

A dramatic sunset over the ocean. The sky is filled with large, dark clouds, with bright sunlight breaking through, creating a golden glow. The sun is low on the horizon, reflecting on the water. In the foreground, a wooden pier extends into the water. A small boat is visible on the water in the distance.

*Assimilation de  
données*

# Estimateur Optimal

## ou: Best Linear Unbiased Estimator (BLUE) Analyse aux moindres carrés Statistical Optimal Interpolation (OI)

Méthode d'analyse utilisée dans tous les grands centres de météo, océano, environnement, etc. sous des noms divers selon les méthodes de calcul (3DVar, 4DVar, filtre de Kalman LETKF, interpolation optimale, etc.)

Résoud les limitations des algorithmes précédents

- Equation générale du BLUE
- Illustration en zéro-D, 1D, 2D
- Calcul par sélection locale d'observations (LETKF)
- Calcul variationnel (3DVar, 4DVar)

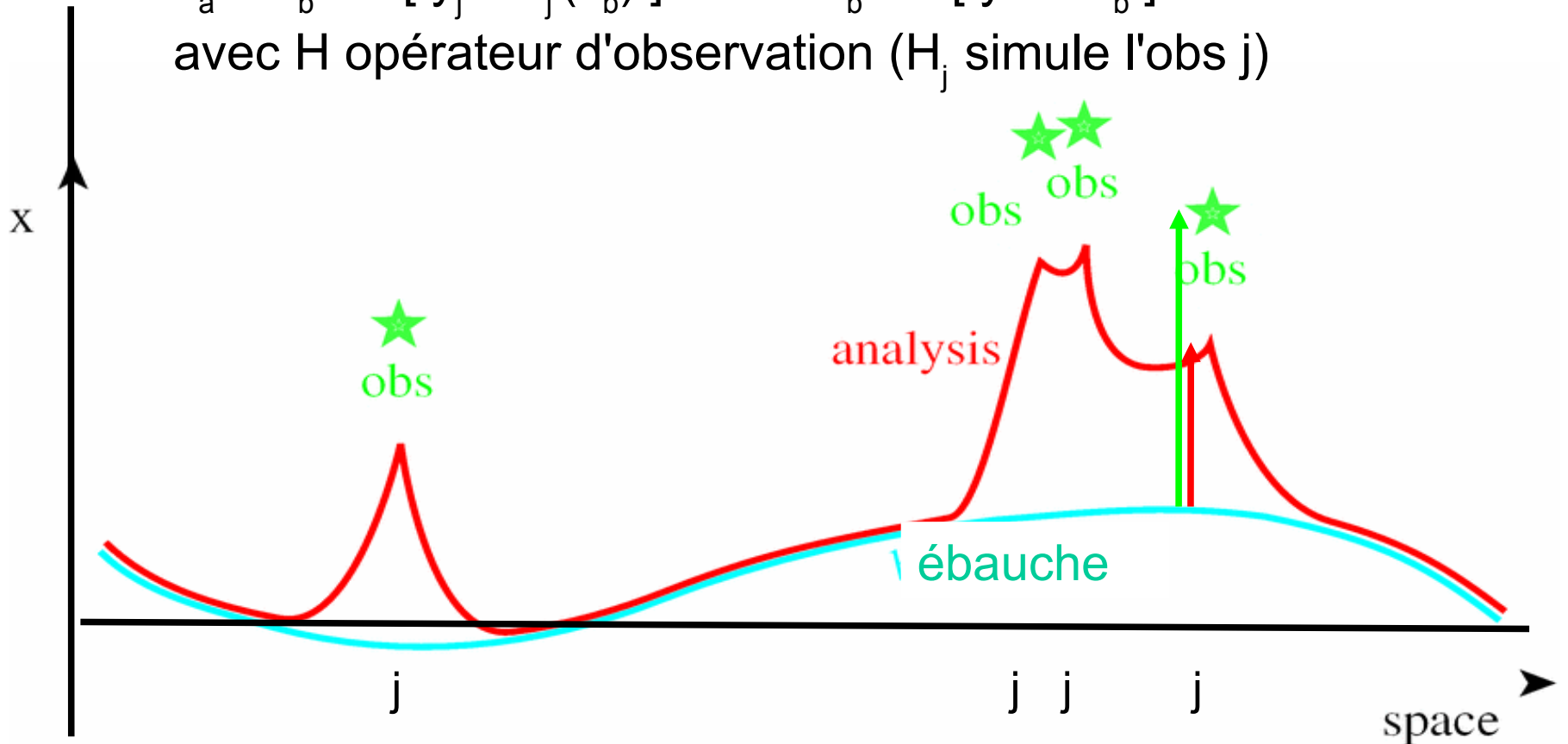
## Pb2: obs qui ne sont pas des variables du modèle

Utiliser un **opérateur d'observation**:

$$x_a = x_b + F[ y_j - x_b(j) ] \quad \text{devient}$$

$$x_a = x_b + F[ y_j - H_j(x_b) ] = x_b + F[ y - Hx_b ]$$

avec H opérateur d'observation ( $H_j$  simule l'obs j)

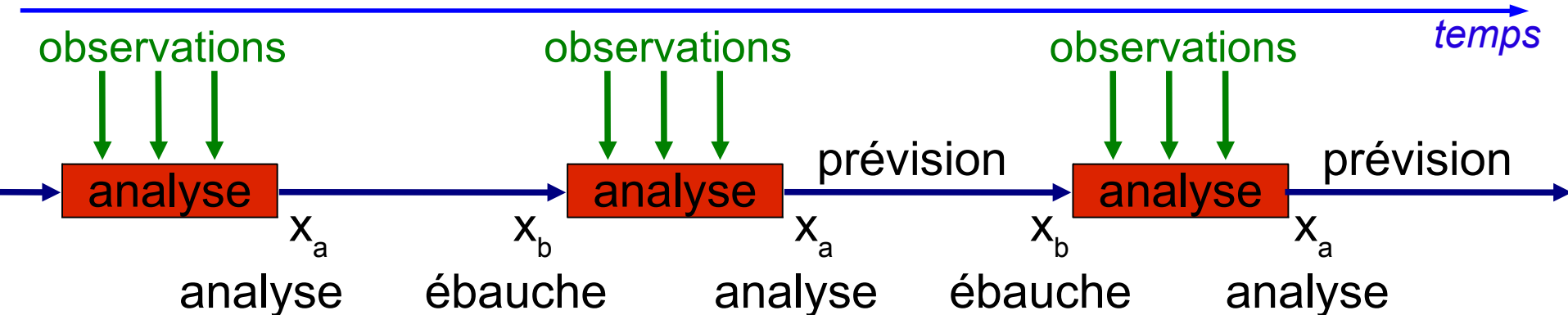
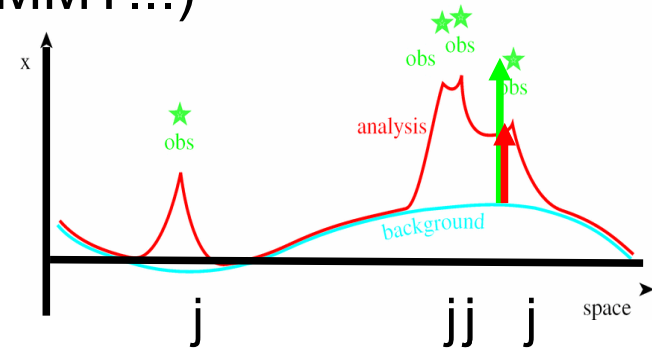


# Pb1: comment combiner les obs avec une ébauche

$x_a = F[ y_j ]$  devient  $x_a = x_b + F[ y_j - x_b(j) ]$   
avec  $x_b(j)$  ébauche de  $x$  interpolée au point d'obs  $j$

Où trouver une ébauche:

- climatologie
- réanalyse (NCEP, ERA...)
- analyse à basse résolution (Mercator, CEPMMT...)
- prévision récente, ou cycle d'assimilation



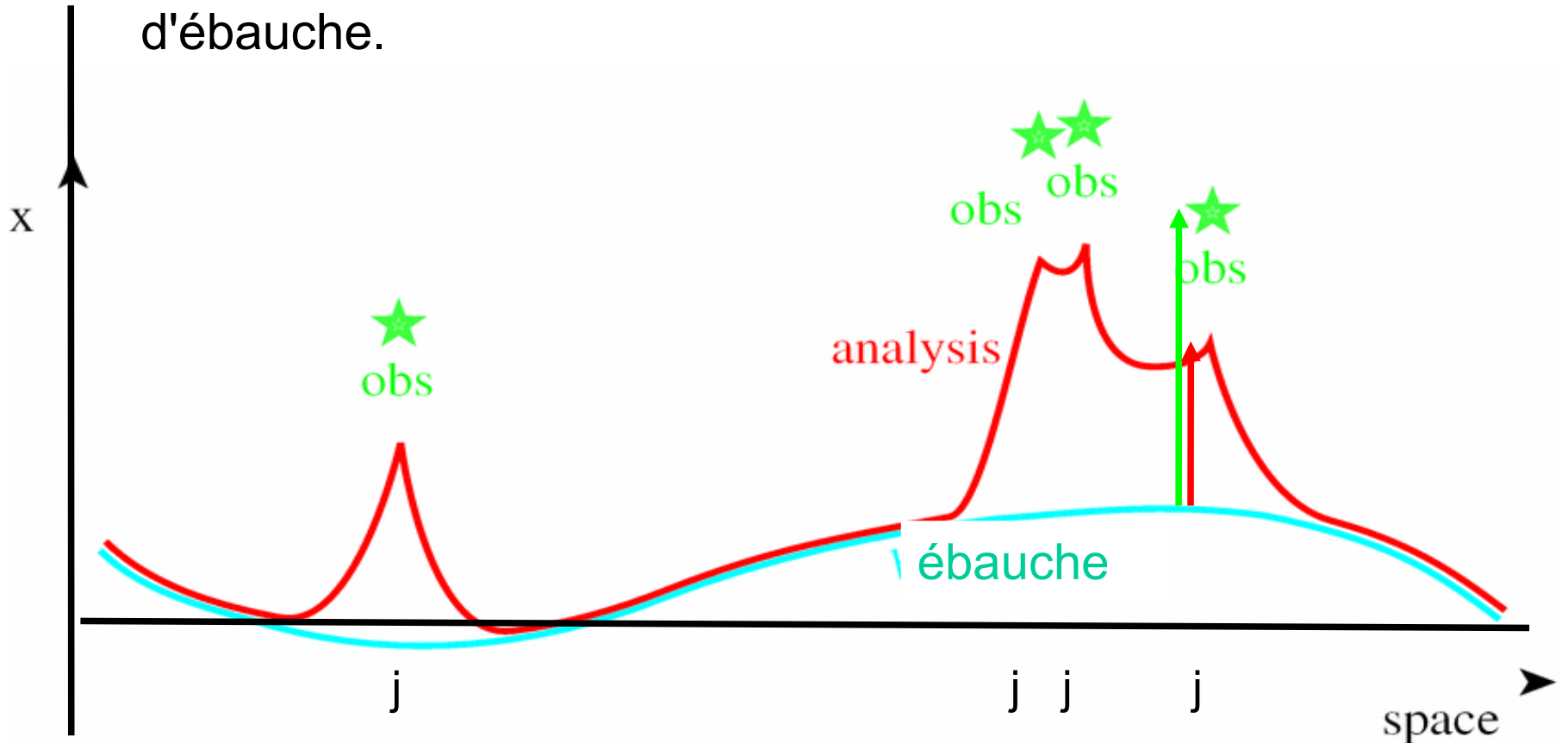
## Pb3: comment construire les fonctions de structure

$$x_a = x_b + F[ y - Hx_b ]$$

avec F qui dépend des covariances d'erreur de  $( y - H x_b )$

et comme  $( y - H x_b ) = ( y - Hx_t ) - H( x_b - x_t )$

on a besoin des covariances d'erreur d'observation et d'ébauche.



# le BLUE (Best Linear Unbiased Estimation)

Cadre mathématique général et pratique pour construire des systèmes d'analyse:

- les hypothèses utilisées ont un *sens physique précis* (pas seulement mathématique)
- nécessite de *modéliser les erreurs de prévision et d'observation de manière statistique*
- permet de fusionner des *sources d'information hétéroclites* (par ex.: ébauche + observations de tous types)
- inclut la notion d'*interpolation spatiale*
- inclut la notion de *compromis entre données contradictoires*
- permet de contrôler le réalisme physique de l'analyse (par ex.: *respecter des équations d'évolution*)
- permet d'analyser des paramètres physiques *non directement observés*
- *efficace en coût calcul*, surtout pour les grands modèles (dim x >1000)

# Notations pour l'équation de l'interpolation optimale

$x$  état du modèle

$y$  vecteur d'observations

$H$  opérateur d'observation:  $H(x)$  sont les valeurs d'obs simulées à partir de l'état du modèle  $x$ .

$x_b$  ébauche de l'analyse

$x_a$  analyse à calculer

$x_t$  réalité (inaccessible)

$B$  covariances d'erreurs d'ébauche =  $\text{cov}(x_b - x_t, x_b - x_t)$

$A$  covariances d'erreurs d'analyse =  $\text{cov}(x_a - x_t, x_a - x_t)$

$R$  covariances d'erreur d'observations =  $\text{cov}(y - H(x_t), y - H(x_t))$

en entrée:  $x_b, y, H, B, R$

en sortie:  $x_a, A$

# **Théorème fondamental du BLUE:** **calcul du gain optimal K**

*sous des hypothèses énoncées ci-après*

L'analyse optimale est  $\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}[\mathbf{x}_b])$

avec  $\mathbf{K} = \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1}$

*optimalité:*  $\mathbf{x}_a$  minimise la variance de l'erreur d'analyse  $\mathbf{x}_a - \mathbf{x}_t$   
et si les erreurs sont gaussiennes et H linéaire, alors  $\mathbf{x}_a$  est aussi  
l'analyse la plus probable (*maximum likelihood estimator*)

*Forme variationnelle:*  $x_a$  calculable par  $\mathbf{x}_a = \mathbf{Arg} \min (\mathbf{J})$

avec  $\mathbf{J}(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{y} - \mathbf{H}[\mathbf{x}])^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{H}[\mathbf{x}])$

Les covariances d'erreur d'analyse sont

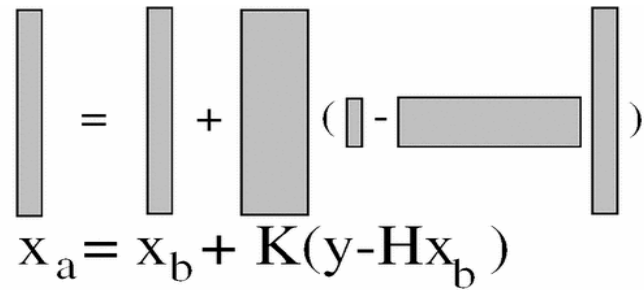
$$\mathbf{A} = (\mathbf{I} - \mathbf{K} \mathbf{H}) \mathbf{B} = 2 (\mathbf{J}'')^{-1}$$



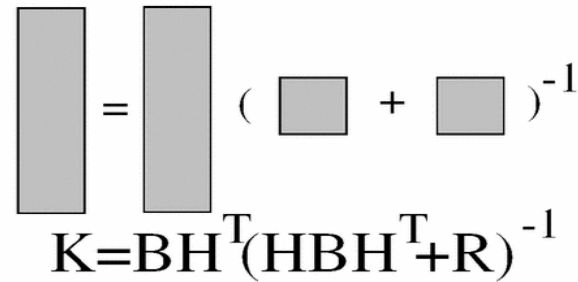
# Remarques sur le BLUE: dimensions des matrices

lorsque  $\dim y \ll \dim x$

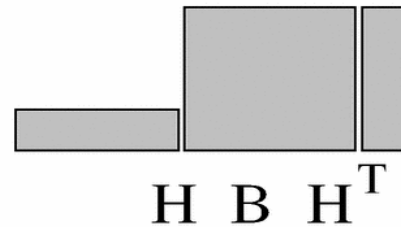
càd  $n_{\text{obs}} \ll \text{taille du modèle}$



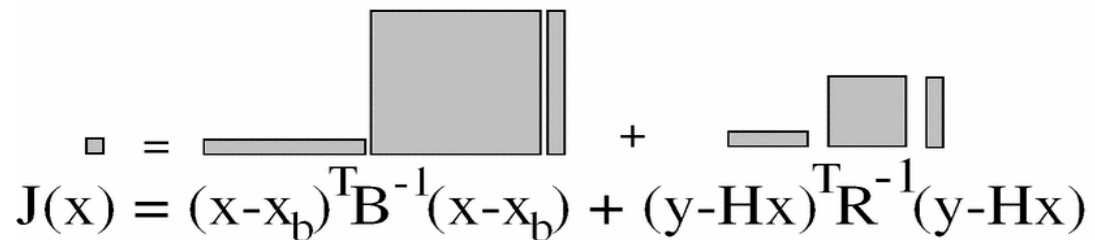
$$x_a = x_b + K(y - Hx_b)$$



$$K = BH^T (HBH^T + R)^{-1}$$



$$HBH^T$$



$$J(x) = (x - x_b)^T B^{-1} (x - x_b) + (y - Hx)^T R^{-1} (y - Hx)$$

## Hypothèses utilisées dans le théorème du BLUE:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K} ( \mathbf{y} - \mathbf{H}[\mathbf{x}_b] ) \quad \text{avec} \quad \mathbf{K} = \mathbf{B} \mathbf{H}^T ( \mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R} )^{-1}$$

- Les biais d'erreur d'ébauche  $\mathbf{x}_b - \mathbf{x}_t$  et d'observation  $\mathbf{y} - \mathbf{H}(\mathbf{x}_t)$  sont nuls (*càd: toutes les erreurs sont, en moyenne, nulles*)
- Les erreurs d'observation ne sont pas corrélées aux erreurs d'ébauche (*càd: ne pas utiliser l'ébauche pour préparer les obs*)
- dans  $\mathbf{K}$ ,  $\mathbf{H}$  est linéarisé par rapport aux variations de  $\mathbf{x}$  :

$$\mathbf{H}(\mathbf{x}) \approx \mathbf{H}(\mathbf{x}_b) + \mathbf{H}' (\mathbf{x} - \mathbf{x}_b) \quad \text{avec} \quad \mathbf{H}' \text{ opérateur linéaire}$$

- $\mathbf{B}$  et  $\mathbf{R}$  sont connus (pas exactement, mais on peut les approximer)
- moins les hypothèses sont vraies, et moins l'analyse est optimale (*càd. loin de la réalité*)

# Remarques sur les équations du BLUE: matrice K

C'est une interpolation avec les poids K:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K} (\mathbf{y} - \mathbf{H}[\mathbf{x}_b]) = (\mathbf{I} - \mathbf{K}\mathbf{H})[\mathbf{x}_b] + \mathbf{K}\mathbf{y}$$

Le *gain de Kalman*  $\mathbf{K} = \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1}$  :

- ne dépend que de H et des covariances d'erreur (d'ébauche et d'obs)
- ne dépend pas des valeurs d'obs
- va de l'espace des obs vers l'espace du modèle: K convertit les *innovations*  $(\mathbf{y} - \mathbf{H}[\mathbf{x}_b])$  en *incréments*  $\mathbf{x}_a - \mathbf{x}_b$

## Calcul algébrique direct de K dans l'analyse BLUE:

On cherche  $x_a = x_b + K (y - H[x_b])$  avec  $K = B H^T (H B H^T + R)^{-1}$

on peut réécrire:

l'innovation :  $d = y - H[x_b] = \text{obs} - \text{ébauche}$

l'incrément :  $x_a - x_b = B H^T w = \text{analyse} - \text{ébauche}$

Il faut donc calculer  $w = (H B H^T + R)^{-1} d$

c'est à dire **résoudre un système linéaire d'équations**:

$$(H B H^T + R) w = d$$

- dimension  $\text{dim}_y \times \text{dim}_y$ , coût numérique en  $(\text{dim}_y)^2$  : c'est cher !
- $H B H^T$  : covariances d'erreurs de  $x_b$  entre les points observés
- $R$  : covariances d'erreurs d'observations

Autrefois on calculait le BLUE ainsi ('interpolation optimale'), en modélisant directement  $H B H^T$ ,  $R$  et  $B H^T$ , et en supposant que  $H B H^T + R$  est creuse.

## Calcul variationnel de $x_a$ avec les équations du BLUE

forme matricielle:  $x_a = x_b + K (y - H[x_b]) = (I-KH)[x_b] + Ky$

forme variationnelle:  $J(x) = (x - x_b)^T B^{-1} (x - x_b) + (y - H[x])^T R^{-1} (y - H[x])$

Les 2 sont mathématiquement équivalents:  $J'(x_a)=0$

Leur calcul exact coûte à peu près pareil

mais

- La forme matricielle est difficile à calculer si *dimx* est grand
- La forme variationnelle est plus pratique, elle permet d'approximer itérativement  $x_a$  avec une précision ajustable: technique de *minimisation* (cf plus loin)

Car il est inutile de résoudre les équations du BLUE à une précision supérieure à celle de H, B ou R.

# Illustration du BLUE avec 1 seule variable

- $x_a - x_b = K (y - H[x_b])$  avec  $K = B H^T (H B H^T + R)^{-1}$

Si on suppose  $\dim x = \dim y = 1$ :

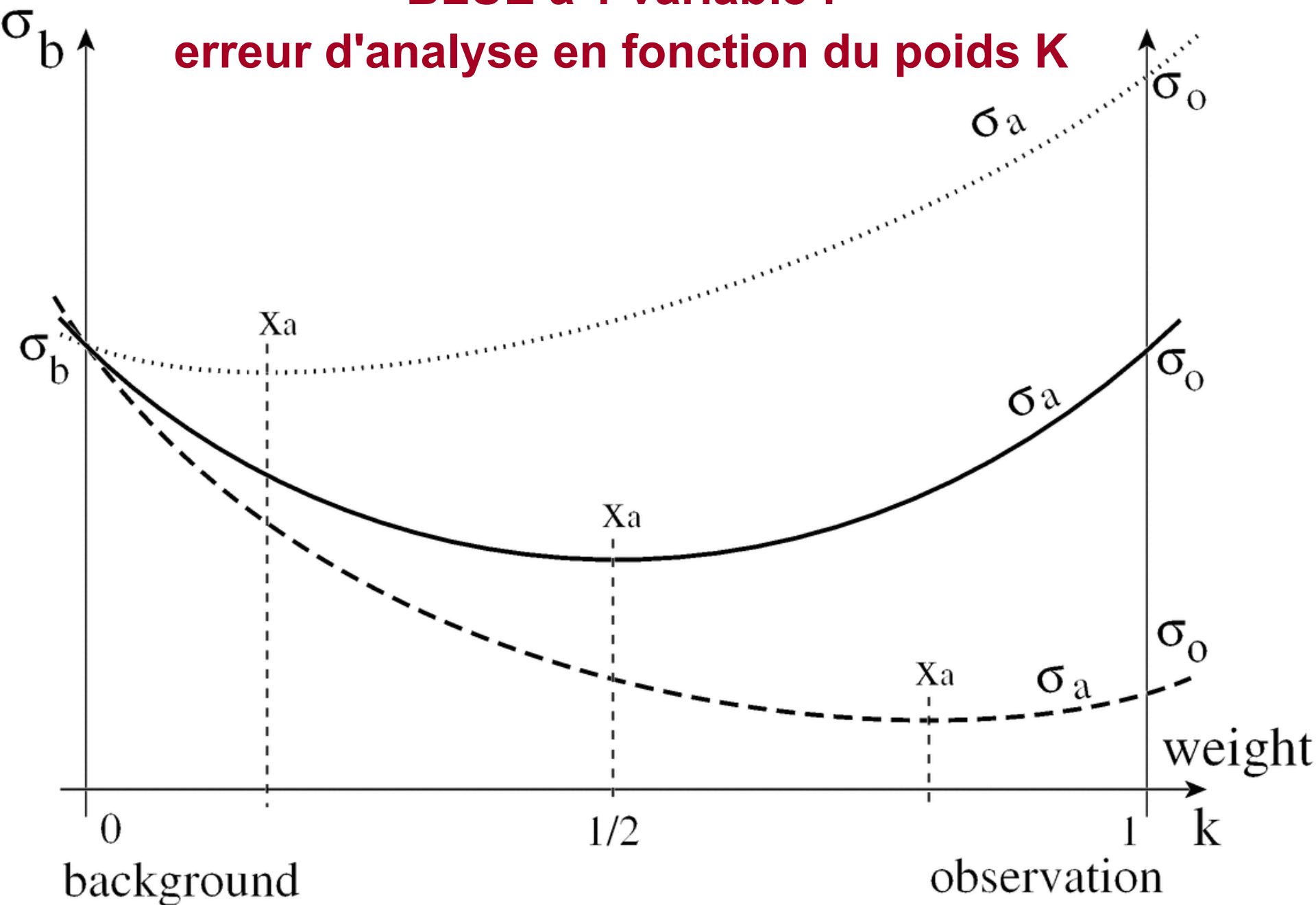
- $H = \text{Identité}$
- $B = \sigma_b^2$  variance d'erreur de l'ébauche
- $R = \sigma_o^2$  variance d'erreur d'observation

alors  $x_a = (1 - K) x_b + K y$  est une interpolation avec le poids

$$K = \sigma_b^2 / (\sigma_b^2 + \sigma_o^2) = (1 + \sigma_o^2 / \sigma_b^2)^{-1}$$

- Cette fonction  $K(\sigma_o/\sigma_b)$  décroît de 1 à zéro:
  - Si l'observation est meilleure que l'ébauche,  $\sigma_o < \sigma_b$  et  $x_a \approx y$
  - Si obs et ébauche sont de même qualité,  $\sigma_o = \sigma_b$  et  $x_a = (x_b + y)/2$
  - Si l'obs est mauvaise, on ne l'utilise pas,  $K = 0$  et  $x_a = x_b$

**BLUE à 1 variable :**  
**erreur d'analyse en fonction du poids K**



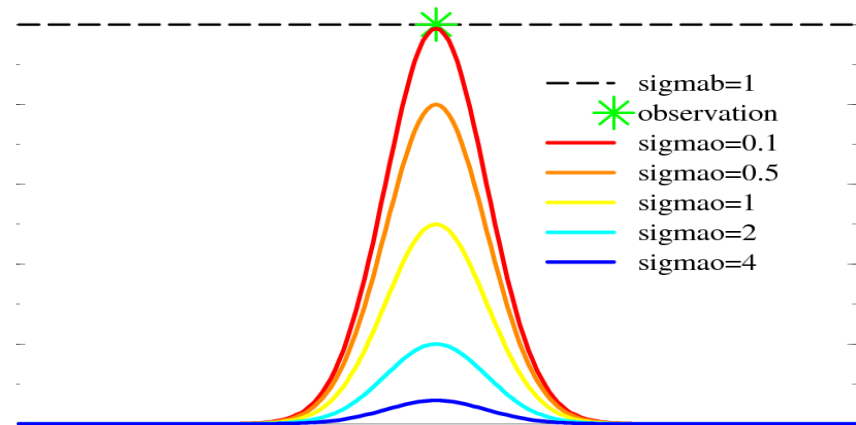
# Illustration du BLUE en 1-D

$x$  est un ensemble de valeurs sur un segment de droite  
 $y$  sont des observations ponctuelles (étoiles)

$x_b$  ébauche = 0 c'est à dire  $x_a = x_a - x_b$  : l'analyse est égale à l'incrément.

on illustre l'analyse avec différents choix de B, de R et de réseau d'obs:

- influence des variances de B
- des corrélations de B
- des corrélations de R...





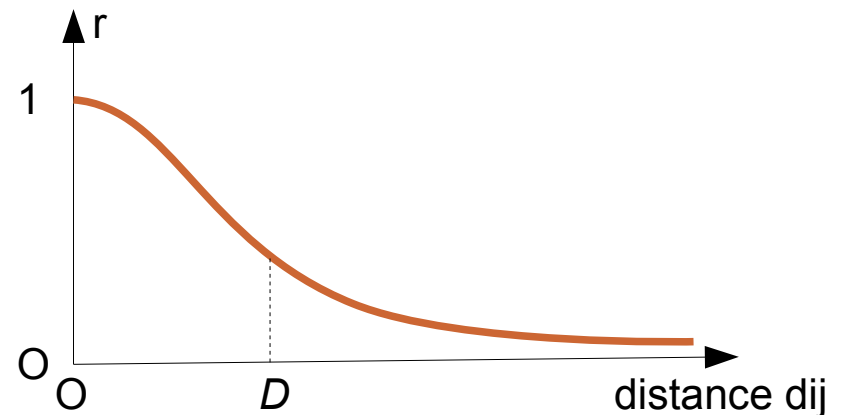
# Construction de B: modèle d'erreur gaussien

Les exemples qui suivent font l'hypothèse d'un modèle simple de covariances d'erreurs B:

- entre tout couple de points d'abscisses  $i$  et  $j$ ,

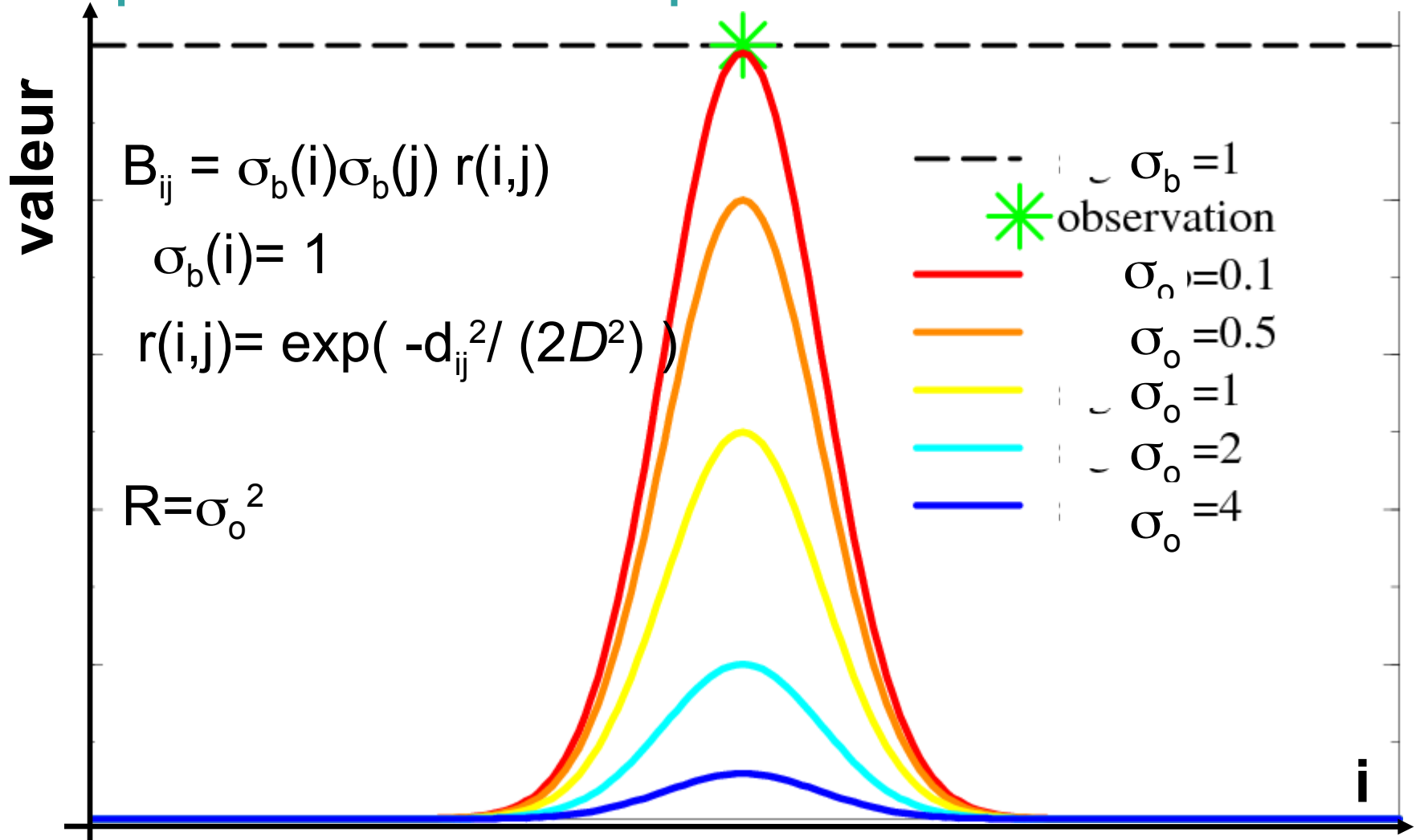
$$B_{ij} = \sigma_b(i)\sigma_b(j) r(i,j)$$

- $\sigma_b(i)$  = un champ d'écart-types très simple
- $r(i,j) = \exp(-d_{ij}^2 / (2D^2))$  modèle de corrélation gaussien de distance caractéristique  $D$



# Analyse BLUE 1D (B Gaussien, ébauche nulle, 1obs)

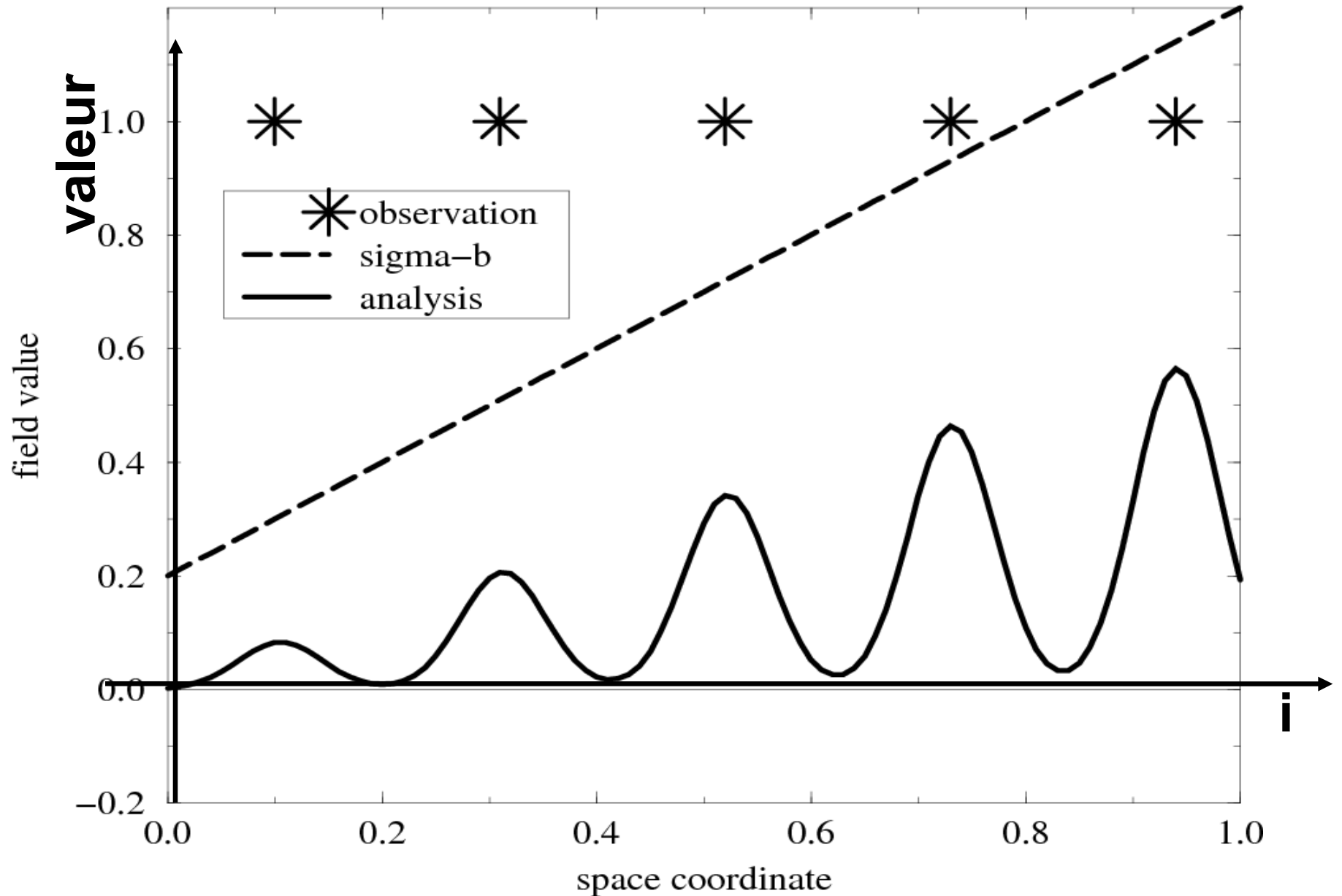
l'amplitude de l'incrément dépend de la variance d'erreur d'obs



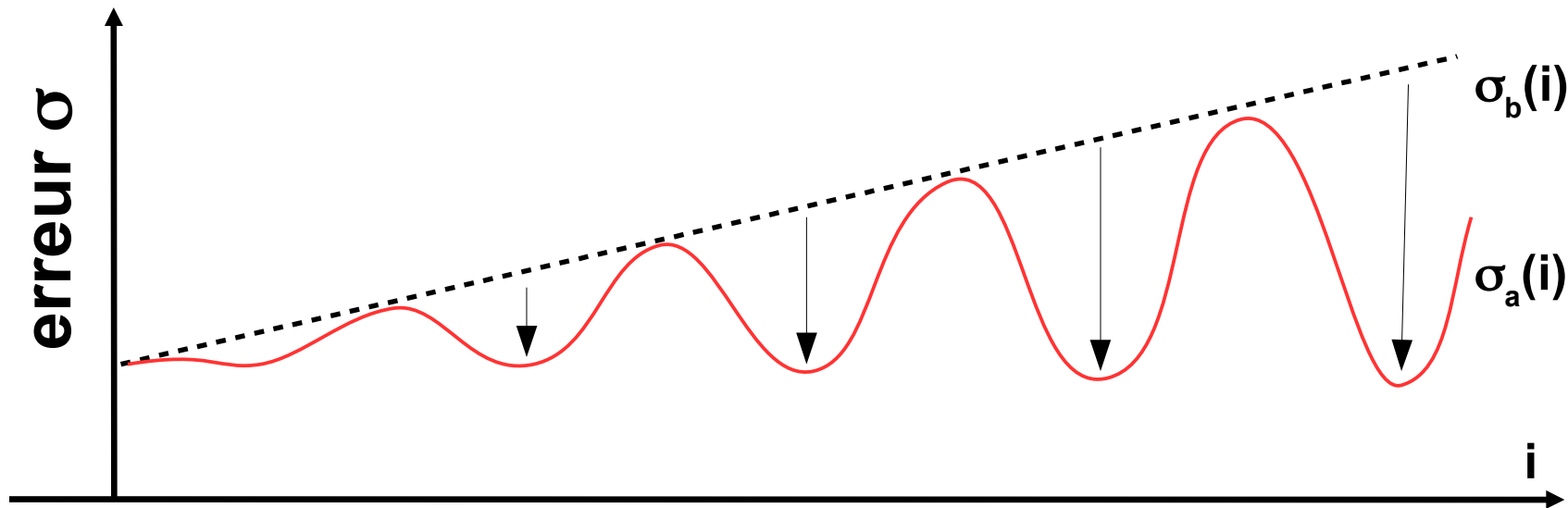
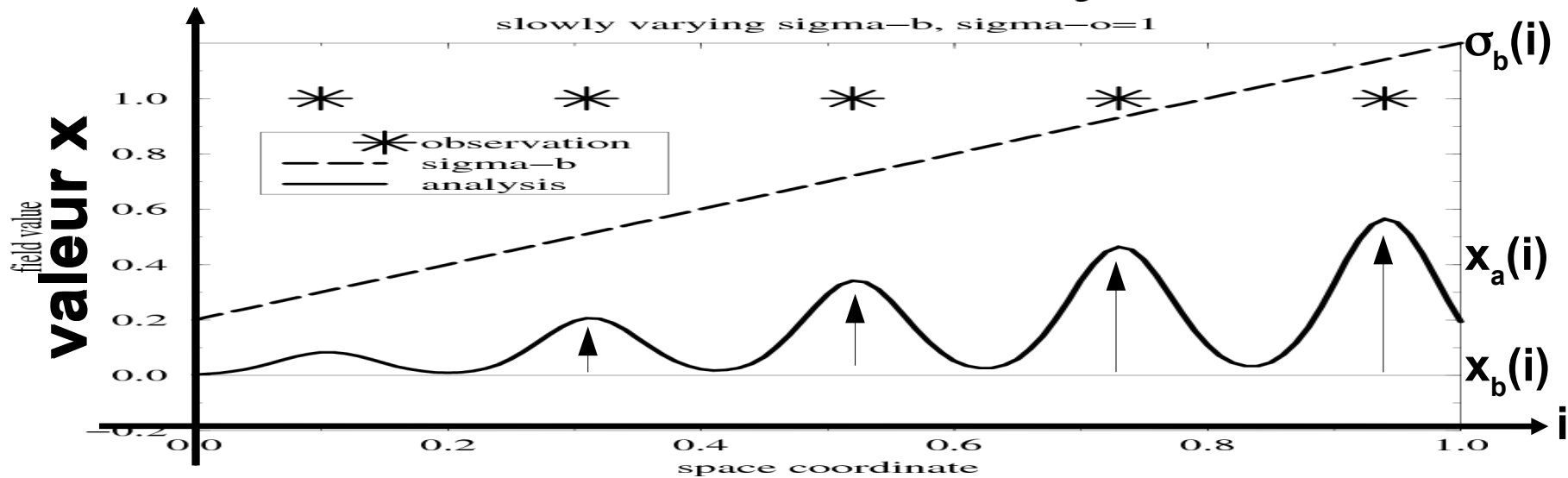
$$x_a - x_b = B H^T (H B H^T + R)^{-1} (y - H[x_b])$$

# Analyse BLUE 1D : variation de l'écart-type de B

slowly varying sigma-b, sigma-o=1



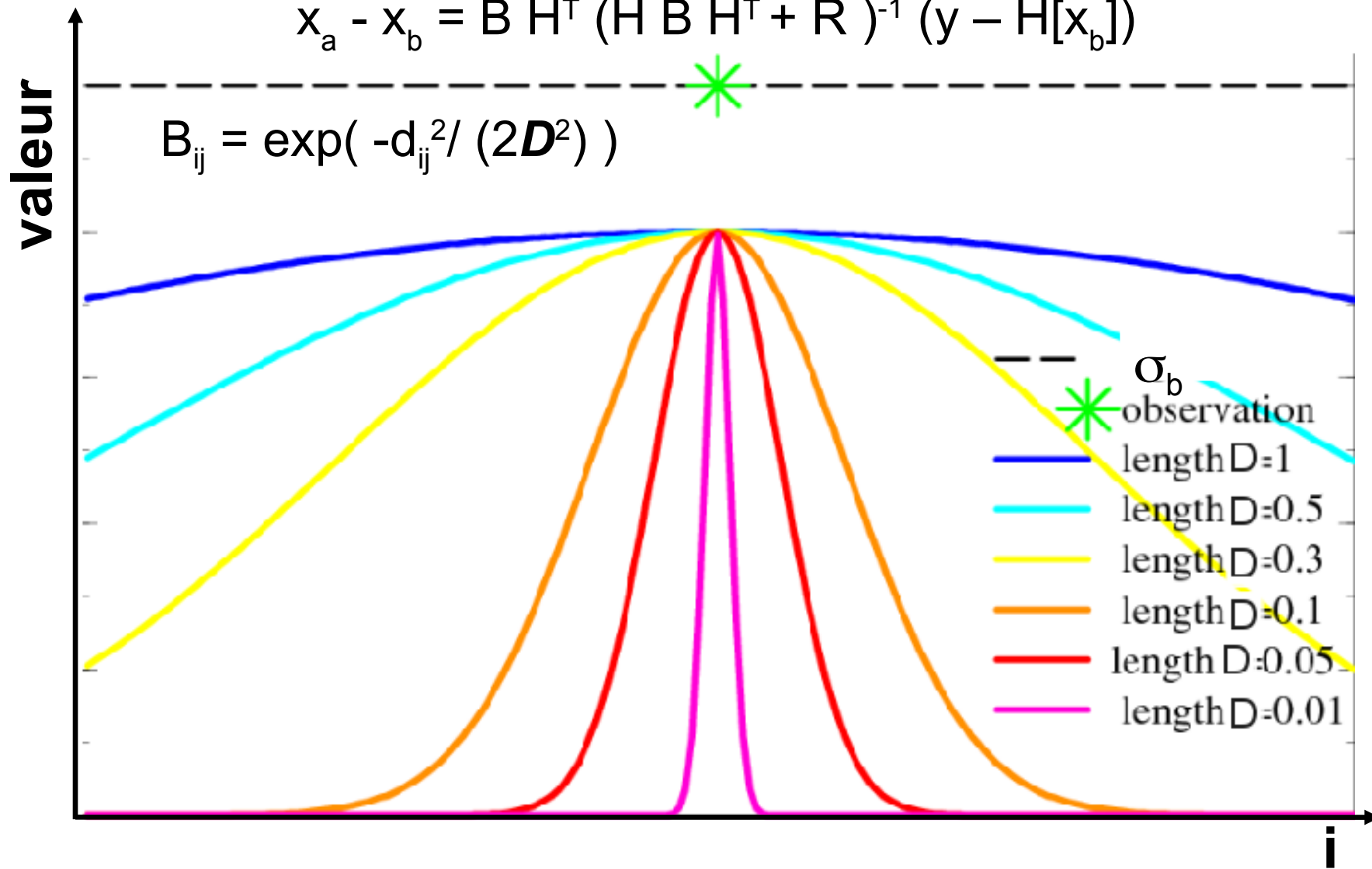
# Analyse BLUE 1D : qualité de l'analyse



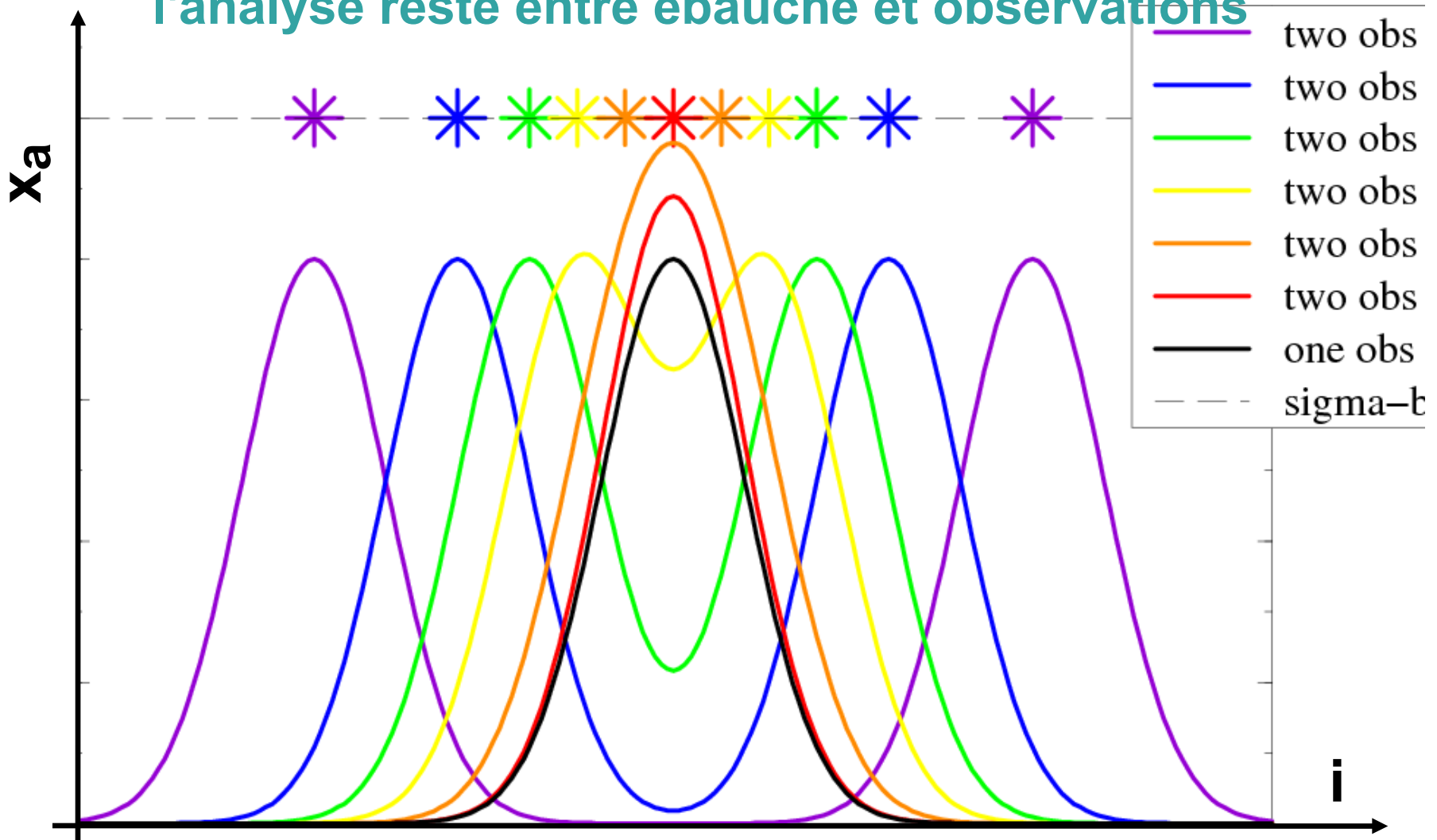
# Analyse BLUE 1D: variation de la distance de corrélation dans B

$$x_a - x_b = B H^T (H B H^T + R)^{-1} (y - H[x_b])$$

$$B_{ij} = \exp(-d_{ij}^2 / (2D^2))$$

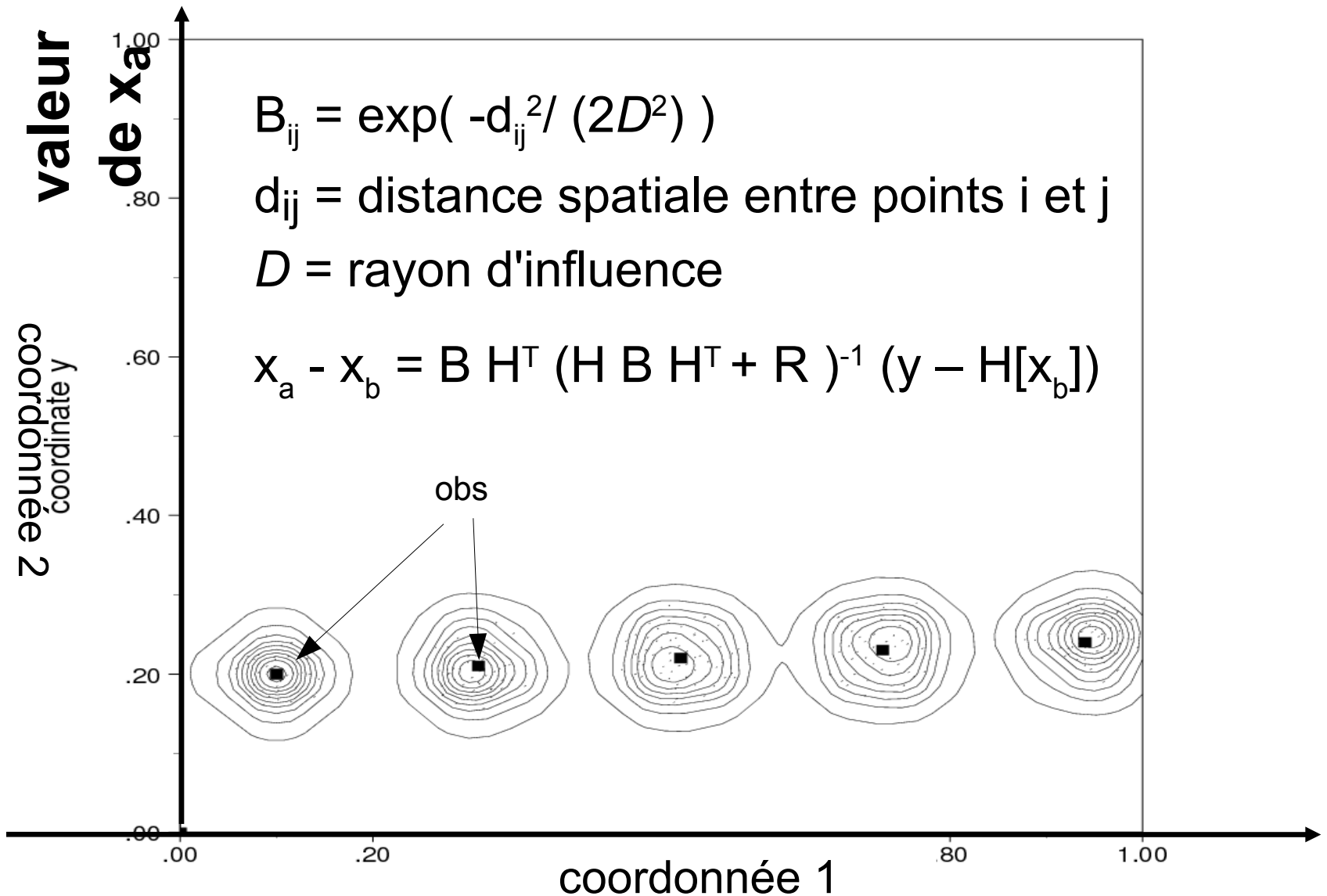


Analyse BLUE 1D (avec 2 obs):  
 2 obs ont plus de poids qu'une seule,  
 l'analyse reste entre ébauche et observations



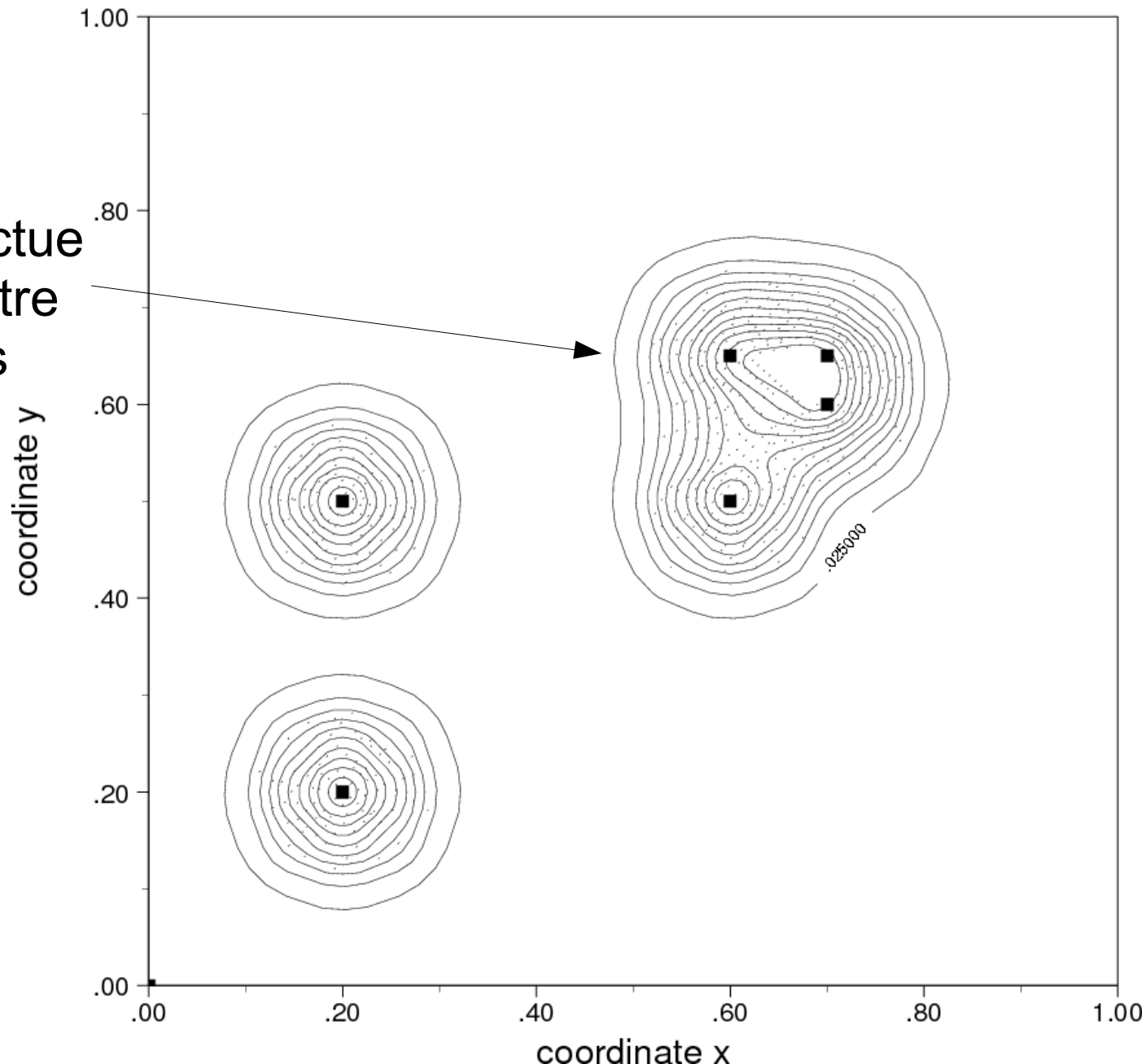
$$x_a - x_b = B H^T (H B H^T + R)^{-1} (y - H[x_b])$$

# Analyse BLUE 2D : les incréments d'obs isolées sont des gaussiennes



# Analyse BLUE 2D à plusieurs obs

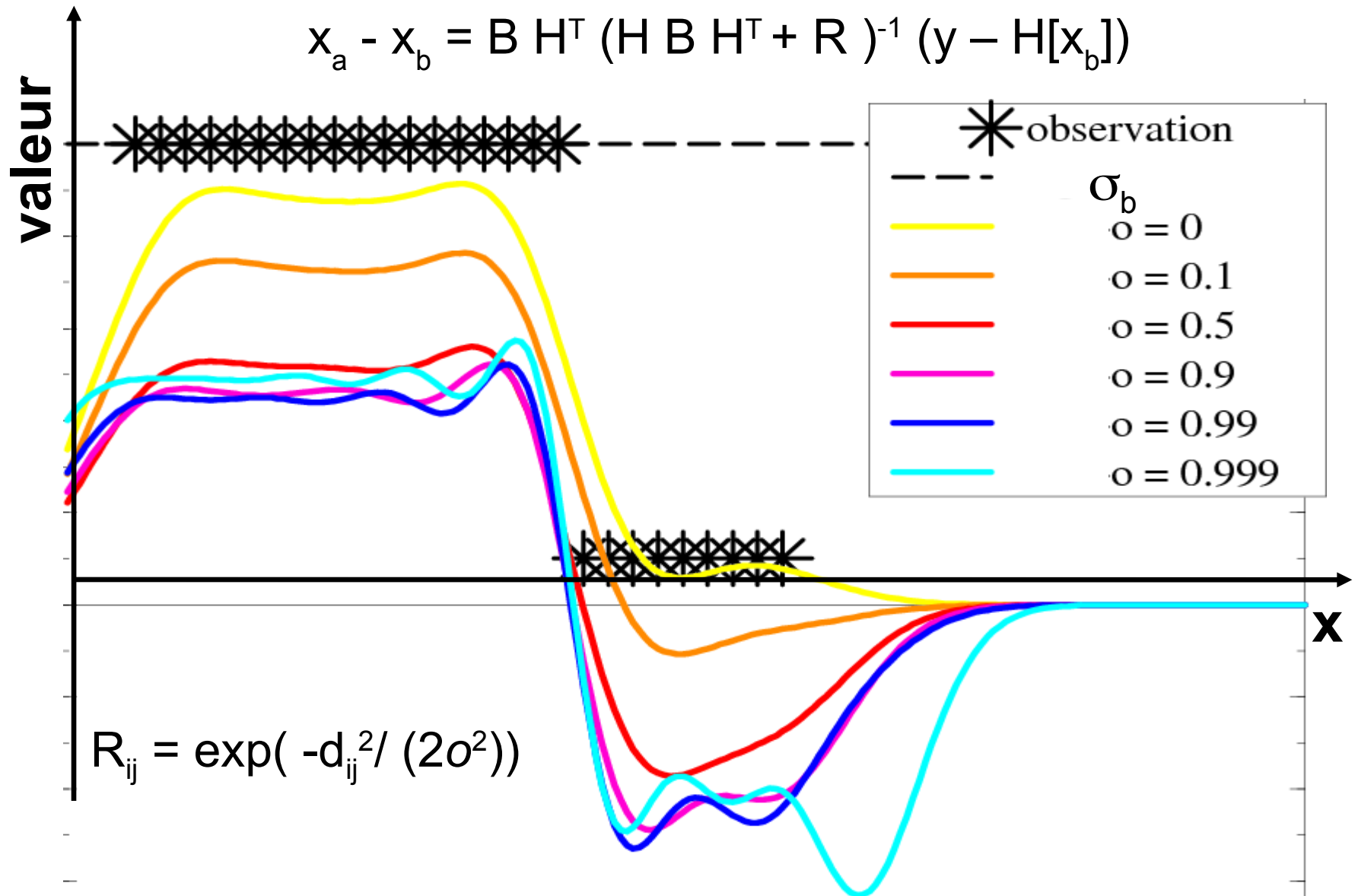
L'analyse effectue un lissage entre obs denses





Effet des corrélations d'erreur d'obs (matrice R):  
 l'analyse BLUE reproduit mieux les gradients entre obs,  
 mais moins bien la moyenne des obs.

$$x_a - x_b = B H^T (H B H^T + R)^{-1} (y - H[x_b])$$



# A retenir sur le BLUE:

- l'analyse BLUE est un **lissage des obs**
- on 'colle' plus aux obs si les écarts-types de B sont grands (par rapport à ceux de R)
- pour des obs isolées (et R diagonale), la forme de l'analyse est celle des corrélations de B: ce sont les **fonctions de structure**
- la structure de H modifie les fonctions de structure si les obs sont des intégrales (pas des points)
- le **lissage spatial est d'autant plus fort que les corrélations de B s'étendent loin**
- l'effet de corrélations d'observations (R) peut être complexe

## Résumé sur le BLUE

### données d'entrée:

$x_b$  : ébauche

B : ses covariances d'erreur  $B = \text{cov}(x_b - x_t)$

y : observations

R : ses covariances d'erreur  $R = \text{cov}(y - H[x_t])$

H : opérateur d'observation reliant les espaces de x et de y

### inconnues:

$x_a$  : analyse

A : ses covariances d'erreur  $A = \text{cov}(x_a - x_t)$

( $x_t$  est la réalité, inaccessible et aléatoire)

### THEOREME 1 - GAIN OPTIMAL

l'analyse de la forme  $x_a = x_b + K(y - H[x_b])$

minimise l'erreur quadratique  $\text{Tr}(A)$  si

$K = BH^T(HBH^T + R)^{-1}$

(sous certaines hypothèses)

et alors  $A = B - KHB$

### THEOREME 2 - FORME VARIATIONNELLE

l'analyse  $x_a$  ci-dessus est l'unique minimum de la fonction-coût:

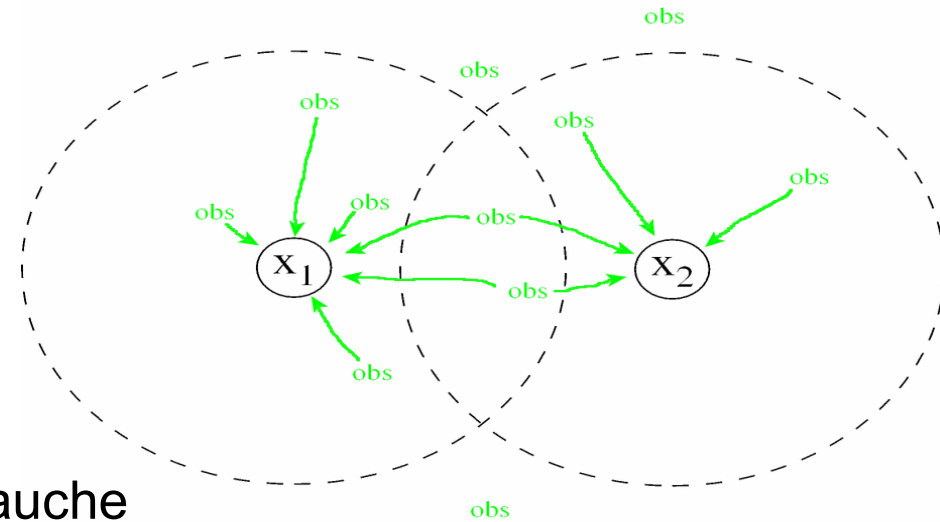
$J(x) = (x - x_b)^T B^{-1} (x - x_b) + (y - Hx)^T R^{-1} (y - Hx)$

et  $A = (1/2 \nabla^2 J)^{-1}$

$\nabla^2 J = J''$  est appelé hessienne de J, c'est une matrice  $\dim x \cdot \dim x$

$\nabla J$  est appelé gradient de J, c'est une fonction vectorielle de x

# Calcul approché du BLUE par 'sélection locale'



**innovation** :  $d = y - H[x_b]$  = obs - ébauche

**incrément** :  $x_a - x_b = B H^t w$  = analyse – ébauche

**système d'équations**:  $(H B H^t + R) w = d$

- en pratique on peut souvent supposer que
  - la  $i$ -ième ligne de  $B H^t$  n'est non-nulle que pour les  $p$  obs proches du  $i$ -ième point de  $x$  : **sélection locale des observations** par 'boîte'
  - les coeffs des matrices  $R$  et  $H B H^t$  sont faibles pour les autres obs
- coût numérique :  $\dim x$  systèmes de taille  $p \times p$ , coût  $\dim x \cdot p^2 \ll (\dim y)^2$
- vieille méthode qui a des inconvénients (sélection arbitraire, manque de cohérence spatiale). Encore utilisée dans le LETKF (Local Transform Ensemble Kalman Filter)

# Calcul approché du BLUE par méthode variationnelle (techniques 3DVar et 4DVar)

- $(x_b, y, H, B, R)$  définissent la **fonction coût**:

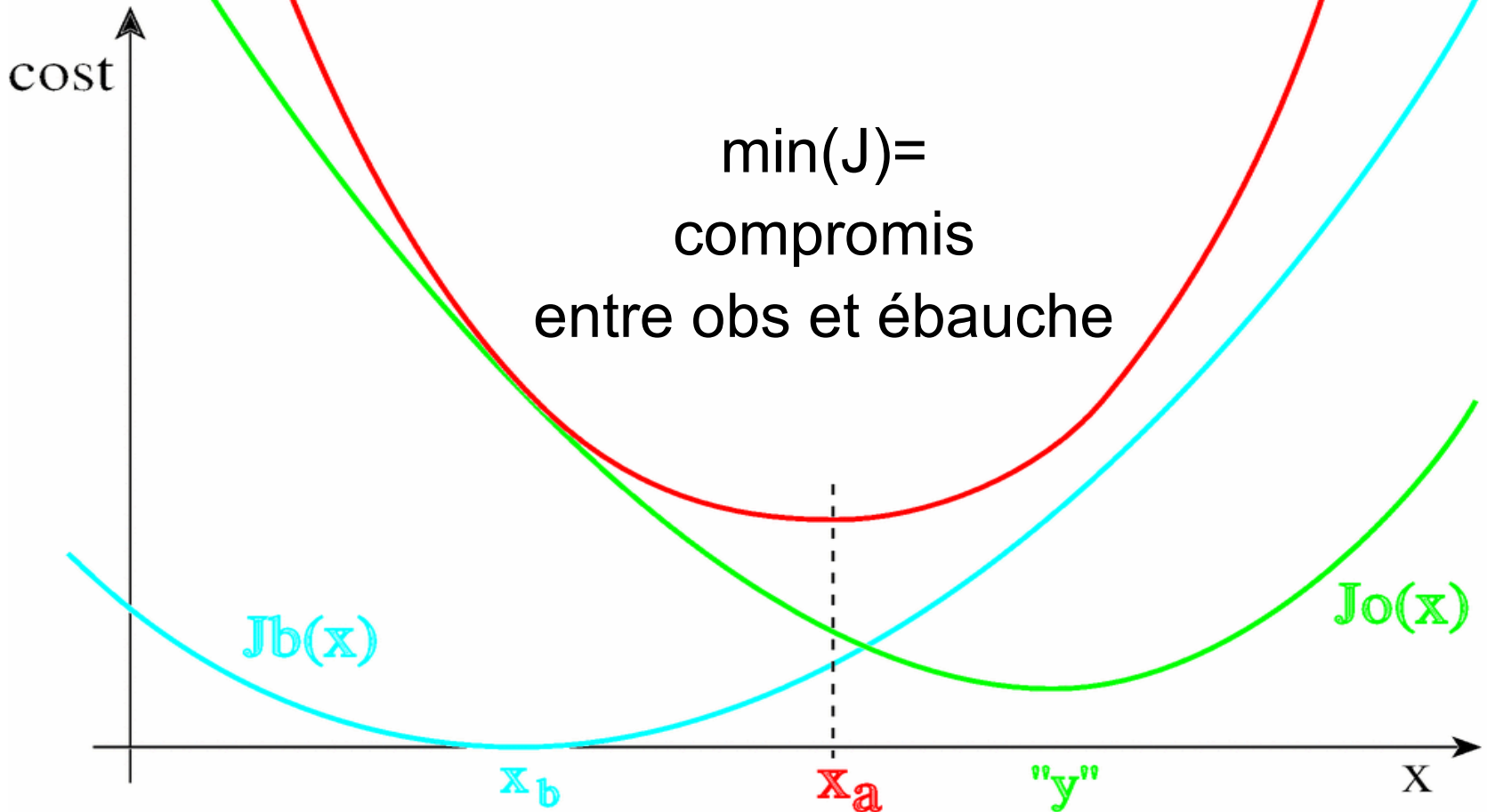
$$x \rightarrow \mathbf{J}(x) = (x - x_b)^T \mathbf{B}^{-1} (x - x_b) + (y - \mathbf{H}[x])^T \mathbf{R}^{-1} (y - \mathbf{H}[x])$$

- $x_a$  est le  $x$  qui **minimise**  $\mathbf{J}$ . On part d'une estimation initiale  $x_0 = x_b$
- répéter Niter fois:
  - calculer  $\mathbf{J}(x_i)$  et  $\mathbf{J}'(x_i)$
  - $x_{i+1} = x_i + a\mathbf{J}'(x_i)$  tel que  $\mathbf{J}(x_{i+1})$  soit minimal (a est calculé par un **minimiseur**: Gradient conjugué ou Quasi-Newton)
  - arrêter lorsque  $\mathbf{J}'(x_i)$  est assez petit: on prend  $x_a = x_{\text{Niter}}$
- L'approximation vient de ce que l'on arrête la minimisation avant d'avoir trouvé exactement  $x_a = \text{Arg min J}$
- En pratique on obtient de bons résultats avec  $\sim 100$  itérations
- Le rapport  $\|\mathbf{J}'(x_a)\| / \|\mathbf{J}'(x_b)\|$  indique de combien on s'est rapproché de  $\text{Arg min J}$
- $\mathbf{J}'(x) = 2 \mathbf{B}^{-1} (x - x_b) - 2 \mathbf{H}^T \mathbf{R}^{-1} (y - \mathbf{H}[x])$  est un vecteur

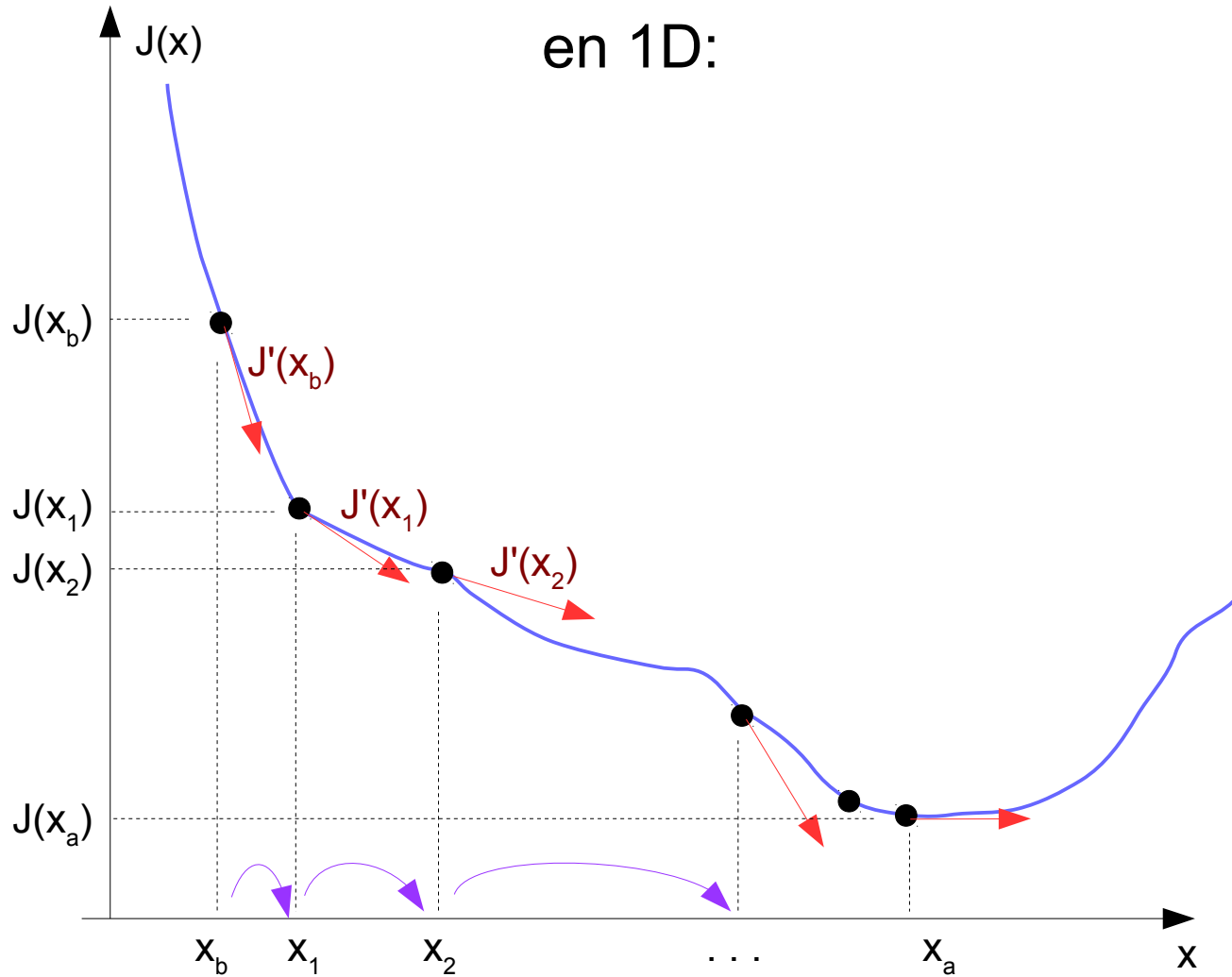
# Schématisation 0-D de

$$J(x) = \underbrace{(x - x_b)^T B^{-1} (x - x_b)}_{J_b(x)} + \underbrace{(y - H[x])^T R^{-1} (y - H[x])}_{J_o(x)}$$

$$J(x) = J_b(x) + J_o(x)$$



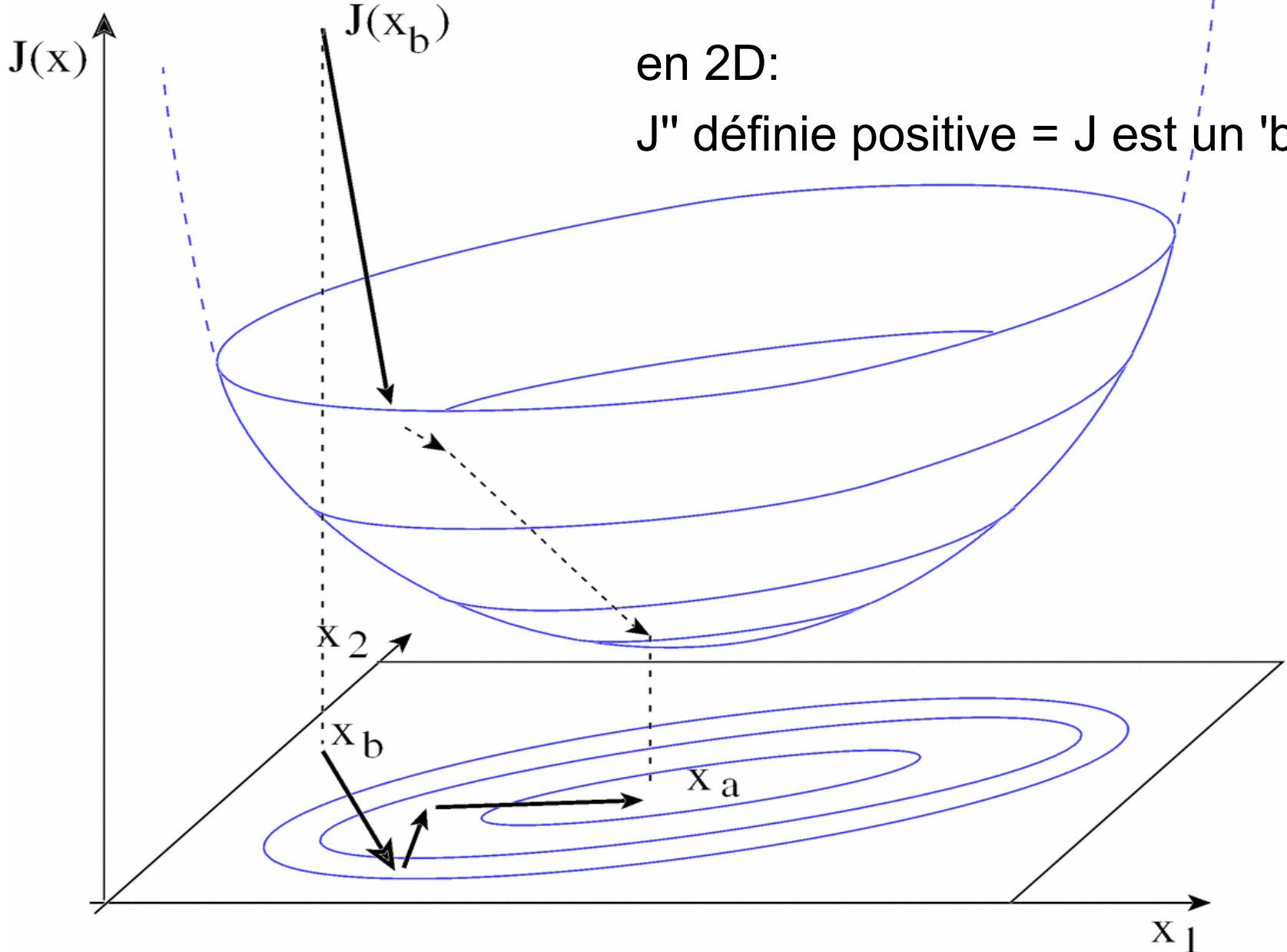
# Optimisation itérative d'une fonction $J: \mathbb{R}^n \rightarrow \mathbb{R}$ à l'aide d'un minimiseur



# Optimisation itérative d'une fonction $J: \mathbb{R}^n \rightarrow \mathbb{R}$ à l'aide d'un minimiseur

en 2D:

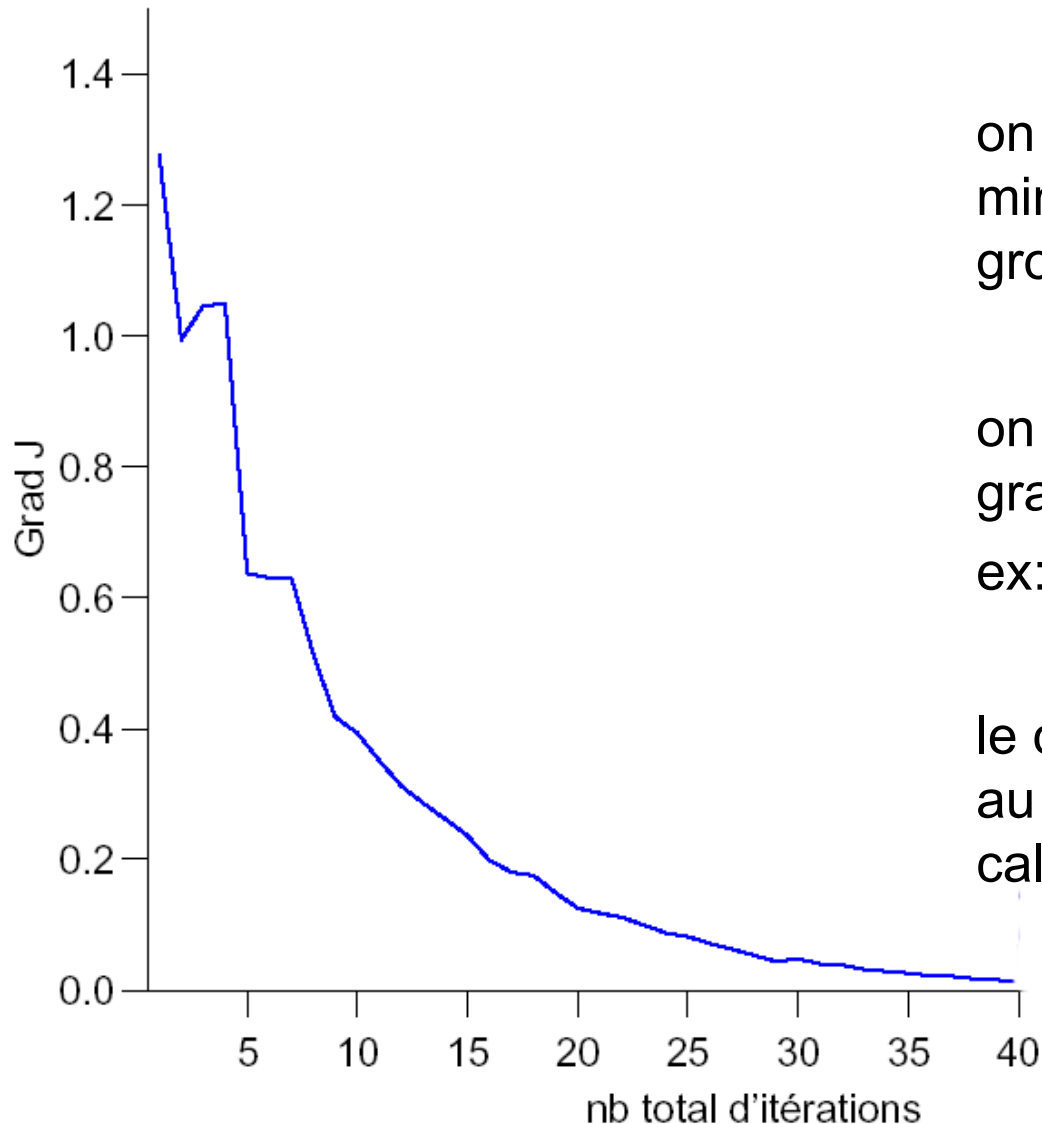
$J''$  définie positive =  $J$  est un 'bol'





# Exemple de minimisation ( $n=10^9$ )

évolution de la norme du gradient de J au cours d'une minimisation



on initialise le processus de minimisation avec une estimation grossière de  $x_a$  (ex:  $x_b$ )

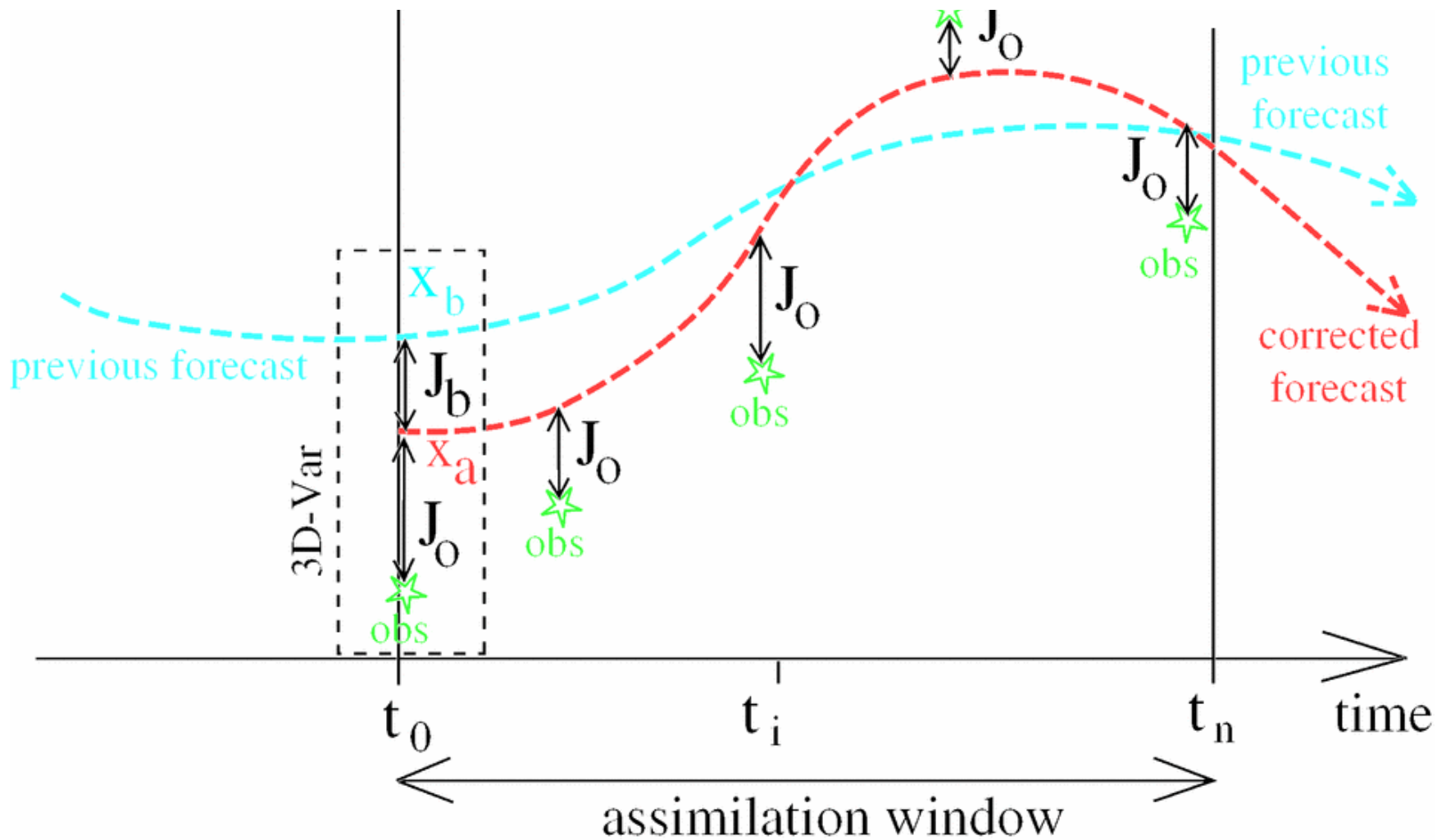
on arrête la minimisation lorsque le gradient de J est suffisamment petit, ex:  $\| J'(x) \| < \| J'(x_b) \| / 100$

le coût du calcul est proportionnel au nombre d'itérations (= nb de calculs de J et de J')

# Assimilation 4DVar: généralisation du BLUE à un historique d'observations, à l'aide d'un modèle de prévision M

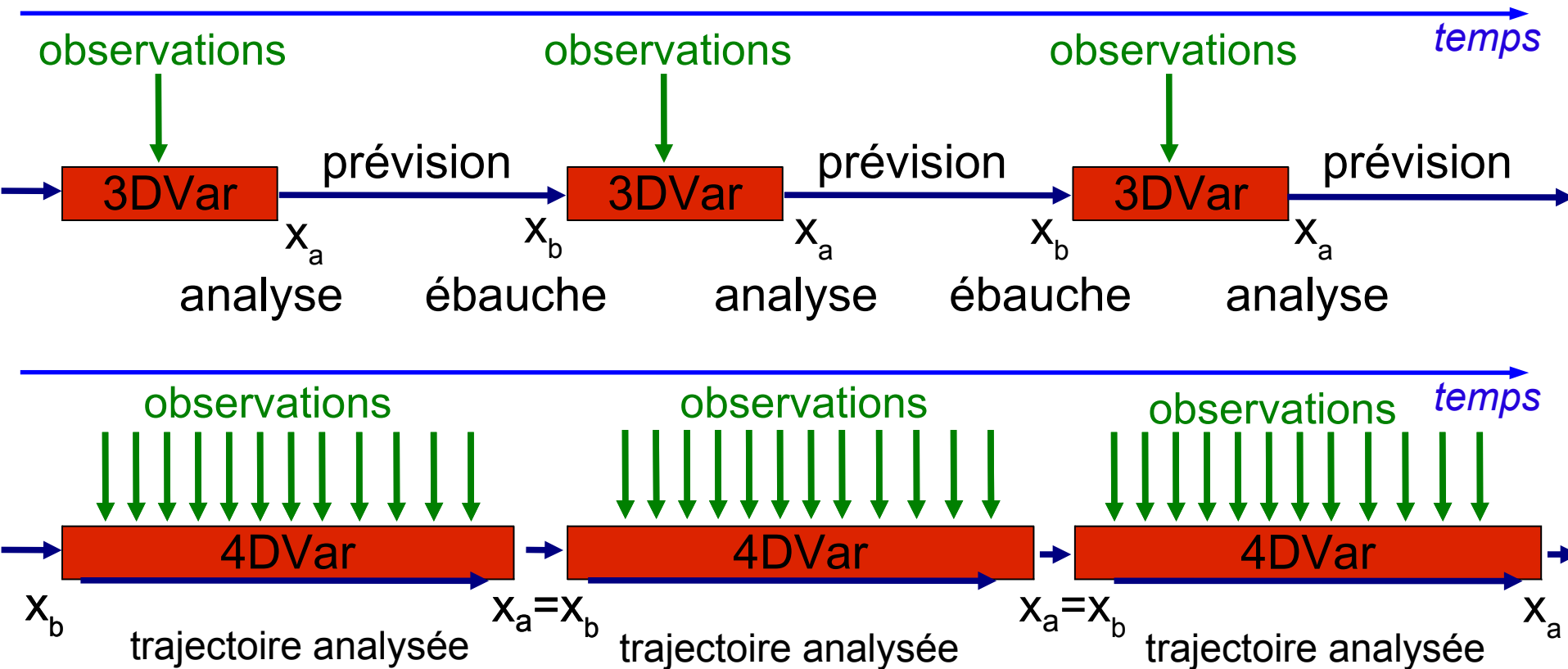
$$J(x) = (x - x_b)^T B^{-1} (x - x_b) + \sum_i (y_i - H_i[\mathbf{M}_i(x)])^T R_i^{-1} (y_i - H_i[\mathbf{M}_i(x)])$$

La variable de minimisation est l'état  $x$  du modèle au début de la fenêtre.



# Comparaison entre 3DVar et 4DVar

- Le 4DVar permet de traiter plus d'obs à la fois
- Il utilise un modèle numérique pour calculer K (pas seulement pour calculer  $x_b$ )
- Il produit des séquences continues d'analyses: des 'trajectoires'



## Assimilation variationnelle 4DVar

- La fonction coût du BLUE est valable à l'instant de l'analyse: c'est le 3DVar

$$x \rightarrow \mathbf{J}(x) = (x - x_b)^T \mathbf{B}^{-1} (x - x_b) + (y - \mathbf{H}[x])^T \mathbf{R}^{-1} (y - \mathbf{H}[x])$$

- Elle modifie  $x$  pour le "rapprocher" des observations à  $t=0$  sans trop s'éloigner de  $x_b$
- La fonction 4DVar "rapproche" une séquence de  $x$  d'une séquence d'observations sur une *fenêtre temporelle*  $t_k$ ,  $k=0\dots K$ :

$$x \rightarrow \mathbf{J}(x) = (x - x_b)^T \mathbf{B}^{-1} (x - x_b) + \sum_k (\mathbf{y}_k - \mathbf{H}_k \circ \mathbf{M}_k[x])^T \mathbf{R}_k^{-1} (\mathbf{y}_k - \mathbf{H}_k \circ \mathbf{M}_k[x])$$

- chaque trajectoire de  $x$  est définie par le modèle de prévision  $\mathbf{M}_k$ :

$$x=x(t_0) \rightarrow \mathbf{x}(t_k)=\mathbf{M}_k(x)$$

- $x$  et  $x_b$  restent définis au début de la fenêtre, car (en général) une prévision se contrôle avec son état initial
- Le 4DVar se minimise comme un 3DVar, mais coûte plus cher: ~200 prévisions pour calculer une analyse
- 4DVar très performant pour les systèmes chaotiques (atmosphère, océan), et les obs distribuées dans le temps (satellites)

A dramatic sunset over the ocean. The sky is filled with large, dark clouds, with bright sunlight breaking through, creating a golden glow and long rays of light. The sun is low on the horizon, reflecting on the water. In the foreground, a wooden pier extends into the water. A small boat is visible on the water in the distance.

*Merci pour votre  
attention*