Simulating the Red Sea Ecology with parallel 1D marine ecosystem models and clustering

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....

- marine ecology models represent biogeochemical interactions as differential equations.
- Can be as minimal as NPZ or as complete as NPZ

...are useful because....

- HAB affect public health, desalination and coastal economy
- predicting chlorophyll can help fisheries
- For research: better understand the large-scale ecosystem
- Especially usefull because we lack data about the subsurface phenomena

...But expensive and difficult to run.

- (*Anderson*, 2005) lot of underdetermination
- circulation model very expensive, because very small grid

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

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

- Divide the Red sea in small regions with similar ecology
- Reduces underdetermination
- Ensure parametrization is better for each region

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

- Red Sea is an interesting environment: extreme temperatures and salinity
- Very rich and preserved ecosystem
- Unexplored environment
- Lack of data: therefore developing models is important

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

- Assimilation constrains the model and reduces underdetermination
- Mitigates the fact that initial conditions are unknown
- SEIK is better than SEEK for strongly nonlinear models
- SEIK is better than EnKF when fewer observations than states
- hybridization reduces the ensemble size and the computational cost

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

- Chl data allows to observe large scale ecological patterns with high spatial and temporal coverage.
- Compared with in situ data that are limited in time and space, and expensive.
- However chl data suffers from missing values due clouds, aerosols, etc.

- Also bad values near the coast, case II waters
- Both problem particularly affect the southern Red Sea, that has nearly no observation in the summer during some months.
- However as lack of in situ data, this is the best we have currently in the Red Sea

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

- Fill the data with DINEOF
- Apply clustering to the Red Sea
- Implement 1D models
- Run models with assimilation
- Analyze results and compare to findings in previous studies

What is new in this paper and why.

- We use CCI data: which is a new dataset, not fully exploited in Red Sea.
- We do eco-region clustering for the first time in the Red Sea.
- We have assimilation of ecological model with hybrid-SEIK published for the first time.
- We improve our understanding of the Red Sea ecology in its different parts.

Introduce sections.

2. Data

2.1. CCI chlorophyll data

We use CCI chlorophyll data because it has more coverage..

- MODIS data has a lot of missing data especially during summer in the South

- CCI data, merges three different sensors and uses the POLYMER algorithm.
- As a result the coverage increases dramatically.
- This is the first dataset that has significant coverage in the southern Red Sea, and that is why we will use it.
- We use 4km resolution L3 CHL product between such and such coordinates.
- We use weekly data for the clustering and 8days data for the assimilation

With a quick look at the data this is what we see....

- plot coverage
- plot average chlorophyll
- plot seasonal chlorophyll

2.2. DINEOF

There are still missing data in CCI, so we use DINEOF for data filling because....

- DINEOF is an EOF based non parametric data filling methods
- introduced orignally by *Beckers and Rixen* (2003)
- Has been applied to geoscience datasets, in particular to chl datasets
- Shown in *Taylor et al.* (2013) to be more efficient that its competition.

More or less this is the way DINEOF works....

- Describe here the algo

This is how we applied DINEOF....

- Choices of parameters and cross validation method and why we made these choices.

Now we show the results of DINEOF.

- Show minimization of error plot.
- Show the filling of a region.

2.3. Clustering

To do 1D models, we cluster the Red Sea using clustering algorithms. We chose GMM because....

- There are many clustering algorithms on the market: for example...
- Discuss advantages and inconvenients of some of them (find reference)
- Finally we tried some of them and found that GMM was given better results, in comparison to what we expected.

This is more or less the way GMM works....

- Describe how GMM algo works

This is how we used it....

- We wanted 3 broad regions in the northern, central and southern Red Sea.
- We ran GMM with k varying from 3 to 7.
- We looked at some of this tests to see if the clusters where good.
- At the end we settled with k=? because it was good for our purposes and the regions where closed to to what *Raitsos et al.* (2013) found.

This is what we got....

- Show plot of clusters
- Comment on clusters, and what was found by *Raitsos et al.* (2013).

3. Model and Assimilation

3.1. 1D-ERSEM model

Description of ERSEM.

- ERSEM develop originally for the northern Sea
- Complete ecology modeling
- Been applied in many ecosystem
- in particular used for the Red Sea simulation *Triantafyllou et al.* (2013)
- Very complex: many parameters and variables

Initialization/Parameters/Forcing.

- We initialize with the values found by *Triantafyllou et al.* (2013)
- The nutrients are initialized using the values of WOA
- Parameters are chosen like this
- Forcing come from the simulation by *Yao et al.* (2014a,b)

3.2. Data Assimilation

We chose hybrid-SEIK DA scheme because....

- Data assimilation is necessary to improve the forecasting skill of complex geophysical models
- It constrains models that are imperfect and whose parametrization is difficult to do.
- SEEK has a long history in assimilation into ecological models.
- Its ensemble variant SEIK, has been shown to behave better for very nonlinear systems
- However, SEIK requires to run the model many time in parallel.
- To reduce the ensemble size and improve the efficiency, we use the hybrid formulation proposed in **?**.

Equations of hybrid-SEIK.

- Show the steps of the hybrid-SEIK scheme

Parameters.

- How we choose the scheme parameters and why.

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.

Plot some models output vs the data and qualitative comments.

Discuss different metrics to evaluate model and DA schemes.

Show a couple metrics on our results.

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.

Discuss differences between climatology and 2003-2004.

Comment on the role of overturning and stratification.

What is the limiting nutrient?.

$$\frac{N3N + N4N}{N1P}$$

vs Redfied ratio

Compute production at different time of years and compare with Weikert1987/Acker2008/Koblentz.

Study the DCM.

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

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.

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.

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.

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, 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.

Yao, F., I. Hoteit, L. J. Pratt, A. S. Bower, P. Zhai, A. Köhl, and G. Gopalakrishnan (2014a), 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.

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