

AFH
Nantes, juin, 2017

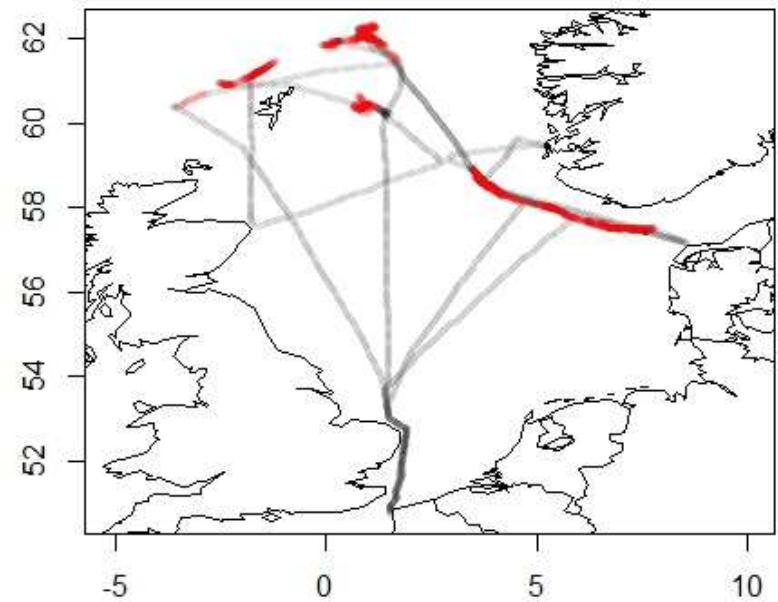
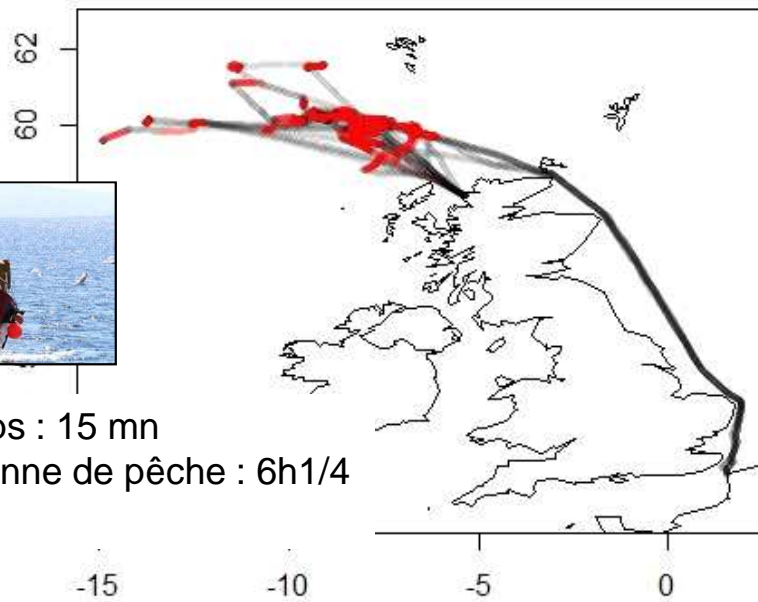
Le diable se cache aussi dans les détails des modèles HMM et HSMM ...

Nicolas Bez, IRD, Sète
Stéphanie Mahévas, Ifremer, Nantes
Marie-Pierre Etienne, AgroParisTech
Pierre Gloaguen, INRIA
Réseau PathTIS





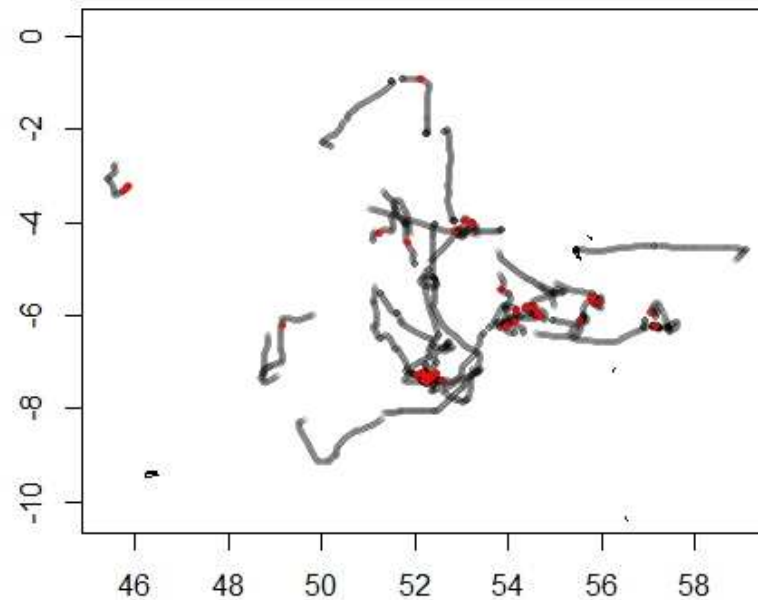
Pas de temps : 15 mn
Durée moyenne de pêche : 6h1/4
Ratio : 25

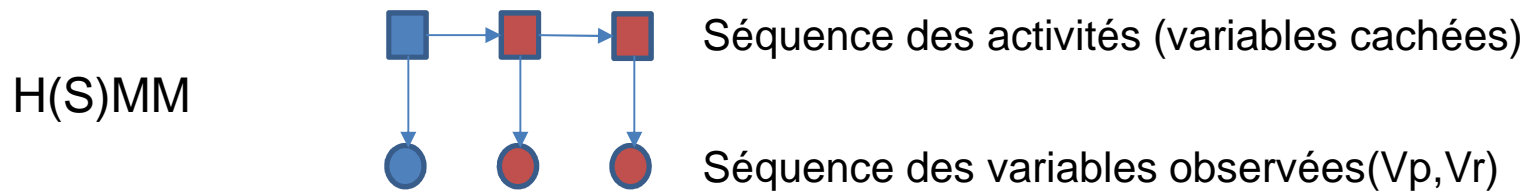


GPS (HD) + observateurs embarqués (données de validation)



Pas de temps : 5 mn
Durée moyenne de pêche : 2h
Ratio : 24





Problème : Les méthodes EM sont sensibles aux paramètres initiaux (et les méthodes Bayésiennes aux priors)

Changement valeurs initiales ~~→~~ changement d'estimation (résultats not shown).

L'EM retrouve correctement le MV.

Le MV dépend des données dont le pouvoir informatif n'est pas suffisant.

→ Vérifier position max. de vraisemblance versus paramètres du modèle par simulation

Plan d'expériences

Structures du modèle (4 modalités)

$V_p, V_r \sim$ gaussiennes indépendantes

2 états (NonPêche/Pêche)

HMM_geom versus HSMM_Poisson x ARO versus AR1

Résolution/Dégradations (4 modalités)

15mn, 30 mn, 1h, 2h ou 5mn, 15mn, 30mn, 1h

Nbr d'alternances (état_1, état_2) observées (5 modalités)



10, 50, 100, 150, 200

Méthodes d'inférence (2 modalités) R Package « mhsmm »

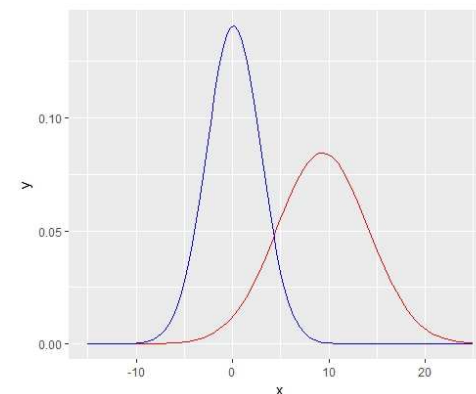
EM avec et sans Viterbi ({état} +probables versus séquence +probable)

Jeux de paramètres initiaux (2 ou 3 modalités selon le bon vouloir du cluster ...)

Bateau 1 : taux de chevauchement $V_p = 0.45$; $V_r = 0.99$

Bateau 3 : taux de chevauchement $V_p = 0.18$; $V_r = 0.96$

Cluster de calcul : 3 jours



Analyses des résultats

Inférences des paramètres

Vp (Gaussienne) : espérance , variance, (éta)
Vr (Gaussienne) : espérance , variance, (éta)
Durée (Géométrique/Poisson) : espérance

} x 2 états = 10 ou 12 paramètres

$\theta_{i,k}$ pour $i = 1, \dots, 12$ et $k = 1, \dots, 400$

$$\tilde{\theta}_{i,k} = \frac{\theta_{i,k}^* - \theta_{i,k}}{\theta_{i,k}}$$

$$\tilde{\theta}_i = \frac{1}{400} \sum_k \theta_{i,k}$$

$$\|\tilde{\theta}\| = \sqrt{\sum_i \tilde{\theta}_i^2}$$

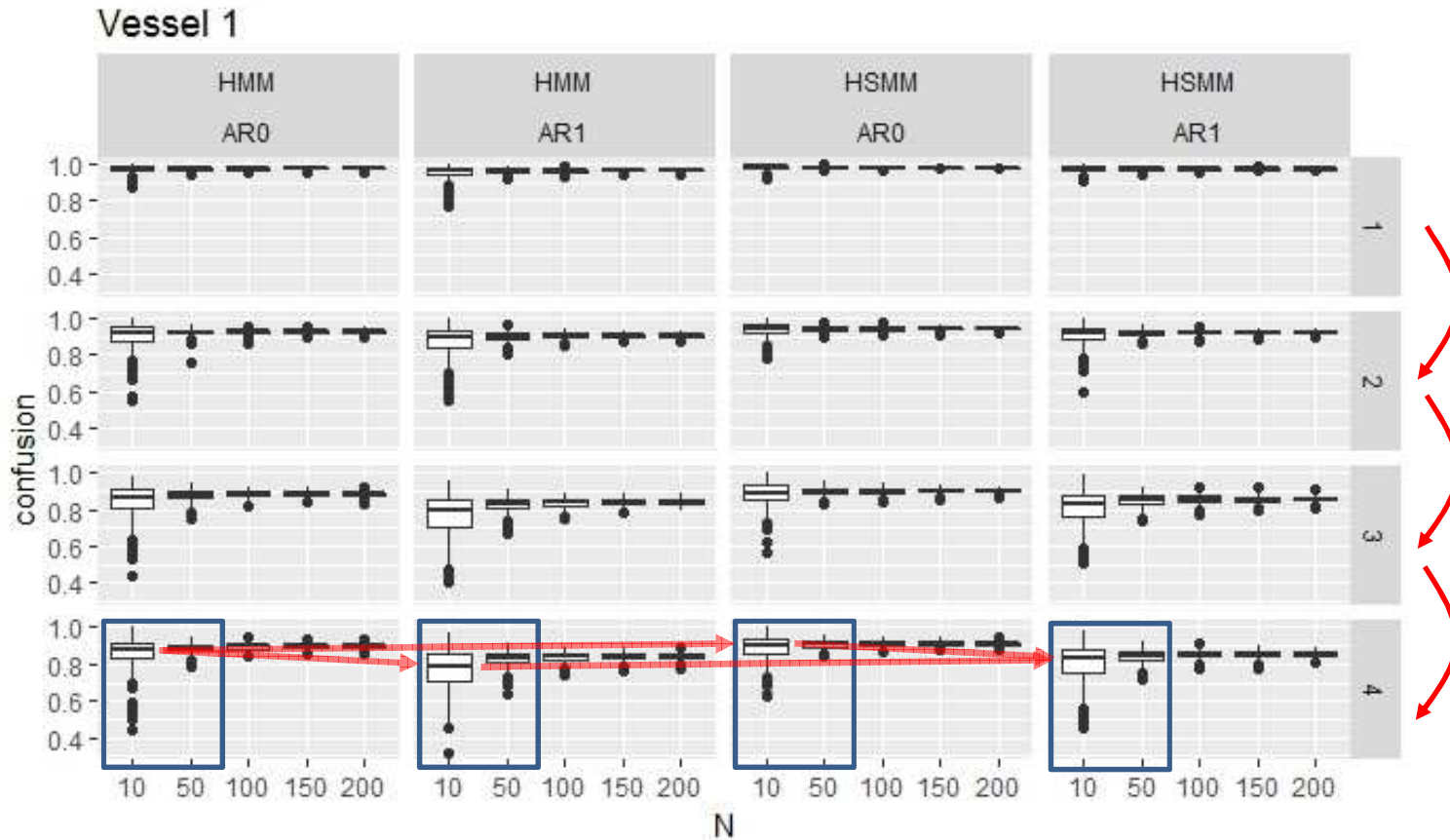
Inférence des états

Matrice de confusion (% de bonnes estimations)

Ajustement des états

Taux de bonnes estimations

HMM/HSMM x AR1/AR0 x nbre de séquences x dégradations



AoV : **HSMM** > **HMM** ; **AR0** > **AR1** ; dégradation 4 < dégradation 1 ; N200 > N10

Conclusion **Le Diable ne se cache pas QUE dans les détails...**

L'EM converge correcte

Vr est une variable

A dégradati
importe e
Rq : c'est

La dégrad

Viterbi $\leq m$

Taux de bonne
(une Lapalissade

en AR1).

rées

état



Conclusions

L'EM converge correctement (et toc !)

Vr est une variable qui perturbe plutôt qu'elle n'aide (moyenne en AR0 et éta en AR1).

A dégradation fixée, i.e. à résolution fixée, le nbr d'alternances (état1,état2) observées importe et ce, d'autant plus que la dégradation est importante.

Rq : c'est presque une Lapalissade mais pas vraiment

La dégradation dégrade mais pas de façon univoque

Viterbi \leq marginal

Taux de bonnes prédictions invers.# taux de confusion entre loi d'émission par état
(*une Lapalissade !*)

Le Diable ne se cache pas QUE dans les détails...

