

# **Inférence, filtrage et sélection de modèle bayésiens**

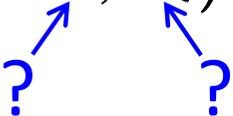
-

## **Quelques applications au PPP**

P. Rebischung, S. Nahmani

# Contexte (1/4)

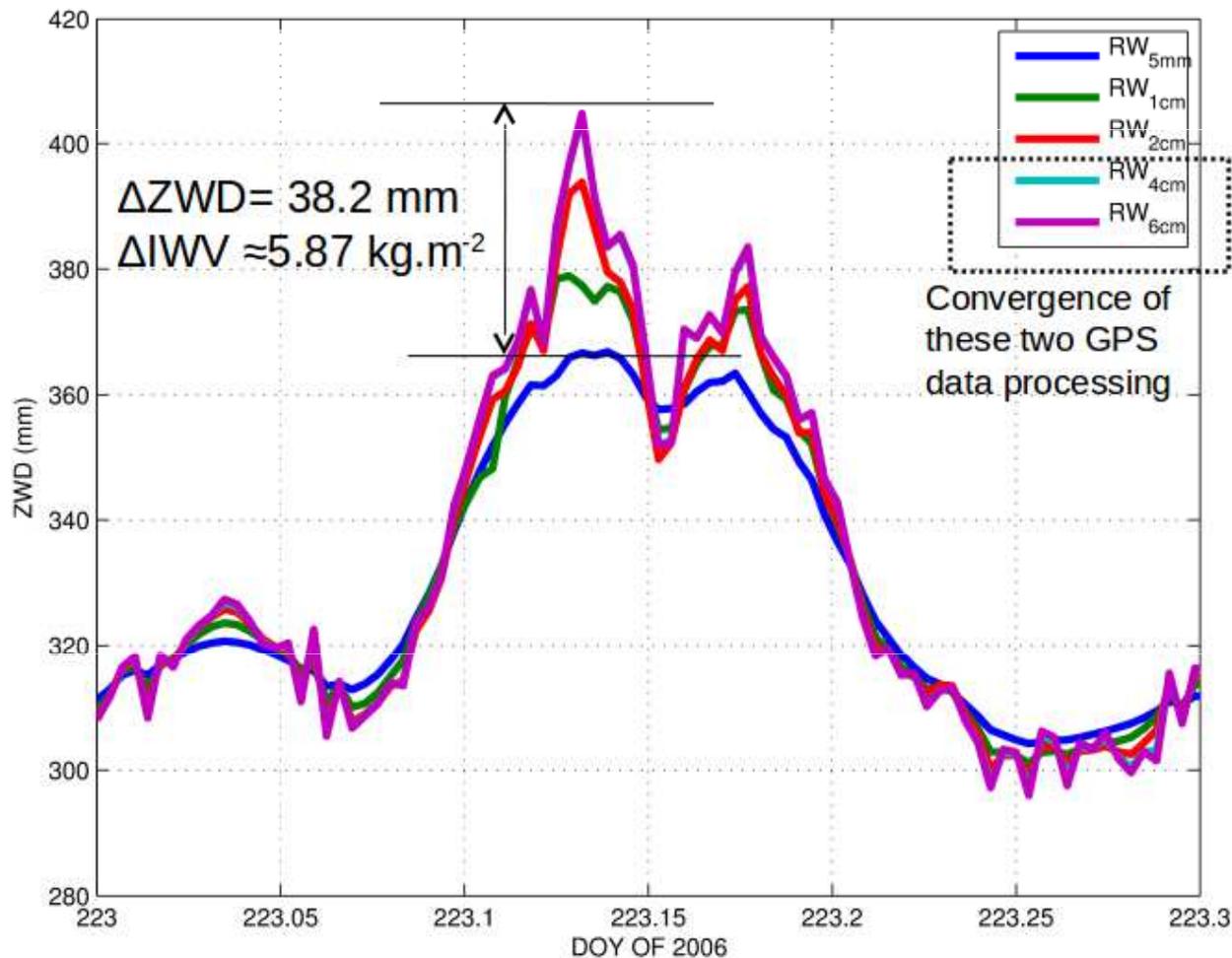
Q : Comment faire quand la partie stochastique du modèle n'est pas (complètement) connue ?

$$\mathbf{y} \sim \mathcal{N}(\mathbf{Ax}, \mathbf{Q})$$


- Ex : En PPP statique, la partie stochastique du modèle est généralement fixée.
  - Observations pondérées en  $\sigma_{\text{OBS}}^2 / \sin^2(\text{el})$
  - ZWD contraints à suivre une marche aléatoire de paramètre  $\sigma_{\text{ZWD}}$
  - Or  $\sigma_{\text{OBS}}$  et  $\sigma_{\text{ZWD}}$  peuvent varier suivant la station, la météo...
  - Et une pondération inappropriée peut influencer les résultats (cf. slide suivant).
- Ex : PPP cinématique, où il faut en plus modéliser l'évolution de la position.
- Ex : Comment pondérer les différentes observations de GEODESIE ?

A : Inférence bayésienne

# Contexte (2/4)



Impact du choix de différents  $\sigma_{ZWD}$  sur les ZWD estimés pendant le passage d'un MCS à Niamey

Lors d'événements intenses, un  $\sigma_{ZWD}$  trop serré empêche les ZWD estimés d'être suffisamment variables.

- biais sur les ZWD estimés (et autres paramètres)
- augmentation des résidus

Estimer  $\sigma_{OBS}$  et  $\sigma_{ZWD}$  peut permettre :

- d'améliorer les ZWD estimés (intérêt météorologique)
- d'améliorer le traitement GNSS (intérêt géodésique)
- de fournir des informations supplémentaires ( $\sigma_{OBS}$  et  $\sigma_{ZWD}$ )

# Contexte (3/4)

Q : Comment choisir le modèle le plus adapté aux données ?

$$\mathbf{y} \sim \mathcal{N}(\mathbf{A}_1 \mathbf{x}_1, \mathbf{Q}_1) \quad \overset{?}{\longleftrightarrow} \quad \mathbf{y} \sim \mathcal{N}(\mathbf{A}_2 \mathbf{x}_2, \mathbf{Q}_2)$$

- Ex : En PPP statique,
  - Pondération des observations :  $\sigma_{\text{OBS}}^2 / \sin^2(\text{el})$  ou  $a^2 + b^2 / \sin^2(\text{el})$  ?
  - ZWD : horaires, semi-horaires, marche aléatoire ?
  - Gradients : constants, linéaires, horaires, marche aléatoire ?

A : Sélection de modèle bayésienne

# Contexte (4/4)

- T2.2 Variance Component Estimator

“establish, test and validate a Variance Component Estimation method to

(i) rigorously weight the measurements

(ii) automatically set the statistical parameters of the random processes at hand in the evolution model of the KFS”

- Comment pondérer de façon optimale les observations et les équations d'évolution dans un filtre ?

- Réponse théorique : filtrage bayésien

- Réponse pratique : ?

# Sommaire

- Inférence bayésienne
  - Principe
  - Estimation des composantes de la variance
- Sélection de modèle bayésienne
  - Principe
  - Calcul de l'évidence d'un modèle
- Quelques applications au PPP
- Filtrage bayésien
  - Principe
  - Problème de l'estimation des composantes de la variance

# 1 - Inférence bayésienne

- Problème :
  - On dispose d'observations  $\mathbf{y}$  dépendant de paramètres inconnus  $\mathbf{x}$ .
  - On connaît la probabilité  $p(\mathbf{y}|\mathbf{x})$  d'observer  $\mathbf{y}$  sachant  $\mathbf{x}$ .
  - On dispose éventuellement d'une probabilité a priori  $p(\mathbf{x})$  pour  $\mathbf{x}$ .
  - Quelle information a-t-on sur  $\mathbf{x}$  connaissant les observations  $\mathbf{y}$  ?
- Solution : théorème de Bayes

$$p(\mathbf{x} | \mathbf{y}) = \frac{p(\mathbf{y} | \mathbf{x}) p(\mathbf{x})}{p(\mathbf{y})} \propto p(\mathbf{y} | \mathbf{x}) p(\mathbf{x})$$

probabilité a posteriori  $\propto$  vraisemblance  $\times$  probabilité a priori

# Moindres carrés à la bayésienne

- Erreurs d'observation normales, avec matrice de variance  $\mathbf{Q}$  connue :

$$\mathbf{y} | \mathbf{x} \sim \mathcal{N}(\mathbf{Ax}, \mathbf{Q}) \rightarrow p(\mathbf{y} | \mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{Q}|}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{Ax})^\top \mathbf{Q}^{-1}(\mathbf{y}-\mathbf{Ax})}$$

- Aucune information a priori sur  $\mathbf{x}$  (*improper prior*) :  $p(\mathbf{x}) = 1$

$$\rightarrow p(\mathbf{x} | \mathbf{y}) \propto e^{-\frac{1}{2}(\mathbf{y}-\mathbf{Ax})^\top \mathbf{Q}^{-1}(\mathbf{y}-\mathbf{Ax})} \propto e^{-\frac{1}{2}(\mathbf{x}-\hat{\mathbf{x}})^\top \mathbf{A}^\top \mathbf{Q}^{-1} \mathbf{A}(\mathbf{x}-\hat{\mathbf{x}})}$$

où  $\hat{\mathbf{x}} = (\mathbf{A}^\top \mathbf{Q}^{-1} \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{Q}^{-1} \mathbf{y}$  est la solution des moindres carrés.

- Sachant  $\mathbf{y}$ , les paramètres  $\mathbf{x}$  suivent une loi normale de moyenne  $\hat{\mathbf{x}}$  et de matrice de variance  $(\mathbf{A}^\top \mathbf{Q}^{-1} \mathbf{A})^{-1}$  :

$$\mathbf{x} | \mathbf{y} \sim \mathcal{N}(\hat{\mathbf{x}}, (\mathbf{A}^\top \mathbf{Q}^{-1} \mathbf{A})^{-1})$$

# Composantes de la variance

- On sait que  $\mathbf{y} \sim \mathcal{N}(\mathbf{Ax}, \mathbf{Q})$ , mais  $\mathbf{Q}$  n'est pas entièrement connue :

$$\mathbf{Q} = \sum \sigma_i^2 \mathbf{Q}_i = \sum e^{\beta_i} \mathbf{Q}_i$$

où les  $\sigma_i^2 = e^{\beta_i}$  sont inconnus.

- On pose  $\boldsymbol{\beta} = [\beta_1, \dots, \beta_{n_v}]^\top = [\ln \sigma_1^2, \dots, \ln \sigma_{n_v}^2]^\top$
- On choisit des probabilités a priori non-informatives sur  $\mathbf{x}$  et  $\boldsymbol{\beta}$  :  $p(\mathbf{x}, \boldsymbol{\beta}) = 1$

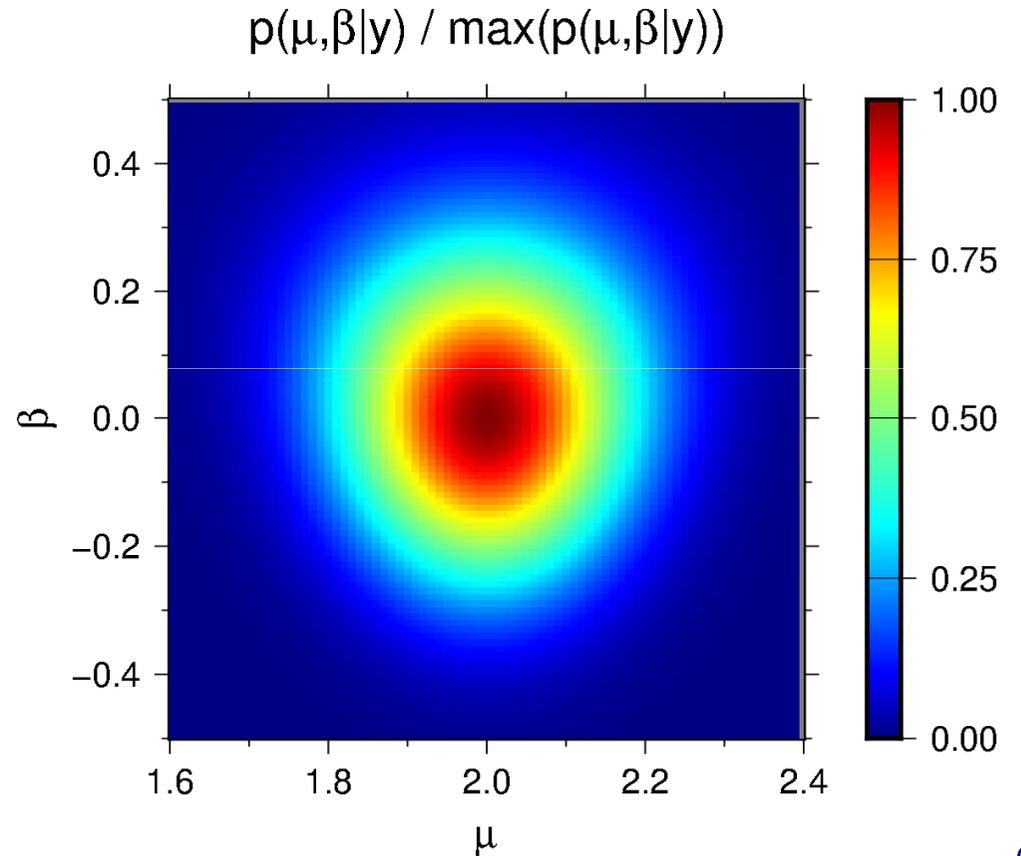
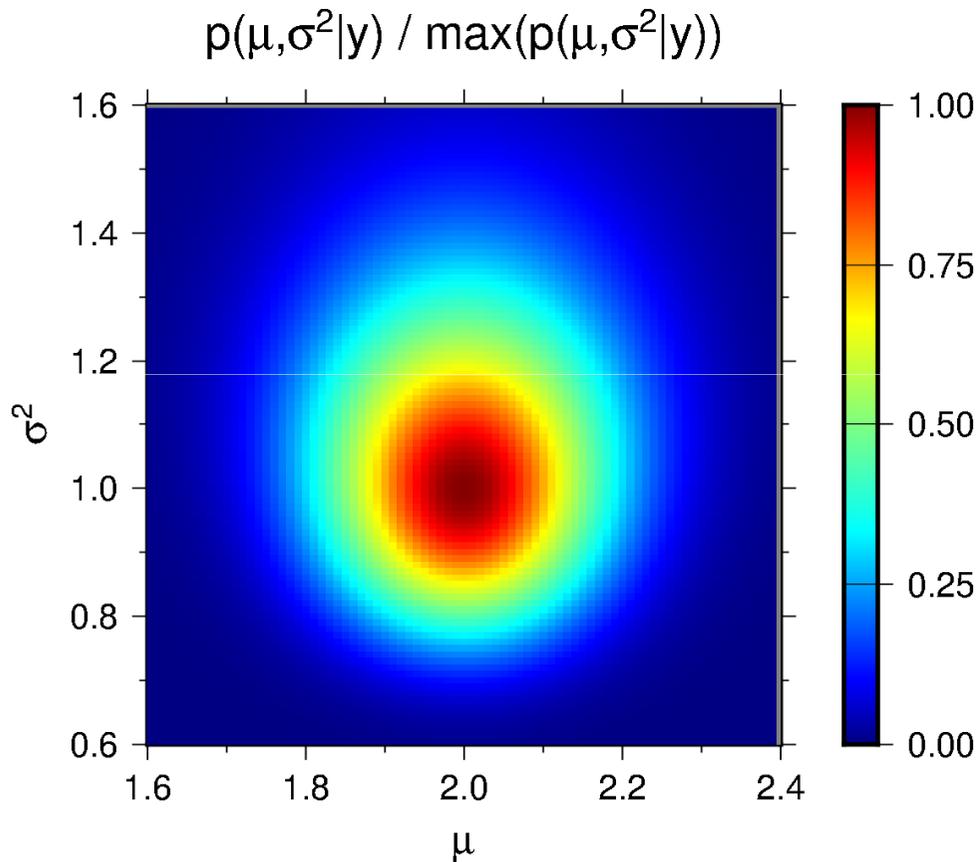
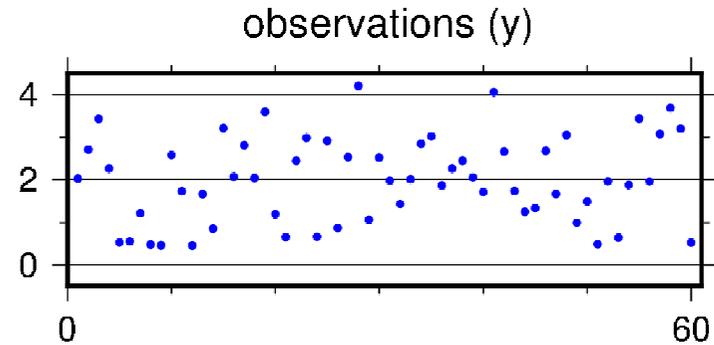
$$\rightarrow p(\mathbf{x}, \boldsymbol{\beta} | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{x}, \boldsymbol{\beta}) p(\mathbf{x}, \boldsymbol{\beta}) \propto \frac{1}{\sqrt{(2\pi)^n |\mathbf{Q}|}} e^{-\frac{1}{2} [\hat{\mathbf{v}}^\top \mathbf{Q}^{-1} \hat{\mathbf{v}} + (\mathbf{x} - \hat{\mathbf{x}})^\top \mathbf{A}^\top \mathbf{Q}^{-1} \mathbf{A} (\mathbf{x} - \hat{\mathbf{x}})]}$$

où  $\mathbf{Q}$ ,  $\hat{\mathbf{x}}$  et  $\hat{\mathbf{v}} = \mathbf{y} - \mathbf{A}\hat{\mathbf{x}}$  dépendent de  $\boldsymbol{\beta}$ .

→ densité de probabilité non-analytique

# Exemple

- On dispose de 60 observations (tirées selon  $\mathcal{N}(2, 1)$ ).
- Quelle information a-t-on sur leur moyenne et leur variance ?



# Quel estimateur pour $\boldsymbol{\beta}$ ?

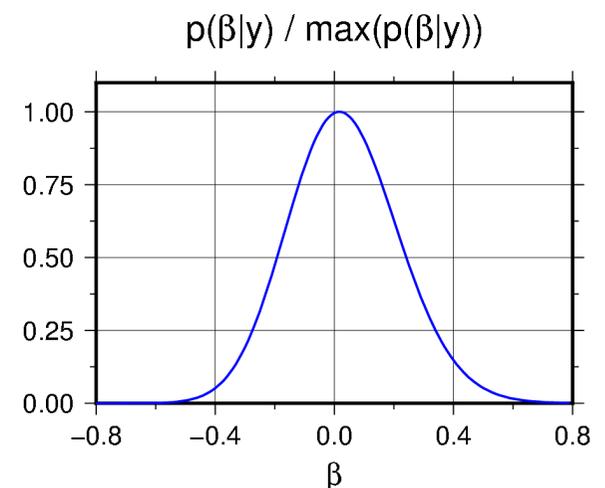
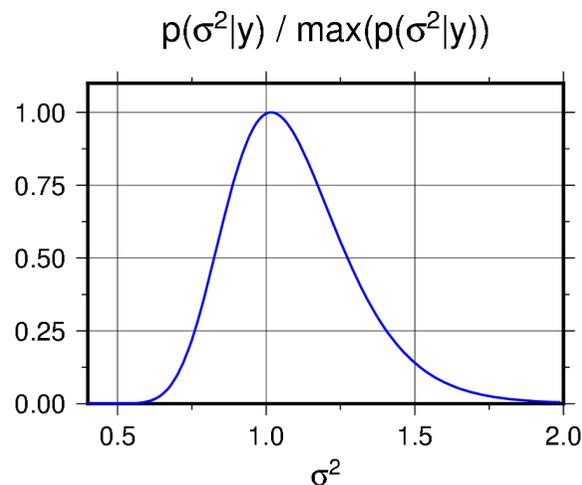
- Si l'on utilise des probabilités a priori non-informatives, maximiser  $p(\mathbf{x}, \boldsymbol{\beta} | \mathbf{y}) \leftrightarrow$  maximum de vraisemblance (bof)
- Mais si l'on maximise la probabilité marginale de  $\boldsymbol{\beta}$  sachant  $\mathbf{y}$  :

$$p(\boldsymbol{\beta} | \mathbf{y}) = \int p(\mathbf{x}, \boldsymbol{\beta} | \mathbf{y}) d\mathbf{x} \propto \frac{1}{\sqrt{(2\pi)^{n-p} |\mathbf{Q}| |\mathbf{A}^\top \mathbf{Q}^{-1} \mathbf{A}|}} e^{-\frac{1}{2} \hat{\mathbf{v}}^\top \mathbf{Q}^{-1} \hat{\mathbf{v}}}$$

alors on retombe sur l'estimateur du « degré de liberté ».

(= MINQUE = BIQUE = Helmert = REML = LS-VCE)

- Retour à l'exemple :



# Quels intérêts ?

- Certes, on retombe sur un résultat classique, mais on a une information supplémentaire : l'incertitude sur les composantes de la variance.
- Si l'on approxime  $p(\boldsymbol{\theta} | \mathbf{y})$  par une gaussienne, on peut en particulier fournir des « facteurs d'erreur » sur les  $\sigma_i^2$ .
- La probabilité marginale de  $\mathbf{x}$  sachant  $\mathbf{y}$  :  $p(\mathbf{x} | \mathbf{y}) = \int p(\mathbf{x}, \boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}$  permet en outre de tenir compte de l'incertitude sur  $\mathbf{x}$  due à l'estimation des composantes de la variance.
- On peut généraliser la méthode à n'importe quel type de paramètres, par exemple les coefficients  $\varphi_i$  d'un AR.

# 2 - Sélection de modèle bayésienne

- Problème :
  - On dispose d'observations  $\mathbf{y}$  pouvant être expliquées par deux modèles  $M_1$  et  $M_2$ .
  - Quelle est le modèle le plus probable connaissant les observations  $\mathbf{y}$  ?

- Solution : facteur de Bayes

$$\frac{p(M_1 | \mathbf{y})}{p(M_2 | \mathbf{y})} = \frac{p(\mathbf{y} | M_1) p(M_1)}{p(\mathbf{y} | M_2) p(M_2)}$$

- Si l'on considère tous les modèles a priori équiprobables, le modèle le plus probable sachant  $\mathbf{y}$  est celui dont l'évidence est maximale.
- Evidence du modèle  $M$  :  $E(M) = p(\mathbf{y} | M) = \int p(\mathbf{y} | \mathbf{x}_M, M) p(\mathbf{x}_M | M) d\mathbf{x}_M$   
où  $\mathbf{x}_M$  contient l'ensemble des paramètres inconnus du modèle  $M$ .

# Calcul de l'évidence

- Modèle comportant des paramètres déterministes uniquement ( $\mathbf{Q}$  connu) :

$$E = \int p(\mathbf{y}|\mathbf{x})p(\mathbf{x})d\mathbf{x} = \int \frac{1}{\sqrt{(2\pi)^n|\mathbf{Q}|}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{A}\mathbf{x})^\top \mathbf{Q}^{-1}(\mathbf{y}-\mathbf{A}\mathbf{x})} d\mathbf{x} = \frac{1}{\sqrt{(2\pi)^{n-p}|\mathbf{Q}||\mathbf{A}^\top \mathbf{Q}^{-1} \mathbf{A}|}} e^{-\frac{1}{2}\hat{\mathbf{v}}^\top \mathbf{Q}^{-1}\hat{\mathbf{v}}}$$

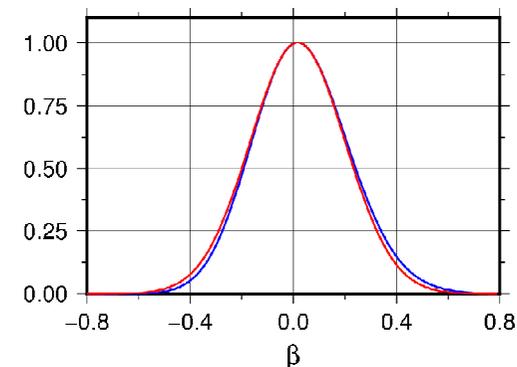
- Modèle comportant paramètres déterministes & composantes de la variance :

$$E = \int p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})p(\mathbf{x}, \boldsymbol{\theta})d\mathbf{x} d\boldsymbol{\theta} = \int \underbrace{\frac{1}{\sqrt{(2\pi)^{n-p}|\mathbf{Q}||\mathbf{A}^\top \mathbf{Q}^{-1} \mathbf{A}|}}}_{L(\boldsymbol{\theta})} e^{-\frac{1}{2}\hat{\mathbf{v}}^\top \mathbf{Q}^{-1}\hat{\mathbf{v}}} d\boldsymbol{\theta}$$

Si  $L(\boldsymbol{\theta}) \propto p(\boldsymbol{\theta}|\mathbf{y})$  est à peu près une gaussienne, on peut calculer E sans intégration numérique grâce à l'approximation de Laplace :

$$L(\boldsymbol{\theta}) \approx L(\hat{\boldsymbol{\theta}}) e^{\frac{1}{2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^\top \mathbf{H}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})} \quad \text{où } \mathbf{H} = \left. \frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \right|_{\hat{\boldsymbol{\theta}}}$$

$$\rightarrow E \approx \sqrt{\frac{(2\pi)^{n_v}}{|\mathbf{H}|}} L(\hat{\boldsymbol{\theta}})$$



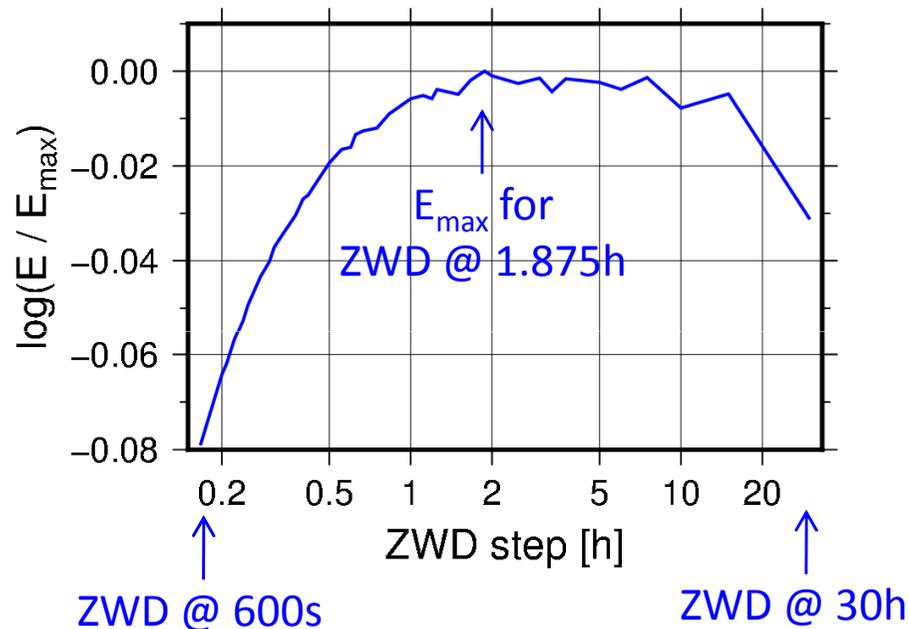
# 3 - Expériences PPP

- Options figées :
  - Sessions de 30h
  - Orbites et horloges satellites « maison » (NAPEOS, arcs de 30h)
  - Observations de phase iono-free
  - GPS uniquement
  - Echantillonnage à 300s
  - Angle de coupure à 5°
  - Ambiguïtés non-fixées
  - Equations d'observation obtenues via NAPEOS
  - Points faux rejetés sur la base des résidus normalisés d'un traitement « standard »
- Options testées :
  - Pondération des observations :  $\sigma_{OBS} / \sin^\alpha(el)$
  - Paramétrage des ZWD : constante / marches d'escalier / RW( $\sigma_{ZWD}$ )
  - Paramétrage des gradients : constants / marches d'escalier / RW( $\sigma_{GRA}$ )
  - Paramétrage de la position : constante / marches d'escalier / RW( $\sigma_{POS}$ )

# Paramétrage des ZWD

- Pondération des observations :  $\sigma_{OBS} / \sin(el)$
- Paramétrage des ZWD : marches d'escalier /  $RW(\sigma_{ZWD})$
- Paramétrage des gradients : constants
- Paramétrage de la position : constante

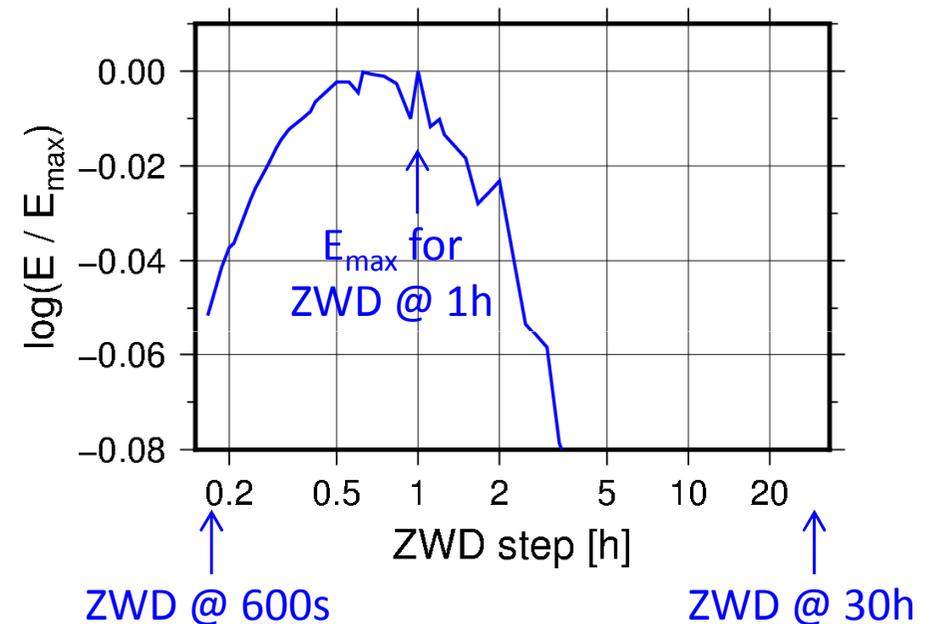
ALGO 08–Nov–2013



$$\log[ E(RW) / E_{\max}(\text{step}) ] = +0.008$$

$$\sigma_{ZWD} = 2.29 \text{ mm/vh } \times / \div 1.12$$

NIAM 11–Aug–2008



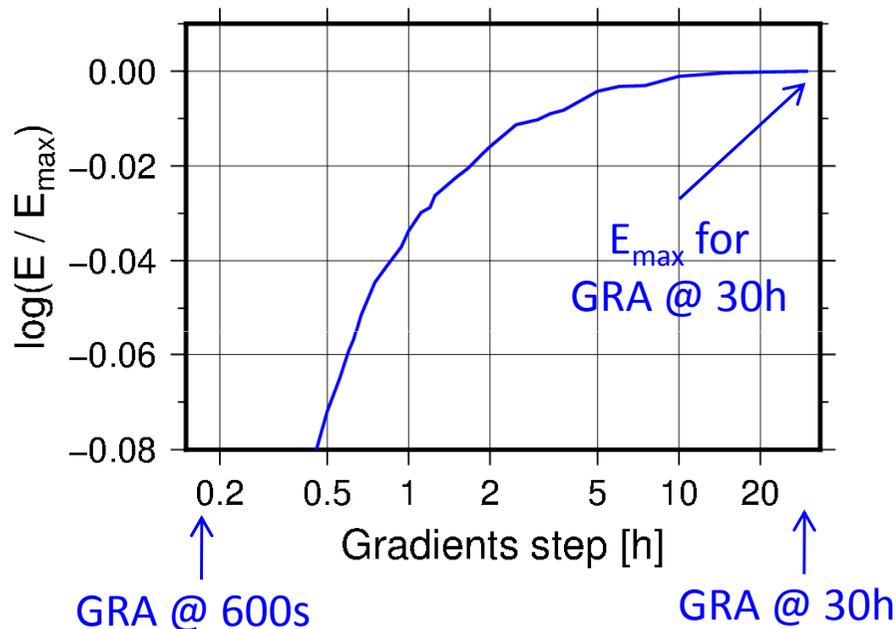
$$\log[ E(RW) / E_{\max}(\text{step}) ] = +0.028$$

$$\sigma_{ZWD} = 10.35 \text{ mm/vh } \times / \div 1.07$$

# Paramétrage des gradients

- Pondération des observations :  $\sigma_{OBS} / \sin(el)$
- Paramétrage des ZWD :  $RW(\sigma_{ZWD})$
- Paramétrage des gradients : marches d'escalier /  $RW(\sigma_{GRA})$
- Paramétrage de la position : constante

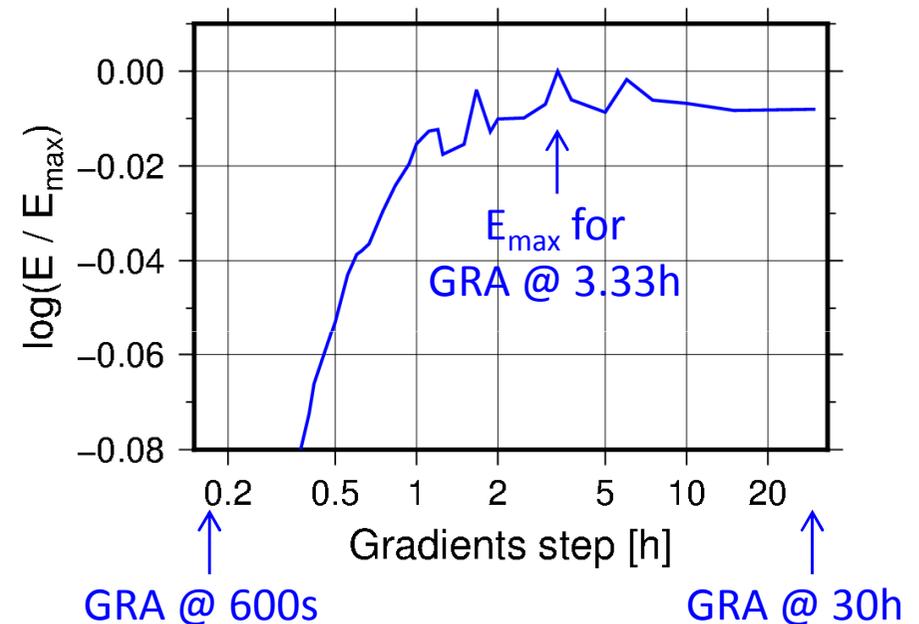
ALGO 08–Nov–2013



$$\log[ E(RW) / E_{\max}(\text{step}) ] = +0.004$$

$$\sigma_{GRA} = 0.33 \text{ mm/vh } \times / \div 1.19$$

NIAM 11–Aug–2008



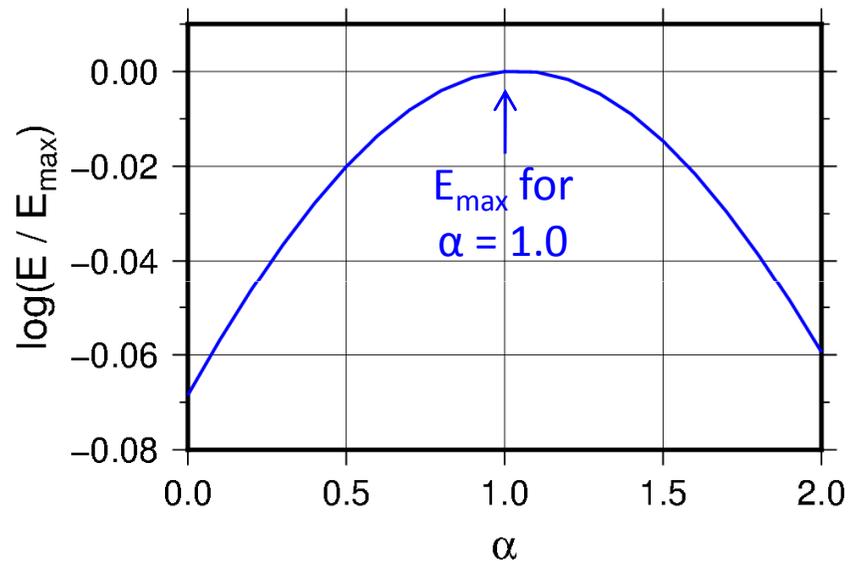
$$\log[ E(RW) / E_{\max}(\text{step}) ] = +0.027$$

$$\sigma_{GRA} = 1.22 \text{ mm/vh } \times / \div 1.08$$

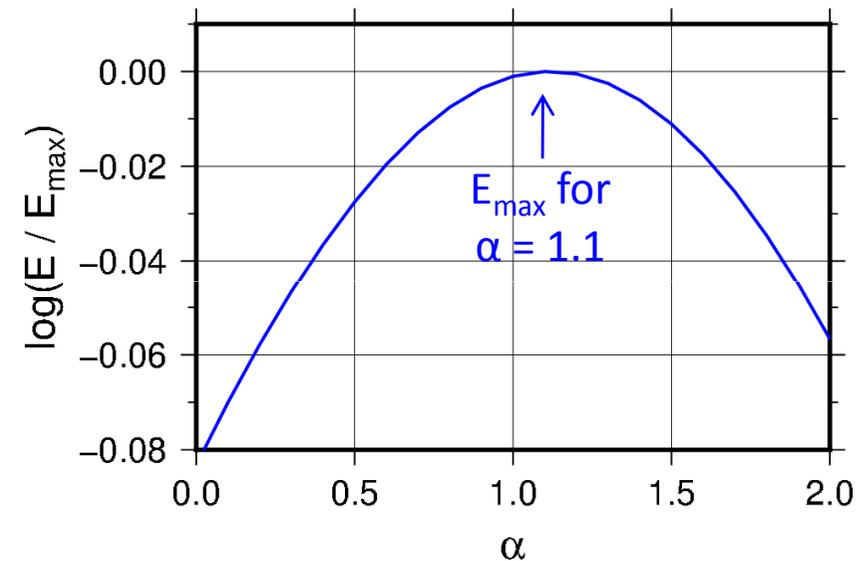
# Pondération des observations

- Pondération des observations :  $\sigma_{\text{OBS}} / \sin^\alpha(e)$
- Paramétrage des ZWD :  $\text{RW}(\sigma_{\text{ZWD}})$
- Paramétrage des gradients :  $\text{RW}(\sigma_{\text{GRA}})$
- Paramétrage de la position : constante

ALGO 08–Nov–2013



NIAM 11–Aug–2008



# ZWD et gradients estimés

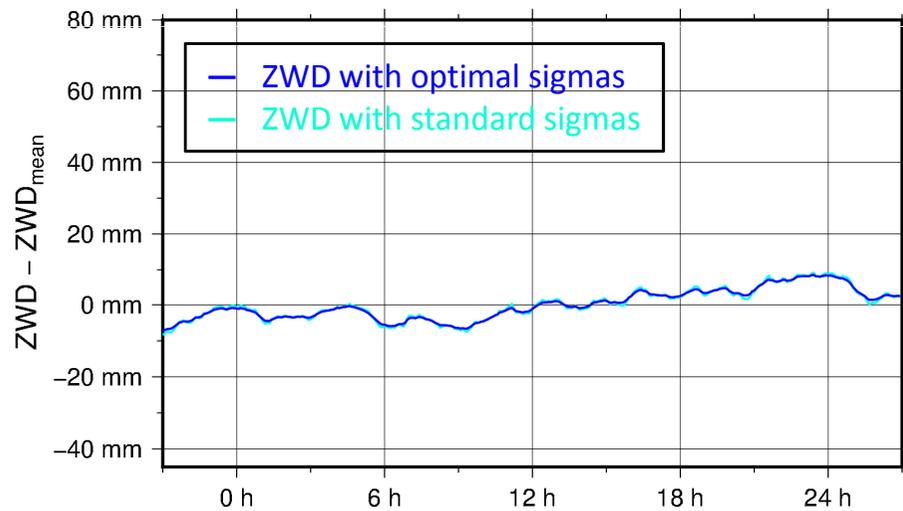
Observations:  $\sigma_{OBS} / \sin(el)$

ZWD : RW( $\sigma_{ZWD}$ )

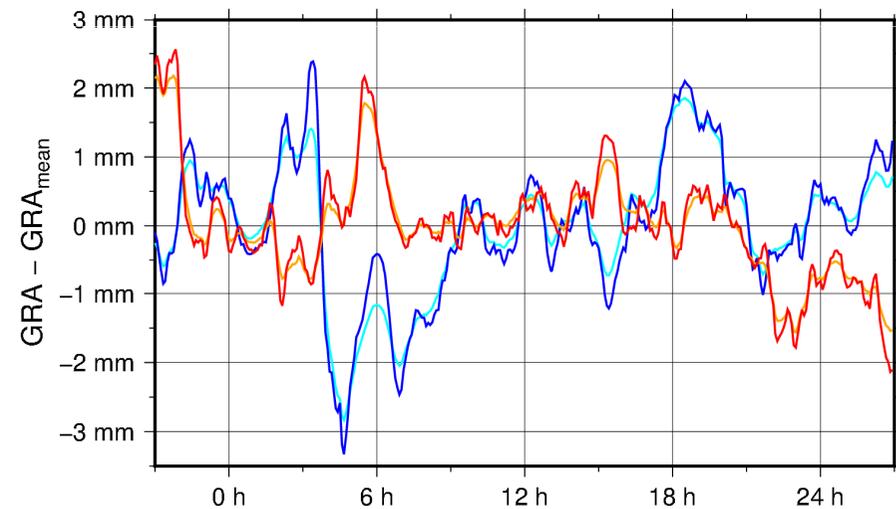
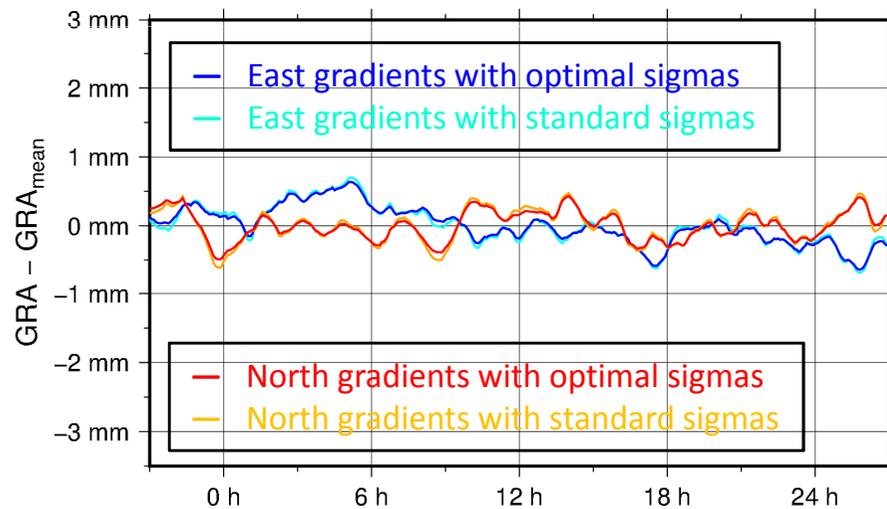
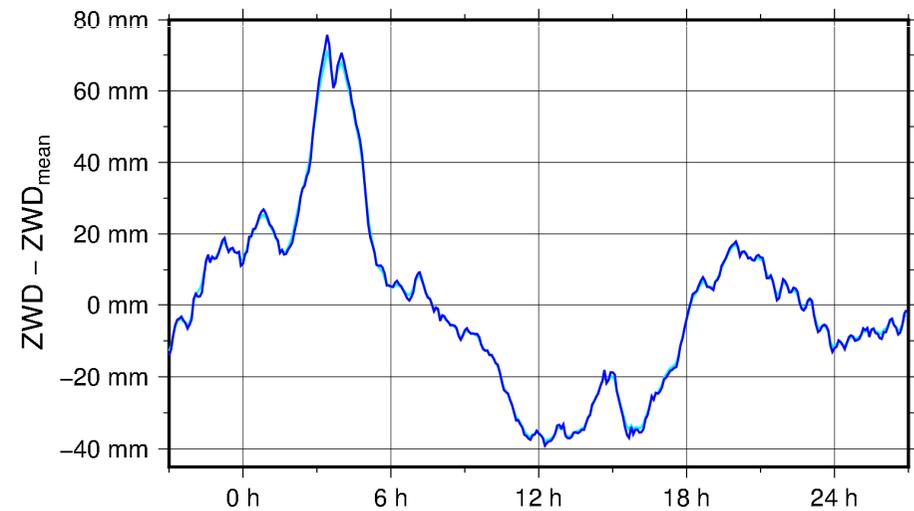
Gradients : RW( $\sigma_{GRA}$ )

Position : constante

ALGO 08–Nov–2013



NIAM 11–Aug–2006



# Comparaison à un radiomètre

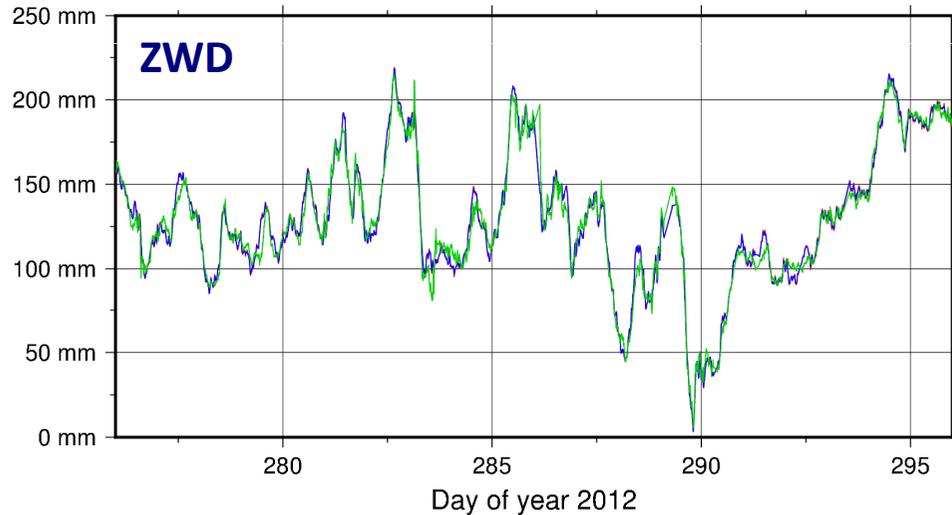
Observations:  $\sigma_{OBS} / \sin(el)$

ZWD :  $RW(\sigma_{ZWD})$

Gradients : constants

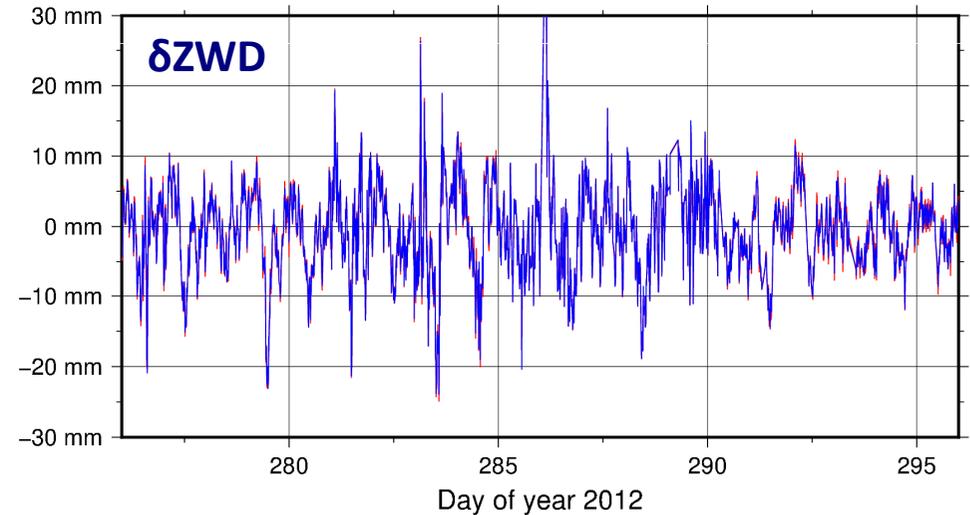
Position : constante

CORT 02-Oct-2012 to 21-Oct-2012

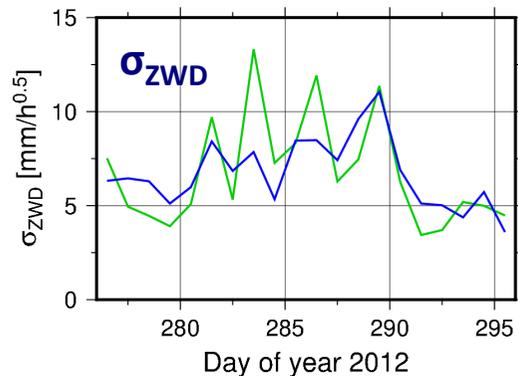


- Bias-corrected radiometer-derived ZWDs
- GPS-derived ZWDs with optimal sigmas
- GPS-derived ZWDs with standard sigmas

CORT 02-Oct-2012 to 21-Oct-2012



- GPS with optimal sigmas - radiometer
- GPS with standard sigmas - radiometer



RMS of GPS – radiometer ZWD differences:

ZWD : RW GRA : constant standard sigmas	ZWD : RW GRA : constant optimal sigmas	ZWD : RW GRA : RW standard sigmas	ZWD : RW GRA : RW optimal sigmas	ZWD : RW GRA : constant « best » sigmas
5.85 mm	5.75 mm	6.15 mm	6.22 mm	5.33 mm

# Questions

- A quel point peut-on faire confiance à l'évidence comme critère de sélection de modèle, et aux sigmas optimaux ?
  - Tous les modèles que l'on compare sont faux (erreurs systématiques).
  - L'évidence a tendance à augmenter pour des modèles de plus en plus libres, où les erreurs systématiques se font absorber (cf. gradients ALGO ; positions des stations).
  - Les sigmas optimaux semblent surévalués (cf. comparaison au radiomètre).
- Comment améliorer la modélisation stochastique des erreurs ?
  - Un modèle stochastique plus réaliste pour les erreurs devrait rendre les évidences et les sigmas optimaux plus fiables.
  - Mais comment construire une modélisation stochastique des multi-trajets ??
- Faudrait-il faire varier  $\sigma_{ZWD}$  et  $\sigma_{GRA}$  au cours du temps ?
  - Au vu des résultats de Samuel et des séries de radiomètre : oui
  - Faisable en principe par filtrage bayésien
  - Difficile à mettre en pratique (voir suite)

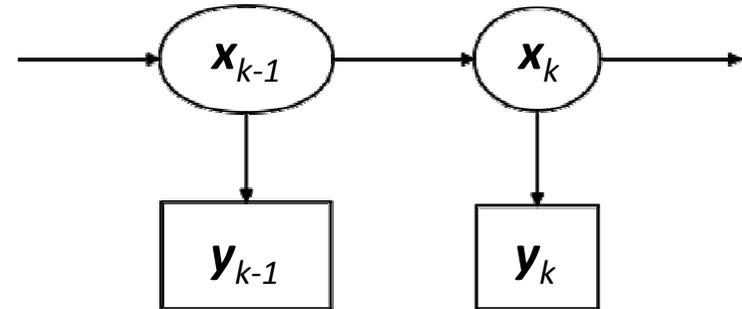
# A faire

- Améliorations PPP
  - Fixation des ambiguïtés
  - Multi-GNSS (impact impressionnant sur les gradients – Li et al., 2015)
  - Echantillonnage à 30s voire 1s – prise en compte des corrélations temporelles
  - GINS ?
- Amélioration des modèles stochastiques
  - Pour les observations (multi-trajets, turbulence)
  - Pour les paramètres troposphériques (marche aléatoire variable, autre ?)
- Etude d'impact sur les paramètres estimés
  - Comparaison des paramètres troposphériques à des données externes (radiomètres, ECMWF)
  - Séries temporelles de positions
- « Climatologie » des  $\sigma_{ZWD}$  et  $\sigma_{GRA}$

# 4 - Filtrage bayésien

- Problème :

- On connaît  $p(\mathbf{x}_k/\mathbf{x}_{k-1})$  et  $p(\mathbf{y}_k/\mathbf{x}_k)$ .
- On cherche  $p(\mathbf{x}_k/\mathbf{y}_{1:k})$ .



- Solution :

- Supposons que l'étape précédente du filtre nous donne  $p(\mathbf{x}_{k-1}/\mathbf{y}_{1:k-1})$ .

- Prédiction (Chapman-Kolmogorov) :

$$p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1}$$

- Mise à jour (théorème de Bayes) :

$$p(\mathbf{x}_k | \mathbf{y}_{1:k}) \propto p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_{1:k-1})$$

# Kalman à la bayésienne (1/2)

- Erreurs d'observation normales, avec matrice de variance  $\mathbf{Q}_k$  connue :

$$\mathbf{y}_k | \mathbf{x}_k \sim \mathcal{N}(\mathbf{A}_k \mathbf{x}_k, \mathbf{Q}_k) \rightarrow p(\mathbf{y}_k | \mathbf{x}_k) = \frac{1}{\sqrt{(2\pi)^{n_k} |\mathbf{Q}_k|}} e^{-\frac{1}{2}(\mathbf{y}_k - \mathbf{A}_k \mathbf{x}_k)^\top \mathbf{Q}_k^{-1} (\mathbf{y}_k - \mathbf{A}_k \mathbf{x}_k)}$$

- Erreurs de transition normales, avec matrice de variance  $\mathbf{R}_k$  connue :

$$\mathbf{x}_k | \mathbf{x}_{k-1} \sim \mathcal{N}(\mathbf{F}_k \mathbf{x}_{k-1}, \mathbf{R}_k) \rightarrow p(\mathbf{x}_k | \mathbf{x}_{k-1}) = \frac{1}{\sqrt{(2\pi)^{p_k} |\mathbf{R}_k|}} e^{-\frac{1}{2}(\mathbf{x}_k - \mathbf{F}_k \mathbf{x}_{k-1})^\top \mathbf{R}_k^{-1} (\mathbf{x}_k - \mathbf{F}_k \mathbf{x}_{k-1})}$$

- On suppose que d'après l'étape précédente du filtre :

$$\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1} \sim \mathcal{N}(\hat{\mathbf{x}}_{k-1}, \mathbf{S}_{k-1}) \rightarrow p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) = \frac{1}{\sqrt{(2\pi)^{p_{k-1}} |\mathbf{S}_{k-1}|}} e^{-\frac{1}{2}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1})^\top \mathbf{S}_{k-1}^{-1} (\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1})}$$

# Kalman à la bayésienne (2/2)

- Prédiction : 
$$p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) = \frac{1}{\sqrt{(2\pi)^{p_k} |\mathbf{F}_k \mathbf{S}_{k-1} \mathbf{F}_k^\top + \mathbf{R}_k|}} e^{-\frac{1}{2}(\mathbf{x}_k - \mathbf{F}_k \hat{\mathbf{x}}_{k-1})^\top (\mathbf{F}_k \mathbf{S}_{k-1} \mathbf{F}_k^\top + \mathbf{R}_k)^{-1} (\mathbf{x}_k - \mathbf{F}_k \hat{\mathbf{x}}_{k-1})}$$

→  $\mathbf{x}_k | \mathbf{y}_{1:k-1} \sim \mathcal{N}(\mathbf{F}_k \hat{\mathbf{x}}_{k-1}, \mathbf{F}_k \mathbf{S}_{k-1} \mathbf{F}_k^\top + \mathbf{R}_k)$

- Mise à jour : 
$$p(\mathbf{x}_k | \mathbf{y}_{1:k}) \propto e^{-\frac{1}{2}[(\mathbf{x}_k - \mathbf{F}_k \hat{\mathbf{x}}_{k-1})^\top (\mathbf{F}_k \mathbf{S}_{k-1} \mathbf{F}_k^\top + \mathbf{R}_k)^{-1} (\mathbf{x}_k - \mathbf{F}_k \hat{\mathbf{x}}_{k-1}) + (\mathbf{y}_k - \mathbf{A}_k \mathbf{x}_k)^\top \mathbf{Q}_k^{-1} (\mathbf{y}_k - \mathbf{A}_k \mathbf{x}_k)]}$$

$$p(\mathbf{x}_k | \mathbf{y}_{1:k}) \propto e^{-\frac{1}{2}(\mathbf{x}_k - \hat{\mathbf{x}}_k)^\top \mathbf{S}_k^{-1} (\mathbf{x}_k - \hat{\mathbf{x}}_k)}$$

→  $\mathbf{x}_k | \mathbf{y}_{1:k} \sim \mathcal{N}(\hat{\mathbf{x}}_k, \mathbf{S}_k)$

avec : 
$$\mathbf{S}_k = \left( (\mathbf{F}_k \mathbf{S}_{k-1} \mathbf{F}_k^\top + \mathbf{R}_k)^{-1} + \mathbf{A}_k^\top \mathbf{Q}_k^{-1} \mathbf{A}_k \right)^{-1}$$

$$\hat{\mathbf{x}}_k = \mathbf{S}_k^{-1} \left( (\mathbf{F}_k \mathbf{S}_{k-1} \mathbf{F}_k^\top + \mathbf{R}_k)^{-1} \mathbf{B}_k \hat{\mathbf{x}}_{k-1} + \mathbf{A}_k^\top \mathbf{Q}_k^{-1} \mathbf{y}_k \right)$$

# Kalman + variance components (1/3)

- Variance des observations partiellement connue :  $\mathbf{Q}_k(\boldsymbol{\theta}) = \sum e^{\beta_i} \mathbf{Q}_{k,i}$

$$\rightarrow p(\mathbf{y}_k | \mathbf{x}_k, \boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^{n_k} |\mathbf{Q}_k(\boldsymbol{\theta})|}} e^{-\frac{1}{2}(\mathbf{y}_k - \mathbf{A}_k \mathbf{x}_k)^\top \mathbf{Q}_k^{-1}(\boldsymbol{\theta})(\mathbf{y}_k - \mathbf{A}_k \mathbf{x}_k)}$$

- Variance de la transition partiellement connue :  $\mathbf{R}_k(\boldsymbol{\theta}) = \sum e^{\beta_i} \mathbf{R}_{k,i}$

$$\rightarrow p(\mathbf{x}_k | \mathbf{x}_{k-1}, \boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^{p_k} |\mathbf{R}_k(\boldsymbol{\theta})|}} e^{-\frac{1}{2}(\mathbf{x}_k - \mathbf{F}_k \mathbf{x}_{k-1})^\top \mathbf{R}_k^{-1}(\boldsymbol{\theta})(\mathbf{x}_k - \mathbf{F}_k \mathbf{x}_{k-1})}$$

- On suppose que l'étape précédente du filtre nous donne :

$$\begin{aligned} p(\boldsymbol{\theta}, \mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) &= p(\boldsymbol{\theta} | \mathbf{y}_{1:k-1}) p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}, \boldsymbol{\theta}) \\ &= p(\boldsymbol{\theta} | \mathbf{y}_{1:k-1}) \frac{1}{\sqrt{(2\pi)^{p_{k-1}} |\mathbf{S}_{k-1}(\boldsymbol{\theta})|}} e^{-\frac{1}{2}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}(\boldsymbol{\theta}))^\top \mathbf{S}_{k-1}^{-1}(\boldsymbol{\theta})(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}(\boldsymbol{\theta}))} \end{aligned}$$

# Kalman + variance components (2/3)

- On obtient la récurrence suivante :

$$S_k(\boldsymbol{\theta}) = \left( \left( F_k S_{k-1}(\boldsymbol{\theta}) F_k^T + R_k(\boldsymbol{\theta}) \right)^{-1} + A_k^T Q_k^{-1}(\boldsymbol{\theta}) A_k \right)^{-1}$$

$$\hat{\boldsymbol{x}}_k(\boldsymbol{\theta}) = S_k^{-1}(\boldsymbol{\theta}) \left( \left( F_k S_{k-1}(\boldsymbol{\theta}) F_k^T + R_k(\boldsymbol{\theta}) \right)^{-1} B_k \hat{\boldsymbol{x}}_{k-1}(\boldsymbol{\theta}) + A_k^T Q_k^{-1}(\boldsymbol{\theta}) \boldsymbol{y}_k \right)$$

$$p(\boldsymbol{\theta} | \boldsymbol{y}_{1:k}) \propto p(\boldsymbol{\theta} | \boldsymbol{y}_{1:k-1}) \sqrt{\frac{|S_k(\boldsymbol{\theta})|}{|Q_k(\boldsymbol{\theta})| |F_k S_{k-1}(\boldsymbol{\theta}) F_k^T + R_k(\boldsymbol{\theta})|}} \\ \times e^{-\frac{1}{2} \left[ (\boldsymbol{y}_k - A_k \hat{\boldsymbol{x}}_k(\boldsymbol{\theta}))^T Q_k^{-1}(\boldsymbol{\theta}) (\boldsymbol{y}_k - A_k \hat{\boldsymbol{x}}_k(\boldsymbol{\theta})) + (\hat{\boldsymbol{x}}_k(\boldsymbol{\theta}) - B_k \hat{\boldsymbol{x}}_{k-1}(\boldsymbol{\theta}))^T (F_k S_{k-1}(\boldsymbol{\theta}) F_k^T + R_k(\boldsymbol{\theta}))^{-1} (\hat{\boldsymbol{x}}_k(\boldsymbol{\theta}) - B_k \hat{\boldsymbol{x}}_{k-1}(\boldsymbol{\theta})) \right]}$$

- 1<sup>er</sup> problème :  $p(\boldsymbol{\theta} | \boldsymbol{y}_{1:k})$  est non-analytique.
  - Mais converge très rapidement vers une gaussienne.
  - Et concerne des paramètres « auxiliaires ».
  - L'approximation de Laplace semble acceptable.

# Kalman + variance components (3/3)

- 2<sup>ème</sup> problème : il n'existe pas de forme fonctionnelle stable par la récurrence en  $\mathbf{S}_k(\boldsymbol{\theta})$  et  $\hat{\mathbf{x}}_k(\boldsymbol{\theta})$ .
- Solution possible : travailler sur des DL en  $\boldsymbol{\theta}$ 
  - A l'ordre 0 :  $\mathbf{S}_k(\boldsymbol{\theta}) \approx \mathbf{S}_k(\hat{\boldsymbol{\theta}}_k)$  et  $\hat{\mathbf{x}}_k(\boldsymbol{\theta}) \approx \hat{\mathbf{x}}_k(\hat{\boldsymbol{\theta}}_k)$   
(Autrement dit, on néglige la dépendance des paramètres déterministes et de leur variance en fonction des composantes de la variance.)
  - Gênant en cas de composantes de la variance mal déterminées
  - Il est envisageable d'aller jusqu'à l'ordre 2.
  - Reste à quantifier les erreurs d'approximation.
- Autres solutions possibles :
  - Itérations Kalman – VCE – Kalman – VCE – Kalman...
  - ???