



Identification automatique des courants et séquences de courants menant aux échouements de bancs d'algues Sargasses dans les Petites Antilles

Emmanuel Biabiany^{1*}, Didier Bernard², Raphaël Cécé², Romual Chery² et Nahoufal Sekkat²

¹LAMIA, ²LARGE, *emmanuel.biabiany@univ-antilles.fr

11^e Biennale du Congrès SMAI

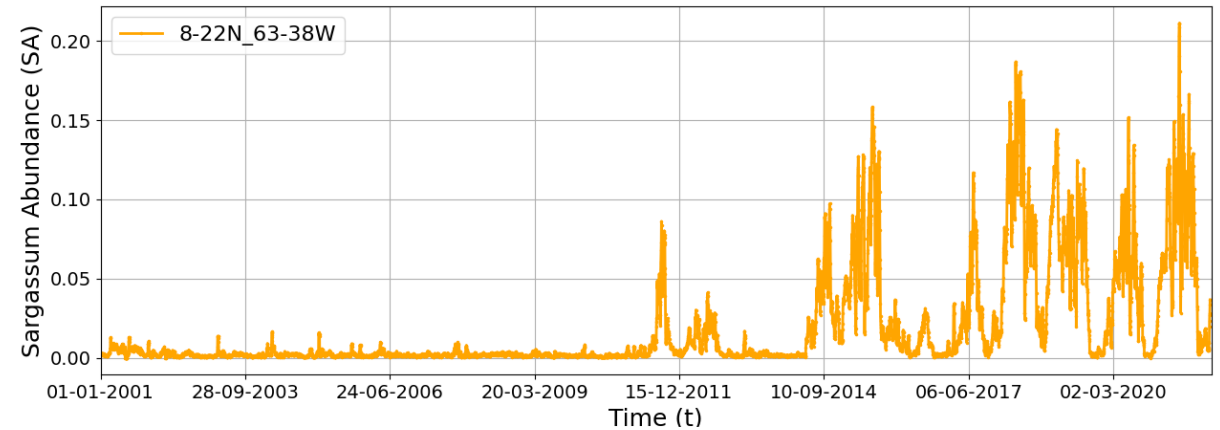
Mercredi 24 mai 2023

Guadeloupe

Contexte

- Depuis 2011 → arrivée massive de sargasses
- Impact sur toutes les îles de la Caraïbes
- Enjeux sanitaire (H2S), environnementale, touristiques → économiques
- Gestion locale → collectivité et EPCI

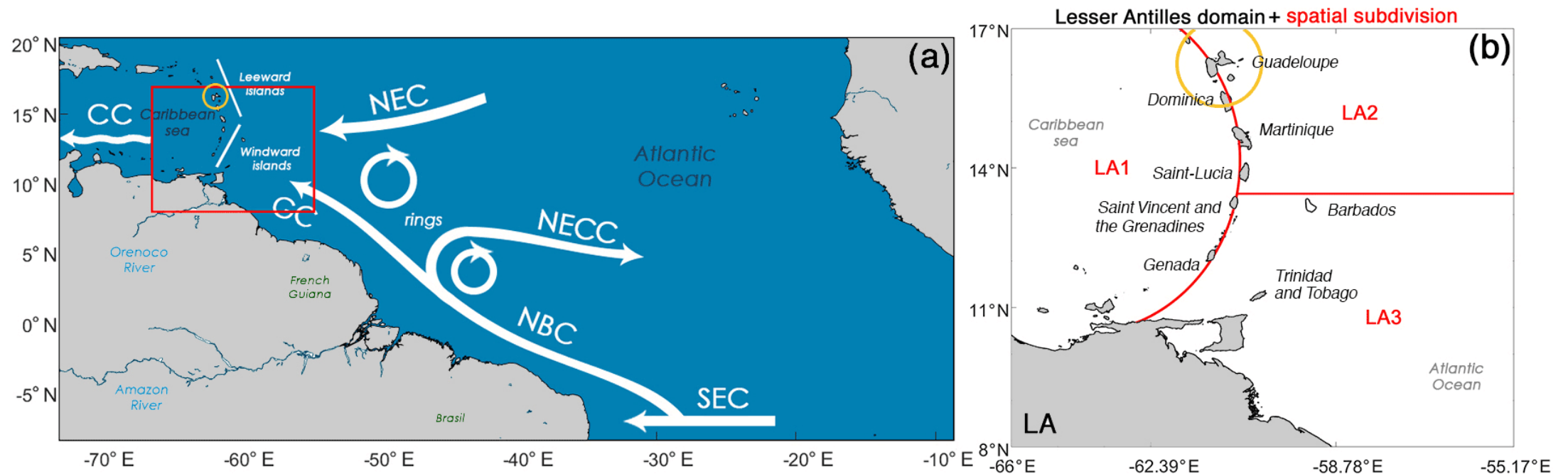
Comprendre → Anticiper (prévisions)



Objectifs

- Recenser les principaux courants de surface → Petites Antilles
- Déterminer la dynamique spatio-temporelle → courants
- Proposer automatisation analyse → prévision
- Identifier séquences courants vecteurs → bancs de sargasses

Données | Zone d'intérêt



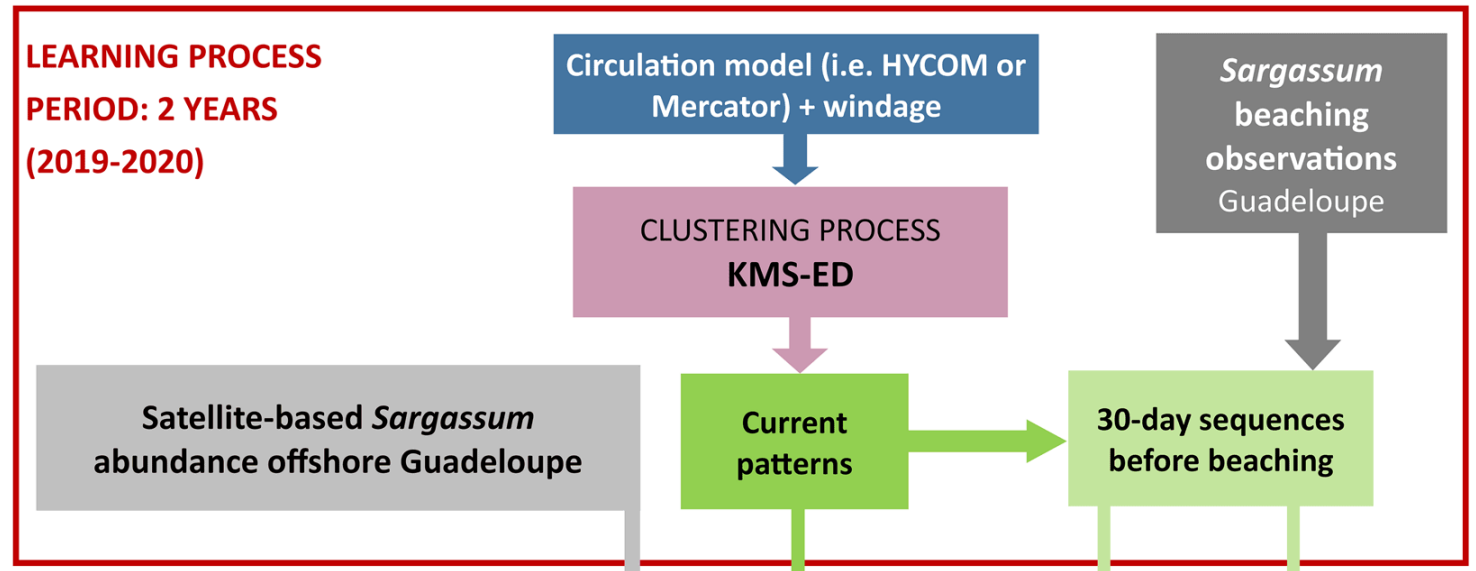
Littérature scientifique → Identification des courants généraux
GC et NEC → introduction des algues dans la zone

Zone d'intérêt → séparation

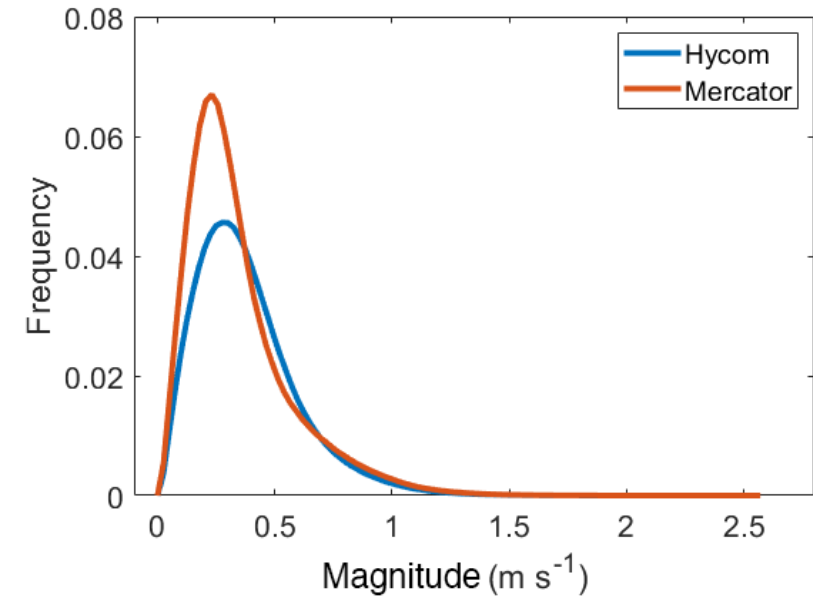
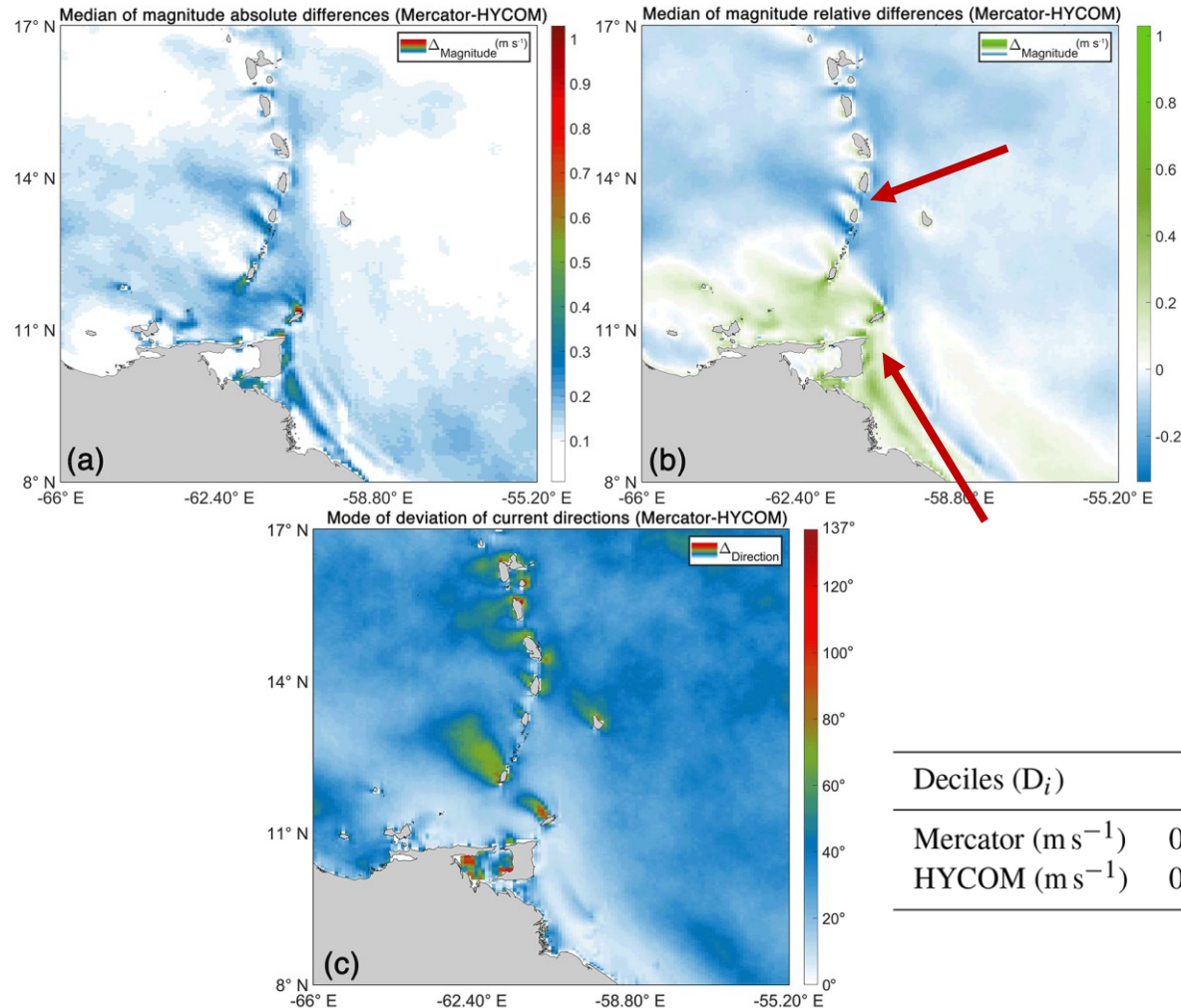
Méthodes | Phase d'apprentissage

Trois sources de données :

- Réanalyse climatique des courants → Clustering
 - Mercator
 - HYCOM



Données | Dynamique globale

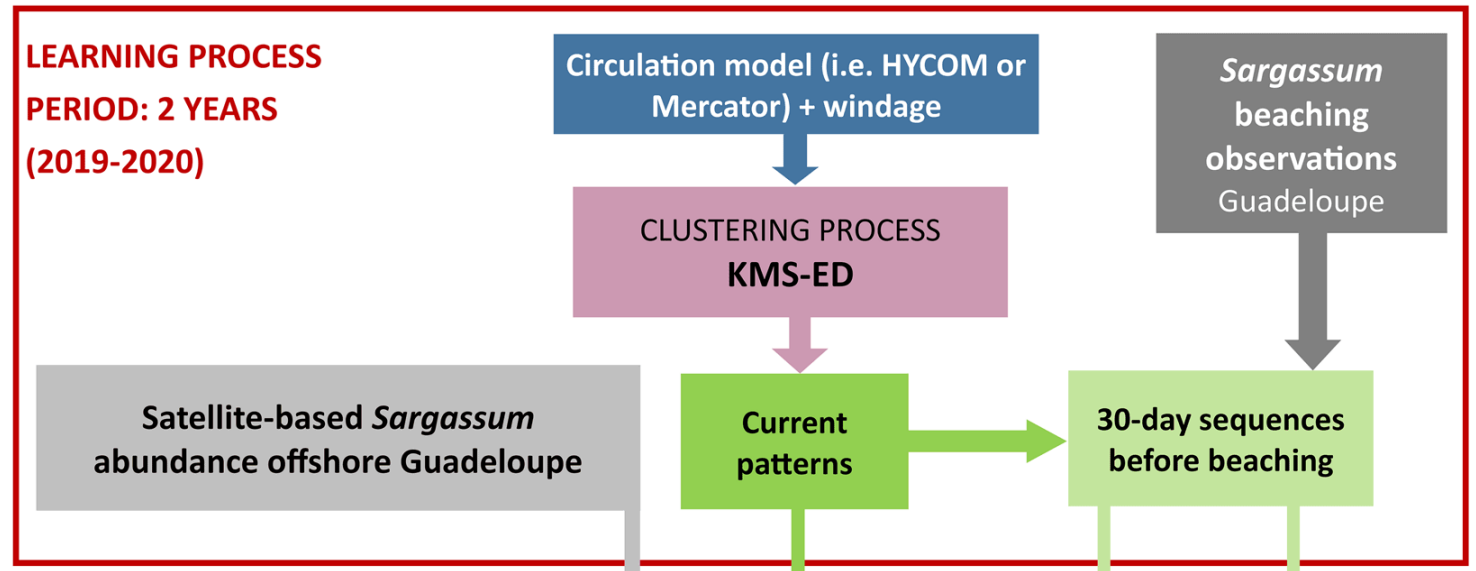


Deciles (D_i)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Max	Mean	σ
Mercator ($m s^{-1}$)	0.11	0.16	0.20	0.24	0.28	0.32	0.39	0.48	0.65	2.57	0.33	0.22
HYCOM ($m s^{-1}$)	0.13	0.18	0.23	0.28	0.32	0.38	0.44	0.52	0.65	2.49	0.36	0.21

Méthodes | Phase d'apprentissage

Trois sources de données :

- Réanalyse climatique des courants → **Clustering**
 - Mercator
 - HYCOM
- Observations → **Validation**
 - DEAL
- Images satellites → **Indicateur**
 - FA Density (Wang and Hu 2016)
 - MODIS
 - Sentinelle I & II



Méthodes | Expert Distance

Méthodes | Expert Distance

Découpe de l'espace → zone pertinente (intégration)

$$D_1 = \{D_1 LA_1, \dots, D_1 LA_3\} \text{ et } D_2 = \{D_2 LA_1, \dots, D_2 LA_3\}$$

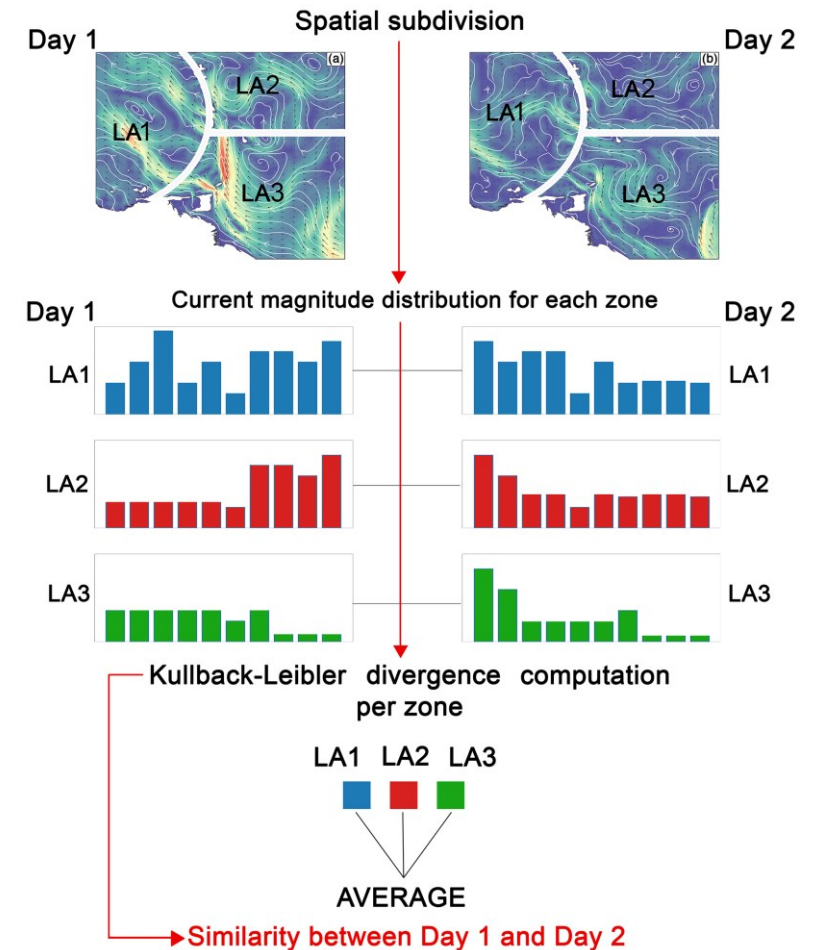
Analyse des dynamiques → flexibilité spatiale

Calcul d'une métrique → comparer deux histogrammes

→ Divergence de Kulback-Leibler Symétrisée (D_{KLS})

(recherche d'information mutuelle)

$$ED(D_1, D_2) = \frac{1}{N} \times \sum_{i=1}^N D_{KLS}(D_1 LA_i, D_2 LA_i)$$

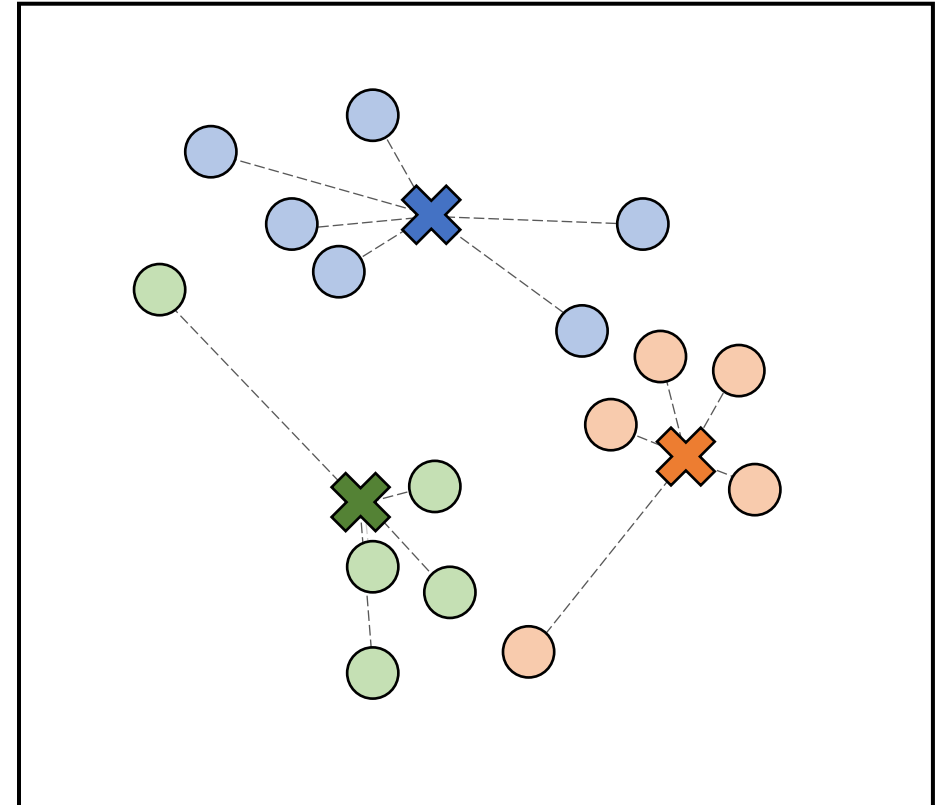


Méthodes | Clustering

ED remplace les métriques classiques lors du clustering

Le regroupement porte une expertise

→ Le choix du nombre de groupe ?



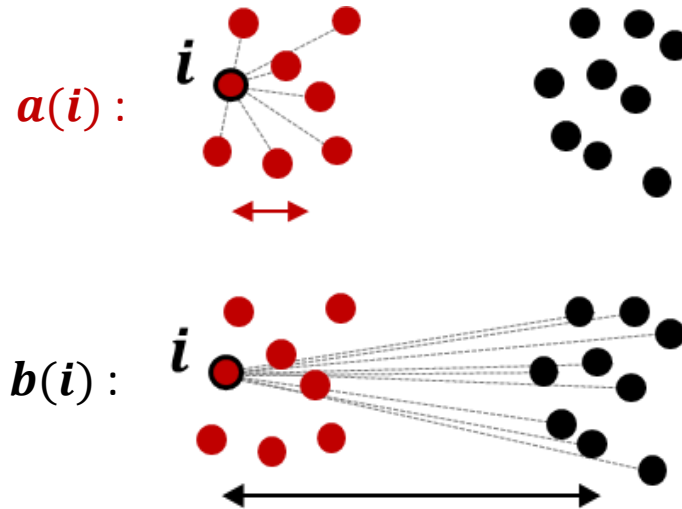
Espace du clustering

Méthodes | Evaluation du clustering

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

ou

$$s(i) = \begin{cases} 1 - a(i)/b(i), & \text{si } a(i) < b(i) \\ 0, & \text{si } a(i) = b(i) \\ b(i)/a(i) - 1, & \text{si } a(i) > b(i) \end{cases}$$



- $Sc(C_i)$ (où $i = n^\circ$ cluster) indice silhouette du cluster C_i

= moyenne des $s(i)$

- $Sa(M_k)$ (où $k = \text{nb}$ cluster) indice du clustering M_k

= moyenne des $Sc(C_i)$

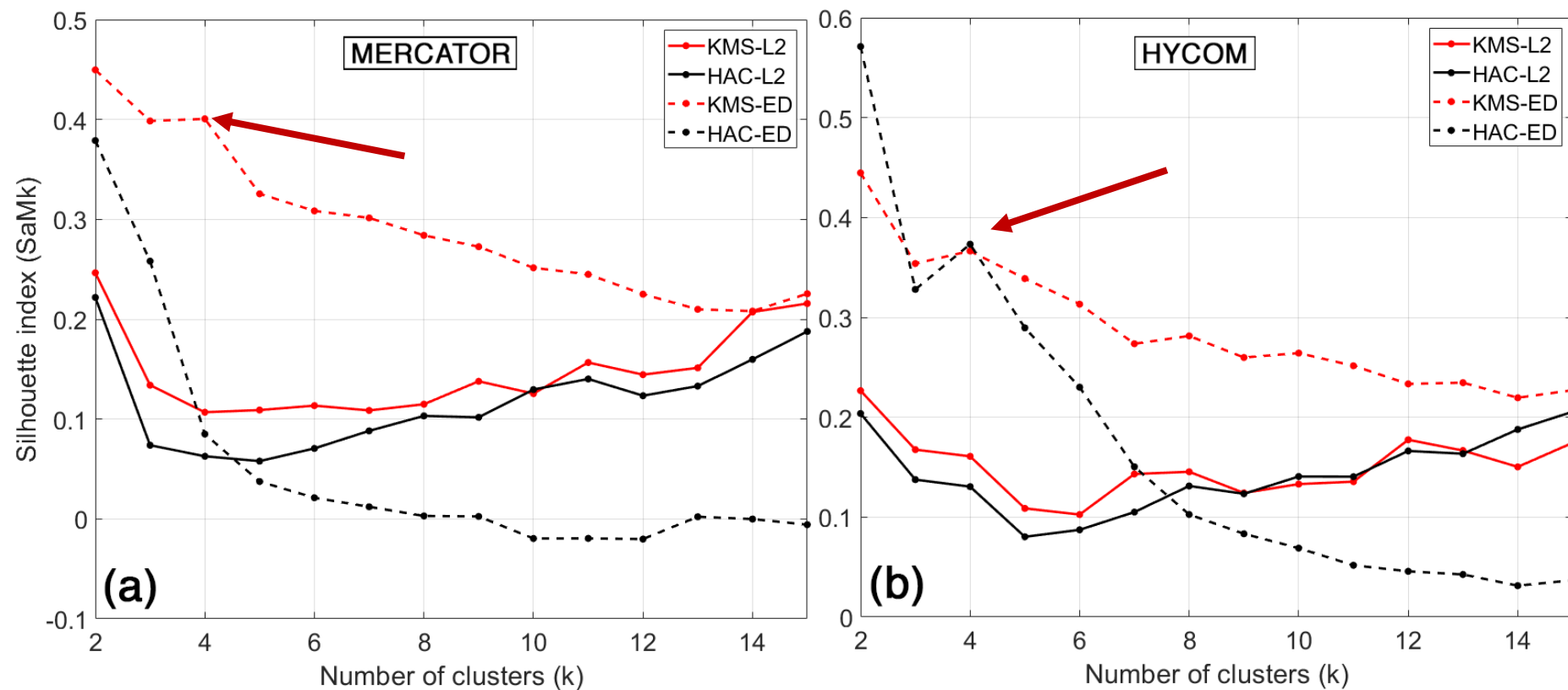
Homogénéité

$a(i) \rightarrow$ distance moyenne entre l'individu i et les autres individus de son cluster
(\rightarrow minimiser)

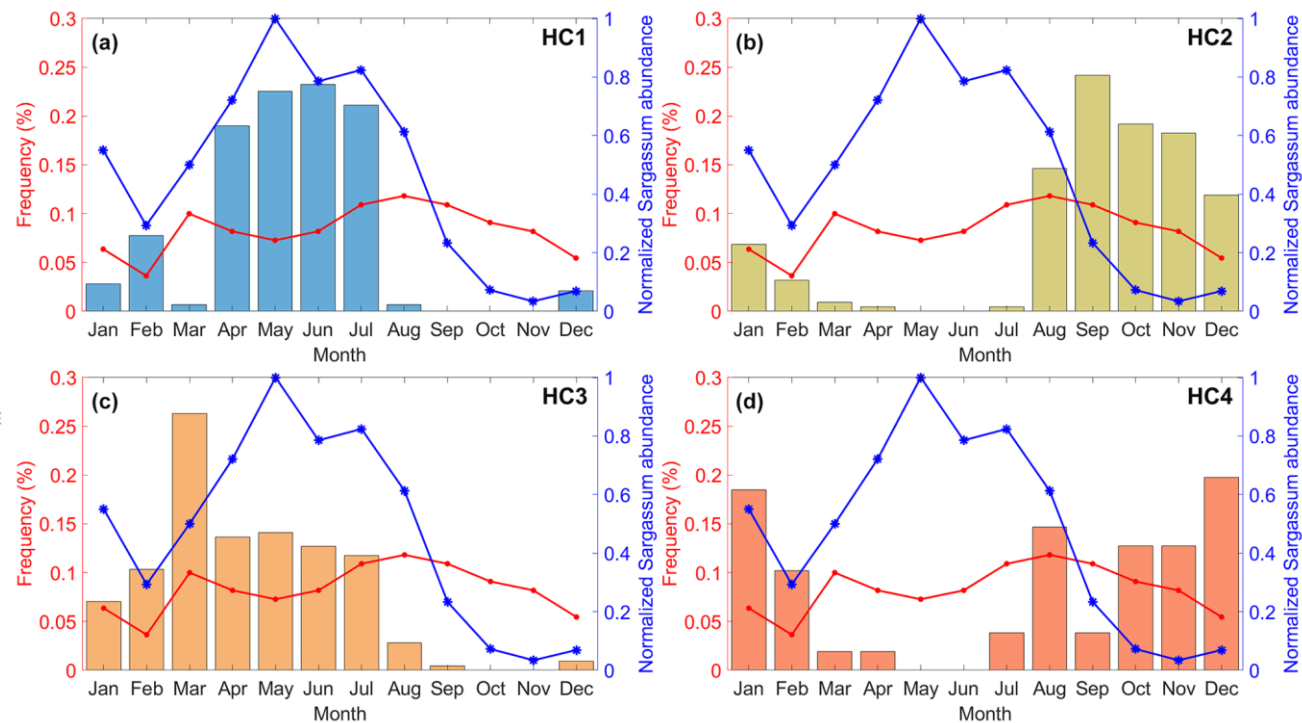
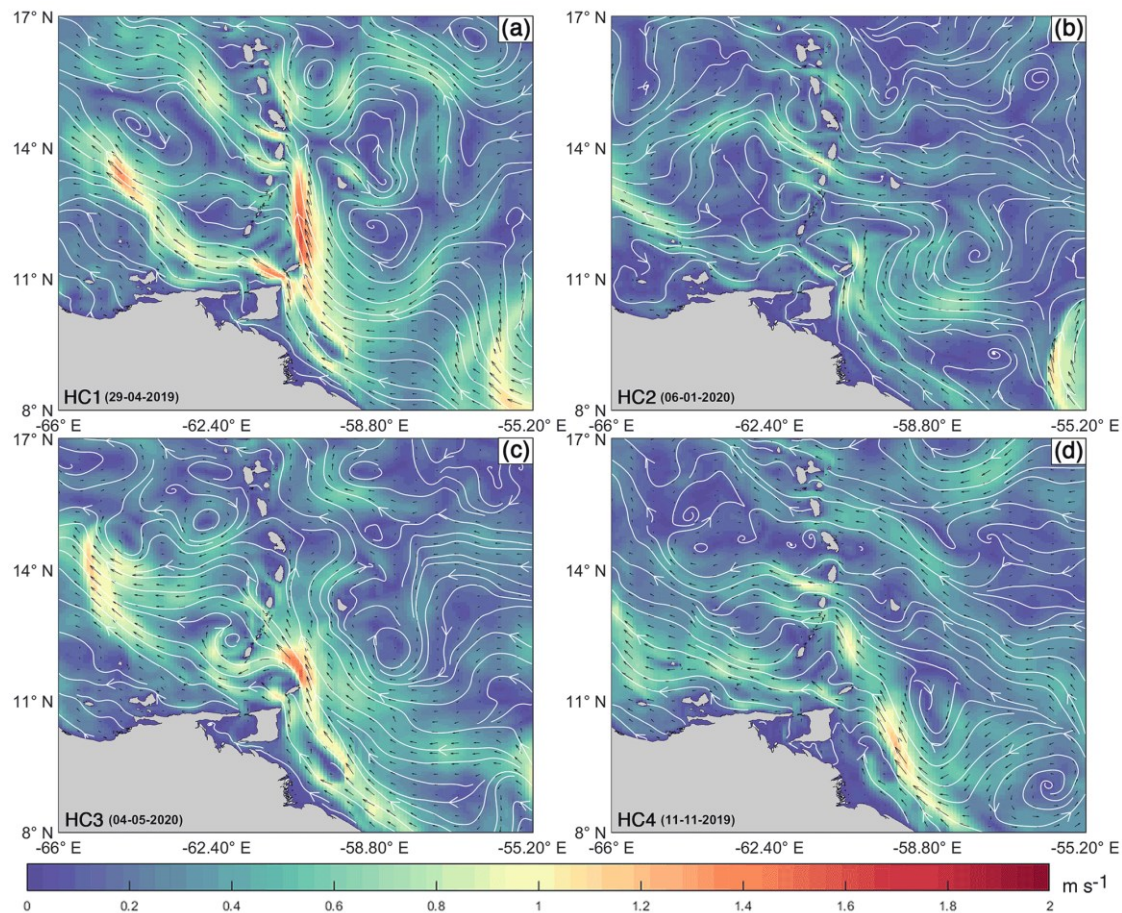
Séparabilité

$b(i) \rightarrow$ plus petite moyenne des distances entre l'individu i et les autres clusters
(\rightarrow maximiser)

Résultats | Evaluation du clustering



Résultats | Courants de surface

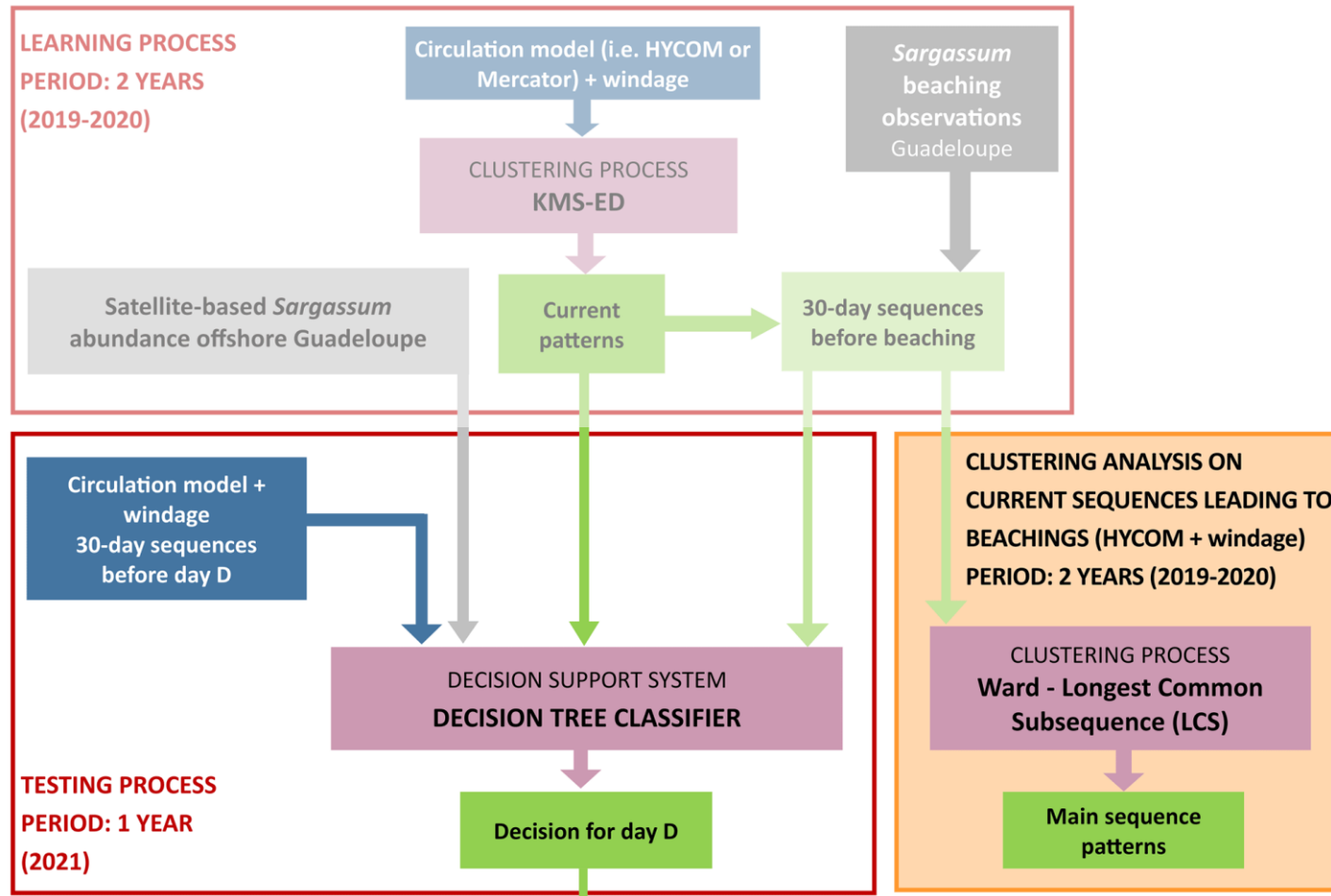


Abondance de sargasses mesurée par satellite

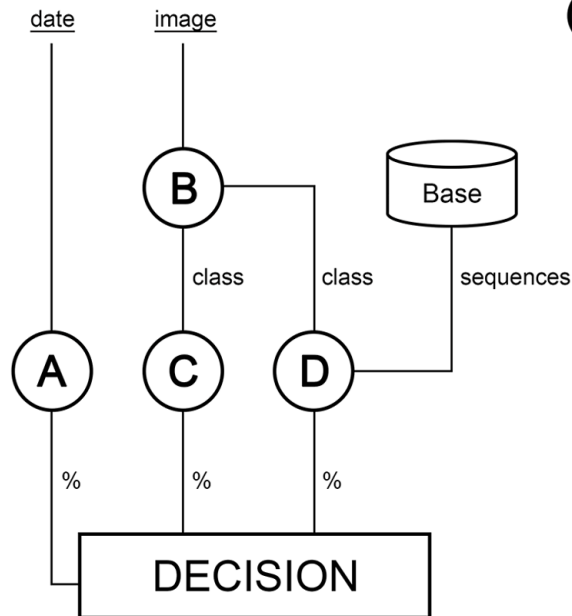
Fréquence d'échouement en GP (2019-2020)

- Décalage entre présence satellite et échouement (GP)
- Plusieurs courants concernés

Méthodes | Phase de test



Méthodes | Aide à la décision



A: daily probability to reach the maximum *Sargassum* abundance in the area 30-100 km offshore Guadeloupe, B: current pattern compare with KMS-ED barycenters, C: *Sargassum* beaching frequency par class (%), D: matching of current sequences with base history on Δt days (%), Base: set of the referenced 30-day current sequences before *Sargassum* beaching.

(a)

Base index	Beaching day	$d - 1$	$d - 2$...	$d - \Delta t$
1	$CI(1, d)$	$CI(1, d - \Delta t)$
2
.
.
.
N	$CI(N, d)$	$CI(N, d - \Delta t)$

$CI(i, d)$ operator gives the cluster label of oceanic surface current (computed by KMS-ED) for the index i of the base of sargassum beaching days and the corresponding date d .

(b)

Résultats | Évaluation des prévisions

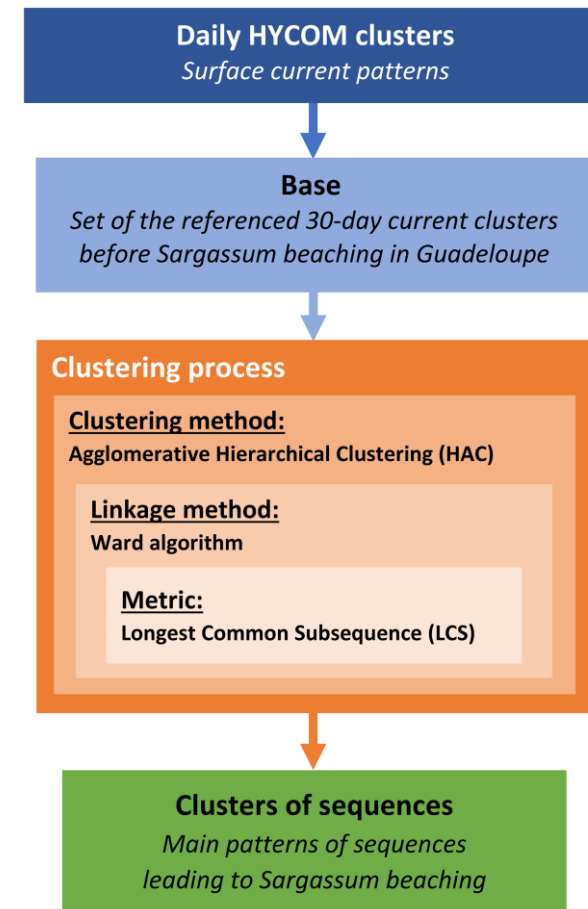
Time range around date (d)	Datasets	True positive (recall %)	True negative (recall %)	Accuracy (ratio %)
0	HYCOM	46 (59.0 %)	151 (52.8 %)	197 (54.1 %)
	Mercator	43 (55.1 %)	141 (49.3 %)	184 (50.6 %)
±1	HYCOM	52 (66.7 %)	175 (61.2 %)	227 (62.4 %)
	Mercator	47 (60.3 %)	151 (52.8 %)	198 (54.4 %)
±2	HYCOM	55 (70.5 %)	189 (66.1 %)	244 (67.0 %)
	Mercator	51 (65.4 %)	155 (54.2 %)	206 (56.6 %)
±3	HYCOM	57 (73.1 %)	198 (69.2 %)	255 (70.1 %)
	Mercator	51 (65.4 %)	161 (56.3 %)	212 (58.2 %)

Méthodes | Rétro-séquences (30 jours)

Regroupement automatique
des séquences de courants

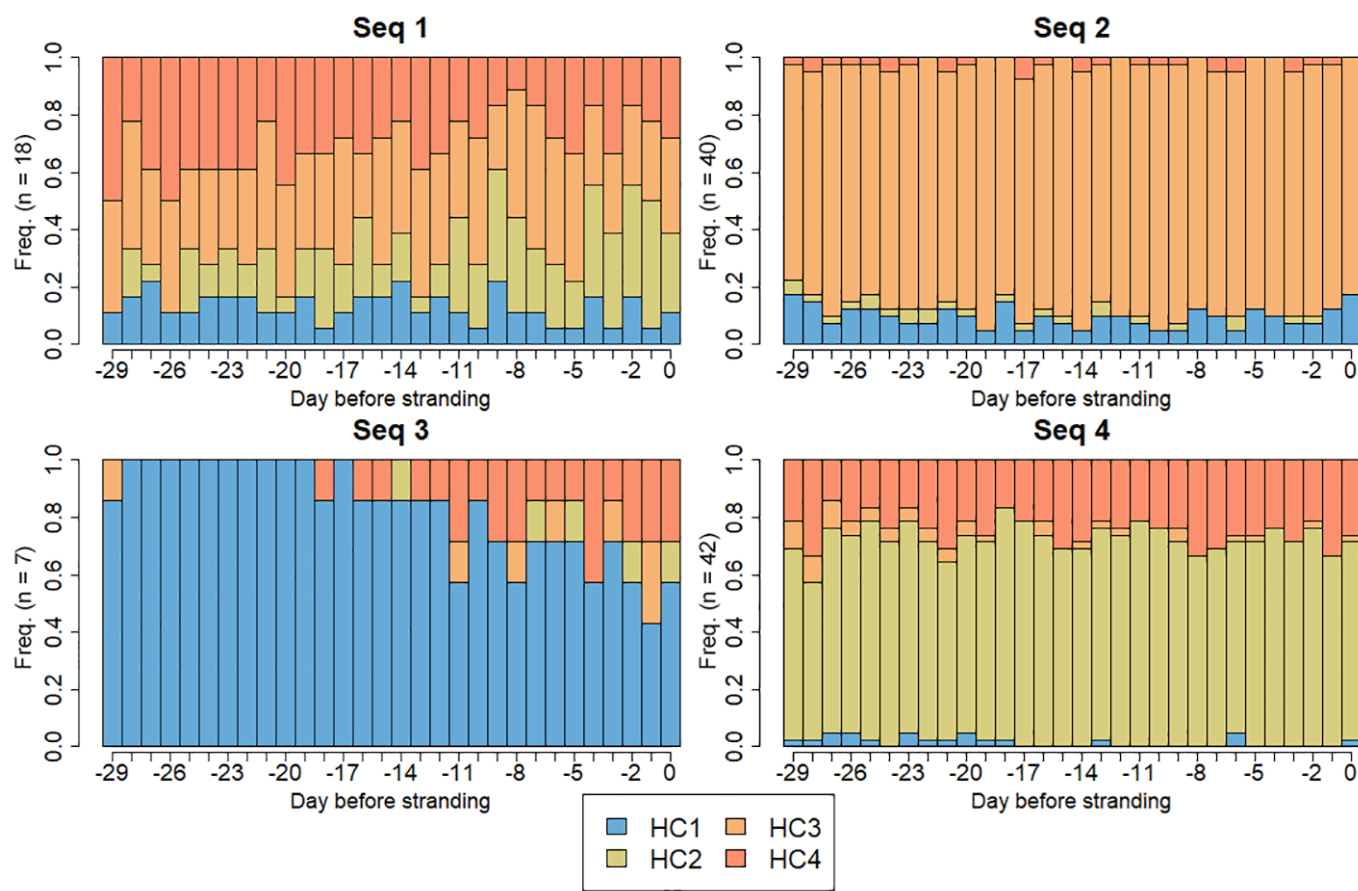
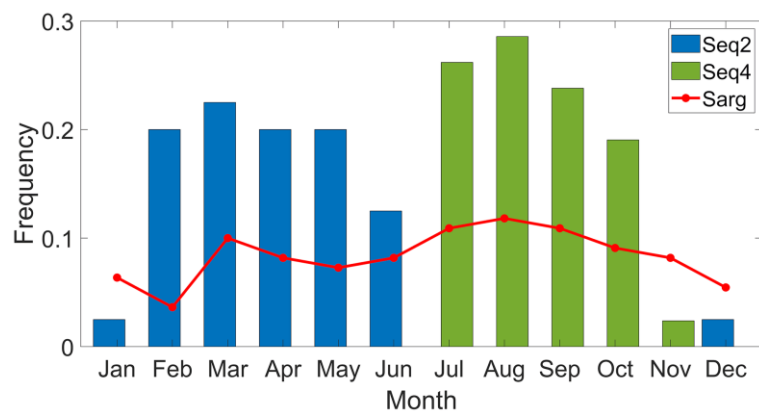
Usage d'un écart basé sur la recherche
de sous séquences communes

→ Identifier des séquences responsables
des échouements



Résultats | Clustering des rétro-séquences

30 d current sequence before beaching (HYCOM)	Seq1	Seq2	Seq3	Seq4
<i>n</i>	18	40	7	42
%	16.8	37.4	6.5	39.3



Merci
de votre attention