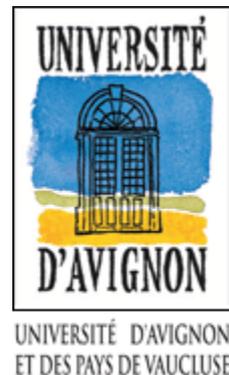


Quelques exemples d'applications de méthodes stochastiques pour l'optimisation de modèles numériques

Hubert Vincent VARELLA



1) Le filtre particulaire appliqué à un modèle de dynamique des populations

Stage Master 2 avec Jean-Pierre Vila, INRA Montpellier
et Roger Pradel, CNRS Montpellier

2) La méthode Importance Sampling appliquée à un modèle de culture

Doctorat avec Martine Guérif et Samuel Buis, INRA Avignon

3) L'assimilation d'ensemble appliquée à un modèle météorologique

Post-Doctorat avec Gérald Desroziers et Loïk Berre, Météo-France

1) Le filtre particulaire appliqué à un modèle de dynamique des populations

Stage Master 2 avec Jean-Pierre Vila, INRA Montpellier
et Roger Pradel, CNRS Montpellier

Compréhension et protection des populations fragiles, cas du flamant rose



Le flamant rose est une espèce fragile sensible à la météorologie, aux prédateurs, à l'homme, ...

La compréhension de cette population peut contribuer à sa protection

Modéliser la dynamique des flamants roses afin de prévoir son évolution et ses caractéristiques démographiques

Modèle stochastique

$$\left\{ \begin{array}{l} x_{t+1} = f_t(x_t, \epsilon_t) \longrightarrow \text{équation d'état} \\ y_t = h_t(x_t, \delta_t) \longrightarrow \text{équation d'observation} \end{array} \right.$$

$$\left\{ \begin{array}{l} N_{t+1}^a \sim \text{Bin}(N_t^a, \phi^a) + \text{Bin}(N_t^p, \phi^p) \\ Y_t^a \sim N(N_t^a, \sigma_{Y^a}^2) \\ Y_t^p \sim N(N_t^p, \sigma_{Y^p}^2) \end{array} \right.$$

vecteur d'états = {nb adultes N_t^a , nb poussins N_t^p }

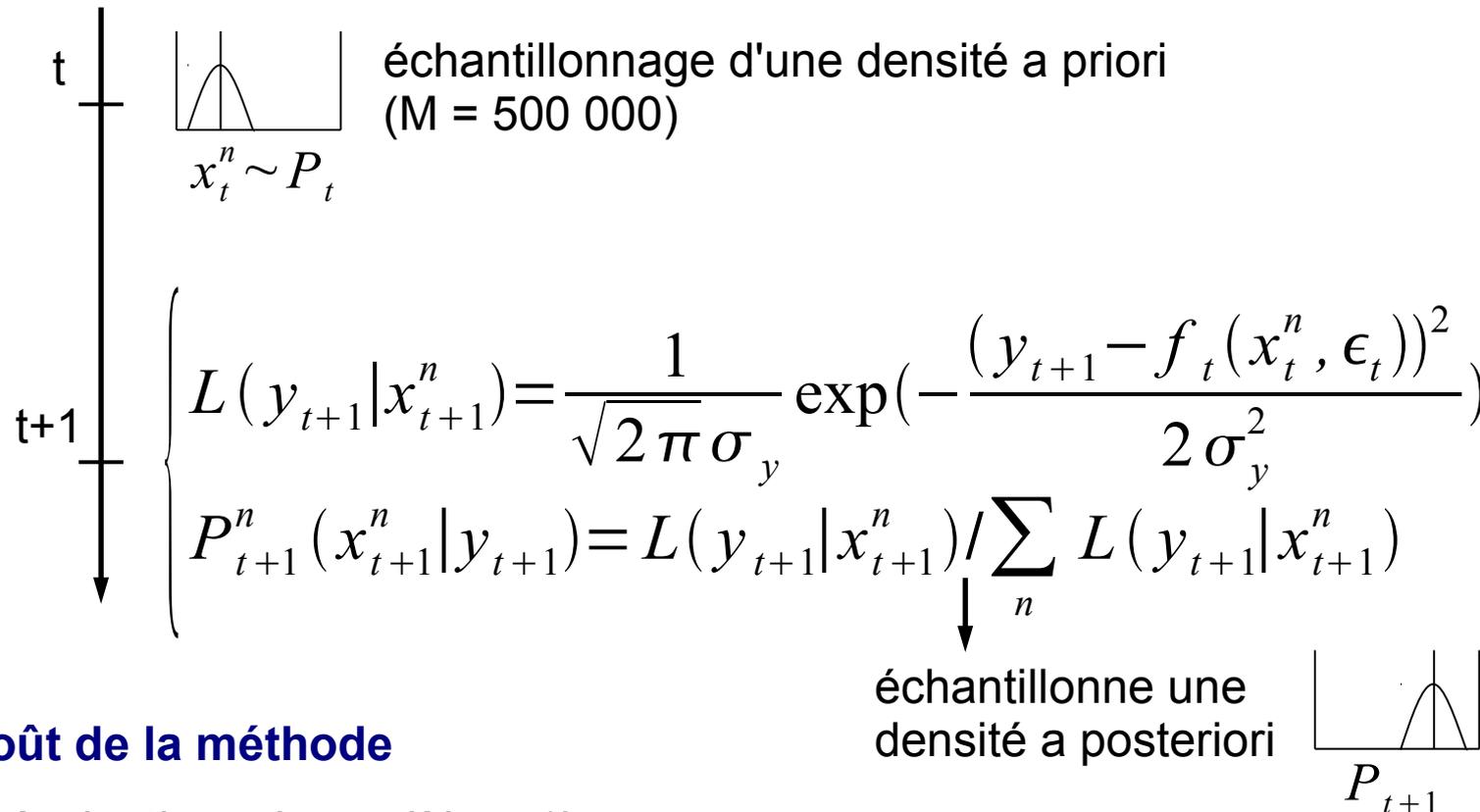
vecteur d'observations = {nb adultes Y_t^a , nb poussins Y_t^p , survie adulte Φ^a , survie poussin Φ^p }

Choix du filtre particulaire

Il relève de la théorie du filtrage non-linéaire en temps discret (Del Moral 1997 ; Doucet et al 2001)

Adapté au modèle de simulation peu coûteux (centième de seconde)

Principe de la méthode

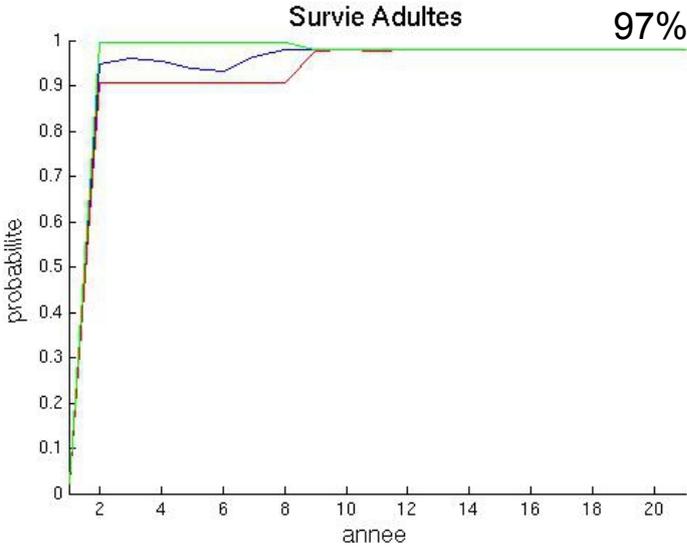
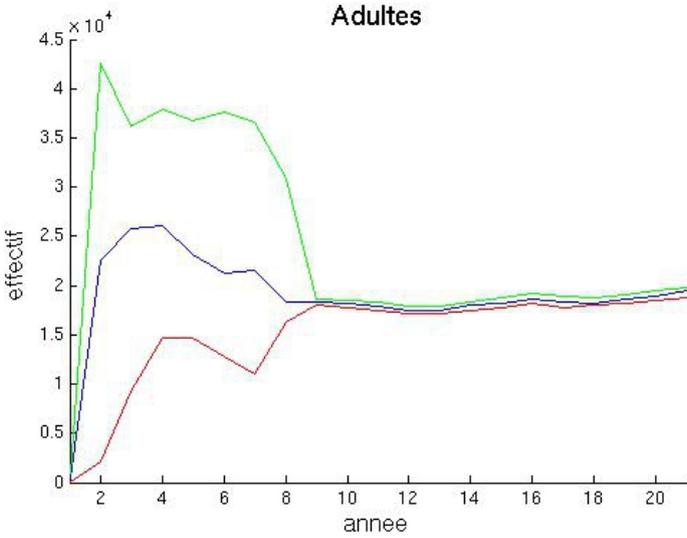
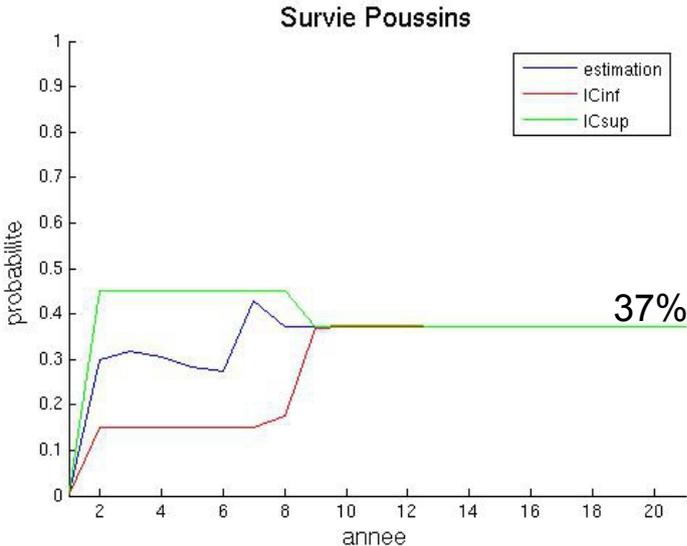
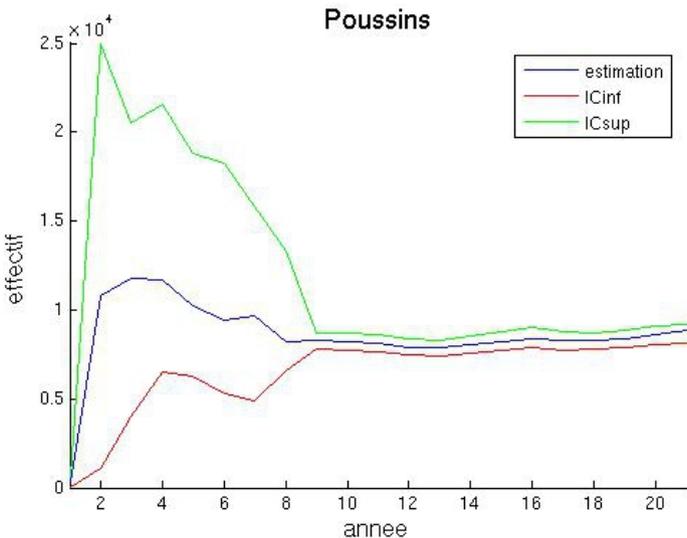


Coût de la méthode

M évaluations du modèle = 1h

Résultats avec des observations réelles

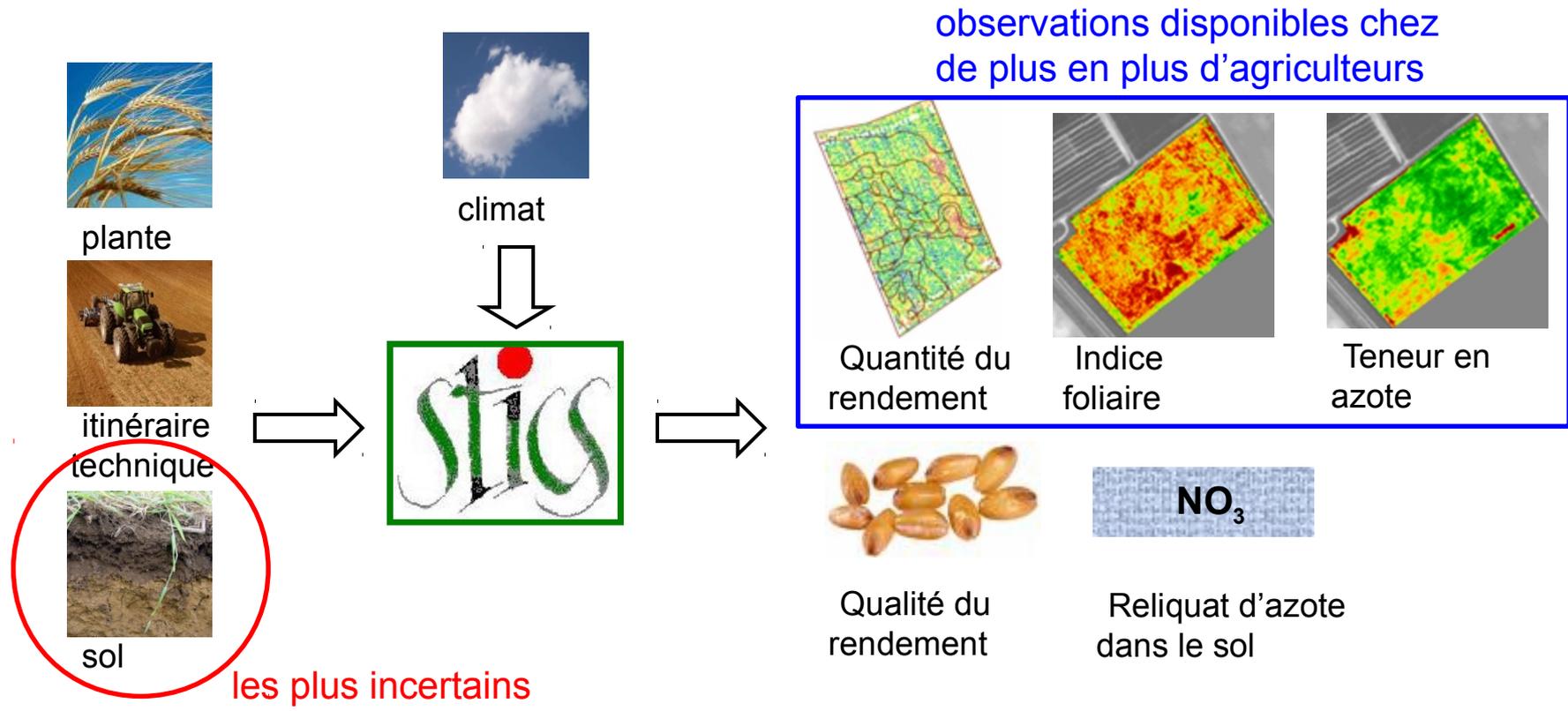
Flamants roses observés en Camargue



2) La méthode Importance Sampling appliquée à un modèle de culture

Doctorat avec Martine Guérif et Samuel Buis, INRA Avignon

Une bonne prévision du système sol-plante permet de raisonner les pratiques agricoles



Inversion du modèle STICS pour estimer les propriétés des sols et améliorer la prévision

STICS est un modèle complexe avec un grand nombre de paramètres sol

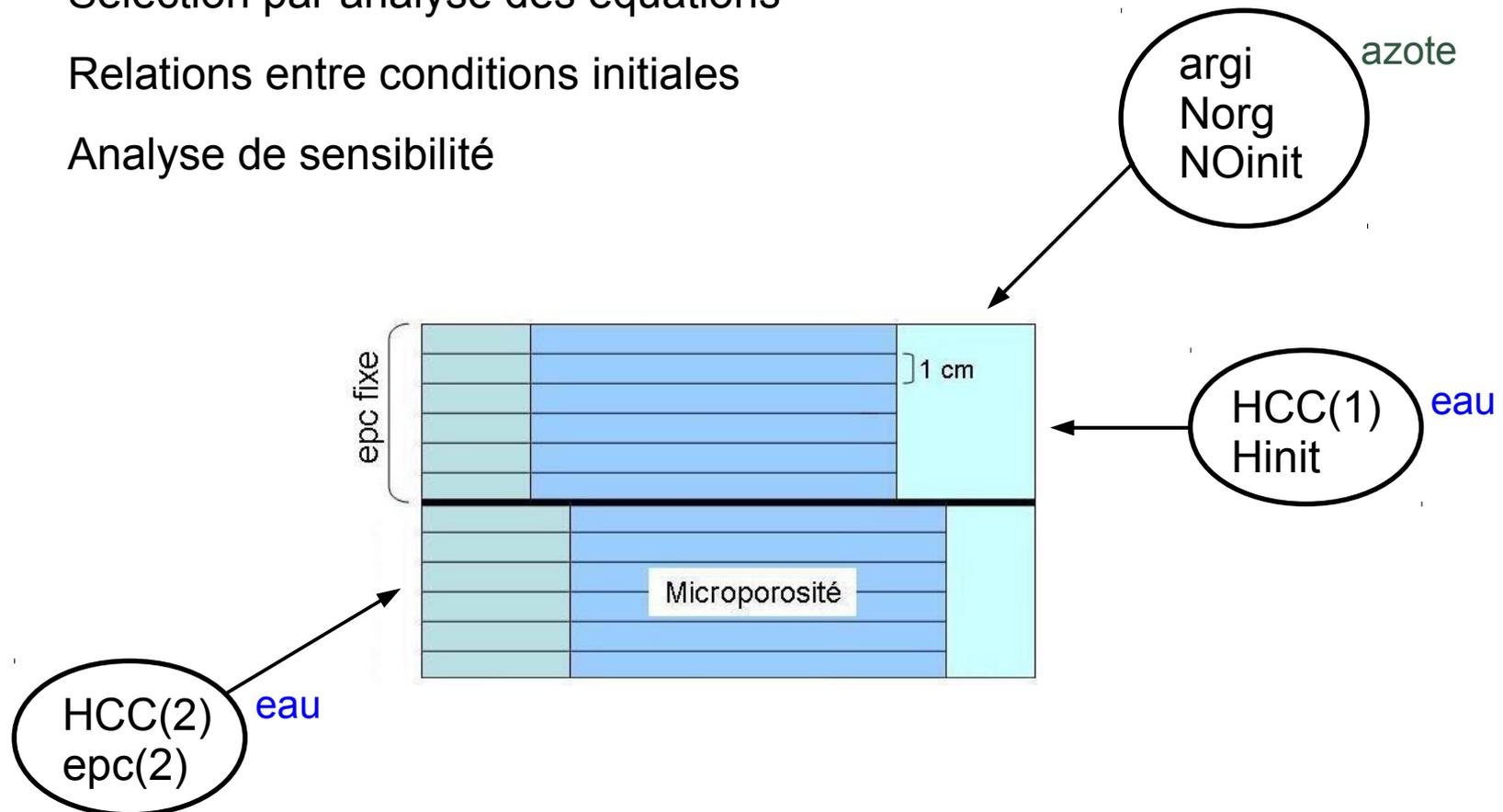
Sélection d'un sous-groupe de paramètres à estimer

Simplification du formalisme du sol

Sélection par analyse des équations

Relations entre conditions initiales

Analyse de sensibilité

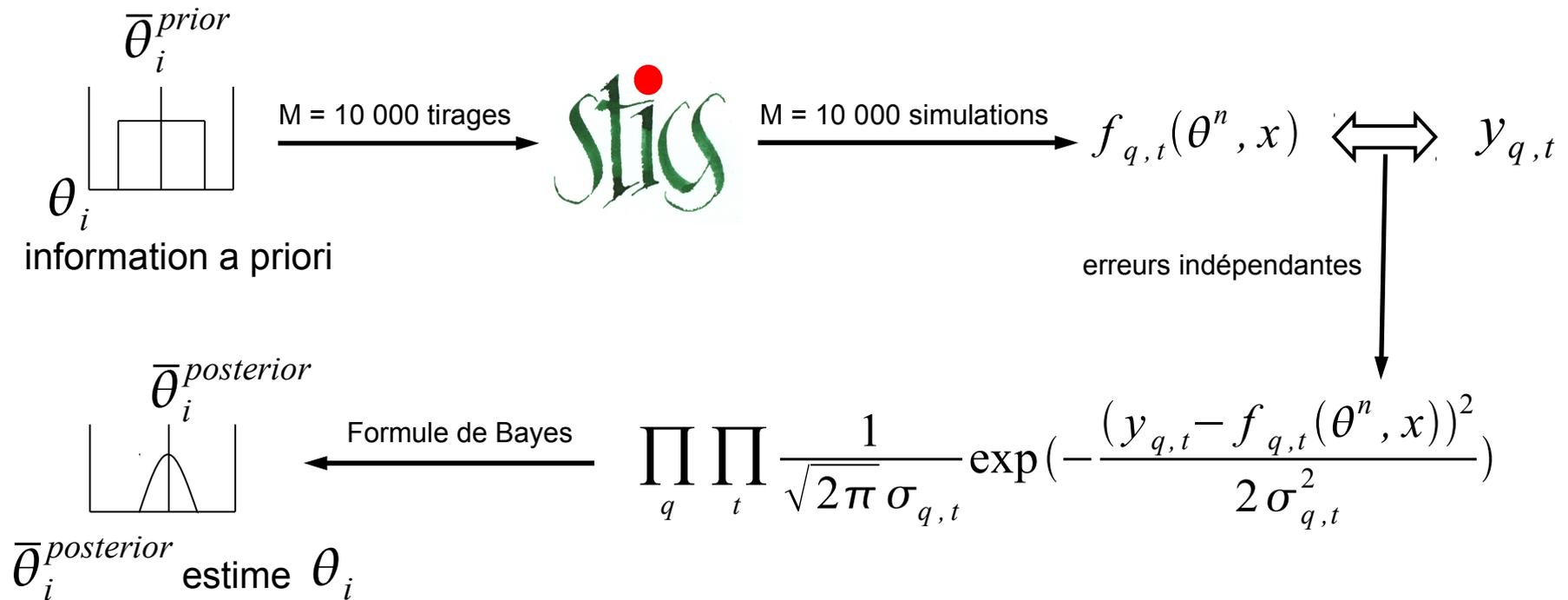


Choix de Importance Sampling (Liu 2001)

Elle considère une information a priori sur la parcelle

Simulations faites une fois pour toute (contrairement à MCMC)

Principe de la méthode



Coût de la méthode

M évaluations du modèle = 2h30min

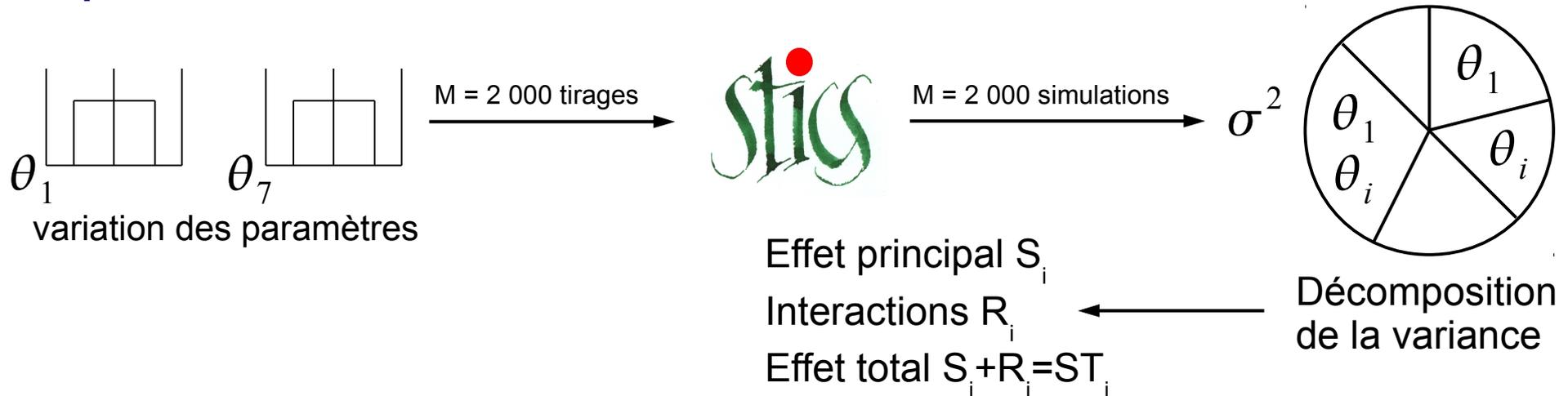
Comment mesurer la quantité d'information permettant d'estimer les paramètres ?

Peut être mesurée par la sensibilité locale (Fisher) mais dépend de la valeur des paramètres

Proposition d'un critère basé sur l'analyse de sensibilité globale (ASG) pour mesurer la quantité d'information sur la gamme de variation des paramètres

Méthode ASG basée sur la décomposition de la variance (Saltelli et al 1999)

Principe de la méthode



Coût de la méthode

M évaluations du modèle = 30min

Critère ASG qui mesure la quantité d'information

$$GMS_i = \frac{1}{k} \sum_k \boxed{(1 - \alpha_k) (S_i^k - R_i^k)}$$

~ 0 si $ST \sim 0$ (peu d'information)

~ 1 si $S \sim 1$ (beaucoup d'information)

~ -1 si $R \sim 1$ (problème d'identifiabilité)

~ 1 si $\alpha \sim 0$ (variable non corrélée aux autres)

~ 0 si $\alpha \sim 1$ (variable corrélée aux autres)

Critère qui mesure la qualité d'estimation

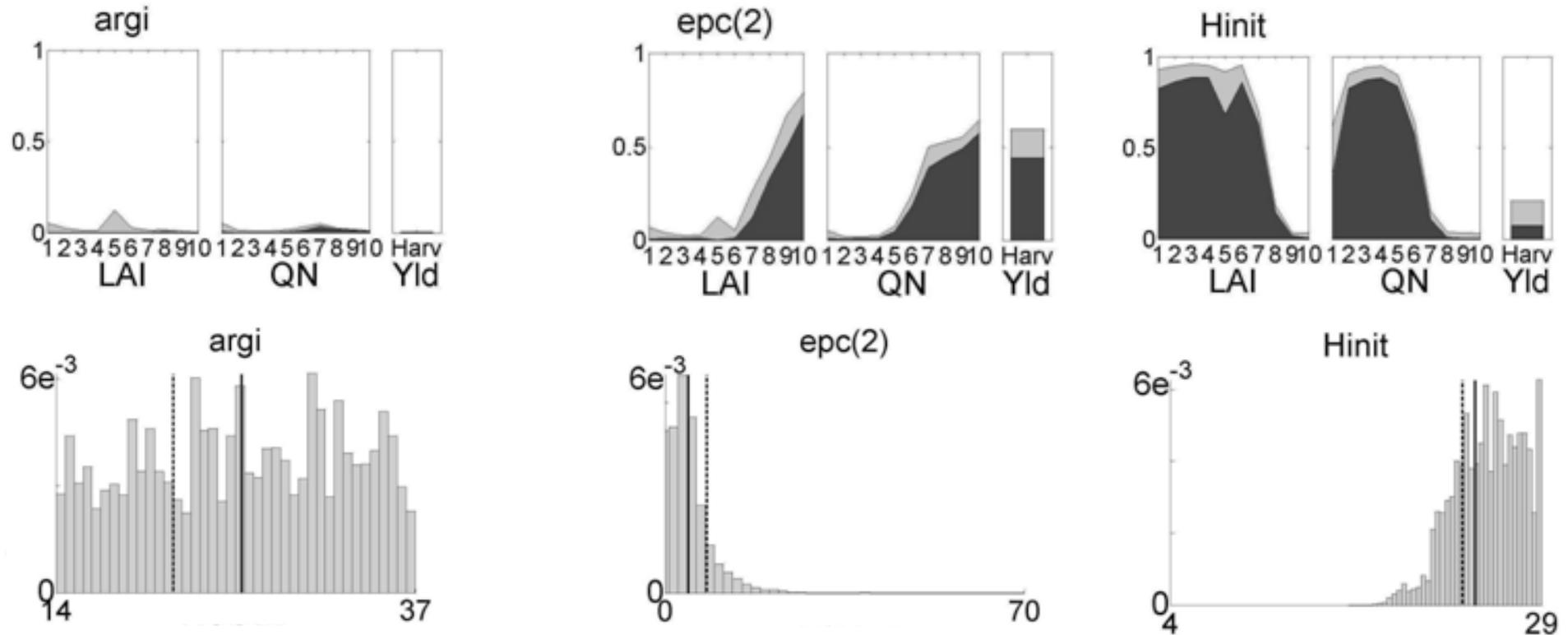
$$RE_i = \frac{RMSE(\bar{\theta}_i^{posterior})}{RMSE(\bar{\theta}_i^{prior})}$$

~ 1 si faible amélioration (qualité d'estimation faible)

~ 0 si grande amélioration (qualité d'estimation grande)

Résultats avec des observations simulées

10 LAI + 10 QN + Yld sur une culture de blé en climat sec et en faible profondeur de sol

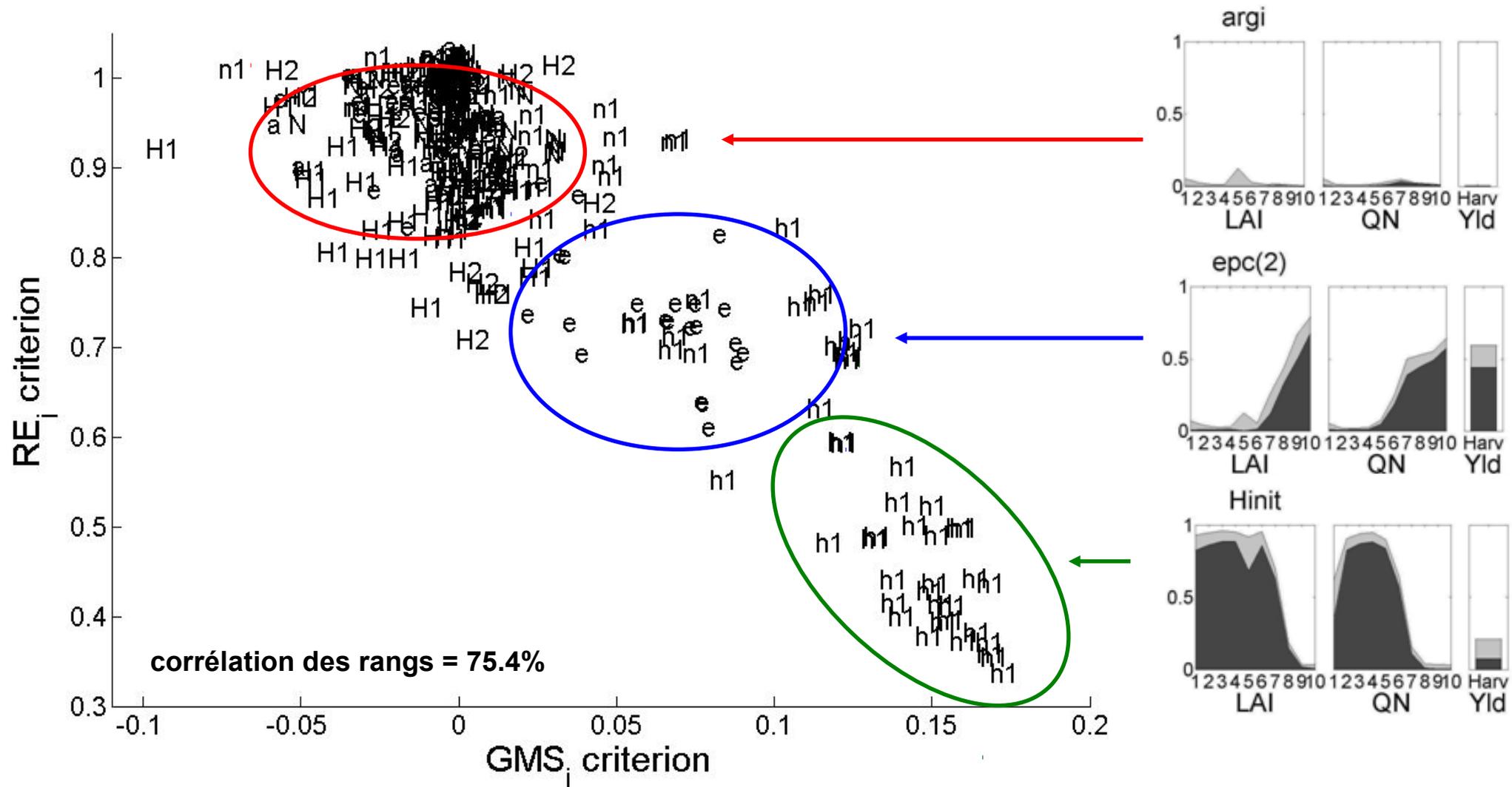


faible sensibilité :
mauvaise qualité d'estimation

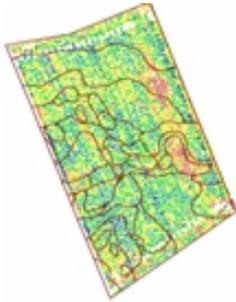
forte sensibilité :
bonne qualité d'estimation

- Effet principal S
- Interactions R
- valeur vraie
- valeur estimée

Résultats avec des observations simulées



Quel est l'effet de la configuration d'observation sur la qualité d'estimation et de prévision ?



quantité du rendement (Yld)



qualité du rendement (Prot)



reliquat d'azote dans le sol(Nit)

Critère qui mesure la qualité d'estimation

$$RE_i = \frac{RMSE(\bar{\theta}_i^{posterior})}{RMSE(\bar{\theta}_i^{prior})}$$

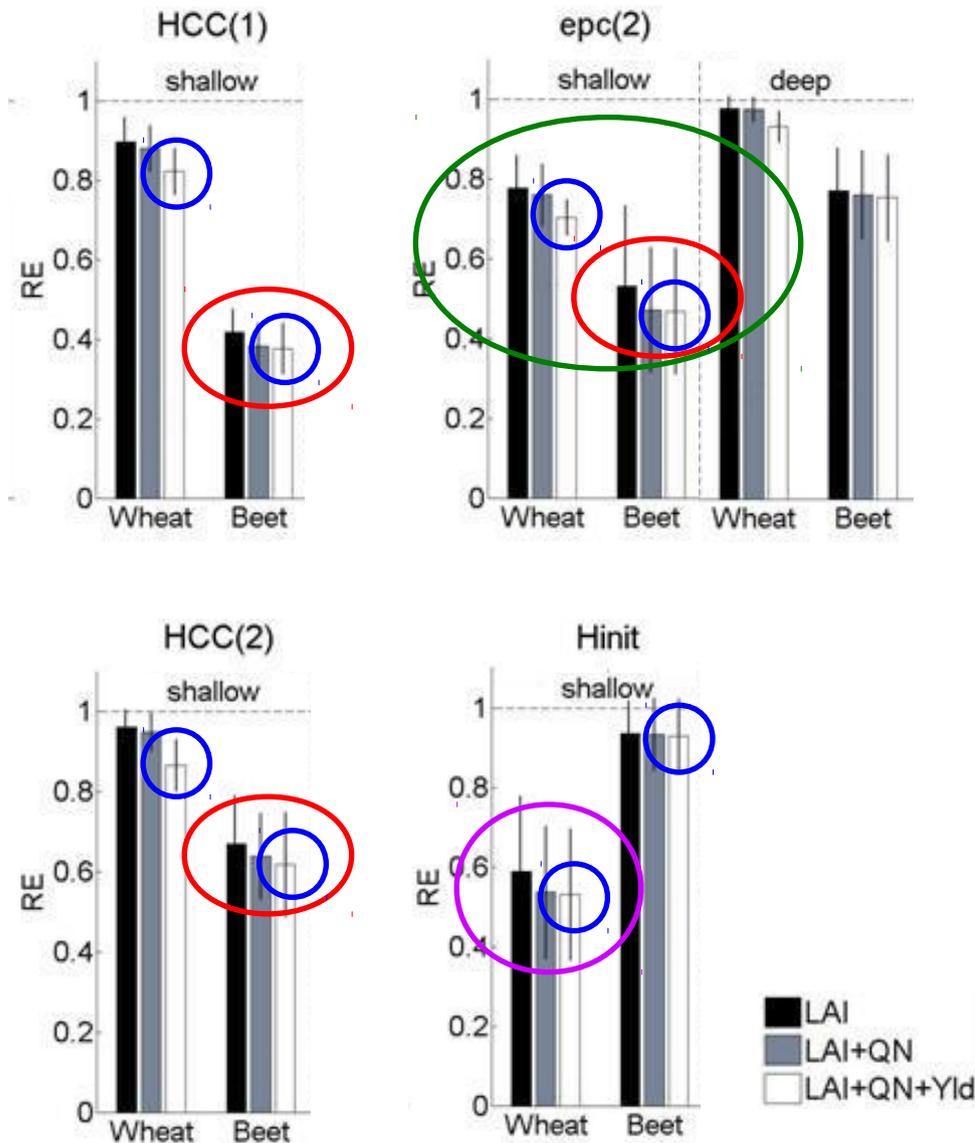
= 1 si faible amélioration (qualité d'estimation faible)
= 0 si grande amélioration (qualité d'estimation grande)

Critère qui mesure la qualité de prévision

$$REP_j = \frac{RMSEP_j(\bar{\theta}^{posterior})}{RMSEP_j(\bar{\theta}^{prior})}$$

= 1 si faible amélioration (qualité de prédiction faible)
= 0 si grande amélioration (qualité de prédiction grande)

Résultats avec des observations simulées (estimation)



Les paramètres liés à l'état hydrique sont améliorés

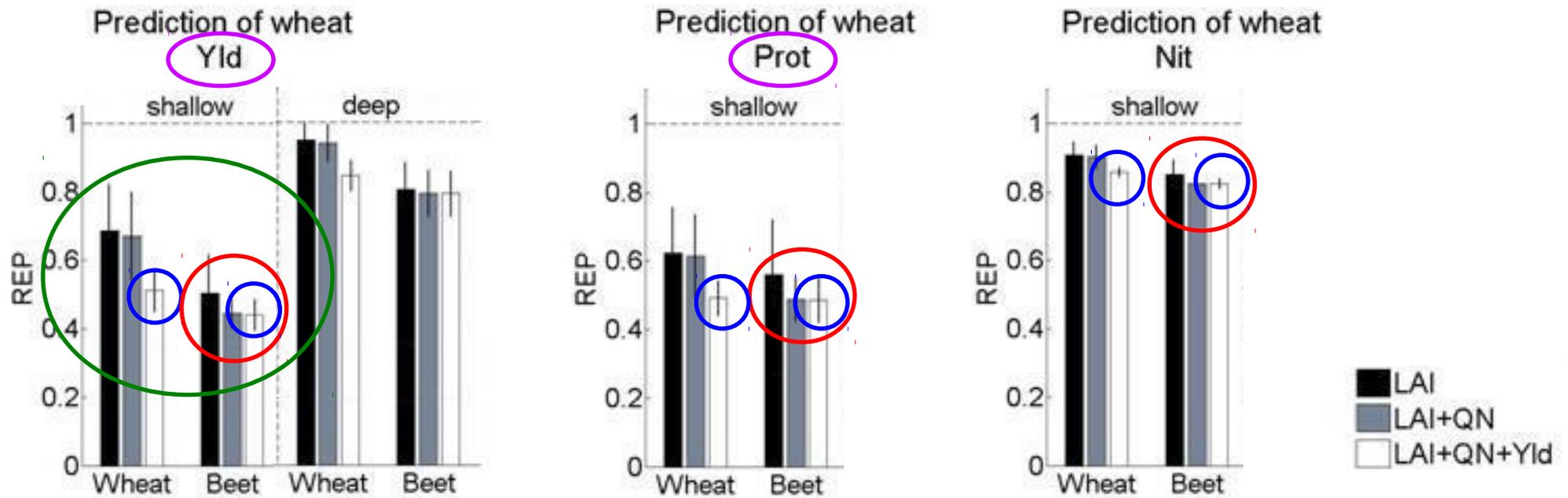
Hinit de meilleure qualité avec des observations de **blé**

La **troisième taille** du jeu d'observations donne la meilleure qualité

L'observation de la **betterave** donne une meilleure qualité d'estimation

epc(2) de meilleure qualité en **faible profondeur de sol**

Résultats avec des observations simulées (prévision)



Possible d'améliorer beaucoup **rendement et qualité** mais pas le reliquat d'azote

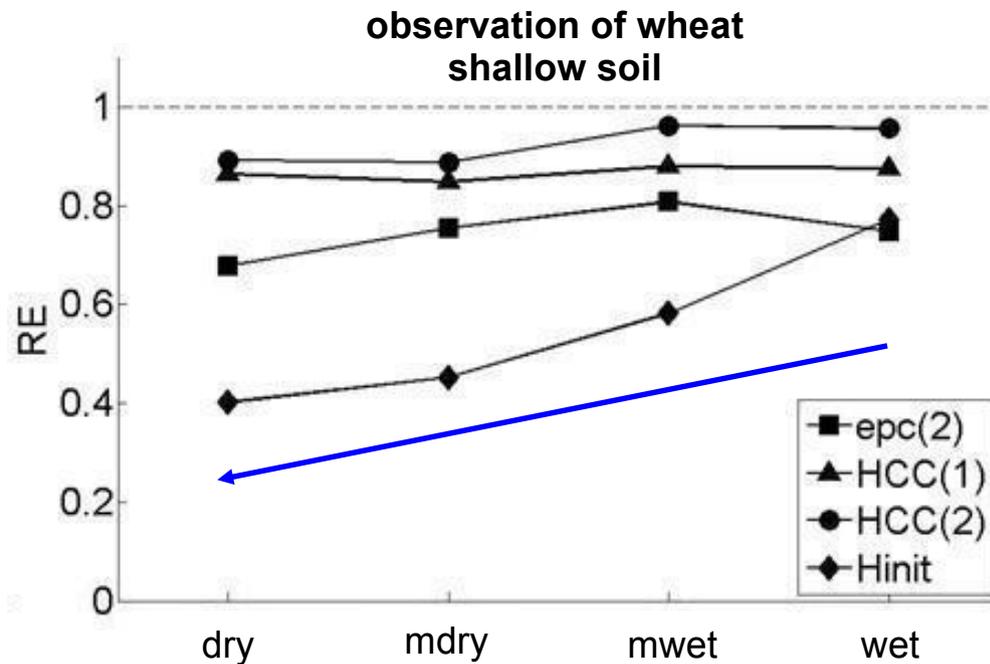
La **troisième taille** du jeu d'observations donne la meilleure qualité de prédiction

L'observation de la **betterave** donne une meilleure qualité de prédiction

Une **faible profondeur de sol** donne une meilleure qualité de prédiction

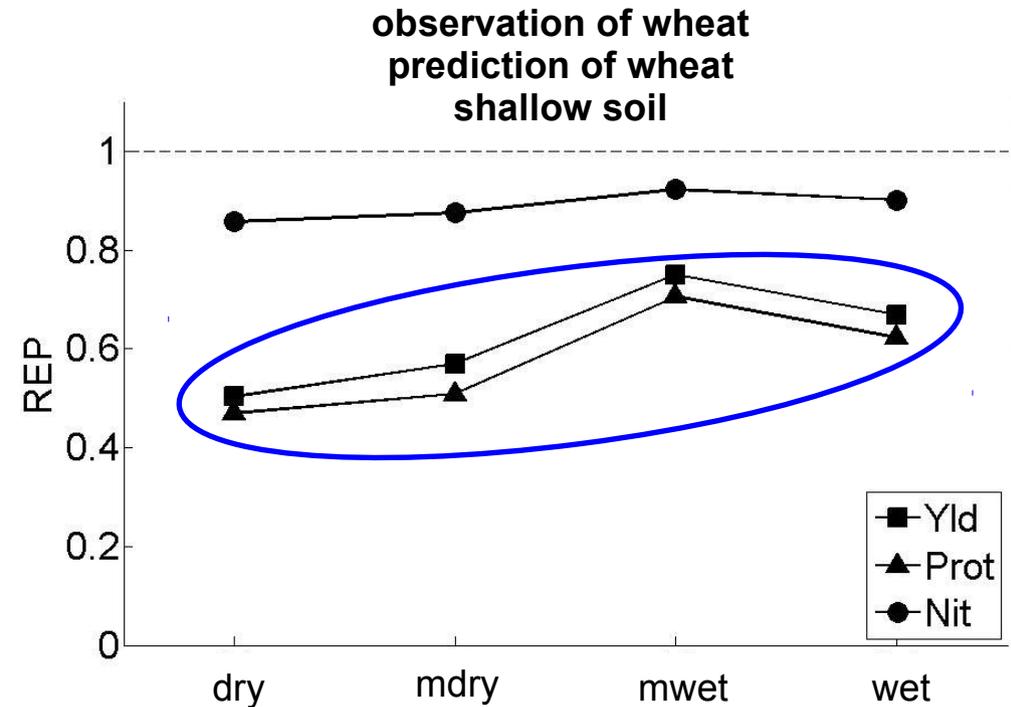
Résultats avec des observations simulées : effet du climat

Estimation



Plus le climat est **sec** plus la qualité d'estimation des paramètres hydriques est bonne

Prévision

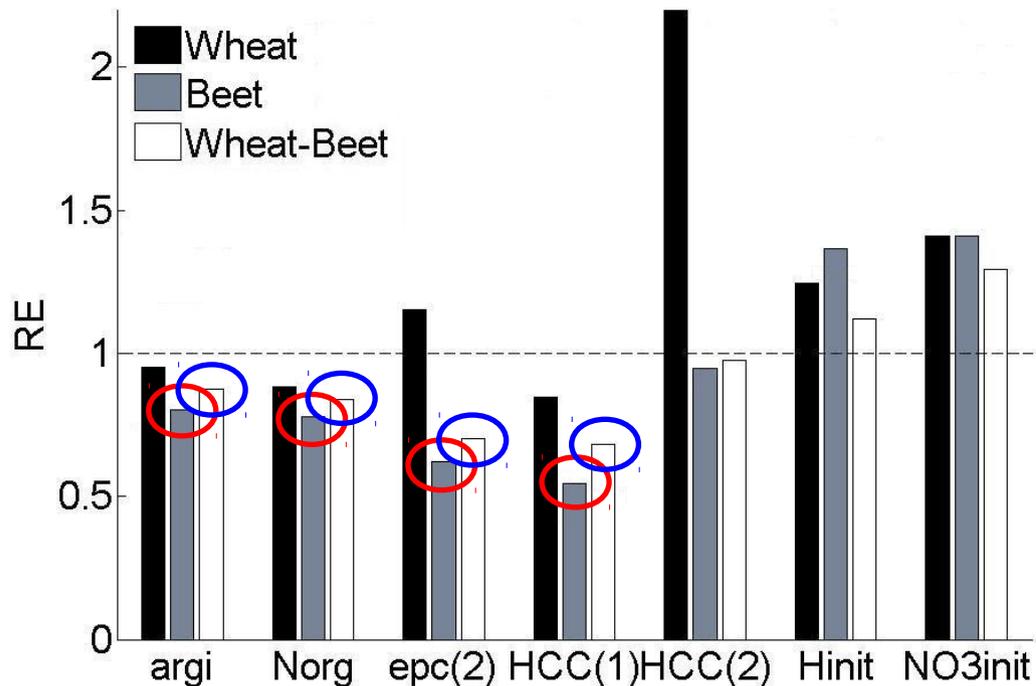


Plus le climat est **sec** plus la qualité de prévision du rendement est bonne

Le climat n'affecte pas la qualité de prévision du reliquat d'azote

Résultats avec des observations réelles (estimation)

Parcelles observées dans le bassin versant de Bruyères



Hinit et *NO3init* non améliorés

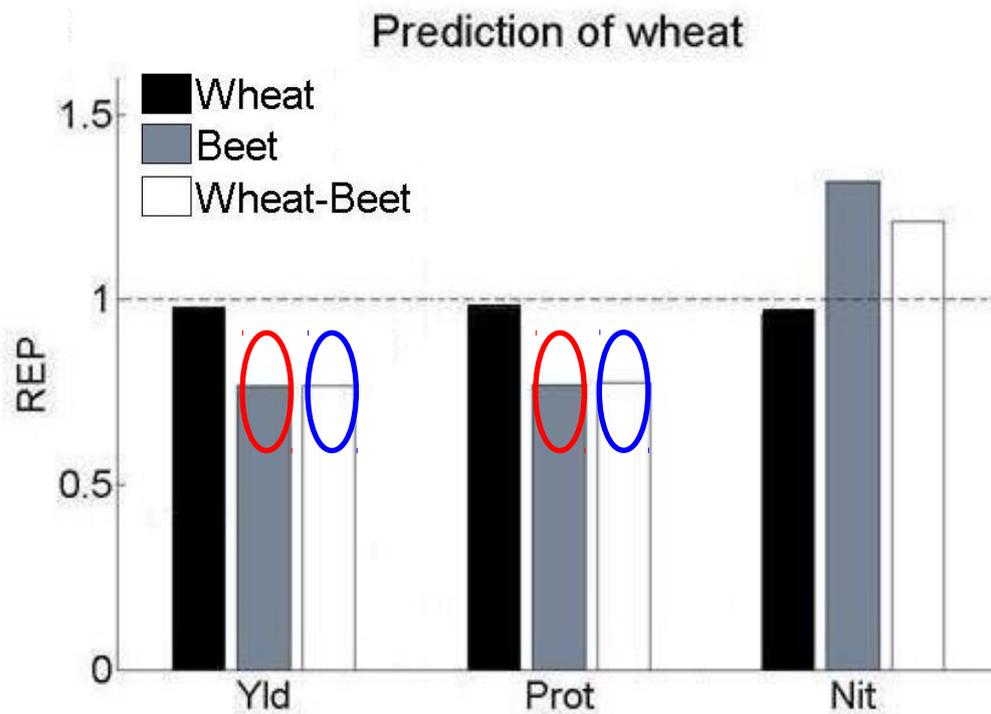
L'observation de **betterave** donne une meilleure qualité d'estimation

epc(2) et *HCC(1)* beaucoup améliorés

L'observation de **blé et betterave** n'améliore pas l'estimation

Résultats avec des observations réelles (prévision)

Parcelles observées dans le bassin versant de Bruyères



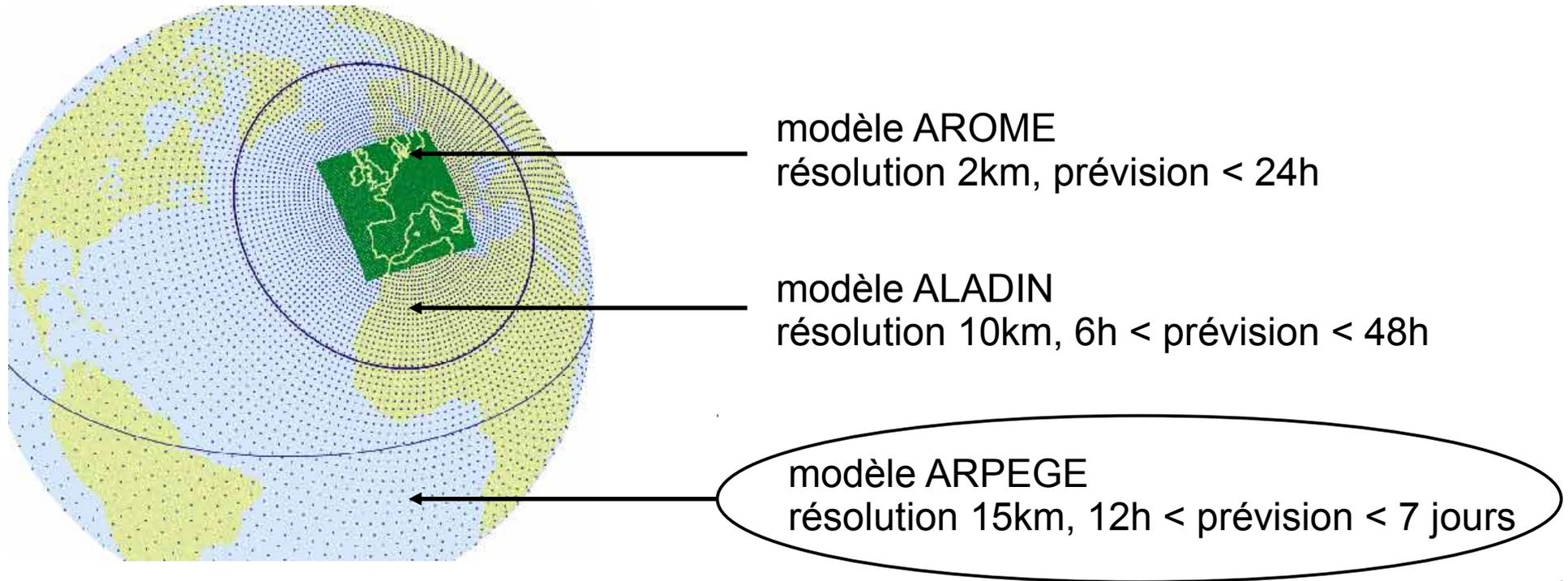
L'observation de **betterave** donne la meilleure qualité de prévision du rendement

L'observation de **blé et betterave** n'améliore pas la prévision

Le reliquat d'azote n'est pas amélioré

3) L'assimilation d'ensemble appliquée à un modèle météorologique Post-Doctorat avec Gérald Desroziers et Loïk Berre, Météo-France

Les modèles de prévision numérique à Météo-France



Le modèle ARPEGE

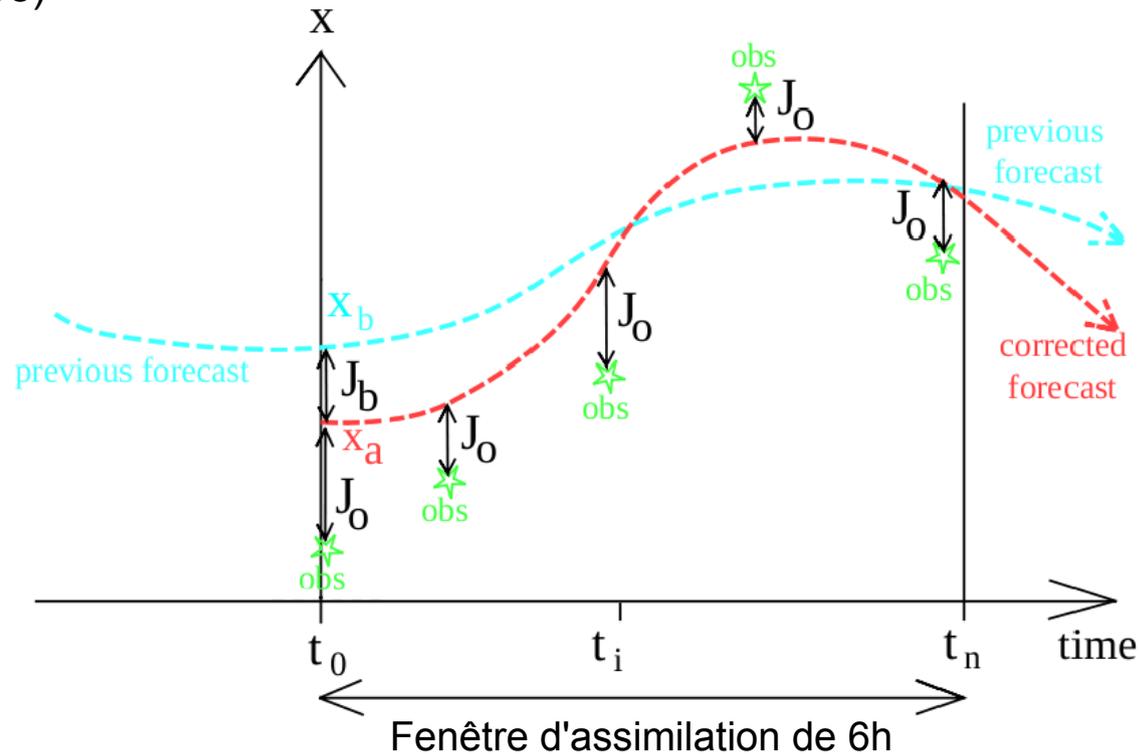
Observations : de surface, aéroportées, satellites, radar, ...

Variables simulées : pression de surface, tourbillon, divergence, température et humidité

Vecteur d'état de dimension 10^8

Les modèles de prévision numérique à Météo-France

La prévision ARPEGE est faite après une étape d'assimilation de données variationnelle 4D (Veersé et Thépaut 1998)



qui consiste à minimiser la fonctionnelle $J(x)$

$$J(x) = \frac{1}{2} (x - x^b)^T B^{-1} (x - x^b) + \frac{1}{2} \sum_{i=0}^n (y_i^o - H_i x(t_i))^T R_i^{-1} (y_i^o - H_i x(t_i))$$

Estimation de la matrice de covariance d'erreur d'ébauche B

$$B = \lim_{N \rightarrow +\infty} \frac{1}{N-1} \sum_{k=1}^N \epsilon_k^b (\epsilon_k^b)^T \quad \text{avec} \quad \epsilon_k^b = x_k^b - x^t$$

x^t inconnu $\hat{\epsilon}_k^b = x_k^b - \bar{x}^b$

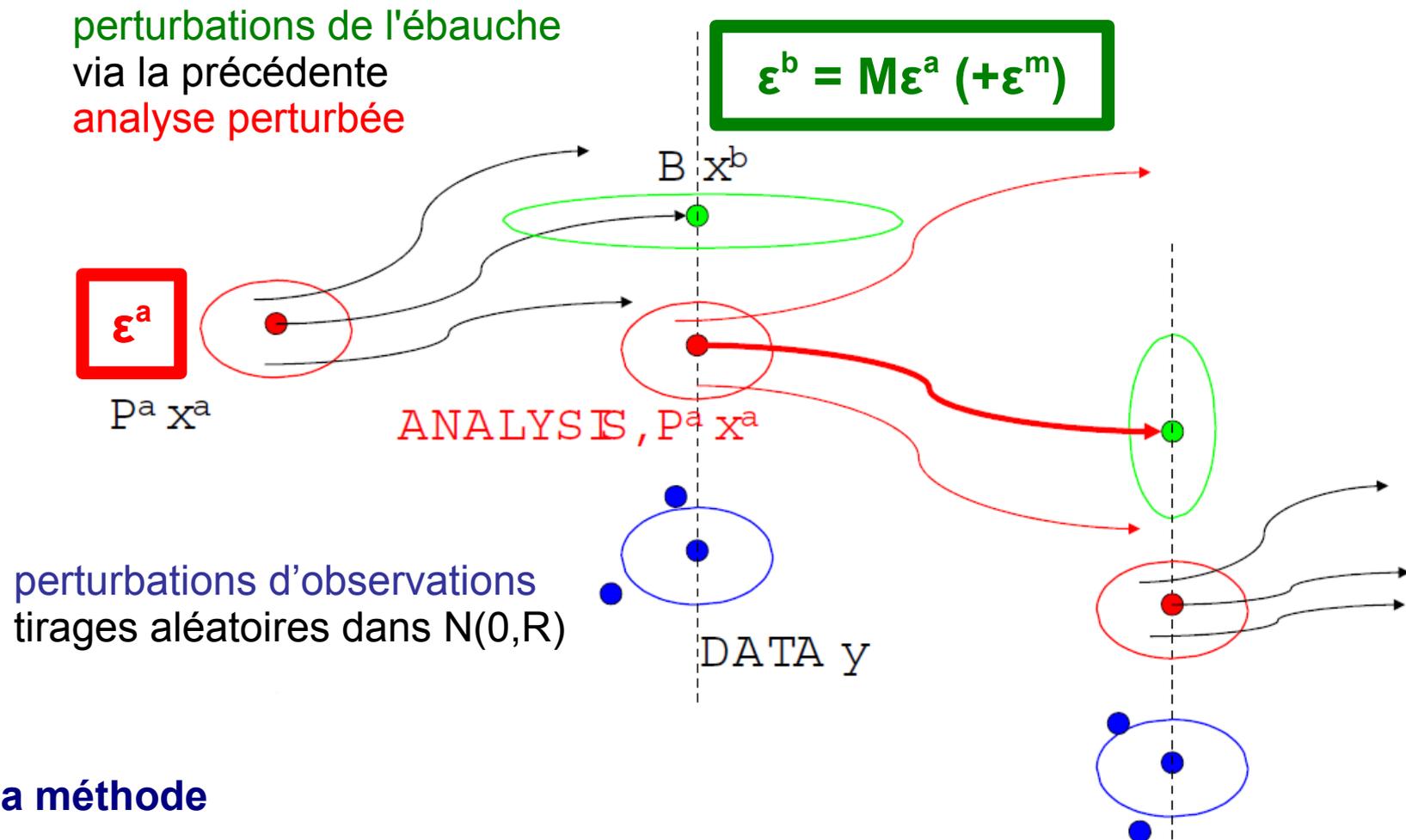
Double problème de simulation

- 1) Comment simuler les erreurs d'ébauche ϵ_k^b pour estimer B ?
- 2) Comment estimer B à partir d'un petit nombre de simulations de ϵ_k^b ?

$$\hat{B} = \frac{1}{N-1} \sum_{k=1}^N \hat{\epsilon}_k^b (\hat{\epsilon}_k^b)^T \quad \text{avec} \quad N \ll +\infty$$

Principe de l'assimilation d'ensemble ARPEGE

simulation stochastique de l'évolution des erreurs au cours du cycle d'assimilation par un ensemble de N perturbations (Houtekamer et al 1996 ; Fisher 2003 ; Ehrendorfer 2006)



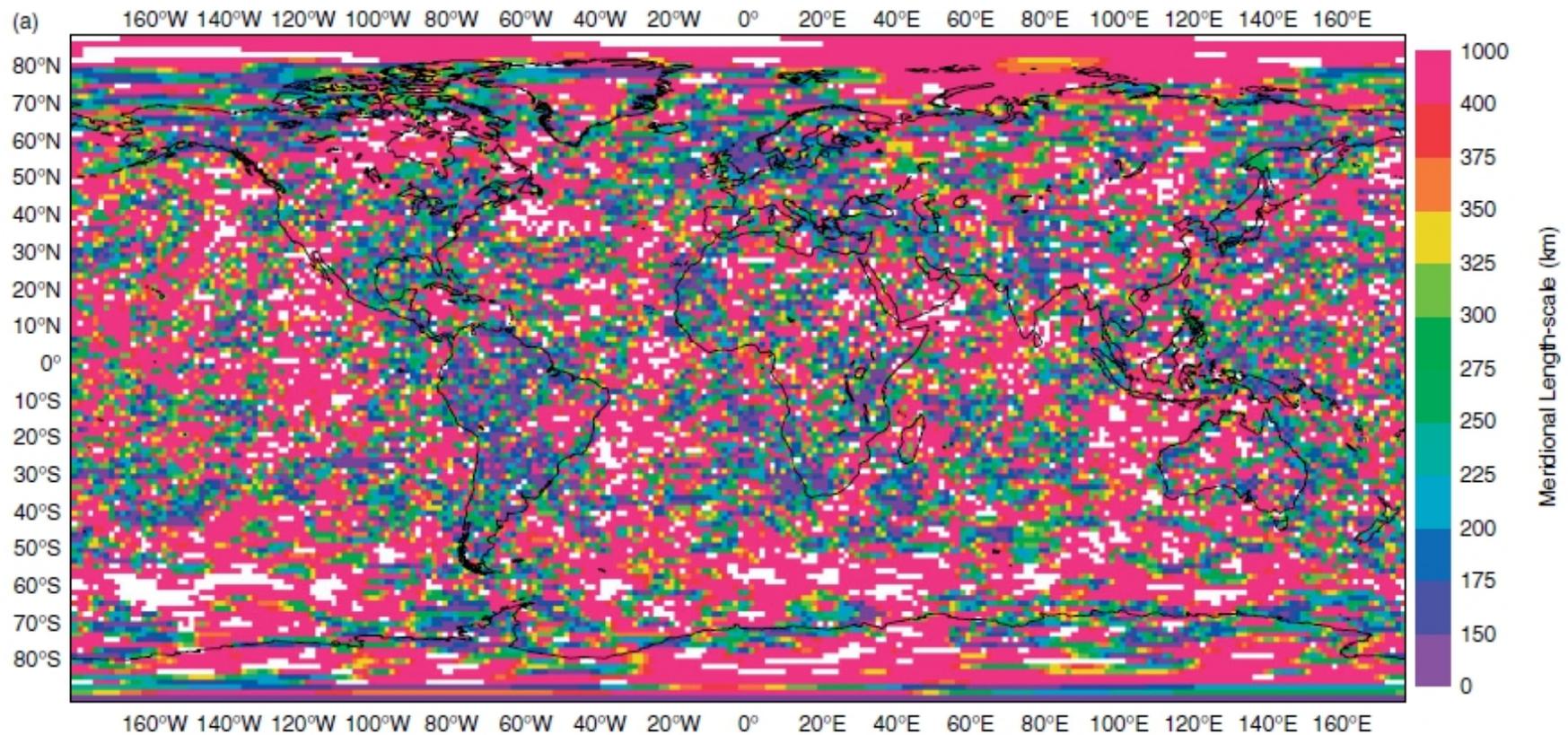
Coût de la méthode

Une simulation de $\epsilon_k^b = 30\text{min}$

Contraintes opérationnelles

Le nombre de simulations de ϵ_k^b par fenêtre d'assimilation est fixé à N=6

Cas des corrélations d'erreur d'ébauche



Estimation brute des portées de corrélation du vent pour une fenêtre d'assimilation et avec N=6

Le bruit est filtré une fois pour toute dans l'opérationnel

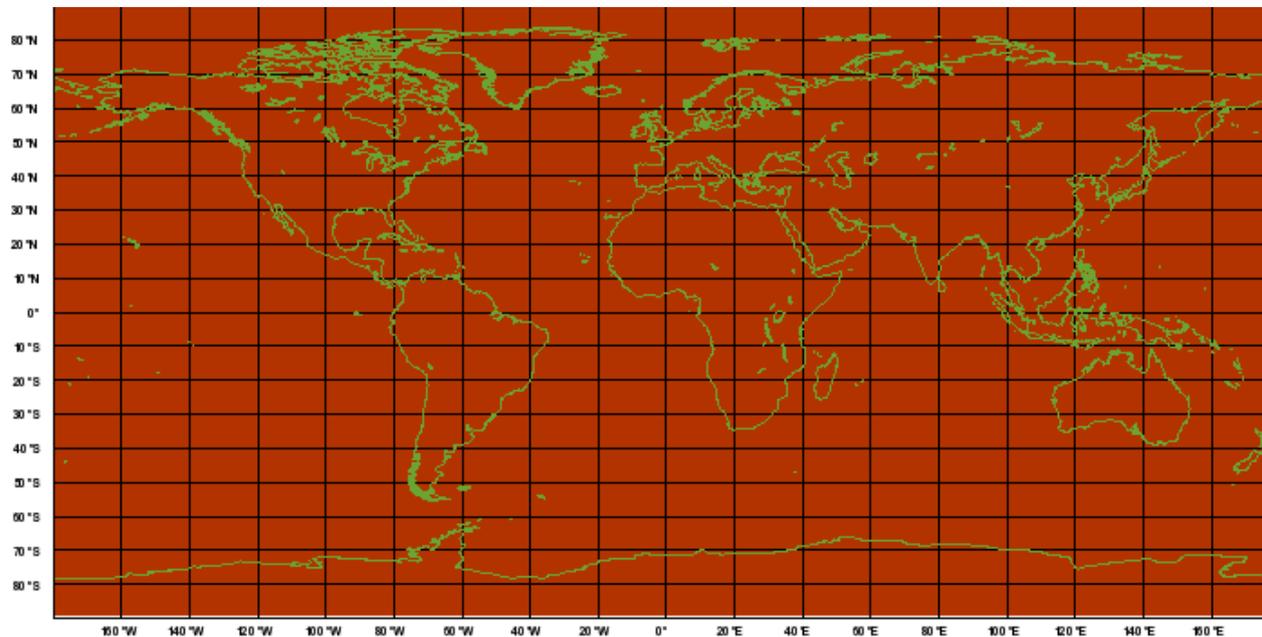
simulation de ε_k^b sur quelques semaines (500 simulations une fois pour toute)

+

calcul d'une moyenne globale (approche spectrale)

=

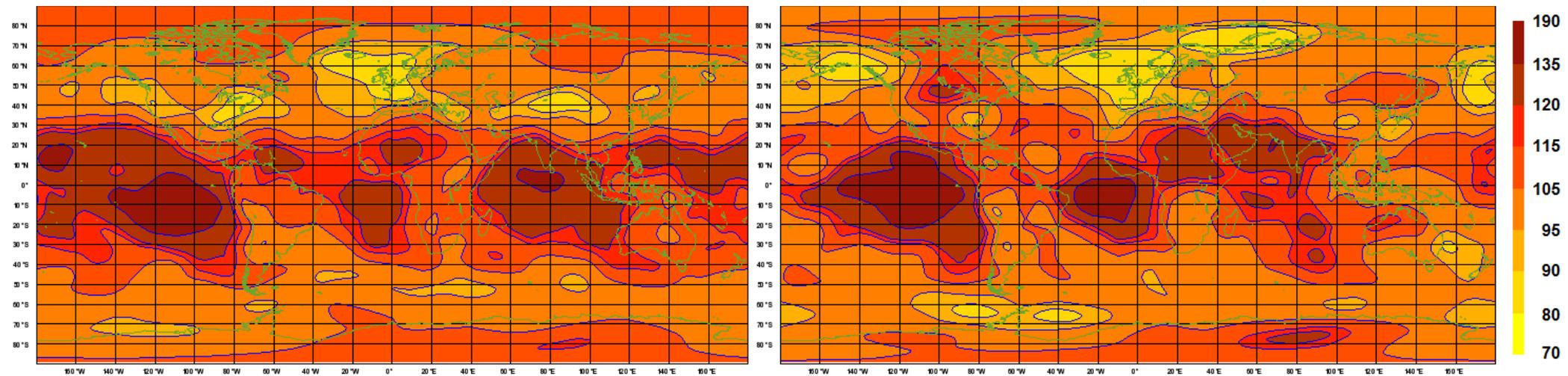
corrélations homogènes et climatologiques



Portées de corrélation du vent utilisées par l'opérationnel (100km)
pour le niveau 500hPa

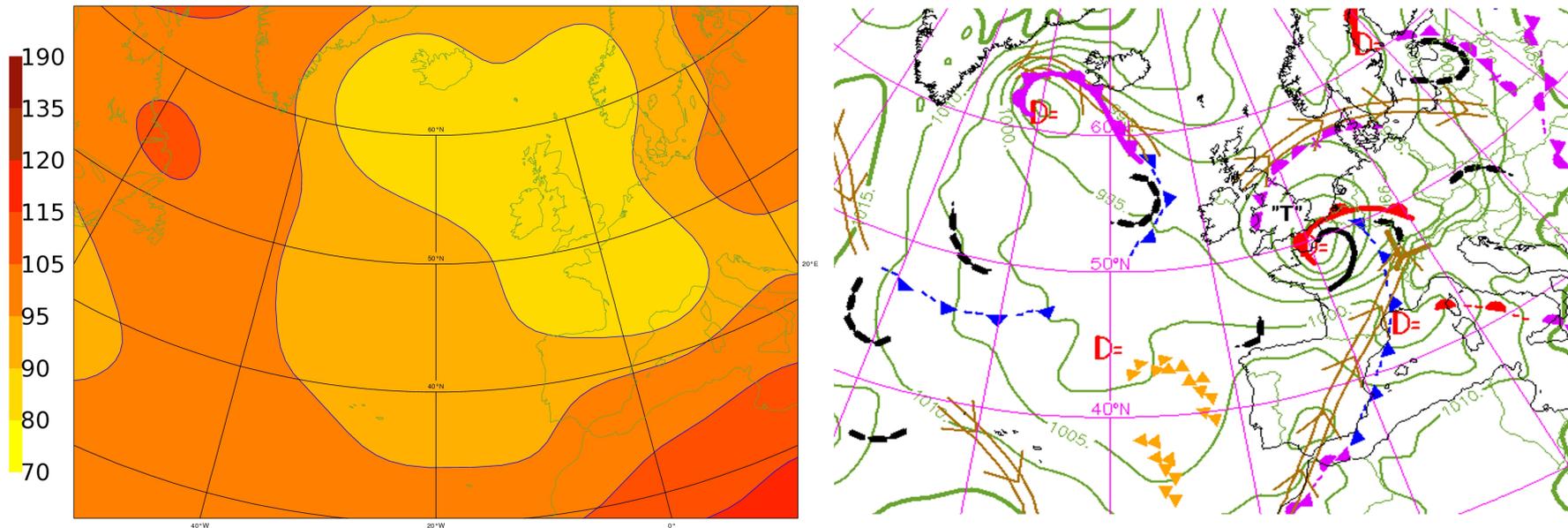
Le bruit est filtré une fois par jour dans une nouvelle méthode

simulation de ε_k^b sur quelques fenêtres (24 simulations par jour)
+
calcul d'une moyenne locale (approche ondelettes)
=
corrélations hétérogènes et dynamiques

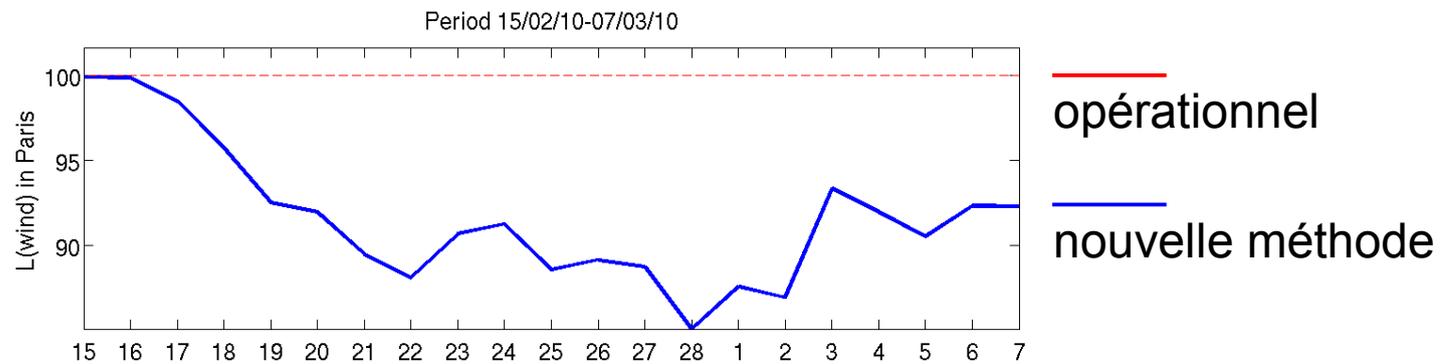


Portées de corrélation du vent estimées par la nouvelle méthode pour le niveau 500hPa et pour les dates 28/02/2010 (à gauche) et 11/12/2011 (à droite)

La nouvelle méthode améliore la représentation des corrélations d'erreur d'ébauche



Les courtes portées de corrélation du vent sont en lien avec des cyclogénèses intenses (cas de la tempête Xynthia du 28/02/2010)

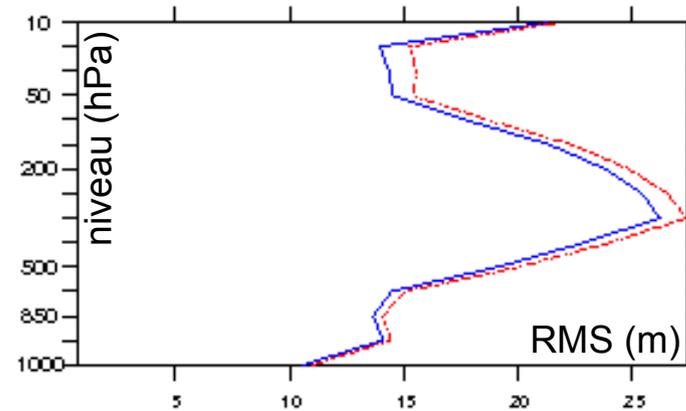
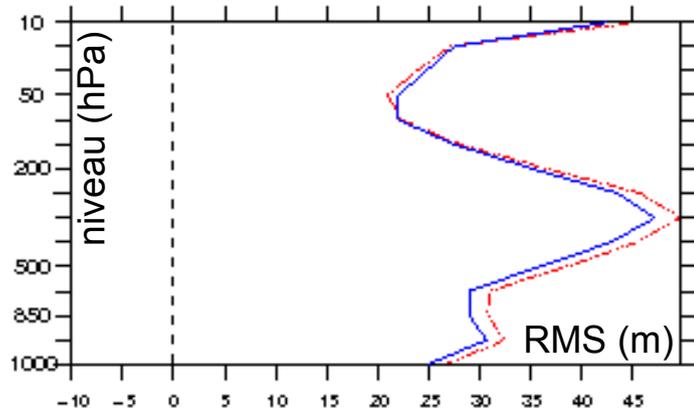


Dynamique des portées de corrélation du vent sur Paris pendant la période 15/02/2010 - 07/03/2010

La nouvelle méthode a un meilleur impact sur les prévisions

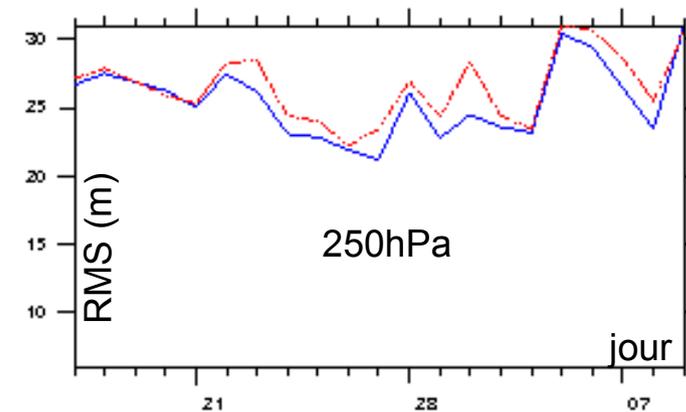
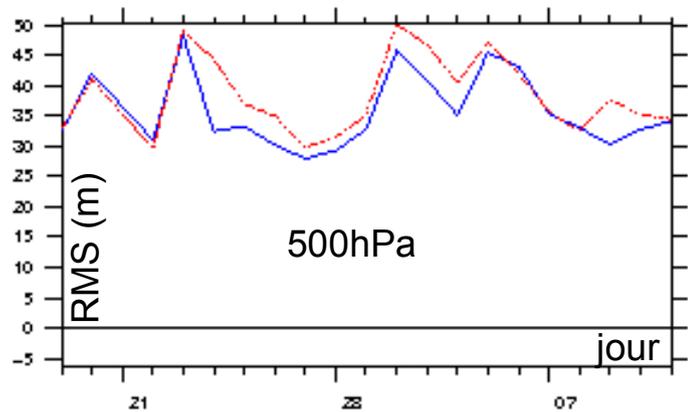
Prévision 96h du géopotential sur EURATL

Prévision 48h du géopotential sur SUD20



— opérationnel

— nouvelle méthode



Erreur de prévision induite par les deux méthodes pendant la période 15/02/2010 - 07/03/2010

Les méthodes stochastiques sont utiles pour l'optimisation de modèles

Elles permettent de prendre en compte une information sous forme probabiliste

- un a priori sur le vecteur d'état ou sur les paramètres
- une observation accompagnée de son erreur

Elles permettent d'optimiser le modèle en réduisant et en quantifiant l'incertitude

- un a posteriori sur le vecteur d'état ou sur les paramètres
- une ébauche accompagnée de son erreur

Ces méthodes doivent cependant être manipulées avec précaution

L'information apportée au système doit être pertinente

- un a priori biaisé donne un a posteriori et une estimation biaisés
- une mauvaise erreur d'observation donne une mauvaise erreur d'ébauche

Le choix et l'adaptation de la méthode dépend de la taille du problème considéré

- le filtre particulaire doit être appliqué à un problème de petite taille et à un modèle peu coûteux
- l'assimilation d'ensemble appliquée à un problème de très grande taille doit être couplée à une autre méthode moins coûteuse