State aggregation for dynamical systems

An information-theoretic approach

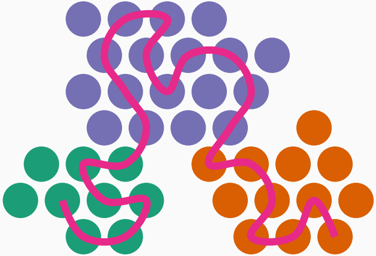
Mauro Faccin
IRD/CEPED, Université de Paris

@ Oxford 2021 (Networks seminar)

Projected Markov Chain

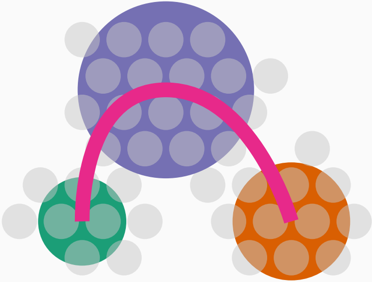
Markov Chain

$$\ldots, x_{\text{past}}, x_{\text{now}}, x_{\text{future}}, \ldots$$

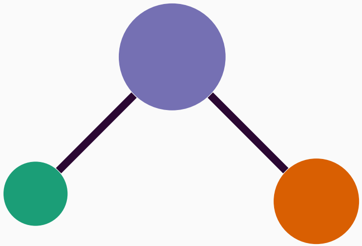


Projection

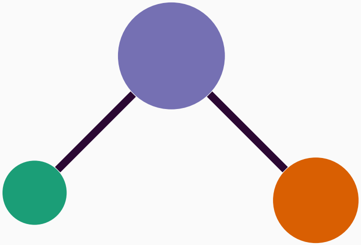
$$\ldots, y_{\text{past}}, y_{\text{now}}, y_{\text{future}}, \ldots$$



Where did the complexity go?



Where did the complexity go?



> Part of the complexity is now hidden in the [projected] dynamics.

> Emergence of **effective memories**.

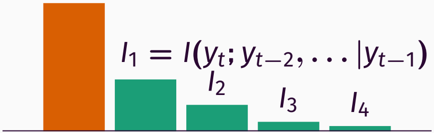
The Entrogram

Entrogram

Information flowing from the **PAST** toward the **FUTURE**.

$$I_0 = I(y_t; y_{t-1}, \dots)$$

$$I_1 = I(y_t; y_{t-2}, \dots | y_{t-1})$$



where $I(X;Y) = H(X) - H(X|Y)$ is the Mutual Information

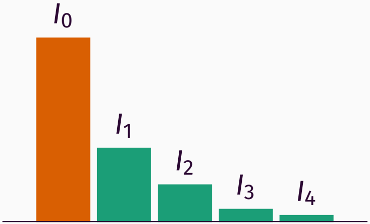
Recall:

$\dots, x_t, \dots$ the Markov Chain on the original space

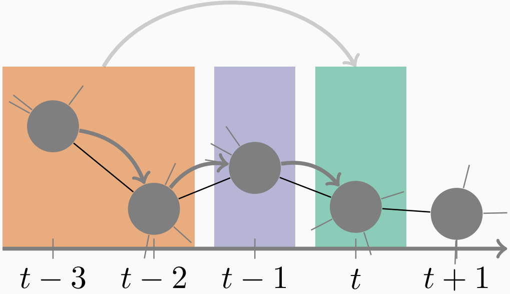
$\dots, y_t, \dots$ the projection of the Markov Chain on the aggregated space

Faccin, Schaub, Delvenne Journal of Complex Networks, 6(5), 2018, p661–678

Entrogram: Information flowing from the PAST to the FUTURE



$$I(y_t;\, y_{t-2}, y_{t-3}, \ldots | y_{t-1})$$

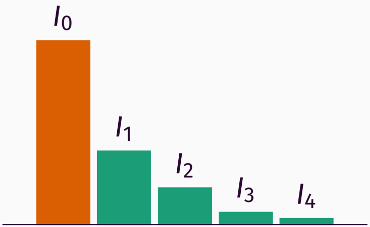


Faccin, Schaub, Delvenne Journal of Complex Networks, 6(5), 2018, p661–678

Crutchfield and Young (1989) PRL, 63, 105.

Crutchfield and Feldman (2003) Chaos, 13, 25–54.

Entrogram: a compact description of the system complexity



> [orange] [Total] Predictability, how the dynamics are aligned to the partition.

> [green] Emergent effective memory

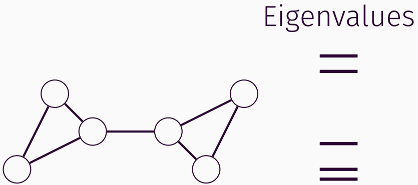
> [green] + [orange] Overall complexity (excess entropy) of the dynamical process

Faccin, Schaub, Delvenne Journal of Complex Networks, 6(5), 2018, p661–678

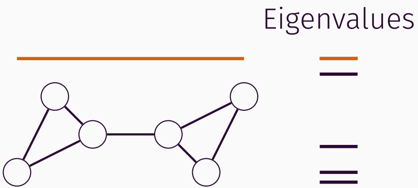
Crutchfield and Young (1989) PRL, 63, 105.

Crutchfield and Feldman (2003) Chaos, 13, 25–54.

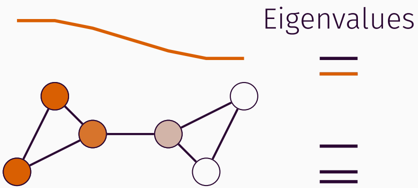
Consider the structure of the eigenvectors of the *transition* matrix (AD^{-1}).



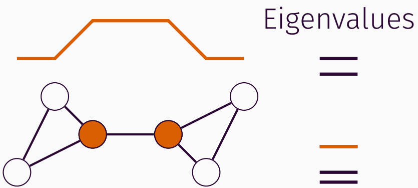
Consider the structure of the eigenvectors of the *transition* matrix (AD^{-1}).



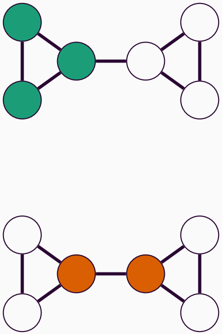
Consider the structure of the eigenvectors of the *transition* matrix (AD^{-1}).

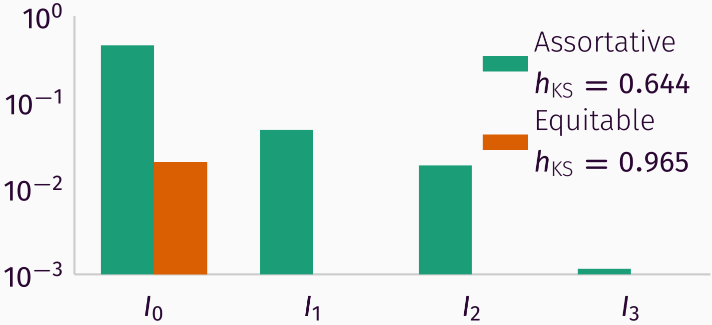


Consider the structure of the eigenvectors of the *transition* matrix (AD^{-1}).



Entrogram of the bow-tie graph





Aggregation strategies

Non-linear correlations

AutoInformation

$$I(y_t; y_{t-\tau})$$

Non-linear *correlation* between successive time-steps

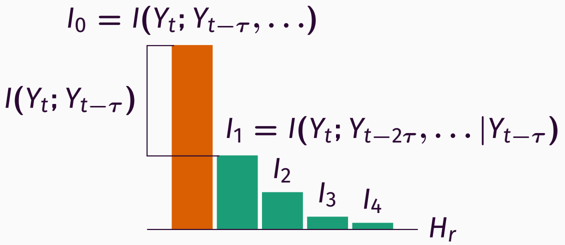
M.F. et al, Journal of Complex Networks, cnx055

Non-linear correlations

AutoInformation

$$I(y_t; y_{t-\tau})$$

Non-linear *correlation* between successive time-steps



where τ represents a time-scale parameter.

A proxy for *Predictability* and *Markovianity*.

M.F. et al, Journal of Complex Networks, cnx055

Random walk covariance

> How much the dynamics are *trapped* by a partition?

Let's consider a partition of nodes into classes where χ_c is the characteristic function of class c.

Partition autocovariance along the dynamics $\longrightarrow$

$$\mathbf{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

$$E(\chi_c(t)\chi_c(t-1)) = \frac{1}{2m}\sum_{ij\in c} A_{ij}$$

$$E(\chi_c(t)) = \frac{1}{2m}\sum_{i\in c} k_i$$

$$\text{where } k_i = \sum_j A_{ij}$$

$$\text{and } m = \frac{1}{2}\sum_{ij} A_{ij}$$

In symmetric networks.

Random walker covariance

χ_c characteristic function of class c

$$Q = \sum_c \mathbf{Cov}\left(\chi_c(t), \chi_c(t+1)\right)$$

Modularity:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m}\right] \delta(c_i, c_j)$$

Linear correlation between consecutive time-steps.

Shen et al. (2010) PRE, 82, 016114

Generative models as particular case

Fitting a generative model (e.g. DC-SBM) to the data through log-likelihood maximization can be seen as maximizing the AutoInformation for paths of lenght $\tau = 1$ (e.g. links).

$$I(Y_t; Y_{t-1}) = H(Y_y) + H(Y_{t-1}) - H(Y_y, Y_{t-1})$$

$$H(Y_t) = -\sum_c \frac{e_c}{2m} \log \frac{e_c}{2m} \qquad e_c = \sum_{i \in c, j} A_{ij}$$

$$H(Y_t, Y_{y-1}) = -\sum_{cd} \frac{e_{cd}}{2m} \log \frac{e_{cd}}{2m} \qquad e_{cd} = \sum_{i \in c, j \in d} A_{ij}$$

In binary symmetric networks

Generative models as particular case

Fitting a generative model (e.g. DC-SBM) to the data through log-likelihood maximization can be seen as maximizing the AutoInformation for paths of lenght $\tau = 1$ (e.g. links).

$$I(Y_t; Y_{t-1}) = H(Y_y) + H(Y_{t-1}) - H(Y_y, Y_{t-1})$$

$$H(Y_t) = -\sum_{c} \frac{e_c}{2m} \log \frac{e_c}{2m} \qquad e_c = \sum_{i \in c, j} A_{ij}$$

$$H(Y_t, Y_{y-1}) = -\sum_{cd} \frac{e_{cd}}{2m} \log \frac{e_{cd}}{2m} \qquad e_{cd} = \sum_{i \in c, j \in d} A_{ij}$$

In binary symmetric networks

DC-SBM

$$\mathcal{S} \propto \frac{1}{2} \sum_{cd} e_{cd} \log \frac{e_{cd}}{e_c e_d}$$

Non-linear communities

AutoInformation and its connections with other approaches.

AutoInformation: $I(Y_t, Y_{t-\tau})$

Modularity

$$Q \propto \sum_c \mathbf{Cov}\left(\chi_c(t), \chi_c(t+1)\right)$$

Spectral Properties

DC-SBM

$$I(Y_t; Y_{t-\tau}) \propto -\sum_{rs} e_{rs} \log \frac{e_{rs}}{e_r e_s}$$

In some cases

Shen et al. (2010) PRE, 82, 016114.
Karrer and Newman (2011), PRE 83, 016107.
Rosvall and Bergstrom (2008) PNAS 105, 1118.

Facts

AutoInformation

$$I(y_t; y_{t-\tau})$$

Non-linear *correlation* between successive time-steps

Maximizing in a naive way is not possible, one need to fix the number of classes or use a model selection:

$$\mathcal{I} = I(y_t; y_{t-1}) - \alpha H(y_t)$$

The parameter τ selects the *time-scale* of the aggregation.

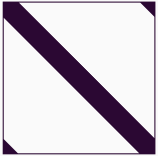
Didactic Examples.

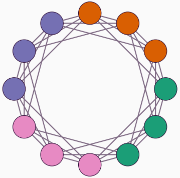
Example 0: One cycle

A regular ring lattice with N nodes, each connected with ***k*** neighbours.

How many classes?

Adj:



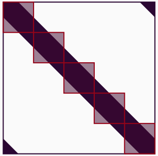


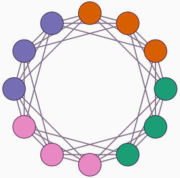
Example 0: One cycle

A regular ring lattice with N nodes, each connected with k neighbours.

How many classes?

Adj:



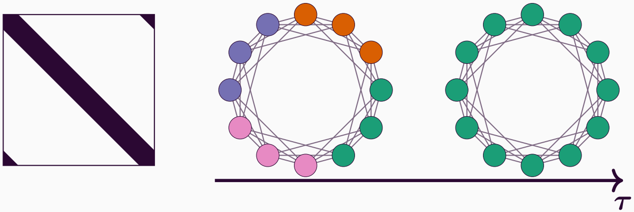


Example 0: One cycle

A regular ring lattice with N nodes, each connected with ***k*** neighbours.

How many classes?

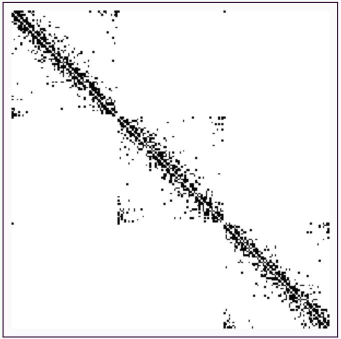
Adj:



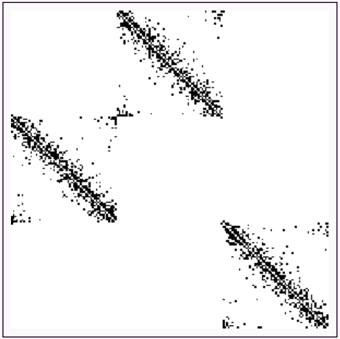
Example 1: Range dependant graphs

$$p_{ij} = \alpha_{c_i c_j} \cdot (\gamma_{c_i c_j})^{d_{ij}}$$
$$\alpha_{c_i c_j}, \gamma_{c_i c_j} \in [0, 1]$$

with d_{ij} a (normalized) distance between nodes aligned on a cycle.

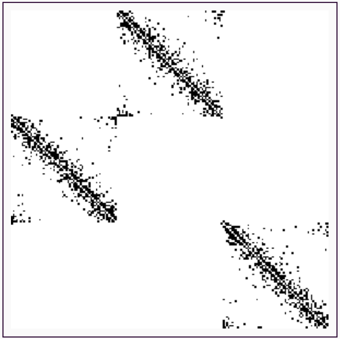


Example 1: Range dependent graphs



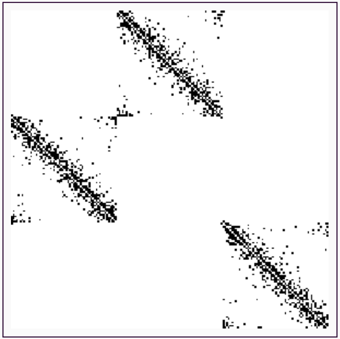
DC-SBM

Example 1: Range dependent graphs

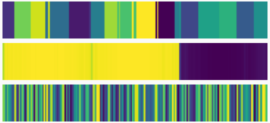


DC-SBM
spectral

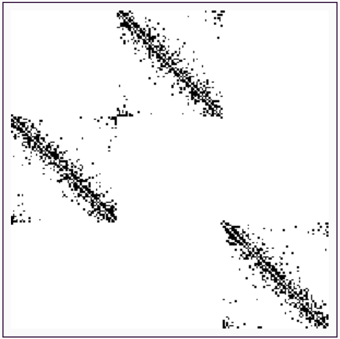
Example 1: Range dependent graphs

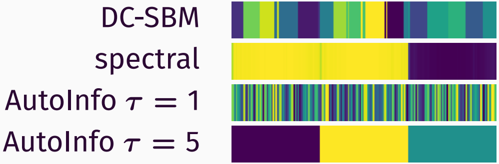


DC-SBM
spectral
AutoInfo $\tau = 1$

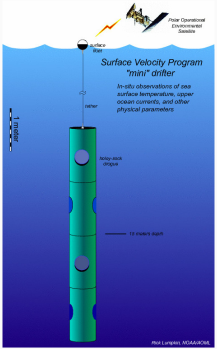


Example 1: Range dependent graphs





Example 2. Ocean buoys



Global Drifter Program

GDP Array

AOML Drifter Data Assembly Center

Mon, 04 Oct 2021

No. of Buoys = 1471

ID WMO

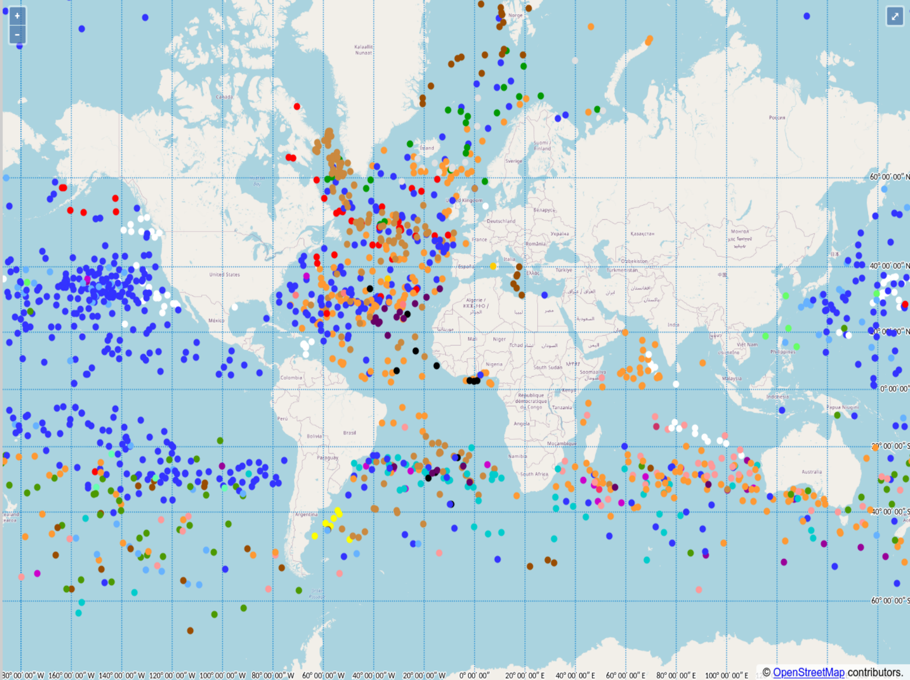
Search..

Map Viewing Options

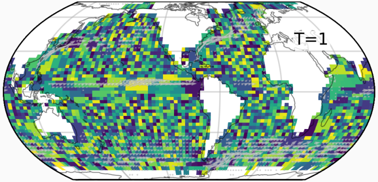
- Deploying Country
- Buoy Type
- Buoy Drogue Status

Deploying Country

Argentina (7)	Australia (48)
Barbados (3)	Brazil (12)
Canada (40)	Chile (4)
China (6)	Denmark (1)
France (272)	Germany (12)
Iceland (23)	India (3)
Indonesia (1)	Italy (51)
Japan (11)	Korea Rep. of (63)
New Zealand (52)	Netherlands (14)
Portugal (19)	Seychelles (1)
South Africa (59)	Spain (2)
Tonga (1)	UK (153)
USA (539)	Unknown (74)

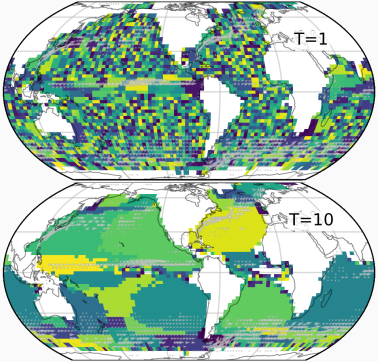


Example 3. Ocean buoys



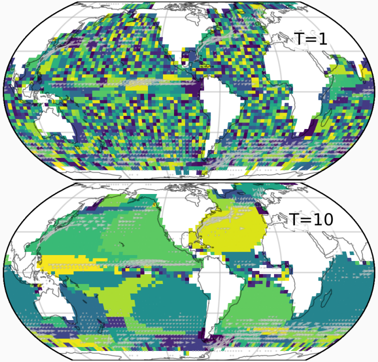
Each time step lasts 16 days.

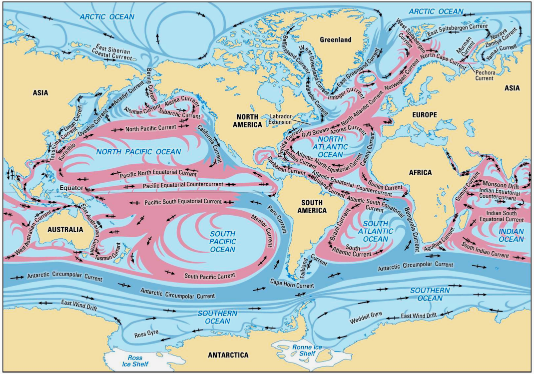
Example 3. Ocean buoys



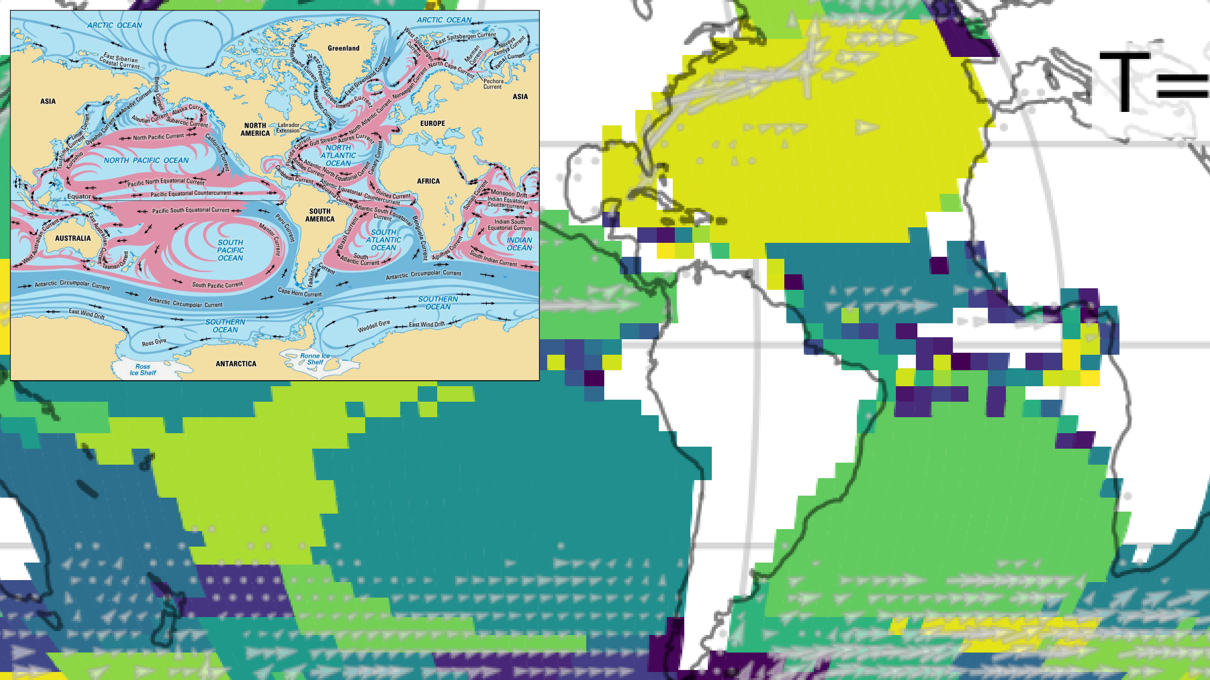
Each time step lasts 16 days.

Example 3. Ocean buoys





Each time step lasts 16 days.


T=
ARCTIC OCEAN
ARCTIC OCEAN
East Siberian Coastal Current
Greenland
Baffin Island Current
West Greenland Current
East Greenland Current
West Spitsbergen Current
East Spitsbergen Current
Novaya Zemlya Current
Murman Current
Yamal Current
North Cape Current
Pechora Current
Norwegian Current
Irminger Current
Labrador Current
ASIA
ASIA
Bering Current
Anadyr Current
Alaska Current
Aleutian Current
Subarctic Current
Oyashio Current
Liman Current
Tsushima Current
Kuroshio
North Pacific Current
California Current
NORTH AMERICA
Labrador Extension
Gulf Stream
North Atlantic Current
Azores Current
EUROPE
NORTH PACIFIC OCEAN
NORTH ATLANTIC OCEAN
Florida Current
Antilles Current
Canary Current
Atlantic North Equatorial Current
Caribbean Current
Pacific North Equatorial Current
Pacific Equatorial Countercurrent
Equator
Guiana Current
Atlantic Equatorial Countercurrent
Guinea Current
AFRICA
Somali Current
Monsoon Drift
Indian Equatorial Countercurrent
Pacific South Equatorial Current
Atlantic South Equatorial Current
Indian South Equatorial Current
SOUTH AMERICA
Peru Current
Mentor Current
Brazil Current
Benguela Current
AUSTRALIA
East Australian Current
West Australian Current
Tasman Current
SOUTH PACIFIC OCEAN
SOUTH ATLANTIC OCEAN
INDIAN OCEAN
Agulhas Current
South Atlantic Current
South Indian Current
Falkland Current
South Pacific Current
Antarctic Circumpolar Current
Antarctic Circumpolar Current
Antarctic Circumpolar Current
Cape Horn Current
East Wind Drift
East Wind Drift
SOUTHERN OCEAN
SOUTHERN OCEAN
Weddell Gyre
Ross Gyre
Ross Ice Shelf
Ronne Ice Shelf
ANTARCTICA

🏃 Finally...

Questions?

FIG. 173.—FRANKLIN'S CHART OF THE GULF STREAM.

Joint work with:

JC Delvenne
UCLouvain

M Schaub
RWTH AACHEN UNIVERSITY

https://maurofaccin.github.io
mauro.fccn@gmail.com

Code at:
https://maurofaccin.github.io/aisa