

Application-specific **model selection and model weighting** of **Earth system models** with application to regional environmental management of **red tide**

Ming Ye (mye@fsu.edu), Florida State University

Ahmed S. Elshall, Florida Gulf Coast University

Sven A. Kranz, Rice University

Julie Harrington, Florida State University

Xiaojuan Yang, Oak Ridge National Laboratory

Yongshan Wan, Environmental Protection Agency

Mathew Maltrud, Los Alamos National Laboratory



National Science Foundation Award 1939994

48th Annual Climate Diagnostics and Prediction Workshop
21st Annual Climate Prediction Applications Science Workshop
March, 28, 2024

Karenia brevis, single-celled, marine dinoflagellate



Image Credit: Sid Perkins



RED TIDE

Red tide in Florida and Texas is caused by the rapid growth of a microscopic algae called *Karenia brevis*. When large amounts of this algae are present, it can cause a harmful algal bloom (HAB) that can be seen from space. NOAA issues HAB forecasts based on satellite imagery and cell counts of *Karenia brevis* collected in the field and analyzed by NOAA partners.

A map of Florida and Texas with a color gradient overlay indicating the location of a red tide bloom. The bloom is shown as a yellow and orange area along the coast of Florida and Texas. A circular inset shows a magnified view of the microscopic algae, *Karenia brevis*, which are small, red, spherical cells with two flagella.

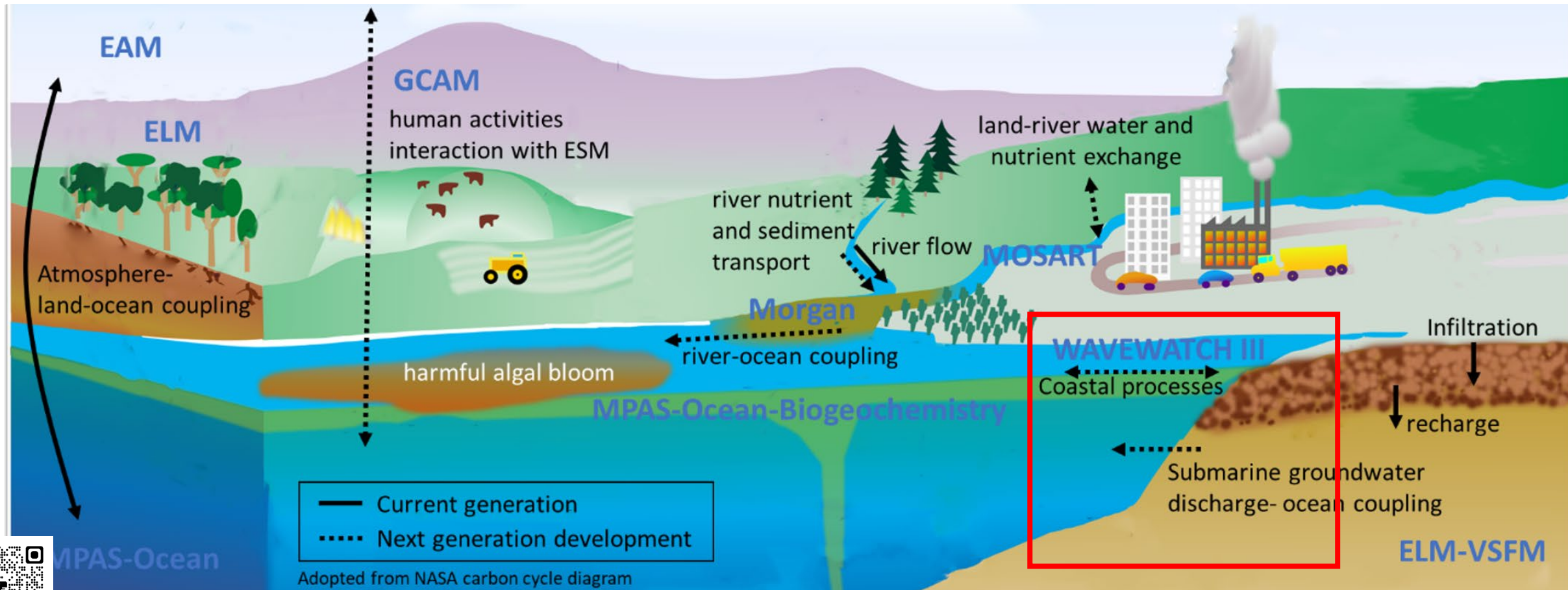
<https://oceanservice.noaa.gov/hazards/hab/gulf-mexico.html>

<https://start1.org/red-tide>

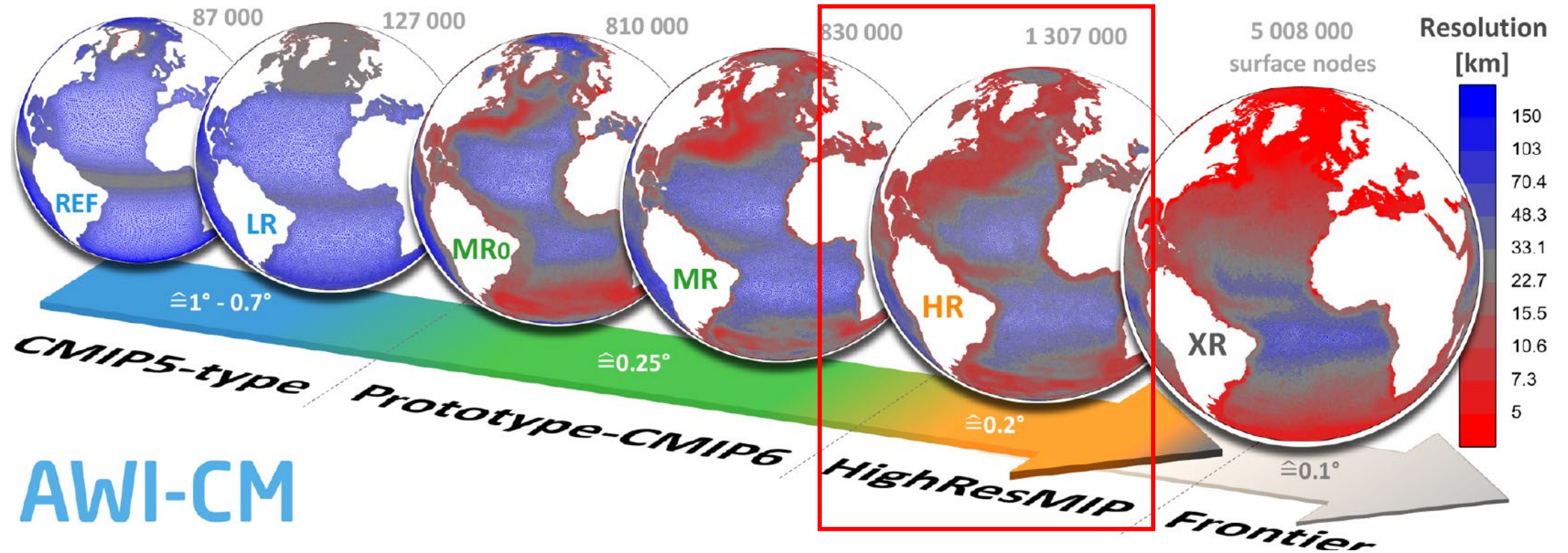
<https://www.sarasotamafitness/2022/07/red-tide->

Red tide projection is challenging as **multiple drivers** can alter the occurrence, intensity, and toxicity of red tides.

Is it possible to forecast red tides by using Earth system models outputs **directly**? **YES and NO.**



Earth system models for red tides in a changing climate based on teleconnections between **global** and **regional** phenomena

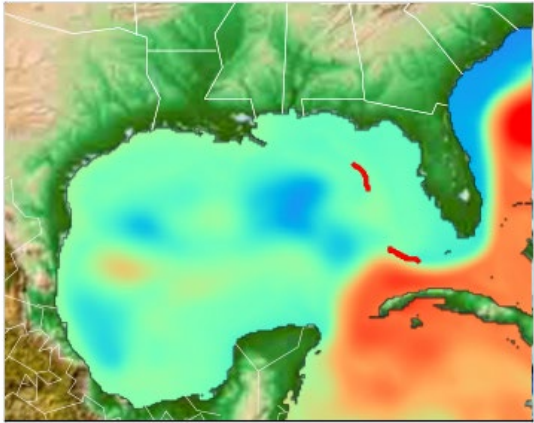


Rackow et al. (2019 GMD)

A high resolution-ESM can reproduce regional phenomena like **Loop Current**, a warm ocean current that drives red tides

Reanalysis Data (8km)

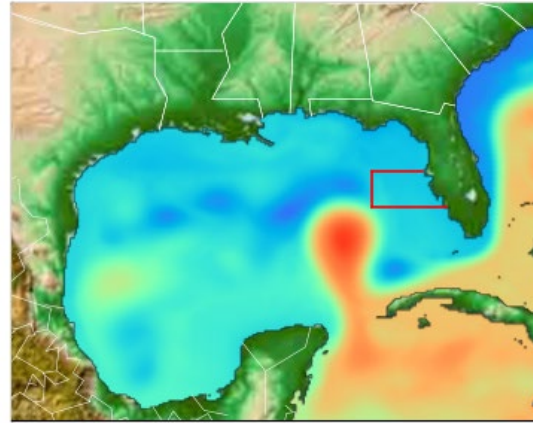
Loop Current South
March 2010



CMEMS global-reanalysis-phy-001-030 (8 km)

Reanalysis Data (8km)

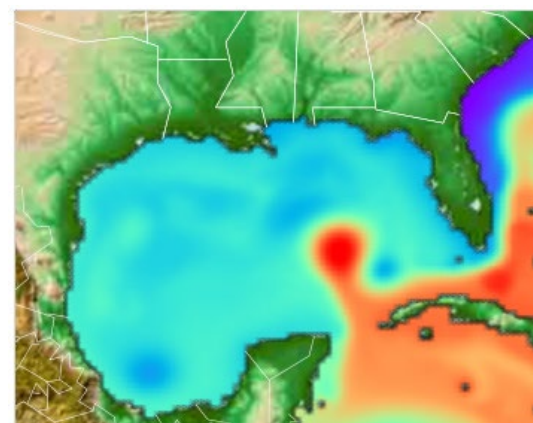
Loop Current North
June 2010



CMEMS global-reanalysis-phy-001-030 (8 km)

High Resolution-ESM (25km)

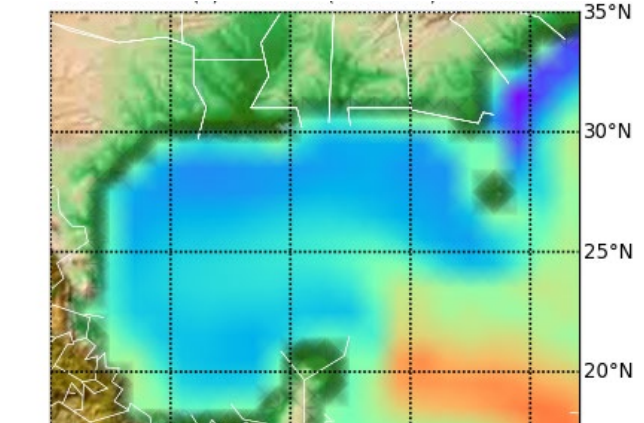
Loop Current North
June 2010



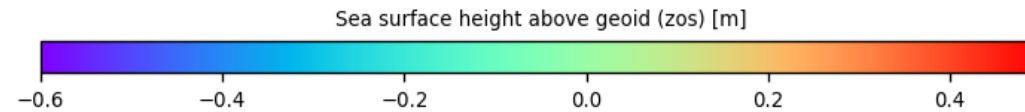
HadGEM3-GC3-MM (25 km)

Low Resolution-ESM (100km)

Loop Current North
June 2010



E3SM (100 km)

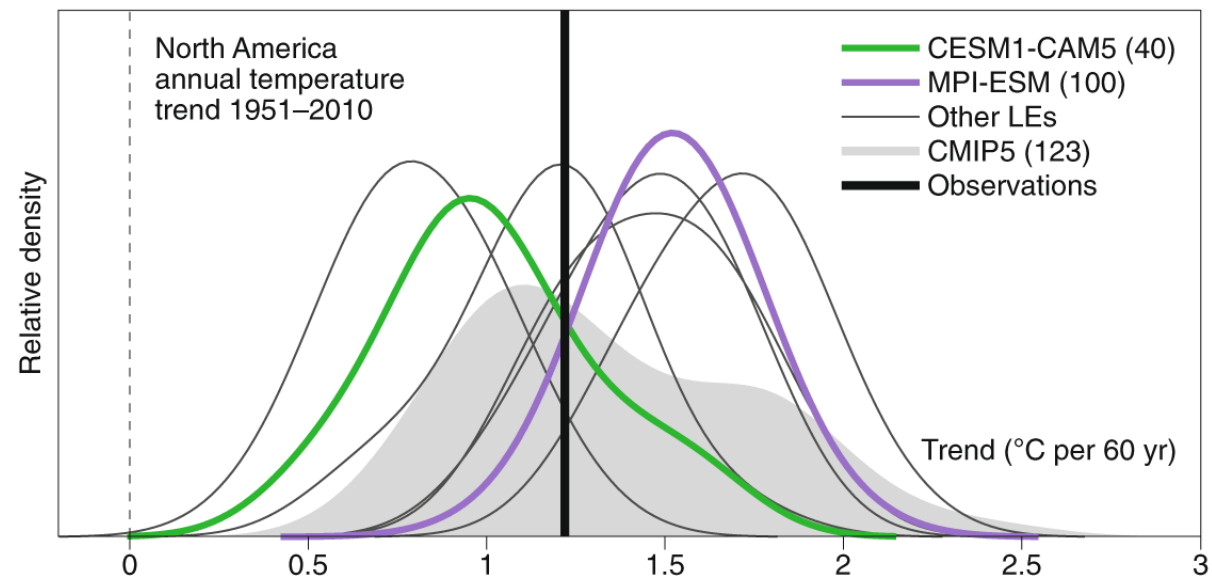


- **Loop Current** is an important factor that controls the occurrence of red tide.
- Maze et al. (2015) showed that the Loop current in a **north position** penetrating through the Gulf of Mexico is a **necessarily condition** for a large red tide bloom to occur (based on red tide data in the red box).
- With approximately 0.3 divisions per day, *Karenia brevis* is a slow growing dinoflagellate that requires an area with mixing slower than the growth rate to form a bloom.

Independent model subset (IMS)	Institution	Country	Model (Reference)	Experiment ID	Members	Ocean model resolution	Ocean Model	Ocean grid	ESM nominal resolution
IMS01	NCAR	USA	CESM1-CAM5-SE-HR (Chang et al. 2020)	hist-1950	r1i1p1f1	0.1° (11 km) nominal resolution	POP2	POP2-HR	25 km
IMS02	CMCC	Italy	CMCC-CM2-HR4 (Cherchi et al. 2019)	hist-1950	r1i1p1f1	0.25° from the Equator degrading at the poles	NEMO v3.6	ORCA025	25 km
			CMCC-CM2-VHR4 (Cherchi et al. 2019)	hist-1950	r1i1p1f1	0.25° from the Equator degrading at the poles	NEMO v3.6	ORCA025	25 km
IMS03	CNRM-CERFACS	France	CNRM-CM6-1-HR (Voltaire et al. 2019)	hist-1950	r(1-3)j1p1f2	0.25° (27-28 km) nominal resolution	NEMO v3.6	eORCA025	25 km
			CNRM-CM6-1-HR (Voltaire et al. 2019)	historical	r1i1p1f2	0.25° (27-28 km) nominal resolution	NEMO v3.6	eORCA025	25 km
IMS04	DOE-E3SM-Project	USA	E3SM-1-0 (Golaz et al. 2109)	historical	r(1-5)j1p1f1	60 km in mid-latitudes and 30 km at the equator and poles	MPAS-O	EC60to30	100 km
IMS05	EC-Earth-Consortium	Europe	EC-Earth3P (Haarsma et al. 2020)	hist-1950	r(1-3)j1p2f1	about 1° (110 km)	NEMO v3.6	ORCA1	100 km
IMS06	EC-Earth-Consortium	Europe	EC-Earth3P-HR (Haarsma et al. 2020)	hist-1950	r(1-3)j1p2f1	about 0.25° (27-28 km)	NEMO v3.6	ORCA025	25 km
IMS07	ECMWF	Europe	ECMWF-IFS-HR (Roberts et al. 2018)	hist-1950	r(1-6)j1p1f1	25 km nominal resolution	NEMO v3.4	ORCA025	25 km
IMS08			ECMWF-IFS-MR (Roberts et al. 2018)	hist-1950	r(1-3)j1p1f1	25 km nominal resolution	NEMO v3.4	ORCA025	25 km
IMS09	NOAA-GFDL	USA	GFDL-CM4 (Held et al 2019)	historical	r1i1p1f1	0.25° (27-28 km) nominal resolution	MOM6	tri-polar grid	50 km
			GFDL-ESM4 (Held et al 2019)	historical	r(2-3)j1p1f1	0.25° (27-28 km) nominal resolution	MOM6	tri-polar grid	50 km
IMS10	NERC	UK	HadGEM3-GC31-HH (Roberts et al. 2019)	hist-1950	r1i1p1f1	8 km nominal resolution	NEMO v3.6	ORCA12	10 km
	MOHC-NERC	UK	HadGEM3-GC31-HM (Roberts et al. 2019)	hist-1950	r1i(1-3)p1f1	25 km nominal resolution	NEMO v3.6	ORCA12	50 km
IMS11	MOHC	UK	HadGEM3-GC31-MM (Roberts et al. 2019)	hist-1950	r1i(1-3)p1f1	25 km nominal resolution	NEMO v3.6	ORCA025	100 km
			HadGEM3-GC31-MM (Roberts et al. 2019)	historical	r(1-4)j1p1f3	25 km nominal resolution	NEMO v3.6	ORCA025	25 km

Ensemble Modeling Approaches

An Illustration



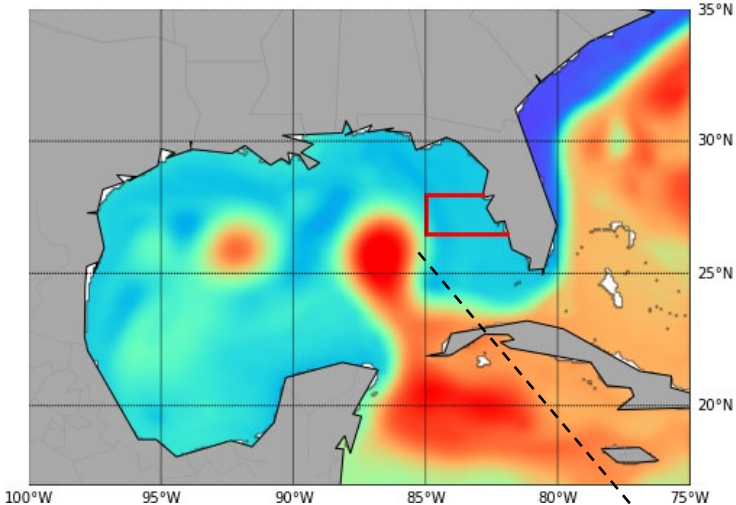
<https://www.nature.com/articles/s41558-020-0731-2>

Ensemble ESMs for loop current

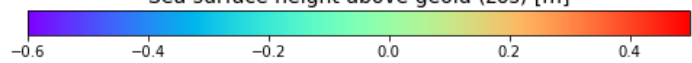
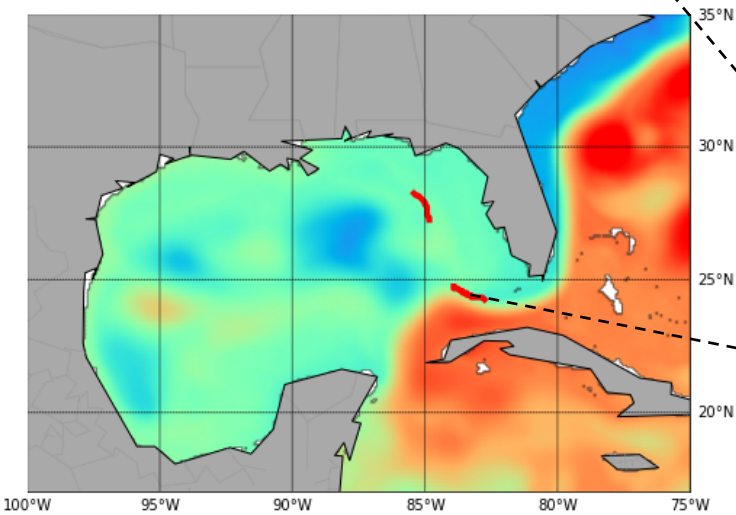
- 41 CMIP6 model runs from 14 different models developed by eight institutes
- **Prescreening-based subset selection**
Excluding non-representing models
- **Application-specific optimal model weighting**

Prescreening-based Subset Selection: Loop Current

(a) Reanalysis Data: Loop Current North Position



(b) Reanalysis Data: Loop Current South Position



CMIP6: zos - Sea Surface Height Above Geoid

Zos anomaly of model M_k (the k-th model in the ensemble)

$$h_t = \max_{h_n} \left(\Delta_m \left[E_l \left[E_k \left(E_j (h_{j,k,l,m,n,t} | M_k) \right) \right] \right] \right)$$

Is the maximum difference of the zos data at the north and the south segment.

LC-N: Loop Current in the north position **LC-S: Loop Current in the south position**

$$H_{LC-N}(h_t) = \begin{cases} 1, & h_t \geq 0 \\ 0, & h_t < 0 \end{cases}$$

$$H_{LC-S}(h_t) = \begin{cases} 1, & h_t < 0 \\ 0, & h_t \geq 0 \end{cases}$$

(a) Reanalysis data



H_{LC-N}

H_{LC-S}

LC-N is a necessary condition of red tide large bloom.

Prescreening

Three prescreening metrics based on a model's ability to reproduce main features of the **physically interpretable relationships of interest (prior info)**.

Oscillating event representation
If the sea surface height is consistently higher at the north segment than at the south segment, then the model is unable to represent alternation of LC-N and LC-S.

Scoring

Score = 1 or 0

$$y_2 = \begin{cases} 1, & 0 < \sum_{t=1}^T H_{LC-N}(h_t) < T \\ 0, & \sum_{t=1}^T H_{LC-N}(h_t) = T \end{cases}$$

Prescreening

a1. Select input data

Raw gridded CMIP6 model output of sea surface height (zos): hist-1950 and historical

(Section 2.2)

a2. Identify independent model subsets

Institutional democracy and ocean grid to identify independent model subsets

(Section 2.3)

a3. Process input data

Loop Current position of independent model subsets following the method of Maze et al. (2015)

(Section 2.4)

a4. Define prescreening metrics

- Physical phenomena simulation (y_1)
- Oscillating event representation (y_2)
- Oscillating event realism (y_3)

(Section 2.5)

a5. Score independent model subsets

Scoring of independent model subsets based on performance ranking given prescreening metrics

(Section 3.1)

a7. Evaluate prediction

Evaluate prediction of independent model subsets given defined predictors

(Section 3.1)

a6. Define prediction metrics

- Oscillating event frequency (y_4)
- Temporal match error (y_5)
- Karenia brevis error (y_6)
- Root-mean-square error (y_7)

(Section 2.5)

Raw gridded observation reanalysis product of sea surface height (zos)

(Section 2.2)

Loop Current position of observation reanalysis data following the method of Maze et al. (2015)

(Section 2.4)

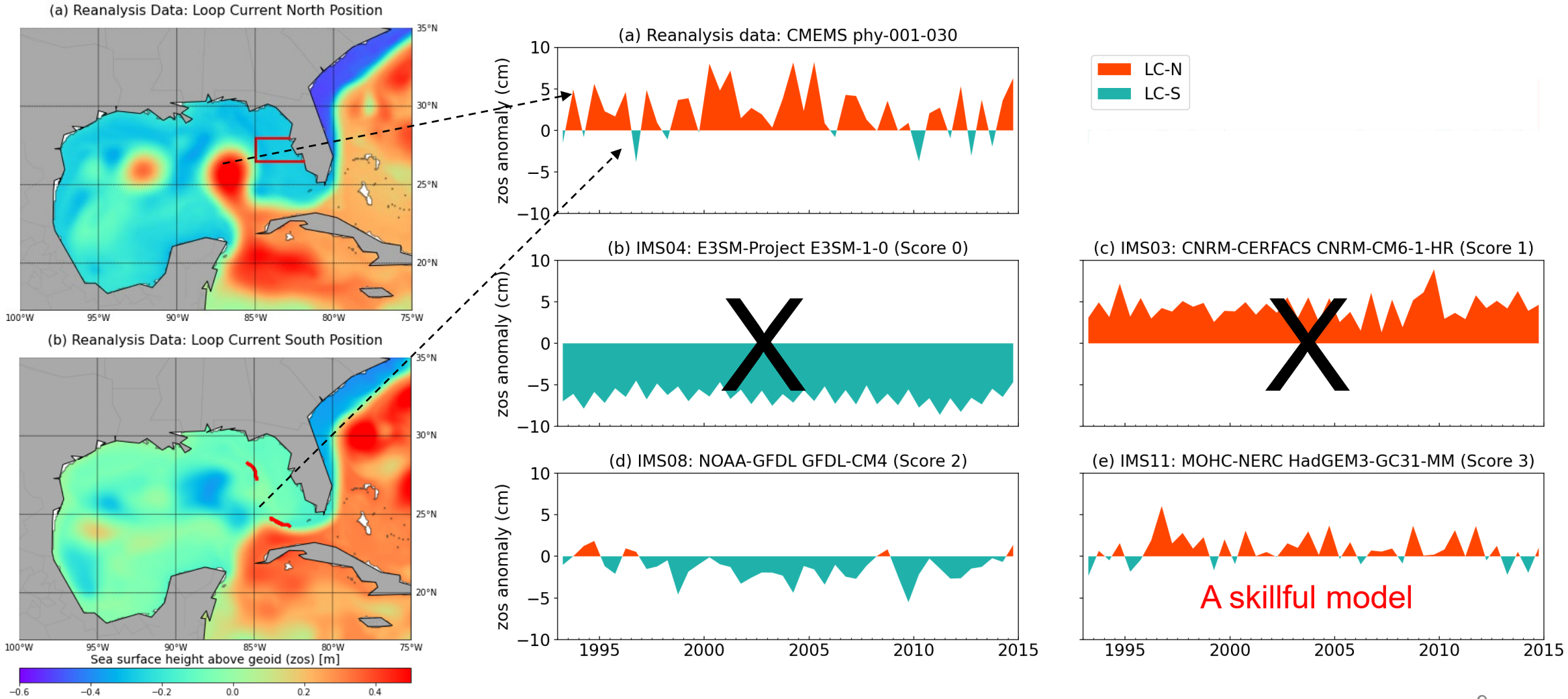
Raw Karenia brevis concentration data

(Section 2.2)

Karenia brevis no bloom or large bloom following the method of Maze et al. (2015)

(Section 2.2)

Prescreening-based subset selection can find representative and skillful models for a given application



Subset selection

b1. Select input data

Raw gridded CMIP6 model output of sea surface height (zos): hist-1950 and historical
(Section 2.2)

Raw gridded observation reanalysis product of sea surface height (zos)
(Section 2.2)

Raw Karenia brevis concentration data
(Section 2.2)

b2. Subset selection

Composition of multi-model ensembles given scores based on performance ranking of independent model subsets
(Section 2.6)

a5. Score independent model subsets

b3. Process input data

Loop Current position of multi-model ensembles following the method of Maze et al. (2015)
(Section 2.4)

Loop Current position of observation reanalysis data following the method of Maze et al. (2015)
(Section 2.4)

Karenia brevis no bloom or large bloom following the method of Maze et al. (2015)
(Section 2.2)

b4. Define prediction metrics

- *Oscillating event frequency* (y_4)
- *Temporal match error* (y_5)
- *Karenia brevis error* (y_6)
- *Root-mean-square error* (y_7)
(Section 2.5)

b5. Evaluate prediction

Evaluate prediction of multi-model ensembles given defined predictors
(Section 3.2)

Subset Selection

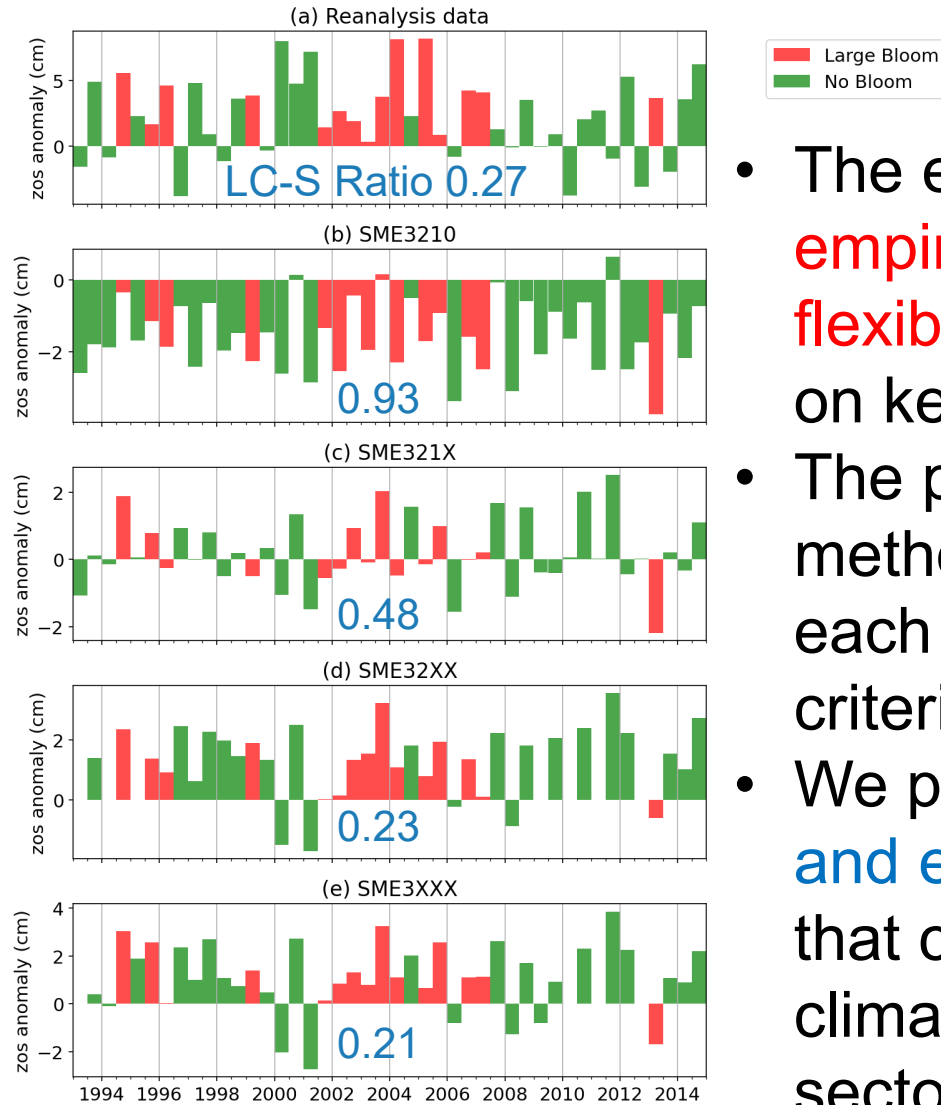
Four metrics to evaluate predictive performance.

Oscillating event frequency
the ratio of the number of a LC south position (LC-S) to the total number of intervals

$$y_4 = \frac{\sum_{t=1}^T H_{LC-S}(h_t)}{T}$$

The LC-S ratio is 0.27 for reanalysis data.

Using the prescreening-based subset selection improves the simulation of Loop Current.



Ensemble including all models
(SME3210)

Excluding models based on prior
information (SME321X)

Non-representative models are
additionally excluded based on
prescreening (SME32XX)

Only representative and skillful
models are included based on
prescreening (SME3XXX)

- The ensemble prescreening is **empirical, but practical and flexible** for using prior knowledge on key features of interest.
- The presented subset-selection method is flexible as it scores each model given multiple binary criteria.
- We provides a **straightforward and easy-to-implement approach** that can be used for many climate services in different sectors as needed.

Application-specific optimal model weighting

Model weight w

$$\sum_{k=1}^K w_k = 1$$

$$\min_{w_k} f = \min_{w_k} \left[\prod_{i=1}^5 (x_i + 1)^{c_i} \right]$$

Oscillating event count error

$$x_1 = \left| \sum_{n=1}^N H_{LC-S}(h_n) - \sum_{n=1}^N H_{LC-S}(h_{n,obs}) \right|$$

Minimize the objective function using the **covariance matrix adaptation evolution strategy (CMA-ES)**

Prescreening-based Subset Selection

a. Select input data

Raw gridded CMIP6 and reanalysis data of sea surface height above geoid (zos) with resolution $\leq 25\text{km}$
(Section 2.1)

b. Model independence

Create independent model subsets using institution democracy and ocean grid as a secondary criterion
(Section 2.1)

c. Process input data

Determine Loop Current position using CMIP6 and reanalysis data following the method of Maze et al. (2015)
(Section 2.2)

d. Subset selection

Score ensemble members, and select ensemble members of each ensemble accordingly
(Section 2.3)

Raw *Karenia brevis* cell count data
(Section 2.1)

Determine red tide no bloom or large bloom, following the method of Maze et al. (2015)
(Section 2.1)

Optimal Model Weighting

e. Estimate model weights

Optimization algorithm

Generate and find optimal weights for ensemble members
(Section 2.4)

Processing input data

Determine the Loop Current position using CMIP6 and reanalysis data given model weights following the method of Maze et al. (2015)
(Section 2.2)

Objective Function (OF)

Evaluate the OF that minimizes

- Oscillating event frequency error (x_1)
- Temporal match error (x_2)
- Temporal match error LC-S (x_3)
- Red tide bloom error (x_4)
- Root-mean-square error (x_5)

(Section 2.4)

Predictive Performance Evaluation

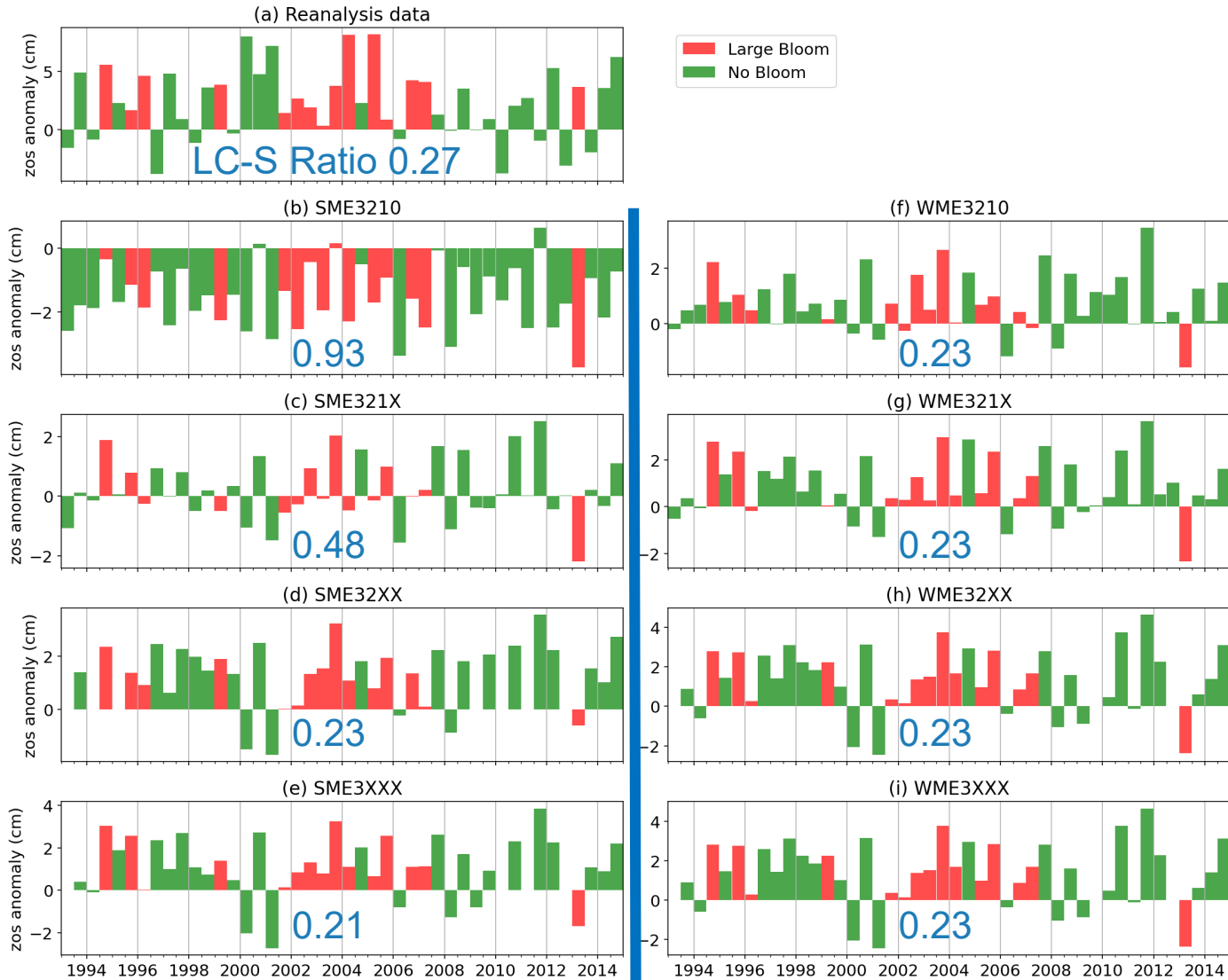
f. Define metrics

- Ensemble performance metrics (x_0 - x_5)
- Ensemble performance and size metrics: AIC (x_6) and BIC (x_7) scores
(Section 2.5)

g. Evaluate prediction

Evaluate model weights and prediction of multi-model ensembles
(Section 3)

While we can use model weighting instead of subset selection, a critical pitfall of model weighting is error cancelation



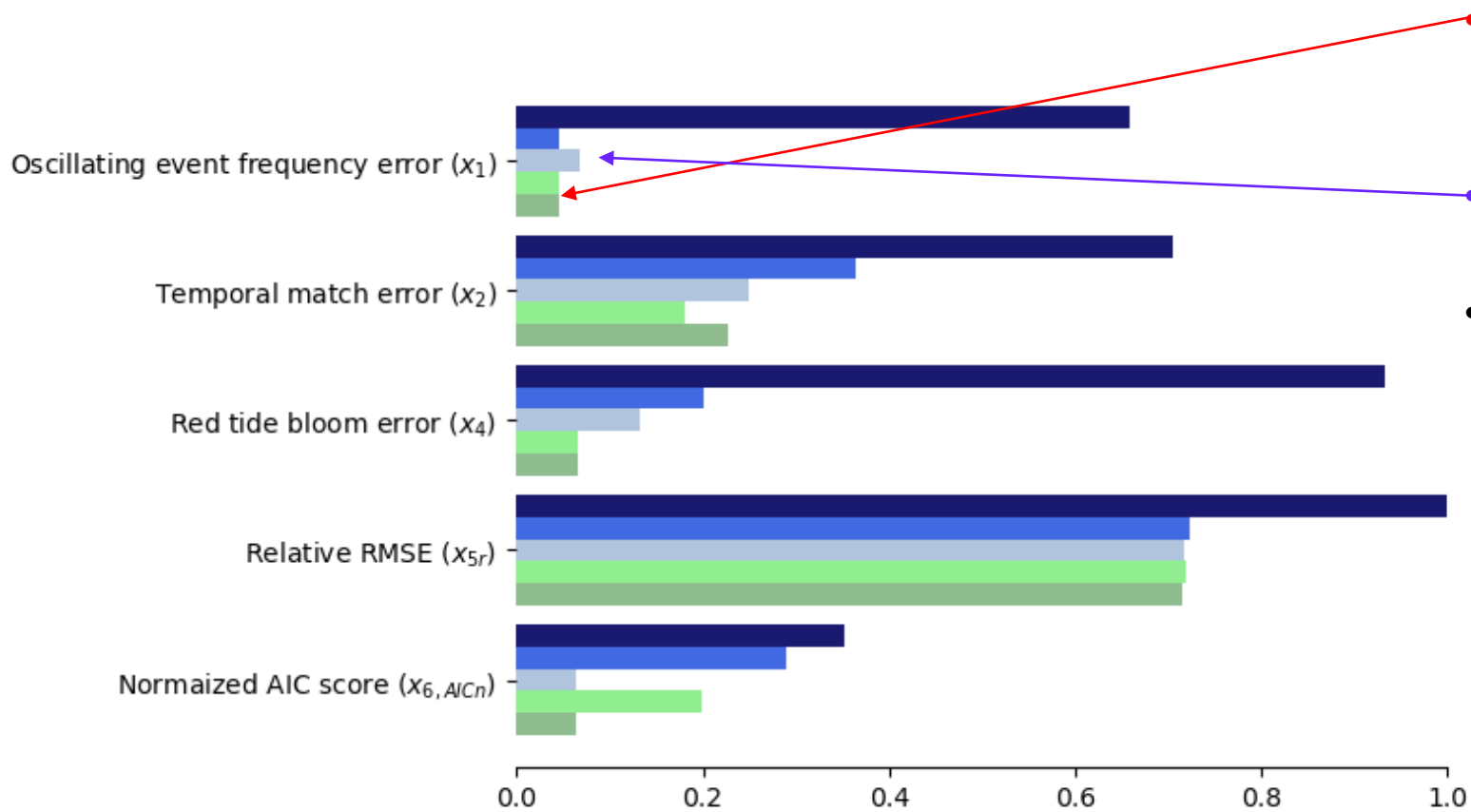
Ensemble including all models
(SME3210)

Excluding models based on prior
information (SME321X)

Non-representative models are
additionally excluded based on
prescreening (SME32XX)

Only representative and skillful
models are included based on
prescreening (SME3XXX)

Application-specific optimal model weighting



Optimal model weighting can improve predictive performance, but be cautious about error cancellation.

Prescreening-based subset selection may be adequate.

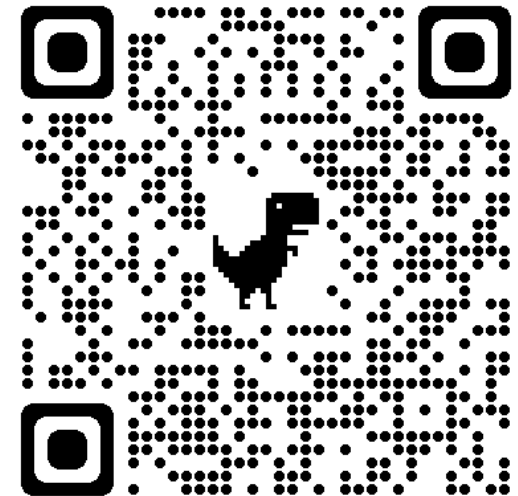
- Practical advantage:
 - ✓ Flexibility in ensemble calibration
 - ✓ Optimization with multiple objectives and multiple metrics
 - ✓ Objectives and metrics can be adaptive to different problems and physically interpretable.

SME3210: (x) Prior information (x) Prescreening-based subset selection (x) Optimal weighting (-) Parsimonious
SME321X: (✓) Prior information (x) Prescreening-based subset selection (x) Optimal weighting (-) Parsimonious
SME3XXX: (✓) Prior information (✓) Prescreening-based subset selection (x) Optimal weighting (-) Parsimonious
WME32XX: (✓) Prior information (✓) Prescreening-based subset selection (✓) Optimal weighting (x) Parsimonious
WME3XXX: (✓) Prior information (✓) Prescreening-based subset selection (✓) Optimal weighting (✓) Parsimonious

Selected results (off-line or maybe online in the future) of Earth System Models can be used directly for regional problems, not red tide yet.

Prescreening-based subset selection is useful for developing an application-specific ensemble given a regional phenomenon.

Questions?



Acknowledgment:
NSF Award #193999