

# 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

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.

Phytoplankton are unicellular, free-floating, photosynthetic algae that live in the upper layers of bodies of water (ocean, lakes, rivers or ponds). There exists a wide diversity of phytoplankton species. Up to date, about 5000 of them have been identified (*Tett and Barton*, 1995). Phytoplankton are also highly variable in sizes, ranging from 0.2  $\mu\text{m}$  for cyanobacteria to 200  $\mu\text{m}$  for the largest species of diatom (*Pal and Choudhury*, 2014). In the oceans, phytoplankton live in the surface layer where there is enough sunlight for photosynthesis.

Phytoplankton play a fundamental role for the ocean ecology. They are at the basis of the marine food web and trap most of

---

\*Corresponding author

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

the energy used by pelagic ecosystems (*Pal and Choudhury, 2014*). Zooplankton graze phytoplankton, which are in turn consumed by higher trophic levels. It has been estimated that nearly 98% of the ocean primary productivity comes from phytoplankton (*Pal and Choudhury, 2014*). Phytoplankton are also responsible for maintaining the dissolved oxygen level necessary for other species to survive. However, high phytoplankton concentration may also impact their environment by creating dead zones. When they die and sink, the bacteria that decompose them can consume all the available oxygen (*Pal and Choudhury, 2014*), and this may cause massive mortality in the fauna. Because of its short life cycle, phytoplankton respond very well to changes in its environment, making it a key parameter to monitor water quality (*Wu et al., 2014*).

Phytoplankton place at the bottom of the marine food chain makes it an important factor for fisheries. Productive fishing zones such as the regions in the Arabian seas, Californian coast, north-west African coast and Chilean coast, are due to the upwelling of cold nutrient rich water favourable to phytoplankton growth. As such, remotely-sensed chlorophyll data have been routinely used since the last decade to help fisheries predict the timing of phytoplankton blooms (*Robinson, 2010*). On the other hand, the El-Nino phenomenon creates less favourable conditions for phytoplankton in the Eastern Pacific, resulting in a dramatic reduction of fish catches of fisheries in the western coast of South America (*Robinson, 2010*). In contrast, in the Red Sea, the MEI (Multivariate ENSO Index) has been found to positively correlate with chlorophyll concentration, a fact that could be of importance for regional fisheries (*Raitsos et al., 2015*).

Phytoplankton also plays the role of a biological CO<sub>2</sub> pump and strongly impact the Earth climate. During photosynthesis, phytoplankton captures carbon and releases oxygen. A part of this organic

material stays in the food web, either transmitted to higher trophic level, or degraded by bacteria. Another part, sinks to the bottom of the ocean to form sediments. It is estimated that phytoplankton accounts for 48% of Earth carbon fixation (*Pal and Choudhury, 2014*).

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

78 threaten the fragile coral reef system (*Raitsos et al.*, 2011).

79 Because of the lack of in-situ data, the large-scale phytoplank-  
80 ton dynamics of the Red Sea remain largely unknown (*Raitsos et al.*,  
81 2013; *Triantafyllou et al.*, 2014). However, in recent studies, remotely-  
82 sensed data and computer simulations have been used to improve our  
83 knowledge of the ecology of this region. The Red Sea is deficient in  
84 major nutrients (*Weikert*, 1987), and the only significant input of  
85 water comes from the Gulf of Aden (*Yao and Hoteit*, 2015). This  
86 explains a general pattern of chlorophyll concentration increase from  
87 north to south (*Raitsos et al.*, 2013), with the lowest concentration  
88 found in the northern central Red Sea. This pattern can be seen in  
89 figure ???. The Red Sea also displays a distinct seasonality, with a  
90 peak in concentration during the winter. A weak summer peak is also  
91 observed around July, everywhere except in the northernmost region  
92 (*Raitsos et al.*, 2013). Despite this regularity, a strong interannual  
93 variability is observed, with blooms that can reach mesotrophic con-  
94 centration levels (*Raitsos et al.*, 2013). According to *Triantafyllou*  
95 *et al.* (2014), the variations in the Red Sea ecology are mainly driven  
96 by physical circulation. In the rest of this section, we explore some  
97 of the mechanisms that have been linked to the major features of  
98 chlorophyll concentration.

99 The exchange of water with the nutrient-rich Gulf of Aden is a  
100 major driving mechanism for the whole Red Sea (*Triantafyllou et al.*,  
101 2014). It is the most important source of nutrients. The maximum  
102 chlorophyll concentration observed in the southern Red Sea during  
103 winter is attributed to wind-driven water intrusion (*Raitsos et al.*,  
104 2013). In Summer, this exchange of water is believed to be the only  
105 significant source of nutrients for the whole Red Sea. The influence  
106 of the water intrusion weakens as the latitude increases, explaining  
107 the low concentration in the northern half of the Red Sea (*Raitsos*  
108 *et al.*, 2013).

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

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

Aerial depositions of dust could also be an important input of nutrients for the Red Sea, but it has been largely left unexplored (*Triantafyllou et al.*, 2014). *Raitsos et al.* (2013) noticed for example 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*

170 *et al.*), as shown in figure ??.

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

179 One solution to missing and bad values is to use a data filling  
180 algorithm, of which one of the most popular is the Date INterpolat-  
181 ing Empirical Orthogonal Functions (DINEOF). It is an EOF based  
182 data filling approach introduced by *Beckers and Rixen* (2003). In  
183 *Sirjacobs et al.* (2011), it has been employed to fill chlorophyll data  
184 with 70% of missing values. *Taylor et al.* (2013) has compared DI-  
185 NEOF with other EOF-based reconstruction algorithms, suggesting  
186 that the former is the best method for data filling. DINEOF has been  
187 employed in several other chlorophyll studies (*Miles and He*, 2010;  
188 *Waite and Mueter*, 2013). It has also been used for multivariate re-  
189 construction of SST fields using chlorophyll data in *Alvera-Azcárate*  
190 *et al.* (2007).

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

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

Model predictions of chlorophyll concentration also have practical applications. Phytoplankton blooms can be harmful to humans and marine life and are closely monitored in many regions of the world (*Pettersson and Pozdnyakov*, 2013). In the Red Sea, where tourism and aquaculture are developing, it is likely to become a concern too. Phytoplankton is also directly, and indirectly through zooplankton, the cause of microfouling that affects desalination plants. In 2008-2009, a red tide forced the shutdown of desalination plants along the Gulf of Oman and the Arabian Gulf (*Richlen et al.*, 2010).

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

Ecological deterministic models in use vary widely in diversity



depending on the number of state parameters and interactions represented. They can be as simple as the nutrient-phytoplankton-zooplankton (NPZ) model (*Anderson, 2005*) that only has three variables representing two trophic levels and nitrate, or as complex as the European regional seas ecosystem model (ERSEM) that has dozens of variables (*Baretta et al., 1995*). NPZ models are extensively used because of its simplicity and its capacity to model the large-scale features of marine ecosystems (*Anderson, 2005*). ERSEM has been use in many studies. It has recently been coupled to the MITgcm circulation model used to simulate the Red Sea ecology (*Triantafyllou et al., 2014*). However the complexity of these models makes them difficult to parametrize if not enough data are available, which is usually the case (*Anderson, 2005*).

Data assimilation is used to improve the simulations of ecological dynamical models and enhance their forecasting capabilities by constraining their predictions with available observations *Edwards et al. (2015)*. Such prediction capabilities are deployed in operational expert systems, for example to study the impact of human activities on the ecosystem of the Gulf of Pagasitikos (*Korres et al., 2012*). 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 er-

ror 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 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  $133 \times 188$  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.

### 381 2.3. Clustering

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

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

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

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

### 407 **3. Model and Assimilation**

#### 408 *3.1. 1D-ERSEM model*

409 We use a 1D coupled ERSEM model. The physical forcing comes from a  
410 3D circulation simulation of the Red Sea [Yao 2014]. The ecological models  
411 are initialized with the results of the 3D Red Sea ecology simulation [Triantfyllou2013].  
412

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

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

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

435 The ecological models are initialized with the results of a 3D  
436 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



467 *Tandeo et al.* (2014), except that the model is non linear. The  
468 scheme will be used to improve the estimates of the observation  
469 and model covariance errors.

## 470 **4. Results**

### 471 *4.1. Model evaluation*

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

### 475 *4.2. Analysis*

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

## 480 **5. Conclusion**

481 Are several 1D parallel 1D models a good alternative to 3D simulations?  
482 What did we learn about the Red Sea ecology?  
483 Future works?

## 484 **Acknowledgment**

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

## 487 **6. Bibliography**

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

- 491 Acker, J., G. Leptoukh, S. Shen, T. Zhu, and S. Kempler (2008), Remotely-  
 492 sensed chlorophyll a observations of the northern Red Sea indicate seasonal  
 493 variability and influence of coastal reefs, *Journal of Marine Systems*, *69*, 191–  
 494 204, doi:10.1016/j.jmarsys.2005.12.006.
- 495 Alvera-Azcárate, A., A. Barth, J.-M. Beckers, and R. H. Weisberg (2007), Multi-  
 496 variate reconstruction of missing data in sea surface temperature, chlorophyll,  
 497 and wind satellite fields, *Journal of Geophysical Research*, *112*, C03008, doi:  
 498 10.1029/2006JC003660.
- 499 Alvera-Azcárate, A., A. Barth, D. Sirjacobs, and J.-M. Beckers (2009), Enhanc-  
 500 ing temporal correlations in EOF expansions for the reconstruction of missing  
 501 data using DINEOF, *Ocean Science*, *5*, 475–485, doi:10.5194/os-5-475-2009.
- 502 Anderson, T. R. (2005), Plankton functional type modelling: running be-  
 503 fore we can walk?, *Journal of Plankton Research*, *27*(11), 1073–1081, doi:  
 504 10.1093/plankt/fbi076.
- 505 Baretta, J. W., W. Ebenhöh, and P. Ruardij (1995), The European regional seas  
 506 ecosystem model, a complex marine ecosystem model, *Netherlands Journal*  
 507 *of Sea Research*, *33*(3/4), 233–246, doi:10.1016/0077-7579(95)90047-0.
- 508 Beckers, J. M., and M. Rixen (2003), EOF calculations and data  
 509 filling from incomplete oceanographic datasets, *Journal of Atmo-*  
 510 *spheric and Oceanic Technology*, *20*(12), 1839–1856, doi:10.1175/1520-  
 511 0426(2003)020;1839:Ecadff;2.0.Co;2.
- 512 Beckers, J.-M., A. Barth, and A. Alvera-Azcárate (2006), DINEOF recon-  
 513 struction of clouded images including error maps - application to the Sea-  
 514 Surface Temperature around Corsican Island, *Ocean Science*, *2*, 183–199, doi:  
 515 10.5194/osd-3-735-2006.
- 516 Brewin, R. J. W., D. E. Raitsos, Y. Pradhan, and I. Hoteit (2013), Com-  
 517 parison of chlorophyll in the Red Sea derived from MODIS-Aqua and

518 in vivo fluorescence, *Remote Sensing of Environment*, 136, 218–224, doi:  
519 10.1016/j.rse.2013.04.018.

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

524 Campbell, J. W. (1995), The lognormal distribution as a model for bio-optical  
525 variability in the sea, *Journal of Geophysical Research*, 100(C7), 13,237–  
526 13,254, doi:10.1029/95JC00458.

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

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

534 Edwards, C. A., A. M. Moore, I. Hoteit, and B. D. Cornuelle (2015), Regional  
535 ocean data assimilation, *Annual Review of Marine Science*, 7, 21–42, doi:  
536 10.1146/annurev-marine-010814-015821.

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

540 Fontana, C., P. Brasseur, and J.-M. Brankart (2013), Toward a multivariate  
541 reanalysis of the North Atlantic Ocean biogeochemistry during 1998-2006  
542 based on the assimilation of SeaWiFS chlorophyll data, *Ocean Science*, 9,  
543 37–56, doi:10.5194/Os-9-37-2013.

- 544 Hamill, T. M., and C. Snyder (2000), A hybrid ensemble Kalman filter-3D  
545 variational analysis scheme, *Monthly Weather Review*, *128*, 2905–2919, doi:  
546 10.1175/1520-0493(2000)128;2905:Ahekfvj2.0.Co;2.
- 547 Hoteit, I., G. Korres, and G. Triantafyllou (2005), Comparison of extended and  
548 ensemble based Kalman filters with low and high resolution primitive equation  
549 ocean models, *Nonlinear Processes in Geophysics*, *12*, 755–765.
- 550 Korres, G., G. Triantafyllou, G. Petihakis, D. E. Raitsos, I. Hoteit, A. Pollani,  
551 S. Colella, and K. Tsiaras (2012), A data assimilation tool for the Pagasitikos  
552 Gulf ecosystem dynamics: Methods and benefits, *Journal of Marine Systems*,  
553 *94*, S102–S117, doi:10.1016/J.Jmarsys.2011.11.004.
- 554 Mann, K. H., and J. R. N. Lazier (2006), *Dynamics of marine ecosystems:*  
555 *Biological-Physical Interactions in the Oceans (3rd edition)*, Blackwell Pub-  
556 lishing, Malden, Maryland.
- 557 McClain, C. R. (2009), A decade of satellite ocean color ob-  
558 servations, *Annual Review of Marine Science*, *1*, 19–42, doi:  
559 10.1146/Annurev.Marine.010908.163650.
- 560 Miles, T. N., and R. He (2010), Temporal and spatial variability of Chl-a and  
561 SST on the South Atlantic Bight: Revisiting with cloud-free reconstructions  
562 of MODIS satellite imagery, *Continental Shelf Research*, *30*, 1951–1962, doi:  
563 10.1016/j.csr.2010.08.016.
- 564 Pal, R., and A. K. Choudhury (2014), *An introduction to phytoplanktons : di-*  
565 *versity and ecology*, Springer, New Delhi, India.
- 566 Pettersson, L. H., and D. Pozdnyakov (2013), *Monitoring of harmful algal*  
567 *blooms*, Springer-Praxis books in geophysical sciences, Springer, Heidelberg,  
568 Germany.
- 569 Racault, M. F., D. E. Raitsos, M. L. Berumen, R. J. Brewin, T. Platt,  
570 S. Sathyendranath, and I. Hoteit (), Phytoplankton phenology indices in coral  
571 reef ecosystems: application to ocean-colour observations in the red sea.

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

575 Raitsos, D. E., Y. Pradhan, R. J. W. Brewin, G. Stenchikov, and I. Hoteit  
576 (2013), Remote sensing the phytoplankton seasonal succession of the Red  
577 Sea, *PLoS One*, *8*(6), e64909, doi:10.1371/journal.pone.0064909.

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

582 Richlen, M. L., S. L. Morton, E. A. Jamali, A. Rajan, and D. M. Ander-  
583 son (2010), The catastrophic 2008-2009 red tide in the Arabian gulf re-  
584 gion, with observations on the identification and phylogeny of the fish-killing  
585 dinoflagellate *Cochlodinium polykrikoides*, *Harmful Algae*, *9*, 163–172, doi:  
586 10.1016/J.Hal.2009.08.013.

587 Robinson, I. S. (2010), *Discovering the ocean from space : the unique applica-*  
588 *tions of satellite oceanography*, Springer-Praxis books on geophysical sciences,  
589 1st ed., Springer, Heidelberg, Germany.

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

595 Steinmetz, F., P. Y. Deschamps, and D. Ramon (2011), Atmospheric correction  
596 in presence of sun glint: application to MERIS, *Optics Express*, *19*(10), 9783–  
597 9800, doi:10.1364/Oe.19.009783.

598 Tandeo, P., M. Pulido, and F. Lott (2014), Offline parameter estimation using  
599 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.
- Tett, P., and E. D. Barton (1995), Why are there about 5000 species of phytoplankton in the sea?, *Journal of Plankton Research*, *17*(8), 1693–1704, doi:10.1093/Plankt/17.8.1693.
- 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.

- 628 Willis, J. K. (2004), Interannual variability in upper ocean heat content, tem-  
629 perature, and thermosteric expansion on global scales, *Journal of Geophysical*  
630 *Research*, *109*, C12036, doi:10.1029/2003jc002260.
- 631 Wu, N., J. Huang, B. Schmalz, and N. Fohrer (2014), Modeling daily chloro-  
632 phyll a dynamics in a German lowland river using artificial neural net-  
633 works and multiple linear regression approaches, *Limnology*, *15*, 47–56, doi:  
634 10.1007/S10201-013-0412-1.
- 635 Yao, F., and I. Hoteit (2015), Thermocline regulated seasonal evolution of sur-  
636 face chlorophyll in the Gulf of Aden, *PLoS One*, in press.
- 637 Yao, F., I. Hoteit, L. J. Pratt, A. S. Bower, A. Köhl, G. Gopalakrishnan, and  
638 D. Rivas (2014a), Seasonal overturning circulation in the Red Sea: 2. winter  
639 circulation, *Journal of Geophysical Research-Oceans*, *119*, 2263–2289, doi:  
640 10.1002/2013jc009331.
- 641 Yao, F., I. Hoteit, L. J. Pratt, A. S. Bower, P. Zhai, A. Köhl, and G. Gopalakr-  
642 ishnan (2014b), Seasonal overturning circulation in the Red Sea: 1. model  
643 validation and summer circulation, *Journal of Geophysical Research-Oceans*,  
644 *119*, 2238–2262, doi:10.1002/2013jc009004.
- 645 Zhai, P., and A. Bower (2013), The response of the red sea to a strong wind  
646 jet near the Tokar Gap in summer, *Journal of Geophysical Research-Oceans*,  
647 *118*, 422–434, doi:10.1029/2012jc008444.
- 648 Zhan, P., A. C. Subramanian, F. C. Yao, and I. Hoteit (2014), Eddies in the  
649 Red Sea: A statistical and dynamical study, *Journal of Geophysical Research-*  
650 *Oceans*, *119*(6), 3909–3925, doi:10.1002/2013jc009563.