Tropical teleconnections in CESM and reanalyses using network tools

Fabrizio Falasca, Annalisa Bracco,
Athanasios Nenes,
Ilias Fountalis, Constantine Dovrolis

Georgia Institute of Technology

Outline

- Motivations
- Proposed approach: δ-Maps
- Robustness on reanalysis datasets
- Application on the CESM-LE
- Summary

Climate Networks - Motivations

Application of data mining algorithms and complex network analysis on climate data for

- Insights: allows for looking into local and non-local interactions in climate solving some of the EOF issues
- Useful tools for model intercomparison
- It can be used in climate attribution studies
- And many others!

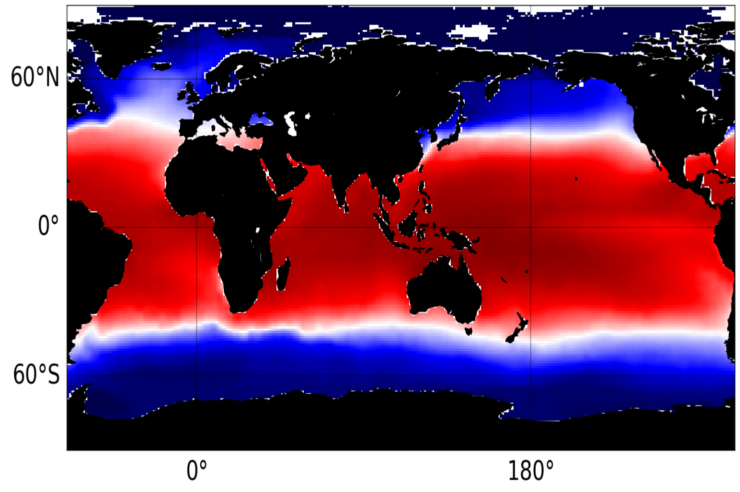
δ–MAPS

Two steps process

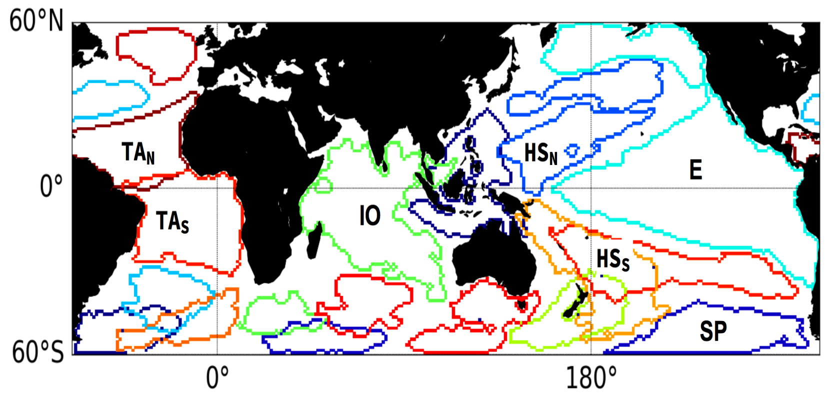
1) Dimensionality reduction

Identification of the functional components (**domains**) of a spatio-temporal system

- Spatially contiguous regions
- Possibly overlapping
- Homogeneous to the underlying variable



(t)



δ-MAPS

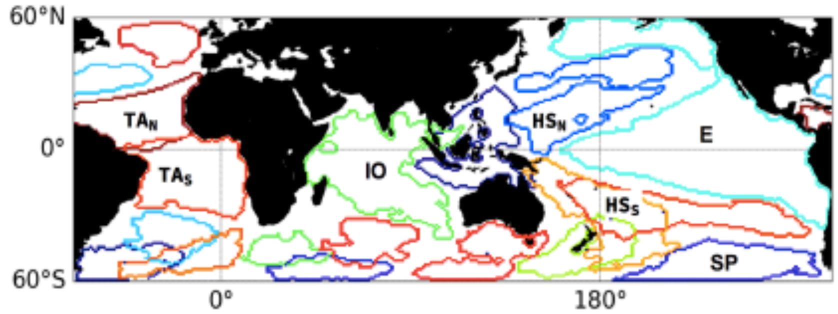
2) Network Inference

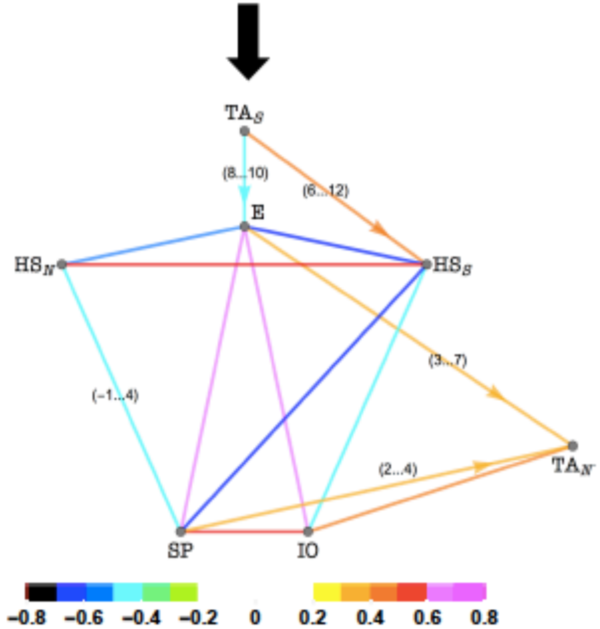
Domains form a network

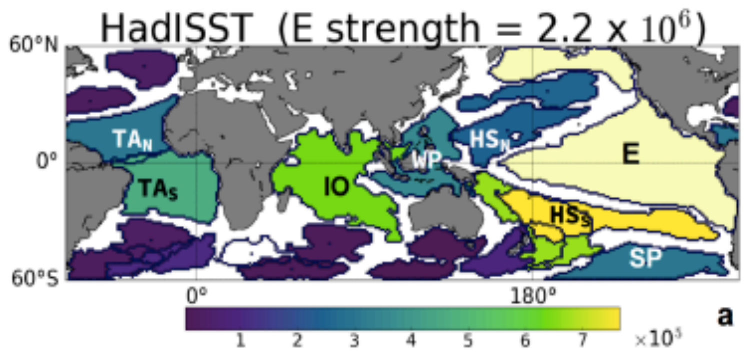
- Every link have a weight that captures the magnitude (covariance) of their interactions

- Lagged interactions are identified: Lag captures the temporal ordering of interactions

$$strength_A = \sum_i |w_i|$$



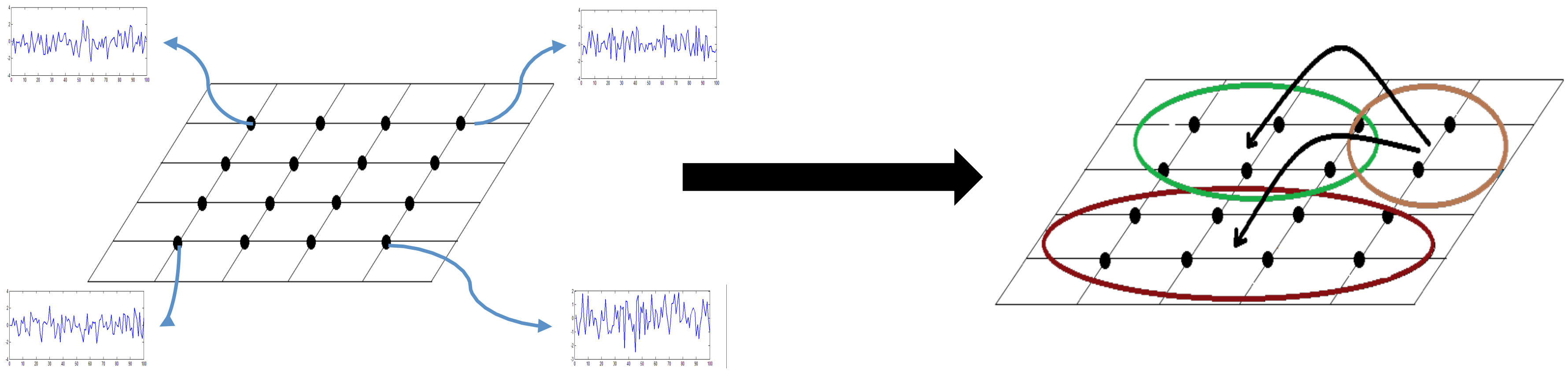




Functional components

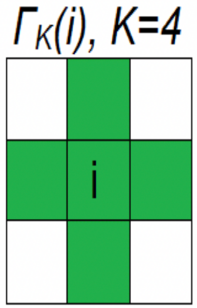
- Major task: identification of the distinct components of the system and the connection between them
- Functional components (domains):

1) Spatially contiguous
2) Functionally homogeneous (the grid cells inside a domain participate in the same dynamic effect)
3) Possible overlap
4) Weighted and lagged interaction



Method: Notation

- Spatio Temporal field **X**(t)
- K-neighborhood $\Gamma_{K(i)}$ of a grid cell
  - K nearest cells to i (including i)



- Local homogeneity of a grid cell i

$$\hat{r}_K(i) = \frac{\sum_{m \neq n \in \Gamma_K(i)} r_{m,n}}{K(K+1)}$$

Where:

$r_{i,j}$ = Pearson correlation between *i* and *j*

$$r_{i,j} = \frac{\sum_{t=1}^{T} (x_i(t) - \tilde{\mu}_i)(x_j(t) - \tilde{\mu}_j)}{T\,\tilde{\sigma}_i \tilde{\sigma}_j}$$

- Homogeneity of a domain

$$\hat{r}(A) = \frac{\sum_{m \neq n \in A} r_{m,n}}{|A|\,(|A| - 1)}$$

Method: Problem Statement

Domain core (epicenter, c): Grid cell at which the local homogeneity is

- a local maximum
- greater than δ

Given an epicenter of action c; given a homogeneity threshold δ:
Domain A is made of grid cells that are spatially contiguous with
$r(A) > \delta$

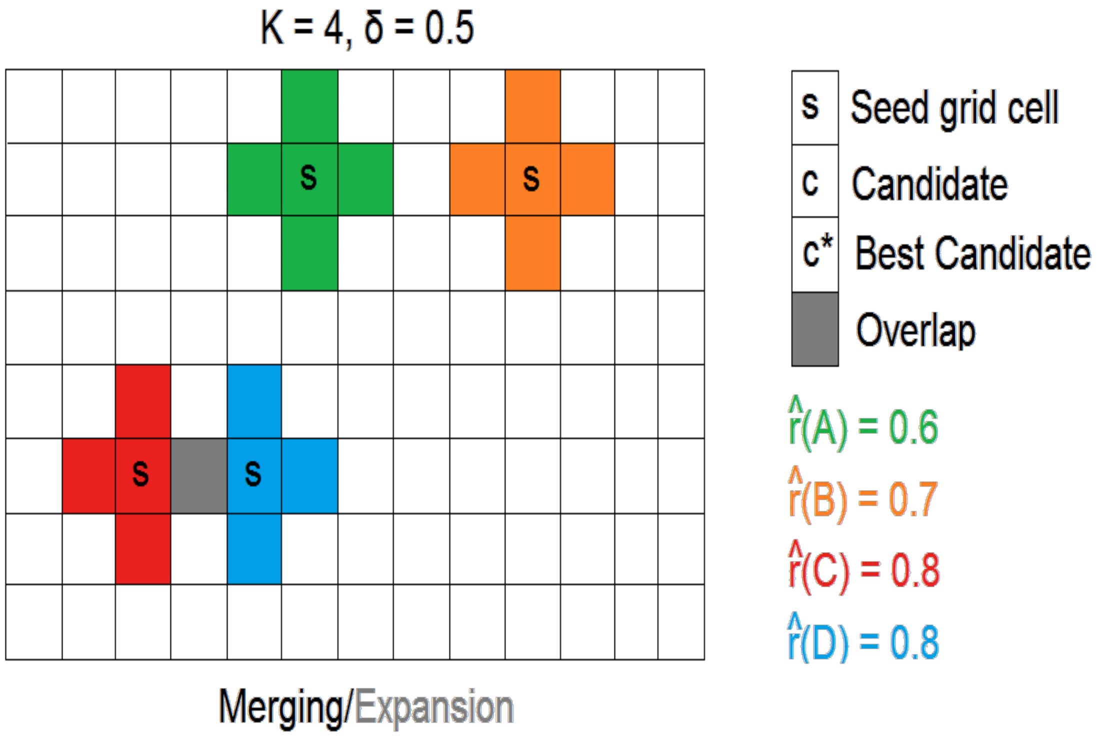
The exact boundaries of a domain are unknown a priori; Domain identification problem. Given a field X, a core cell c and a threshold δ, *A* is the maximum sized set of cells that satisfies $r(A) > \delta$

'Greedy' algorithm (that makes the locally optimal choice at each stage with the hope of finding a global optimum) in two steps:

- Identify all seeds or cores c
- Iteratively expand and merge seeds to identify domains

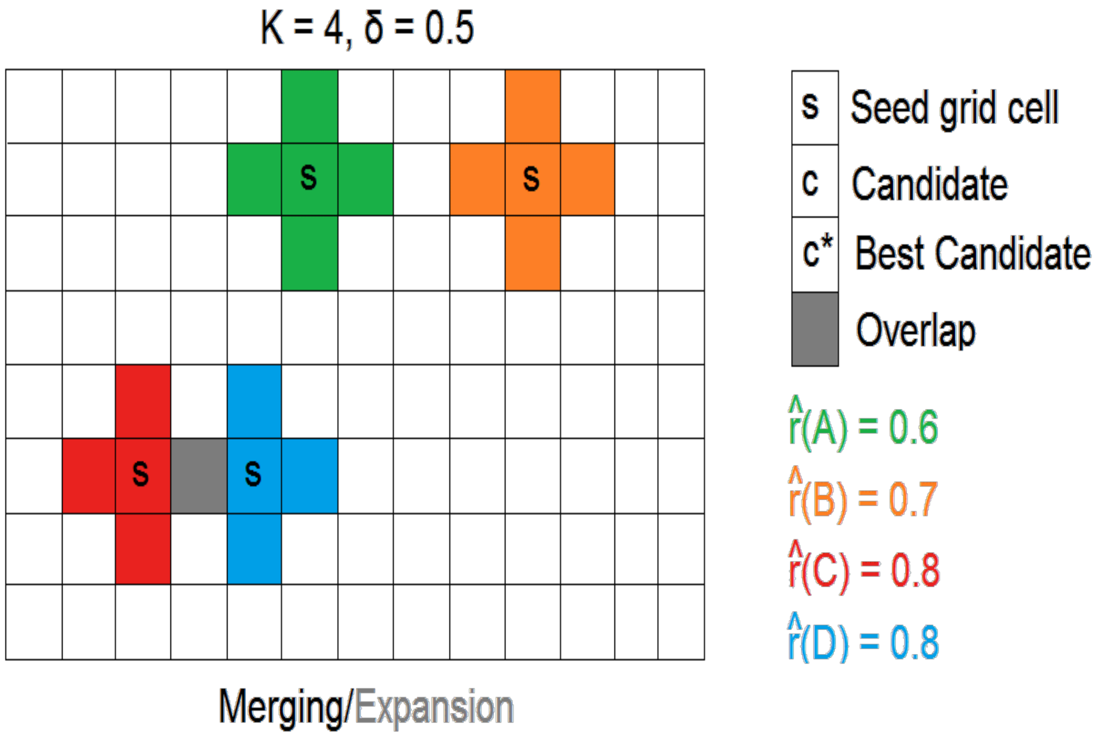
$$c \in A, \quad I_G(A) = 1, \quad \hat{r}(A) > \delta$$

Method: Domain Identification



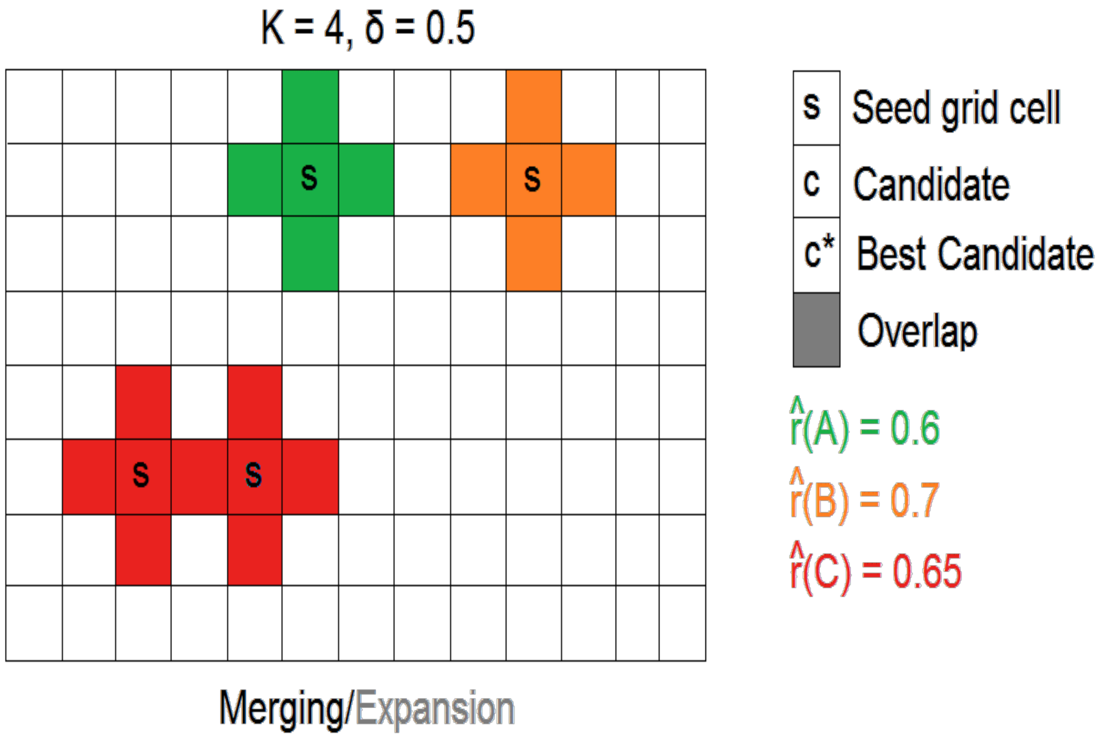
- Input: set of seeds
- Iterative process: merging and expansion

Method: Domain Identification

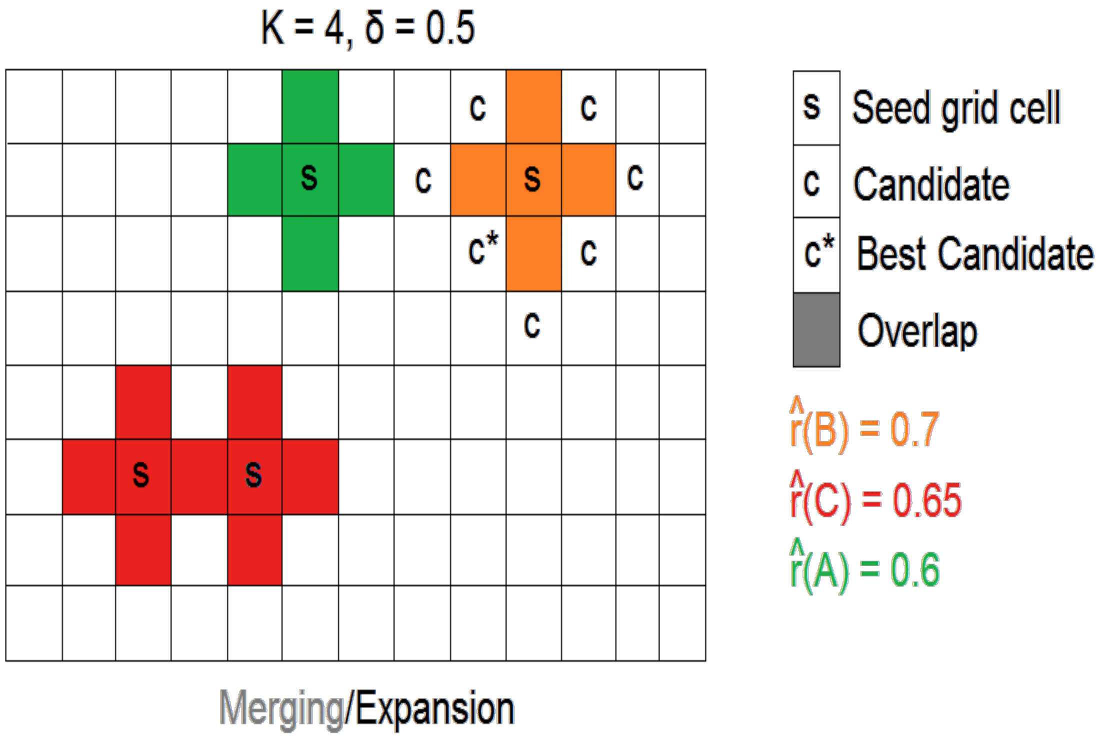


- Merging:
  - Two domains can be merged if
    - Spatially adjacent
    - $r\,(A \cup B) > \delta$
  - Merge first the two domains with max $r\,(A \cup B)$
  - Terminate when no merging is possible

Method: Domain Identification

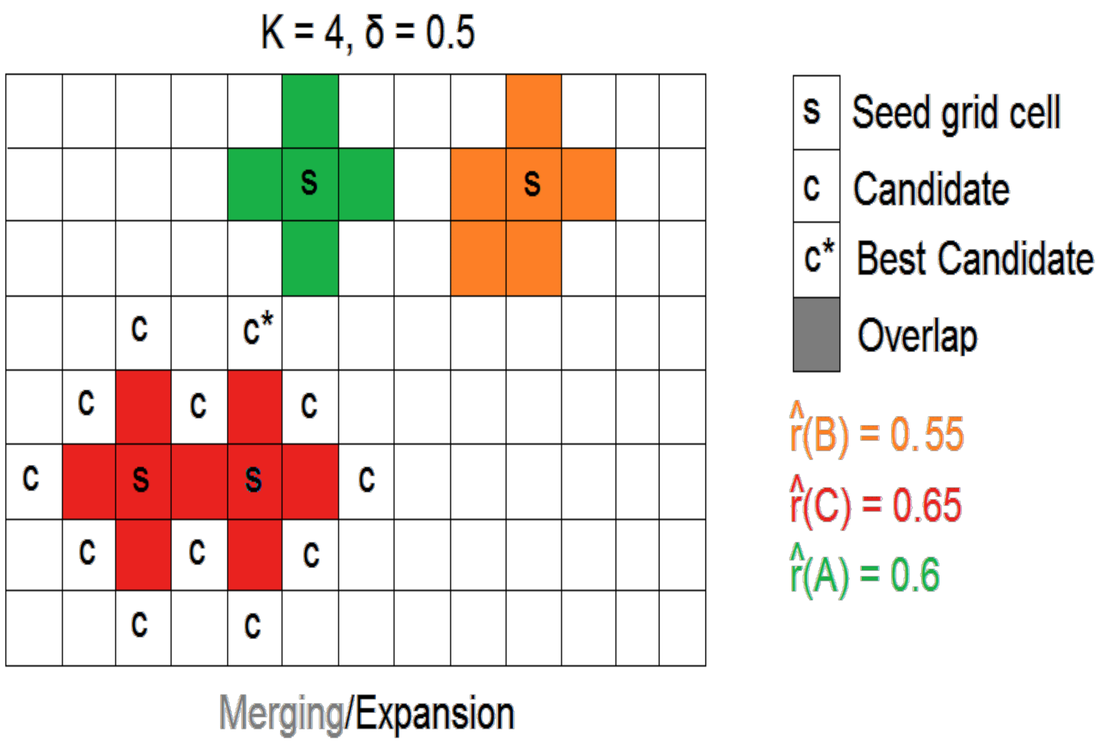


Method: Domain Identification



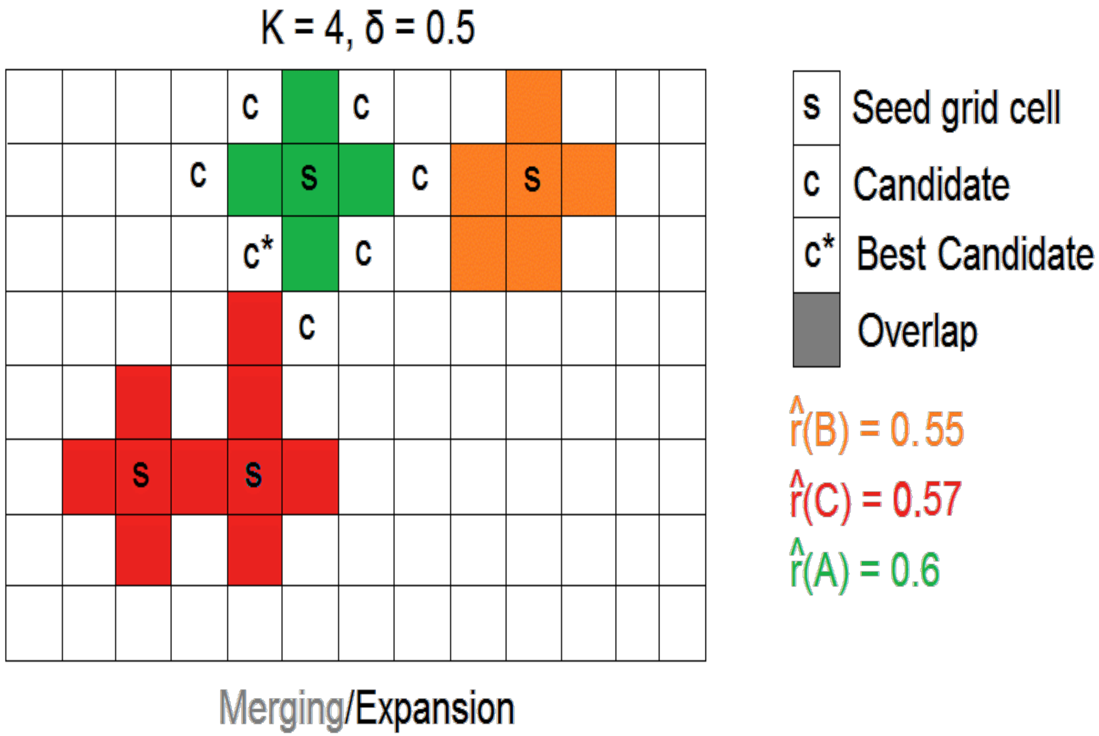
- Expansion:
  - Domains are sorted by homogeneity
  - Expand by considering all adjacent cells
  - Expand by adding grid cell with max $r\ (A \cup \{i\})$
  - After each expansion check if merging is possible

Method: Domain Identification



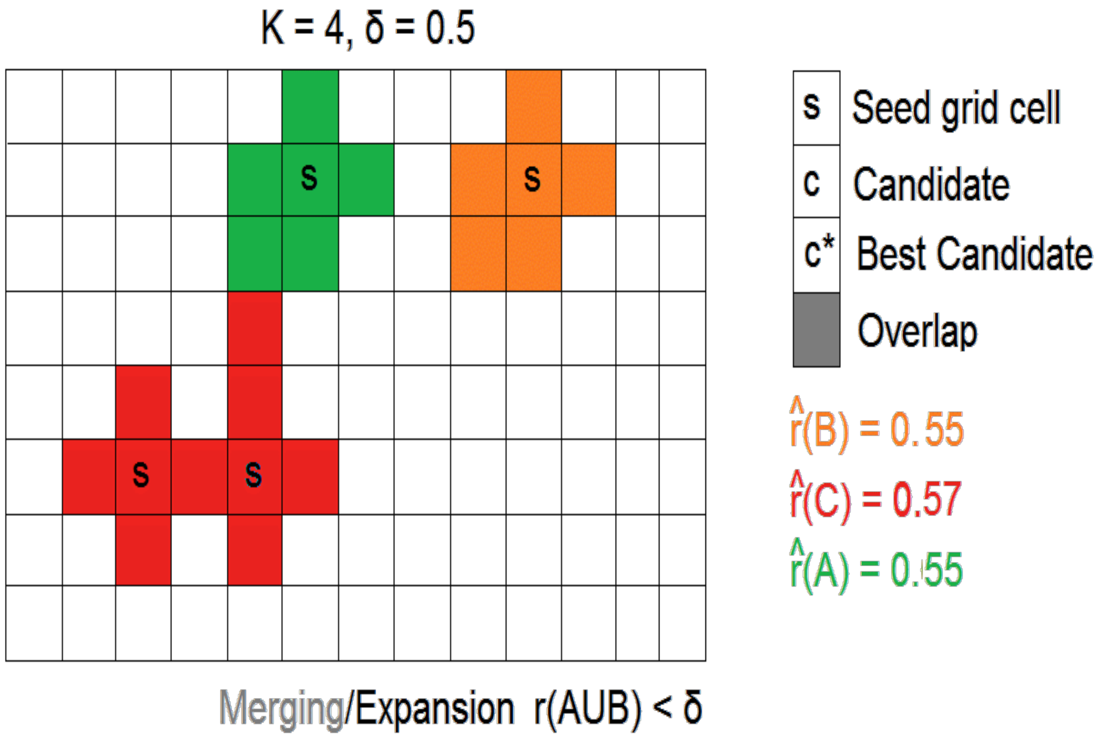
- Expansion:
  - Domains are sorted by homogeneity
  - Expand by considering all adjacent cells
  - Expand by adding grid cell with max $r\,(A \cup \{i\})$
  - After each expansion check if merging is possible

Method: Domain Identification



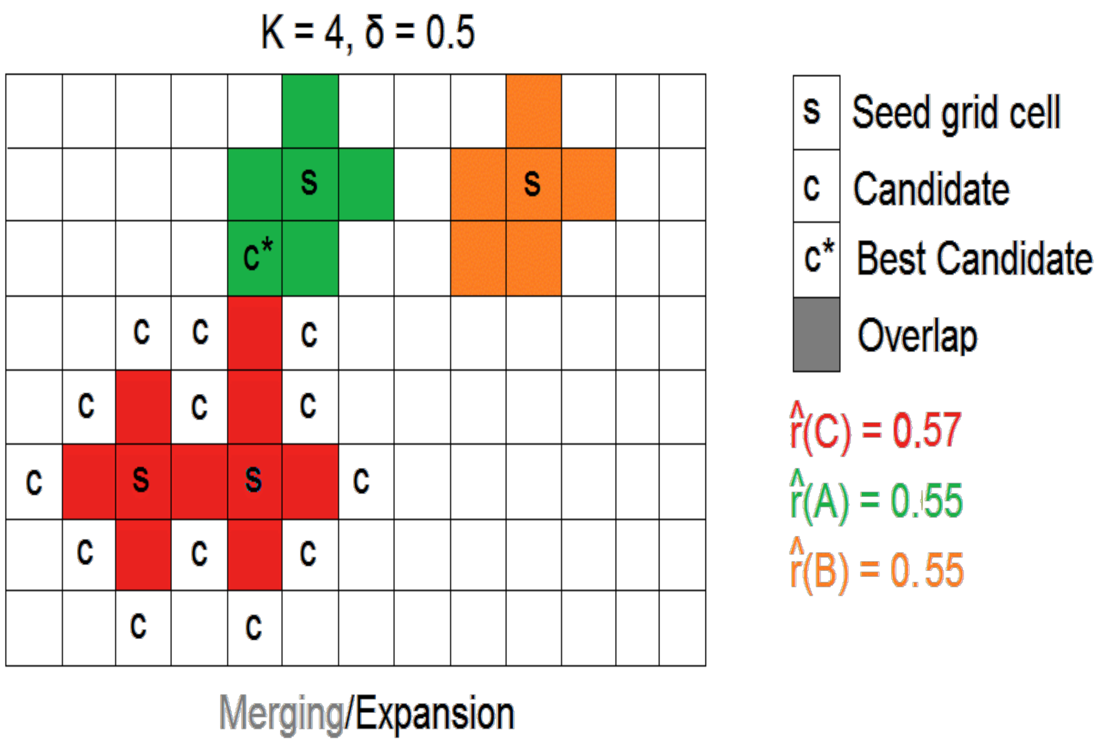
- Expansion:
  - Domains are sorted by homogeneity
  - Expand by considering all adjacent cells
  - Expand by adding grid cell with max $r\,(A \cup \{i\})$
  - After each expansion check if merging is possible

Method: Domain Identification



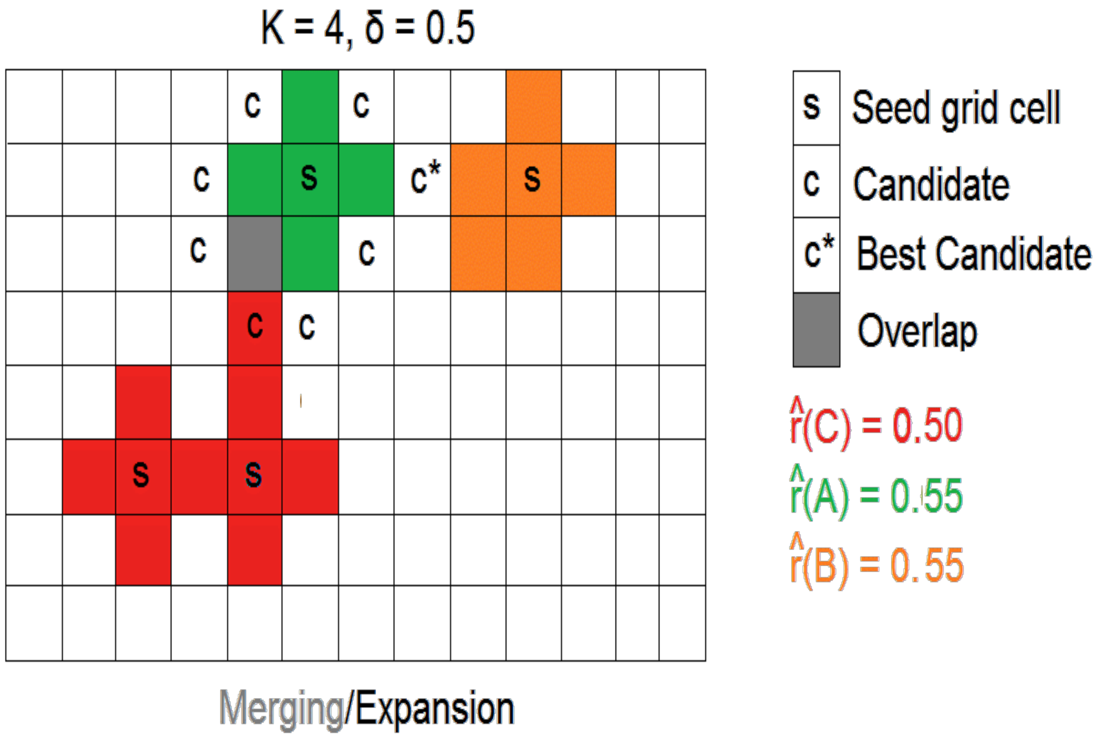
- Expansion:
    - Domains are sorted by homogeneity
    - Expand by considering all adjacent cells
    - Expand by adding grid cell with max $r(A \cup \{i\})$
    - After each expansion check if merging is possible

Method: Domain Identification



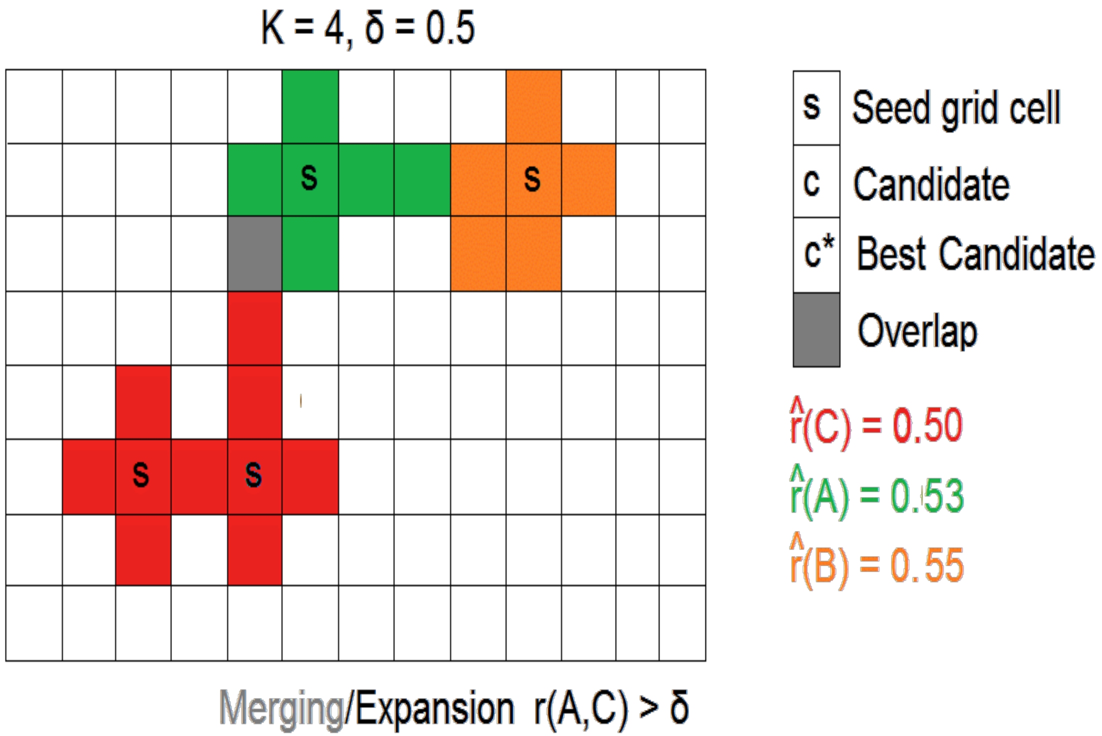
- Expansion:
  - Domains are sorted by homogeneity
  - Expand by considering all adjacent cells
  - Expand by adding grid cell with max $r\,(A \cup \{i\})$
  - After each expansion check if merging is possible

Method: Domain Identification



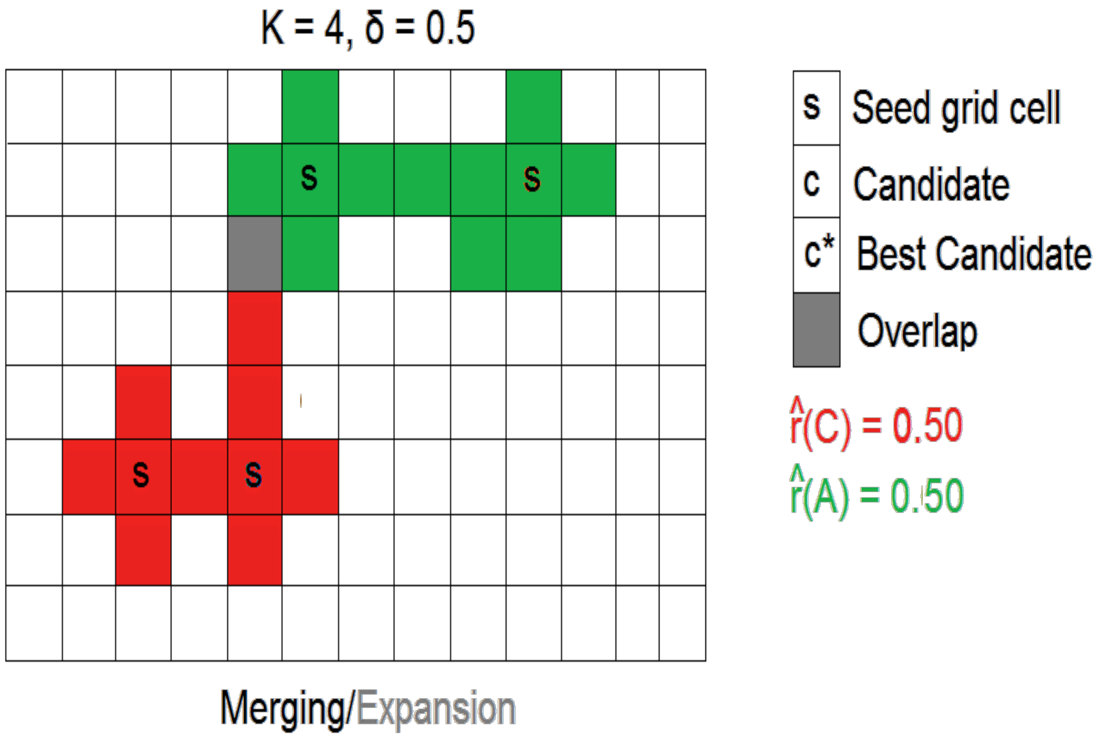
- Expansion:
  - Domains are sorted by homogeneity
  - Expand by considering all adjacent cells
  - Expand by adding grid cell with max $r\ (A \cup \{i\})$
  - After each expansion check if merging is possible

Method: Domain Identification



- Expansion:
  - Domains are sorted by homogeneity
  - Expand by considering all adjacent cells
  - Expand by adding grid cell with max $r\,(A \cup \{i\})$
  - After each expansion check if merging is possible

Method: Domain Identification



- Termination:
    - NO FURTHER MERGING OR EXPANSION IS POSSIBLE

Method: Network Inference

- Domain signal: $X_A(t) = \sum_{i \in A} x_i(t) cos(\phi_i)$

- Functional components might be correlated at a non zero lag
  - For each pair of domains A,B compute Pearson correlation for a range of lags

    $[-\tau_{max} \leq t \leq \tau_{max}]$

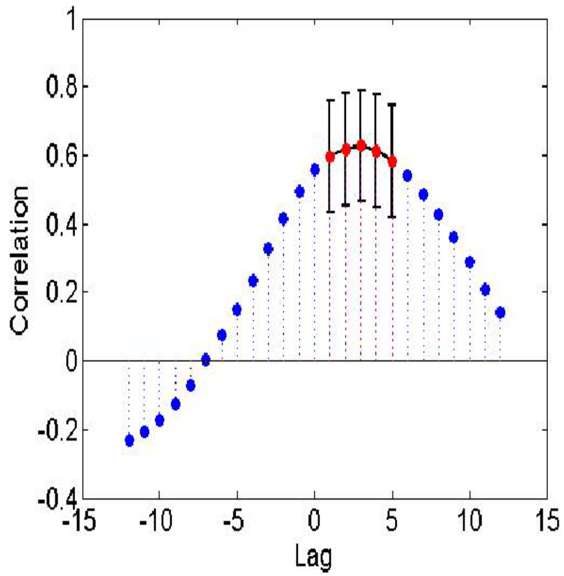
$$r_{A,B}(\tau) = \frac{\sum_{t=1}^{T-\tau} (X_A(t) - \tilde{\mu}_A)(X_B(t+\tau) - \tilde{\mu}_B)}{T \tilde{\sigma}_A \tilde{\sigma}_B}$$

- Statistical significance: uncorrelated signals can produce spurious correlations if they have a strong autocorrelation structure.

BARTLETT'S FORMULA to test significance

Method: Lag Inference

- Domains are connected if there exists at least one significant correlation
- Appropriate lag?
  - Range of lags at which the correlation is significant
  - Located within one standard deviation from the max absolute correlation
  - Catalog edges as undirected or direct (+ or -) depending on lags

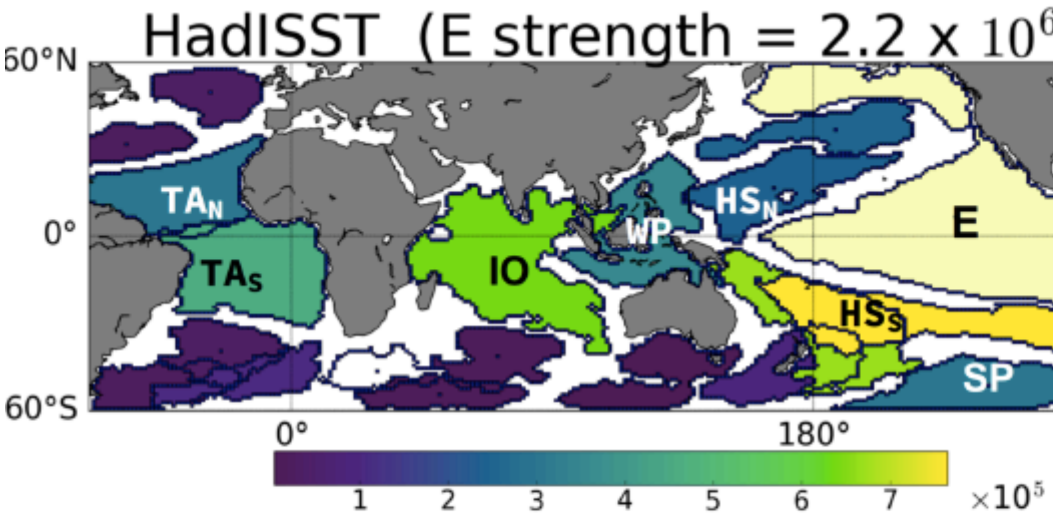


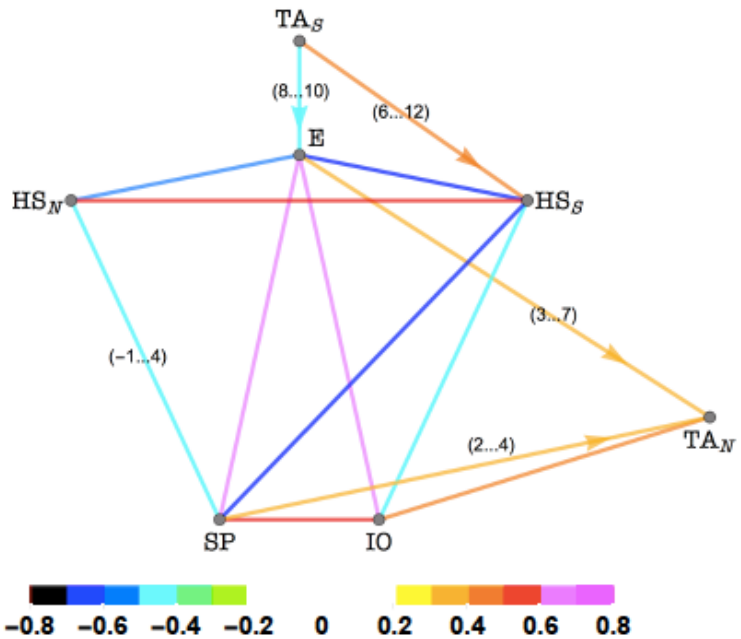
Method: Lag Inference

- One more step: WEIGHTING THE LINKS
- Edge Weight: $w(A, B) = \text{cov}[X_A(t), X_B(t)] = \tilde{\sigma}_A \, \tilde{\sigma}_B \, r^*_{A,B}$
- where
  - r^*_{AB} ı correlation in absolute sense
- Edge weight capture the magnitude of the signal of the two domains
- Weights can be positive or negative
- FINAL NETWORK: weighted and lagged network

δ−MAPS

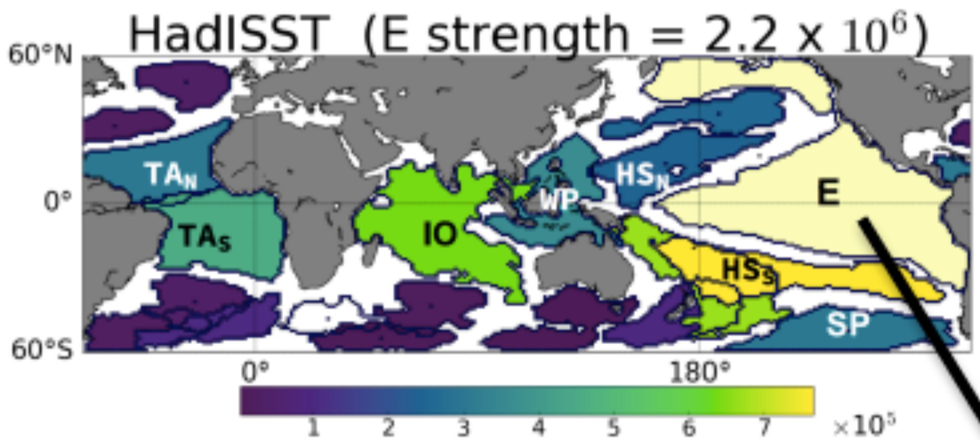
- **Sea Surface Temperature anomalies**
- **Monthly time step**
- **Networks inferred using 45 years period**





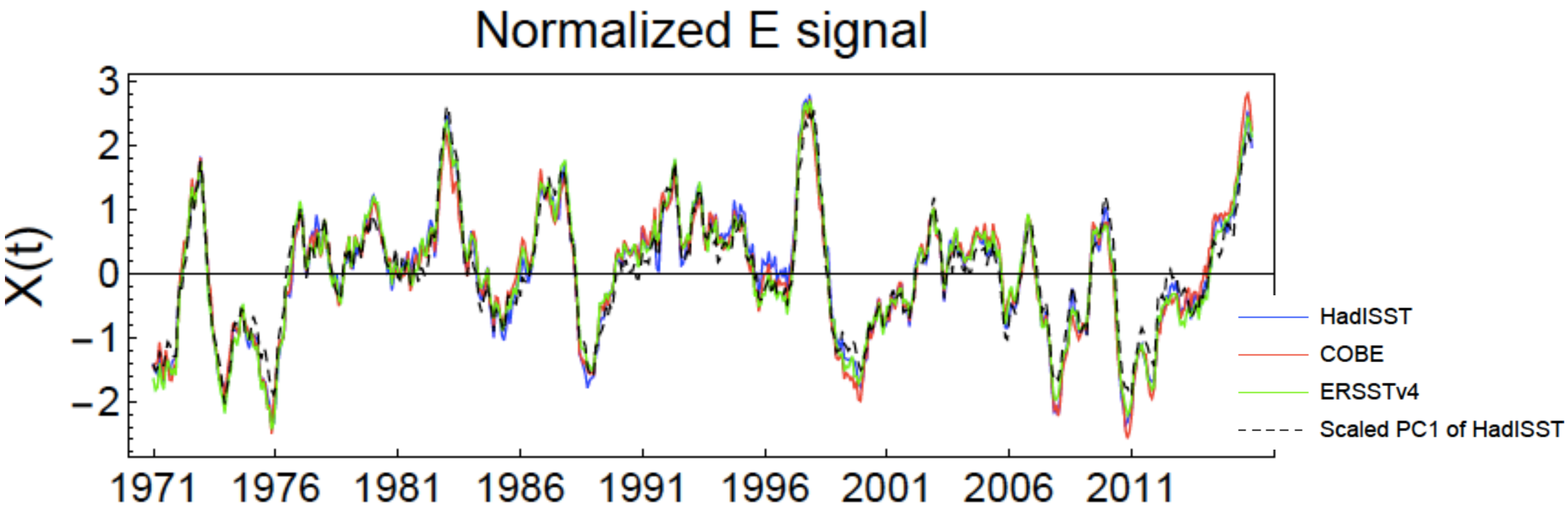
Period considered: 1971-2015

Domains as proxy for climate modes

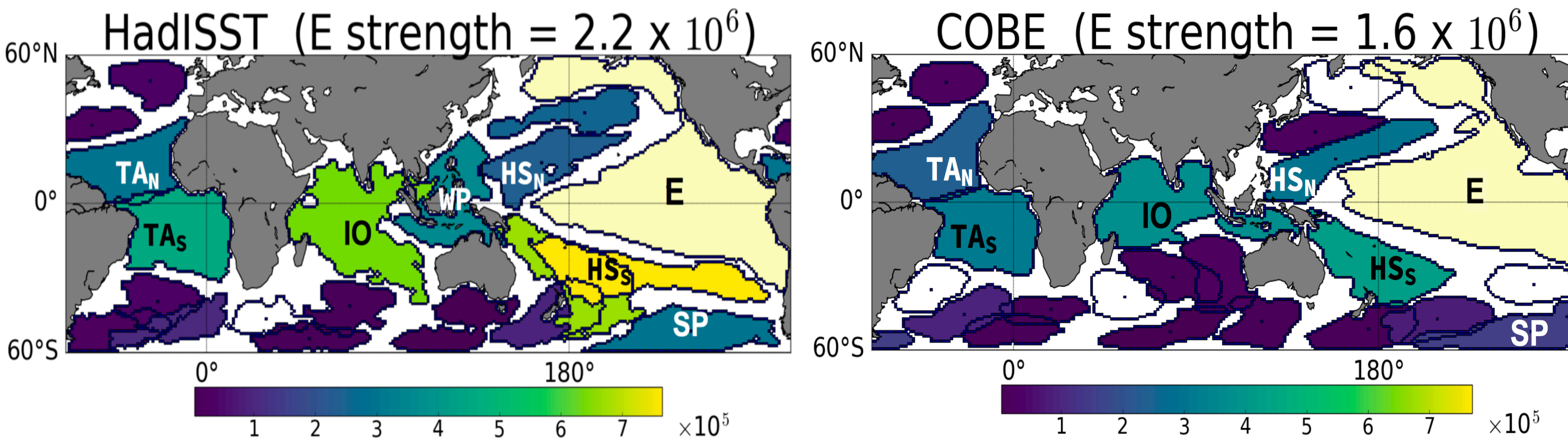


Domain Signal

$$X_E(t) = \sum_{i \in E} x_i(t) \cos(\varphi_i)$$



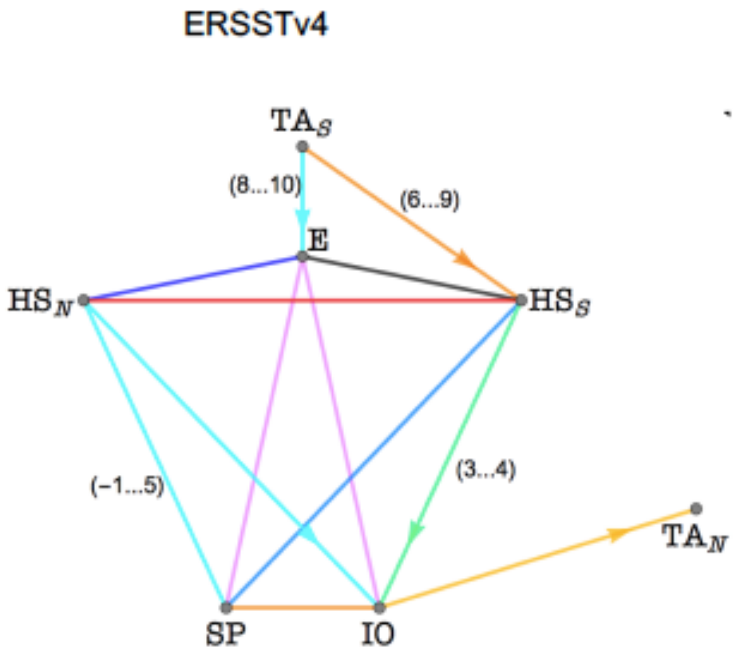
Reanalyses: strength maps

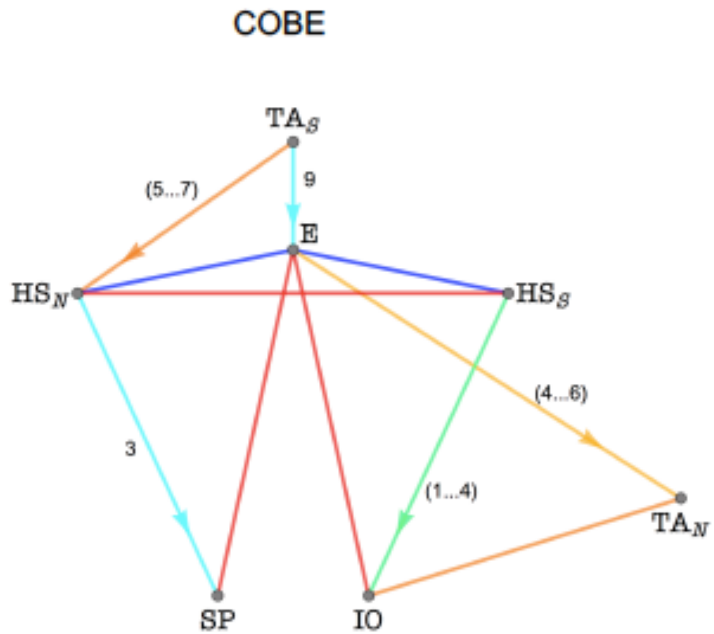


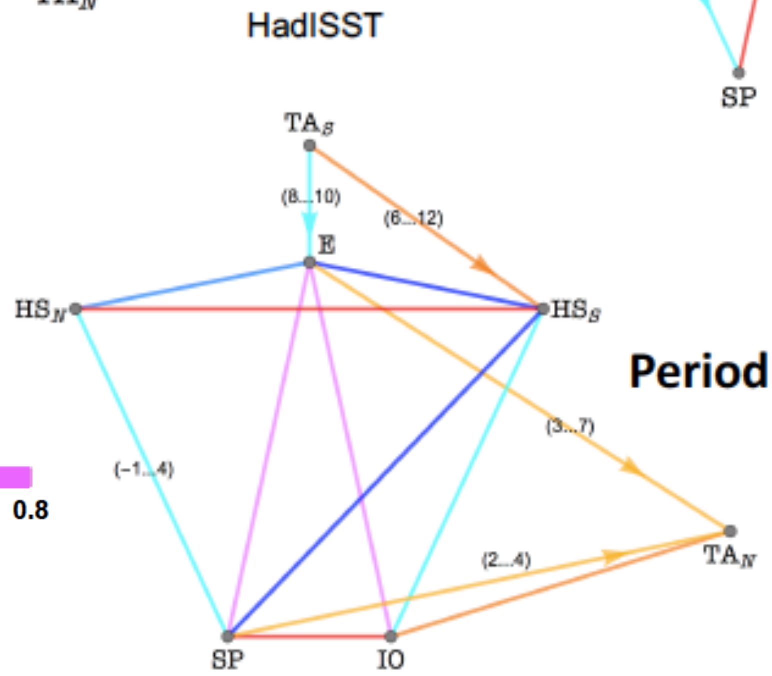
Resolution: 1° X 1.25°

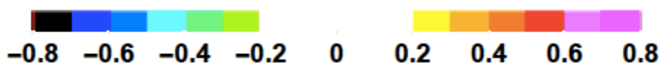
Period considered: 1971-2015

Reanalyses: Networks





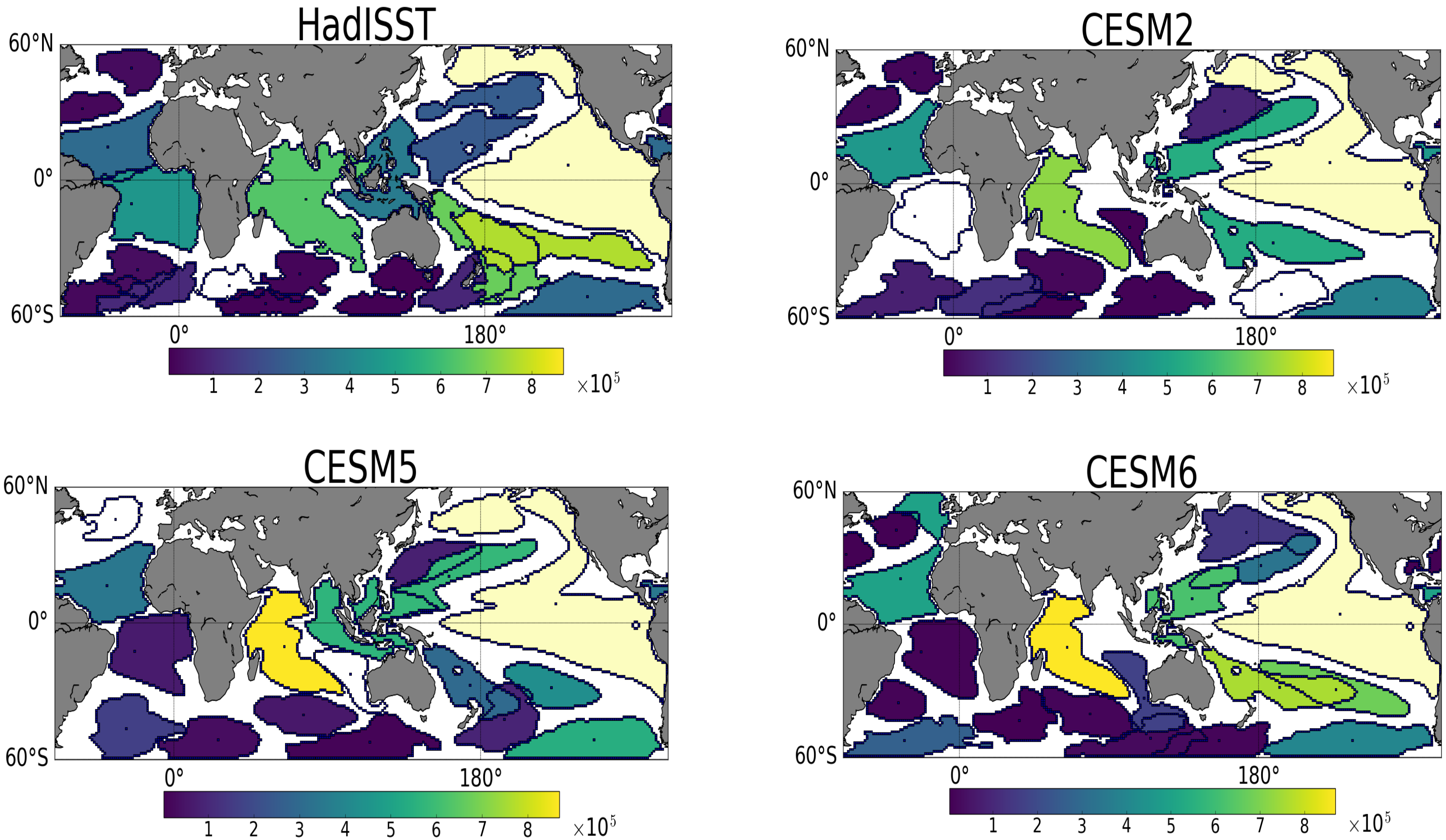




Period considered: 1971-2015

CESM-LE: strength maps

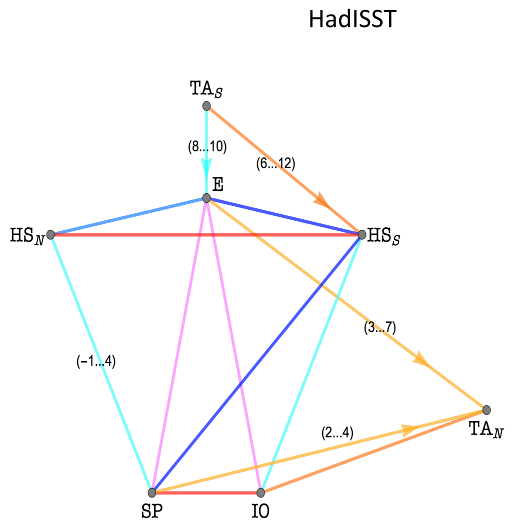
Period considered: 1971-2015

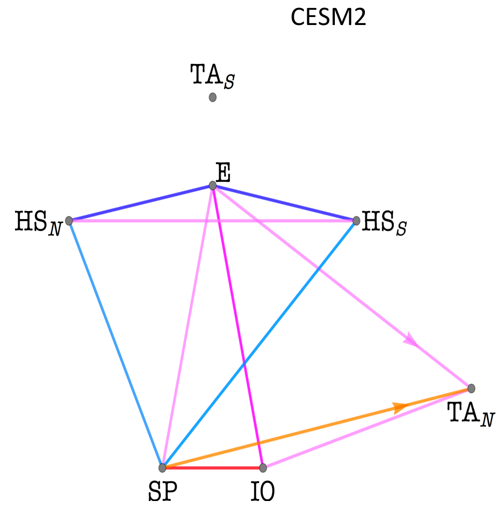


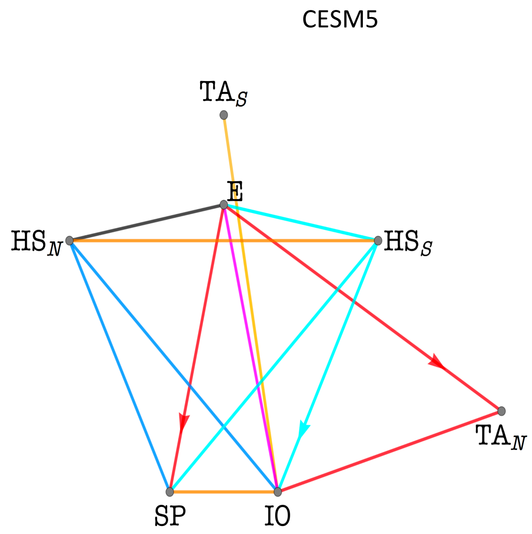
CESM-LE: networks

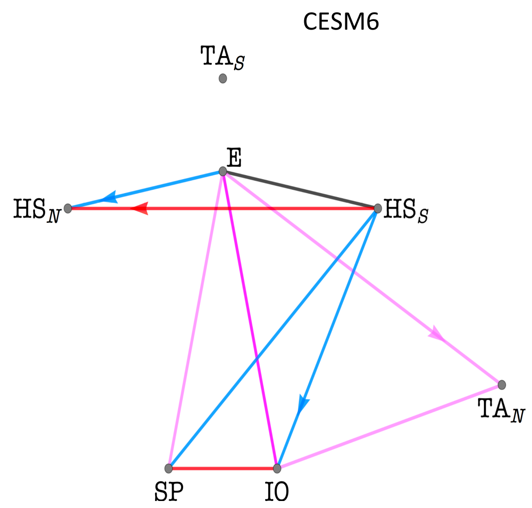
Period considered: 1971-2015

- **Modeled (anti-) correlations too strong**
- **South Tropical Atlantic lead on ENSO never captured**
- **Structure of the network generally well identified**





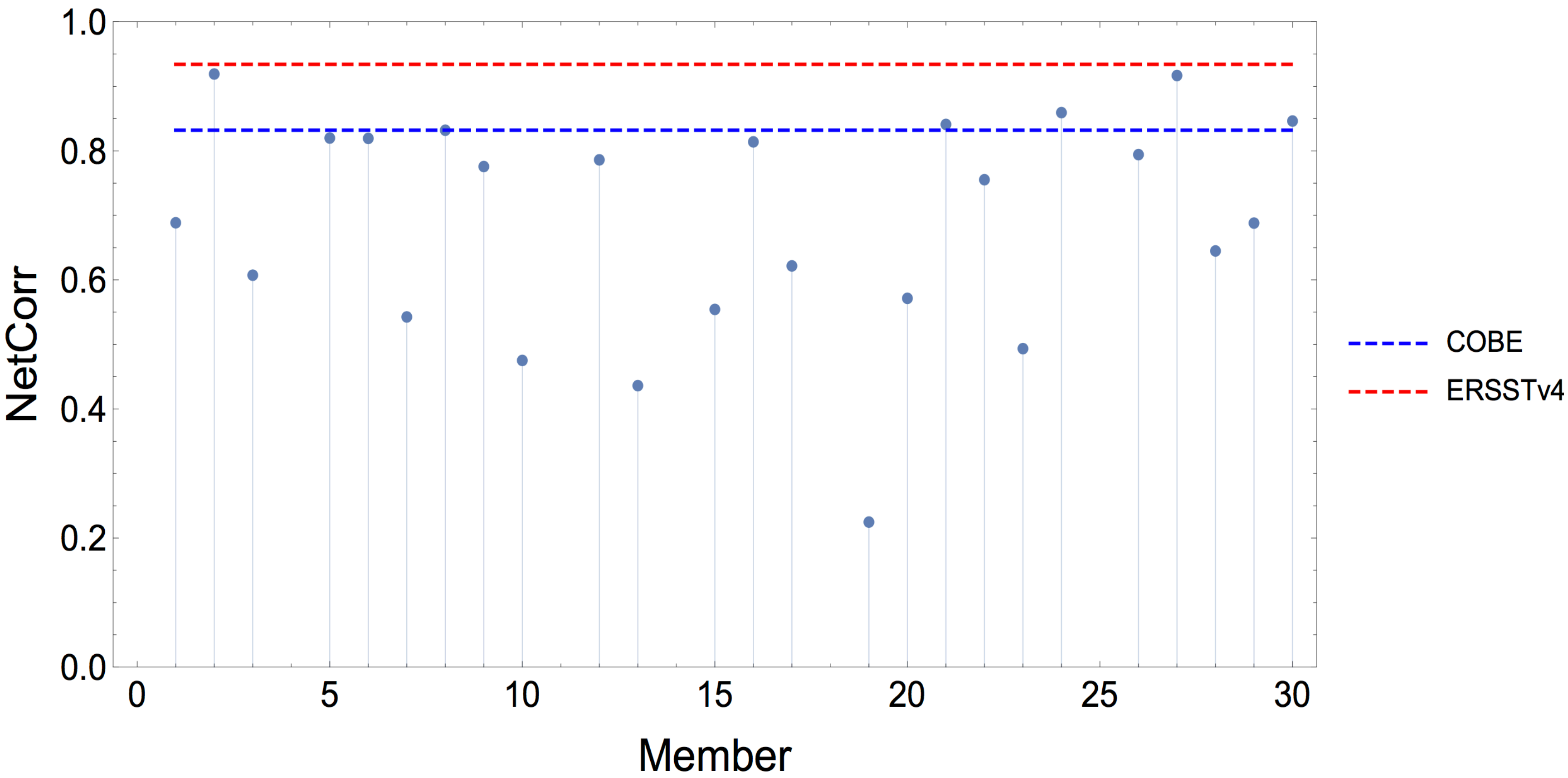




Network Structure

- **NetCorr: Correlation between different network structure (correlation between two adjacency matrix)**
- **Network structure (teleconnections) captured well by many members (despite flaws in Tropical Atlantic areas)**

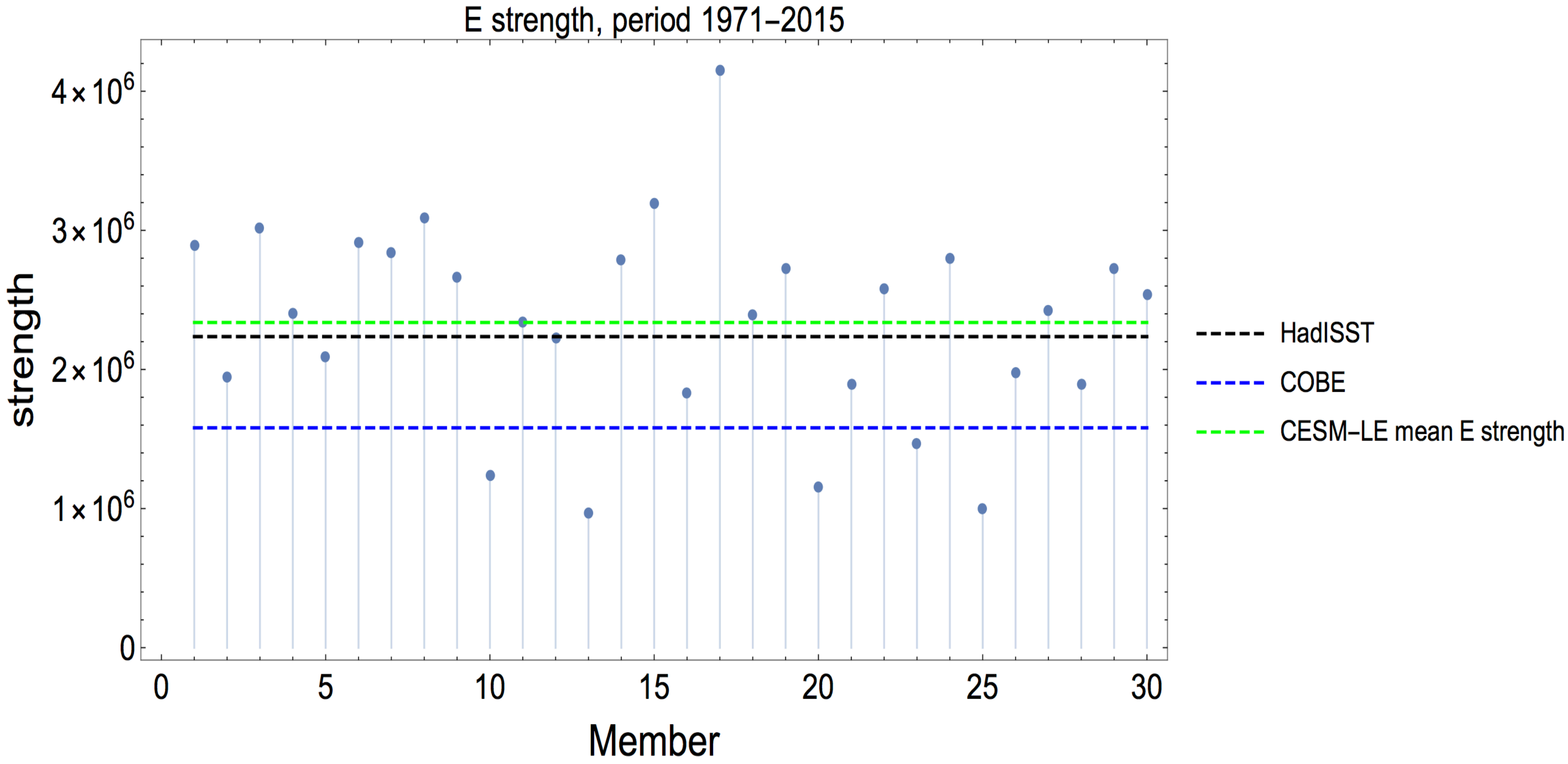
Ground Truth: HadISST



E strength

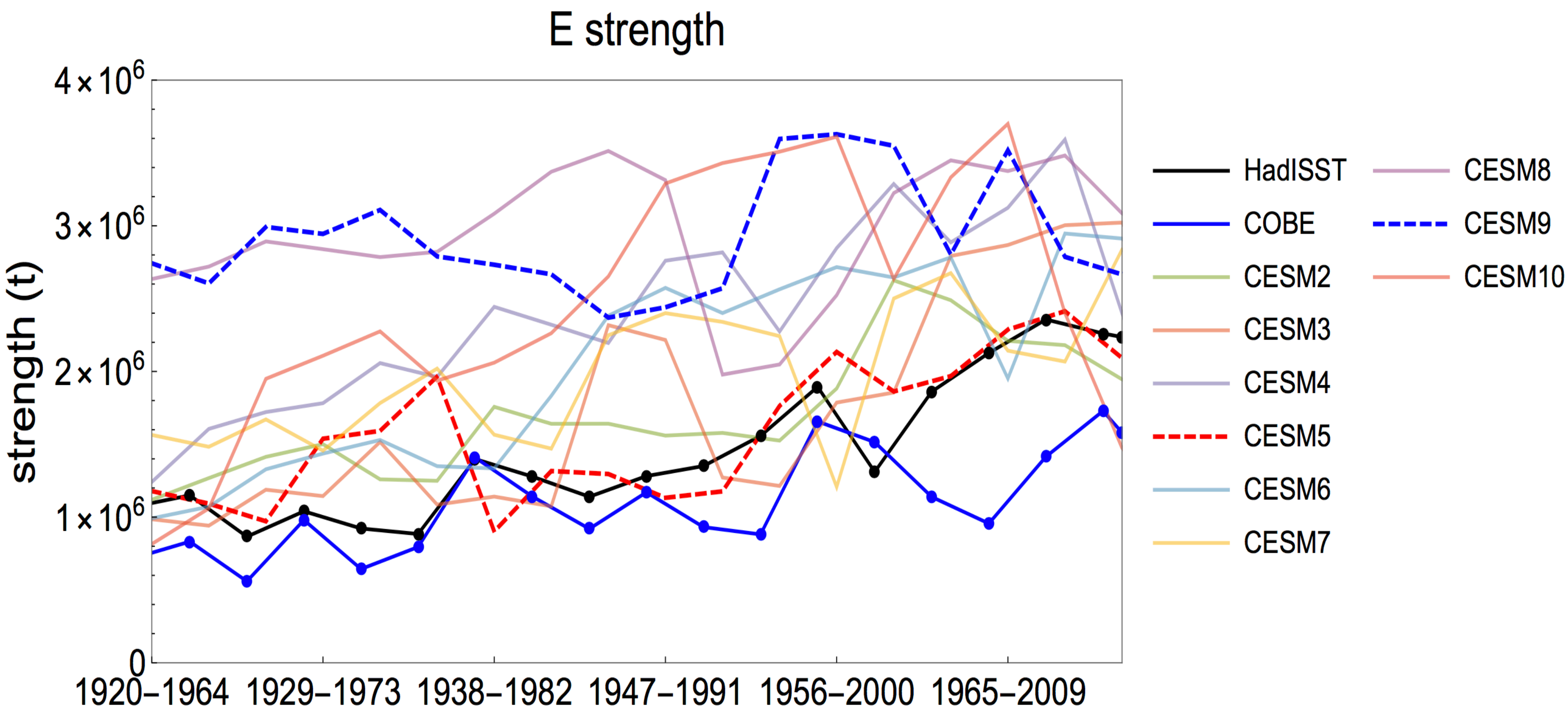
Given all the weights w_i associated with a domain A

$$strength_A = \sum_i |w_i|$$



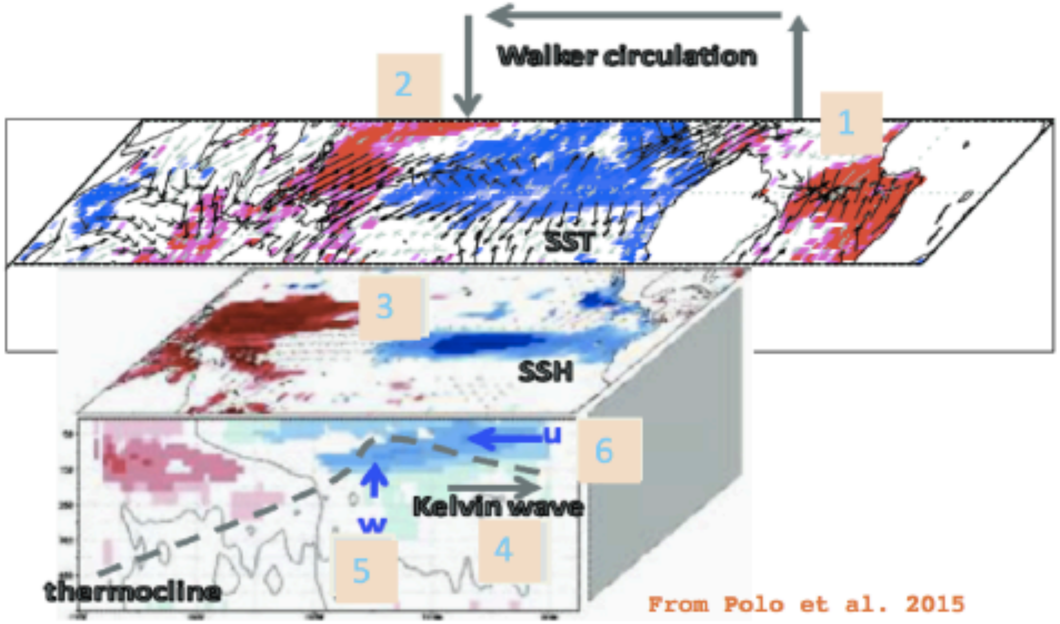
E strength in time

- **Networks are calculated every 3 years over 45 year intervals**
- **First period: 1920-1964**
- **Last period: 1971-2015**



The Atlantic problem

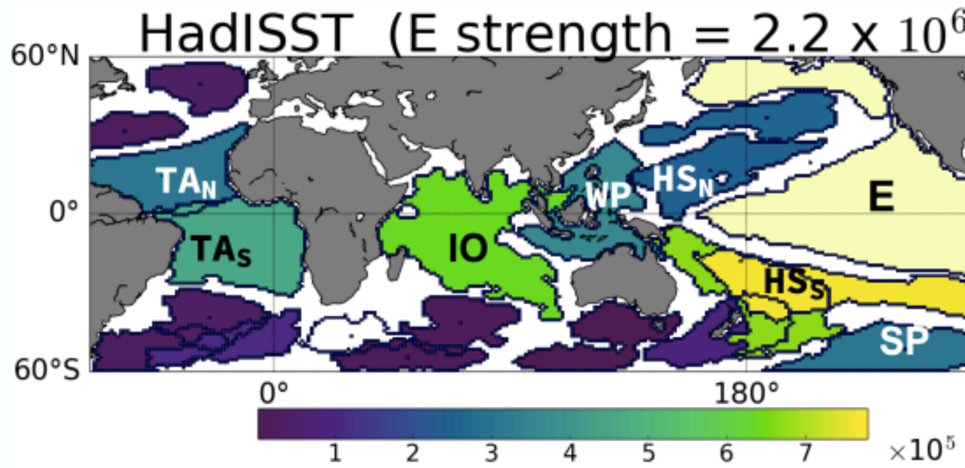
... through an air-sea coupled mechanism

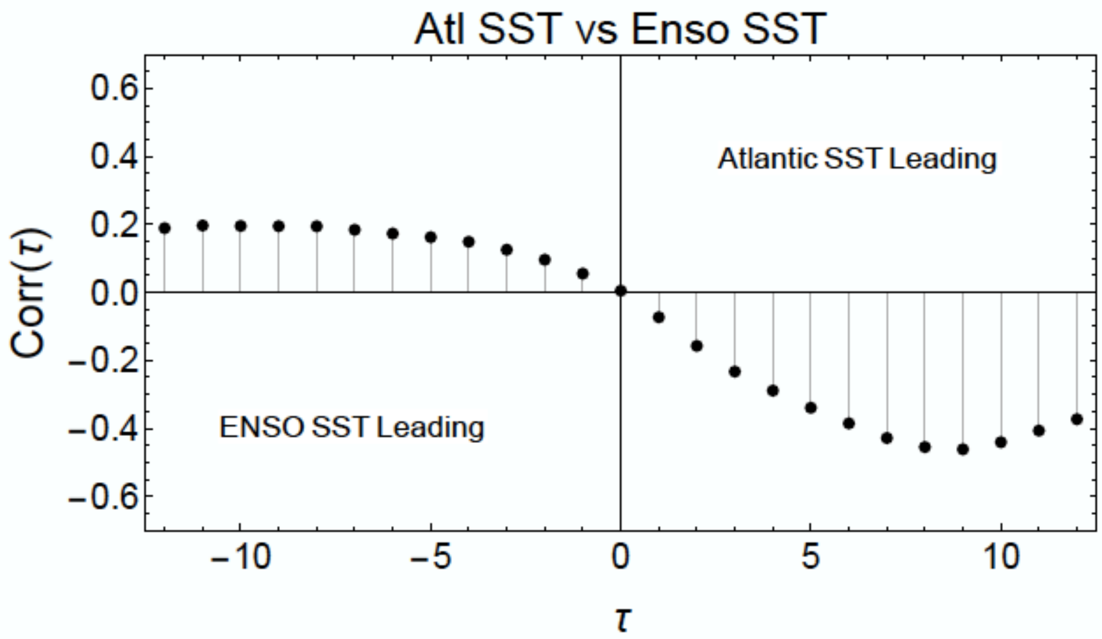


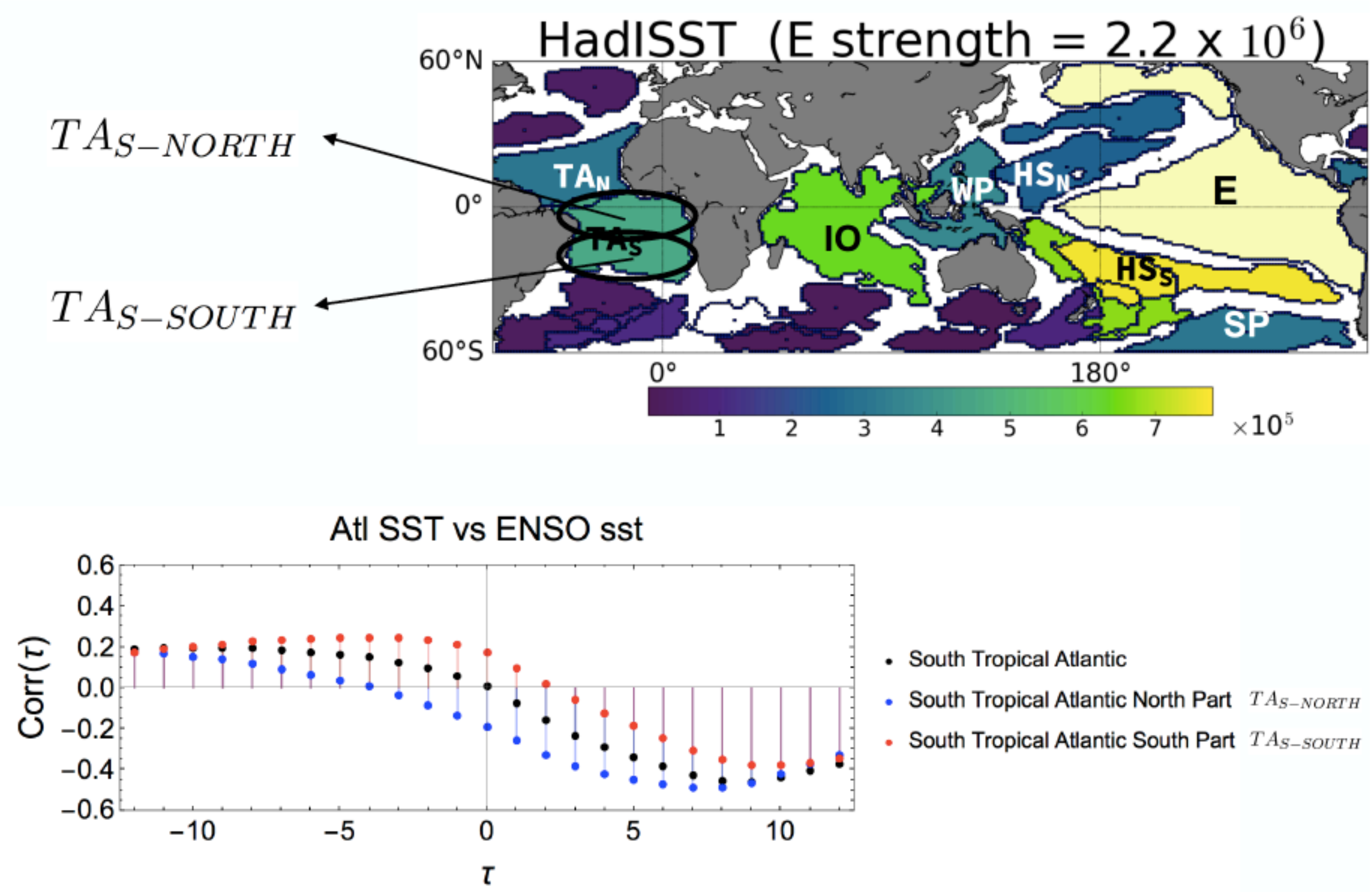
From Polo et al. 2015

1. Atlantic Niño enhances the convection over the Atlantic
2. Alteration of the Walker circulation
3. Surface wind divergence induces westerlies in west Pacific
4. Triggering an equatorial Kelvin wave propagating eastward
5. A Shallower thermocline favours the cooling of the sea surface through anomalous temperature advection by anomalous zonal currents by mean vertical entrainment velocity
6. Reinforcement of the surface winds, establishing the Bjerknes feedback

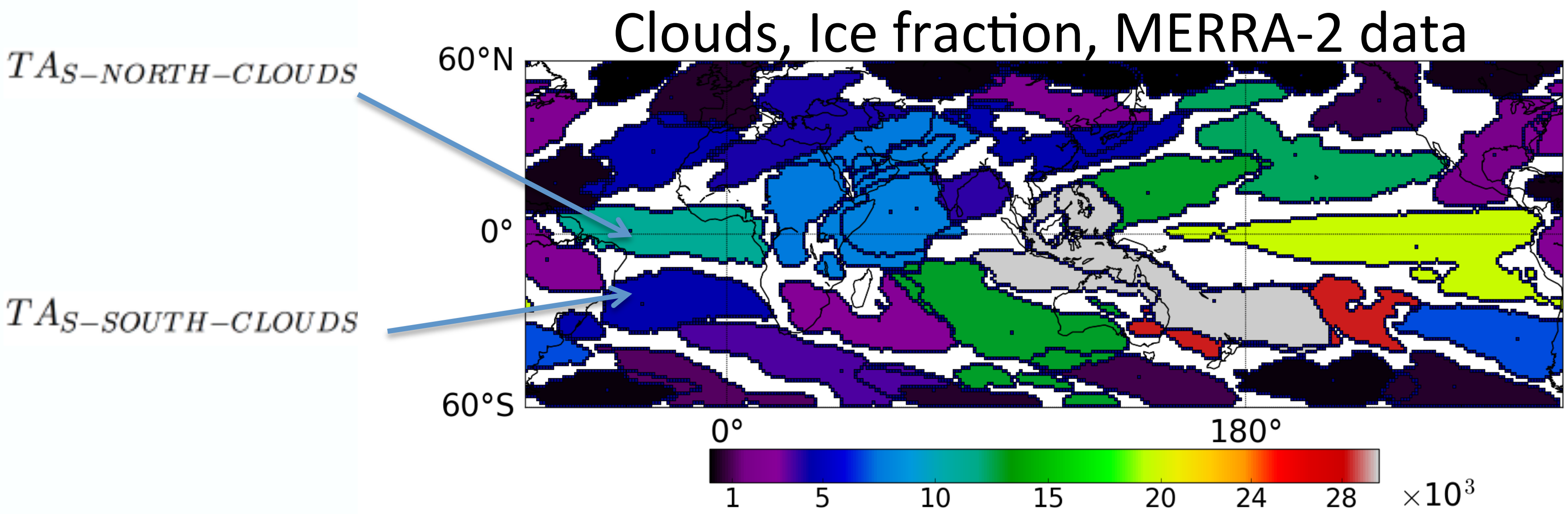
The Atlantic problem



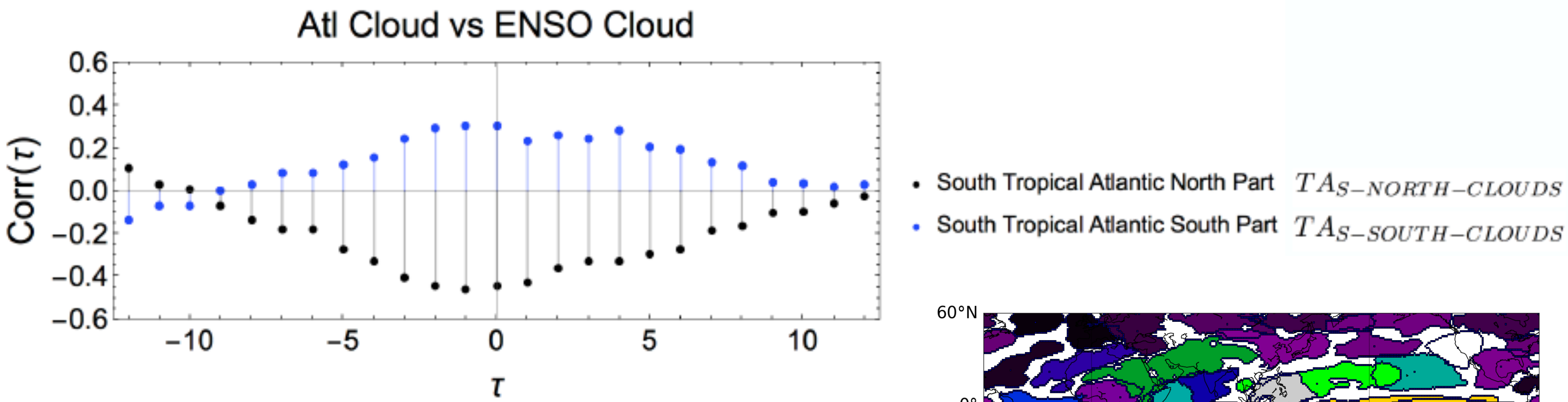




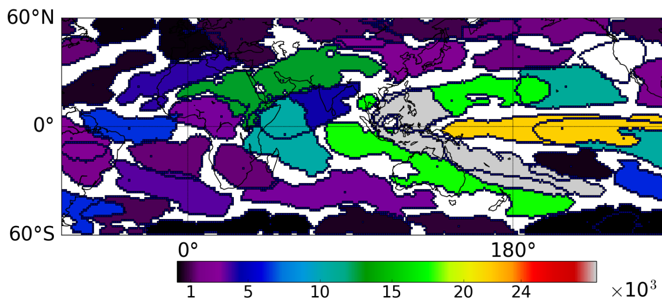
HadISST (E strength = 2.2 x 10^6)
$TA_{S-NORTH}$
$TA_{S-SOUTH}$
60°N
0°
60°S
0°
180°
TA_N
TA_S
IO
WP
HS_N
E
HS_S
SP
1 2 3 4 5 6 7
$\times 10^5$
Atl SST vs ENSO sst
Corr(τ)
0.6 0.4 0.2 0.0 −0.2 −0.4 −0.6
−10 −5 0 5 10
τ
South Tropical Atlantic
South Tropical Atlantic North Part $TA_{S-NORTH}$
South Tropical Atlantic South Part $TA_{S-SOUTH}$



Clouds, Ice fraction, MERRA-2 data
$TA_{S-NORTH-CLOUDS}$
$TA_{S-SOUTH-CLOUDS}$
60°N
0°
60°S
0°
180°
1 5 10 15 20 24 28
$\times 10^3$

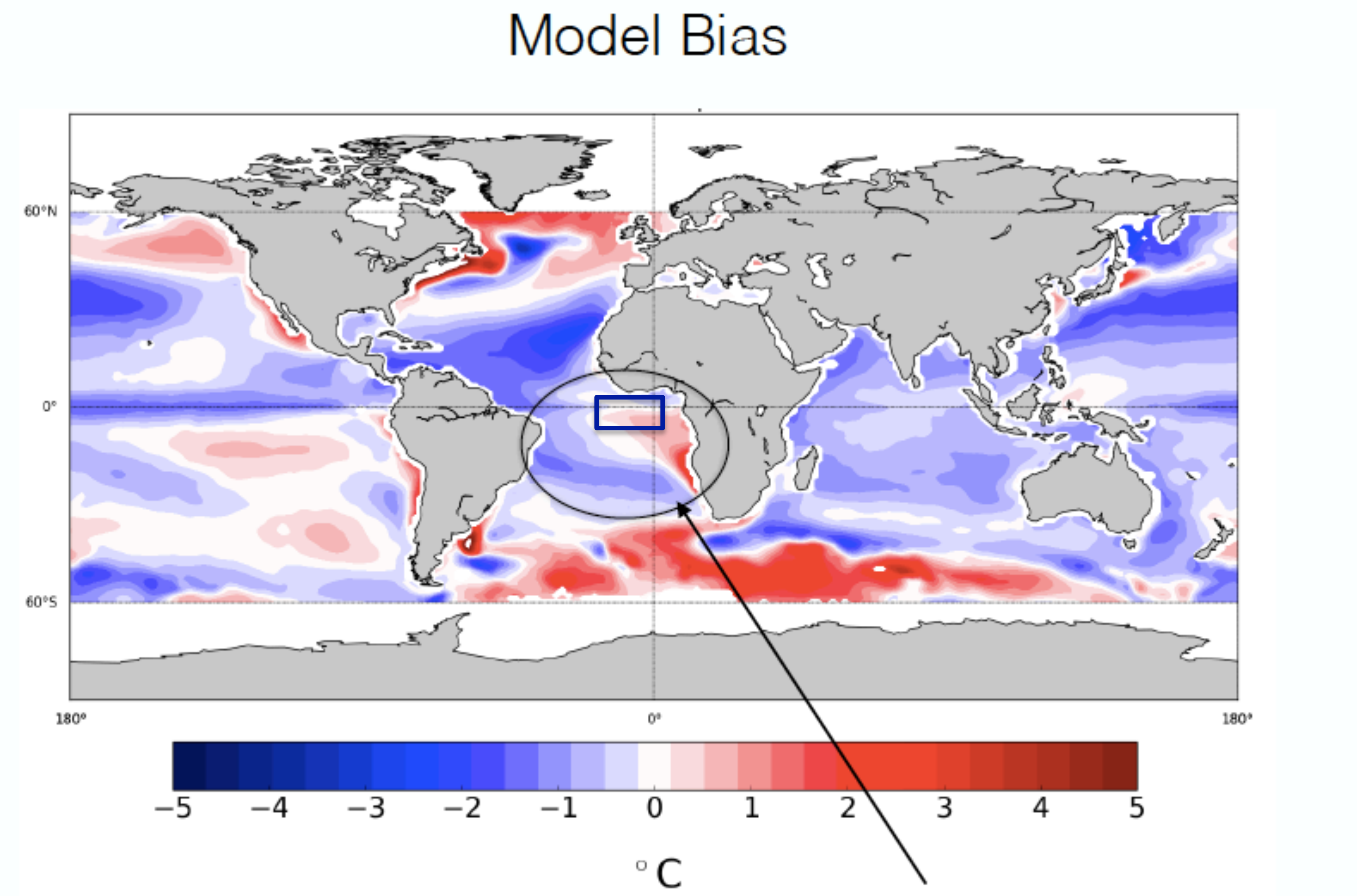


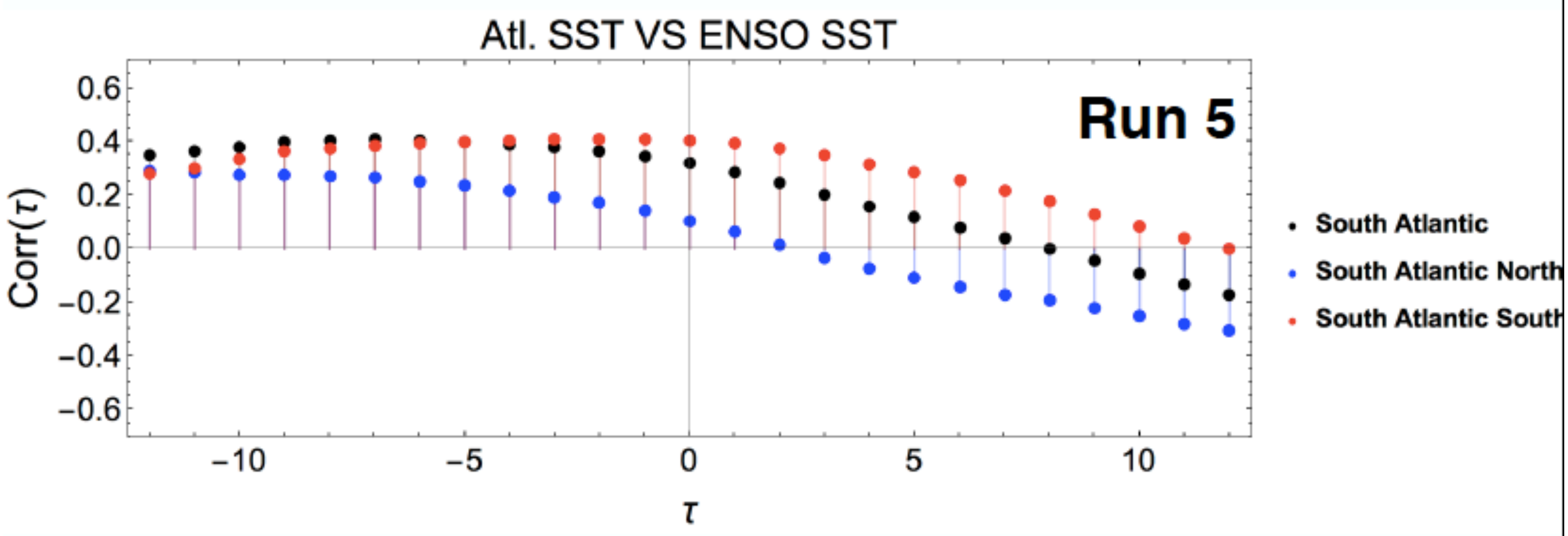
Atl Cloud vs ENSO Cloud
Corr(τ)
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−10
−5
0
5
10
τ
South Tropical Atlantic North Part $TA_{S-NORTH-CLOUDS}$
South Tropical Atlantic South Part $TA_{S-SOUTH-CLOUDS}$

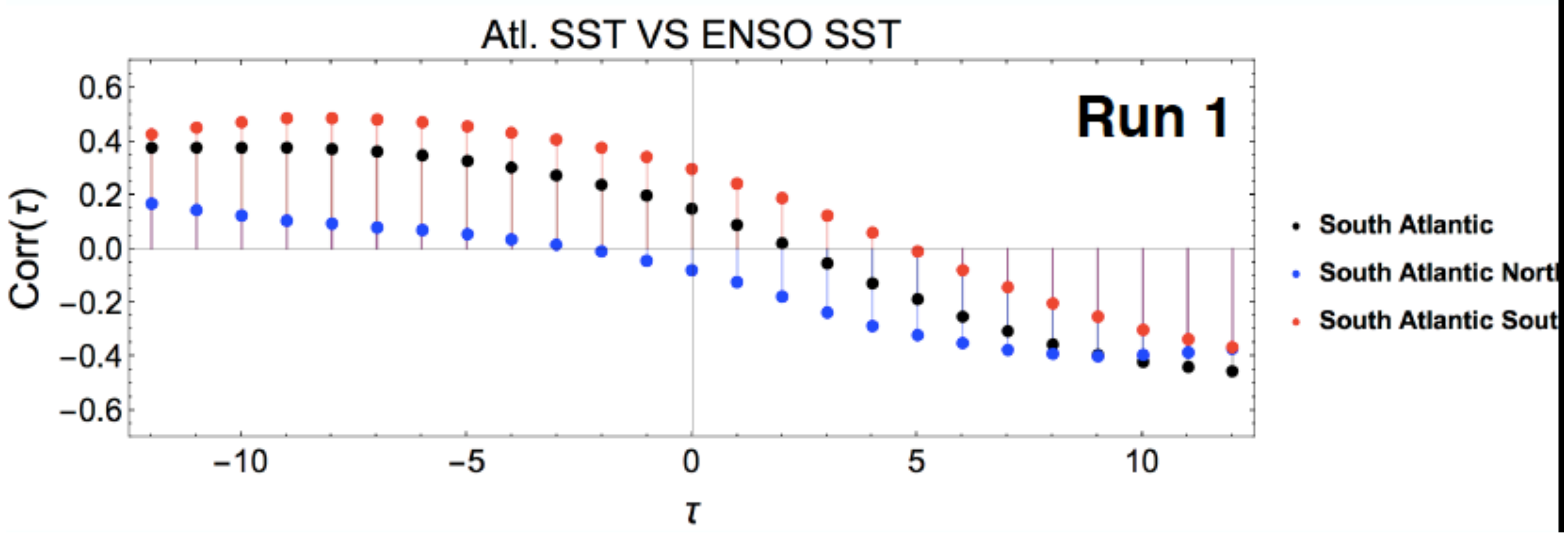


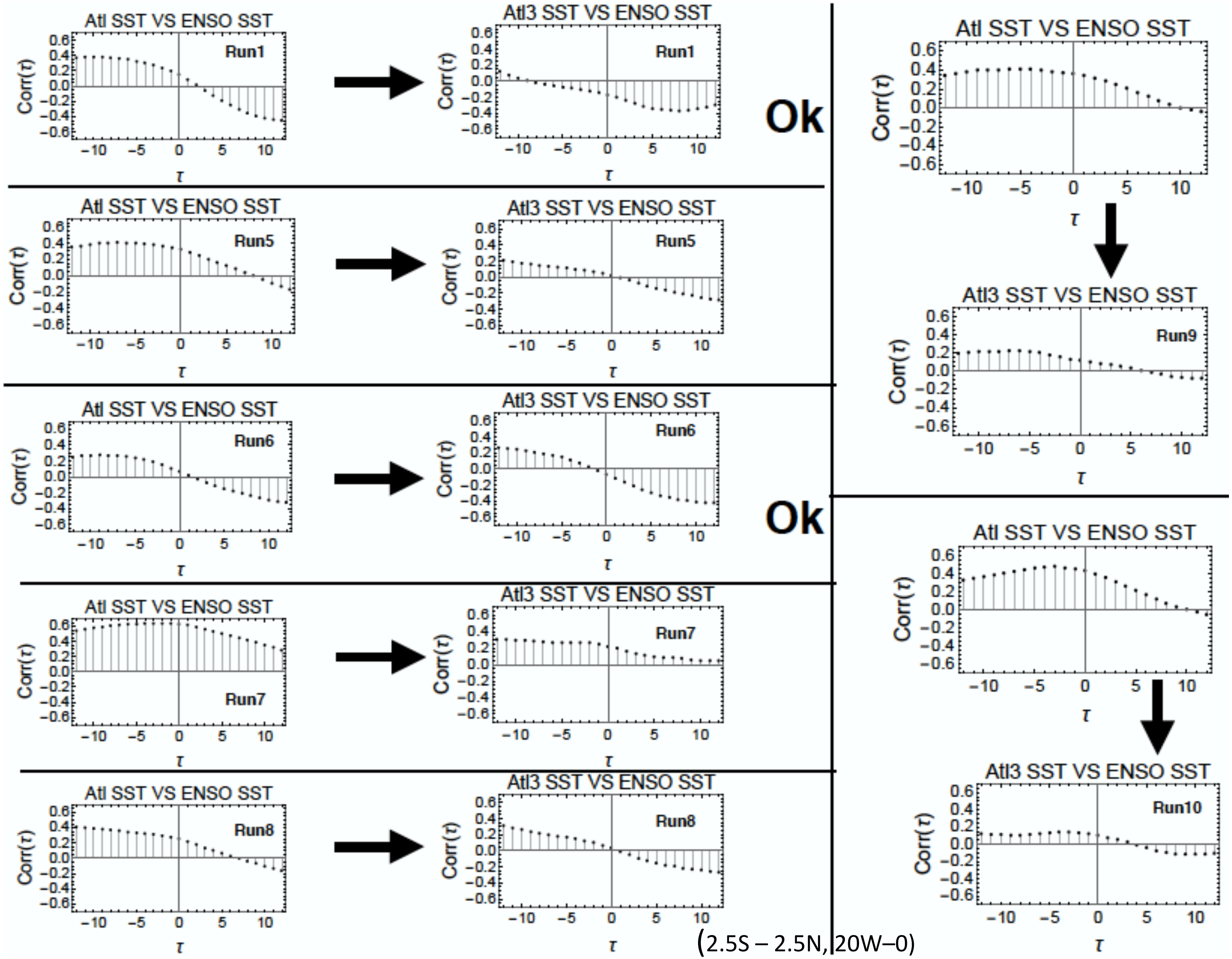
60°N
0°
60°S
0°
180°
1 5 10 15 20 24
$\times 10^3$

In CESM




Atl. SST VS ENSO SST
Run 5
Corr(τ)
τ
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−10
−5
0
5
10
South Atlantic
South Atlantic North
South Atlantic South

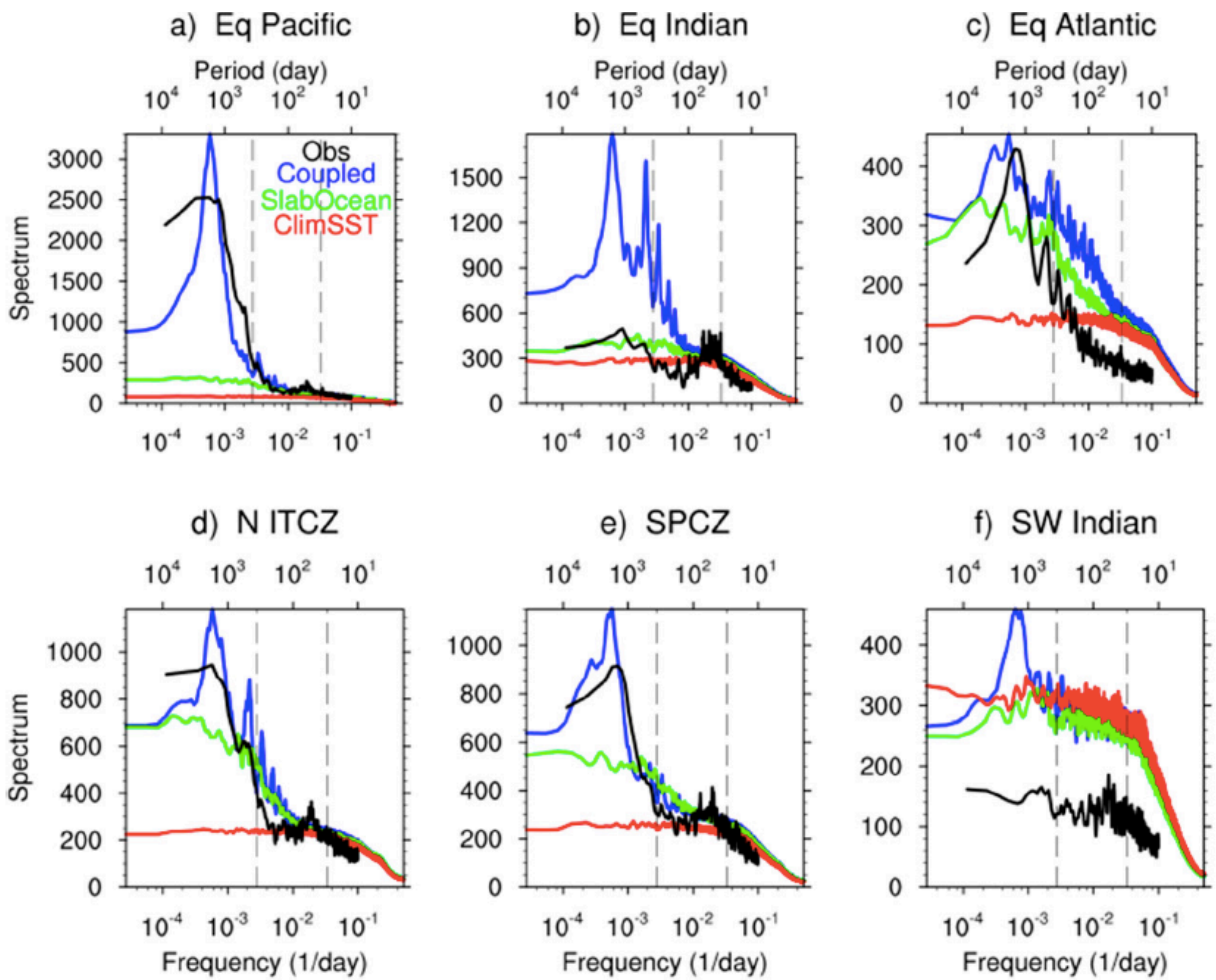

Atl. SST VS ENSO SST
Run 1
Corr(τ)
τ
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−10
−5
0
5
10
South Atlantic
South Atlantic North
South Atlantic South



Atl SST VS ENSO SST
Run1
Atl3 SST VS ENSO SST
Run1
Ok
Atl SST VS ENSO SST
Run5
Atl3 SST VS ENSO SST
Run5
Atl SST VS ENSO SST
Run6
Atl3 SST VS ENSO SST
Run6
Ok
Atl SST VS ENSO SST
Run7
Atl3 SST VS ENSO SST
Run7
Atl SST VS ENSO SST
Run8
Atl3 SST VS ENSO SST
Run8
Atl SST VS ENSO SST
Atl3 SST VS ENSO SST
Run9
Atl SST VS ENSO SST
Atl3 SST VS ENSO SST
Run10
Corr(τ)
τ

(2.5S – 2.5N, 20W–0)

Confirmed in precipitation power spectra from He et al., 2017



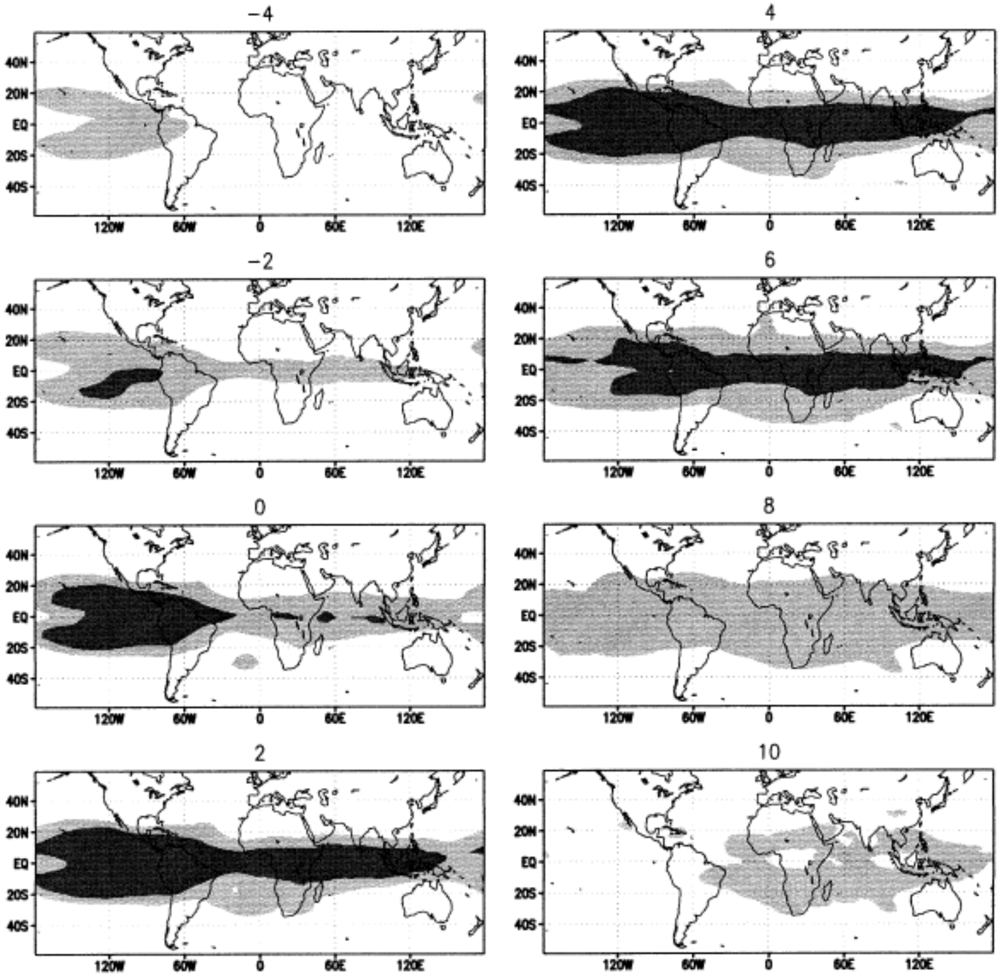


FIG. 2. Lag correlation between the Niño-3 index (SST anomalies averaged between 5°S–5°N, and 150°–90°W), and MSU channel-2 temperatures for 1979–99. Light shading is for $0.3 < r < 0.6$, and dark is $r > =0.6$. The number above each panel indicates the lag or lead in months: a −2 implies MSU channel 2 leads Niño-3 by 2 months.

Overestimation of thermodynamic response to ENSO

Chiang and Sobel, 2002

The Atlantic problem: partial conclusions

- This teleconnection is fragile (Chang et al.,2006). There are two Niños in different basins (the usual Eastern Pacific and the one in the Atlantic) and they are both influencing each others.
- These two modes can have a "destructive" (or "constructive") interference (in CESM always canceling each other).
- A key ingredient to simulate well the Atlantic variability is to be able to reproduce the ocean-atmosphere dynamics (waves teleconnections for the atmospheric part; issues with EBUS bias).
- For CESM the thermodynamic influence of ENSO is overestimated: the **thermodynamics response to El Niño events dominates** over the ocean-atmosphere dynamical feedback → also explains strong correlation between the ENSO domain and the North Tropical Atlantic (for which the tropospheric heating mechanism is the key mechanism)