King Abdullah University of Science and Technology

KAUST

On the Development of an Integrated Data-Driven Modelling and Forecasting System for the Red Sea

Ibrahim Hoteit

Earth Sciences and Engineering

Applied Mathematics and Computational Sciences

King Abdullah University of Science and Technology (KAUST)

Annecy, June 7, 2017

QS WORLD UNIVERSITY RANKINGS®

Ranked #1 in
Citations Per Faculty

nature INDEX 2016
RISING STARS

Ranked #19 in high
quality research output

Quick facts

*as of Sept 2016

940 MS & PhD Students

150 Faculty

400 Post-Docs

310 Research Scientists & Staff

6,700 Community Members

1,200 Alumni

1,460 School Children

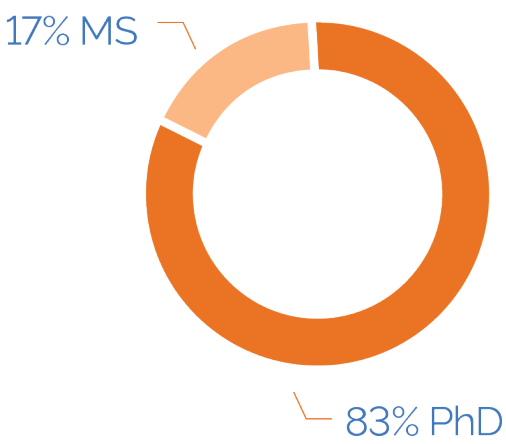
2,200 Workforce

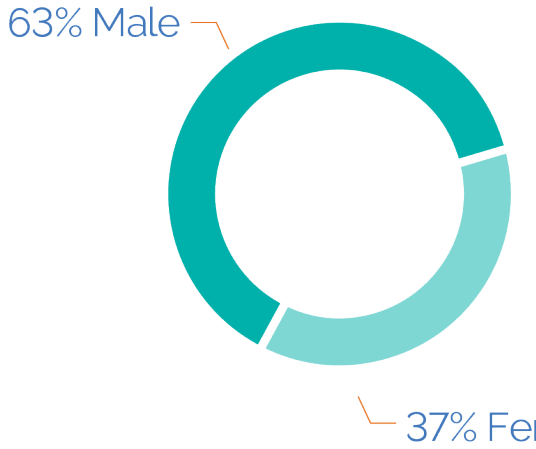
Nationalities

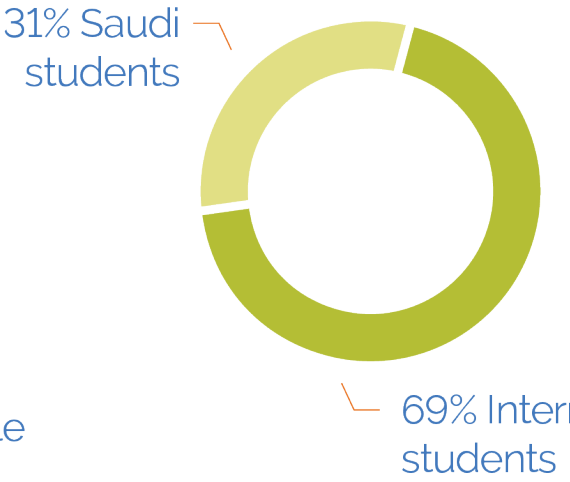
100 Community

80 Workforce

Our students







Our facilities

The Red Sea

King Abdullah University of Science and Technology KAUST

- ~2000 km long & ~200 km wide
- Semi-enclosed basin, connect with Indian Ocean through Bab-El-Mandeb
- Two small gulfs: Aqaba and Suez, important for deep water formation
- One of the warmest and saltiest water masses in the world

  Exemplifies conditions that are predicted to occur in other Marine Ecosystems few decades from now

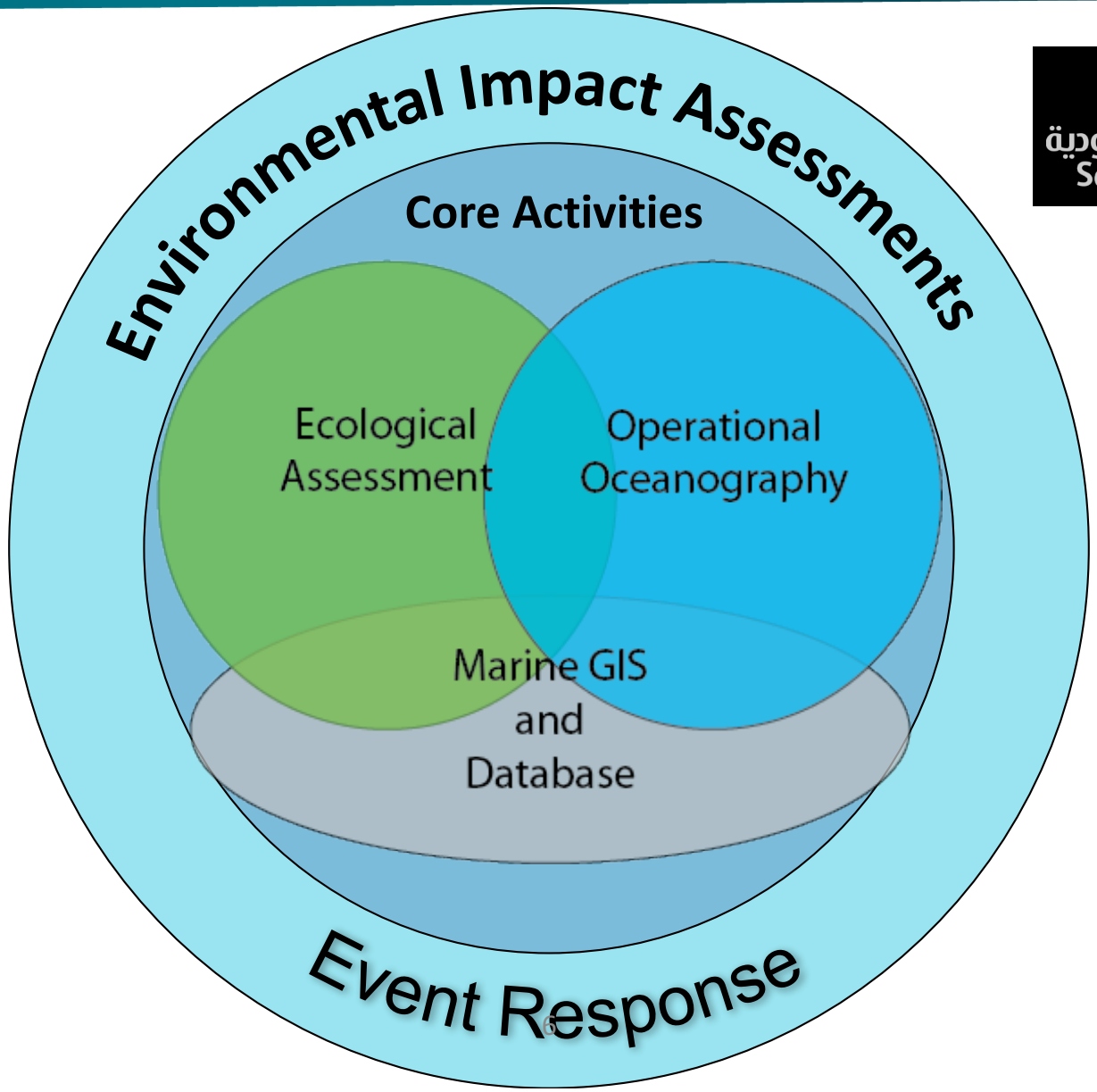
Why the Red Sea?

King Abdullah University of Science and Technology KAUST

- It is our backyard; and we know little about it
- 2nd longest and 3rd largest coral reef system
- Commercial highway
- Source of food (fishery and aquaculture), water (desalination), and energy
- Governmental and industrial developments
- Vulnerable to climate change
- *ARAMCO is exploring it …*

SAKMEO

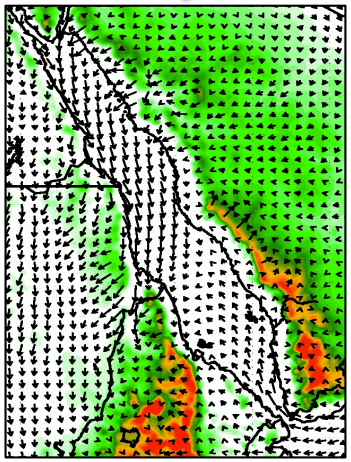
King Abdullah University of Science and Technology
KAUST



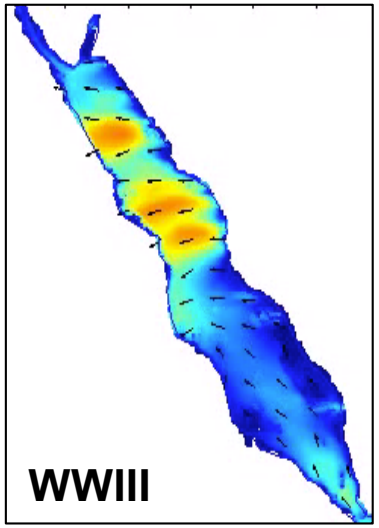
ارامكو السعودية
Saudi Aramco

The Data-Driven Red Sea Modelling System

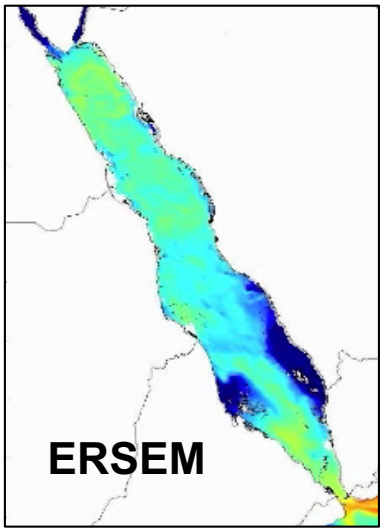
Atmosphere



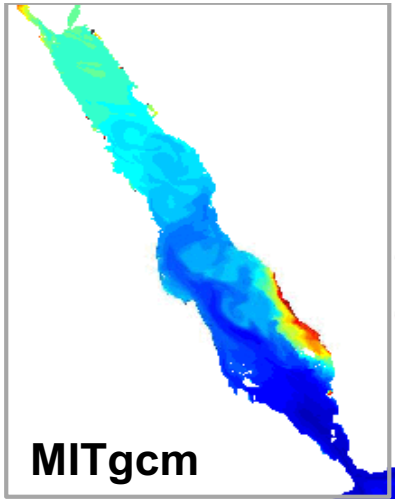
Waves

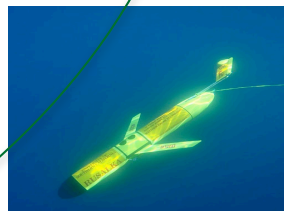


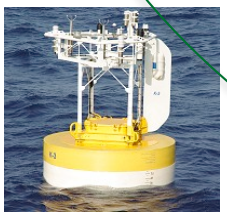
Ecosystem



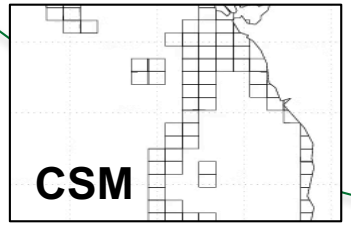
Circulation



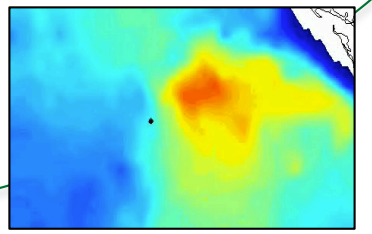




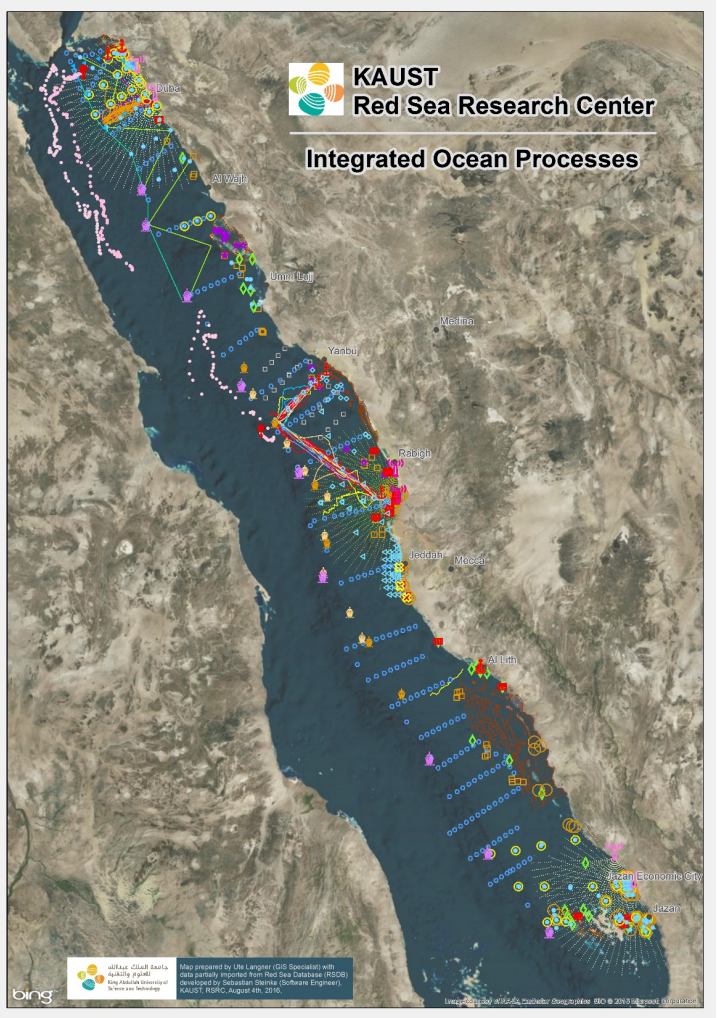
Connectivity

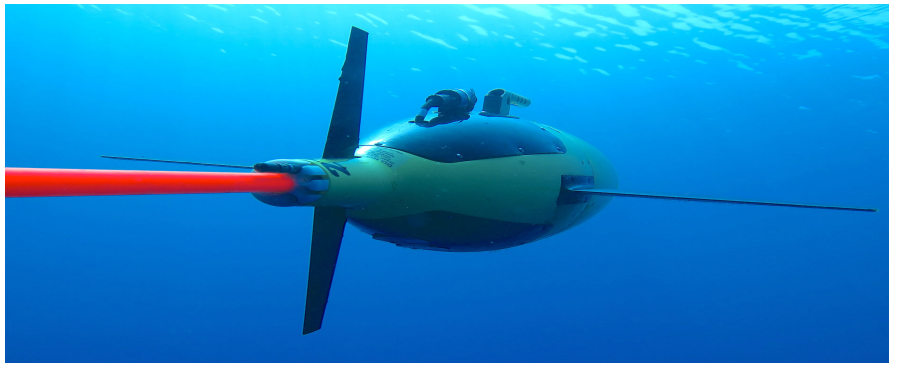


Oil Spill



Observations Efforts in the Red Sea





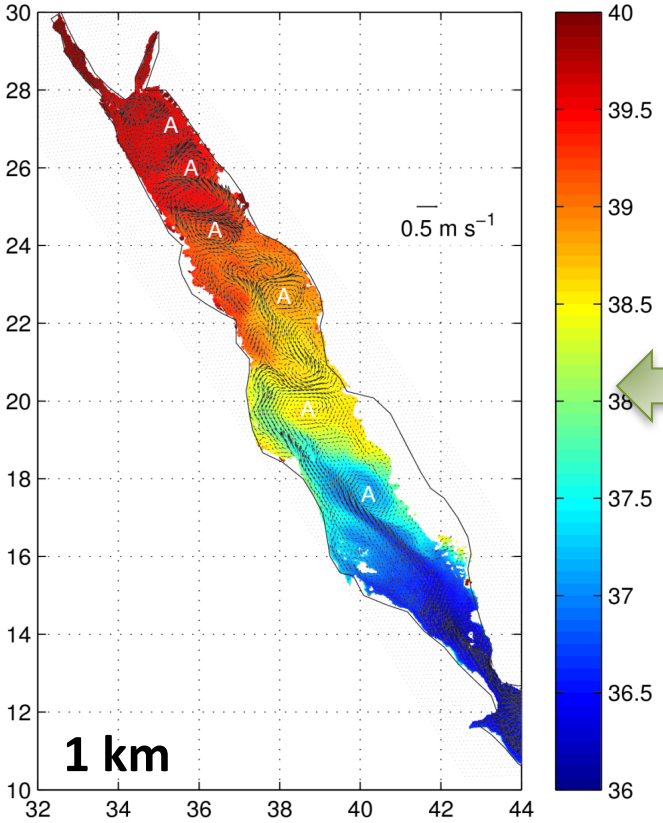
Reconstructing Red Sea Circulation

King Abdullah University of Science and Technology

KAUST

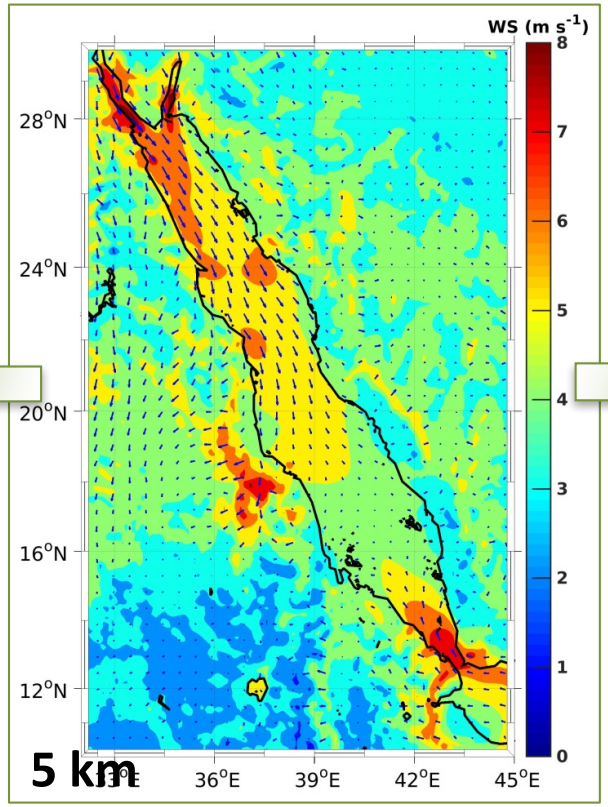
Generate long term Red Sea reanalyses using models and observations

OCEAN



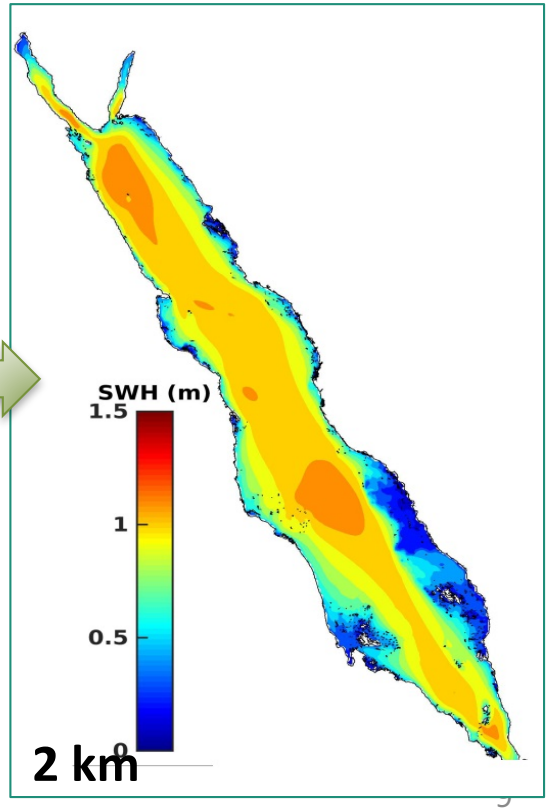
Kartadikaria et al. (2017)

ATMOSPHERE



Yesubabu al. (2016)

WAVES



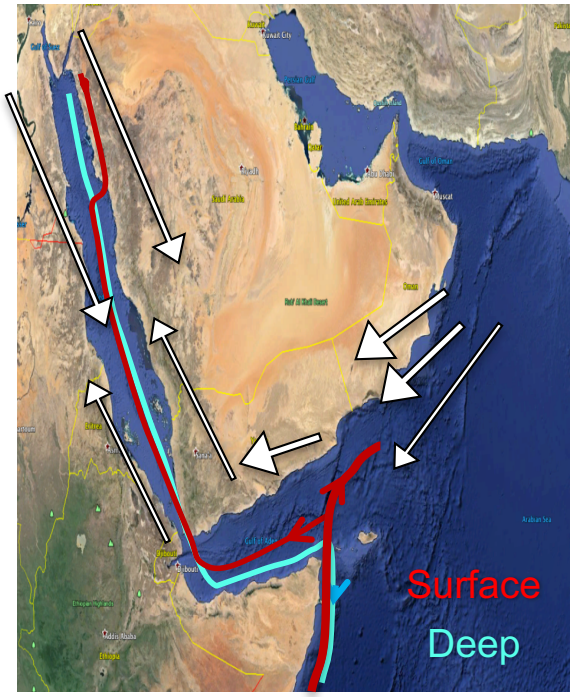
Langodan et al. (2017)

Applications Supported by the System

King Abdullah University of Science and Technology

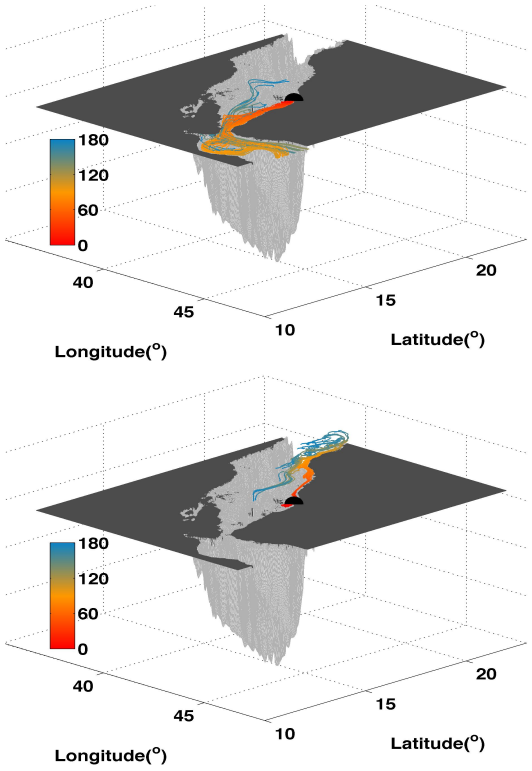
KAUST

Science



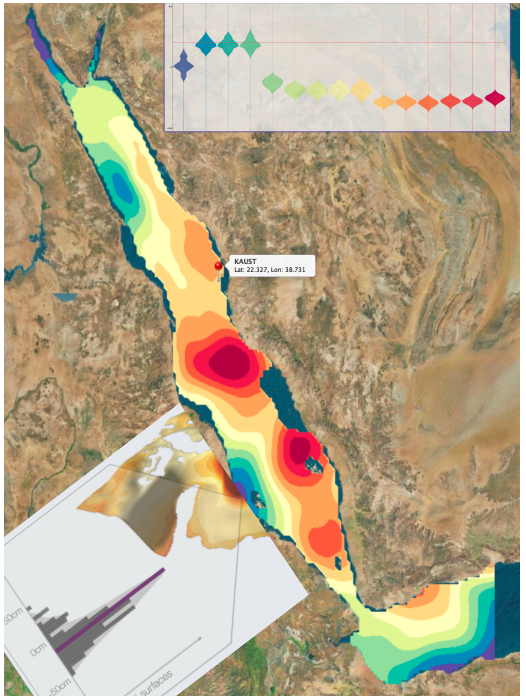
Understand variability and climate impact

Engineering



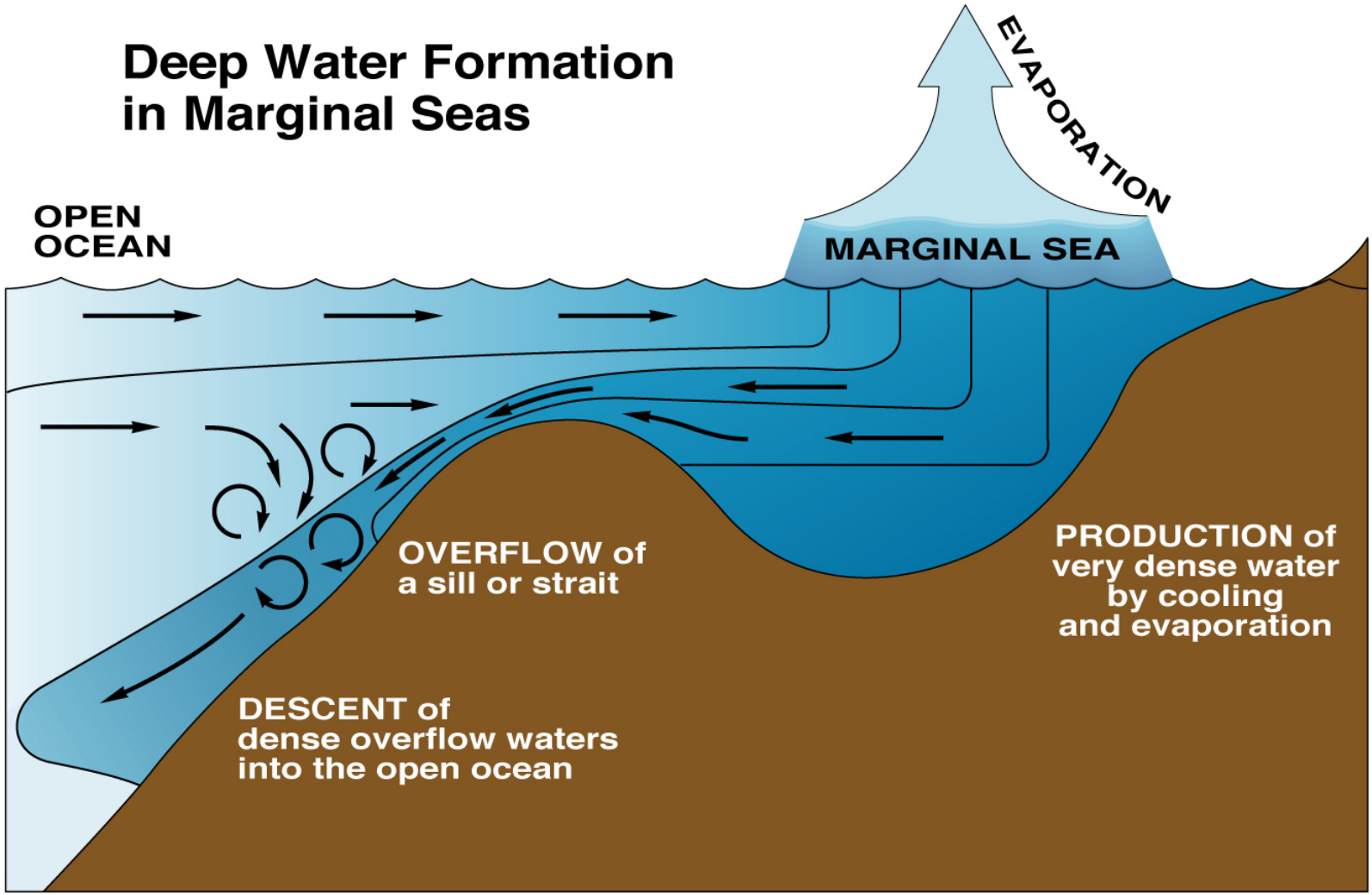
Support governmental and industrial activities

Technology



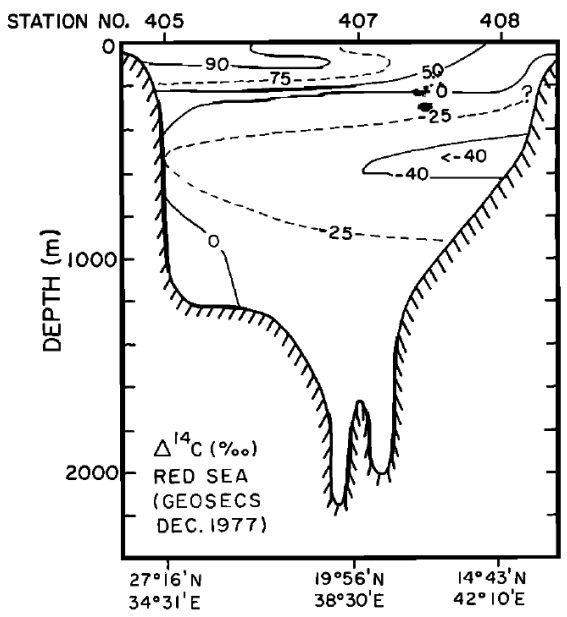
New paradigms for Assimilation, UQ and information science

Red Sea a Miniature of the Atlantic MOC

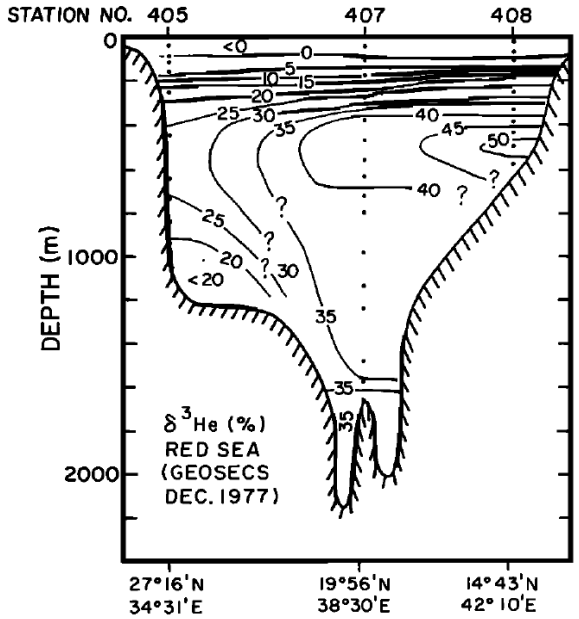


Overturning Cells as Depicted by Geochemical Tracers

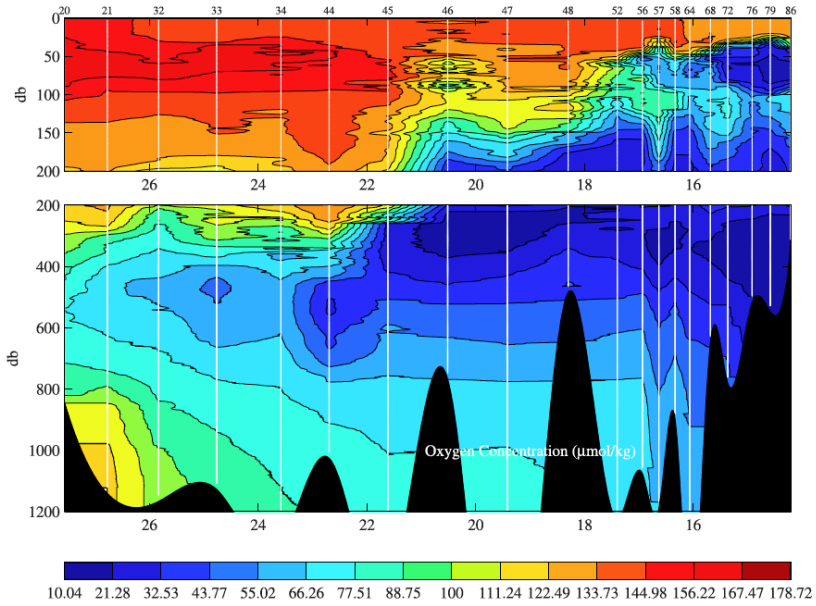
^{14}C



^{3}He



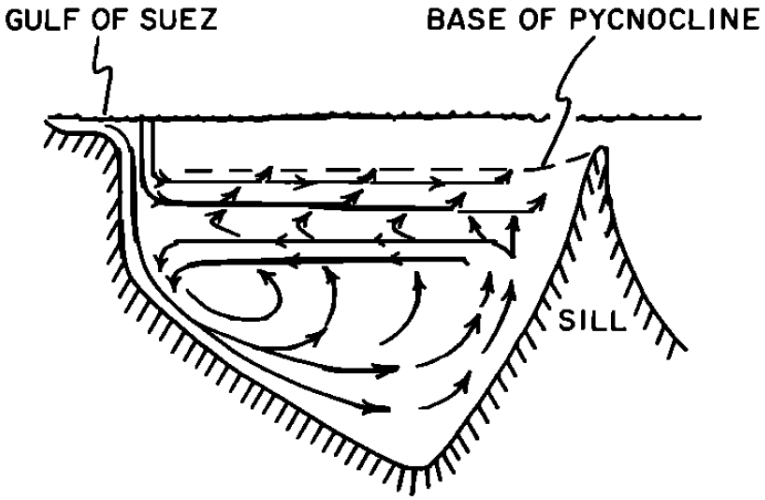
Dissolved Oxygen



Cember (1988)

Sofianos and Johns (2007)

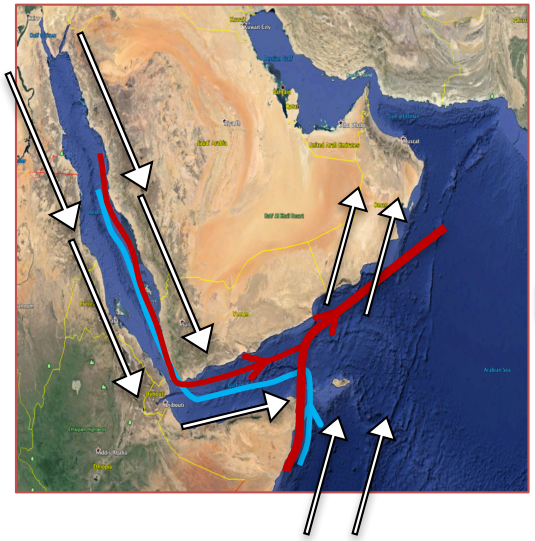
Schematics of the Circulation



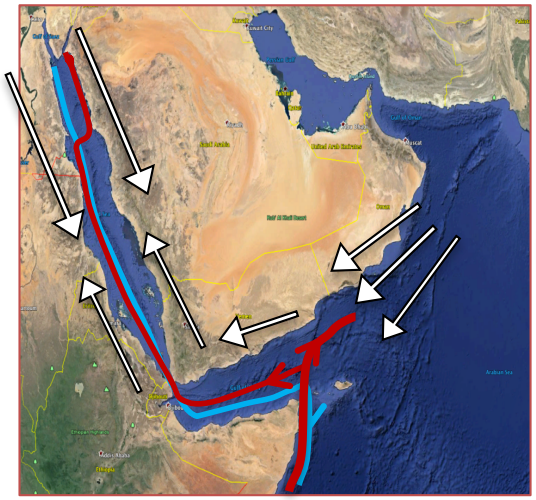
- Deep water formed in the north, possibly from Gulf of Suez, Gulf of Aqaba, or the northern Red Sea
- Two vertical cell structure: shallow and deep overturning cells
- A southward bottom flow and upwelling in the south
- A northward intermediate return flow at 500 m depth

IMPACT ON RED SEA PRODUCTIVITY

Summer Monsoon

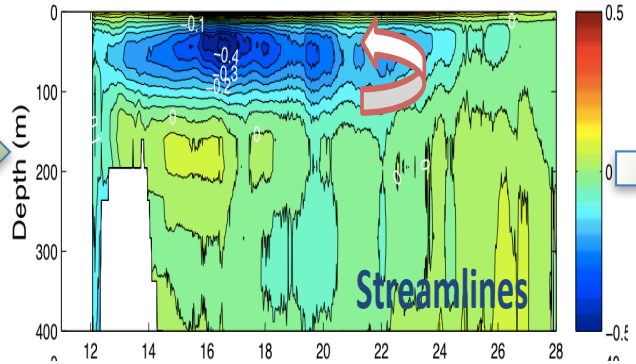


Winter Monsoon



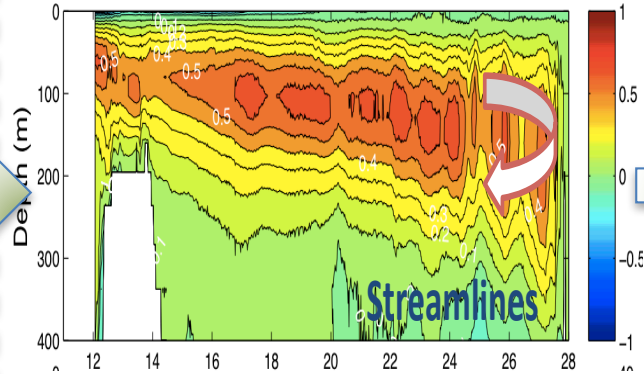
Seasonally Reversing Overturning Circulation

Summer



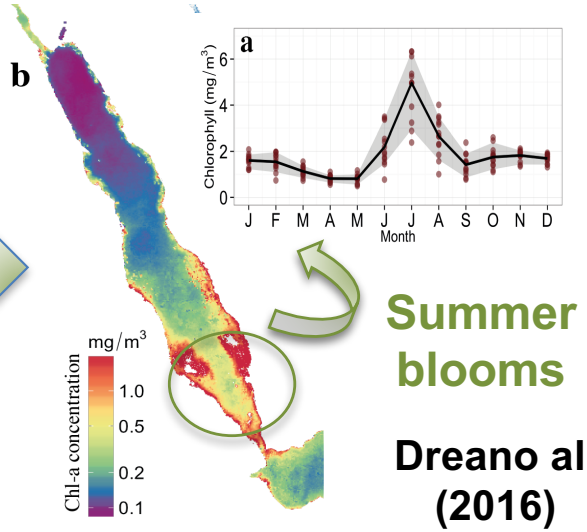
Yao et al. (2014a)

Winter



Yao et al. (2014b)

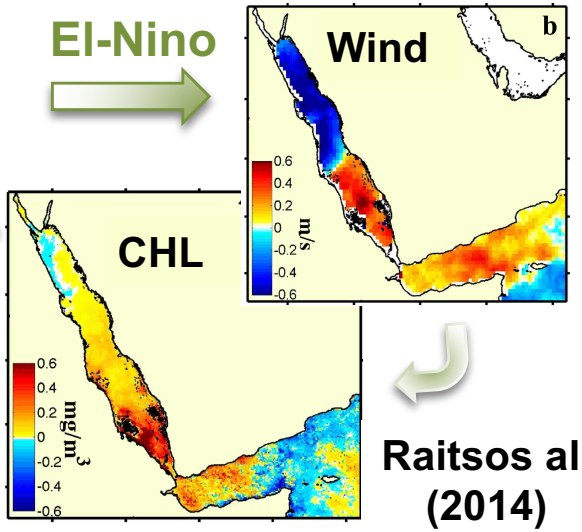
Impact on Ecosystem Productivity



Summer blooms

Dreano al. (2016)

El-Nino



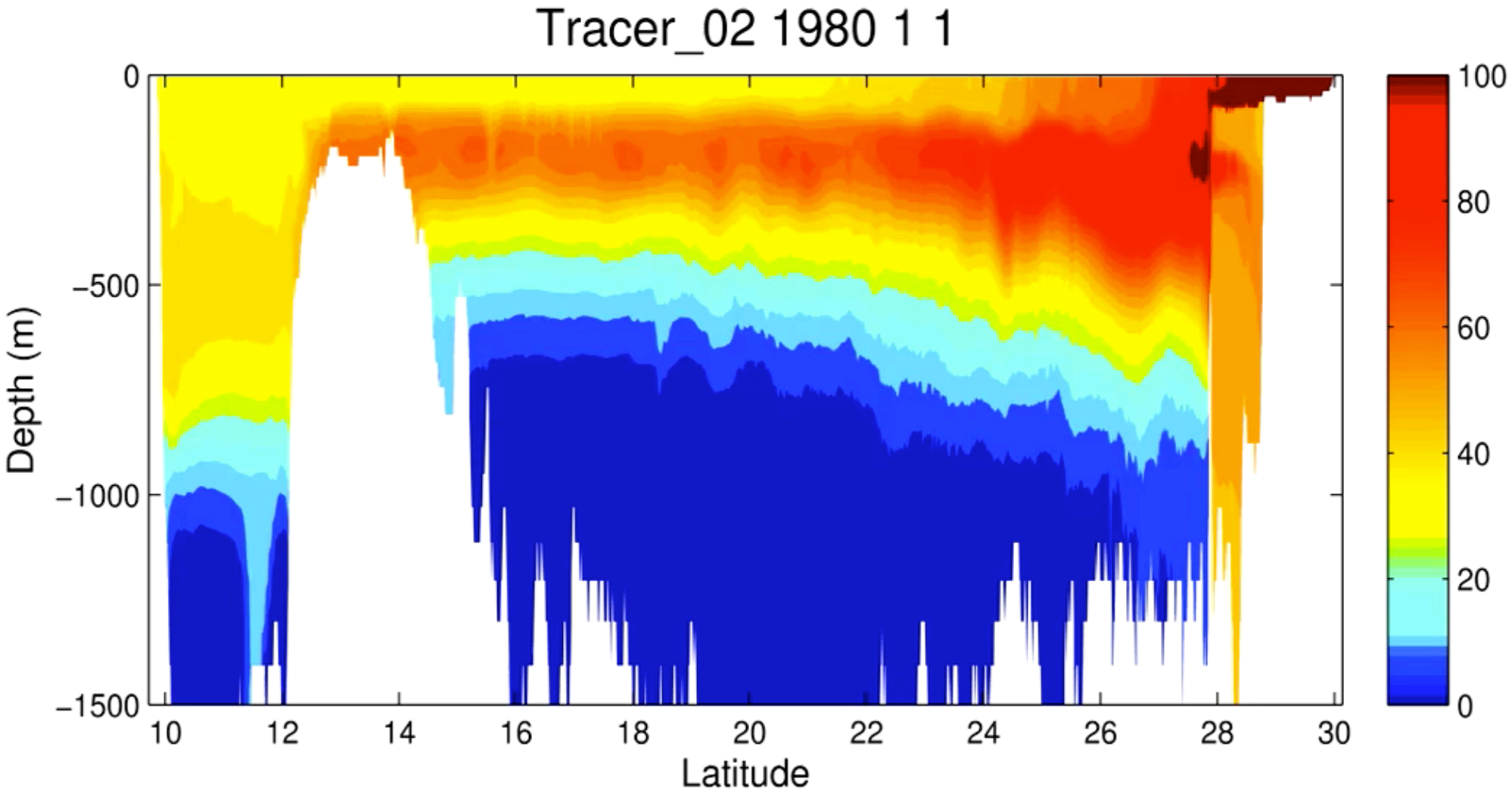
Raitsos al. (2014)

Deep Water Formation and Circulation

King Abdullah University of Science and Technology

KAUST

Tracers Released in the Gulf of Suez



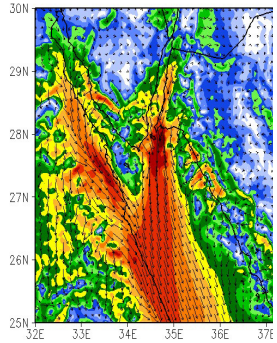
Currently working on

King Abdullah University of Science and Technology

KAUST

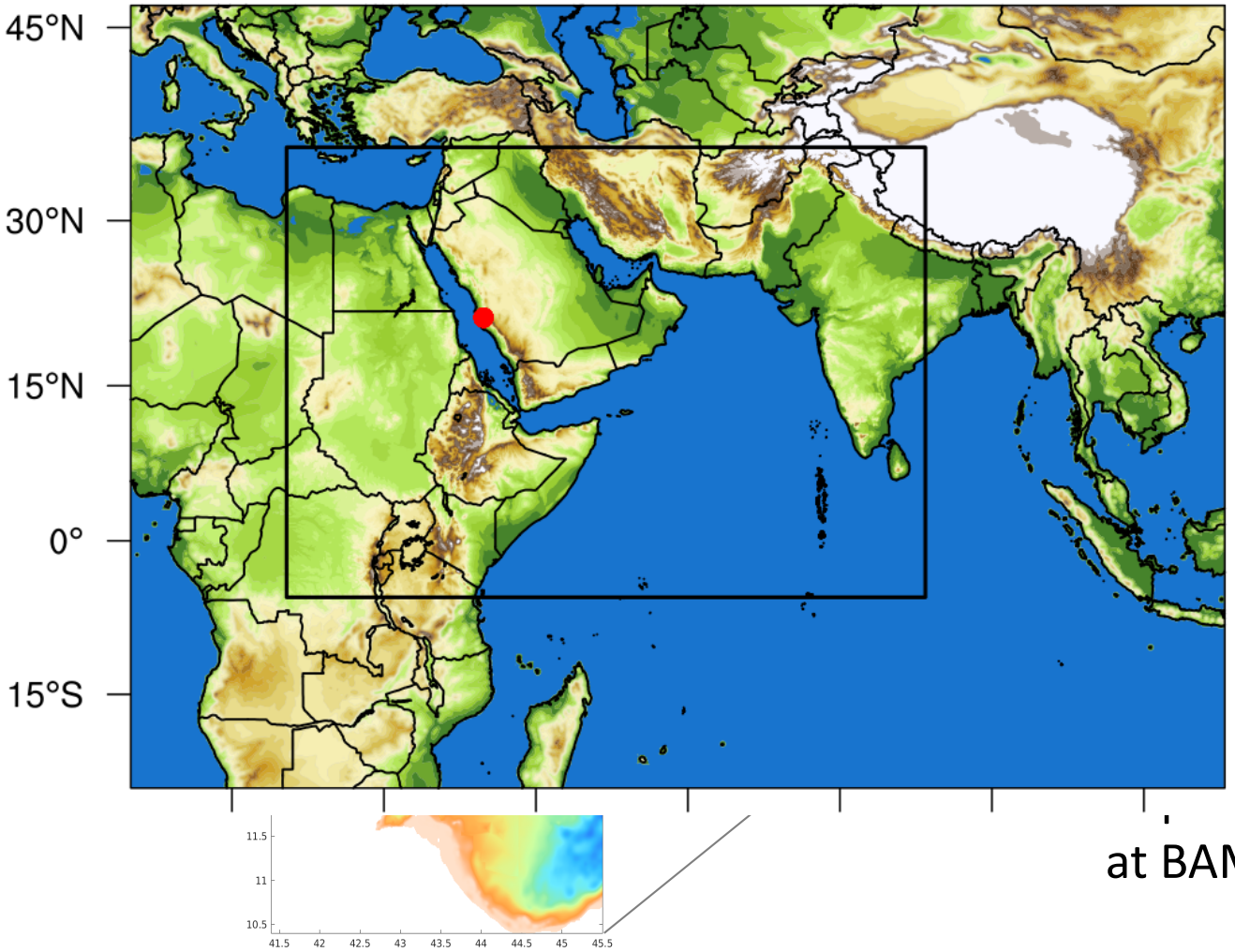
Wind

WRF (5kr



South (20

- Variabili controll of excha
- Hydraul
- ISWs



Scripps

)m)

processes of ents

d shallow eractions

al impact

n exchanges

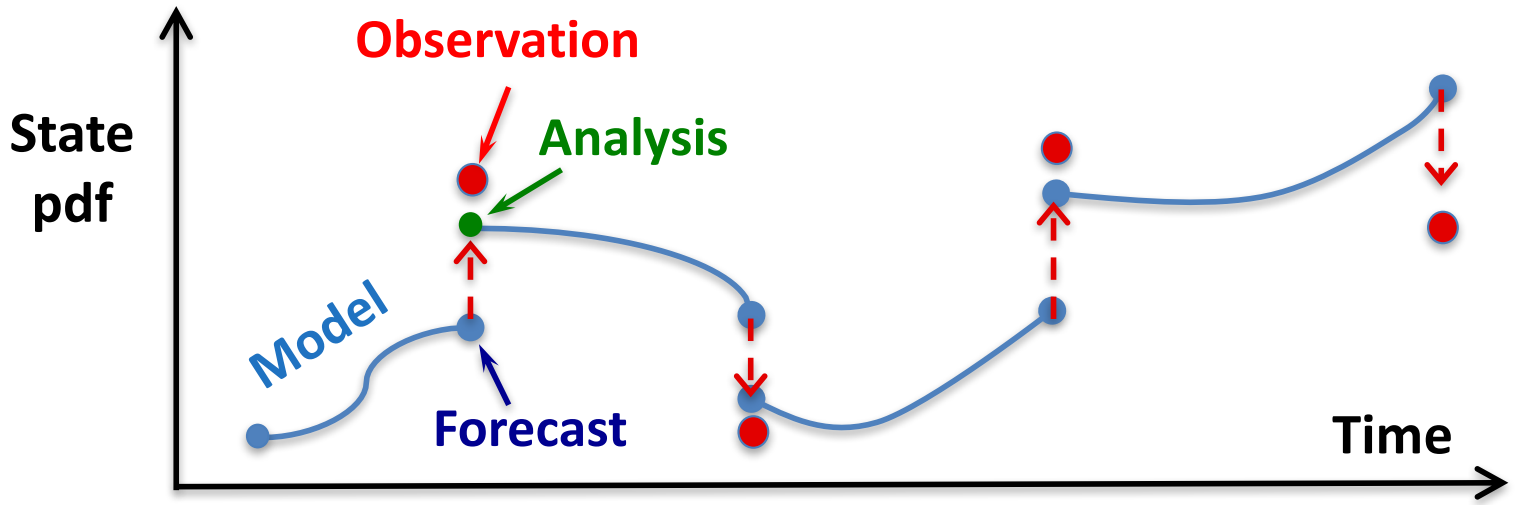
at BAM

Data Assimilation: Bayesian Formulation

- State-space modeling

$$x_k = M_k(x_{k-1}) + \eta_k$$
$$y_k = H_k(x_x) + \varepsilon_k$$

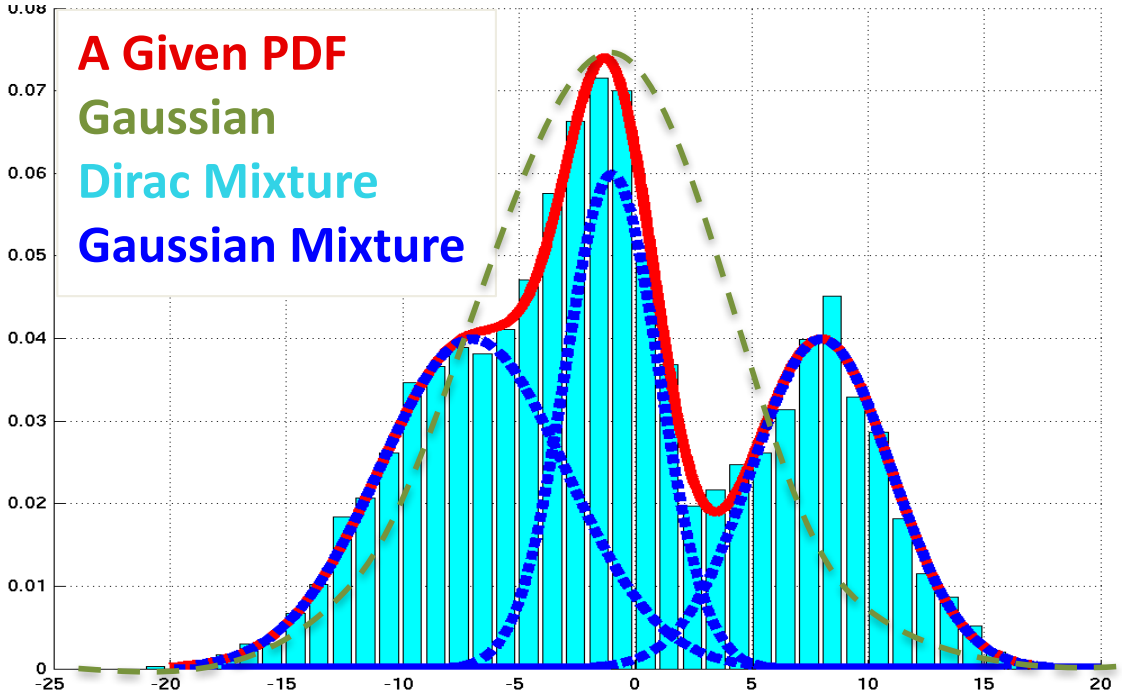
- Compute probability distribution (*pdf*) of the state given available observations up to the estimation time $p_k(x \mid y_{1:k})$



Numerical Implementation

King Abdullah University of Science and Technology KAUST

- *Some sort of discretization/parameterization of the pdfs are required; the most common are*

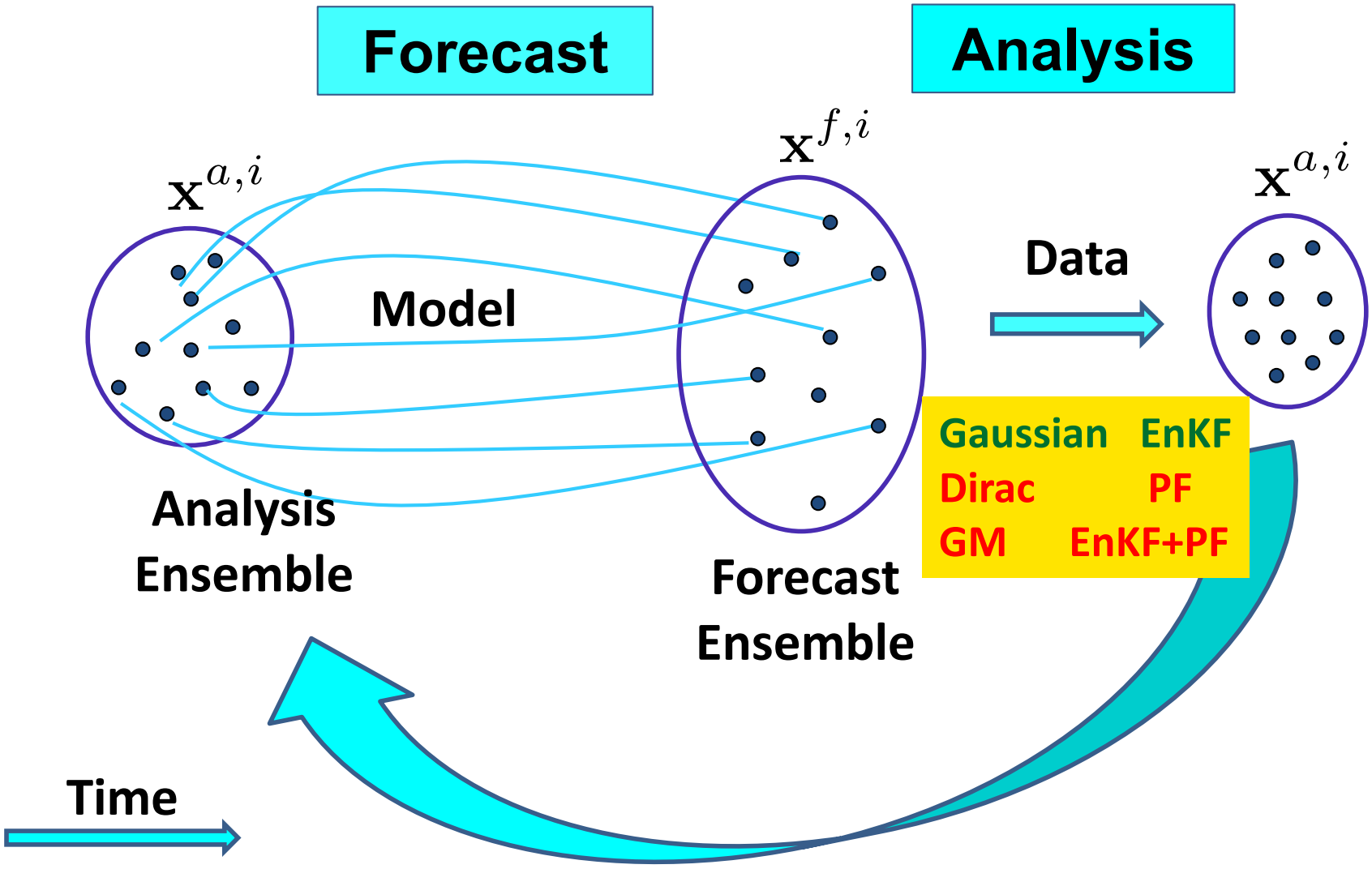


$$\sum_{i=1}^{N} w_i \delta(x - x_i)$$

$$\sum_{i=1}^{N} w_i \mathcal{N}\left(x - x_i; \mathrm{P}_i\right)$$

$$\mathrm{P}_i = \alpha \cdot \rho \circ \mathrm{Cov}(x_i) + \beta \cdot \mathrm{B}$$

Ensemble Filtering

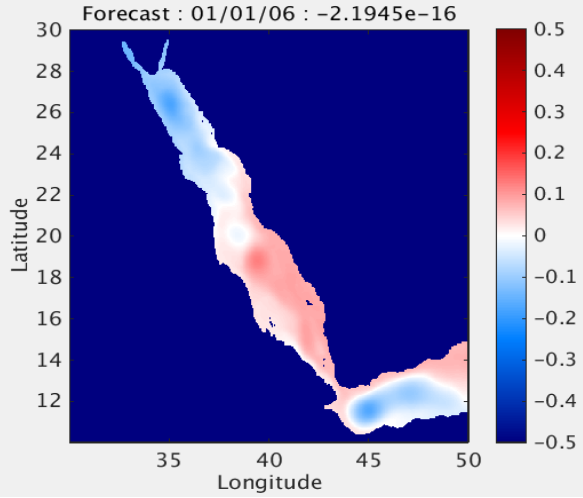


EnKF Assimilation in the Red Sea (2km)

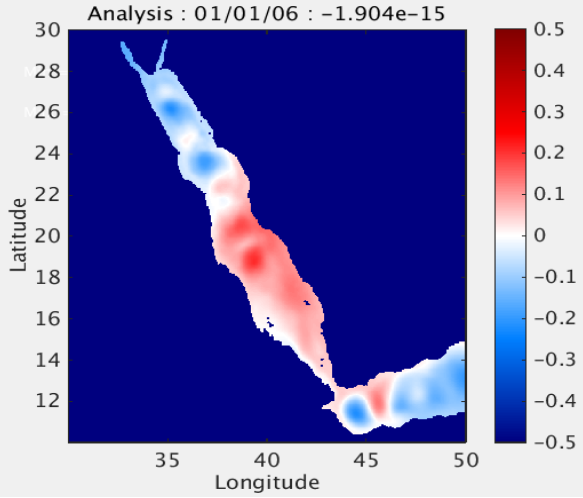
King Abdullah University of Science and Technology

KAUST

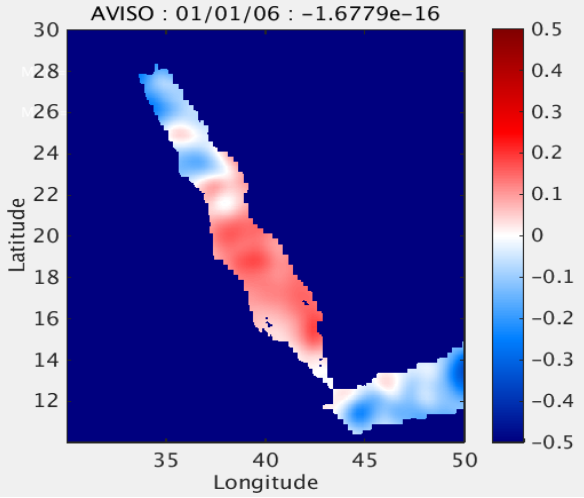
Forecast

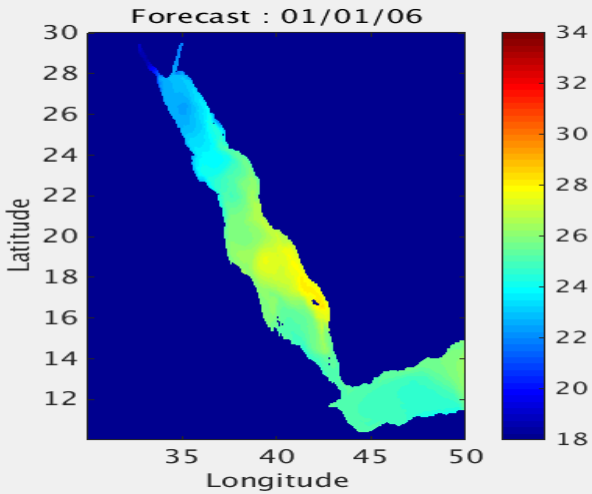


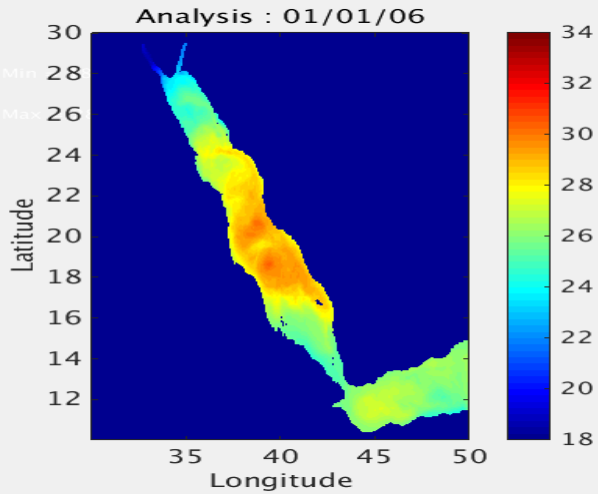
Analysis

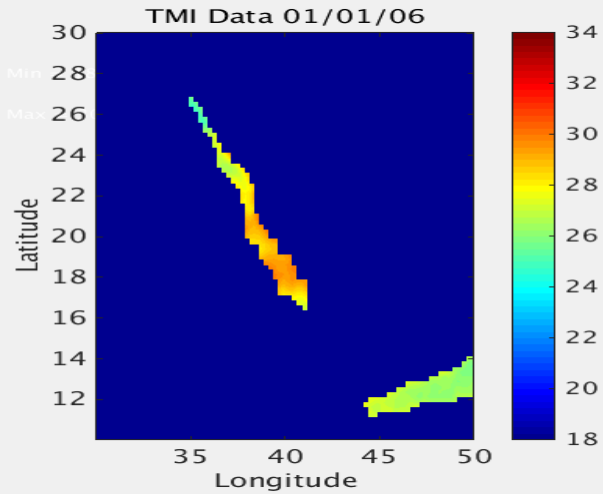


AVISO









Moving Forward with Ensemble Assimilation

King Abdullah University of Science and Technology KAUST

EnOI ⟸ **EnKF** ⟹ **Non-Gaussian**

Stationary Covariances

- Seasonal
  (Toye et al., 2017)

Why EnOI?

King Abdullah University of Science and Technology KAUST

Remove label | More

[Supercomputing Lab #13437] Proposal submission ITs x

Rooh Khurram via RT 12/7/15

to me

Dear Ibrahim,

RCAC approved your project allocation. We have added **49M core-hours** after adjusting for the temporary allocations. This allocation is equivalent to a grant of **USD 636K**, based on amortizing the acquisition and operating costs of Shaheen over a five-year period and dividing by the number of core-hours available during that span.

```
Project k1028: Ensemble Kalman and non-Gaussian Filtering
PI: Ibrahim Hoteit
```

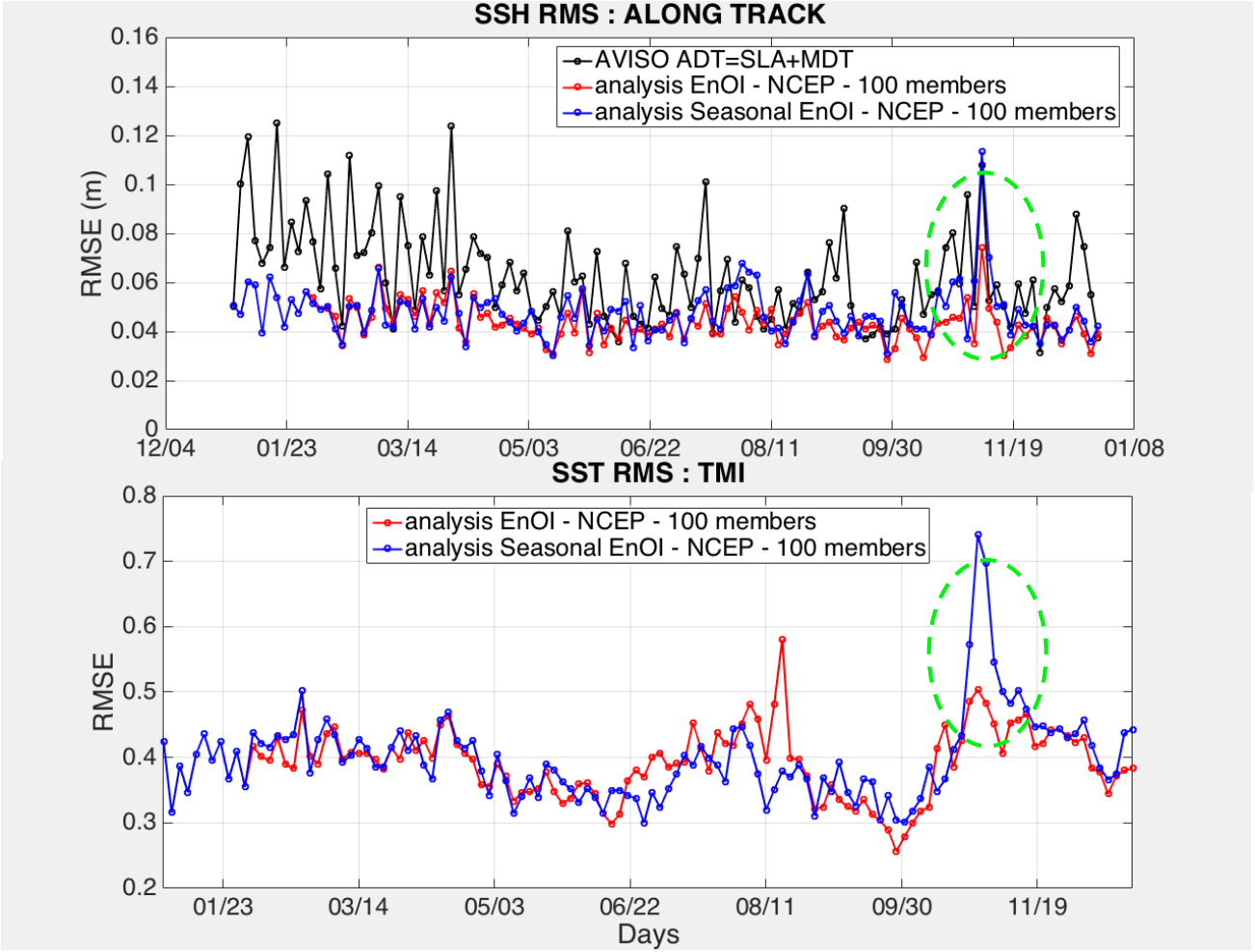
EnOI uses a stationary ensemble-based covariance:
reduce computing cost by a factor N (ensemble size)

EnOI vs. Seasonal-EnOI

King Abdullah University of Science and Technology

KAUST

- A stationary ensemble-covariance may not be appropriate for Red Sea → *"Hand-picked" stationary-variant covariances adapted for each "season"*



Moving Forward with Assimilation

King Abdullah University of Science and Technology KAUST

EnOI ⟵ **EnKF** ⟶ **Non-Gaussian**

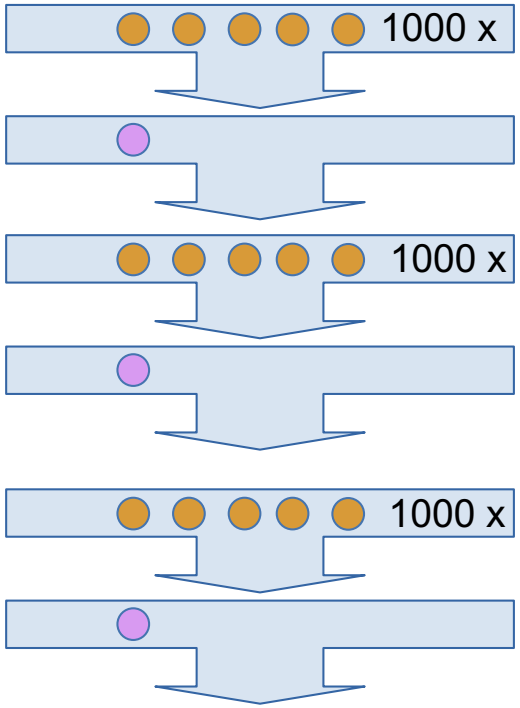
Stationary Covariances

- Seasonal (Toye et al., 2017)
- Adaptive - dictionary (Toye et al., 2017)
- *EnOI - GM*

- Fault-Free System

Typical MITgcm/DART workflow

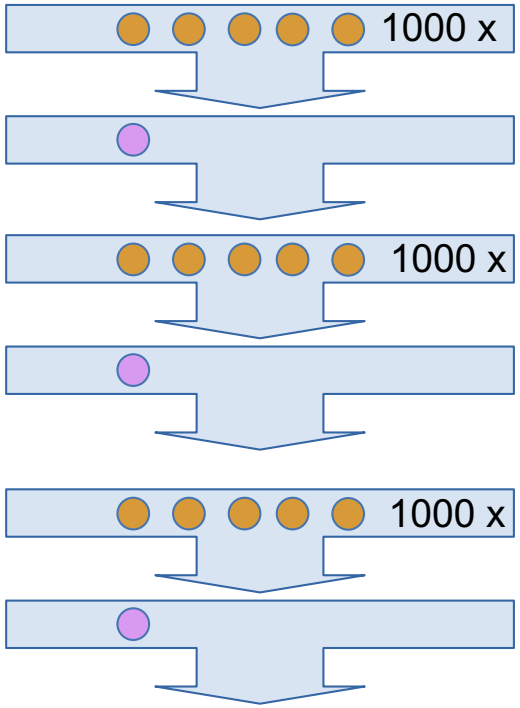
King Abdullah University of Science and Technology KAUST



First set of MITgcm (1000 x 6-node runs)
– barrier –
Apply the filter (DART) (1 x 32-node run)
– barrier –
second set of MITgcm (1000 x 6-node runs)
– barrier –
Apply the filter (DART) (1 x 32-node run)
– barrier –
3rd set of MITgcm (1000 x 6-node runs)
– barrier –
Apply the filter (DART) (1 x 32-node run)

Typical MITgcm/DART workflow

King Abdullah University of Science and Technology KAUST



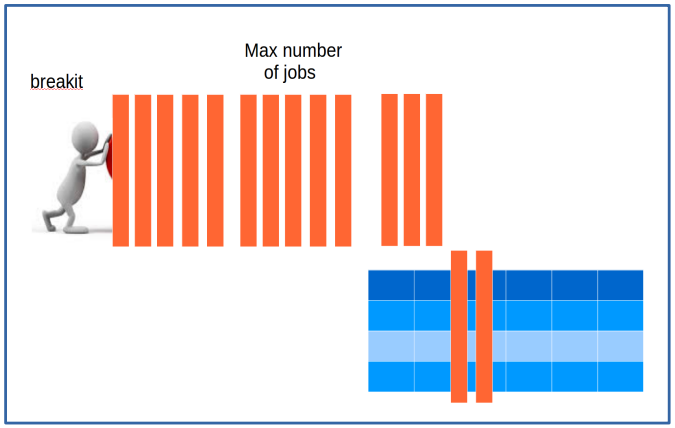
- Only 800 jobs can be submitted to Shaheen
- In case of hardware or numerical convergence failure, the whole workflow hangs and needs to be *manually fixed and restarted*
- Currently written as bash scripts: *hard to maintain, evolve and scale…*

Decimate

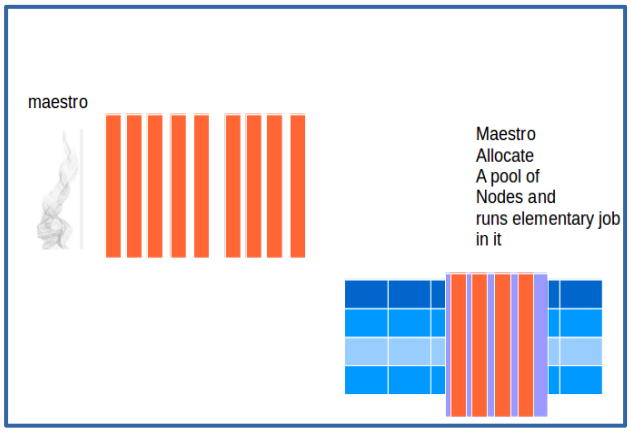
King Abdullah University of Science and Technology

KAUST

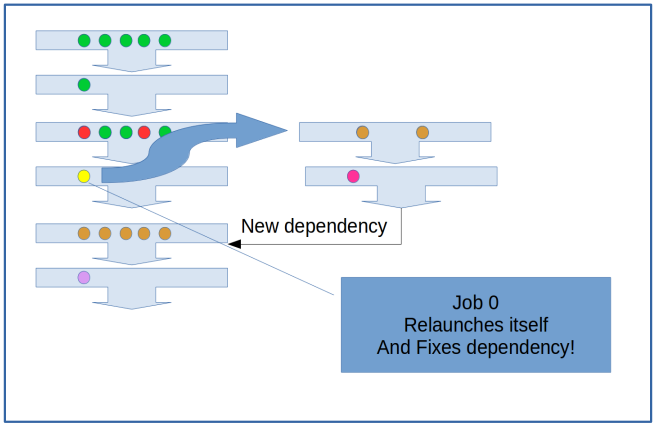
- we designed and developed **Decimate**, a Python scheduler extension to overcome these limitations



Breaks the 800 jobs limit by auto-submitting chunks of less than 800 jobs



Handles 1000s of jobs in a container seen as one SLURM job



Adds fault-tolerance: user can design his proper check & restart

- Tested and validated different replacement strategies

```
kortass@cdl3:/project/k1029/Sam/run_1k> d -sa

         ########################################
         #                                      #
         # Welcome to dart_mitgcm v  0.4.1!  #
         #   (using DECIMATE Framework 0.9)     #
         #                                      #
         ########################################

        running on cdl3 (shaheen)
                python /project/k1029/Sam/dart_mitgcm/0.4.1/sles11.3_gnu5.1.0/dart_mitgcm/dart_mitgcm.py -sa

[INFO ] !!!! WARNING WARNING mismatched tag single_restart_file_in values:
                                    >.false.< in ensemble_manager_nml.
                                    >.true.< in restart_file_tool_nml.
[INFO ] -------------> deleting non namespaced value........ "single_restart_file_in"
[INFO ] !!!! WARNING WARNING mismatched tag restart_in_file_name valu
                                    >perfect_ics< in perfect_mod
                                    >assim_model_state_ud< in fi
                                    >smoother_ics< in smoother_n
[INFO ] -------------> deleting non namespaced value........ "restart
[INFO ] input,nml.template did not change since last time, no need to
[INFO ] launch-0!0:2 unconsistent steps were found: [12-mitgcm-3,11-mitgcm-8]
 100%...[INFO ] launch-0!0:no active job in the queue, changing all WAITING in ABORTED???
[MSG  ] step 2-filter-0:1-1                    SUCCESS   SUCCESS:  100%   FAILURE:   0% -> []
[MSG  ] step 2-mitgcm-0:1-999                  MIXED     SUCCESS:   99%   FAILURE:   0% -> [614,617-620,622,624-625]
[MSG  ] step 2-mitgcm-1:614,617-620,622,624-625 SUCCESS   SUCCESS:  100%        FAILURE:   0% -> []
[MSG  ] step 3-filter-0:1-1                    SUCCESS   SUCCESS:  100%   FAILURE:   0% -> []
[MSG  ] step 3-mitgcm-0:1-999                  SUCCESS   SUCCESS:  100%   FAILURE:   0% -> []
[MSG  ] step 4-filter-0:1-1                    SUCCESS   SUCCESS:  100%   FAILURE:   0% -> []
[MSG  ] step 4-mitgcm-0:1-999                  MIXED     SUCCESS:   74%   FAILURE:  25% -> [124,235-236,238,241,243-245,751-999]
[MSG  ] step 4-mitgcm-1:124,235-236,238,241,243-245 SUCCESS   SUCCESS:  100%    FAILURE:   0% -> []
[MSG  ] step 5-filter-0:1-1                    SUCCESS   SUCCESS:  100%   FAILURE:   0% -> []
[MSG  ] step 5-mitgcm-0:1-999                  MIXED     SUCCESS:   99%   FAILURE:   0% -> [861,875]
[MSG  ] step 5-mitgcm-1:861,875                SUCCESS   SUCCESS:  100%   FAILURE:   0% -> []
[MSG  ] step 6-filter-0:1-1                    SUCCESS   SUCCESS:  100%   FAILURE:   0% -> []
[MSG  ] step 6-mitgcm-0:1-999                  MIXED     SUCCESS:   99%   FAILURE:   0% -> [345,350,352]
[MSG  ] step 6-mitgcm-1:345,350,352            SUCCESS   SUCCESS:  100%   FAILURE:   0% -> []
```

Second step of MITgcm was restarted once after 1% of failure

With no manual intervention…

5 ensemble assimilation steps made in 3h30 on a crowded machine

Simulation Footprint

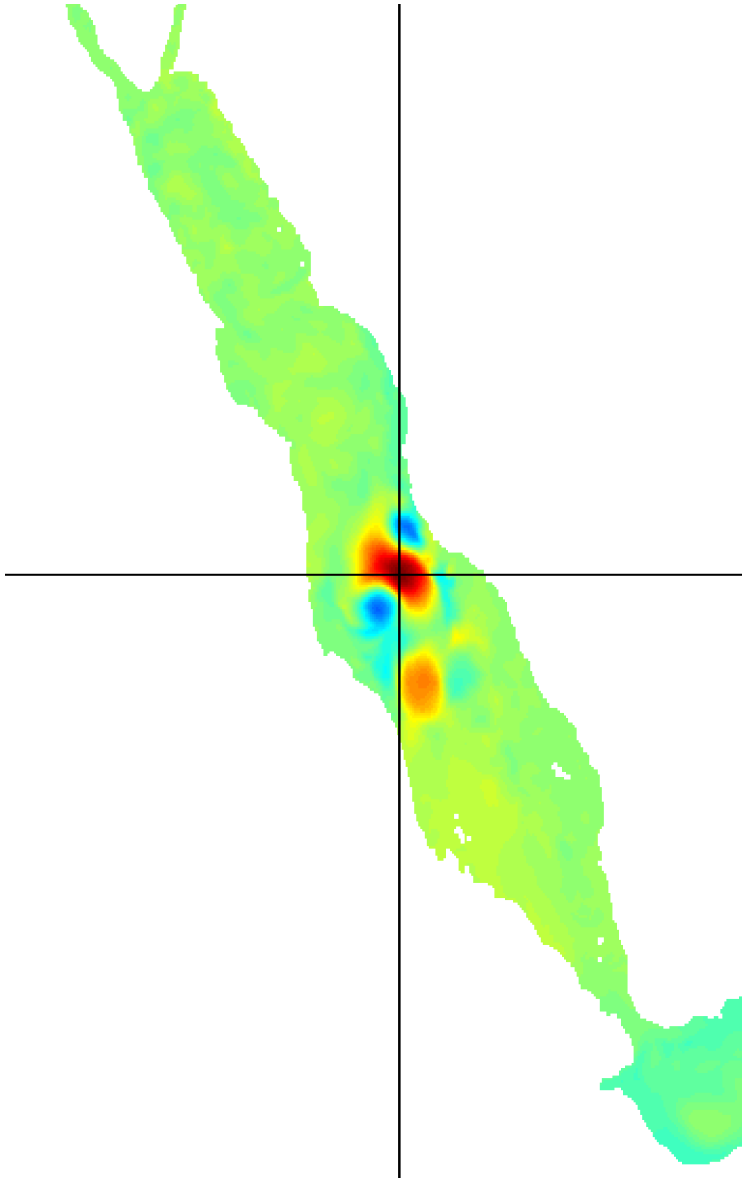
King Abdullah University of Science and Technology KAUST

- **100 or 1000 members ensemble assimilation:**
  - 22000 independent successful runs of MITgcm
  - 617 failed MITgcm runs due to model failures followed by a change of members (< 3%)
  - Roughly 10% of failed jobs due to hardware failures (needs a more precise survey); before maintenance of Sept 3, filesystem was highly sollicitated
  - Experiments run from August 21 to Sept 7 2017 on a machine busy at more than 90%

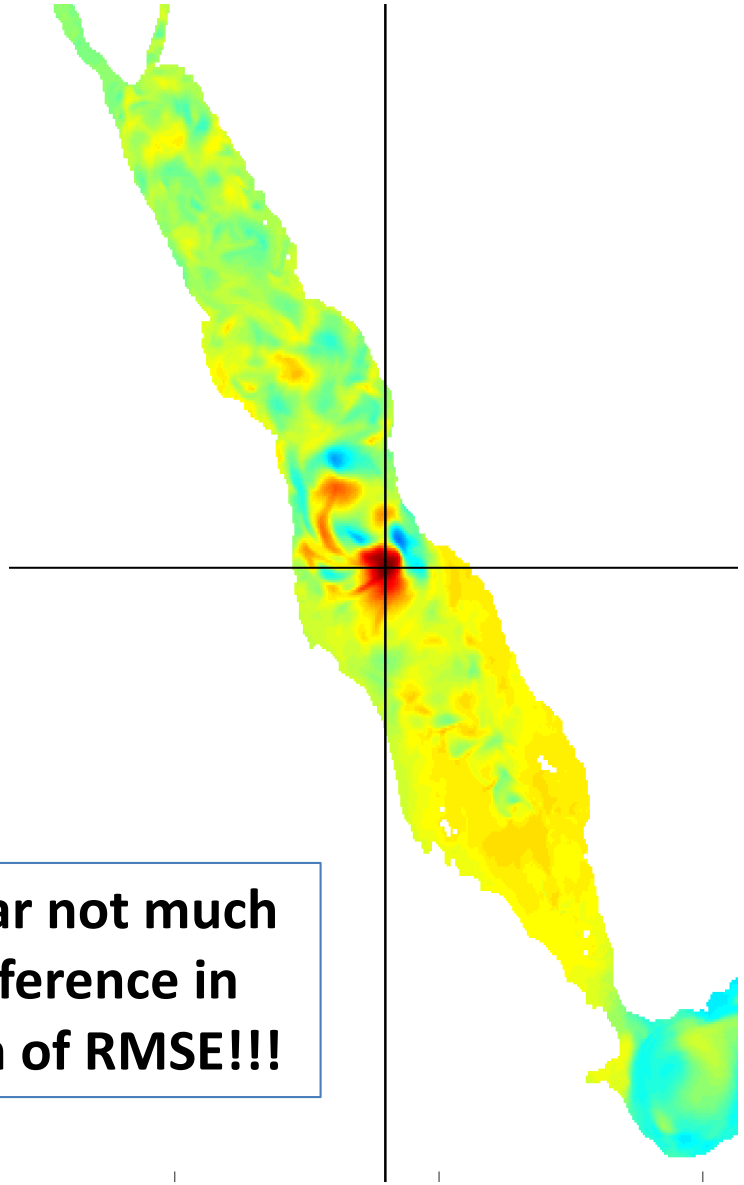
Correlations

King Abdullah University of Science and Technology KAUST

1000 members



100 members



So far not much difference in term of RMSE!!!

Moving Forward with Assimilation

King Abdullah University of Science and Technology KAUST

EnOI ⟵ **EnKF** ⟶ **Non-Gaussian**

Stationary Covariances

- Seasonal (Toye et al., 2017)
- Adaptive - dictionary (Toye et al., 2017)
- *EnOI - GM*

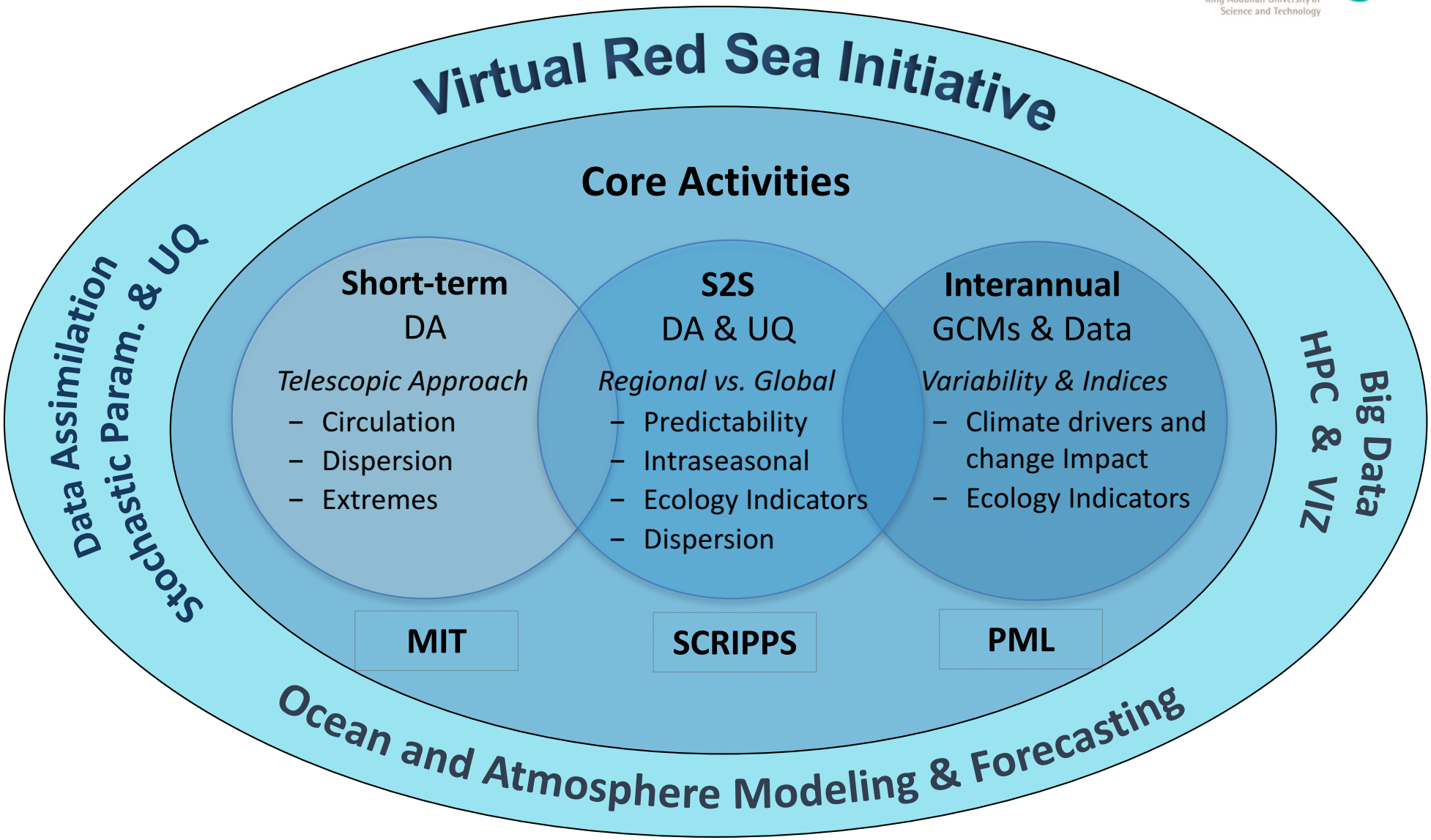
- Fault-Free System (Kortas et al., 2017)
- *One-Step-Ahead Smoothing* (Raboudi et al., 2017)
- *Parameters, Bias, ...*

Small Ensembles???

- *GM + Clustering + "targeted sampling"* (Lui et al., 2015; 2016)
- *Variational PF* (Ait-El-Fquih al., 2016)

Hybrid ***???***

Next 5 Years …


Virtual Red Sea Initiative
Core Activities
Short-term
DA
Telescopic Approach
– Circulation
– Dispersion
– Extremes
S2S
DA & UQ
Regional vs. Global
– Predictability
– Intraseasonal
– Ecology Indicators
– Dispersion
Interannual
GCMs & Data
Variability & Indices
– Climate drivers and change Impact
– Ecology Indicators
MIT
SCRIPPS
PML
Data Assimilation
Stochastic Param. & UQ
Big Data
HPC & VIZ
Ocean and Atmosphere Modeling & Forecasting

Atlas of Ocean, Atmosphere, Bio,

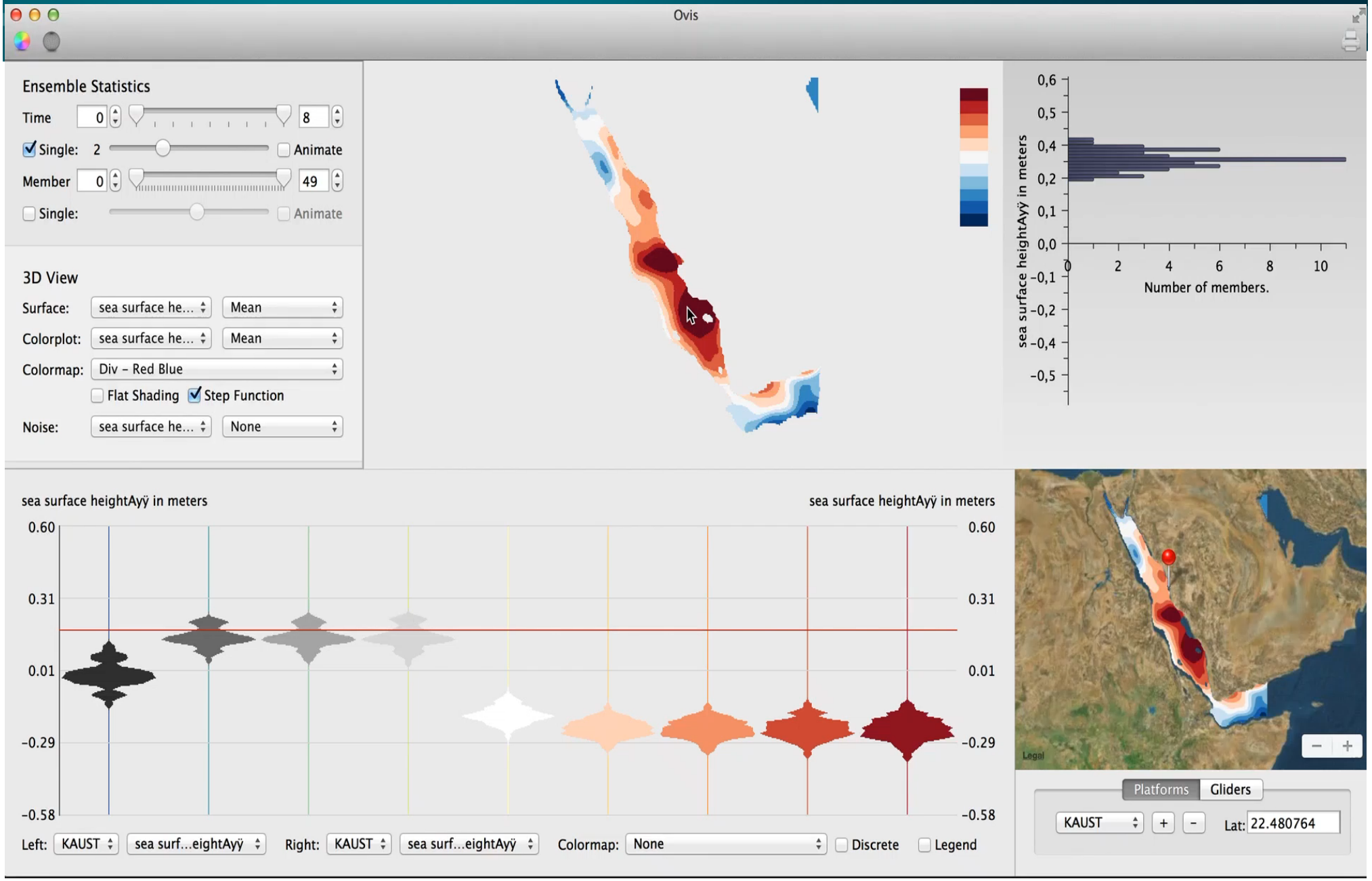
King Abdullah University of Science and Technology

KAUST

Interactive 5D Visualization

King Abdullah University of Science and Technology

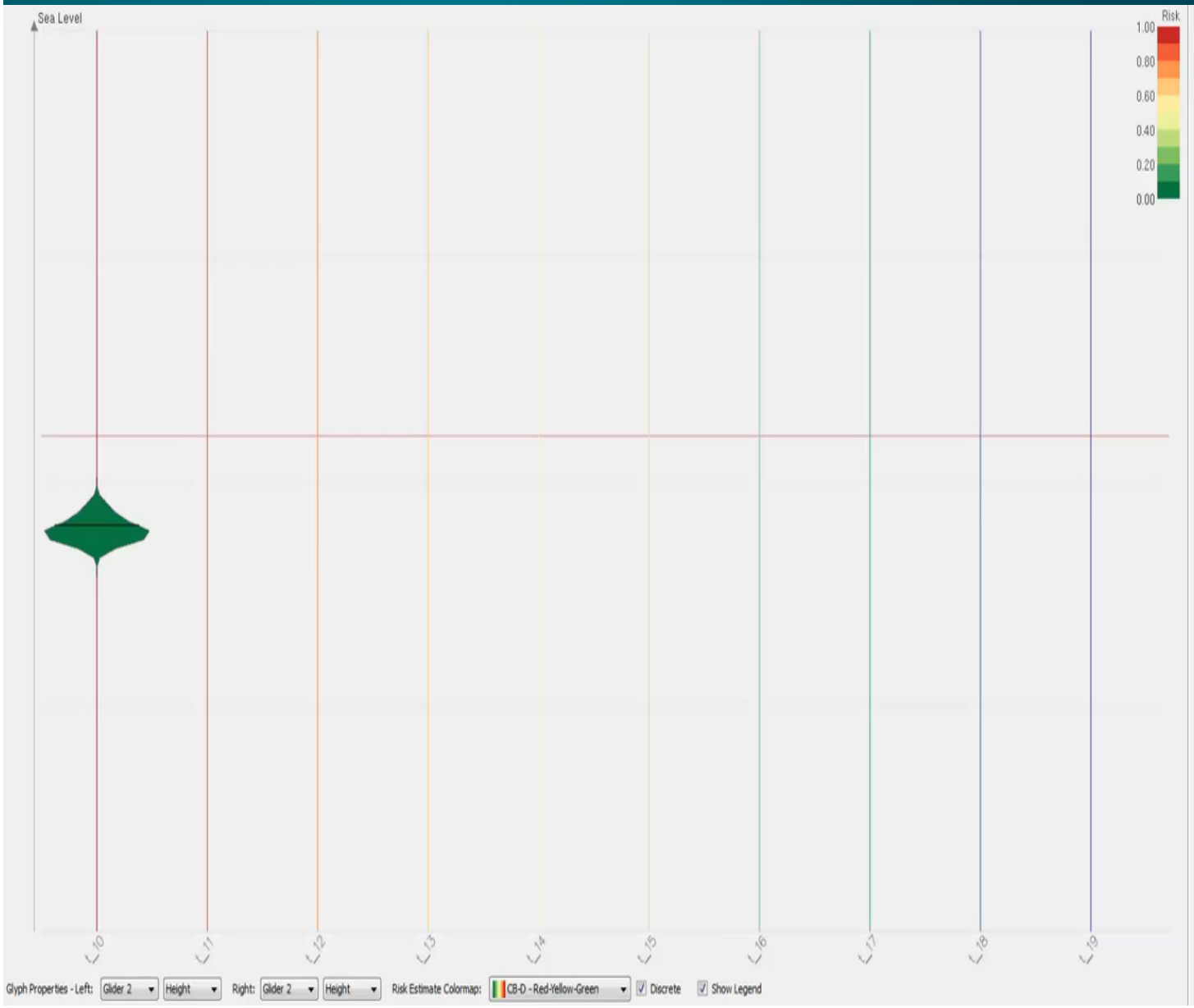
KAUST

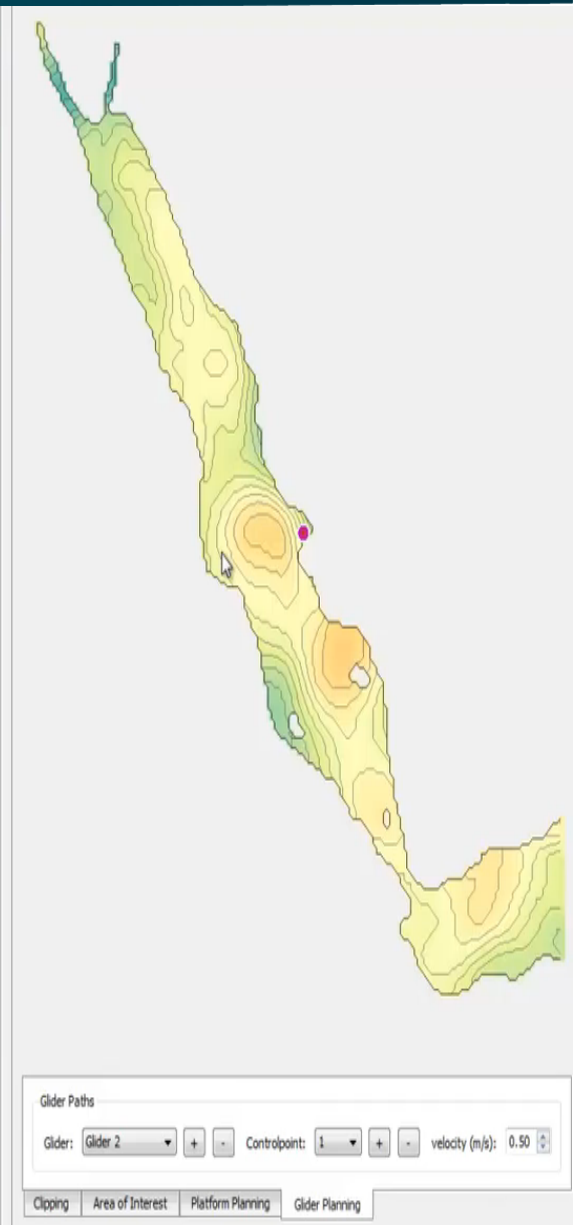


Gliders Pathways Planing

King Abdullah University of Science and Technology

KAUST

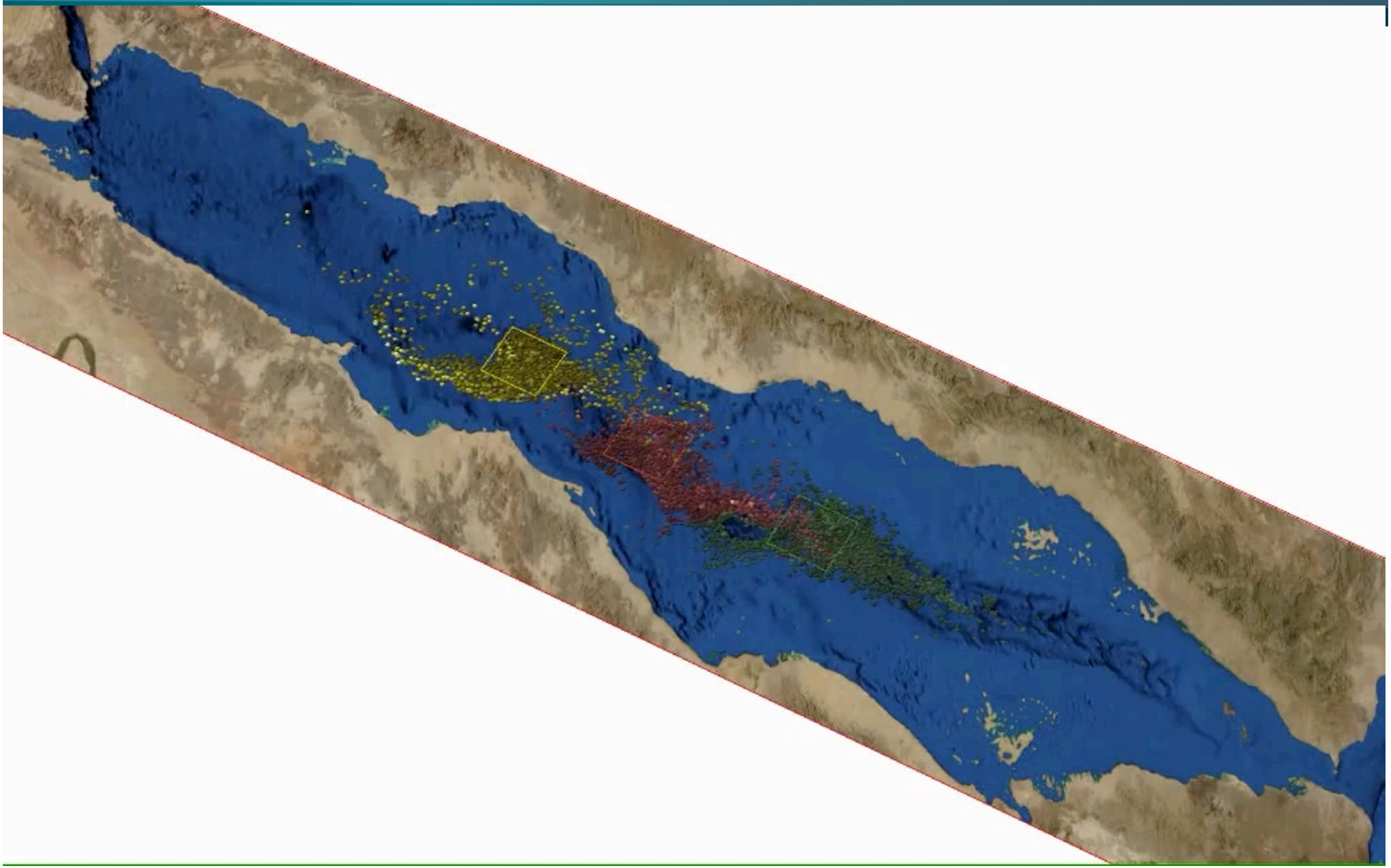




Interactive Particles Tracking

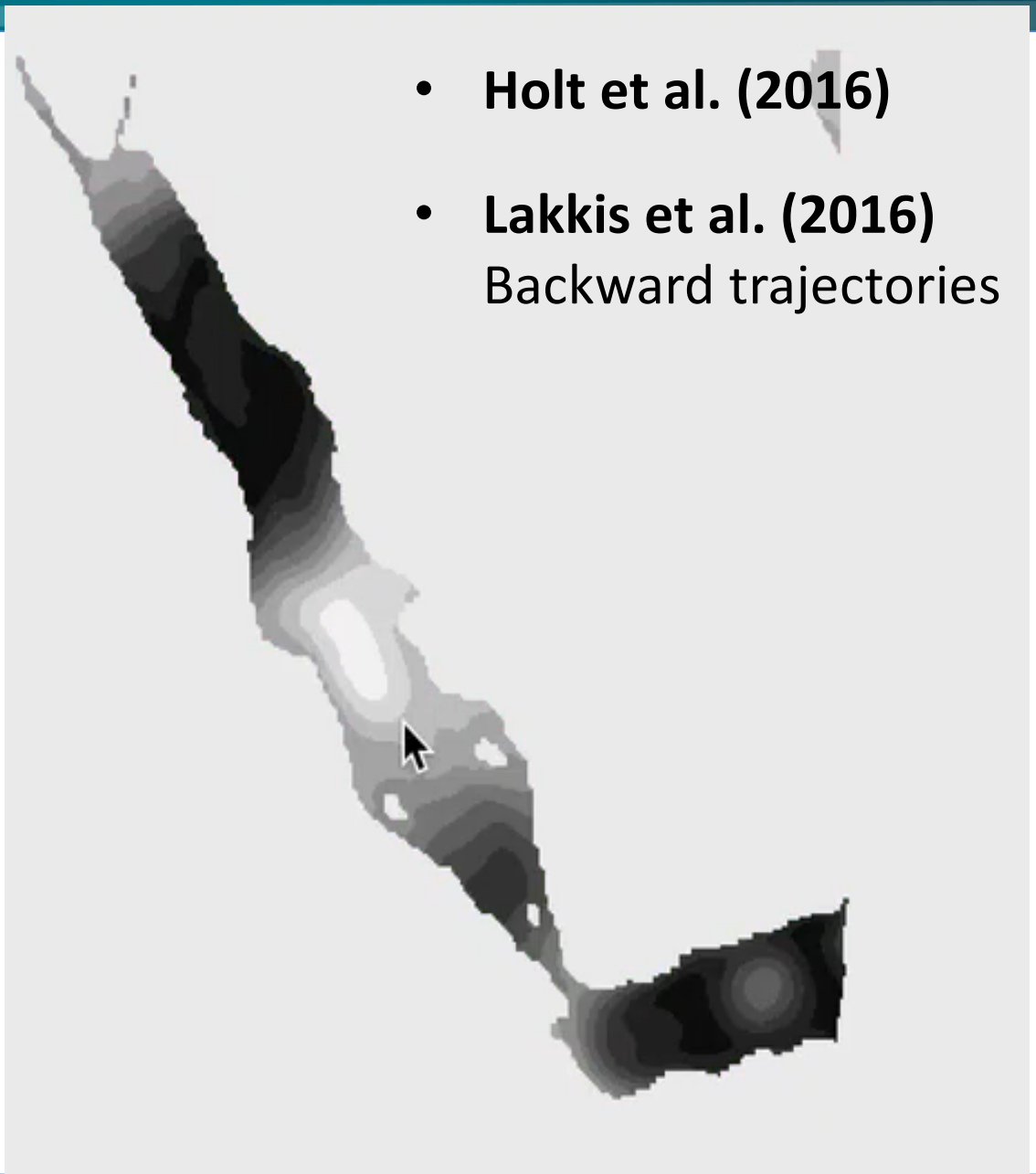
King Abdullah University of Science and Technology

KAUST



Probabilistic Particles Tracking

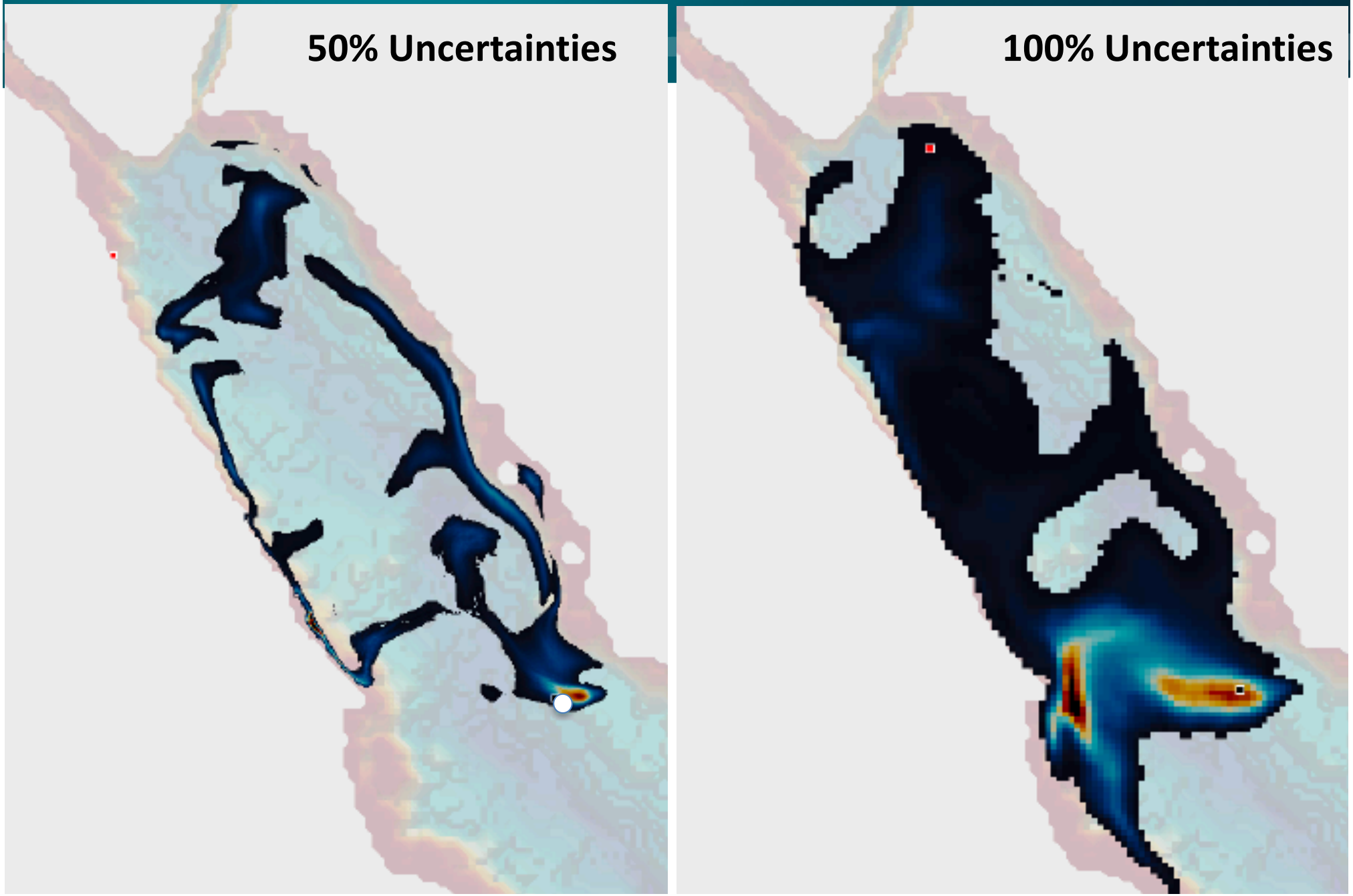
King Abdullah University of Science and Technology
KAUST

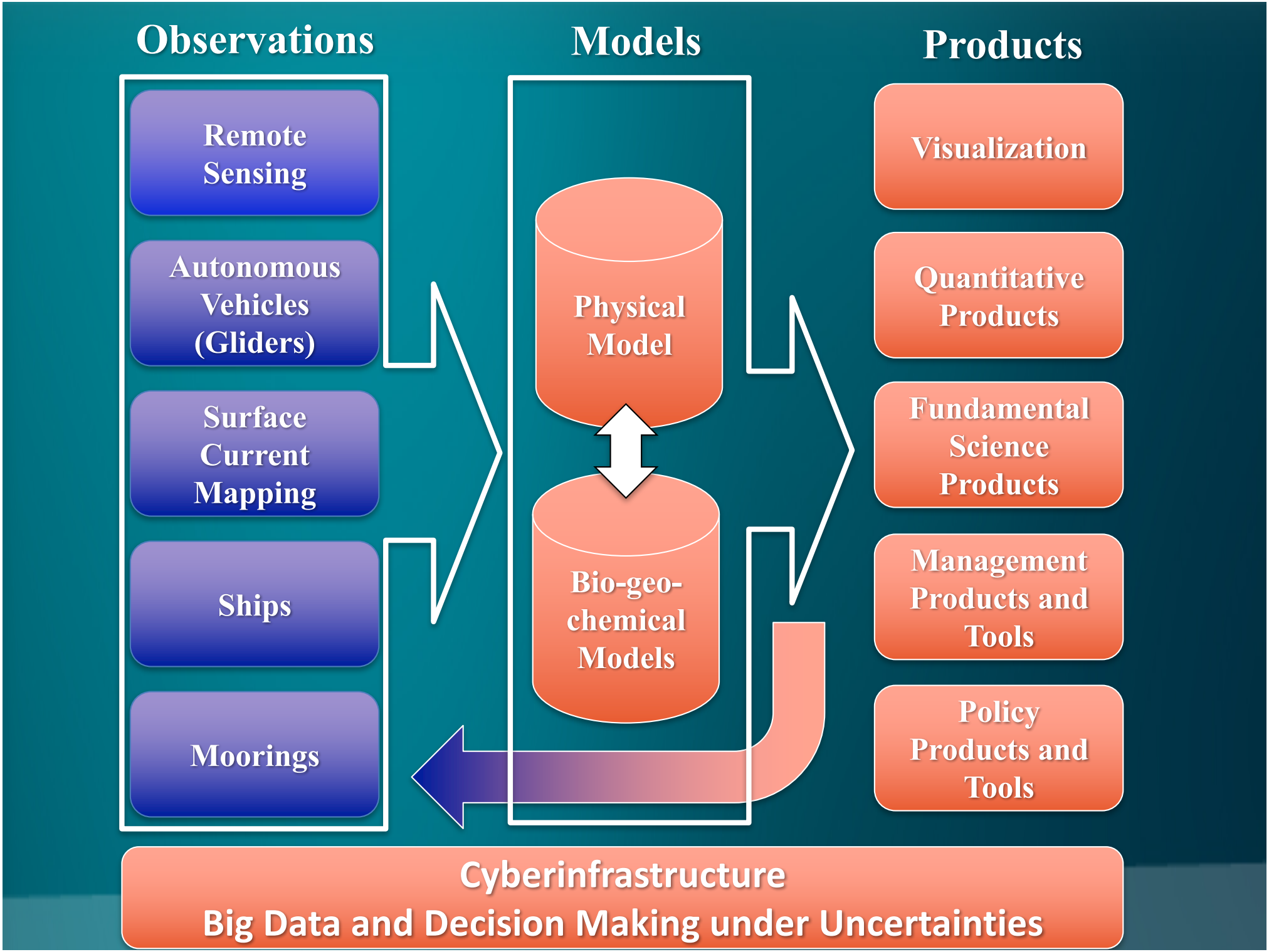


- **Holt et al. (2016)**
- **Lakkis et al. (2016)**
  Backward trajectories

Probabilistic BackwardTracking

King Abdullah University of Science and Technology KAUST




Observations
Models
Products
Remote Sensing
Autonomous Vehicles (Gliders)
Surface Current Mapping
Ships
Moorings
Physical Model
Bio-geo-chemical Models
Visualization
Quantitative Products
Fundamental Science Products
Management Products and Tools
Policy Products and Tools
Cyberinfrastructure
Big Data and Decision Making under Uncertainties

THANK YOU