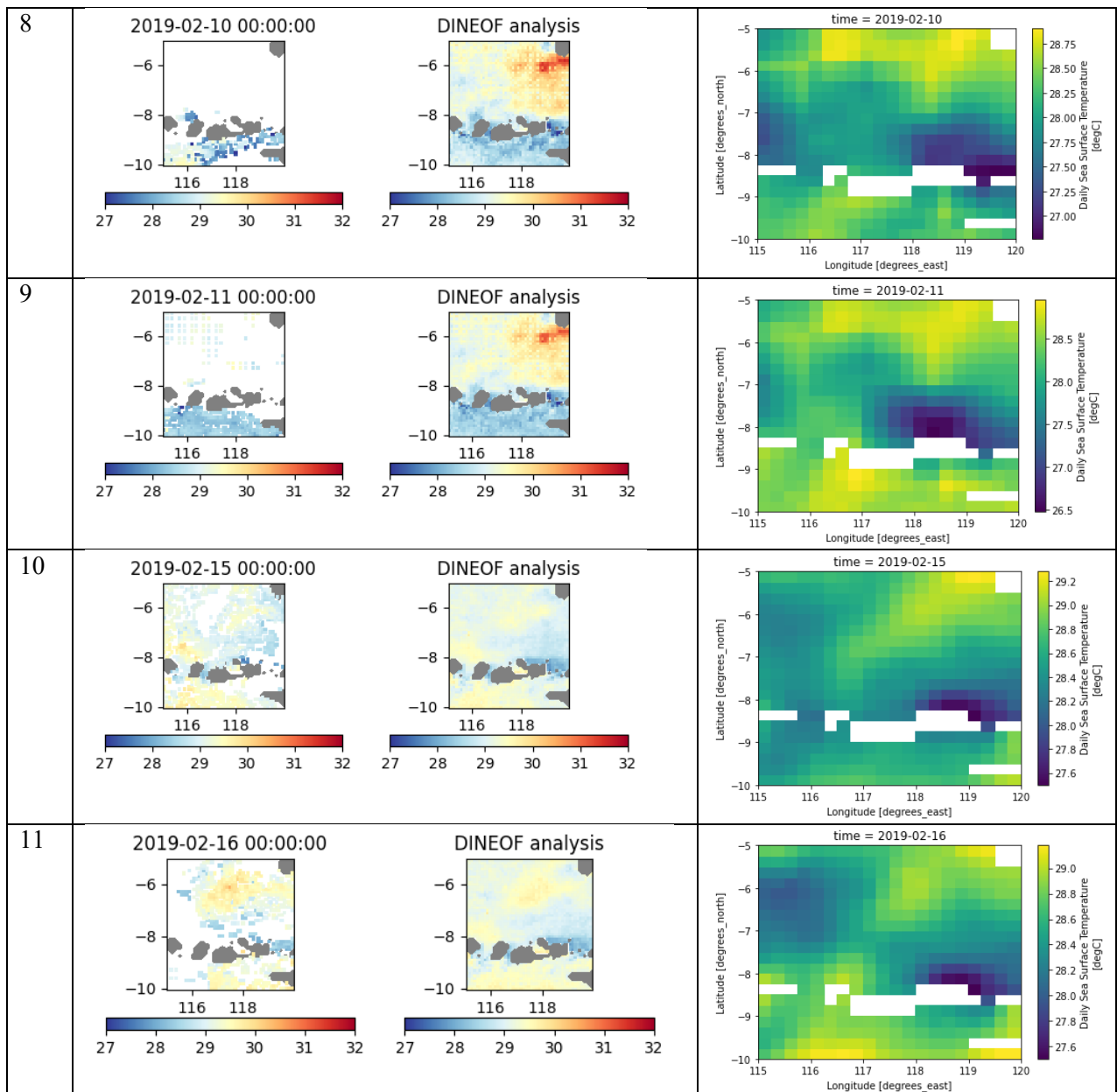
Figure 5. The comparison of DINEOF method and OI visualization

Regarding Figure 5. there is a significant change of data which had been reconstructed over the spatial and temporal, which presented that this DINEOF method had successfully reconstructed the data, despite the missing data were quite huge as it in the 30 January 2019 and 10 February 2019 which almost there is blank data in the northern sea of Lombok island. This blank or the gap of data due to this study area located in the tropical which sometimes wrongly interpreted as valid data, can be from cloud edges, haze areas, contrails, cloud shadows. The condensation is also quite huge related to clouds or monsoon, and in the equatorial, sun-glints or cloud shadow may lead to a gap or missing data while the satellite was trying to retrieve the SST signal over this area. However, DINEOF can obtain a good result of reconstruction data due to the truncation of DINEOF over temporal. It is due to this multivariate version of the number of matrices EOFs, will give a missing point of one variable that can be useful from the presence of another variable at the same time [2,7,9]. Therefore, the correlation between both variables is considered by the number of modes in EOFs to reconstruct missing data.

In order to recheck the values of the output data of DINEOF, this script also making a graph of the temperature trend over time (Figure 6b.) which the extreme data found in time 200 there is a depletion of the temperature of those areas until 22.19° C following after the downtrend, in the time nearly 300, there is a peak temperature up to 31 ° C. However, 32.96 ° C is the extreme value that had been found in the Lombok Strait overall year of 2019.

a.   b.
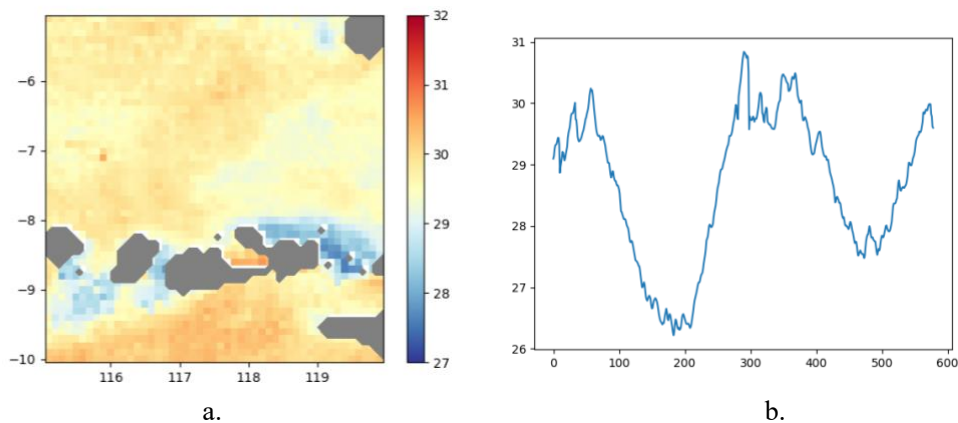
Figure 6. The temperature distribution maps (a) in the specific matrix of the reconstructed data and the fluctuation graph of temperature trend during the one-year data

After calculating the EOFs from the covariance matrix by using non-missing values, the reconstruction will resolve over time and space, which later determine the optimal number of EOFs by cross-validation. In this case, the error number will be estimated.
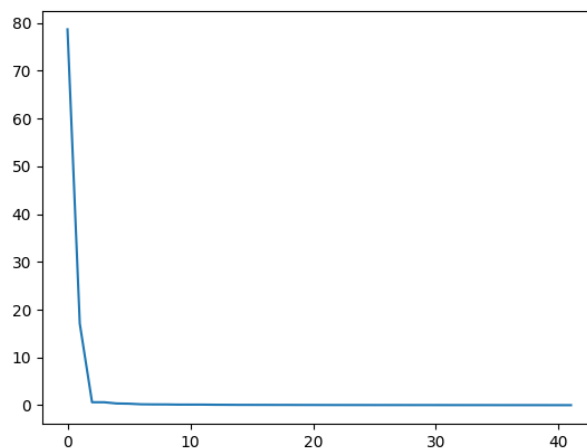
Figure 7. The optimal number of EOFs by cross-validation.

This figure above depicts the optimal number of EOFs by using cross-validation, which will be depleted after EOFs approximately around three. After the diffusion of covariance matrix activated with the alpha is 0.10 and the number of iterations is three, it obtained *779590 out of 1340385 data*. It means about *58.16%* of valid data flagged as missing. Therefore, the cross-validation data will be re-introduced in the matrix as it is shown in Figure 7. which EOFs optimal will minimizes this error due to gappy data. This result had processed iteratively on a whole matrix (missing data value will be improved at each iteration) in order to overcome the missing data problem[2,15].
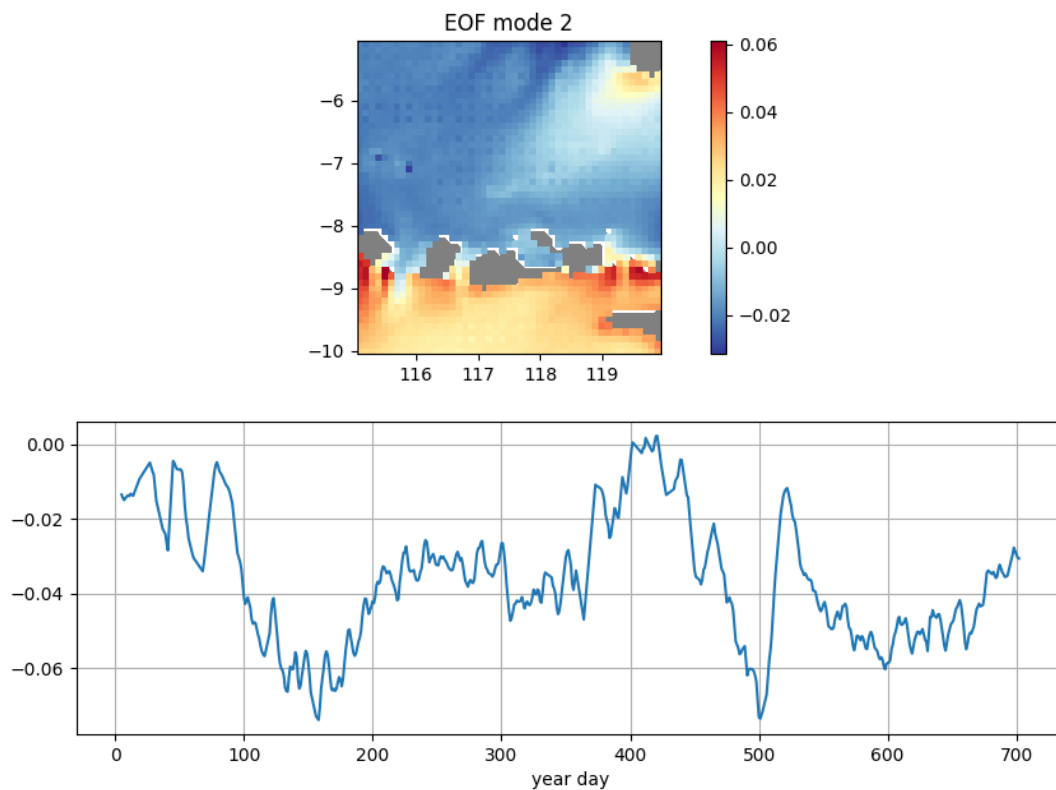
Figure 8. The EOF mode 2

Finally, one of the DINEOF interpretations can be seen in (Figure 8) the EOF mode two which there is a good reconstruction of the data over spatial and space, in which a series of upwelling/ downwelling events were detected in this location. This EOF mode had encountered some process behind minimizing the error in maps by using the EOFs as background covariance in order to get the optimal interpolation. This covariance matrix is accurately calculated from DINEOF SVD decomposition, and it allows an easy combination between variables. It is highly to note that DINEOF is a more objective reconstruction technique because it calculates these statistics internally and based on available data [2,6,16]

4. Conclusion

The presented work shows DINEOF is a reliable technique to reconstruct missing data with a wide variety of data, and platforms can be used this method with either Univariate or Multivariate approaches. Therefore, it is needed to detect the outliers within DINEOF as pixels for which the analysis-observations difference is larger than the statistically expected misfit calculated during the analysis. DINEOF generates the error maps to assess the confidence of the reconstruction, filtering the time temporally in order to avoid spikes and to improve the smoothness of the matrix, and the last was detecting the outlier to remove the suspect data. Moreover, this approach can reduce the computational cost, and the spatial correlation is also taken into account over the different geographical locations.

DINEOF technique supports the filling gap of two-year length data. This study in Lombok Strait showed missing values up to 58.16% from January 2019 until November 2020. The result has also shown the previous set-aside data for cross-validation. The optimum number of EOF modes for the data length is 42 iterations, with the convergence achieved up to 0.9806×10^{-3}. The results highlight DINEOF method can effectively reconstruct SST data in Lombok Strait even-though this massive gap of missing data due to the atmosphere dynamic in the tropical area. Despite, longer timescales from the data shown will

provide a better representation of the technique. To comprehend the result of this study, validation by using *in situ* data is also recommended in the future.

References
[1]   Alvera-Azca´rate, A., A. Barth, M. Rixen, and J. M. Beckers 2005 Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: Application to the Adriatic Sea (*Ocean Modell/Elsevier*) vol 9 pp 325 – 346
[2]   Alvera-Azca´rate, A., A. Barth, J.-M. Beckers, and R. H. Weisberg 2007 Multivariate reconstruction of missing data in sea surface temperature, chlorophyll, and wind satellite fields (*J. Geophys. Res 112*) C03008 doi:10.1029/2006JC003660
[3]   Alvera-Azcárate, A., Barth, A., Sirjacobs, D., Lenartz, F., & Beckers, J. M. 2011 Data interpolating empirical orthogonal functions (DINEOF): A tool for geophysical data analyses (*Mediterranean Marine Science*) vol *12*(3) pp 5–11. https://doi.org/10.12681/mms.64
[4]   Alvera-Azcárate, A., Sirjacobs, D., Barth, A., & Beckers, J. M. 2012 Outlier detection in satellite data using spatial coherence (*Remote Sensing of Environment*) vol *119* pp 84–91 https://doi.org/10.1016/j.rse.2011.12.009.
[5]   Autret, E., Pioll, J., Date, A., Assurance, Q., & Group, R. (n.d.). For High Resolution SST multi sensor over the Global Ocean SST _ GLO _ SST _ L3S _ NRT _ OBSERVATIONS _ 010 _ 010. *CMEMS-SST-QUID-010-010 Date:10/09/2020 Issue 1.3*
[6]   Beckers, J. M., & Rixen, M. 2003 *EOF* Calculations and Data Filling from Incomplete Oceanographic Datasets *(Journal of Atmospheric and Oceanic Technology)* vol 20(12) pp 1839-1856. https://journals.ametsoc.org/view/journals/atot/20/12/1520
[7]   Beckers, J. M., Barth, A., & Alvera-Azcárate, A. 2006 DINEOF reconstruction of clouded images including error maps application to the Sea-Surface Temperature around Corsican Island (*Ocean Science*) vol *2*(2) pp 183–199 https://doi.org/10.5194/os-2-183-2006
[8]   Beckers, J.-M., A. Barth, and A. Alvera-Azca´rate 2006 DINEOF reconstruction of clouded images including error maps: Application to the sea surface temperature around Corsican Island (*Ocean Science*) vol 3(4) pp 735 – 776
[9]   Igor T, Aïda Alvera A, Barth, A., & Beckers, J. 2012 Dineof-based bias correction of seviri sea surface temperature using metop-a/avhrr and envisat/aatsr sst (Belgium: EUMESAT) oai:orbi.ulg.ac.be:2268/165571
[10]   Everson R, Cornillon P, Sirovich L, Webber A. 1997 An empirical eigenfunction analysis of sea surface temperatures in the western North Atlantic. (*J. Phys.Oceanogr.*) vol 27(3) pp 468–479.doi:10.1175/1520-0485(1997)027
[11]   Fieguth P, Menemenlis D, Ho T, Willsky A, Wunsch C. 1998 Mapping Mediterranean altimeter data with a multiresolution optimal interpolation algorithm *J. Atmos. Ocean. Tech.* vol 15(2) pp *535–546 doi:10.1175/1520-0426(1998)*015
[12]   Matthews, J. P., H. Aiki, S. Masuda, T. Awaji, and Y. Ishikawa 2011 Monsoon regulation of Lombok Strait internalwaves. *J. Geophys. Res. 116.* C05007. doi:10.1029/2010JC006403
[13]   Marquis, J. W., Bogdanoff, A. S., Campbell, J. R., Cummings, J. A., Westphal, D. L., Smith, N. J., Zhang, J. 2017 Estimating Infrared Radiometric Satellite Sea Surface Temperature Retrieval Cold Biases in the Tropics due to Unscreened Optically Thin Cirrus Clouds. *Journal of Atmospheric and Oceanic Technology*, 34: 355-373. doi: 10.1175/JTECH-D-15-0226.1.

[14]   Nikolaidis, A., Georgiou, G.C., Hadjimitsis, D., & Akylas, E. 2014 Filling in Missing Sea-Surface Temperature Satellite Data Over The Eastern Mediterranean Sea Using the DINEOF Algorithm.   *Central European Journal of Geosciences* vol 6(1) pp 27-41. doi: 10.2478/s13533-012-0148-1tral.

[15]   Ping, B., Su, F., & Meng Y. 2016 DINEOF Algorithm for Filling Missing Values in Spatio-Temporal Sea Surface Temperature Data. *PLoS ONE* vol 11(5) pp e0155928 doi: 10.1371/journal.pone.0155928.

[16]   Wang, Y., Gao, Z., & Liu, D. 2019 Multivariate DINEOF reconstruction for creating long-term cloud-free chlorophyll-a data records from seaWiFS and MODIS: A case study in Bohai and Yellow Seas, China. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* vol *12*(5) pp 1383–1395. https://doi.org/10.1109/JSTARS.2019.2908182