

# Lissage particulière non paramétrique dans les modèles à espace d'état

A. Cuzol, LMBA - Université de Bretagne Sud  
avec P. Ailliot, R. Fablet, V. Monbet, P. Tandéo.

Réunion SEACS, Landéda, 26-27 mai 2015.

# Introduction

Modèle à espace d'état :

- état caché

$$d\mathbf{x}(t) = f(\mathbf{x}(t))dt + \sigma(\mathbf{x}(t))d\mathbf{B}(t)$$

- observé à des instants discrets  $t_1, \dots, t_N$

$$\mathbf{y}(t_k) = g(\mathbf{x}(t_k)) + \gamma_{t_k}$$

# Introduction

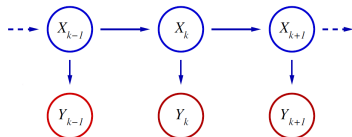
Modèle à espace d'état :

- état caché

$$d\mathbf{x}(t) = f(\mathbf{x}(t))dt + \sigma(\mathbf{x}(t))d\mathbf{B}(t)$$

- observé à des instants discrets  $t_1, \dots, t_N$

$$\mathbf{y}(t_k) = g(\mathbf{x}(t_k)) + \gamma_{t_k}$$



Dépendances du modèle à espace d'état.

# Introduction

Filtrage et lissage : deux méthodes d'assimilation de données.

- **Filtrage** : on cherche à estimer  $p(\mathbf{x}_{t_k} | \mathbf{y}_{t_1:t_k}) \forall t_k$
- **Lissage** : on cherche à estimer  $p(\mathbf{x}_t | \mathbf{y}_{t_1:t_N}) \forall t$

# Introduction

Filtrage et lissage : deux méthodes d'assimilation de données.

- **Filtrage** : on cherche à estimer  $p(\mathbf{x}_{t_k} | \mathbf{y}_{t_1:t_k}) \forall t_k$
- **Lissage** : on cherche à estimer  $p(\mathbf{x}_t | \mathbf{y}_{t_1:t_N}) \forall t$

Motivation principale : applications environnementales

- phénomènes spatio-temporels non linéaires ;
- état de grande dimension ;
- fréquence temporelle des observations faible.

## Filtrage particulière avec analogues

Etapes de la méthode :

- $\hat{p}(\mathbf{x}_{t_{k-1}} | \mathbf{y}_{t_1:t_{k-1}})$  décrit par un ensemble de  $N$  **particules pondérées** :

$$\hat{p}(\mathbf{x}_{t_{k-1}} | \mathbf{y}_{t_1:t_{k-1}}) = \sum_{i=1}^N w_{t_{k-1}}^{(i)} \delta_{\mathbf{x}_{t_{k-1}}^{(i)}}(\mathbf{x}_{t_{k-1}})$$

## Filtrage particulière avec analogues

Etapes de la méthode :

- $\hat{p}(\mathbf{x}_{t_{k-1}} | \mathbf{y}_{t_1:t_{k-1}})$  décrit par un ensemble de  $N$  **particules pondérées** :

$$\hat{p}(\mathbf{x}_{t_{k-1}} | \mathbf{y}_{t_1:t_{k-1}}) = \sum_{i=1}^N w_{t_{k-1}}^{(i)} \delta_{\mathbf{x}_{t_{k-1}}^{(i)}}(\mathbf{x}_{t_{k-1}})$$

- On **prédit** ces particules :
  - Tirage d'un voisin de la particule  $\mathbf{x}_{t_{k-1}}^{(i)}$  parmi les  $k$  plus proches dans le catalogue ;
  - On prédit la particule  $\mathbf{x}_{t_k}^{(i)}$  grâce au futur du voisin.

## Filtrage particulière avec analogues

Etapes de la méthode :

- $\hat{p}(\mathbf{x}_{t_{k-1}} | \mathbf{y}_{t_1:t_{k-1}})$  décrit par un ensemble de  $N$  **particules pondérées** :

$$\hat{p}(\mathbf{x}_{t_{k-1}} | \mathbf{y}_{t_1:t_{k-1}}) = \sum_{i=1}^N w_{t_{k-1}}^{(i)} \delta_{\mathbf{x}_{t_{k-1}}^{(i)}}(\mathbf{x}_{t_{k-1}})$$

- On **prédit** ces particules :
  - Tirage d'un voisin de la particule  $\mathbf{x}_{t_{k-1}}^{(i)}$  parmi les  $k$  plus proches dans le catalogue ;
  - On prédit la particule  $\mathbf{x}_{t_k}^{(i)}$  grâce au futur du voisin.
- On **pondère** les particules grâce à l'observation courante  $\mathbf{y}_{t_k}$  (vraisemblance  $p(\mathbf{y}_{t_k} | \mathbf{x}_{t_k}^{(i)})$ )



## Filtrage particulière avec analogues

Etapes de la méthode :

- $\hat{p}(\mathbf{x}_{t_{k-1}} | \mathbf{y}_{t_1:t_{k-1}})$  décrit par un ensemble de  $N$  **particules pondérées** :

$$\hat{p}(\mathbf{x}_{t_{k-1}} | \mathbf{y}_{t_1:t_{k-1}}) = \sum_{i=1}^N w_{t_{k-1}}^{(i)} \delta_{\mathbf{x}_{t_{k-1}}^{(i)}}(\mathbf{x}_{t_{k-1}})$$

- On **prédit** ces particules :
  - Tirage d'un voisin de la particule  $\mathbf{x}_{t_{k-1}}^{(i)}$  parmi les  $k$  plus proches dans le catalogue ;
  - On prédit la particule  $\mathbf{x}_{t_k}^{(i)}$  grâce au futur du voisin.
- On **pondère** les particules grâce à l'observation courante  $\mathbf{y}_{t_k}$  (vraisemblance  $p(\mathbf{y}_{t_k} | \mathbf{x}_{t_k}^{(i)})$ )
- On **rééchantillonne** les particules (suppression des faibles et duplication des fortes)

## Filtrage particulière avec analogues

Etapes de la méthode :

- $\hat{p}(\mathbf{x}_{t_{k-1}} | \mathbf{y}_{t_1:t_{k-1}})$  décrit par un ensemble de  $N$  **particules pondérées** :

$$\hat{p}(\mathbf{x}_{t_{k-1}} | \mathbf{y}_{t_1:t_{k-1}}) = \sum_{i=1}^N w_{t_{k-1}}^{(i)} \delta_{\mathbf{x}_{t_{k-1}}^{(i)}}(\mathbf{x}_{t_{k-1}})$$

- On **prédit** ces particules :
  - Tirage d'un voisin de la particule  $\mathbf{x}_{t_{k-1}}^{(i)}$  parmi les  $k$  plus proches dans le catalogue ;
  - On prédit la particule  $\mathbf{x}_{t_k}^{(i)}$  grâce au futur du voisin.
- On **pondère** les particules grâce à l'observation courante  $\mathbf{y}_{t_k}$  (vraisemblance  $p(\mathbf{y}_{t_k} | \mathbf{x}_{t_k}^{(i)})$ )
- On **rééchantillonne** les particules (suppression des faibles et duplication des fortes)
- On obtient  $\hat{p}(\mathbf{x}_{t_k} | \mathbf{y}_{t_1:t_k})$

## Lissage particulière

Première idée naturelle : le **lissage généalogique**. En théorie le poids  $w_{t_k}^{(i)}$  est associé à toute la trajectoire d'une particule.

## Lissage particulière

Première idée naturelle : le **lissage généalogique**. En théorie le poids  $w_{t_k}^{(i)}$  est associé à toute la trajectoire d'une particule.

On peut donc écrire à l'instant final  $t_N$  :

$$\hat{p}(\mathbf{x}_{t_1}, \dots, \mathbf{x}_{t_N} | \mathbf{y}_{t_1}, \dots, \mathbf{y}_{t_N}) = \sum_{i=1}^N w_{t_N}^{(i)} \delta_{\mathbf{x}_{t_1}^{(i)}, \dots, \mathbf{x}_{t_N}^{(i)}}(\mathbf{x}_{t_1}, \dots, \mathbf{x}_{t_N})$$

$\Rightarrow$  on peut alors extraire les marginales  $\hat{p}(\mathbf{x}_t | \mathbf{y}_{t_1}, \dots, \mathbf{y}_{t_N})$ .

## Lissage particulière

Première idée naturelle : le **lissage généalogique**. En théorie le poids  $w_{t_k}^{(i)}$  est associé à toute la trajectoire d'une particule.

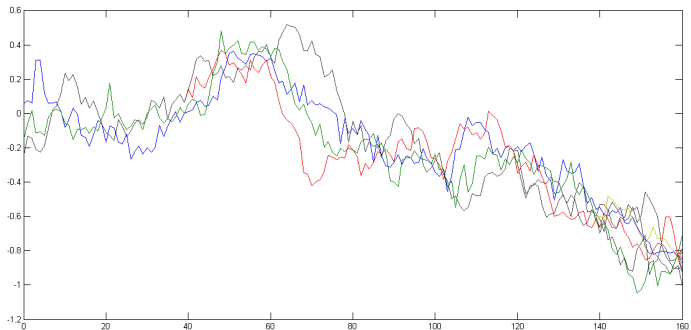
On peut donc écrire à l'instant final  $t_N$  :

$$\hat{p}(\mathbf{x}_{t_1}, \dots, \mathbf{x}_{t_N} | \mathbf{y}_{t_1}, \dots, \mathbf{y}_{t_N}) = \sum_{i=1}^N w_{t_N}^{(i)} \delta_{\mathbf{x}_{t_1}, \dots, \mathbf{x}_{t_N}}^{(i)}(\mathbf{x}_{t_1}, \dots, \mathbf{x}_{t_N})$$

$\Rightarrow$  on peut alors extraire les marginales  $\hat{p}(\mathbf{x}_t | \mathbf{y}_{t_1}, \dots, \mathbf{y}_{t_N})$ .

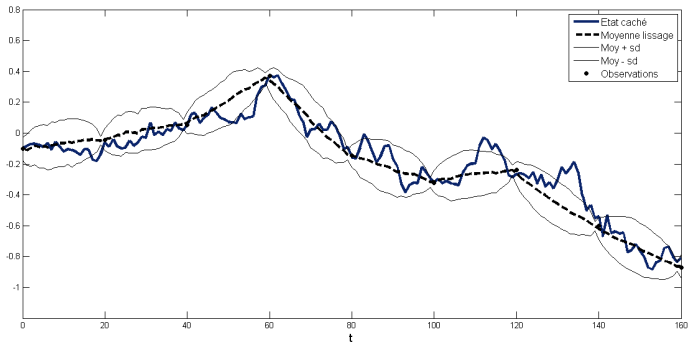
**Mais** : en pratique ceci ne marche pas si  $[t_1, t_N]$  est grand (et  $N$  pas très grand) à cause du rééchantillonnage : les particules ont très rapidement les mêmes ancêtres communs.

Exemple :



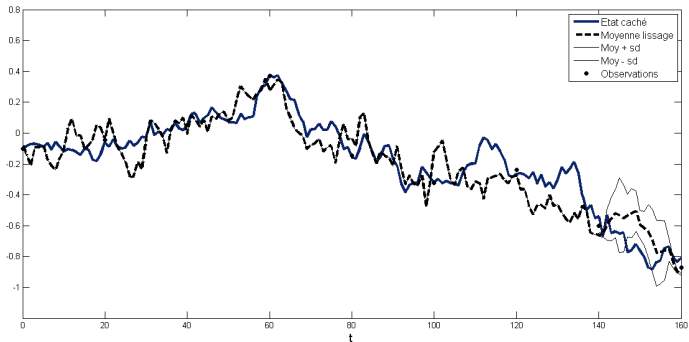
Trajectoires ancestrales des particules

Illustration pour  $N$  très grand :



Lissage généalogique pour  $N = 10000$

... et pour  $N$  petit :



Lissage généalogique pour  $N = 20$



## Lissage particulière FFBS

FFBS = Forward filtering Backward Smoothing

$$\begin{aligned} p(\mathbf{x}_t | \mathbf{y}_{1:T}) &= \int p(\mathbf{x}_t, \mathbf{x}_{t+1} | \mathbf{y}_{1:T}) d\mathbf{x}_{t+1} \\ &= \int p(\mathbf{x}_{t+1} | \mathbf{y}_{1:T}) p(\mathbf{x}_t | \mathbf{y}_{1:t}, \mathbf{x}_{t+1}) d\mathbf{x}_{t+1} \end{aligned}$$

# Lissage particulière FFBS

FFBS = Forward filtering Backward Smoothing

$$\begin{aligned}
 p(\mathbf{x}_t | \mathbf{y}_{1:T}) &= \int p(\mathbf{x}_t, \mathbf{x}_{t+1} | \mathbf{y}_{1:T}) d\mathbf{x}_{t+1} \\
 &= \int p(\mathbf{x}_{t+1} | \mathbf{y}_{1:T}) p(\mathbf{x}_t | \mathbf{y}_{1:t}, \mathbf{x}_{t+1}) d\mathbf{x}_{t+1} \\
 &= \int \underbrace{p(\mathbf{x}_{t+1} | \mathbf{y}_{1:T})}_{\text{Lissage } t+1} \frac{p(\mathbf{x}_{t+1} | \mathbf{x}_t) \overbrace{p(\mathbf{x}_t | \mathbf{y}_{1:t})}^{\text{Filtrage } t}}{p(\mathbf{x}_{t+1} | \mathbf{y}_{1:t})} d\mathbf{x}_{t+1}
 \end{aligned}$$

## Lissage particulière FFBS

FFBS = Forward filtering Backward Smoothing

$$\begin{aligned}
 p(\mathbf{x}_t | \mathbf{y}_{1:T}) &= \int p(\mathbf{x}_t, \mathbf{x}_{t+1} | \mathbf{y}_{1:T}) d\mathbf{x}_{t+1} \\
 &= \int p(\mathbf{x}_{t+1} | \mathbf{y}_{1:T}) p(\mathbf{x}_t | \mathbf{y}_{1:t}, \mathbf{x}_{t+1}) d\mathbf{x}_{t+1} \\
 &= \int \underbrace{p(\mathbf{x}_{t+1} | \mathbf{y}_{1:T})}_{\text{Lissage } t+1} \frac{p(\mathbf{x}_{t+1} | \mathbf{x}_t) \overbrace{p(\mathbf{x}_t | \mathbf{y}_{1:t})}^{\text{Filtrage } t}}{p(\mathbf{x}_{t+1} | \mathbf{y}_{1:t})} d\mathbf{x}_{t+1}
 \end{aligned}$$

où :

$$p(\mathbf{x}_{t+1} | \mathbf{y}_{1:t}) = \int p(\mathbf{x}_{t+1} | \mathbf{x}_t) \underbrace{p(\mathbf{x}_t | \mathbf{y}_{1:t})}_{\text{Filtrage } t} d\mathbf{x}_t$$

## Lissage particulière FFBS

Finalement :

- Partie FF : on stocke les  $\{\hat{p}(\mathbf{x}_t|\mathbf{y}_{1:t})\}_{t=1,\dots,T}$  décrits par les particules  $\mathbf{x}_t^{(i)}$  et leurs poids  $w_t^{(i)}$

# Lissage particulière FFBS

Finalement :

- Partie FF : on stocke les  $\{\hat{p}(\mathbf{x}_t|\mathbf{y}_{1:t})\}_{t=1,\dots,T}$  décrits par les particules  $\mathbf{x}_t^{(i)}$  et leurs poids  $w_t^{(i)}$
- Partie BS : on repondère les trajectoires existantes :

$$\hat{p}(\mathbf{x}_t|\mathbf{y}_{1:T}) = \sum_{i=1}^N w_{t|T}^{(i)} \delta_{\mathbf{x}_t^{(i)}}(\mathbf{x}_t)$$

avec :

$$w_{t|T}^{(i)} = w_t^{(i)} \sum_{j=1}^N w_{t+1|T}^{(j)} \frac{p(\mathbf{x}_{t+1}^{(j)}|\mathbf{x}_t^{(i)})}{\sum_{l=1}^N w_t^{(l)} p(\mathbf{x}_{t+1}^{(j)}|\mathbf{x}_t^{(l)})}$$

(en partant de  $w_{T|T}^{(i)} = \frac{1}{N} \quad \forall i$ ).

## Lissage particulière 2-filter

On combine un filtrage "avant" et un filtre "arrière" :

$$\begin{aligned}
 p(\mathbf{x}_t | \mathbf{y}_{1:T}) &\propto \frac{p(\mathbf{x}_t | \mathbf{y}_{1:t-1}) p(\mathbf{x}_t | \mathbf{y}_{t:T})}{p(\mathbf{x}_t)} \\
 &\quad \text{Filtrage arrière } t \\
 &\quad \underbrace{p(\mathbf{x}_t | \mathbf{y}_{t:T})}_{\text{Filtrage arrière } t} \int \underbrace{p(\mathbf{x}_{t-1} | \mathbf{y}_{1:t-1})}_{\text{Filtrage avant } t-1} p(\mathbf{x}_t | \mathbf{x}_{t-1}) d\mathbf{x}_{t-1} \\
 &\propto \frac{\int \underbrace{p(\mathbf{x}_{t-1} | \mathbf{y}_{1:t-1})}_{\text{Filtrage avant } t-1} p(\mathbf{x}_t | \mathbf{x}_{t-1}) d\mathbf{x}_{t-1}}{p(\mathbf{x}_t)}
 \end{aligned}$$

## Lissage particulière 2-filter

On combine un filtrage "avant" et un filtre "arrière" :

$$\begin{aligned}
 p(\mathbf{x}_t | \mathbf{y}_{1:T}) &\propto \frac{p(\mathbf{x}_t | \mathbf{y}_{1:t-1}) p(\mathbf{x}_t | \mathbf{y}_{t:T})}{p(\mathbf{x}_t)} \\
 &\quad \text{Filtrage arrière } t \\
 &\quad \underbrace{p(\mathbf{x}_t | \mathbf{y}_{t:T})}_{\text{Filtrage arrière } t} \int \underbrace{p(\mathbf{x}_{t-1} | \mathbf{y}_{1:t-1}) p(\mathbf{x}_t | \mathbf{x}_{t-1})}_{\text{Filtrage avant } t-1} d\mathbf{x}_{t-1} \\
 &\propto \frac{\hspace{10em}}{p(\mathbf{x}_t)}
 \end{aligned}$$

Filtrage arrière :

$$\hat{p}(\mathbf{x}_t | \mathbf{y}_{t:T}) = \sum_{i=1}^N \tilde{w}_t^{(i)} \delta_{\tilde{\mathbf{x}}_t^{(i)}}(\tilde{\mathbf{x}}_t)$$

## Lissage particulière 2-filter

On combine un filtrage "avant" et un filtre "arrière" :

$$\begin{aligned}
 p(\mathbf{x}_t | \mathbf{y}_{1:T}) &\propto \frac{p(\mathbf{x}_t | \mathbf{y}_{1:t-1}) p(\mathbf{x}_t | \mathbf{y}_{t:T})}{p(\mathbf{x}_t)} \\
 &\quad \text{Filtrage arrière } t \\
 &\quad \underbrace{p(\mathbf{x}_t | \mathbf{y}_{t:T})}_{\text{Filtrage avant } t-1} \int \underbrace{p(\mathbf{x}_{t-1} | \mathbf{y}_{1:t-1})}_{\text{Filtrage avant } t-1} p(\mathbf{x}_t | \mathbf{x}_{t-1}) d\mathbf{x}_{t-1} \\
 &\propto \frac{\hspace{10em}}{p(\mathbf{x}_t)}
 \end{aligned}$$

Filtrage arrière :

$$\hat{p}(\mathbf{x}_t | \mathbf{y}_{t:T}) = \sum_{i=1}^N \tilde{w}_t^{(i)} \delta_{\tilde{\mathbf{x}}_t^{(i)}}(\mathbf{x}_t)$$

On obtient alors :

$$\hat{p}(\mathbf{x}_t | \mathbf{y}_{1:T}) = \sum_{i=1}^N \frac{\tilde{w}_t^{(i)} \sum_{j=1}^N w_{t-1}^{(j)} p(\tilde{\mathbf{x}}_t^{(i)} | \mathbf{x}_{t-1}^{(j)})}{p(\tilde{\mathbf{x}}_t^{(i)})} \delta_{\tilde{\mathbf{x}}_t^{(i)}}(\mathbf{x}_t)$$



## Estimation non paramétrique des probas de transition

Pour les deux lissages, on doit estimer des termes du type  $p(\mathbf{x}_{t+1}^{(j)} | \mathbf{x}_t^{(i)})$ .

## Estimation non paramétrique des probas de transition

Pour les deux lissages, on doit estimer des termes du type  $p(\mathbf{x}_{t+1}^{(j)} | \mathbf{x}_t^{(i)})$ .

⇒ on les estime de façon non paramétrique (estimateur kNN) à partir du catalogue.

## Estimation non paramétrique des probas de transition

Pour les deux lissages, on doit estimer des termes du type  $p(\mathbf{x}_{t+1}^{(j)} | \mathbf{x}_t^{(i)})$ .

⇒ on les estime de façon non paramétrique (estimateur kNN) à partir du catalogue.

Principe de l'estimation kNN :

$$\hat{p}(\mathbf{x}) = \frac{k}{nc^h d(\mathbf{x}, \mathbf{x}^{[k]})}$$

où :

- $n$  est le nombre total de points ;
- $c^h$  est le volume de la sphère en dimension  $h$  ;
- $d(\mathbf{x}, \mathbf{x}^{[k]})$  est la distance de  $\mathbf{x}$  au  $k^{eme}$  voisin ;
- $k$  est le paramètre à fixer.

Finalement :

$$\hat{p}(\mathbf{x}_{t+1}^{(j)} | \mathbf{x}_t^{(i)}) \propto \frac{d(\mathbf{x}_t^{(i)}, \mathbf{x}_t^{(i),[k]})}{d((\mathbf{x}_t^{(i)}, \mathbf{x}_{t+1}^{(j)}), (\mathbf{x}_t^{(i),[k]}, \mathbf{x}_{t+1}^{(j),[k]}))}$$

où :

- $\mathbf{x}_t^{(i),[k]}$  est le  $k^{eme}$  plus proche voisin de  $\mathbf{x}_t^{(i)}$  dans le catalogue ;
- $(\mathbf{x}_t^{(i),[k]}, \mathbf{x}_{t+1}^{(j),[k]})$  est le  $k^{eme}$  plus proche voisin de  $(\mathbf{x}_t^{(i)}, \mathbf{x}_{t+1}^{(j)})$  dans le catalogue.

## Résultats du lissage 2-filter

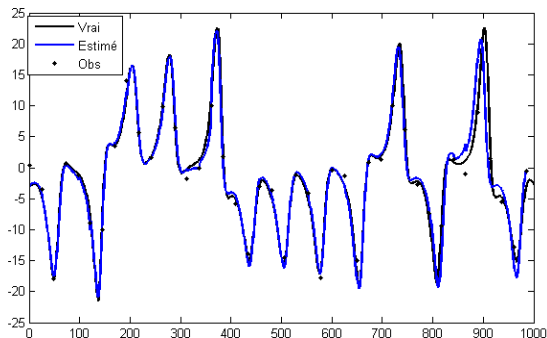
Exemple 1 :

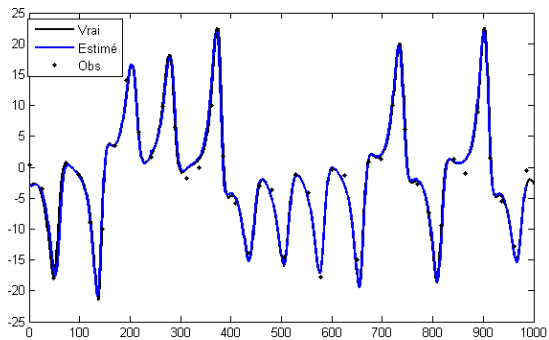
- $dt = 0.01$  et  $dt_{obs} = 24$ .
- Taille du catalogue :  $10^4$
- Première composante observée
- $N = 20$

# Résultats du lissage 2-filter

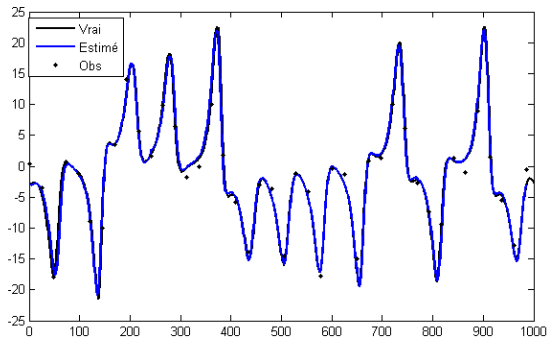
Exemple 1 :

- $dt = 0.01$  et  $dt_{obs}=24$ .
- Taille du catalogue :  $10^4$
- Première composante observée
- $N = 20$



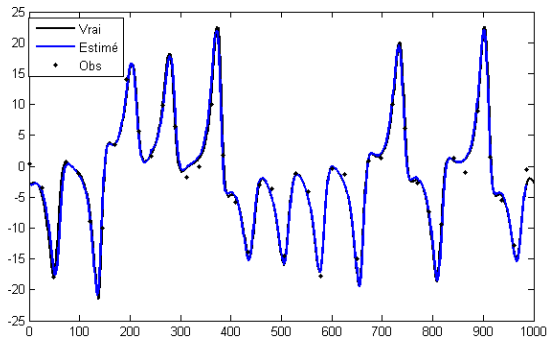


Filtrage arrière  
(moy. deuxième composante)



Lissage 2-filter  
(moy. deuxième composante)

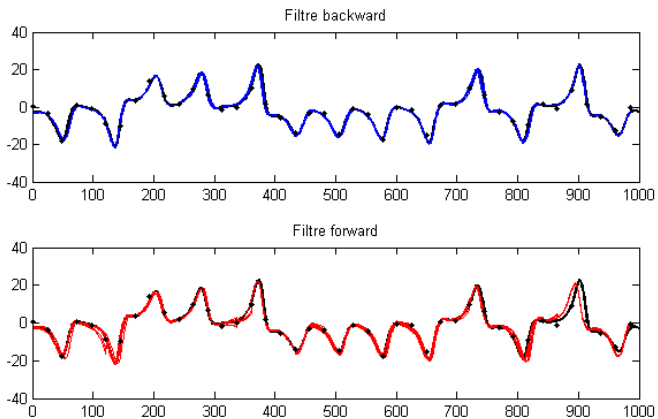




Lissage 2-filter  
(moy. deuxième composante)

Erreurs pour les 3 composantes :

- Filtrage avant : 1,226 1,965 2,467
- Filtrage arrière : 0,623 0,875 1,19
- Lissage 2-filter : 0,622 0,875 1,189



Toutes les trajectoires avant et arrière  
(deuxième composante)

## Résultats du lissage 2-filter

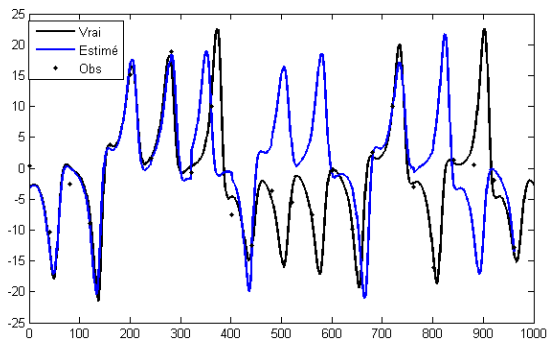
### Exemple 2

- $dt = 0.01$  et  $dt_{obs}=40$
- Taille du catalogue :  $10^4$
- Première composante observée
- $N = 20$

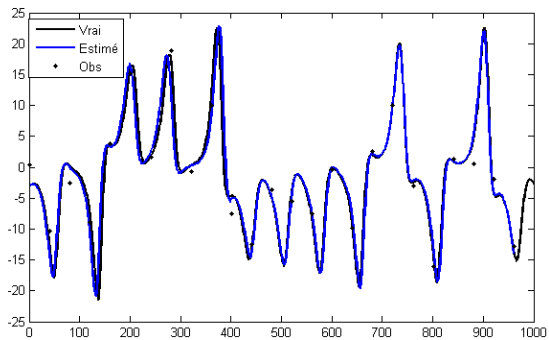
# Résultats du lissage 2-filter

## Exemple 2

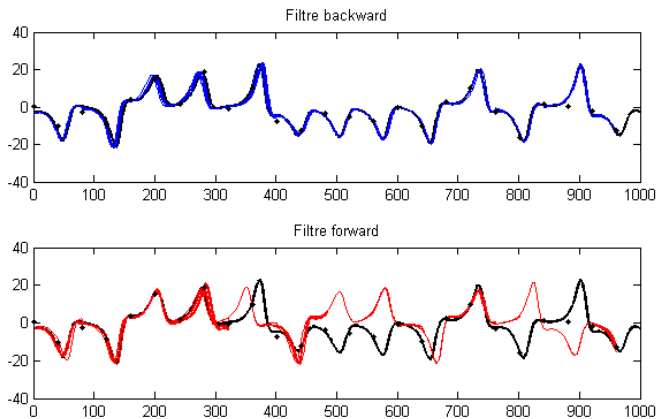
- $dt = 0.01$  et  $dt_{obs}=40$
- Taille du catalogue :  $10^4$
- Première composante observée
- $N = 20$



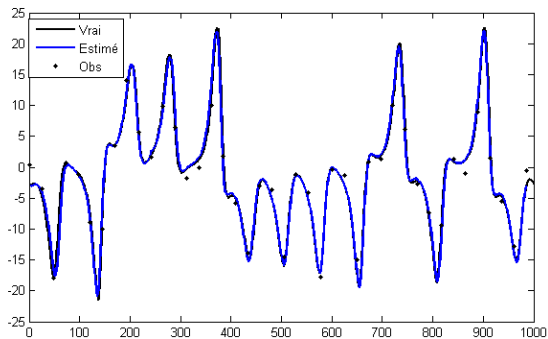
Filtrage avant  
(moy. deuxième composante)



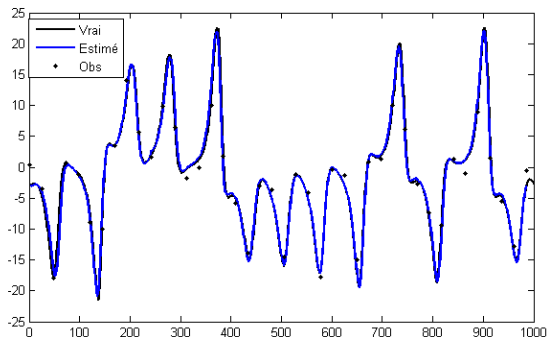
Filtrage arrière  
(moy. deuxième composante)



Toutes les trajectoires avant et arrière  
(deuxième composante)



Lissage 2-filter  
(moy. deuxième composante)



Lissage 2-filter  
(moy. deuxième composante)

Erreurs pour les 3 composantes :

- Filtrage avant : 9,064 10,445 7,432
- Filtrage arrière : 1,249 1,919 2,367
- Lissage 2-filter : 1,246 1,916 2,363



## Résultats du lissage 2-filter

Exemple 3 :

- $dt = 0.01$  et  $dt_{obs}=24$
- Taille du catalogue :  $10^4$
- Première composante observée
- $N = 100$

## Résultats du lissage 2-filter

Exemple 3 :

- $dt = 0.01$  et  $dt_{obs}=24$
- Taille du catalogue :  $10^4$
- Première composante observée
- $N = 100$

Erreurs pour les 3 composantes :

- Filtrage avant : 0,952 1,450 1,786
- Filtrage arrière : 0,804 1,143 1,516
- Lissage 2-filter : 0,799 1,135 1,506

## Résultats du lissage FFBS

Même cadre que l'exemple 1 précédent :

- $dt = 0.01$  et  $dt_{obs}=24$
- Taille du catalogue :  $10^4$
- Première composante observée
- $N = 20$

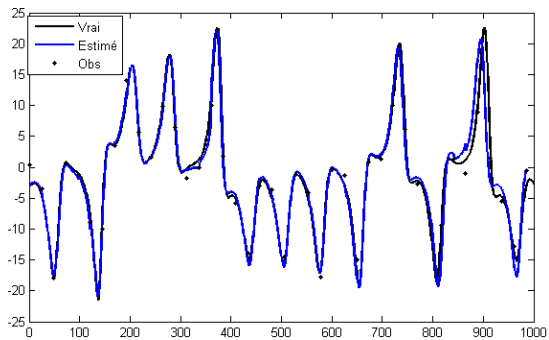
## Résultats du lissage FFBS

Même cadre que l'exemple 1 précédent :

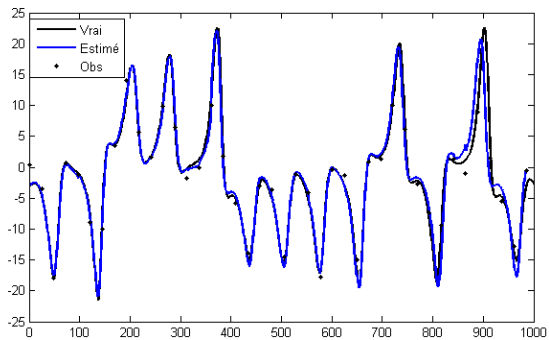
- $dt = 0.01$  et  $dt_{obs}=24$
- Taille du catalogue :  $10^4$
- Première composante observée
- $N = 20$

Erreurs pour les 3 composantes :

- Filtrage : 1,226 1,965 2,467
- Lissage FFBS : 1,198 1,920 2,438



Filtrage  
(moy. deuxième composante)



Lissage FFBS  
(moy. deuxième composante)