MDPI

Article

TIEOF: Algorithm for recovery of missing multidimensional satellite data on water bodies based on higher-order tensor decompositions

Leonid Kulikov [1,2], Natalia Inkova [2], Daria Cherniuk [1,2], Anton Teslyuk [2,3] and Zorigto Namsaraev [3,*]

[1] Skolkovo Institute of Science and Technology, Moscow, Russia
[2] Moscow Institute of Physics and Technology, Moscow, Russia
[3] NRC "Kurchatov Institute", Moscow, Russia
* Correspondence: zorigto@gmail.com

Abstract: Satellite research methods are actively involved in observations of water bodies. One of the most important problems in satellite observations is the presence of missing data due to internal malfunction of satellite sensors and poor atmospheric conditions. We proceeded on the assumption that the use of data recovery methods based on spatial relationships in data can increase the recovery accuracy. In this paper, we present a method for missing data reconstruction from remote sensors. We refer our method to as Tensor Interpolating Empirical Orthogonal Functions (TIEOF). The method relies on the two-dimensional nature of sensor images and organizes the data into three-dimensional tensors. We use high-order tensor decomposition to interpolate missing data. Using MODIS and SeaWiFS satellite data of lake Baikal we show that the observed improvement of TIEOF was 69 % on average compared to the current state-of-the-art DINEOF algorithm measured in various preprocessing data scenarios including thresholding and different interpolating schemes.

Keywords: Satellite observations of water bodies; missing data reconstruction; higher-order tensor decomposition; chlorophyll; lake Baikal

Citation: Kulikov L., Inkova N., Cherniuk D., Teslyuk A. and Namsaraev Z. TIEOF: Algorithm for recovery of missing multidimensional satellite data on water bodies based on higher-order tensor decompositions. *Water* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

1. Introduction

Satellite research methods are actively involved in observations of water bodies. They are used for a wide variety of tasks, including study of lake morphodynamic characteristics (water level, surface, volume) [1], surface temperature [2], chlorophyll content, dynamics of phytoplankton and primary production [3], protection of endangered species [4] etc. However, satellite methods for observing water bodies have their own drawbacks. One of the most important problems in satellite observations is the presence of missing data due to internal malfunction of satellite sensors and poor atmospheric conditions.

Missing value reconstruction is problem known for decades which emerges in a vast range of research areas. When analyzing measurements which contain partially missing information a researcher has to adjust data analysis methods to take into account the missing data. A common approach for dealing with missing data is data imputation, where missing values are filled with new data reconstructed from observations. To make the reconstruction one needs to define a model which will capture particular patterns in observed data and use the model to fill the gaps in measurements. Various approaches exist to model the missing data including rich set of statistical and probabilistic methods [5], factor analysis [6], deep learning networks [7], autoencoders [8] and many more.

One of the most widely used method to model remote sensing data with missing information is Data Interpolating Empirical Orthogonal Functions (DINEOF) algorithm [9]. It relies on empirical orthogonal functions (EOF) method which approximates the data

using low dimensional linear decomposition. The DINEOF algorithm is closely related to principal component analysis method (PCA) [10] which is one the most common methods for low dimensional data representation. A number of modifications for PCA were proposed to account missing data, a detailed review of them was presented in [11]. PCA with missing data allows various formulations including: least square problem [12], alternating algorithm [13], probabilistic model (PPCA) [14] and variational Bayesian method (VBPCA) [15].

When applied to spacio-temporal data EOF- and PCA-based methods represent the data using two dimensional matrix where single dimensions correspond to spatial and temporal measurements. When dealing with remote sensing data like satellite images three dimensional tensors are more natural way of data representation. One dimension is used for the timeline and the other two for two-dimensional images. In this case the information about 2D spatial data structure doesn't get lost and one can take advantage of it in further analysis. Although there is no single straightforward generalization of singular matrix decomposition underlying the EOF and PCA methods to the case of multidimensional tensors, a number of methods have been developed for the latter to find low-rank representations of multidimensional data [16] which include canonical polyadic decomposition (CPD) [17], Tucker decomposition [17,18] and its generalization as a higher-order singular value decomposition (HOSVD) [18,19].

In this paper we show that the use of the tensors for the data representation and high order tensor decomposition can significantly improve the quality of missing information recovery compared to the DINEOF algorithm. We refer our algorithm to as Tensor Interpolating Empirical Orthogonal Functions (TIEOF).

We tested the developed algorithm using satellite data on the concentration of chlorophyll a in Lake Baikal obtained within the period from 1998 to 2020. Lake Baikal is a UNESCO World Heritage Site and holds 20 % of the world's freshwater reserves. Since 2011 the blooms of filamentous green algae were regularly observed in the nearshore zone of the lake. Besides, since 2016 severe cyanobacterial blooms were observed in the coastal and pelagic zones of the lake [20]. Given the large size of the lake, satellite data provide valuable information for studying the long-term trends in lake eutrophication, but the amount of missing satellite data makes this challenging. We hope that our algorithm will make it possible to assess the long-term trends in the development of lake ecosystems more accurately, which in turn is necessary to take effective measures to preserve them.

2. Materials and Methods

2.1. Recap of the multi-way tensor decompositions

Tensor decompositions is a well-studied field. There are several multi-way tensor decompositions. In this work we employed three popular decomposition methods: Higher Order Singular Value Decomposition (HOSVD), Higher Order Orthogonal Iteration (HOOI) and , PARAllel FACtors (PARAFAC).

HOSVD is the simplest generalization of renowned Singular Value Decomposition (SVD) on tensors of order higher than 2. The main idea here is to find Empirical Orthogonal Functions (EOFs) and a kernel tensor of the same order as the original data which contains singular values of different tensor unfoldings. When combining all that information with mode products we get a reconstruction according to the Figure 1.

PARAFAC is a decomposition of a multi-way tensor on sum of tensors of rank 1 (i.e. sum of outer products of vectors) like on the Figure 2. To calculate those vectors Alternating Least Squares (ALS) algorithm [21] can be used. It is simple yet an effective algorithm which iteratively fine-tune part of the vectors while keeping others fixed [22]. Such a sum of outer products can actually be formulated as a mode product with diagonal kernel in fashion similar to HOSVD.

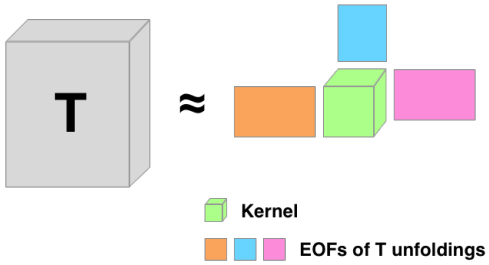

Figure 1. HOSVD

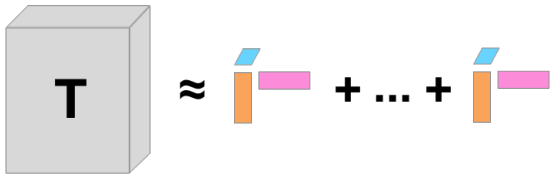

Figure 2. PARAFAC

HOOI algorithm simply combines ideas from both decompositions: HOSVD and PARAFAC. Firstly, we use HOSVD to initialize a kernel and EOFs of unfoldings and then apply ALS algorithm to further improve approximation.

2.2. Description of the TIEOF pipeline

In this section we describe the TIEOF algorithm for missing data reconstruction in spacio-temporal datasets. The pipeline consists of two stages: data preprocessing and missing data recovery. In the first stage we preprocess original data from the satellite and interpolate to the fixed grid to combine two-dimensional shots for different days together into three-dimensional tensor. In the second stage we apply the iterative method TIEOF to reconstruct the tensor, i.e. fill unknown values. Pseudo-code of the TIEOF algorithm is depicted in the scheme 1. Visual representation of TIEOF is presented on Figure 3

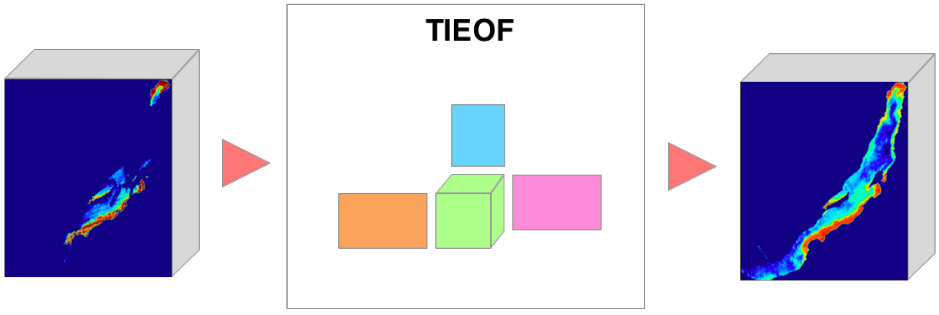

Figure 3. Simplified visual representation of TIEOF

2.3. Data preprocessing

TIEOF takes an array of images as its input. Every image contains a set of points with measurements made along the path of the satellite at a certain date and time. Coordinates of these points are not constant in time and change with every image. As a first step of data preprocessing interpolation of the measurements taken on different days to the fixed spacial coordinates grid was made. We have used two methods of interpolation: regression with K nearest neighbours with distance weighting and regression with nearest neighbours within fixed radius with distance weighting as well [23].

Data thresholding. Next preprocessing step is satellite signal thresholding, clipping signal with high chlorophyll-a concentration. In situ studies [24,25] showed that satellite estimations for high chlorophyll-a values in Lake Baikal can be significantly overestimated. In particular, a large discrepancy is observed starting from the concentration of $2mg/m^3$. To reduce the effect of satellite chlorophyll-a overestimation we have clipped satellite measurements exceeding $2mg/m^3$ to the threshold value. The value of the threshold was motivated by the field Baikal measurements by Heim [24]. In this way we introduce some mistrust to the raw satellite data, but at the same time reduce the noise coming from satellite chlorophyll-a estimation algorithm OC3M [26]. In the next section we show that such a threshold increases the accuracy of every algorithm on validation subset.

The performance of all the models used in reconstruction was measured by calculating the error on a separate validation subset with known chlorophyll-a measurements. As chlorophyll-a concentration tends to be zero in a majority of points we had to elevate the validation points collection from the standard uniform random method. Instead we used stratified sampling approach: data points were grouped into clusters according to their chlorophyll-a values with K-Means clustering method. The number of clusters were chosen to be **10**. Then equal number of points were sampled for validation dataset within each cluster providing stable representation of measurements with different chlorophyll-a levels in the validation dataset.

2.4. Missing data reconstruction

The core of TIEOF is iterative data reconstruction algorithm. Our method shares with DINEOF algorithm the general approach of low dimensional data representation using linear decomposition and iterative estimation of decomposition parameters. The principal difference between methods is that DINEOF is based upon SVD matrix decomposition to find empirical orthogonal functions while we use three dimensional tensors for data representation and tensor decomposition methods to find low dimensional data features. In our work we have used three different multi-way tensor decompositions: Truncated Higher Order Singular Value Decomposition (HOSVD), Higher Order Orthogonal Iteration (HOOI) and Parallel Factors (PARAFAC) [22]. The TIEOF pseudo-code is presented in the scheme 1.

It takes an array of images as its input as a three dimensional tensor *Data* with dimensions (lat, lon, t). At the initial step we calculate the averaged image along time axis as well as average signal for every image. Then substract it from initial data to get it centered along spatial and time axes. This is a standard procedure for PCA and EOF algorithms.

At first step we fill missing data with zero values. Then iterative algorithm starts until convergence: Tensor decomposition of current representation of *Data* tensor using one of three algorithms (HOSVD, HOOI, PARAFAC) is calculated. Tensor decomposition is used to estimate reconstructed tensor $Data_r$. Reconstructed tensor is used to fill unknown values of the data. Reconstruction error is calculated as a normalized root mean squared error (NRMSE) difference of reconstructed values for missing data between successive iterations defined by 1 ($y^{'}$ - reconstructed values and y - ground-truth values) of reconstructed tensor between iterations in the known points according to the mask.

Algorithm 1 TIEOF

Data - input set of images (lat, lon, t)
Mask - boolean mask of known pixels
r - number of components in reconstruction
procedure TIEOF(*Data*, *Mask*, r)
    $Data \leftarrow Center(Data)$ ▷ Substract averages on (lat, lon) and (t) dimensions
    $Data[!mask] = 0$ ▷ Initial guess for unknown values
    while $Error > 1e-3$ **do**
        K_r, F_r = [HOSVD | HOOI | PARAFAC]($Data, r$)
        $Data_r = Reconstruct(K_r, F_r)$
        $Data_r[mask] = Data[mask]$
        $Error \leftarrow NRMSE(Data[!mask], Data_r[!mask])$
        $Data \leftarrow Data_r$
    return *Data*

$$NRMSE = \frac{\sqrt{E[(y' - y)^2]}}{\sigma[y]} \tag{1}$$

Early stopping. Iterations stop when reconstruction error goes below predefined threshold ($1e-3$ in all our experiments). Large number of iterations with low amount of training data can lead to severe over-fitting. To avoid over-fitting we used the early stopping technique that is typically used in deep learning routines [27]. As DINEOF algorithm is iterative, in the new setup we track not only the normalized-root-mean-squared-error (NRMSE) between successive reconstructions in the known points but also the error gradients (absolute difference between errors in successive iterations). When gradients become too low we stop iterations and return the reconstructed tensor which potentially has larger reconstruction error between iterations, but, as we will show in the next section, has lower error on validation points.

2.5. Dataset

To test TIEOF algorithm we used three publicly available datasets distributed by NASA's Ocean Biology Processing Group. These three datasets contain products obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS) located aboard the Terra (EOS AM) and Aqua (EOS PM) satellites and from the Sea-Viewing Wide Field-of-View Sensor (SeaWiFS) located aboard the OrbView-2 (SeaStar) satellite. Data are available for the period from 24 February 2000 to the present for MODIS Terra, from 04 July 2002 to the present for MODIS Aqua and from 04 September 1997 to 11 December 2010 for SeaWiFS. We selected data on chlorophyll-a concentration (Level 2) for 93 summer days (June 1 to September 1) 2003-2020 for MODIS Aqua, 2001-2020 for MODIS Terra and 1998-2010 for SeaWiFS.

During preprocessing stage we calculated interpolation of the satellite images into static spacial grid with fixed coordinates. Coordinates of the static grid were chosen to cover the whole region of interest (Baikal) with a kilometer density. The size of the grid was chosen to be **latitude: 482 points, longitude: 406 points**. Taking into account a sample of 93 summer days for each year we sum up to the dataset tensors of shape **(482, 406, 93)**. For nearest neighbors interpolation we have used the number of neighbors $K = 3$ and radius range equal to 5 km.

The amount of missing information differs between satellites. MODIS Aqua and MODIS Terra usually have relatively large amount of the points with measured value of chlorophyll-a concentration (25%) while SeaWiFS is almost empty and contain about 5% of informative points in Baikal region. The data missing ratio was varying between satellites and years as depicted on Figure 4.

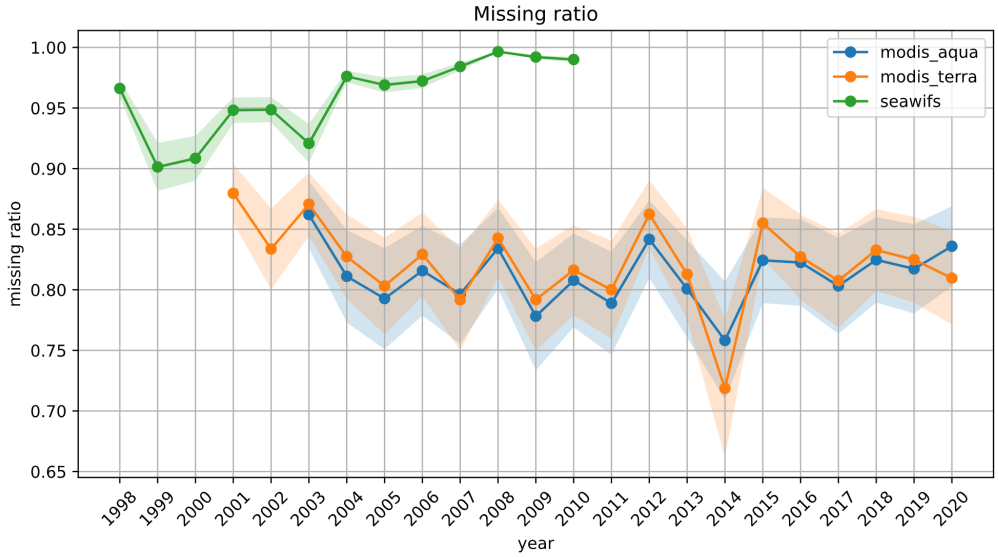

Figure 4. Amount of missing data for different years and satellites

To collect reconstruction and missing ratio confidence intervals and means we make 30 independent reconstructions for every dataset using different random choices of known data points with replacement which is known as bootstrapping technique.

To estimate the performance of our models every dataset was split into two parts with ratio 0.95/0.05: training dataset that was used for model learning and validation dataset to evaluate model performance.

To extract only the Baikal points in the downloaded pieces of data shapefile with Baikal coastline [28][1] was used.

3. Results

3.1. Effect of early stopping

To check the effect of early stopping technique we reconstructed the tensors for this experiment from Aqua subset and for all of the algorithms, TIEOF included and measured NRMSE. Before reconstruction the raw data provided by the satellite was mapped onto a static grid using neighbours within five km radius interpolation. The results are depicted on Figure 5. In case of DINEOF we got a consistent improvement using early stopping mode while for TIEOF the difference is negligible. On average early stopping mode is better (leading to lower NRMSE) than full convergence for DINEOF on 8.4 %, for TIEOF early stopping mode is slightly worse and NRMSE error grows on a factor of 0.3 - 1.8 %. TIEOF with HOOI decomposition has the lowest factor. Overall both error curves catch the error trend. Thus we conclude that TIEOF has the better resistance to over-fitting phenomenon than DINEOF, because full convergence mode is very close to early stopping mode (max difference is only 1.8%). Consideration of more complex feature space definitely helps even when amount of examples is very limited.

The stopping criteria value for both regular error and gradients between iterations was chosen to be 0.001.

All further experiments were conducted within early stopping mode to slightly alleviate the computation burden.

3.2. Data thresholding reduces reconstruction errors

As described above satellite signal of chlorophyll-a tends to be greatly overestimated when the concentration goes beyond $1.5 - 2mg/l^3$. To check the effect of data clipping

[1] https://www.marineregions.org/gazetteer.php?p=details&id=5659

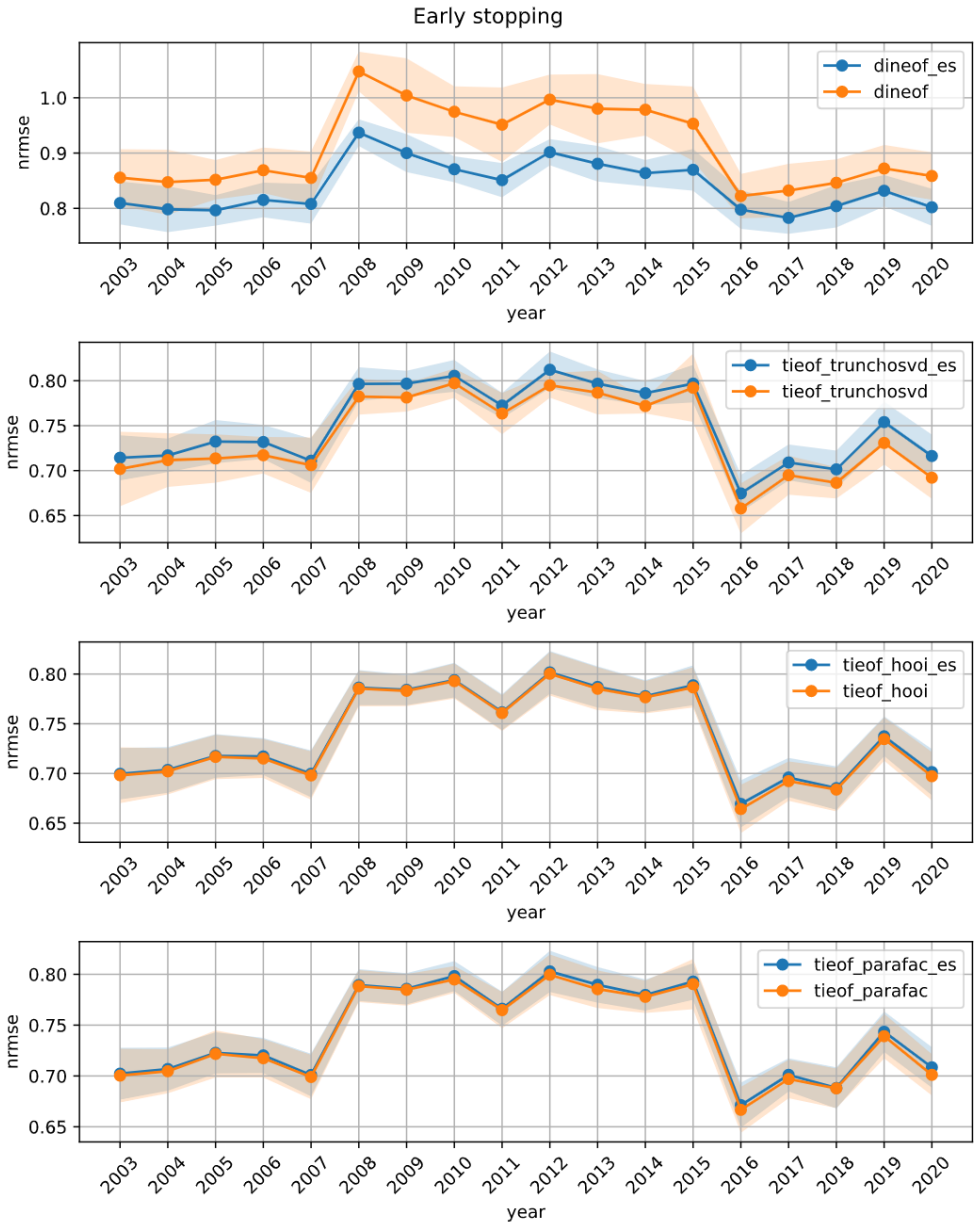

Figure 5. Error levels with and without using early stopping (es)

Table 1: Improvement when using thresholding

METHOD	IMPROVEMENT FACTOR
DINEOF	1.24
TIEOF: HOSVD	1.56
TIEOF: PARAFAC	1.56
TIEOF: HOOI	1.57

we reconstructed both thresholded to 2.0mg/m^3 and non-thresholded data using all models.

For missing data reconstruction we have used Aqua subset mapped to the static grid using neighbours interpolation within 5 km radius and early-stopping for model convergence strategy. Reconstruction of initially thresholded tensor led to significantly lower NRMSE on the validation subset which is demonstrated on Figure 6. The average factor of an improvement (Equation 2) between all years available for all methods is presented in the Table 1.

$$\frac{NRSME}{NRMSE_{thresh2}} \tag{2}$$

3.3. Comparison of the methods

On all the datasets available all TIEOF variations showed superiority above DINEOF as depicted on Figure 7.

The difference between different TIEOF variations is not as significant as comparing to DINEOF. This shows us that TIEOF is robust to the choice of tensor decomposition and its main benefit not in the specific algorithm but in the usage of original multidimensional complex feature spaces. We depicted the mean errors between all years for all datasets and methods in Table 2.

HOOI variation of TIEOF in all cases shows the best performance while PARAFAC and HOSVD are slightly behind. This behavior is reasonable because HOOI algorithm combines alternating least squares (ALS [21]) optimization and Tucker decomposition behind the scenes. Tucker decomposition is used as an initialization of kernel and factors and then ALS is applied for fine-tuning. HOSVD utilizes only Tucker decomposition method while PARAFAC employs only ALS. The combination of both components leads to better results. The gap between methods tends to be low because of high similarity of components used in the algorithms.

We observe that the improvement of TIEOF over DINEOF varies between 9 - 309% (69% on average). High improvement is noticed when the data is poor, i.e. the error (NRMSE) is much larger (> 25%) than a median error calculated on all years and low improvement is observed when the error is close to the median. Therefore, TIEOF provides a stable (69% on average) improvement comparing to DINEOF, and the difference between methods grow on poor data when DINEOF level of NRMSE increase.

4. Discussion

Our results demonstrate that using tensor decompositions give consistently better results than DINEOF matrix-based algorithm. In this section we try to answer the question of why this might be so. Images contain lots of features. Even if the image is of moderate resolution, e.g. 400x400 pixels, there is a lot of information scattered thinly. When transformed to a vector it contains more than 160000 features. Matrix decomposition-based methods are able to seek linear dependencies between all pairs of features giving no preference to particular pairs, there is no way for the method to prefer to look specific patterns in close points. Due to the high number of features the number of pairs of features is dramatic which must inevitably lead to over-fitting. In the same

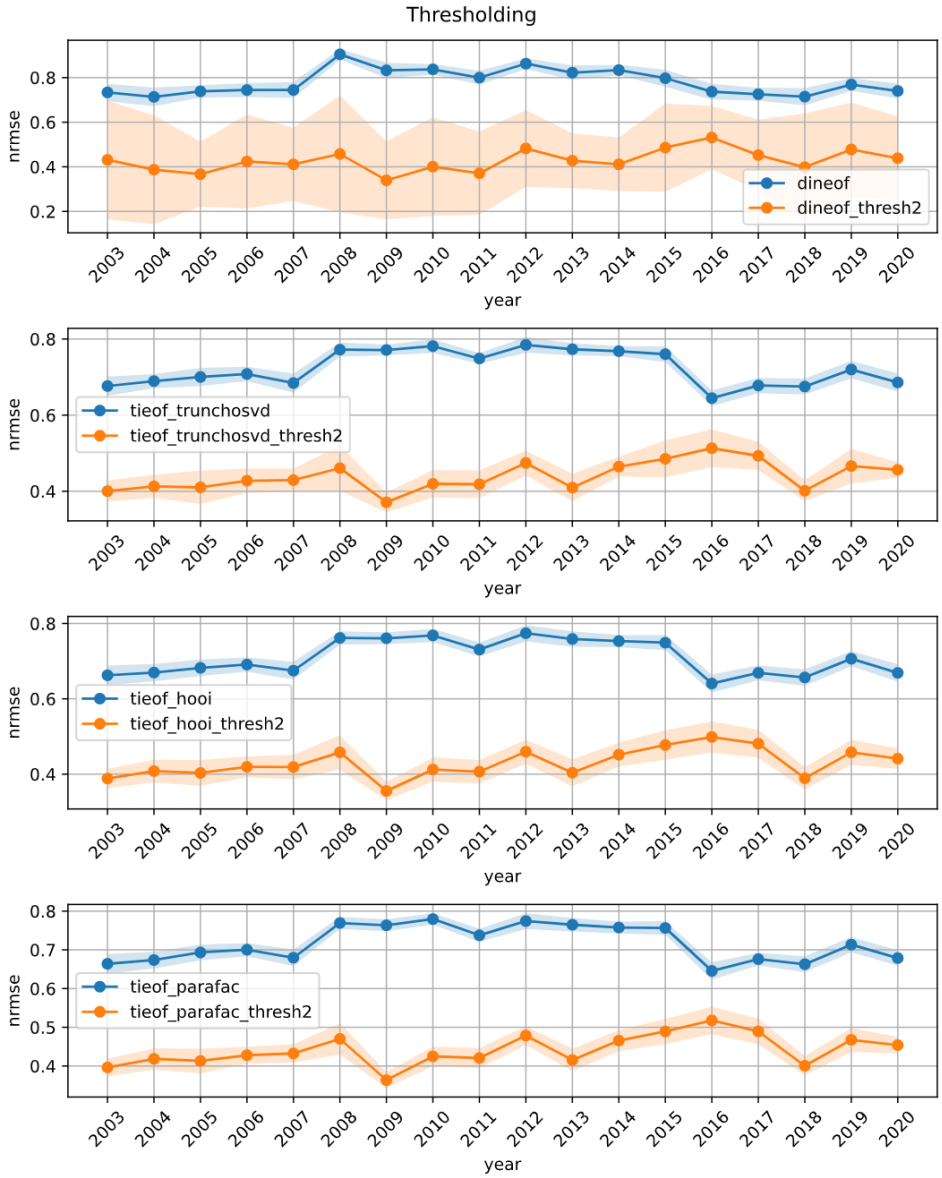

Figure 6. Effect of clipping data signal above $2mg/m^3$

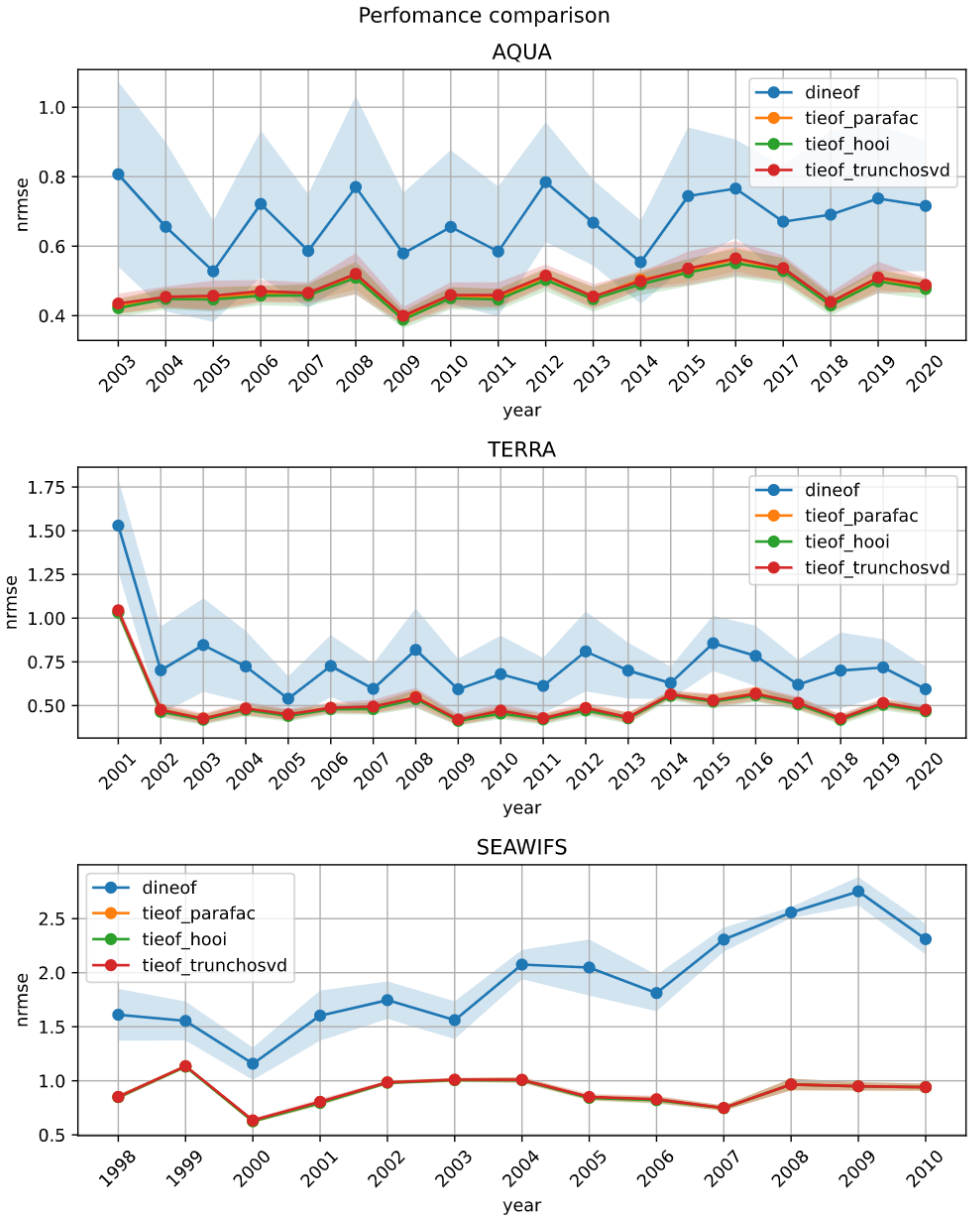

Figure 7. Performance comparison

Table 2: Mean errors (NRMSE)

DATA	METHOD	MEAN NRMSE
AQUA	DINEOF	0.6788
	TIEOF: HOSVD	0.4814
	TIEOF: PARAFAC	0.4774
	TIEOF: HOOI	0.4710
TERRA	DINEOF	0.7388
	TIEOF: HOSVD	0.5120
	TIEOF: PARAFAC	0.5091
	TIEOF: HOOI	0.5022
SEAWIFS	DINEOF	1.9304
	TIEOF: HOSVD	0.9014
	TIEOF: PARAFAC	0.8983
	TIEOF: HOOI	0.8968

time it is reasonable that points that are far from each other are barely related to each other and give a tiny contribution to the patterns algorithm try to learn. So the majority of point pairs are distinct from each other and provide more noise than informative signal for the model. Common sense shows that narrowing the feature space should make the problem easier and thus lead to more accurate results.

The intuition also stems from the renowned convolutional neural networks. The convolution operation (or more precisely cross-correlation) uses typically small patches to calculate features on the next layer. Thus the new features are constructed from only some local part of the information from the previous layer of a neural network. For high-level tasks (e.g. image classification) the features that are based on local regions surely aren't enough. A stacking of convolutional layers leads to gradual increase of a receptive field which is important for abstractness of features on deep levels. Based on these abstract features it is intuitively easier for a simple function (e.g. linear function) to classify the image. Yet for low-level tasks (e.g. super-resolution, image completion) even small shallow convolutional neural networks bring surprisingly good results [29] where features are calculated from some small local parts of the input. Therefore we make a conclusion that the accurate gradual treatment of close points and rejection of faraway points are of crucial importance, especially for low-level tasks and it's a good move to somehow empower the algorithm with this prior information to potentially enhance its speed and quality.

Another approach is to consider patches in the original data and apply DINEOF-based algorithm to them and then combine the results. Nevertheless, it is always a question how large the patch should be and how to choose it. Multi-way tensor decomposition gives us a method to select reasonable subset of features. Multi-way tensor decomposition allows us to not discard original complex feature space. Instead we search there linear dependencies within low amount of examples in high-dimensional feature space there. We hypothesize that lower amount of points (barely related too) makes the problem solving easier and extraction of underlying dependency is more precise and accurate. The algorithm also potentially benefits from high-dimensional feature space as it is allows to arrange data points in a more understandable way and find latent relations.

It is worth to mention that just considering multi-way tensor decompositions within the framework of DINEOF instead of SVD doesn't necessarily remove all the barely related points. Instead we consider the points only within vertical and horizontal dimensions in a two-dimensional slice of data (e.g. satellite shot) separately. The points on the poles probably also don't have much of a relation, thus we consider careful treatment of that fact as a potential improvement direction.

5. Conclusions

In this work we proposed the new set of methods to reconstruct gapped multi-way tensors. Comparing to the renowned DINEOF our methods have lower error (69 % improvement) which is statistically proven and lower confidence intervals. They also don't suffer from the over-fitting phenomenon as significantly as DINEOF.

We have also investigated additional preprocessing techniques: data thresholding, early stopping and different interpolation methods to bind the tensors to the static grid to better deal with satellite-related problems.

Besides, we consider further investigation of treating local regions more carefully as a potential improvement direction.

Author Contributions: Conceptualization, L.K., A.T. and Z.N.; methodology, L.K., D.C. and A.T.; software, L.K., D.C. and N.I.; writing—original draft preparation, L.K., N.I., A.T. and Z.N.; writing—review and editing, L.K., N.I., A.T. and Z.N.; visualization, L.K.; supervision, A.T. and Z.N.; project administration, Z.N.; funding acquisition, Z.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Ministry of Science and Higher Education of the Russian Federation, grant #075-15-2019-1659.

Data Availability Statement: The source code has been made available on github [2]

Acknowledgments: The work has been carried out using computing resources provided by NRC Kurchatov institute project "Development of modular platform for scientific data processing and mining" (Project No. 1571). Thanks to the developers of Algorithm: The Logic Game [3] for consulting us on the topic of algorithms.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DINEOF	Data Interpolating Empirical Orthogonal Functions algorithm
PCA	Principal component analysis
TIEOF	Tensor Interpolating Empirical Orthogonal Functions
HOSVD	Truncated Higher Order Singular Value Decomposition
HOOI	Higher Order Orthogonal Iteration
PARAFAC	Parallel Factors
SVD	Singular Value Decomposition (SVD)
EOFs	Empirical Orthogonal Functions
NRMSE	normalized root mean squared error
SeaWiFS	Sea-Viewing Wide Field-of-View Sensor
MODIS	Moderate Resolution Imaging Spectroradiometer

References

1. Crétaux, J.F.; Abarca-del Río, R.; Berge-Nguyen, M.; Arsen, A.; Drolon, V.; Clos, G.; Maisongrande, P. Lake volume monitoring from space. *Surveys in Geophysics* **2016**, *37*, 269–305.
2. Ganzedo, U.; Alvera-Azcarate, A.; Esnaola, G.; Ezcurra, A.; Saenz, J. Reconstruction of sea surface temperature by means of DINEOF: a case study during the fishing season in the Bay of Biscay. *International Journal of Remote Sensing* **2011**, *32*, 933–950.
3. Bergamino, N.; Horion, S.; Stenuite, S.; Cornet, Y.; Loiselle, S.; Plisnier, P.D.; Descy, J.P. Spatio-temporal dynamics of phytoplankton and primary production in Lake Tanganyika using a MODIS based bio-optical time series. *Remote sensing of environment* **2010**, *114*, 772–780.

[2] https://github.com/theleokul/tieof
[3] https://apps.apple.com/us/app/algorithm-the-logic-game/id1475410194

4. Breece, M.W.; Oliver, M.J.; Fox, D.A.; Hale, E.A.; Haulsee, D.E.; Shatley, M.; Bograd, S.J.; Hazen, E.L.; Welch, H. A satellite-based mobile warning system to reduce interactions with an endangered species. *Ecological Applications* **2021**, p. e02358.
5. Little, R.J.; Rubin, D.B. *Statistical analysis with missing data*; Vol. 793, John Wiley & Sons, 2019.
6. Kamakura, W.A.; Wedel, M. Factor analysis and missing data. *Journal of Marketing Research* **2000**, *37*, 490–498.
7. Zhang, Q.; Yuan, Q.; Zeng, C.; Li, X.; Wei, Y. Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing* **2018**, *56*, 4274–4288.
8. Jaques, N.; Taylor, S.; Sano, A.; Picard, R. Multimodal autoencoder: A deep learning approach to filling in missing sensor data and enabling better mood prediction. 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII). IEEE, 2017, pp. 202–208.
9. Beckers, J.M.; Rixen, M. EOF calculations and data filling from incomplete oceanographic datasets. *Journal of Atmospheric and oceanic technology* **2003**, *20*, 1839–1856.
10. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science* **1901**, *2*, 559–572.
11. Ilin, A.; Raiko, T. Practical approaches to principal component analysis in the presence of missing values. *The Journal of Machine Learning Research* **2010**, *11*, 1957–2000.
12. Wiberg, T. Umea, Computation of Principal Components when Data are Missing. Proc. Second Symp. Computational Statistics, 1976, pp. 229–236.
13. Grung, B.; Manne, R. Missing values in principal component analysis. *Chemometrics and Intelligent Laboratory Systems* **1998**, *42*, 125–139.
14. Tipping, M.E.; Bishop, C.M. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **1999**, *61*, 611–622.
15. Bishop, C.M. Pattern recognition. *Machine learning* **2006**, *128*.
16. Sidiropoulos, N.D.; De Lathauwer, L.; Fu, X.; Huang, K.; Papalexakis, E.E.; Faloutsos, C. Tensor decomposition for signal processing and machine learning. *IEEE Transactions on Signal Processing* **2017**, *65*, 3551–3582.
17. Hitchcock, F.L. The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics* **1927**, *6*, 164–189.
18. Tucker, L.R. Some mathematical notes on three-mode factor analysis. *Psychometrika* **1966**, *31*, 279–311.
19. De Lathauwer, L.; De Moor, B.; Vandewalle, J. A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications* **2000**, *21*, 1253–1278.
20. Namsaraev, Z.; Melnikova, A.; Ivanov, V.; Komova, A.; Teslyuk, A. Cyanobacterial bloom in the world largest freshwater lake Baikal. IOP Conference Series: Earth and Environmental Science. IOP Publishing, 2018, Vol. 121, p. 032039.
21. Kroonenberg, P.M.; De Leeuw, J. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika* **1980**, *45*, 69–97.
22. Rabanser, S.; Shchur, O.; Günnemann, S. Introduction to Tensor Decompositions and their Applications in Machine Learning, 2017, [arXiv:stat.ML/1711.10781].
23. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Müller, A.; Nothman, J.; Louppe, G.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Édouard Duchesnay. Scikit-learn: Machine Learning in Python, 2018, [arXiv:cs.LG/1201.0490].
24. Heim, B. Qualitative and quantitative analyses of Lake Baikal's surface-waters using ocean colour satellite data (SeaWiFS). PhD thesis, Universität Potsdam, 2005.
25. Abbas, M.M.; Melesse, A.M.; Scinto, L.J.; Rehage, J.S. Satellite Estimation of Chlorophyll-a Using Moderate Resolution Imaging Spectroradiometer (MODIS) Sensor in Shallow Coastal Water Bodies: Validation and Improvement. *Water* **2019**, *11*, 1621.
26. Hooker, S.B.; Firestone, E.R.; OReilly, J.E.; Maritorena, S.; OBrien, M.C.; Siegel, D.A.; Toole, D.; Mueller, J.L.; Mitchell, B.G.; Kahru, M.; others. SeaWiFS Postlaunch Technical Report Series. Volume 11; SeaWIFS Postlaunch Calibration and Validation Analyses **2000**.
27. Prechelt, L. Early stopping-but when? In *Neural Networks: Tricks of the trade*; Springer, 1998; pp. 55–69.
28. Claus, S.; De Hauwere, N.; Vanhoorne, B.; Deckers, P.; Souza Dias, F.; Hernandez, F.; Mees, J. Marine regions: towards a global standard for georeferenced marine names and boundaries. *Marine Geodesy* **2014**, *37*, 99–125.

388 29. Hui, Z.; Gao, X.; Yang, Y.; Wang, X. Lightweight Image Super-Resolution with Information
389 Multi-distillation Network. *Proceedings of the 27th ACM International Conference on Multimedia*
390 **2019**. doi:10.1145/3343031.3351084.