Ecological clustering of the Red Sea and parallel 1D-ecological simulations

Denis Dreano[a], George Triantafyllou[c], Bani Mallick[b], Ibrahim Hoteit[a,*]

[a] *Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology*
[b] *Department of Statistics, Texas A&M University*
[c] *Hellenic Center for Marine Research*

Abstract

Abstract

1. Introduction

3D marine ecological models are useful....

...But expensive and difficult to run.

In this article we look at ways to simulate 3D ecosystems more cheaply by running many parallel 1D regional models.

We are going to test that idea on the Red Sea because....

We will also the hybrid-SEIK data assimilation scheme, because....

We will assimilate Chl data even if it is imperfect because it is the best available data for the Red Sea.

What we are going to do in this paper step by step.

What is new in this paper and why.

Introduce sections.

*Corresponding author
Email address: ibrahim.hoteit@kaust.edu.sa (Ibrahim Hoteit)

2. Data

2.1. Chlorophyll data

We use CCI monthly and 8-days CHL data.

> Satellite data provide chlorophyll (CHL) concentrations with a spatial and temporal resolution not achievable with in situ observations, making them particularly relevant to the Red Sea, where very few in situ data collection are conducted.
>
> Level-3 mapped data from the NASA SeaWiFS (Sea-Viewing Wide Field-of-View Sensor) satellite sensor are used in this study. The dataset is publicly available at `http://oceancolor.gsfc.nasa.gov`. In this study, we use the 9km resolution mapped weekly averages from January 1998 to December 2007 (460 time steps). At each time step, a 133×188 pixel map is available for a domain extending from longitudes between 33°E and 44°E and latitudes between 12°N and 28°N, of which 5635 pixels correspond to actual Red Sea surface (see Figure **??**(a)). A log-transformation was applied in order to obtain an approximately Gaussian distribution *Campbell* (1995). Pixels with too few observations were discarded, and a control quality check was applied to remove outliers *Willis* (2004).
>
> Remotely sensed CHL may have missing data because of cloud coverage. The cloud variability in the Red Sea follows a seasonal cycle. Figure **??**(c) shows that the cloud coverage is particularly pronounced during summers because of the monsoon and it is sparse during winters. The cloud coverage is, however, not homogenous over the Red Sea. It is much more pronounced in the south (figure **??**(b)). In this region, almost no data are available during summers.

2.2. DINEOF

CCI data present missing data, in particular, in the southern Red Sea during summer. In order to have a complete dataset on which can apply a clustering algorithm, we use DINEOF, a data filling algorithm. The Chl data is averaged over each region to give a data time-series for each of them.

> The DINEOF (Data Interpolating Empirical Orthogonal Function) is an EOF-based, recursive method for the reconstruction of data matrices with missing values *Beckers and Rixen* (2003); *Alvera-Azcárate et al.* (2009). It estimates the values of the missing data by successive singular values decompositions (SVD) of a given data matrix and truncated reconstructions. The advantage of this method is that it does not require any a priori information about the data. It has been successfully used for reconstruction of incomplete chlorophyll datasets in different regions of the ocean *Miles and He* (2010); *Sirjacobs et al.* (2011); *Waite and Mueter* (2013).
>
> Let $\mathbf{X}$ be an $m \times k$ centered data matrix with missing values initially filled with 0s. Then, until the missing values have converged, the following steps are repeated. An SVD is first applied to the data matrix: $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, with $\mathbf{U}$ an $m \times m$ unitary matrix, $\mathbf{\Sigma}$ an $m \times k$ diagonal matrix and $\mathbf{V}$ a $k \times k$ unitary matrix. The missing values are then replaced by the truncated reconstruction order n of the data matrix: $\{\mathbf{X}\}_{i,j} = \{\mathbf{U}^{(n)}\mathbf{\Sigma}^{(n)}(\mathbf{V}^{(n)})^T)\}_{i,j}$, for i, j indices of the missing values, with $\mathbf{U}^{(n)}$ the $m \times n$ matrix composed of the n first columns of $\mathbf{U}$, $\mathbf{V}^{(n)}$ the $k \times n$ matrix composed of the n first columns of $\mathbf{V}$, and $\mathbf{\Sigma}^{(n)}$ the $n \times n$ diagonal matrix with the n largest eigenvalues on its diagonal. It is assumed that the eigenvalues and eigenvector are sorted by decreasing order of eigenvalues. In *Alvera-Azcárate et al.* (2009), the authors introduced the filtering of the temporal covariance matrix as a way of reducing spurious oscillations that may appear when the data are sparsely sampled in time. This filtering is controlled by the parameter of the Laplacian filter and the number of times the filter is applied.
>
> The values of the DINEOF parameters are determined following

the method outlined in *Alvera-Azcárate et al.* (2009). The smoothing parameter of the Laplacian filter is set to 0.005. The number of modes in the truncation and the number of times the filter is applied are chosen following a cross-validation technique. A random subset of observed values is taken from X and assumed to be missing before the DINEOF is applied. The algorithm is then run with different numbers of iterations (1, 3, 10, 30, 100) and orders of truncation (from 2 to 50). The set of parameters minimizing the RMS error over the cross-validation data is chosen as the best number of iterations and order of truncation. The approach of *Beckers et al.* (2006) is followed to select a cross-validation dataset. Instead of selecting it by sampling the dataset point by point, contiguous regions are set aside. These regions correspond to regions of missing data from the original dataset and are selected randomly until 3% of the data have been extracted.

2.3. Clustering

We use clustering algorithm to divide the Red Sea into regions with similar behavior. We tried K-means and Gaussian Mixture Model, a generalization of the former.GMM was found to give better results.

I used clustering algorithms in order to derive the Red Sea eco-regions. These were applied to monthly log-concentration of chlorophyll. I used SeaWiFS data, that has been filled using DINEOF. I used the popular K-means, and the Gaussian Mixture Model (GMM) clustering algorithms.

I found that GMM provides more robust results. With any number of clusters, we obtain a division of the Red Sea into regions of comparable sizes. With 5 clusters, the regions (shown in figure **??**) are very similar to those identified by *Raitsos et al.* (2013). Contrary to the purely latitudinal division proposed by the former, we observe that the separation between clusters is curved at the position of major Red Sea eddies. The fact that the curvature is oriented toward the south suggests that most nutrients propagate northward from the Gulf of Aden.

In Chapter 2, I plan to use the dataset constructed in Chapter 1. By using CCI chlorophyll data instead of SeaWiFS, the need for data filling is minimized. This is desirable, as data filling can introduce biases. It will also be possible to use additional variables. For example, we can expect the temperature and the bathymetry to have a large impact on the Red Sea phytoplankton biology. Sea level anomaly can be useful in that it indicates the presence of mesoscale eddies. Finally, alternative clustering algorithms will be tested.

3. Model and Assimilation

3.1. 1D-ERSEM model

We use a 1D coupled ERSEM model. The physical forcing comes from a 3D circulation simulation of the Red Sea [Yao 2014]. The ecological models are initialized with the results of the 3D Red Sea ecology simulation [Triantfyllou2013].

The 1D regional ecological models used for this thesis have been configured and are operational. Three models will be used: for the northern, central and southern Red Sea. The extreme south of the Red Sea is not modeled, as its dynamics is poorly understand and we miss in situ data. The ecology is modeled with ERSEM, and the hydrodynamics is modeled with the MITgcm.

The results of the MITgcm are those from *Yao et al.* (2014a,b), in which a simulation of the Red Sea and part of the Gulf of Aden circulation was run over 50 years. The NCEP data were used for atmospheric forcing, and the ocean ECCO data for the open boundary conditions in the Gulf of Aden. The output of the 50 years run are used for the temperature and vertical circulation at the modeled points.

ERSEM simulates the complete water column with the pelagic and benthic ecosystems, as well a their coupling. The equations model the flow of carbon, nitrogen, phosphorus and silicon in the ecosystem. Living organisms are modeled in terms of population processes (growth and mortality) and physiological processes (ingestion, respiration, excretion, and egestion). The biota is divided into functional groups according to their trophic levels: producers (phytoplankton), consumers (zooplankton) and decomposers (bacteria), and further subdivided according to their sizes (*Baretta et al.*, 1995).

The ecological models are initialized with the results of a 3D ecological simulation of the Red Sea (*Triantafyllou et al.*, 2014). The nutrient concentrations are initialized using values from the World Ocean Atlas 2005 (WOA 2005).

3.2. Data Assimilation

To improve the results of the simulation we use the hybrid-SEIK assimilation scheme, detailed in this subsection.

The assimilation scheme for the ecological models has been implemented and is operational. The chosen scheme is the hybrid-SEIK, described in *Hamill and Snyder* (2000). It can be seen as a variant of the 3DVAR variational assimilation scheme. 3DVAR assumes that the error forecast covariance is fixed in time. In the case of the hybrid, the covariance is a linear combination of the 3DVAR covariance and the time-evolving SEIK covariance matrix. Figure **??** shows the assimilation scheme improves the fit of the model to the chlorophyll data.

The problem of optimal filtering can be solved exactly by the Kalman Filter for linear systems. For nonlinear models, one can use the Extended Kalman (EK) filter, in which the model is linearized by computing the error covariance function. However, when the state is large, as is often the case for oceanographic applications, the EK is intractable. In that case, SEEK can be used, where the error covariance function is projected into a smaller subspace. This subspace evolves to ensure that most of the error is represented and filtered out. SEIK can be viewed as an ensemble variant of the SEEK, where the error covariance function is represented exactly by an ensemble of states. This avoids the computation of model gradients, and allows the assimilation scheme to perform better when this model is strongly non-linear. SEIK has been shown to be efficient for large-scale 3D ecosystem simulations (*Triantafyllou et al.*, 2003).

The Expectation-Maximization scheme to estimate the filter parameters has also been derived. It is similar to that proposed by *Tandeo et al.* (2014), except that the model is non linear. The scheme will be used to improve the estimates of the observation and model covariance errors.

4. Results

4.1. Model evaluation

Here, we compare the results of the free-run with the assimilated-run. We show that we have a good prediction skill, and that the assimilation improves the model.

4.2. Analysis

Here we look at the results and interpret them biologycally. Do we find comparable results as Acker, Raitsos, Weiker, etc. What can we say about the hypothesis that they made about he process that drive primary productivity in the Red Sea.

5. Conclusion

Are several 1D paralled 1D models a good alternative to 3D simulations?

What did we learn about the Red Sea ecology?

Future works?

Acknowledgment

The research reported in this publication was supported by King Abdullah University of Science and Technology (KAUST).

6. Bibliography

Abualnaja, Y., V. P. Papadopoulos, S. A. Josey, I. Hoteit, H. Kontoyiannis, and D. E. Raitsos (2015), Impacts of climate modes on air-sea heat exchange in the Red Sea, *Journal of Climate*, doi:10.1175/JCLI-D-14-00379.1, in press.

Acker, J., G. Leptoukh, S. Shen, T. Zhu, and S. Kempler (2008), Remotely-sensed chlorophyll a observations of the northern Red Sea indicate seasonal variability and influence of coastal reefs, *Journal of Marine Systems*, *69*, 191–204, doi:10.1016/j.jmarsys.2005.12.006.

Alvera-Azcárate, 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, *Journal of Geophysical Research*, *112*, C03008, doi:10.1029/2006JC003660.

Alvera-Azcárate, A., A. Barth, D. Sirjacobs, and J.-M. Beckers (2009), Enhancing temporal correlations in EOF expansions for the reconstruction of missing data using DINEOF, *Ocean Science*, *5*, 475–485, doi:10.5194/os-5-475-2009.

Anderson, T. R. (2005), Plankton functional type modelling: running before we can walk?, *Journal of Plankton Research*, *27*(11), 1073–1081, doi:10.1093/plankt/fbi076.

Baretta, J. W., W. Ebenhöh, and P. Ruardij (1995), The European regional seas ecosystem model, a complex marine ecosystem model, *Netherlands Journal of Sea Research*, *33*(3/4), 233–246, doi:10.1016/0077-7579(95)90047-0.

Beckers, J. M., and M. Rixen (2003), EOF calculations and data filling from incomplete oceanographic datasets, *Journal of Atmospheric and Oceanic Technology*, *20*(12), 1839–1856, doi:10.1175/1520-0426(2003)020¡1839:Ecadff¿2.0.Co;2.

Beckers, J.-M., A. Barth, and A. Alvera-Azcárate (2006), DINEOF reconstruction of clouded images including error maps - application to the Sea-Surface Temperature around Corsican Island, *Ocean Science*, *2*, 183–199, doi: 10.5194/osd-3-735-2006.

Brewin, R. J. W., D. E. Raitsos, Y. Pradhan, and I. Hoteit (2013), Comparison of chlorophyll in the Red Sea derived from MODIS-Aqua and in vivo fluorescence, *Remote Sensing of Environment*, *136*, 218–224, doi: 10.1016/j.rse.2013.04.018.

Butenschön, M., and M. Zavatarelli (2012), A comparison of different versions of the SEEK filter for assimilation of biogeochemical data in numerical models of marine ecosystem dynamics, *Ocean Modelling*, *54-55*, 37–54, doi:10.1016/j.ocemod.2012.06.003.

Campbell, J. W. (1995), The lognormal distribution as a model for bio-optical variability in the sea, *Journal of Geophysical Research*, *100*(C7), 13,237–13,254, doi:10.1029/95JC00458.

Ciavatta, S., R. Torres, S. Saux-Picart, and J. I. Allen (2011), Can ocean color assimilation improve biogeochemical hindcasts in shelf seas?, *Journal of Geophysical Research*, *116*, C12043, doi:10.1029/2011JC007219.

Ciavatta, S., R. Torres, V. Martinez-Vicente, T. Smyth, G. Dall'Olmo, L. Polimene, and J. I. Allen (2014), Assimilation of remotely-sensed optical properties to improve marine biogeochemistry modelling, *Progress in Oceanography*, *127*, 74–95, doi:10.1016/J.Pocean.2014.06.002.

Edwards, C. A., A. M. Moore, I. Hoteit, and B. D. Cornuelle (2015), Regional ocean data assimilation, *Annual Review of Marine Science*, *7*, 21–42, doi: 10.1146/annurev-marine-010814-015821.

Fennel, W., and T. Neumann (2004), *Introduction to the Modelling of Marine Ecosystems*, *Elsevier Oceanography Series*, vol. 72, Elsevier, Amsterdam, The Netherlands.

Fontana, C., P. Brasseur, and J.-M. Brankart (2013), Toward a multivariate reanalysis of the North Atlantic Ocean biogeochemistry during 1998-2006 based on the assimilation of SeaWiFS chlorophyll data, *Ocean Science*, *9*, 37–56, doi:10.5194/Os-9-37-2013.

Hamill, T. M., and C. Snyder (2000), A hybrid ensemble Kalman filter-3D variational analysis scheme, *Monthly Weather Review*, *128*, 2905–2919, doi: 10.1175/1520-0493(2000)128¡2905:Ahekfv¿2.0.Co;2.

Hoteit, I., G. Korres, and G. Triantafyllou (2005), Comparison of extended and ensemble based Kalman filters with low and high resolution primitive equation ocean models, *Nonlinear Processes in Geophysics*, *12*, 755–765.

Korres, G., G. Triantafyllou, G. Petihakis, D. E. Raitsos, I. Hoteit, A. Pollani, S. Colella, and K. Tsiaras (2012), A data assimilation tool for the Pagasitikos Gulf ecosystem dynamics: Methods and benefits, *Journal of Marine Systems*, *94*, S102–S117, doi:10.1016/J.Jmarsys.2011.11.004.

Mann, K. H., and J. R. N. Lazier (2006), *Dynamics of marine ecosystems: Biological-Physical Interactions in the Oceans (3rd edition)*, Blackwell Publishing, Malden, Maryland.

McClain, C. R. (2009), A decade of satellite ocean color observations, *Annual Review of Marine Science*, *1*, 19–42, doi: 10.1146/Annurev.Marine.010908.163650.

Miles, T. N., and R. He (2010), Temporal and spatial variability of Chl-a and SST on the South Atlantic Bight: Revisiting with cloud-free reconstructions of MODIS satellite imagery, *Continental Shelf Research*, *30*, 1951–1962, doi: 10.1016/j.csr.2010.08.016.

Pal, R., and A. K. Choudhury (2014), *An introduction to phytoplanktons : diversity and ecology*, Springer, New Delhi, India.

Pettersson, L. H., and D. Pozdnyakov (2013), *Monitoring of harmful algal blooms*, Springer-Praxis books in geophysical sciences, Springer, Heidelberg, Germany.

Racault, M. F., D. E. Raitsos, M. L. Berumen, R. J. Brewin, T. Platt, S. Sathyendranath, and I. Hoteit (), Phytoplankton phenology indices in coral reef ecosystems: application to ocean-colour observations in the red sea.

Raitsos, D. E., I. Hoteit, P. K. Prihartato, T. Chronis, G. Triantafyllou, and Y. Abualnaja (2011), Abrupt warming of the Red Sea, *Geophysical Research Letters*, *38*, L14601, doi:10.1029/2011gl047984.

Raitsos, D. E., Y. Pradhan, R. J. W. Brewin, G. Stenchikov, and I. Hoteit (2013), Remote sensing the phytoplankton seasonal succession of the Red Sea, *PLoS One*, *8*(6), e64909, doi:10.1371/journal.pone.0064909.

Raitsos, D. E., X. Yi, T. Platt, M.-F. Racault, R. J. W. Brewin, Y. Pradhan, V. P. Papadopoulos, S. Sathyendranath, and I. Hoteit (2015), Monsoon oscillations regulate fertility of the Red Sea, *Geophysical Research Letters*, *42*, doi:10.1002/2014gl062882, in press.

Richlen, M. L., S. L. Morton, E. A. Jamali, A. Rajan, and D. M. Anderson (2010), The catastrophic 2008-2009 red tide in the Arabian gulf region, with observations on the identification and phylogeny of the fish-killing dinoflagellate Cochlodinium polykrikoides, *Harmful Algae*, *9*, 163–172, doi: 10.1016/J.Hal.2009.08.013.

Robinson, I. S. (2010), *Discovering the ocean from space : the unique applications of satellite oceanography*, Springer-Praxis books on geophysical sciences, 1st ed., Springer, Heidelberg, Germany.

Sirjacobs, D., A. Alvera-Azcárate, A. Barth, G. Lacroix, Y. Park, B. Nechad, K. Ruddick, and J.-M. Beckers (2011), Cloud filling of ocean colour and sea surface temperature remote sensing products over the southern north sea by the data interpolating empirical orthogonal functions methodology, *Journal of Sea Research*, *65*, 114–130, doi:10.1016/j.seares.2010.08.002.

Steinmetz, F., P. Y. Deschamps, and D. Ramon (2011), Atmospheric correction in presence of sun glint: application to MERIS, *Optics Express*, *19*(10), 9783–9800, doi:10.1364/Oe.19.009783.

Tandeo, P., M. Pulido, and F. Lott (2014), Offline parameter estimation using EnKF and maximum likelihood error covariance estimates: Application to a subgrid-scale orography parametrization, *Quarterly Journal of the Royal Meteorological Society*, doi:10.1002/qj.2357.

Taylor, M. H., M. Losch, M. Wenzel, and J. Schröter (2013), On the sensitivity of field reconstruction and prediction using empirical orthogonal functions derived from gappy data, *Journal of Climate*, *26*, 9194–9205, doi:10.1175/Jcli-D-13-00089.1.

Triantafyllou, G., I. Hoteit, and G. Petihakis (2003), A singular evolutive interpolated Kalman filter for efficient data assimilation in a 3-D complex physical-biogeochemical model of the Cretan Sea, *Journal of Marine Systems*, *40-41*, 213–231, doi:10.1016/S0924-7963(03)00019-8.

Triantafyllou, G., I. Hoteit, X. Luo, K. Tsiaras, and G. Petihakis (2013), Assessing a robust ensemble-based Kalman filter for efficient ecosystem data assimilation of the Cretan Sea, *Journal of Marine Systems*, *125*, 90–100, doi:10.1016/J.Jmarsys.2012.12.006.

Triantafyllou, G., F. Yao, G. Petihakis, K. P. Tsiaras, D. E. Raitsos, and I. Hoteit (2014), Exploring the Red Sea seasonal ecosystem functioning using a three-dimensional biophysical model, *Journal of Geophysical Research-Oceans*, *119*, 1791–1811, doi:10.1002/2013jc009641.

Waite, J. N., and F. J. Mueter (2013), Spatial and temporal variability of chlorophyll-a concentrations in the coastal Gulf of Alaska, 1998-2011, using cloud-free reconstructions of SeaWiFS and MODIS-Aqua data, *Progress in Oceanography*, *116*, 179–192, doi:10.1016/j.pocean.2013.07.006.

Weikert, H. (1987), Plankton and the pelagic environment, in *Red Sea*, edited by J. . E. Alasdair, pp. 90–111, Pergamon, Oxford, United Kingdom, doi:10.1016/B978-0-08-028873-4.50010-4.

Willis, J. K. (2004), Interannual variability in upper ocean heat content, temperature, and thermosteric expansion on global scales, *Journal of Geophysical Research*, *109*, C12036, doi:10.1029/2003jc002260.

Yao, F., and I. Hoteit (2015), Thermocline regulated seasonal evolution of surface chlorophyll in the Gulf of Aden, *PLoS One*, in press.

Yao, F., I. Hoteit, L. J. Pratt, A. S. Bower, A. Köhl, G. Gopalakrishnan, and D. Rivas (2014a), Seasonal overturning circulation in the Red Sea: 2. winter circulation, *Journal of Geophysical Research-Oceans*, *119*, 2263–2289, doi:10.1002/2013jc009331.

Yao, F., I. Hoteit, L. J. Pratt, A. S. Bower, P. Zhai, A. Köhl, and G. Gopalakrishnan (2014b), Seasonal overturning circulation in the Red Sea: 1. model validation and summer circulation, *Journal of Geophysical Research-Oceans*, *119*, 2238–2262, doi:10.1002/2013jc009004.

Zhai, P., and A. Bower (2013), The response of the red sea to a strong wind jet near the Tokar Gap in summer, *Journal of Geophysical Research-Oceans*, *118*, 422–434, doi:10.1029/2012jc008444.

Zhan, P., A. C. Subramanian, F. C. Yao, and I. Hoteit (2014), Eddies in the Red Sea: A statistical and dynamical study, *Journal of Geophysical Research-Oceans*, *119*(6), 3909–3925, doi:10.1002/2013jc009563.