Filtre de Kalman pour l'assimilation de données

Pierre Ailliot

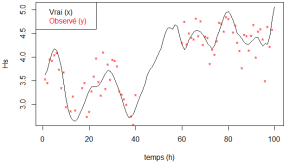
Laboratoire de mathématiques de Bretagne Atlantique
Université de Bretagne Occidentale

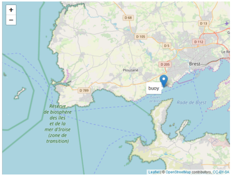
26 janvier 2023

Plan

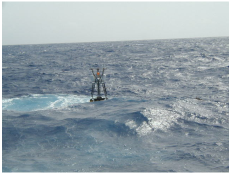
Motivations

Hauteur significative des vagues (Hs) à Ste Anne du Porzic (France)



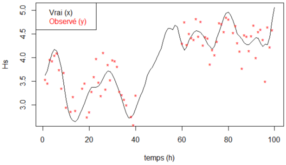


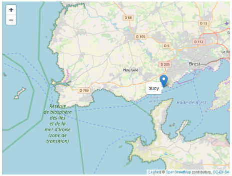
- Les bouées mesurent la hauteur significative des vagues (Hs) avec une certaine erreur
  - X_t "vraie" valeur de Hs, Y_t valeur mesurée à l'instant t
  - **Hypothèse 1 (équation d'observation) :** $Y_t = X_t + \epsilon_t$ avec ϵ_t l'erreur de mesure
- Modélisation de l'évolution temporelle de Hs ?
  - **Hypothèse 2 (modèle dynamique) :** $X_t = MX_{t-1} + \eta_t$



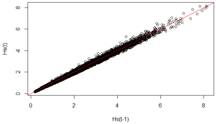
Motivations

Hauteur significative des vagues (Hs) à Ste Anne du Porzic (France)



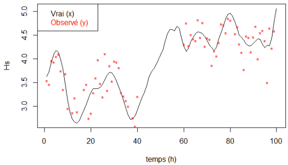


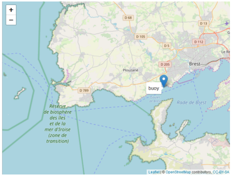
- Les bouées mesurent la hauteur significative des vagues (Hs) avec une certaine erreur
  - X_t "vraie" valeur de Hs, Y_t valeur mesurée à l'instant t
  - **Hypothèse 1 (équation d'observation) :** $Y_t = X_t + \epsilon_t$ avec ϵ_t l'erreur de mesure
- Modélisation de l'évolution temporelle de Hs ?
  - **Hypothèse 2 (modèle dynamique) :** $X_t = MX_{t-1} + \eta_t$



Motivations

Hauteur significative des vagues (Hs) à Ste Anne du Porzic (France)





- Les bouées mesurent la hauteur significative des vagues (Hs) avec une certaine erreur
  - X_t "vraie" valeur de Hs, Y_t valeur mesurée à l'instant t
  - **Hypothèse 1 (équation d'observation) :** $Y_t = X_t + \epsilon_t$ avec ϵ_t l'erreur de mesure
- Modélisation de l'évolution temporelle de Hs ?
  - **Hypothèse 2 (modèle dynamique) :** $X_t = MX_{t-1} + \eta_t$
- **Objectifs ?** "Reconstruire" le vrai Hs à partir des observations bruitées et parfois manquantes

Plan

Modèles espace état général, vision "système dynamique"

- Modèle avec deux composantes X_t et Y_t
- X_t décrit le vrai état du système à l'instant t.
  - On suppose que cet état est "'caché"' ou "'latent"' ou "'non-observable"'.
  - Evolution dynamique décrite par **l'équation d'état**

$$X_t = \mathcal{M}_t(X_{t-1}) + \eta_t$$

  - $\mathcal{M}_t$ décrit la partie déterministe de l'évolution temporelle de l'état (en assimilation de données, typiquement un modèle numérique basé sur des EDP)
  - η_t un suite de variables aléatoire (**v.a.**) qui décrit les erreurs (approximations physiques et numériques, échelles non-résolues, etc)
  - On suppose que η_t est indépendant de $X_{0:t-1} := (X_0, \ldots, X_{t-1})$ et $Y_{0:t-1}$
- Y_t décrit les observations disponibles à l'instant t.
  - On suppose que les observations sont liées à l'état caché par l'**équation d'observation**

$$Y_t = \mathcal{H}_t(X_t) + \epsilon_t$$

  - $\mathcal{H}_t$ décrit le lien entre l'état et les observations (typiquement, $\mathcal{H}_t = I$ ou projection sur sous-espace)
  - ϵ_t un suite de v.a. indépendantes qui décrit les erreurs d'observations (erreurs de capteur, différences de support entre le modèle et les observations, ...)
  - On suppose que ϵ_t est indépendant de $X_{0:t}$ et $Y_{0:t-1}$.

Modèles espace état général, vision "probabiliste"

- **Notation.** $p(x_t|x_{0:t-1}, y_{1:t-1})$ est la densité de la loi conditionnelle $P(X_t|X_0 = x_0, ..., X_{t-1} = x_{t-1}, Y_1 = x_0, ..., Y_{t-1} = y_{t-1})$ (idem pour autres lois conditionnelles).
- Le processus $\{X_t, Y_t\}$ vérifie les propriétés d'indépendance conditionnelle suivantes
  - $p(x_t|x_{0:t-1}, y_{1:t-1}) = p(x_t|x_{t-1})$
    - Correspond à l'équation $X_t = \mathcal{M}_t(X_{t-1}) + \eta_t$ avec η_t indépendant de $X_{0:t-1}$ et $Y_{0:t-1}$.
    - L'état du système à l'instant t dépend uniquement de l'état du système à l'instant $t-1$
    - Propriété Markovienne avec oubli de l'état du système et des observations aux instants précédents
  - $p(y_t|x_{1:t}, y_{1:t-1}) = p(y_t|x_t)$
    - Correspond à l'équation $Y_t = \mathcal{H}_t(X_t) + \epsilon_t$ avec $\epsilon_t$indépendant de $X_{0:t}$ et $Y_{0:t-1}$.
    - L'observation à l'instant t dépend uniquement de l'état du système à l'instant t
- La structure probabiliste du modèle espace-état peut être résumée par le diagramme d'indépendance conditionnelle (DAG) suivant

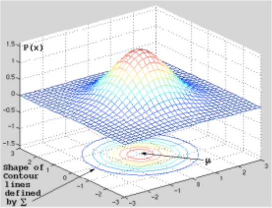
$$\begin{array}{lccccccccc}
\text{Etat} & \cdots & \rightarrow & X_{t-1} & \rightarrow & X_t & \rightarrow & X_{t+1} & \rightarrow & \cdots \\
 & & & \downarrow & & \downarrow & & \downarrow & & \\
\text{Observation} & \cdots & & Y_{t-1} & & Y_t & & Y_{t+1} & & \cdots
\end{array}$$

- La loi normale joue un rôle important en assimilation de données

Rappels sur la loi normale multivariée

- Définition : $X = (X_1, ..., X_d)' \sim \mathcal{N}(\mu, \Sigma)$ si sa densité est donnée par

$$\phi(x;\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{det(\Sigma)}} exp\left(-\frac{1}{2}(X-\mu)'\Sigma^{-1}(X-\mu)\right)$$



- $\mu = E[X] = (E[X_1], ..., E[X_d])'$ et $\Sigma = var(X) = E[XX'] - E[X]E[X]'$
- Fonction caractéristique : $E[e^{it'X}] = \int_{\mathbb{R}^d} e^{it'x}\phi(x;\mu,\Sigma)dx = e^{it'\mu - \frac{1}{2}t'\Sigma t}$
- N.B. Loi normale = loi gaussienne = loi de Gauss
- Loi la plus utilisée en statistique, machine learning, traitement du signal...Pourquoi ?
  - Théorème central limite...
  - et possibilités de faire de nombreux calculs de manière analytique !

Rappels sur la loi normale multivariée
Stabilité par transformation linéaire

Proposition 1

On suppose que $Z_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$ et soit M une matrice à coefficients réels. Alors $Z_1 = MZ_0 \sim \mathcal{N}(\mu_1, \Sigma_1)$ avec $\mu_1 = M\mu_0$ et $\Sigma_1 = M\Sigma_0 M'$.

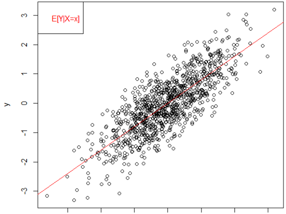
Preuve Utiliser les fonctions caractéristiques.

Rappels sur la loi normale multivariée
Stabilité par conditionnement

Proposition 2

Soit $(X, Y) \sim \mathcal{N}\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{pmatrix}\right)$. Si Σ_Y est inversible alors la loi conditionnelle $\mathcal{L}_{X|Y=y}$ est une loi normale de moyenne $\mu_X + \Sigma_{XY}\Sigma_Y^{-1}(y - \mu_Y)$ et de matrice de covariance $\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}$

Preuve $p(x|y) = \frac{p(x,y)}{p(y)}$ avec $p(x, y)$ densité de (X, Y), $p(x)$ densité de X, $p(x|y)$ densité de $\mathcal{L}_{X|Y=y}$



Rappels sur la loi normale multivariée
Stabilité par conditionnement

Proposition 2

Soit $(X, Y) \sim \mathcal{N}\left(\left(\begin{array}{c}\mu_X \\ \mu_Y\end{array}\right), \left(\begin{array}{cc}\Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y\end{array}\right)\right)$. Si Σ_Y est inversible alors la loi conditionnelle $\mathcal{L}_{X|Y=y}$ est une loi normale de moyenne $\mu_X + \Sigma_{XY}\Sigma_Y^{-1}(y - \mu_Y)$ et de matrice de covariance $\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}$

- **Conséquence** : $E[X|Y] = \mu_X + \Sigma_{XY}\Sigma_Y^{-1}(Y - \mu_Y)$ est le meilleur prédicteur au sens des moindres carrés de X à partir de Y (**estimateur des moindres carrés**)
- **Remarque** : dans le cas général (non-gaussien), $\mu_X + \Sigma_{XY}\Sigma_Y^{-1}(Y - \mu_Y)$ reste le meilleur prédicteur **linéaire sans biais** au sens des moindres carrés de X à partir de Y (**Best Linear Unbiased Prediction**). Ceci peut justifier l'utilisation de ce prédicteur dans le cas non-gaussien

Modèle espace-état linéaire gaussien

- Le filtre de Kalman est optimal pour le **modèle linéaire gaussien** : modèle espace-état dans lequel les opérateurs sont linéaires et les erreurs sont gaussiennes.
- $X_t = MX_{t-1} + \eta_t$ avec $\eta_t \sim \mathcal{N}(0, Q)$ et donc

$$\mathcal{L}(X_t | X_{t-1} = x_{t-1}) \sim \mathcal{N}(Mx_{t-1}, Q)$$

  - M modélise la dynamique et Q l'incertitude associée.
- $Y_t = HX_t + \epsilon_t$ avec $\epsilon_t \sim \mathcal{N}(0, R)$ et donc

$$\mathcal{L}(Y_t | X_t = x_t) \sim \mathcal{N}(Hx_t, R).$$

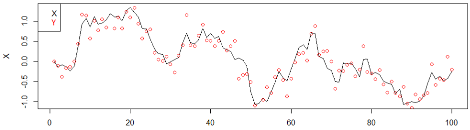
  - H modélise le mécanisme d'observations et R l'incertitude associée.
- **Remarques**
  - M , H , Q et R peuvent varier au cours du temps (supposés constants dans la suite du cours pour simplifier les notations)
  - On suppose dans la suite que la loi initiale est telle que $X_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$. On en déduit que$(X_0, X_1, Y_1, ..., X_T, Y_T)$ est un vecteur gaussien

Modèle espace-état linéaire gaussien

- Le filtre de Kalman est optimal pour le **modèle linéaire gaussien** : modèle espace-état dans lequel les opérateurs sont linéaires et les erreurs sont gaussiennes.
- $X_t = MX_{t-1} + \eta_t$ avec $\eta_t \sim \mathcal{N}(0, Q)$ et donc
$$\mathcal{L}(X_t | X_{t-1} = x_{t-1}) \sim \mathcal{N}(Mx_{t-1}, Q)$$
- $Y_t = HX_t + \epsilon_t$ avec $\epsilon_t \sim \mathcal{N}(0, R)$ et donc
$$\mathcal{L}(Y_t | X_t = x_t) \sim \mathcal{N}(Hx_t, R).$$

Algorithme pour simuler un modèle espace-état linéaire gaussien

1. Simuler $X_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$, $(\eta_1, ..., \eta_T) \sim_{iid} \mathcal{N}(0, Q)$ et $(\epsilon_1, ..., \epsilon_T) \sim_{iid} \mathcal{N}(0, R)$
2. Pour $t \in \{1, ..., T\}$
   1. Simuler $X_t = MX_{t-1} + \eta_t$
   2. Simuler $Y_t = HX_t + \epsilon_t$



Prédiction/filtrage/lissage dans les modèles espace-état

- **Objectif** : estimer l'état latent X_t à l'instant t connaissant une séquence d'observations $\{y_1, \cdots, y_T\}$
- Information contenue dans la loi conditionnelle $p(x_t|y_{1:T})$
  - En particulier $E[X_t|y_{1:T}] = \int x_t p(x_t|y_{1:T})dx_t$ donne la meilleur prévision de X_t au sens des moindres carrés. On peut aussi en déduire une incertitude sur la prévision
- On distingue trois cas selon la valeur de T
  - **Prédiction (forecast)**
    - On cherche à estimer l'état du système à l'instant t connaissant les observations passées
    - La quantité d'intérêt est la distribution de prédiction $p(x_t|y_1, \cdots, y_{t-1})$
  - **Filtrage (nowcast, analyse)**
    - On cherche à estimer l'état du système à l'instant t connaissant les observations passées et présentes
    - La quantité d'intérêt est la distribution de filtrage $p(x_t|y_1, \cdots, y_t)$
  - **Lissage (hindcast ou réanalyse)**
    - On cherche à estimer l'état du système à l'instant t connaissant les observations passées et présentes et futures
    - La quantité d'intérêt est la distribution de lissage $p(x_t|y_1, \cdots, y_T)$ avec $T > t$
    - Lissage pas discuté dans la suite du cours

Récursion forward-backward

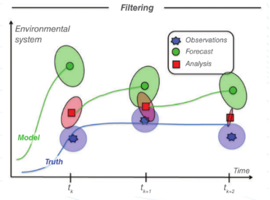
- Dans le modèle espace-état général, les probabilités de prédiction et de filtrage vérifient les récursions "forward"
  - **Prédiction :**

$$p(x_t|y_1,\cdots,y_{t-1}) = \int p(x_{t-1}|y_1,\cdots,y_{t-1})p(x_t|x_{t-1})dx_{t-1}$$

    - Interprétation : l'état du système est prédit en appliquant le modèle dynamique $p(x_t|x_{t-1})$ sur la loi de filtrage $p(x_{t-1}|y_1,\cdots,y_{t-1})$
  - **Correction :** l'état du système est corrigé en utilisant l'observation y_t à l'instant t

$$\begin{aligned} p(x_t|y_1,\cdots,y_t) &= \frac{p(y_t|x_t)p(x_t|y_1,\cdots,y_{t-1})}{\int p(y_t|x)p(x|y_1,\cdots,y_{t-1})dx} \\ &\propto p(y_t|x_t)p(x_t|y_1,\cdots,y_{t-1}) \end{aligned}$$

    - Interprétation : correction bayésienne de la loi a priori $\mathcal{L}(X_t|y_1,\cdots,y_{t-1})$



- Des expressions exactes sont connues pour ces intégrales seulement dans les deux

Récursion forward-backward

- Dans le modèle espace-état général, les probabilités de prédiction et de filtrage vérifient les récursions "forward"
  - **Prédiction :**

$$p(x_t|y_1, \cdots, y_{t-1}) = \int p(x_{t-1}|y_1, \cdots, y_{t-1}) p(x_t|x_{t-1}) dx_{t-1}$$

    - Interprétation : l'état du système est prédit en appliquant le modèle dynamique $p(x_t|x_{t-1})$ sur la loi de filtrage $p(x_{t-1}|y_1, \cdots, y_{t-1})$
  - **Correction :** l'état du système est corrigé en utilisant l'observation y_t à l'instant t

$$\begin{aligned} p(x_t|y_1, \cdots, y_t) &= \frac{p(y_t|x_t)p(x_t|y_1, \cdots, y_{t-1})}{\int p(y_t|x)p(x|y_1, \cdots, y_{t-1})dx} \\ &\propto p(y_t|x_t)p(x_t|y_1, \cdots, y_{t-1}) \end{aligned}$$

    - Interprétation : correction bayésienne de la loi a priori $\mathcal{L}(X_t|y_1, \cdots, y_{t-1})$
- Des expressions exactes sont connues pour ces intégrales seulement dans les deux cas suivants :
  - X_t à valeurs dans un espace fini (Chaîne de Markov cachée) : algorithme Forward-Backward.
  - Modèle linéaire Gaussien : **filtre de Kalman**
- Alternatives dans les autres cas : approximer les intégrales par des simulations -> méthodes de Monte-Carlo séquentielles : EnKF, filtre particulaire,...

Plan

Filtre de Kalman

Résultat principal du cours. On se place sous les hypothèses du modèle linéaire gaussien.

- $X_t = MX_{t-1} + \eta_t$ avec $\eta_t \sim_{iid} \mathcal{N}(0, Q)$
- $Y_t = HX_t + \epsilon_t$ avec $\epsilon_t \sim_{iid} \mathcal{N}(0, R)$

Alors les probabilités de prédiction et de filtrage suivent des lois normales

- $\mathcal{L}(X_t|y_1, ..., y_{t-1}) \sim \mathcal{N}(x_t^f, P_t^f)$
- $\mathcal{L}(X_t|y_1, ..., y_t) \sim \mathcal{N}(x_t^a, P_t^a)$

avec x^f, x^a, P^f, P^a qui sont donnés par les récursions de Kalman (1966) ci-dessous

- **Prédiction**
  $x_t^f = Mx_{t-1}^a$ (espérance de l'état prédit)
  $P_t^f = MP_{t-1}^a M' + Q$ (variance de l'état prédit)
- **Correction (ou mise à jour)**
  $K_t = P_t^f H' (HP_t^f H' + R)^{-1}$ (gain de Kalman)
  $x_t^a = x_t^f + K_t(y_t - Hx_t^f)$ (espérance de l'état filtré)
  $P_t^a = (I - K_t H)P_t^f$ (variance de l'état filtré)

Schéma de la preuve

Raisonnement par récurrence. Supposons que $\mathcal{L}(X_{t-1}|y_1, ..., y_{t-1}) \sim \mathcal{N}(x^a_{t-1}, P^a_{t-1})$

1. On va déduire de l'expression

$$p(x_t|y_1, \cdots, y_{t-1}) = \int p(x_{t-1}|y_1, \cdots, y_{t-1})p(x_t|x_{t-1})dx_{t-1}.$$

   que $\mathcal{L}(X_t|y_1, ..., y_{t-1}) \sim \mathcal{N}(x^f_t, P^f_t)$.

2. On va déduire de l'expression

$$p(x_t|y_1, \cdots, y_t) = \frac{p(y_t|x_t)p(x_t|y_1, \cdots, y_{t-1})}{\int p(y_t|x)p(x|y_1, \cdots, y_{t-1})dx}$$

   que $\mathcal{L}(X_t|y_1, ..., y_t) \sim \mathcal{N}(x^a_t, P^a_t)$.

Preuve du point 1.

- Supposons que $Z_{t-1} \sim \mathcal{N}(\mu_{t-1}, \Sigma_{t-1})$ et $Z_t = MZ_{t-1} + \eta$ avec $\eta \sim \mathcal{N}(0, Q)$ indépendant de Z_{t-1}. En utilisant la proposition 1, on obtient que $Z_t \sim \mathcal{N}(\mu_t, \Sigma_t)$ avec $\mu_t = M\mu_{t-1}$ et $\Sigma_t = Q + M\Sigma_{t-1}M'$.
- On peut réécrire ce résultat avec les probabilités conditionnelles : si $p(z_{t-1}) = \phi(z_{t-1}; \mu_{t-1}, \Sigma_{t-1})$ et $p(z_t|z_{t-1}) = \phi(z_t; Mz_{t-1}, Q)$ alors

$$p(z_t) = \int p(z_t|z_{t-1})p(z_{t-1})dz_{t-1} = \phi(z_t; \mu_t, \Sigma_t)$$

- L'étape de prédiction du filtre de Kalman se déduit alors de l'expression

$$p(x_t|y_1, \cdots, y_{t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|y_1, \cdots, y_{t-1})dx_{t-1}$$

  avec
  - $p(x_{t-1}|y_1, \cdots, y_{t-1}) = \phi(x_{t-1}; x^a_{t-1}, P^a_{t-1})$ d'après l'hypothèse de récurrence
  - $p(x_t|x_{t-1}) = \phi(x_t; Mx_{t-1}, Q)$ d'après la définition du modèle linéaire gaussien

Preuve du point 2.

Proposition 2 (rappel)

Soit $(X, Y) \sim \mathcal{N}\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{pmatrix}\right)$. Si Σ_Y est inversible alors la loi conditionnelle $\mathcal{L}_{X|Y=y}$ est une loi normale de moyenne $\mu_X + \Sigma_{XY}\Sigma_Y^{-1}(y - \mu_Y)$ et de matrice de covariance $\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}$.

Conséquences

- Si $X \sim \mathcal{N}(\mu_X, \Sigma_X)$ et $Y = HX + \epsilon$ avec $\epsilon \sim \mathcal{N}(0, R)$ indépendant de X alors $(X, Y) \sim \mathcal{N}\left(\begin{pmatrix} \mu_X \\ \mu_X \end{pmatrix}, \begin{pmatrix} \Sigma_X & \Sigma_X H' \\ H\Sigma_X & H\Sigma_X H' + R \end{pmatrix}\right)$. Donc

$$\mathcal{L}_{X|Y=y} \sim \mathcal{N}(\mu_{X|y}, \Sigma_{X|y})$$

  avec $\mu_{X|y} = \mu_x + \Sigma_X H'(H\Sigma_X H' + R)^{-1}(y - \mu_x)$ et $\Sigma_{X|y} = \Sigma_X - \Sigma_X H'(H\Sigma_X H' + R)^{-1} H\Sigma_X$.
- Avec les probabilités conditionnelles : si $p(x) = \phi(x; \mu_X, \Sigma_X)$ et $p(y|x) = \phi(y; Hx, R)$ alors

$$p(x|y) \propto \int p(y|x)p(x)dx = \phi(x; \mu_{X|y}, \Sigma_{X|y}).$$

Preuve du point 2.

Proposition 2 (rappel)

Soit $(X, Y) \sim \mathcal{N}\left(\left(\begin{array}{c}\mu_X\\ \mu_Y\end{array}\right), \left(\begin{array}{cc}\Sigma_X & \Sigma_{XY}\\ \Sigma_{YX} & \Sigma_Y\end{array}\right)\right)$. Si Σ_Y est inversible alors la loi conditionnelle $\mathcal{L}_{X|Y=y}$ est une loi normale de moyenne $\mu_X + \Sigma_{XY}\Sigma_Y^{-1}(y - \mu_Y)$ et de matrice de covariance $\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}$.

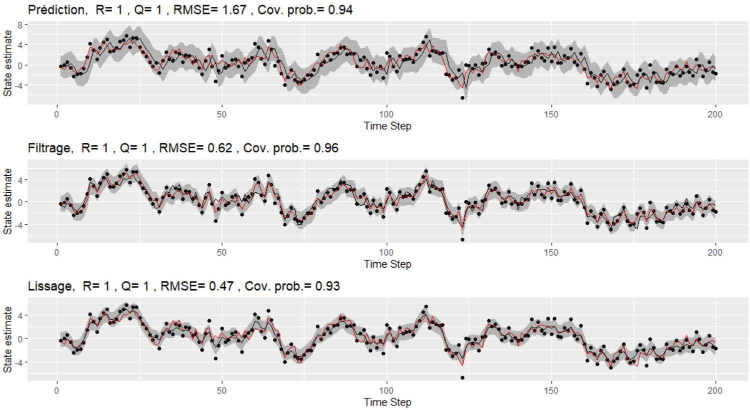
Conséquences

- Avec les probabilités conditionnelles : si $p(x) = \phi(x; \mu_X, \Sigma_X)$ et $p(y|x) = \phi(y; Hx, R)$ alors
$$p(x|y) \propto p(y|x)p(x)dx = \phi(x; \mu_{X|y}, \Sigma_{X|y}).$$

- L'étape de correction du filtre de Kalman se déduit alors de l'expression
$$p(x_t|y_1, \cdots, y_t) \propto p(y_t|x_t)p(x_t|y_1, \cdots, y_{t-1})$$
avec $p(x_t|y_1, \cdots, y_{t-1}) = \phi(x_t; x_t^f, P_t^f)$ et $p(y_t|x_t) = \phi(y_t; Hx_t, R)$.

Quelques remarques sur le filtre de Kalman

- On peut expliciter le lisseur de Kalman de manière similaire



Time series $\{x_t\}$ (black curve) and $\{y_t\}$ (black dots) simulated with $Q^t = 1$ and $R^t = 1$, and reconstructions with a Kalman filter/smoother (red lines and gray 95% confidence interval).

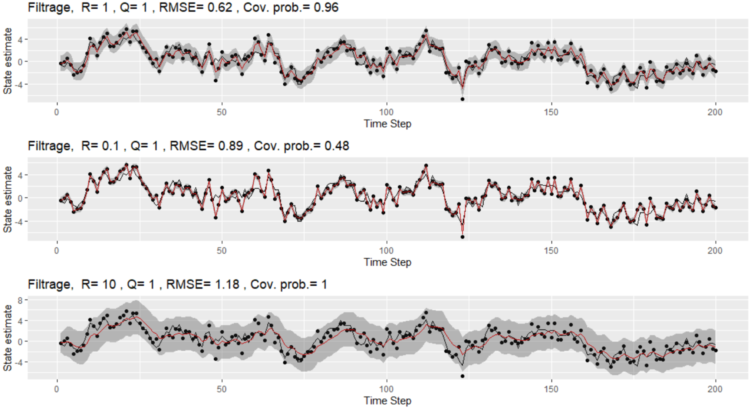
Quelques remarques sur le filtre de Kalman

- On peut expliciter le lisseur de Kalman de manière similaire
- Avantages du filtre de Kalman
  - Solution exacte
  - Coût de calcul faible en petite dimension
  - Implémenté dans les langages usuels (Python, R, Matlab,...)
- Inconvénients du filtre de Kalman
  - Linéaire+Gaussien : hypothèses fortes !
  - Coûts de calcul (produit et inversion de matrices) et de stockage en grande dimension
- Interprétation
  - Le filtre et le lisseur de Kalman sont des interpolateurs linéaires
  - Dans le filtre de Kalman, l'étape de correction s'écrit $x_t^a = (I - K_t H)x_t^f + K_t y_t$. L'état analysé est donc une moyenne pondérée entre la prédiction et l'observation et le gain de Kalman K_t s'interprète comme le "poids" donné à l'observation y_t
  - Le gain de Kalman est donné par $K_t = P_t^f H' (H P_t^f H' + R)^{-1}$. Intuitivement, K_t décroit lorsque le bruit d'observation R augmente : on donne moins de poids aux observations lorsque le bruit d'observation est important (cf transparent suivant)

Plan

Estimation des paramètres

- En général les paramètres du modèles sont inconnus
- Les valeurs de Q et R ont un impact sur la moyenne et la variance du filtre



Time series $\{x_t\}$ (black curve) and $\{y_t\}$ (black dots) simulated with $Q^t = 1$ and $R^t = 1$, and reconstructions with a Kalman filter (red lines and gray 95% confidence interval) with different values of Q and R.

Estimation des paramètres

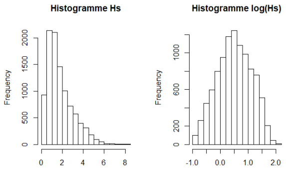
- En général les paramètres du modèles sont inconnus
- Les valeurs de Q et R ont un impact sur la moyenne et la variance du filtre
- Il est important d'estimer les paramètres avant d'utiliser le filtre de Kalman
- Méthode la plus classique : **maximum de vraisemblance** via l'**algorithme EM** ou optimisation numérique
- Alternative classique en assimilation de données : **méthode des moments** (cf exercice 1)

Plan

Résultats sur les données de Hs à Sainte Anne

Méthodologie :

- La transformation logarithme permet de "gaussianiser" Hs
- Centrage de la série temporelle
- Modélisation par un modèle espace-état linéaire gaussien
- Estimation des paramètres par maximum de vraisemblance
- Utilisation du lisseur de Kalman pour reconstruire la série temporelle



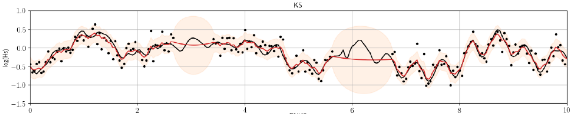


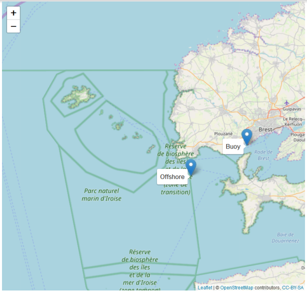
Figure – Séries temporelles $\{x_t\}$ (noir), $\{y_t\}$ (point), et probabilité de lissage (moyenne en rouge et intervalle de prédiction à 95%)

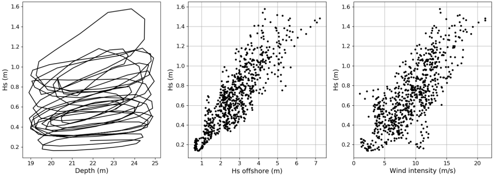
Extension : inclusion de covariables, non-linéarités (1)

La hauteur des vagues à Sainte Anne dépend de

- la marée (Hs plus grand à marée montante et à marée haute)
- la hauteur des vagues au large (propagation)
- vent local (mer du vent)

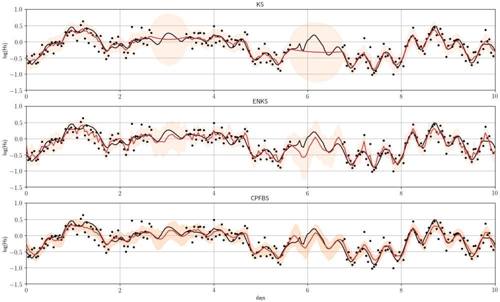
et certaines relations sont non-linéaires





Extension : inclusion de covariables, non-linéarités (2)

$$
\begin{array}{lccccccccc}
\text{Covariables} & \cdots & \rightarrow & Z_{t-1} & \rightarrow & Z_t & \rightarrow & Z_{t+1} & \rightarrow & \cdots \\
 & & (m,Q) & \downarrow & & \downarrow & & \downarrow & & \\
\text{Etat caché} & \cdots & \rightarrow & X_{t-1} & \rightarrow & X_t & \rightarrow & X_{t+1} & \rightarrow & \cdots \\
 & & (H,R) & \downarrow & & \downarrow & & \downarrow & & \\
\text{Observations} & \cdots & & Y_{t-1} & & Y_t & & Y_{t+1} & & \cdots
\end{array}
$$



Plan

En résumé

- Le filtre de Kalman permet de reconstruire de manière optimale l'état latent dans le cadre du modèle linéaire gaussien
- Il consiste à appliquer une étape de prédiction et une étape de correction à chaque pas de temps
- Il permet de calculer non seulement l'espérance (conditionelle) de l'état mais aussi sa variance, et donc de quantifier l'incertitude sur l'état reconstruit
- Son implémentation est facile et numériquement efficace
- Nombreuses applications : assimilation de données, mais aussi navigation, contrôle, économétrie, séries temporelles, ...
- Nombreuses généralisations : lisseur de Kalman, filtre de Kalman étendu, EnKF, filtres particulaires,...

Exercices

Exercice 1

On considère le modèle espace-état linéaire gaussien univarié (X_t et Y_t à valeurs dans $\mathbb{R}$).

1. *Calculer $E[X_t]$ et $var(X_t)$ en fonction de $E[X_{t-1}]$ et $var(X_{t-1})$.*
2. *On suppose que $X_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$ et que $E[X_t]$ et $var(X_t)$ ne dépendent pas de t. Exprimer alors μ_0 et Σ_0 en fonction de M et Q.*
3. *Calculer $cov(X_t, X_{t+k})$ (on pourra commencer par $k = 1$). En déduire que le processus $\{X_t\}$ est stationnaire (sens second ordre).*
4. *Exprimer $cov(Y_t, Y_{t+k})$ pour $k \in \mathbb{N}$ en fonction des paramètres du modèle. En déduire que le processus $\{Y_t\}$ est stationnaire (sens second ordre).*
5. *En inversant ces relations, en déduire des estimateurs de M, Q et R.*
6. *Montrer que $(Y_1, ..., Y_T)$ est un vecteur gaussien dont on précisera les caractéristiques. En déduire une expression de la fonction de vraisemblance $p(y_1, ..., y_T)$.*
7. *Montrer que $p(y_1, ..., y_T) = p(y_1) \prod_{t=2}^{T} p(y_t | y_1, ..., y_{t-1})$. Décrire une méthode pour calculer la fonction de vraisemblance à partir du filtre de Kalman.*

Exercices

Exercice 2

Soit $(X, Y) \sim \mathcal{N}\left(\left(\begin{array}{c}\mu_X \\ \mu_Y\end{array}\right), \left(\begin{array}{cc}\Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y\end{array}\right)\right)$*. On suppose que* Σ_Y *est inversible et on admet que* $E[X|Y] = \beta_0 + AY$.

1. *En utilisant que* $E[E[X|Y]] = E[X]$ *et* $E[(X - E[X|Y])'Y] = 0$*, donner des expressions pour* β_0 *et* A.
2. *Calculer* $var(Y - E[X|Y])$
3. *Vérifier qu'on retrouve l'expression de la loi conditionnelle de* X *sachant* $Y = y$ *donnée dans la Proposition 2.*

Exercice 3

On considère le processus ARMA(1,1) défini par

$$Y_t = \alpha_1 Y_{t-1} + E_t + \beta_1 E_{t-1}$$

avec $E_t \sim_{iid} \mathcal{N}(0, \sigma^2)$*. Montrer que* Y_t *a une représentation sous la forme d'un modèle espace-état.*
Indication : poser $\eta_t = \alpha_1 \eta_{t-1} + E_t$ *et montrer par récurrence que* $Y_t = \epsilon_t + \beta_1 \epsilon_{t-1}$.

Exercices

Exercice 4

On se place dans le cadre d'un modèle linéaire gaussien sans dynamique

- $X \sim \mathcal{N}(\mu_X, \Sigma_X)$ *(loi a priori sur l'état)*
- $Y = X + \epsilon$ *avec* $\epsilon \sim \mathcal{N}(0, R)$ *indépendant de* X *(équation d'observation)*

Pour simplifier, on suppose que X *et* Y *sont des variables aléatoires réelles.*

1. *On pose* $F(b_0, b_1) = E[(X - (b_0 + b_1 Y))^2]$. *Expliciter* $(\beta_0, \beta_1) = argmin_{(b_0, b_1) \in \mathbb{R}^2} F(b_0, b_1)$. $\hat{X} = \beta_0 + \beta_1 Y$ *s'interprète comme le prédicteur linéaire de* X *qui minimise l'erreur quadratique de prédiction. Montrer que c'est également le prédicteur sans biais qui est de variance minimale parmi les prédicteurs linéaires sans biais (BLUE).*
2. *Quelle est la loi du vecteur* (X, Y) *? Quelle est la loi conditionnelle de* X *sachant* $Y = y$ *? Montrer que le résultat de la question précédente se déduit de la Proposition 2. Quel résultat plus fort donne la Proposition 2 ?*

Références

Certaines figures et résultats sont issus des deux références ci-dessous.

- Chau, T. T. T., Ailliot, P., & Monbet, V. (2021). An algorithm for non-parametric estimation in state-space models. Computational Statistics & Data Analysis, 153, 107062.
- Tandeo, P., Ailliot, P., Bocquet, M., Carrassi, A., Miyoshi, T., Pulido, M., & Zhen, Y. (2020). A review of innovation-based methods to jointly estimate model and observation error covariance matrices in ensemble data assimilation. Monthly Weather Review, 148(10), 3973-3994.