

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

Denis Dreano<sup>a</sup>, George Triantafyllou<sup>c</sup>, Bani Mallick<sup>b</sup>, Ibrahim Hoteit<sup>a,\*</sup>

<sup>a</sup>*Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah  
University of Science and Technology*

<sup>b</sup>*Department of Statistics, Texas A&M University*

<sup>c</sup>*Hellenic Center for Marine Research*

---

## Abstract

Abstract

---

## 1. Introduction

*What are marine ecosystem models.* ERSEM, NPZ.

*Marine ecosystem models have scientific and practical applications.* Forecasting,  
harmful algae blooms, scientific inquiry when not enough data.

*Complex 3D ecological models have limitations.* Indetermination and computa-  
tional cost.

*Assimilation is necessary but increases the computational cost.* Ecological sys-  
tems are chaotic system that require assimilation to maintain a good prediction  
skill. But assimilation increases the computational cost. Assimilation of eco  
models remains a challenge.

*The hybrid-SEIK.* Has never been use for eco models before.

---

\*Corresponding author

Email address: [ibrahim.hoteit@kaust.edu.sa](mailto:ibrahim.hoteit@kaust.edu.sa) (Ibrahim Hoteit)

12 *Ocean colour data has been assimilated a lot to eco models.* chlorophyll is a green  
13 colored pigment critical for photosynthesis and found in plants and algae (*Pal*  
14 *and Choudhury*, 2014). Give a detectable green coloration to the water when  
15 phytoplankton is present (*Robinson*, 2010). Remotely sensed ocean colour data  
16 give highly available data with high coverage both in time and space (*Robinson*,  
17 2010). Ocean colour data products like chl are very good proxy for phytoplank-  
18 ton concentration.

19 *But has some limitations.* CHL dataset have missing data because of clouds.  
20 Which is a big problem in the southern Red Sea in summer where the coverage  
21 is almost zero *Racault et al.*. Moreover the chl is difficult to estimate in case II  
22 optically complex waters, especially near the coasts. It also particularly affects  
23 the southern Red Sea which is very shallow.

24 *Solutions to missing data.* One of the most popular approach for data filling of  
25 remotely-sensed chlorophyll data is DINEOF introduced by *Beckers and Rixen*  
26 (2003).

27 *The Red Sea is a relatively unexplored sea.* Not a lot of studies, not a lot of in  
28 situ data. Must use models and remotely-sensed data. RS is TTS with strong  
29 stratification that limits vertical diffusion of nutrients. (*Mann and Lazier*, 2006).  
30 Other than the gulf of Aden (*Yao and Hoteit*, 2015), the RS has no known  
31 significant input of nutrients and is oligotrophic (*Raitsos et al.*, 2013; *Weikert*,  
32 1987). Red Sea has a rich ecosystem and unique ecosystem that has adapted  
33 its extreme environment (*Raitsos et al.*, 2011). RS relatively well preserved but  
34 increasingly fragilized by human activities. Sharp increase of temperature in  
35 the past decade threatend the RS environment. (*Raitsos et al.*, 2011).

36 *Major biological patterns.* Current hypotheses about primary production in Red  
37 Sea. There is a lack of missing data therefore the large scale ecological dynamics  
38 is poorly known (*Raitsos et al.*, 2013; *Triantafyllou et al.*, 2014). The role  
39 of aerosol deposition could be important but has not been investigated yet  
40 (*Raitsos et al.*, 2013). Eddies are believed to play important role (*Raitsos et al.*,

2013; Zhan *et al.*, 2014). Chlorophyll increase from north to south (Raitsos *et al.*, 2013). Secondary summer bloom (but not in NRS) (Raitsos *et al.*, 2013). Strong Interannual variability (Raitsos *et al.*, 2013). Exchange of water with GOA is a major driver of productivity for the whole Red Sea (Triantafyllou *et al.*, 2014). SRS winter bloom attributed to wind-driven intrusion (Raitsos *et al.*, 2013). Deep convection in winter plays a big role in the northern Red Sea (Raitsos *et al.*, 2013). Red Sea circulation impacted by eddies that could impact productivity (Zhai and Bower, 2013). Central red Sea anti cyclonic eddy is believed to control June peak (Raitsos *et al.*, 2013). Climate mode indices have impact on the Red Sea (Raitsos *et al.*, 2015).

*Objective:*. 3D ecological models are expensive to run. Can we divide the Red Sea into regions and have 1D models running in each of them in parallel? In this article we cluster the Red Sea in 3 different eco-regions using automatic unsupervised learning algorithms. We then run an assimilative 1D ecological model on each of the region and analyze the results.

*Introduce chapters.*

Typical tropical seas (TTS), like the Red Sea, are characterized by a highly stratified structure, where warm nutrient-depleted surface water is separated from the cold nutrient-rich deep water by a steep gradient of temperature zone called pycnocline. The pycnocline acts as barrier that limits the upward nutrients flow (Mann and Lazier, 2006). As a result, TTS are oligotrophic and have low chlorophyll concentrations. Until recently, marine biologists believed that tropical and subtropical seas have therefore a very low productivity. However, recent investigations have contested this idea, suggesting that different upwelling mechanisms (winter deep mixing, storms, eddies, etc) exist, which bring new nutrients to the surface water (Mann and Lazier, 2006).

Despite being an oligotrophic and challenging environment for marine life, the Red Sea presents a surprisingly rich and diverse

ecosystem (*Raitsos et al.*, 2011), and a very well developed coral reef system (*Racault et al.*). The source of nutrient for sustaining such a developed ecosystem is not well understood yet, but the exchange with the open ocean, the atmospheric depositions and transport through the mesoscale eddies are believed to play an important role (*Raitsos et al.*, 2013; *Zhan et al.*, 2014).

Although the Red Sea environment is still relatively well preserved, it is increasingly stressed by human activities. The continuous urbanization and fishing activity contribute to the fragilization of this unique ecosystem (*Acker et al.*, 2008). An abrupt increase of temperature has further occurred in the last decade, which may threaten the fragile coral reef system (*Raitsos et al.*, 2011).

Because of the lack of in-situ data, the large-scale phytoplankton dynamics of the Red Sea remain largely unknown (*Raitsos et al.*, 2013; *Triantafyllou et al.*, 2014). However, in recent studies, remotely-sensed data and computer simulations have been used to improve our knowledge of the ecology of this region. The Red Sea is deficient in major nutrients (*Weikert*, 1987), and the only significant input of water comes from the Gulf of Aden (*Yao and Hoteit*, 2015). This explains a general pattern of chlorophyll concentration increase from north to south (*Raitsos et al.*, 2013), with the lowest concentration found in the northern central Red Sea. This pattern can be seen in figure ???. The Red Sea also displays a distinct seasonality, with a peak in concentration during the winter. A weak summer peak is also observed around July, everywhere except in the northernmost region (*Raitsos et al.*, 2013). Despite this regularity, a strong interannual variability is observed, with blooms that can reach mesotrophic concentration levels (*Raitsos et al.*, 2013). According to *Triantafyllou et al.* (2014), the variations in the Red Sea ecology are mainly driven by physical circulation. In the rest of this section, we explore some of the mechanisms that have been linked to the major features of

102 chlorophyll concentration.

103 The exchange of water with the nutrient-rich Gulf of Aden is a  
104 major driving mechanism for the whole Red Sea (*Triantafyllou et al.*,  
105 2014). It is the most important source of nutrients. The maximum  
106 chlorophyll concentration observed in the southern Red Sea during  
107 winter is attributed to wind-driven water intrusion (*Raitsos et al.*,  
108 2013). In Summer, this exchange of water is believed to be the only  
109 significant source of nutrients for the whole Red Sea. The influence  
110 of the water intrusion weakens as the latitude increases, explaining  
111 the low concentration in the northern half of the Red Sea (*Raitsos*  
112 *et al.*, 2013).

113 Deep convection also plays an important role in allowing nutrient-  
114 rich deep water to mix with water of the euphotic zone. Vertical  
115 mixing is the most vigorous in the northern extremity of the Red  
116 Sea during the winter. This explains its higher chlorophyll concen-  
117 tration compared to the north-central Red Sea, a region of weak  
118 mixing (*Raitsos et al.*, 2013). The northern Red Sea mixing is be-  
119 lieved to be driven by wind (*Raitsos et al.*, 2013).

120 The Red Sea circulation is strongly influenced by mesoscale ed-  
121 dies (*Yao et al.*, 2014a,b; *Zhan et al.*, 2014) that could impact pri-  
122 mary production (*Zhai and Bower*, 2013). In particular, the anti-  
123 cyclonic eddy in the central Red Sea is believed to control the June  
124 concentration peak and the summer productivity of this region, by  
125 transporting nutrients and/or phytoplankton from the adjacent coral  
126 reefs (*Raitsos et al.*, 2013). In the northern Red Sea, a cold-core eddy  
127 plays a role in enhancing the vertical mixing in this region (*Raitsos*  
128 *et al.*, 2013).

129 Aerial depositions of dust could also be an important input of  
130 nutrients for the Red Sea, but it has been largely left unexplored  
131 (*Triantafyllou et al.*, 2014). *Raitsos et al.* (2013) noticed for example  
132 that sand storms in the Red Sea most frequently happen in June and

July, which coincides with the summer chlorophyll peak. Finally, climate mode indices have been shown to be strongly correlated with air-sea heat exchanges in the Red Sea (*Abualnaja et al.*, 2015), and might therefore influence its biology. This has been recently confirmed by *Raitsos et al.* (2015), who have shown that El Nino has a positive impact on the chlorophyll concentration, by strengthening the wind transporting nutrients into the Red Sea from the Gulf of Aden.

Chlorophyll is a molecule present in algae, phytoplankton and plants that is critical for photosynthesis. It is a poor absorber of green light, and is responsible for the coloration of plants (*Pal and Choudhury*, 2014). When phytoplankton are present in high concentrations, the water also takes a detectable green coloration (it can also take a red or blue coloration depending on the type of dominating phytoplankton) (*Robinson*, 2010). This offers an efficient way to observe the phytoplankton concentration from space.

In-situ measurement of chlorophyll concentration can be gathered through scientific cruises, buoy stations or gliders (unmanned submarines). These methods are expensive to deploy and therefore generally have limited temporal and spatial coverage (*Robinson*, 2010). Political issues, as in the Red Sea, as well as security issues, as in the Arabian Sea, set also barriers to in-situ measurements.

Satellite measurements of chlorophyll provide excellent proxies for phytoplankton concentrations with a good temporal and spatial coverage (*Robinson*, 2010). The SeaWiFS, MODIS and MERIS missions have provided an uninterrupted coverage of the world since 1997. High-resolution maps of daily chlorophyll concentration are freely accessible to the scientific community (*McClain*, 2009). Despite some limitations, like missing data due to cloud coverage and sunglint, or problematic values in coastal areas, remotely-sensed

chlorophyll concentrations are intensively used by the scientific community. In regions, like in the Red Sea, where little in-situ measurements are available (*Raitsos et al.*, 2013; *Brewin et al.*, 2013), these constitute the most important data source.

The quality of remotely-sensed chlorophyll data products such as MODIS and SeaWiFS in the Red Sea is comparable with that of the rest of the world for case I waters (open sea) (*Brewin et al.*, 2013). However, the data contains a large amount of missing values because of persistent clouds, sun-glint and sensor saturation (*Racault et al.*). This problem is particularly acute during the summer in the southern Red Sea where the data coverage is almost null (*Racault et al.*), as shown in figure ??.

Chlorophyll concentration estimation in optically complex case II waters is a recurrent problem in this remotely-sensed data that particularly affects the southern Red Sea. In this region, the remotely sensed chlorophyll data could be overestimated (*Raitsos et al.*, 2013). However, all high values are not necessarily bad, as highly productive coral reefs are also present in this region (*Raitsos et al.*, 2013). However, these values have not been validated yet, due to the lack of in situ measurements (*Raitsos et al.*, 2013).

One solution to missing and bad values is to use a data filling algorithm, of which one of the most popular is the Date INterpolating Empirical Orthogonal Functions (DINEOF). It is an EOF based data filling approach introduced by *Beckers and Rixen* (2003). In *Sirjacobs et al.* (2011), it has been employed to fill chlorophyll data with 70% of missing values. *Taylor et al.* (2013) has compared DINEOF with other EOF-based reconstruction algorithms, suggesting that the former is the best method for data filling. DINEOF has been employed in several other chlorophyll studies (*Miles and He*, 2010; *Waite and Mueter*, 2013). It has also been used for multivariate reconstruction of SST fields using chlorophyll data in *Alvera-Azcárate*

194 *et al.* (2007).

195 The OC-CCI is a new chlorophyll data product that considerably  
196 increases the Red Sea coverage. It merges the data from sensors Sea-  
197 WiFS, MODIS and MERIS. Overall, it achieves a 75-80% coverage in  
198 the entire Red Sea basin against 50-65% for a single sensor (*Racault*  
199 *et al.*). During the summer, the improvement is dramatic, as shown  
200 in figure ???. This is mostly due to the use of the POLYMER al-  
201 gorithm (*Steinmetz et al.*, 2011) that allows to exploit MERIS data  
202 collected during hazy conditions. However, this new dataset has not  
203 been fully explored to revisit the assumptions made on the large-  
204 scale Red Sea phytoplankton productivity.

205 Models could be useful to identify causes behind the chlorophyll  
206 patterns we observe in the Red Sea. Many hypotheses have been  
207 made about the drivers of chlorophyll concentration in this regions,  
208 but some of them have not been yet investigated through models.  
209 The role played by the exchange of water with the Gulf of Aden and  
210 winter overturning in the northern Red Sea have been successfully  
211 modeled with a 3D coupled ecological model (*Triantafyllou et al.*,  
212 2014). However, the interaction between the open sea and coral  
213 reefs, and the role of atmospheric depositions have not been investi-  
214 gated yet. Models, can also be helpful for understanding governing  
215 dynamics affecting the chlorophyll concentration. In particular, the  
216 interaction between the productivity level of the different regions of  
217 the Red Sea is yet to be explored.

218 Model predictions of chlorophyll concentration also have practi-  
219 cal applications. Phytoplankton blooms can be harmful to humans  
220 and marine life and are closely monitored in many regions of the  
221 world (*Pettersson and Pozdnyakov*, 2013). In the Red Sea, where  
222 tourism and aquaculture are developing, it is likely to become a con-  
223 cern too. Phytoplankton is also directly, and indirectly through zoo-  
224 plankton, the cause of microfouling that affects desalination plants.



225 In 2008-2009, a red tide forced the shutdown of desalination plants  
226 along the Gulf of Oman and the Arabian Gulf (*Richlen et al.*, 2010).

227 There is a rich literature on the modeling of marine ecosystem  
228 using differential equations (see *Fennel and Neumann* (2004) for an  
229 introduction. In these models the interactions of complex physical,  
230 chemical and biological processes are modeled by differential equa-  
231 tions that represent the flow of carbon, nitrogen, phosphate and  
232 silicon. The biota is divided into trophic levels, and can be further  
233 divided by feeding methods and size classes (*Triantafyllou et al.*,  
234 2014).

235 Ecological deterministic models in use vary widely in diversity  
236 depending on on the number of state parameters and interactions  
237 represented. They can be as simple as the nutrient-phytoplankton-  
238 zooplankton (NPZ) model (*Anderson*, 2005) that only has three vari-  
239 ables representing two trophic levels and nitrate, or as complex as the  
240 European regional seas ecosystem model (ERSEM) that has dozens  
241 of variables (*Baretta et al.*, 1995). NPZ models are extensively used  
242 because of its simplicity and its capacity to model the the large-scale  
243 features of marine ecosystems (*Anderson*, 2005). ERSEM has been  
244 use in many studies. It has recently been coupled to the MITgcm  
245 circulation model used to simulate the Red Sea ecology (*Triantafyl-  
246 lou et al.*, 2014). However the complexity of these models makes  
247 them difficult to parametrize if not enough data are available, which  
248 is usually the case (*Anderson*, 2005).

249 Data assimilation is used to improve the simulations of ecological  
250 dynamical models and enhance their forecasting capabilities by con-  
251 straining their predictions with available observations *Edwards et al.*  
252 (2015). Such prediction capabilities are deployed in operational ex-  
253 pert systems, for example to study the impact of human activities  
254 on the ecosystem of the Gulf of Pagasitikos (*Korres et al.*, 2012).  
255 The deployment of a similar forecasting system in the Red Sea is

currently under development (*Triantafyllou et al.*, 2014). Hindcasting, the estimation of unobserved variables, is another application of assimilation scheme. *Ciavatta et al.* (2011) showed that they could improve the seasonal and annual hindcast of non-assimilated biogeochemical properties in the shelf area of Western English Channel. Data assimilation can also be used for reanalysis, to provide estimates of past years biogeochemical variables (*Fontana et al.*, 2013).

In the marine ecology modeling community, three assimilation schemes have been widely used: the Ensemble Kalman filters (EnKF), the Singular Evolutive Extended Kalman filter (SEEK), and its ensemble variant, the Singular Evolutive Interpolated Kalman filter (SEIK). The stochastic EnKF, a Monte-Carlo approximation of the Kalman Filter, has been used in *Ciavatta et al.* (2011, 2014). This scheme may however suffer from sampling errors when the ensemble size is smaller than the number of observations, as is usually the case when assimilating remotely-sensed data. SEEK is a reduced-rank variant of the Extended Kalman filter (EK). It was introduced to run efficiently when the state dimension is very large, as is the case in ocean applications. It is based on the projection of the error covariance onto a low dimensional space. SEEK has a long history in data assimilation for marine ecology models and is still extensively used in recent studies (*Fontana et al.*, 2013; *Korres et al.*, 2012; *Butenschön and Zavatarelli*, 2012). SEIK is an ensemble variant of the SEEK and a deterministic version of the EnKF that do not suffer from observations sampling errors, as it updates the filter and forecast exactly as in the Kalman filter, but requires a resampling step to generate a new ensemble for the next forecast step. SEIK has been used by (*Triantafyllou et al.*, 2013; *Korres et al.*, 2012). *Korres et al.* (2012) shows that SEIK and SEEK are both comparably robust methods for highly nonlinear systems. *Hoteit et al.* (2005) has shown that SEIK outperforms SEEK when using a high-resolution

non linear model.

Ecological models are challenging applications for state of the art data assimilation schemes (*Edwards et al.*, 2015). First, biogeochemical variables are usually positive concentration, whereas Kalman filters expect Gaussian variables, and log-transformation may fail at avoiding this issue (*Ciavatta et al.*, 2011). In an attempt to mitigate this problem, *Fontana et al.* (2013) has introduced Gaussian anamorphosis transformations. Second, ecological blooms are intermittent and highly nonlinear, conditions that are challenging for Kalman filter-based assimilation schemes (*Triantafyllou et al.*, 2013; *Korres et al.*, 2012). Third, SEIK, EnKF and SEEK project the error covariance onto some subspace, resulting in an underestimation of the estimation error. *Butenschön and Zavatarelli* (2012) studied different ways to propagate the error covariance in order to alleviate this issue. Finally, the model error statistics are required by Kalman-derived filters, but are difficult to estimate. *Triantafyllou et al.* (2013) proposed to use the  $H_\infty$  method with SEIK in order to alleviate this requirement.

Particle filters represent a class of data assimilation scheme that are not derived from the Kalman filter, and do not make any linearity or Gaussianity assumption (*Edwards et al.*, 2015). They have been studied in the case of 0D and 1D ecological models (*Edwards et al.*, 2015). Their application to 3D model is an active field of research (*Edwards et al.*, 2015).

## 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 observa-

tions, 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  $133 \times 188$  pixel map is available for a domain extending from longitudes between  $33^\circ\text{E}$  and  $44^\circ\text{E}$  and latitudes between  $12^\circ\text{N}$  and  $28^\circ\text{N}$ , 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 understood 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 as 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 biologically. Do we find comparable results as Acker, Raitsos, Weiker, etc. What can we say about the hypothesis that they made about the process that drive primary productivity in the Red Sea.

## 5. Conclusion

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

What did we learn about the Red Sea ecology?

Future works?

## 488 **Acknowledgment**

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

## 491 **6. Bibliography**

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

495 Acker, J., G. Leptoukh, S. Shen, T. Zhu, and S. Kempler (2008), Remotely-  
496 sensed chlorophyll a observations of the northern Red Sea indicate seasonal  
497 variability and influence of coastal reefs, *Journal of Marine Systems*, 69, 191–  
498 204, doi:10.1016/j.jmarsys.2005.12.006.

499 Alvera-Azcárate, A., A. Barth, J.-M. Beckers, and R. H. Weisberg (2007), Multi-  
500 variate reconstruction of missing data in sea surface temperature, chlorophyll,  
501 and wind satellite fields, *Journal of Geophysical Research*, 112, C03008, doi:  
502 10.1029/2006JC003660.

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

506 Anderson, T. R. (2005), Plankton functional type modelling: running be-  
507 fore we can walk?, *Journal of Plankton Research*, 27(11), 1073–1081, doi:  
508 10.1093/plankt/fbi076.

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

512 Beckers, J. M., and M. Rixen (2003), EOF calculations and data  
513 filling from incomplete oceanographic datasets, *Journal of Atmo-*

- spheric 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.

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

544 Fontana, C., P. Brasseur, and J.-M. Brankart (2013), Toward a multivariate  
545 reanalysis of the North Atlantic Ocean biogeochemistry during 1998-2006  
546 based on the assimilation of SeaWiFS chlorophyll data, *Ocean Science*, 9,  
547 37–56, doi:10.5194/Os-9-37-2013.

548 Hamill, T. M., and C. Snyder (2000), A hybrid ensemble Kalman filter-3D  
549 variational analysis scheme, *Monthly Weather Review*, 128, 2905–2919, doi:  
550 10.1175/1520-0493(2000)128;2905:Ahekfvl2.0.Co;2.

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

554 Korres, G., G. Triantafyllou, G. Petihakis, D. E. Raitsos, I. Hoteit, A. Pollani,  
555 S. Colella, and K. Tsiaras (2012), A data assimilation tool for the Pagasitikos  
556 Gulf ecosystem dynamics: Methods and benefits, *Journal of Marine Systems*,  
557 94, S102–S117, doi:10.1016/J.Jmarsys.2011.11.004.

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

561 McClain, C. R. (2009), A decade of satellite ocean color ob-  
562 servations, *Annual Review of Marine Science*, 1, 19–42, doi:  
563 10.1146/Annurev.Marine.010908.163650.

564 Miles, T. N., and R. He (2010), Temporal and spatial variability of Chl-a and  
565 SST on the South Atlantic Bight: Revisiting with cloud-free reconstructions  
566 of MODIS satellite imagery, *Continental Shelf Research*, 30, 1951–1962, doi:  
567 10.1016/j.csr.2010.08.016.

- 568 Pal, R., and A. K. Choudhury (2014), *An introduction to phytoplanktons : di-*  
569 *versity and ecology*, Springer, New Delhi, India.
- 570 Pettersson, L. H., and D. Pozdnyakov (2013), *Monitoring of harmful algal*  
571 *blooms*, Springer-Praxis books in geophysical sciences, Springer, Heidelberg,  
572 Germany.
- 573 Racault, M. F., D. E. Raitsos, M. L. Berumen, R. J. Brewin, T. Platt,  
574 S. Sathyendranath, and I. Hoteit (), Phytoplankton phenology indices in coral  
575 reef ecosystems: application to ocean-colour observations in the red sea.
- 576 Raitsos, D. E., I. Hoteit, P. K. Prihartato, T. Chronis, G. Triantafyllou, and  
577 Y. Abualnaja (2011), Abrupt warming of the Red Sea, *Geophysical Research*  
578 *Letters*, *38*, L14601, doi:10.1029/2011gl047984.
- 579 Raitsos, D. E., Y. Pradhan, R. J. W. Brewin, G. Stenchikov, and I. Hoteit  
580 (2013), Remote sensing the phytoplankton seasonal succession of the Red  
581 Sea, *PLoS One*, *8*(6), e64909, doi:10.1371/journal.pone.0064909.
- 582 Raitsos, D. E., X. Yi, T. Platt, M.-F. Racault, R. J. W. Brewin, Y. Pradhan,  
583 V. P. Papadopoulos, S. Sathyendranath, and I. Hoteit (2015), Monsoon os-  
584 cillations regulate fertility of the Red Sea, *Geophysical Research Letters*, *42*,  
585 doi:10.1002/2014gl062882, in press.
- 586 Richlen, M. L., S. L. Morton, E. A. Jamali, A. Rajan, and D. M. Ander-  
587 son (2010), The catastrophic 2008-2009 red tide in the Arabian gulf re-  
588 gion, with observations on the identification and phylogeny of the fish-killing  
589 dinoflagellate *Cochlodinium polykrikoides*, *Harmful Algae*, *9*, 163–172, doi:  
590 10.1016/J.Hal.2009.08.013.
- 591 Robinson, I. S. (2010), *Discovering the ocean from space : the unique applica-*  
592 *tions of satellite oceanography*, Springer-Praxis books on geophysical sciences,  
593 1st ed., Springer, Heidelberg, Germany.
- 594 Sirjacobs, D., A. Alvera-Azcárate, A. Barth, G. Lacroix, Y. Park, B. Nechad,  
595 K. Ruddick, and J.-M. Beckers (2011), Cloud filling of ocean colour and sea

596 surface temperature remote sensing products over the southern north sea by  
597 the data interpolating empirical orthogonal functions methodology, *Journal*  
598 *of Sea Research*, 65, 114–130, doi:10.1016/j.seares.2010.08.002.

599 Steinmetz, F., P. Y. Deschamps, and D. Ramon (2011), Atmospheric correction  
600 in presence of sun glint: application to MERIS, *Optics Express*, 19(10), 9783–  
601 9800, doi:10.1364/Oe.19.009783.

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

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

610 Triantafyllou, G., I. Hoteit, and G. Petihakis (2003), A singular evolutive inter-  
611 polated Kalman filter for efficient data assimilation in a 3-D complex physical-  
612 biogeochemical model of the Cretan Sea, *Journal of Marine Systems*, 40-41,  
613 213–231, doi:10.1016/S0924-7963(03)00019-8.

614 Triantafyllou, G., I. Hoteit, X. Luo, K. Tsiaras, and G. Petihakis (2013), As-  
615 sessing a robust ensemble-based Kalman filter for efficient ecosystem data  
616 assimilation of the Cretan Sea, *Journal of Marine Systems*, 125, 90–100, doi:  
617 10.1016/J.Jmarsys.2012.12.006.

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

622 Waite, J. N., and F. J. Mueter (2013), Spatial and temporal variability of  
623 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 thermocline 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.