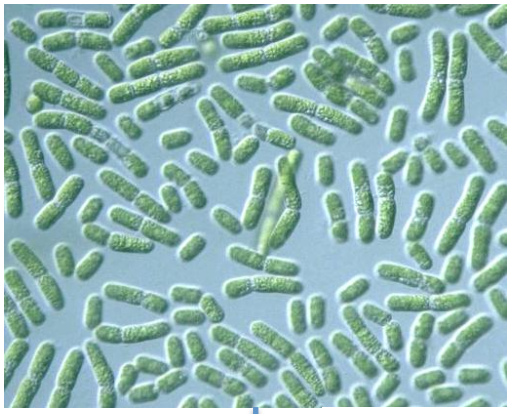
Data and Dynamics Driven Approaches for Modeling and Forecasting the Red Sea Chlorophyll

Thesis Proposal

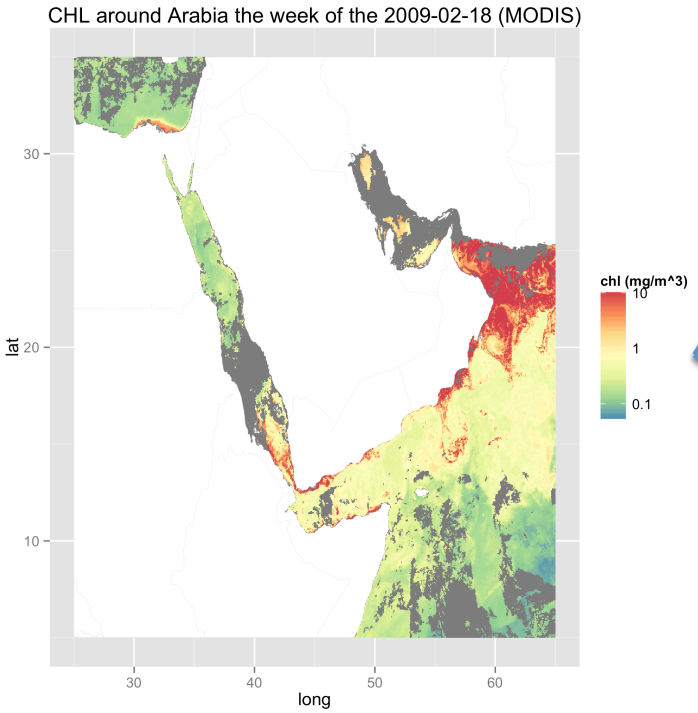
Denis Dreano

April 21st, 2015

Phytoplankton, Chlorophyll, the Red Sea and Applications



Small, unicellular, photosynthetic algae



Ocean color remote sensing

Desalination

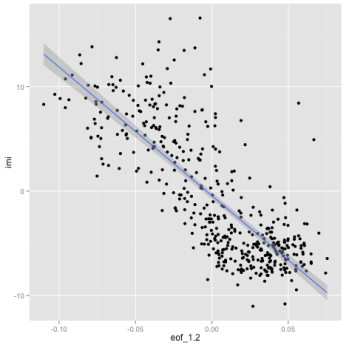
Marine Ecology

Climate Study

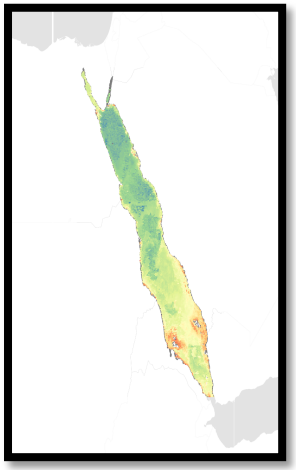
Thesis Objectives

Developing and comparing dynamical and data-driven methods for chlorophyll modeling and forecasting in the Red Sea

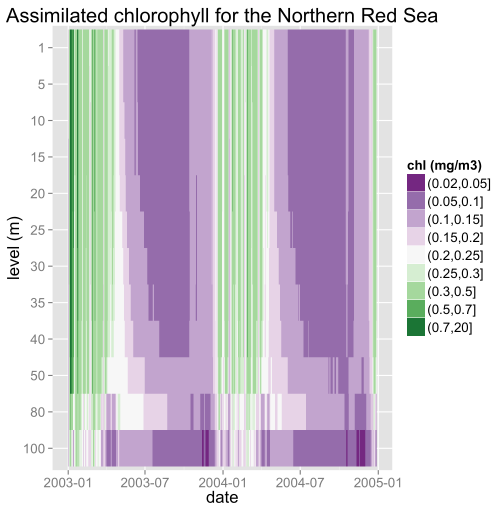
Machine Learning



Geostatistics



1D Assimilated Ecological models (ERSEM)



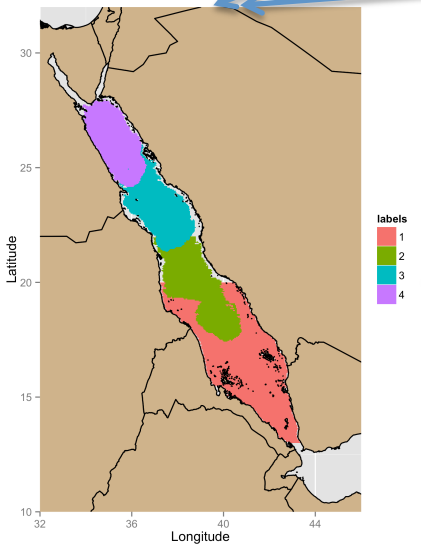
Chapter 1: Dataset building and Exploration

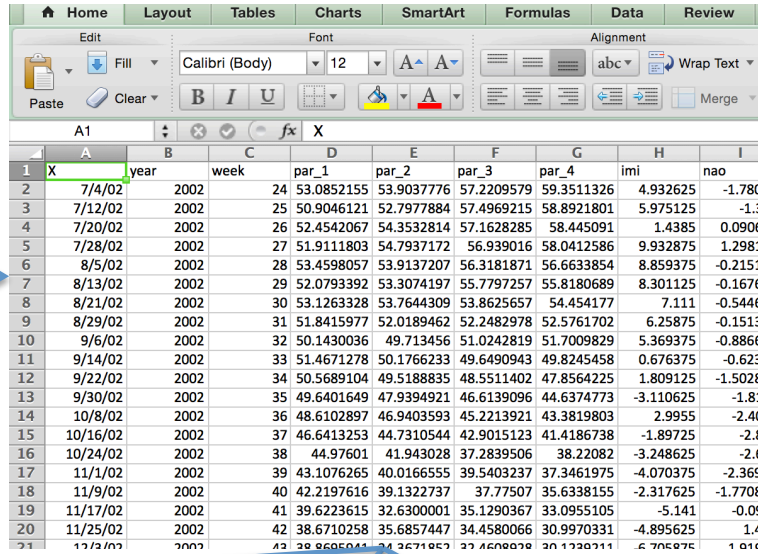
Raw Data

Rain SST SLA CHL PAR IMI AOT

NAO EAWR MEI SOI

Projection & Aggregation





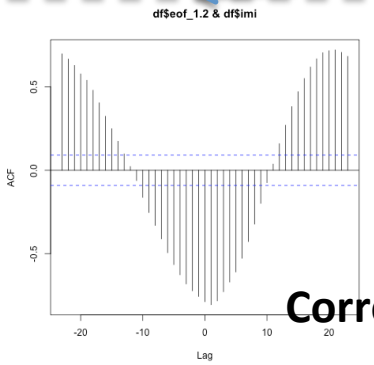
Home | Layout | Tables | Charts | SmartArt | Formulas | Data | Review

Edit: Fill, Clear, Paste — Font: Calibri (Body), 12 — Alignment: Wrap Text, Merge

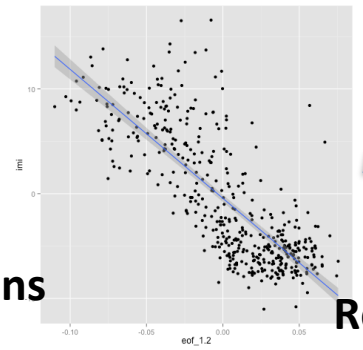
A1: X

	A	B	C	D	E	F	G	H	I
1	X	year	week	par_1	par_2	par_3	par_4	imi	nao
2	7/4/02	2002	24	53.0852155	53.9037776	57.2209579	59.3511326	4.932625	-1.78(
3	7/12/02	2002	25	50.9046121	52.7977884	57.4969215	58.8921801	5.975125	-1.3
4	7/20/02	2002	26	52.4542067	54.3532814	57.1628285	58.445091	1.4385	0.090(
5	7/28/02	2002	27	51.9111803	54.7937172	56.939016	58.0412586	9.932875	1.298:
6	8/5/02	2002	28	53.4598057	53.9137207	56.3181871	56.6633854	8.859375	-0.215:
7	8/13/02	2002	29	52.0793392	53.3074197	55.7797257	55.8180689	8.301125	-0.167(
8	8/21/02	2002	30	53.1263328	53.7644309	53.8625657	54.454177	7.111	-0.544(
9	8/29/02	2002	31	51.8415977	52.0189462	52.2482978	52.5761702	6.25875	-0.151:
10	9/6/02	2002	32	50.1430036	49.713456	51.0242819	51.7009829	5.369375	-0.886(
11	9/14/02	2002	33	51.4671278	50.1766233	49.6490943	49.8245458	0.676375	-0.62:
12	9/22/02	2002	34	50.5689104	49.5188835	48.5511402	47.8564225	1.809125	-1.502(
13	9/30/02	2002	35	49.6401649	47.9394921	46.6139096	44.6374773	-3.110625	-1.8:
14	10/8/02	2002	36	48.6102897	46.9403593	45.2213921	43.3819803	2.9955	-2.4(
15	10/16/02	2002	37	46.6413253	44.7310544	42.9015123	41.4186738	-1.89725	-2.:
16	10/24/02	2002	38	44.97601	41.943028	37.2839506	38.22082	-3.248625	-2.(
17	11/1/02	2002	39	43.1076265	40.0166555	39.5403237	37.3461975	-4.070375	-2.36(
18	11/9/02	2002	40	42.2197616	39.1322737	37.77507	35.6338155	-2.317625	-1.770(
19	11/17/02	2002	41	39.6223615	32.6300001	35.1290367	33.0955105	-5.141	-0.0:
20	11/25/02	2002	42	38.6710258	35.6857447	34.4580066	30.9970331	-4.895625	1.4

Analysis

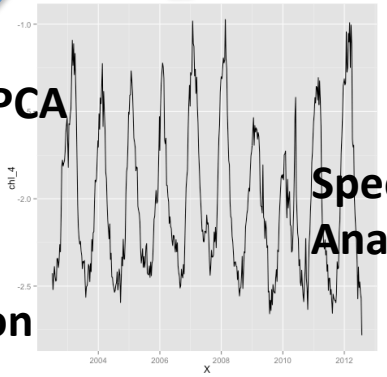


Correlations



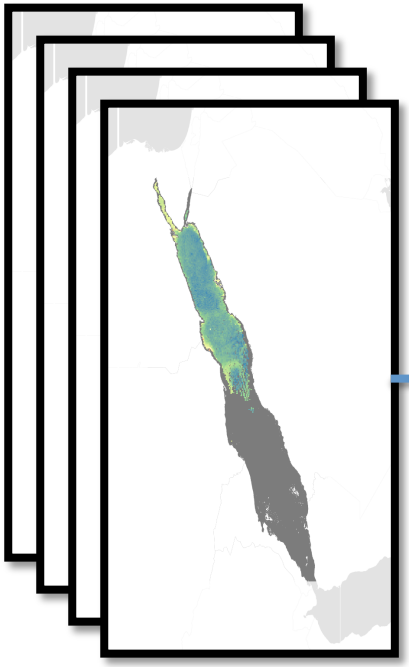
Linear Regression

PCA

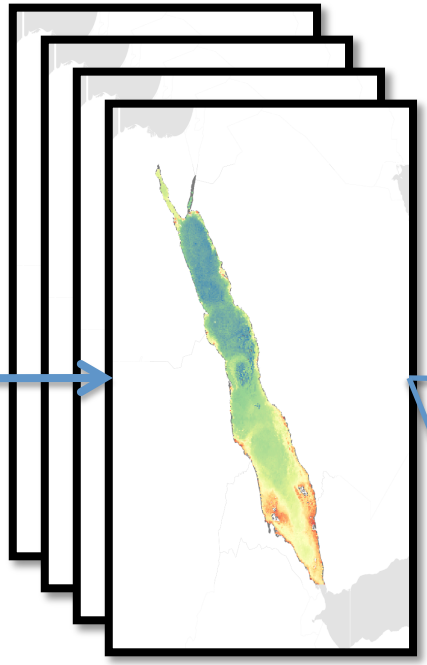


Spectral Analysis

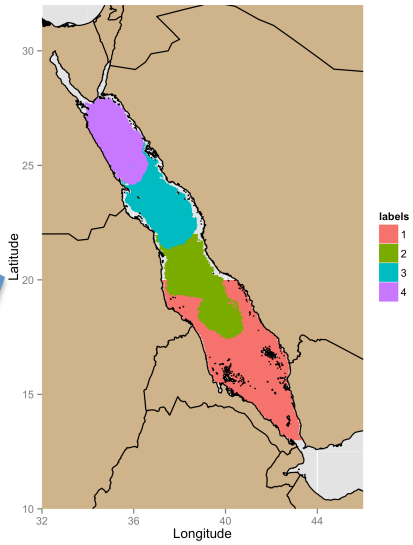
Chapter 1: Clustering and EOF computations



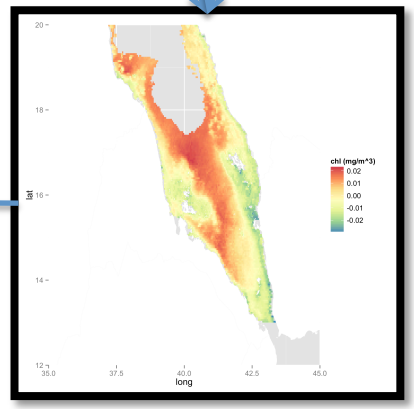
CCI Raw Data



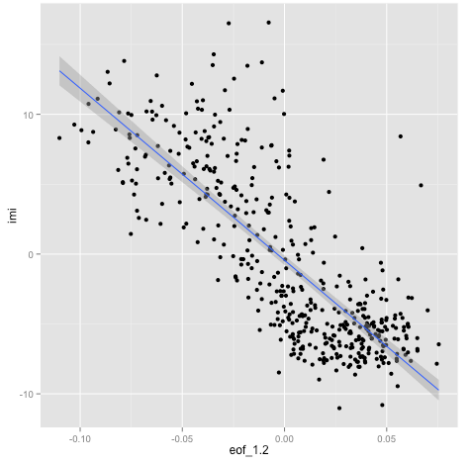
Climatological filling



Gaussian Mixture Model clustering



EOF Analysis



Analysis

Chapter 2: Statistical Models for Chlorophyll Forecasting

Goal: 8-days ahead forecasting

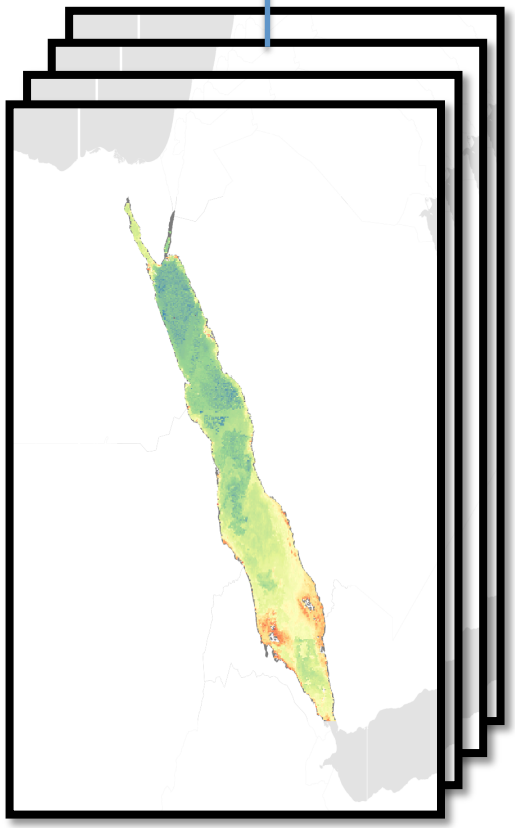
CHL(t) = f(CHL(t-1), SST(t-1), PAR(t-1), …)

Dataset Preparation:
- Variable Selection
- Cross-validation

Models:
- Linear Regression
- GAM
- Tree Regression
- SVM
- Neural Networks

Evaluation:
- Prediction Computation
- Model interpretation

Chapter 3: Global Geostatistical Model for Chlorophyll Forecasting



Modeled as a space-time stationary Gaussian Process

Estimate mean function

Estimate Covariance function

Forecasting using space-time Kriging:
CHL(t) = M.CHL(t-1) + e

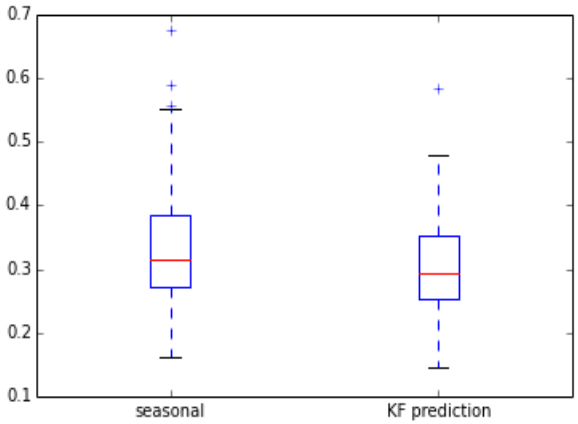
Reformulate in a state-space framework

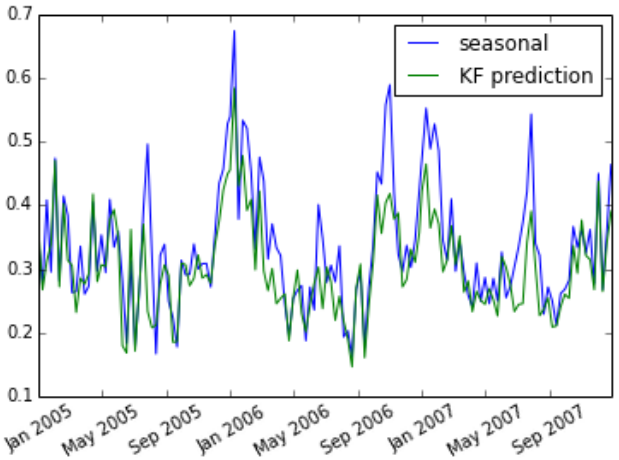
Kalman filtering

Chapter 3: Global Geostatistical Model for Chlorophyll Forecasting

$$C(\mathbf{h};u) = \begin{cases} \sigma^2 \exp(-c||\mathbf{h}||^{2\gamma}) & , \text{ if } u = 0, \\ \frac{\sigma^2 \tau}{a|u|^{2\alpha}+1} \exp(-\frac{c||\mathbf{h}||^{2\gamma}}{(a|u|^{2\alpha}+1)^{\beta\gamma}}) & , \text{ otherwise,} \end{cases}$$

Covariance model fitted with WLS

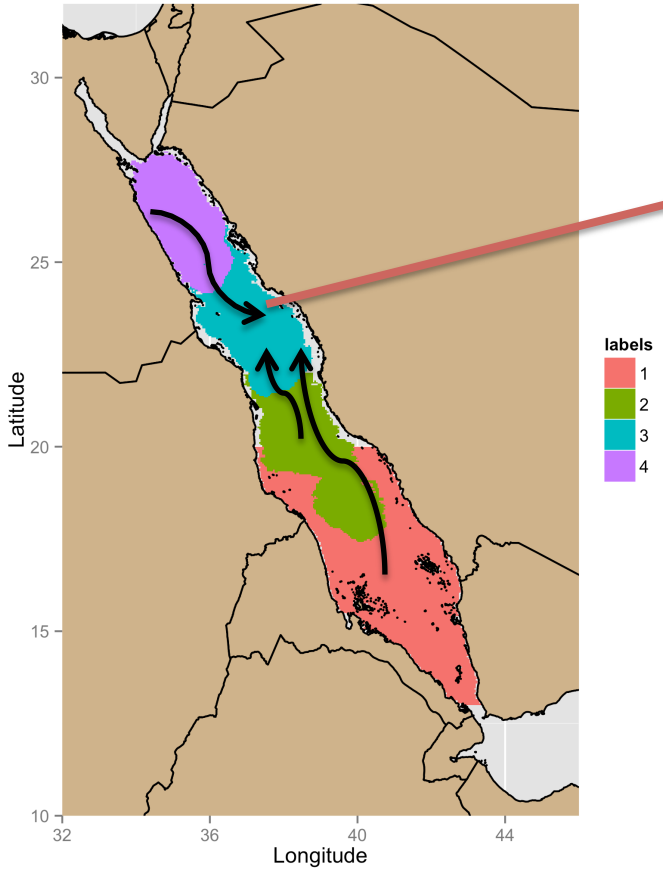




The model reduces prediction errors over the climatological prediction

Chapter 4: Local Geostatistical Models for Chlorophyll Forecasting

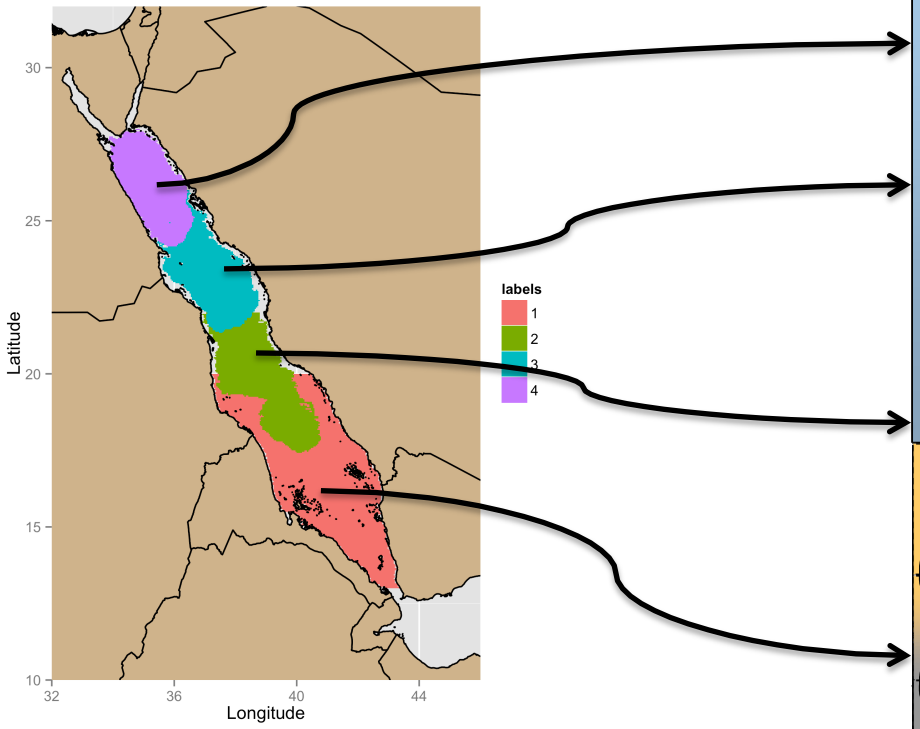
Goal: Improve stationary assumption

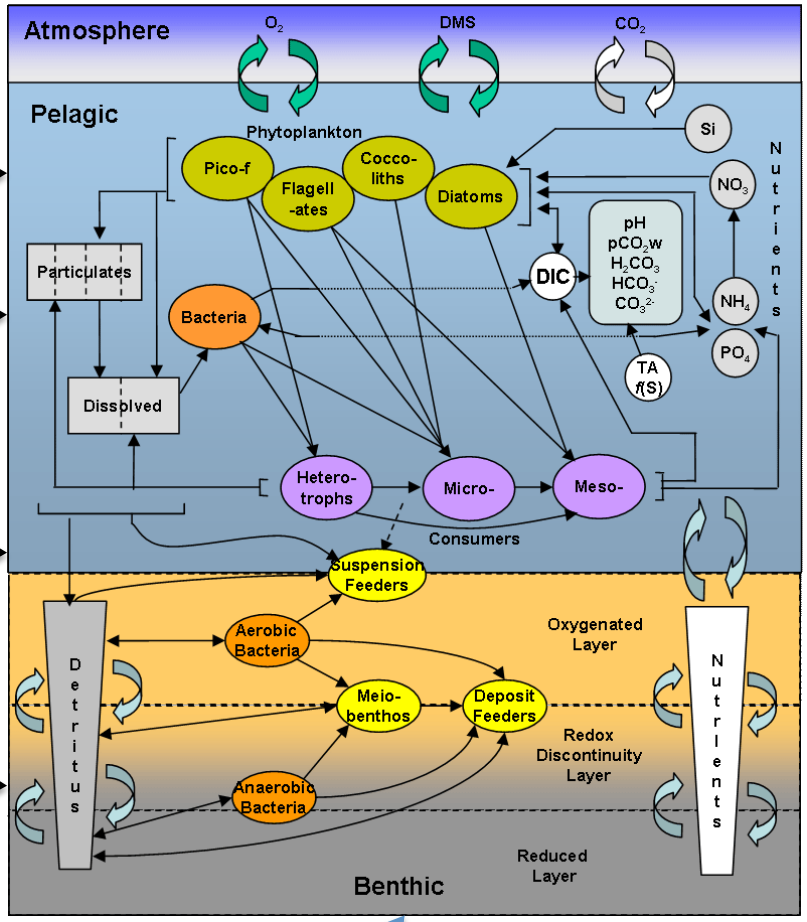


$CHL_3(s;t)$
Locally a stationary Gaussian process

Challenge: Incorporate other clusters information

Chapter 5: Assimilation of Regional 1D Ecological Models and Comparison with statistical Models





Developing 1D ERSEM ecological models for each cluster

Physical forcing from MITgcm

Chapter 5: Assimilation of Regional 1D Ecological Models and Comparison with statistical Models

$$\begin{cases} \mathbf{x}_k = f_k(\mathbf{x}_{k-1}|\Theta) + \eta_k \\ \mathbf{y}_k = h_k(\mathbf{x}_k|\Theta) + \varepsilon_k \end{cases}$$

Modeling ecological processes and chlorophyll observation as a State-Space system

Examples of ensemble filtering and smoothing: SEIK, SEEK, EnKF, hybrids, etc.

An Expectation-Maximization scheme for state/parameters estimations

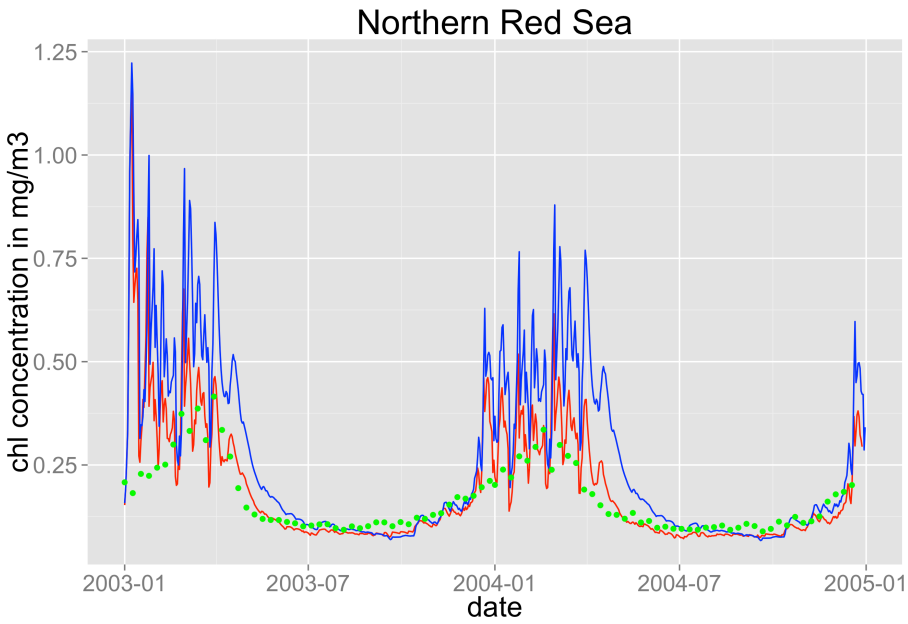
E-step $$G(\Theta) = E_{|\Theta=\Theta^{(r)}}\left[\log p(\mathbf{x}_{0:K}|\Theta, y_{1:K})\right]$$

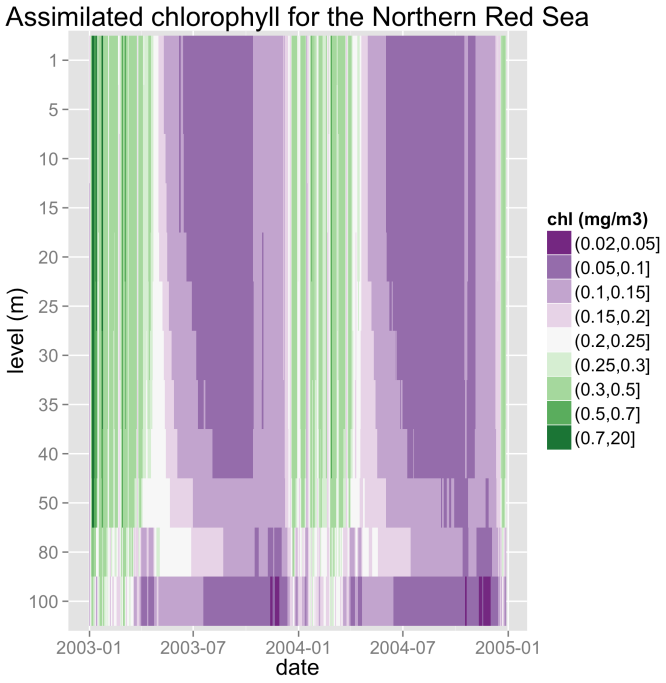
M-step $$\Theta^{(r+1)} = \arg\max_{\Theta} G(\Theta).$$

Ensemble Kalman smoother

Chapter 5: Assimilation of Regional 1D Ecological Models and Comparison with statistical Models

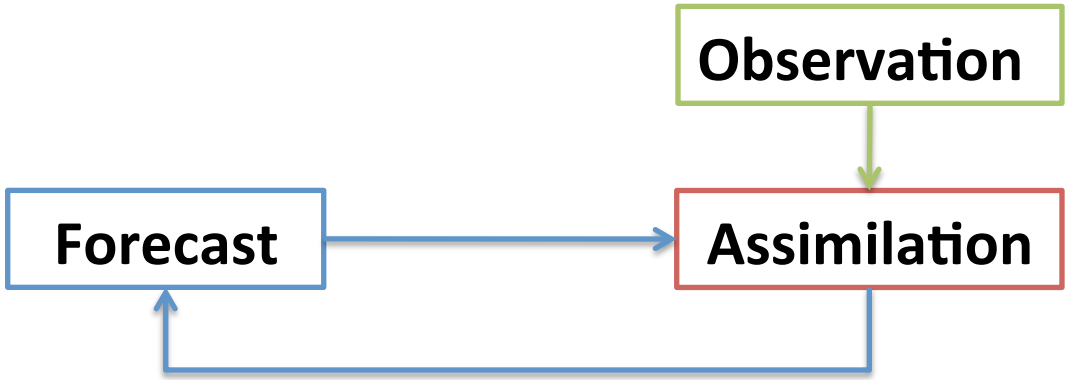
Chlorophyll Assimilation improves the model forecasts





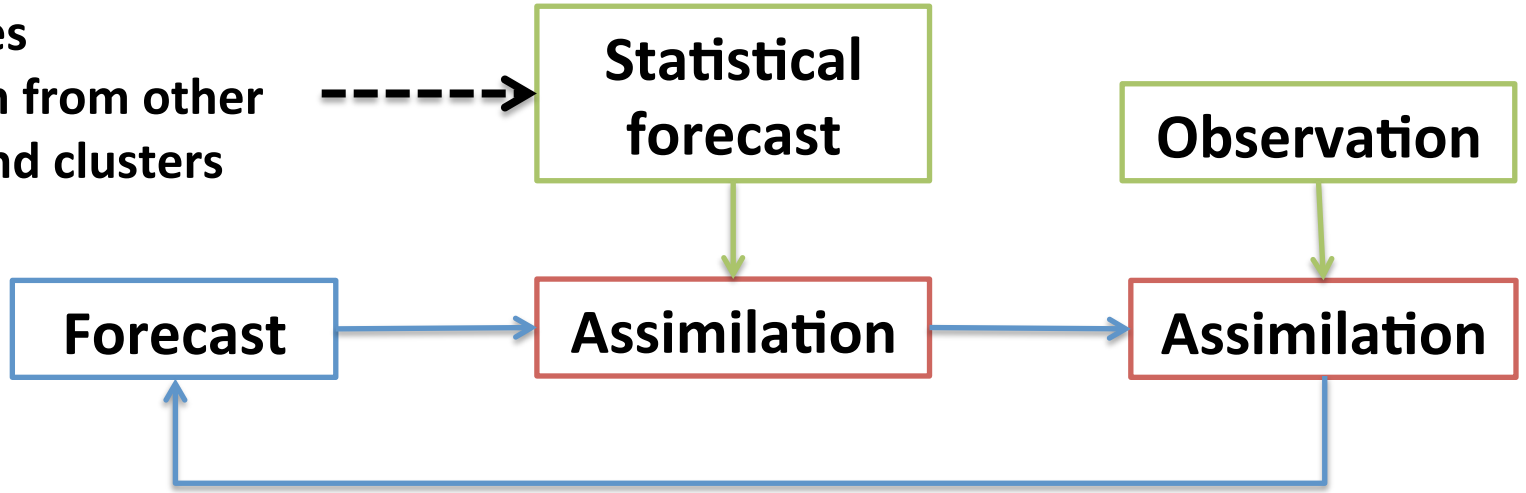
1D ecological models give information about subsurface processes

Chapter 6: Combining Statistical and Data Assimilative Predictive Models



Regular Assimilation Scheme

Incorporates information from other variables and clusters



Combined with statistical model

Chapter	Contribution
1) Dataset building and Exploration	• Exhaustive dataset • Improved knowledge of the Red Sea Ecology
2) Statistical Models for Chlorophyll Forecasting	• A benchmark for model comparison • Improved knowledge of the chlorophyll dynamics
3) Global Geostatistical Model for Chlorophyll Forecasting	• Methodology to employ geostatistical modeling in a filtering framework • Characterization of the space-time chlorophyll data
4) Local Geostatistical Models for Chlorophyll Forecasting	• A methodology to have different geostatistical models communicate
5) Assimilation of Regional 1D Ecological Models and Comparison with statistical Models	• An efficient method to estimate the parameters of a 1D ecological model • An efficient alternative to 3D ecological systems
6) Combining Statistical and Data Assimilative Predictive Models	• A method to improve the forecasts of assimilative ecological models • A methods to exchange information between 3D ecological models

Schedule and Conclusion

Timeline	Chapter	What has been done
	Chapter 3	**Published**
June 2015	Chapter 1	- Data Filling - Clusters - Dataset
August 2015	Chapter 2	- Dataset
November 2015	Chapter 5	- Models and assimilation implemented - Clusters
February 2016	Chapter 6	
May 2016	Chapter 4	- Clusters - Data Filling - Dataset