
Asimilación de datos

Manuel Pulido

*(1) Department of Physics, Universidad Nacional del Nordeste,
Corrientes, Argentina*

(2) Data Assimilation Research Centre. University of Reading. UK

<http://pulidom.github.io>

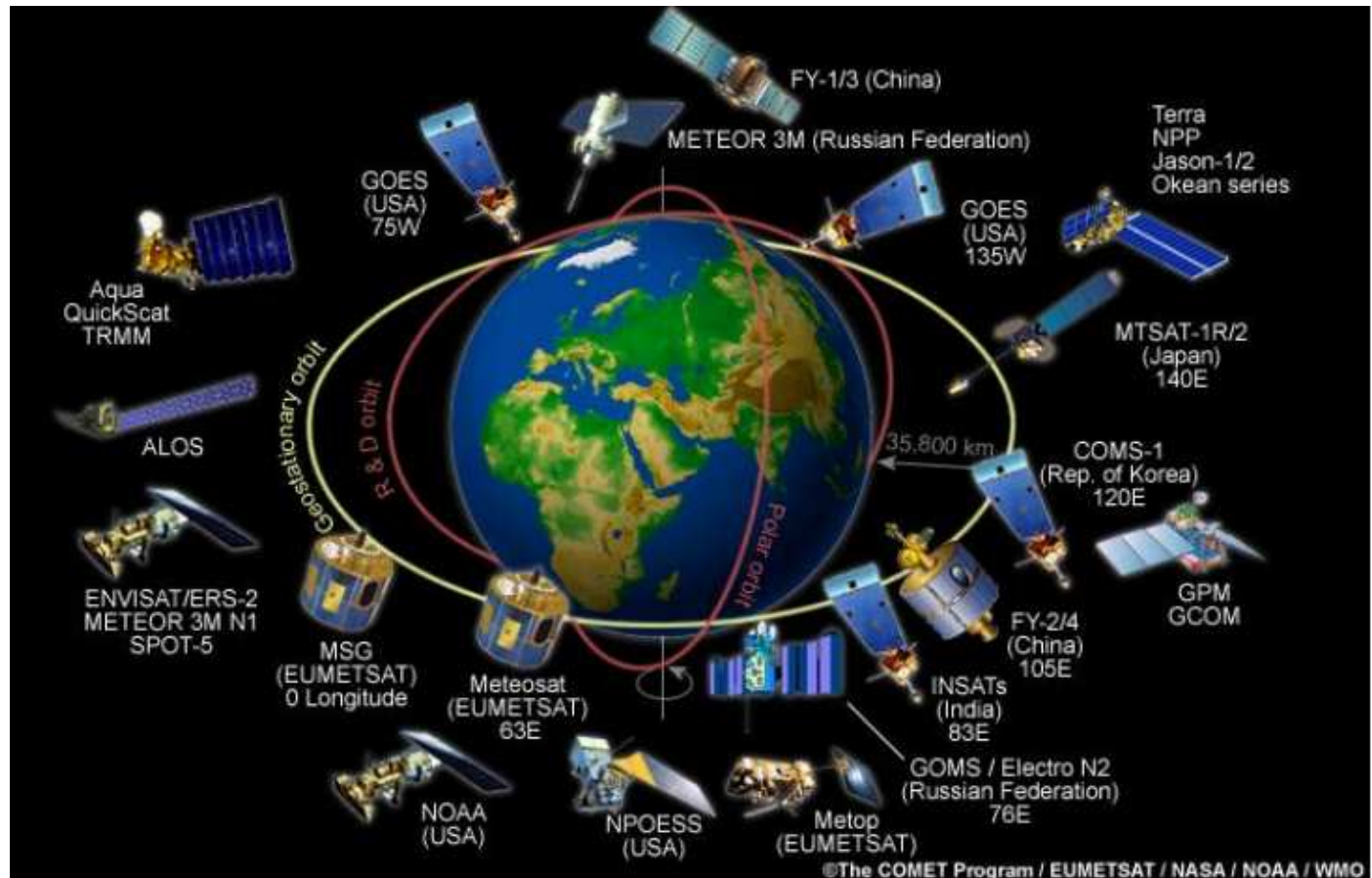
Organización

.

1. Introducción y motivación
2. Inferencia en un modelo de Markov escondido
3. Ejemplos 1D
4. Interpolación óptima
5. Asimilacion variacional. 3DVar
6. Filtro de Kalman
7. 4DVar
8. Filtro de Kalman por ensambles
9. Filtro de particulas

¿Cuál es la necesidad de asimilar?

- Sistemas muy complejos, i.e. la atmósfera, son imposibles de “medir” con un solo instrumento.
- Mediciones **localizadas y parciales**.

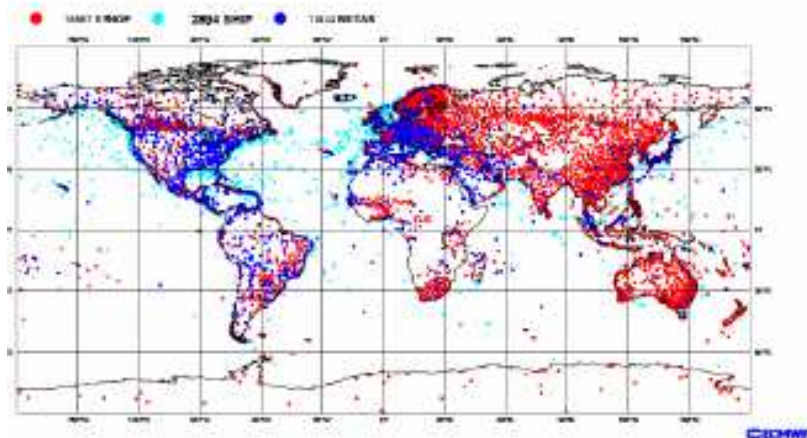


- Los instrumentos realizan mediciones **indirectas** de variables.
- Mediciones **ruidosas**.
- Múltiples instrumentos.

Observaciones en un ciclo (6h)

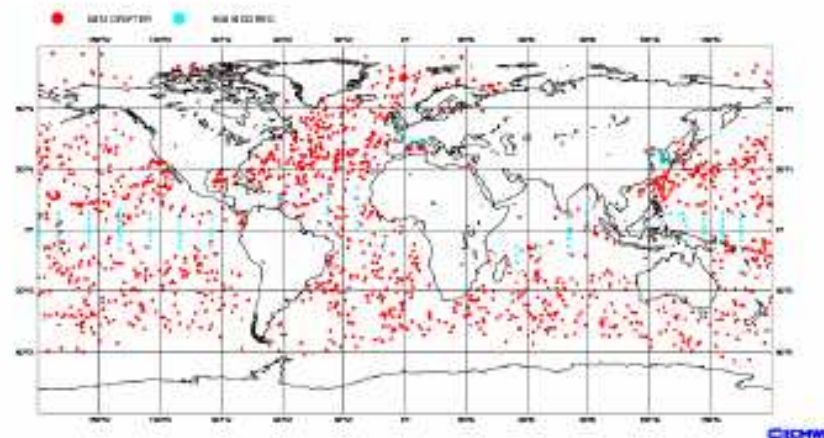
surface stations

ECMWF Data Coverage (All obs DA) - SYNOP/SHIP
12/NOV/2010; 00 UTC
Total number of obs = 31923



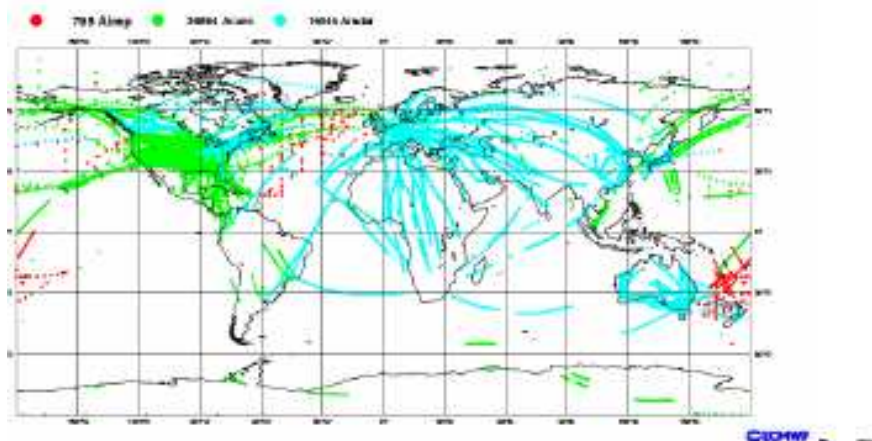
buoys

ECMWF Data Coverage (All obs DA) - BUOY
12/NOV/2010; 00 UTC
Total number of obs = 9423



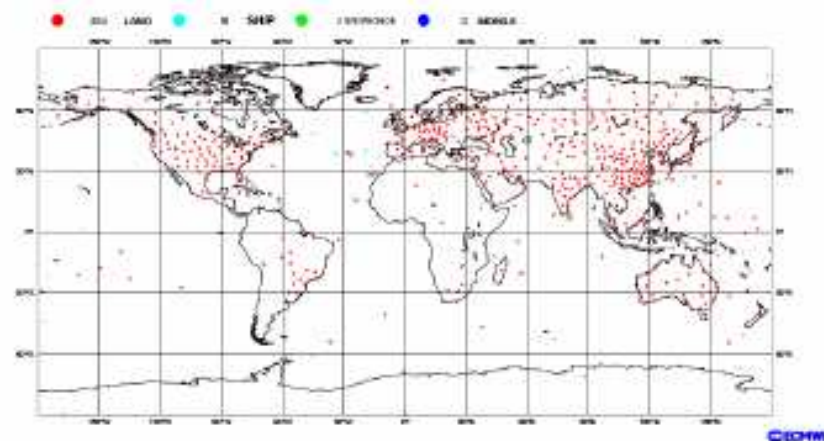
aircraft

ECMWF Data Coverage (All obs DA) - AIRCRAFT
12/NOV/2010; 00 UTC
Total number of obs = 53704



radiosondes

ECMWF Data Coverage (All obs DA) - TEMP
12/NOV/2010; 00 UTC
Total number of obs = 644



Definición de asimilación de datos

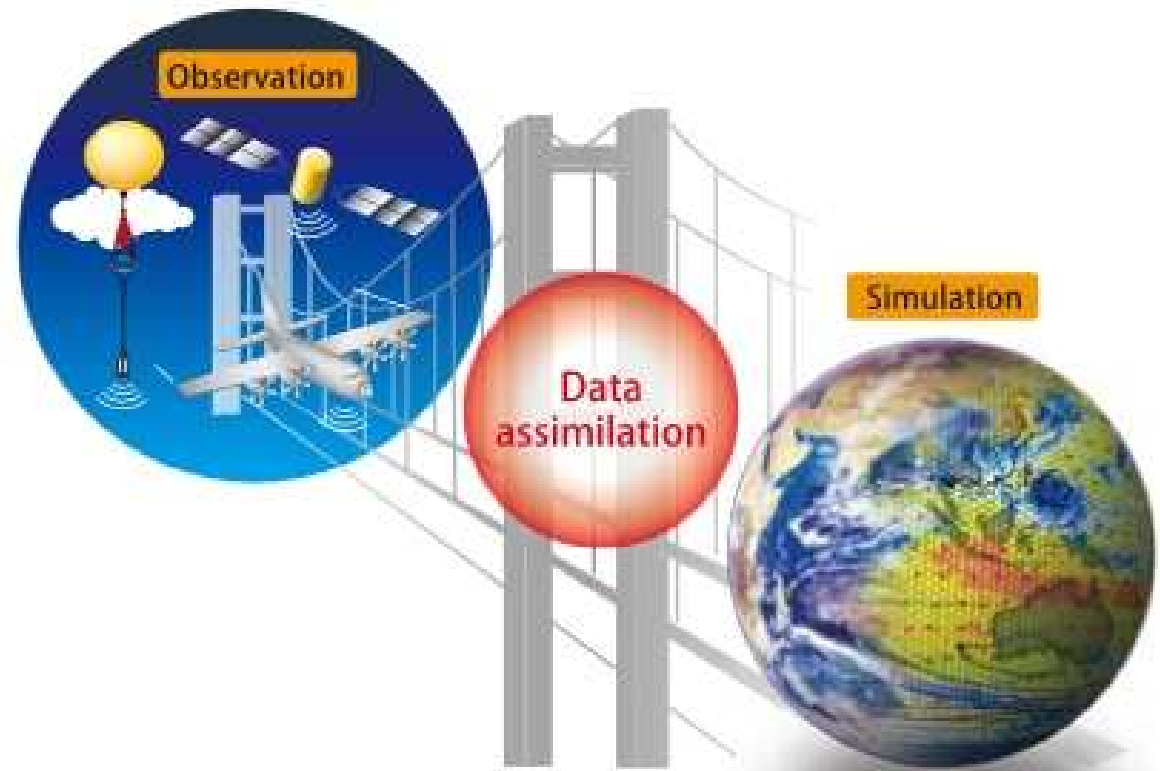
Cambiamos el tradicional lema de la unión hace la fuerza:

LEMA ADAPTADO: La unión **valorada** de información independiente hace la fuerza...

Las observaciones **no** son la única fuente de información.

Conocemos las **leyes físicas**.

Podemos desarrollar modelos que nos **simulen** la evolución del sistema.

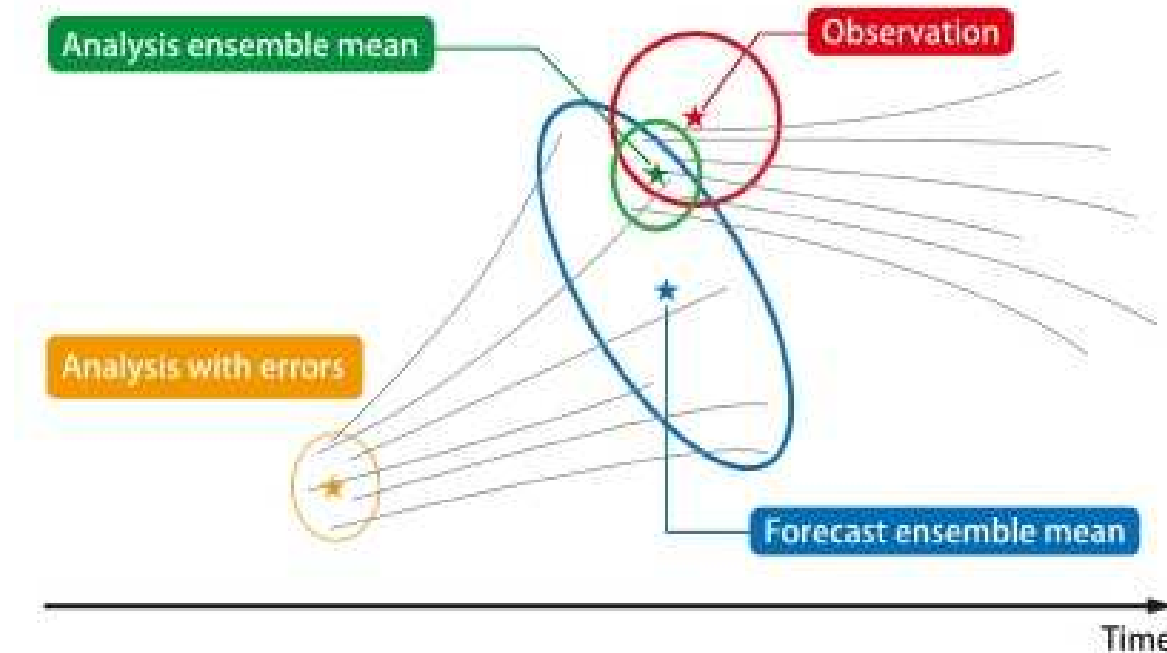


Definición: El objetivo de la asimilación de datos es combinar fuentes de información, mediciones y modelos físicos, para obtener el estado del sistema en un dado tiempo lo mas preciso posible.

Definición de asimilación de datos II

El fin último de la asimilación es la estimación del estado del sistema
(Esta es la parte visible del negocio)

Asimilación de datos es: Cuantificar incertezas de múltiples fuentes de datos (Esta la parte de atras o cocina del negocio).



Hitos de la asimilación de datos

Gauss estudio de la mecánica celestial, e.g. determinación de la órbita de planetas, quien utilizó la técnica actualmente conocida por **cuadrados mínimos**.

Wiener y Kalman en los 50-60 quienes fueron incentivados por la carrera aeroespacial. Estimación estadística.

Talagrand 1986. Asimilación de datos variacional (4DVar). Meteorología

Gordon et al 1993. Filtro de partículas. Radares de defensa. Tracking.

Evensen 1994. Filtro de Kalman por ensambles. Oceanografía.

Aplicaciones de la asimilación de datos

1. Sistemas de navegación
2. Oceanografía
3. Meteorología
5. Contaminación. Flujos de CO₂.
6. Suelos.
7. Sensores remotos (Modelado inverso)
8. Geofísica. Pozos de petróleo. (Modelado inverso)
9. Sistemas biológicos y ecología

Grandes desafíos de la asimilación de datos

- Problema de muy altas dimensiones 10^7 .
- Problema indeterminado. Estado 10^7 . Observaciones 10^5 .
- \rightarrow Esencial un tratamiento o aprendizaje de error de modelo
- Modelo dinámico y operador observación altamente no-lineales. \mathcal{M} y \mathcal{H} . \rightarrow Estadística no-gaussiana.

Todas estas forman parte de áreas de investigación actuales en la parte teórica de la asimilación.

Bibliografía recomendada

La bibliografía fundamental recomendada, siendo la utilizada para el armado de este curso, es:

- Bouttier F. and P. Courtier, 1999: *Data assimilation concepts and methods*. Lecture Notes ECMWF.
- Evensen, G. 2006: *Data assimilation: The ensemble Kalman filtering*. Springer.
- Jazwinski A. H. 1970: *Stochastic processes and filtering theory*. Dover Publications (republished 2007).
- Tarantola A. 2005: *Inverse problem theory*. SIAM.
- Toddling R. 1999: *Estimation theory and foundations of data assimilation*. NASA DAO Office Note 1999-01.

2. Inferencia en un modelo de Markov escondido

Modelo de Markov escondido (HMM) o modelo state-space

Suponemos el estado escondido evoluciona a través de un modelo de Markov

$$\mathbf{x}_k = \mathcal{M}(\mathbf{x}_{k-1}) + \boldsymbol{\eta}_k$$

Las observaciones

$$\mathbf{y}_k = \mathcal{H}(\mathbf{x}_k) + \boldsymbol{\nu}_k$$

\mathbf{x}_k estado (escondido) del sistema (variables aleatorias!!!). k tiempo. N_x dimensiones.

\mathcal{M} modelo dinámico. Ej. WRF.

$\boldsymbol{\eta}_k$ error de modelo. Ej. $\boldsymbol{\eta}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$, \mathbf{Q}_k matriz de covarianza del error de modelo.

\mathbf{y}_k observación (variables aleatorias solo conocemos una realización)

\mathcal{H} operador o modelo observacional. Ej. Radiancia \rightarrow temperatura.

$\boldsymbol{\nu}_k$ error observacional. Ej. $\boldsymbol{\nu}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$, \mathbf{R}_k matriz de covarianza del error de observacional.

La principal hipótesis del HMM es que: **Los verdaderos \mathcal{M} y \mathcal{H} son conocidos. \rightarrow No existen errores sistemáticos.**

Solución general del HMM: Inferencia Bayesiana

Dada la densidad de probabilidad de \mathbf{x}_{k-1} , $p(\mathbf{x}_{k-1})$, y dada una realización de la observación en k , queremos inferir la densidad de probabilidad de \mathbf{x}_k ,

$$p(\mathbf{x}_k | \mathbf{y}_k)$$

El dato es $p(\mathbf{x}_{k-1})$ y necesitamos $p(\mathbf{x}_k)$ se requiere del modelo (densidad de transición):

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

Solución general del HMM: Inferencia Bayesiana

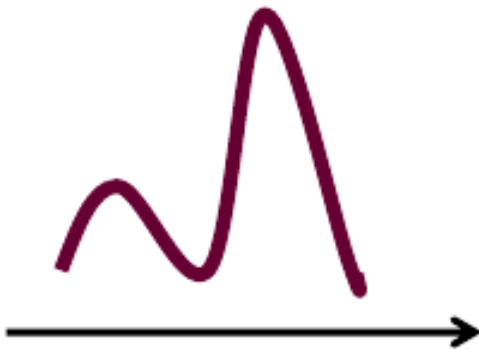
Para determinar el objetivo de nuestro problema podemos usar la regla de Bayes:

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

Diagram illustrating the components of Bayes' rule:

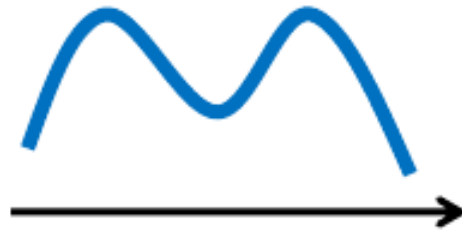
- Verosimilitud** (Likelihood) points to $p(\mathbf{y}_k | \mathbf{x}_k)$
- Densidad posterior** (Posterior density) points to $p(\mathbf{x}_k | \mathbf{y}_k)$
- densidad a priori** (Prior density) points to $p(\mathbf{x}_k)$
- verosimilitud marginal** (Marginal likelihood) points to $p(\mathbf{y}_k)$

posterior



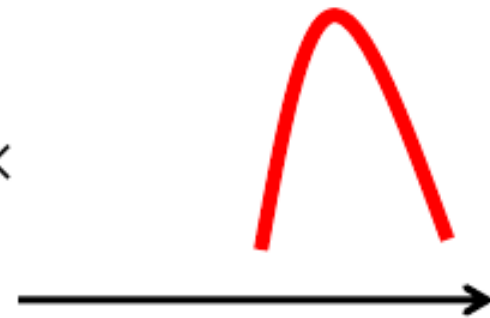
=

prior



×

likelihood



Inferencia Bayesiana

La inferencia Bayesiana es la solución mas completa que podemos obtener de nuestro problema.

En general, es muy costoso determinar las densidades de probabilidad en problemas de muy alta dimensionalidad.

- Se determina solo la media o moda y la covarianza de la densidad.
- Técnicas de Monte Carlo o ensambles. Se representa a la densidad a través de muestras.

3. Ejemplo 1D

Mundo Gaussiano 1D

Supongamos que tenemos una observación de ozono en un punto geográfico y y además tenemos un conocimiento a priori x^b .

Dados: un apriori Gaussiano con media x^b (background) y varianza P^b .

una observación y^o con errores de la observación también Gaussianos y de varianza R .

La densidad posterior es

$$p(x|y) = \frac{p(y|x)p(x)}{\int p(y|x)p(x)dx} = N \exp \left(-\frac{1}{2} \|y - x\|_R \right) \exp \left(-\frac{1}{2} \|x - x^b\|_B \right)$$

Con un poco de álgebra (completando cuadrados en el exponente) se puede obtener

$$p(x|y) = N \exp \left(-\frac{1}{2} \|x - x^a\|_{P^a} \right)$$

Es decir el resultado es una Gaussiana con media y varianza:

$$x^a = x^b + \frac{P^b}{P^b + R} (y - x^b) \text{ y } P^a = (P^{b^{-1}} + R^{-1})^{-1} = \frac{P^b R}{P^b + R}$$

Mundo Gaussiano 1D

La media se interpreta por $x^a = x^b + W(y - x^b)$ con $W = \frac{P^b}{P^b + R}$ el peso que le estamos dando a las observaciones, ergo $1 - W$ es el peso del background.

- Si el background es muy preciso con respecto a la observación $P^b \ll R$ entonces $W \rightarrow 0$, i.e. todo el peso se esta dando al background $x^a = x^b$.
- Si la observación es muy precisa $R \ll P^b$ entonces $W = 1$ y por lo tanto $x^a = y$.

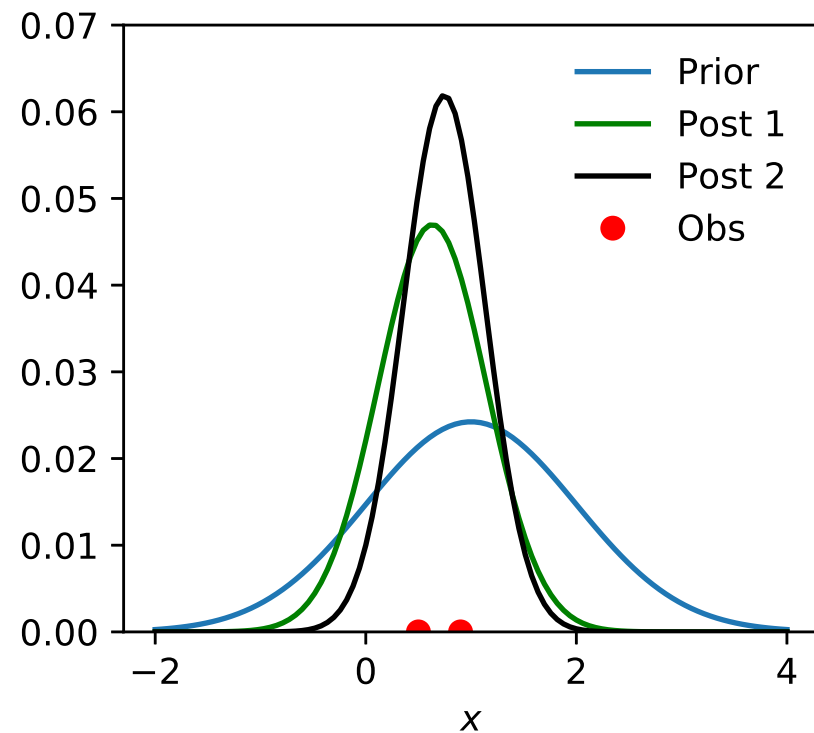
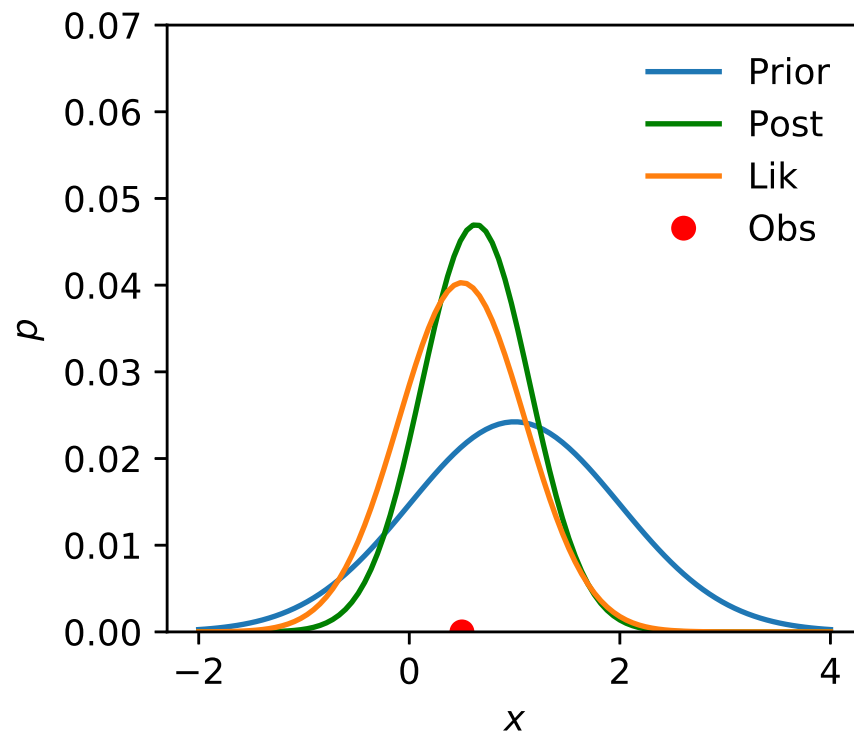
La incerteza de la estimación

$$P^a = (P^{b^{-1}} + R^{-1})^{-1} = \frac{P^b R}{P^b + R}$$

satisface que: $P^a < P^b$ y $P^a < R$.

Como se espera la estimación disminuye la incerteza.

Mundo Gaussiano 1D



$$p(x) \sim \mathcal{N}(1, 1)$$

$$\nu \sim \mathcal{N}(0, R) = \mathcal{N}(0, 0.6^2)$$

Observaciones: $y_1 = 0.5$, $y_2 = 0.9$.

La asimilación reduce la incerteza!!!!

Ejemplo 1D. Varianza mínima

¿Podemos determinar el peso W minimizando la varianza del análisis?

El error del análisis, $e^a = x^a - x^t$, $e^b = x^b - x^t$ es

$$e^a = e^b + W(e^o - e^b)$$

Si asumimos que los errores de las observaciones no están correlacionados con los errores del background entonces $\mathcal{E}\{e^b e^o\} = 0$ entonces

$$\mathcal{E}\{(e^a)^2\} = \mathcal{E}\{(e^b)^2\} - 2W\mathcal{E}\{(e^b)^2\} + W^2\mathcal{E}\{(e^o)^2 + (e^b)^2\}$$

Se quiere minimizar la varianza de los análisis con respecto al peso W por lo que

$$\frac{d\mathcal{E}\{(e^a)^2\}}{dW} = 2[-\mathcal{E}\{(e^b)^2\} + W\mathcal{E}\{(e^o)^2 + (e^b)^2\}] = 0$$

Ejemplo 1D. Varianza mínima

El peso que da la varianza del análisis mínima es:

$$W = \frac{\mathcal{E}\{(e^b)^2\}}{\mathcal{E}\{(e^o)^2 + (e^b)^2\}} = \frac{P^b}{P^b + R}$$

entonces el peso que tendrán las observaciones es igual a la razón de la varianza del background sobre la suma de las varianzas.

Mientras la varianza para el caso del peso óptimo

$$\mathcal{E}\{(e^a)^2\} = P^a = \mathcal{E}\{(e^b)^2\} - 2W\mathcal{E}\{(e^b)^2\} + W^2\mathcal{E}\{(e^o)^2 + (e^b)^2\} = \frac{P^b R}{P^b + R}$$

Esto es el mismo resultado que lo que habíamos obtenido en el marco Bayesiano.

La única diferencia es que aquí determinamos los dos primeros momentos mientras el marco Bayesiano busca determinar las densidades de probabilidad.

Ejemplo 1D. Propagación de la información

Si queremos tener la precipitación de un campo pero la medición del INTA mas cercana está a 300 km.

$$x_{ca}^a = x_{ca}^b + W(y_{INTA} - x_{ca}^b)$$

Optimizando nuevamente la varianza, y despejando el peso,

$$W = \frac{\mathcal{E}\{e_{ca}^b e_{INTA}^b\}}{P_{INTA}^b + R_{INTA}} = \frac{\rho \sigma_{INTA}^b \sigma_{ca}^b}{P_{INTA}^b + R_{INTA}}$$

donde $\mathcal{E}\{e_{ca}^b e_{INTA}^b\} = \rho \sigma_{INTA}^b \sigma_{ca}^b$, ρ es la correlación entre los puntos, σ_{INTA}^b es la desviación standard del background $P_{INTA}^b = (\sigma_{INTA}^b)^2$ en el INTA y σ_{ca}^b en el campo.

Las correlaciones en la covarianza del background nos permiten propagar la información observacional espacialmente.

4. Interpolación óptima

Interpolación óptima

Supongamos nuevamente un problema de inferencia estadística en el cual tenemos:

- una densidad a priori Gaussiana de media \mathbf{x}^b de dimensión n y covarianza $\mathbf{P}^b (n \times n)$.
- tenemos observaciones \mathbf{y} (n dim) con error observacional Gaussiano y una matriz de covarianza del error observacional \mathbf{R} .

Queremos encontrar la media y covarianza del análisis. La solución de este problema es lo que se denomina interpolación óptima.

Nuevamente existen dos opciones encontrar la densidad posterior en el marco Bayesiano o el estado y la covarianza que minimizan la varianza.

$$p(\mathbf{x}|\mathbf{y}) = A \exp \left[-\frac{1}{2} (\|\mathbf{x} - \mathbf{x}^b\|_{\mathbf{P}^b} + \|\mathbf{y} - \mathbf{x}\|_{\mathbf{R}}) \right]$$

Esto es totalmente equivalente al caso 1D, solo que ahora tenemos matrices.
Completando cuadrados se obtiene

$$p(x|y) = A \exp \left[-\frac{1}{2} (\|\mathbf{x} - \mathbf{x}^a\|_{\mathbf{P}^a}) \right]$$

con media el análisis:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{P}^b(\mathbf{P}^b + v\mathbf{R})^{-1}(\mathbf{y} - \mathbf{x}^b)$$

y la covarianza del análisis es

$$\mathbf{P}^a = [(\mathbf{P}^b)^{-1} + (\mathbf{R})^{-1}]^{-1}$$

Aquí tenemos correlaciones entre los distintos componentes de \mathbf{x} y de \mathbf{y} .

Interpolación óptima

Supongamos que tenemos una ecuación de la forma

$$\hat{\mathbf{x}}^a = \hat{\mathbf{x}}^b + \mathbf{W}(\mathbf{y} - \hat{\mathbf{x}}^b)$$

donde \mathbf{x}^b es el background, \mathbf{x}^o son las observaciones. Asumimos que las observaciones son realizadas en las mismas ubicaciones que los puntos de grilla.

La verdad, que la desconocemos, la denotamos por \mathbf{x}^t . Los errores vienen dados por

$$\mathbf{e}^a = \mathbf{x}^a - \mathbf{x}^t, \quad \mathbf{e}^b = \mathbf{x}^b - \mathbf{x}^t, \quad \mathbf{e}^o = \mathbf{x}^o - \mathbf{x}^t$$

Las matrices de covarianza son:

$$\mathbf{P}^a = \mathcal{E}\{\mathbf{e}^a(\mathbf{e}^a)^T\}, \quad \mathbf{P}^b = \mathcal{E}\{\mathbf{e}^b(\mathbf{e}^b)^T\}, \quad \mathbf{R} = \mathcal{E}\{\mathbf{e}^o(\mathbf{e}^o)^T\}$$

Por $\mathcal{E}\{\}$ representamos la esperanza.

El análisis es insesgado

Restando por \mathbf{x}^t en la ecuación de análisis

$$\mathbf{e}^a = \mathbf{e}^b + \mathbf{W}(\mathbf{e}^o - \mathbf{e}^b)$$

Asumimos que los errores son insesgados:

$$\mathcal{E}\{\mathbf{e}^o\} = \mathcal{E}\{\mathbf{e}^b\} = 0$$

Si los errores de las observaciones y del background son insesgados se deduce que $\mathcal{E}\{\mathbf{e}^a\} = 0$.

Además asumimos que no hay correlaciones entre los errores:

$$\mathcal{E}\{\mathbf{e}^b(\mathbf{e}^R)^T\} = \mathcal{E}\{\mathbf{e}^R(\mathbf{e}^b)^T\} = 0$$

Covarianza del análisis

Calculando la covarianza del análisis

$$\mathcal{E}\{\mathbf{e}^a(\mathbf{e}^a)^T\} = \mathcal{E}\{[\mathbf{e}^b + \mathbf{W}(\mathbf{e}^R - \mathbf{e}^b)][\mathbf{e}^b + \mathbf{W}(\mathbf{e}^R - \mathbf{e}^b)]^T\}$$

Teniendo en cuenta las hipótesis de que los errores no tienen correlaciones entre sí:

$$\mathbf{P}^a = \mathbf{P}^b + \mathbf{W}(\mathbf{R} + \mathbf{P}^b)\mathbf{W}^T - \mathbf{P}^b\mathbf{W}^T - \mathbf{W}\mathbf{P}^b$$

Para minimizar la covarianza se deriva la traza con respecto a la matriz peso \mathbf{W} :

$$\frac{d\text{Tr}(\mathbf{P}^a)}{d\mathbf{W}} = 2\mathbf{W}(\mathbf{R} + \mathbf{P}^b) - 2\mathbf{P}^b = 0$$

donde se usó que \mathbf{P}^b y \mathbf{R} son matrices simétricas.

Se deduce que el peso que minimiza la varianza del análisis es:

$$\mathbf{W} = \mathbf{P}^b(\mathbf{R} + \mathbf{P}^b)^{-1}$$

Optimal interpolation is un BLUE (best linear unbiased estimator).

Interpolación óptima

El análisis viene dado por

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{P}^b(\mathbf{R} + \mathbf{P}^b)^{-1}(\mathbf{x}^o - \mathbf{x}^b)$$

La covarianza del análisis es

$$\mathbf{P}^a = [(\mathbf{P}^b)^{-1} + (\mathbf{R})^{-1}]^{-1}$$

Si los errores del background son pequeños se da que el peso del incremento $\mathbf{P}^b(\mathbf{R} + \mathbf{P}^b)^{-1}$ es prácticamente nulo, luego el resultado del análisis es $\mathbf{x}^a \approx \mathbf{x}^b$.

Si los errores de las observaciones son pequeños todo el peso estará en la innovación $\mathbf{x}^a \approx \mathbf{x}^o$.

Si $\mathbf{P}^b = \mathbf{R}$ se tiene que $\mathbf{x}^a = \mathbf{x}^b + 1/2(\mathbf{x}^o - \mathbf{x}^b) = 0.5(\mathbf{x}^b + \mathbf{x}^o)$.

Ejemplo n=2

Supongamos que tenemos mediciones de temperatura en dos ciudades.

Una hipótesis habitual es que no hay correlaciones entre las mediciones, además si es el mismo tipo de instrumento se tiene el mismo error por lo que

$$\mathbf{R} = \begin{bmatrix} (\sigma^o)^2 & 0 \\ 0 & (\sigma^o)^2 \end{bmatrix}$$

En el caso de la covarianza del background se define a la correlación por

$$\rho = \frac{\mathcal{E}\{e_1^b e_2^b\}}{\sqrt{\mathcal{E}\{(e_1^b)^2\} \mathcal{E}\{(e_2^b)^2\}}}$$

La covarianza del background es

$$\mathbf{P}_b = \begin{bmatrix} (\sigma_1^b)^2 & \rho \sigma_1^b \sigma_2^b \\ \rho \sigma_1^b \sigma_2^b & (\sigma_2^b)^2 \end{bmatrix}$$

Ejemplo n=2

Luego se obtiene

$$(\mathbf{R} + \mathbf{P}_b)^{-1} = \frac{1}{|\mathbf{R} + \mathbf{P}^b|} \begin{bmatrix} (\sigma^o)^2 + (\sigma_2^b)^2 & -\rho\sigma_1^b\sigma_2^b \\ -\rho\sigma_1^b\sigma_2^b & (\sigma^R)^2 + (\sigma_1^b)^2 \end{bmatrix}$$

Si la correlación es nula $\rho = 0$ los análisis 1 y 2 son independientes. Sin embargo si $\rho \neq 0$ entonces la medición en 1 incide en el análisis de 2.

La covarianza del background, a través de sus correlaciones, hace que la información de las observaciones sea propagada.

La matriz de peso:

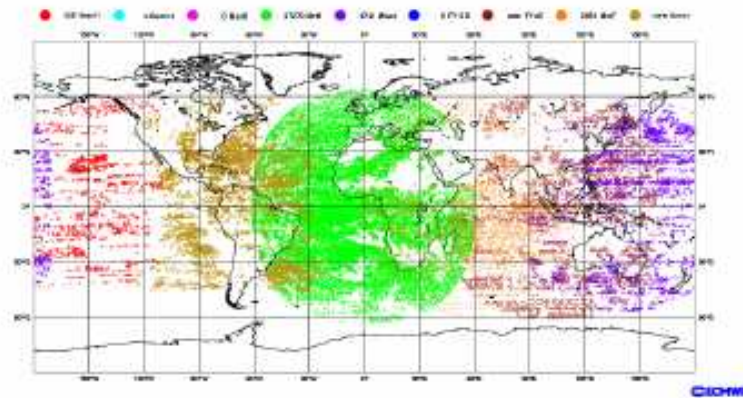
$$\mathbf{W} = [(1 + \alpha_2)(1 + \alpha_1) + \rho^2]^{-1} \begin{bmatrix} (1 + \alpha_2) - \rho^2 & \rho\sqrt{\alpha_1\alpha_2} \\ \rho\sqrt{\alpha_1\alpha_2} & (1 + \alpha_1) - \rho^2 \end{bmatrix}$$

$$\text{donde } \alpha_1 = \frac{(\sigma^R)^2}{(\sigma_1^b)^2}, \alpha_2 = \frac{(\sigma^R)^2}{(\sigma_2^b)^2}$$

Variables observadas vs variables de estado

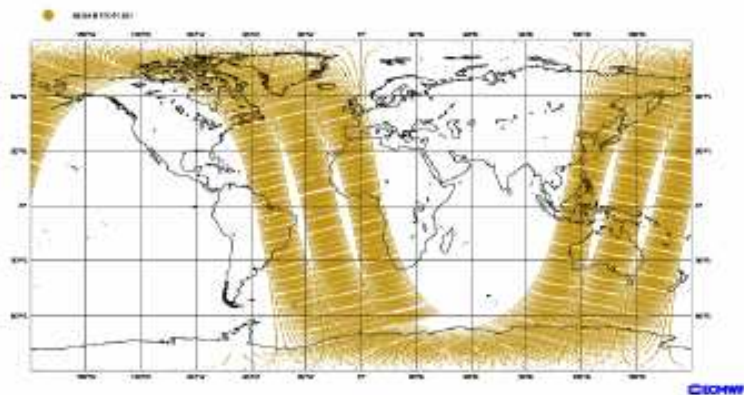
Infrared AMVs

ECMWF Data Coverage (All obs DA) - AMV IR
12/NOV/2010; 00 UTC
Total number of obs = 50083



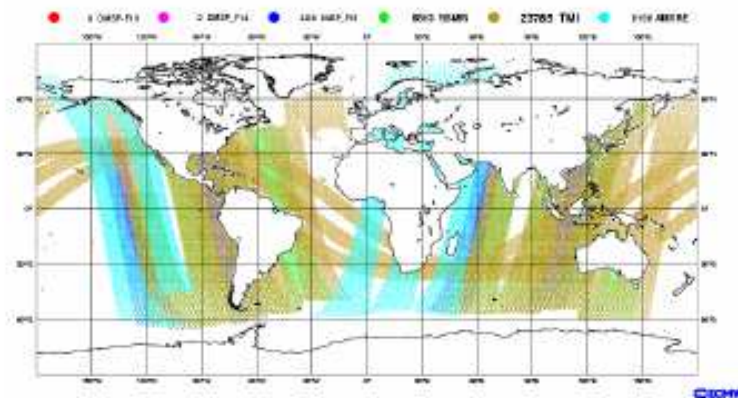
IASI

ECMWF Data Coverage (All obs DA) - IASI
12/NOV/2010; 00 UTC
Total number of obs = 56039



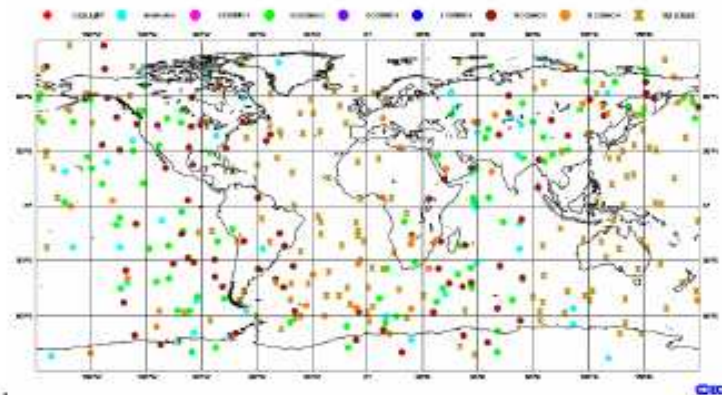
Microwave imagers

ECMWF Data Coverage (All obs DA) - Microwave imager
12/NOV/2010; 00 UTC
Total number of obs = 43153



Bending angle

ECMWF Data Coverage (All obs DA) - GPSRO
12/NOV/2010; 00 UTC
Total number of obs = 379



Interpolación óptima con operador observacional

En general se requiere:

- Interpolación espacial.
- La variable que se mide no es la misma que la del modelo. Eg los satélites miden radianza y la variable es la temperatura.

Definimos al operador observación/forward model \mathcal{H} que transforma las variables del espacio del estado al espacio observacional.

El análisis en este caso viene dado por

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y} - \mathcal{H}(\mathbf{x}^b))$$

\mathbf{K} es la matriz de peso.

$\mathbf{y} - \mathcal{H}(\mathbf{x}^b)$ es la innovación o incremento.

Error de representación

Supongamos que las observaciones tienen un error estocástico debido al instrumento ν ,

$$\mathbf{y} = \mathbf{y}^t + \mathbf{v}$$

Sin embargo ahora tenemos además un error al hacer la proyección:

$$\begin{aligned}\mathbf{y} &= \mathbf{y}^t + \mathcal{H}(\mathbf{x}^t) - \mathcal{H}(\mathbf{x}^t) + \mathbf{v} \\ &= \mathcal{H}(\mathbf{x}^t) + \nu\end{aligned}$$

Entonces el error total de la observación es

$$\mathbf{v} = [\mathbf{y}^t - H(\mathbf{x}^t)] + \nu$$

donde $[\mathbf{y}^t - H(\mathbf{x}^t)]$ es el error de representación.

La matriz de covarianza del error de las observaciones es

$$\mathbf{R} = \mathcal{E}\{(\mathbf{v} - \mathcal{E}\{\mathbf{v}\})(\mathbf{v} - \mathcal{E}\{\mathbf{v}\})^T\}$$

R incluye el error de observación y el de representación.

Nuevamente, es sencillo encontrar la densidad posterior:

$$p(\mathbf{x}|\mathbf{y}) = A \exp \left[-\frac{1}{2} \left(\|\mathbf{x} - \mathbf{x}^b\|_{\mathbf{P}^b} + \|\mathbf{y} - \mathcal{H}(\mathbf{x})\|_{\mathbf{R}} \right) \right]$$

Sin embargo, si \mathcal{H} tiene una dependencia no lineal con \mathbf{x} , **la densidad posterior NO es Gaussiana.**

¿Qué cosas pueden suceder cuando dejamos de vivir en el mundo Gaussiano?

Ejemplo: Mediciones del módulo de la velocidad

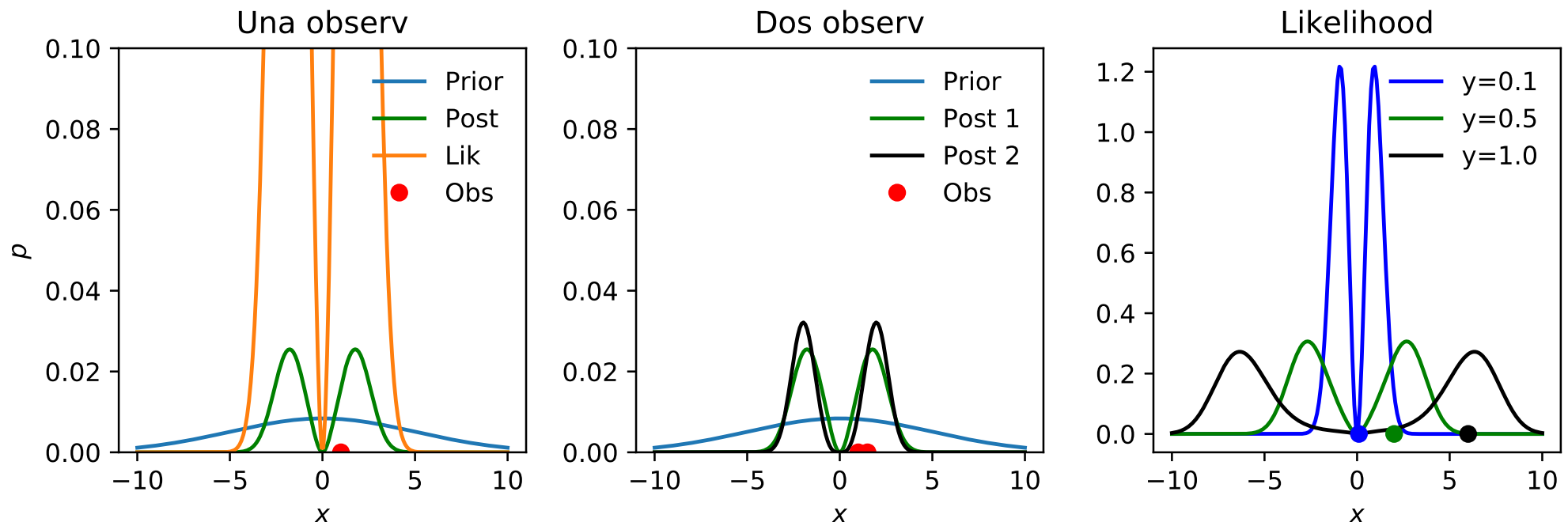
Supongamos la variable de estado es la velocidad y el instrumento me permite medir el modulo de la velocidad.

$$\mathcal{H}(x) = |x|$$

Fuerte desafio cuando el mapeo no es inyectivo!

Densidad a priori: $p(x^b) = \mathcal{N}(0, (5\text{m/s})^2)$

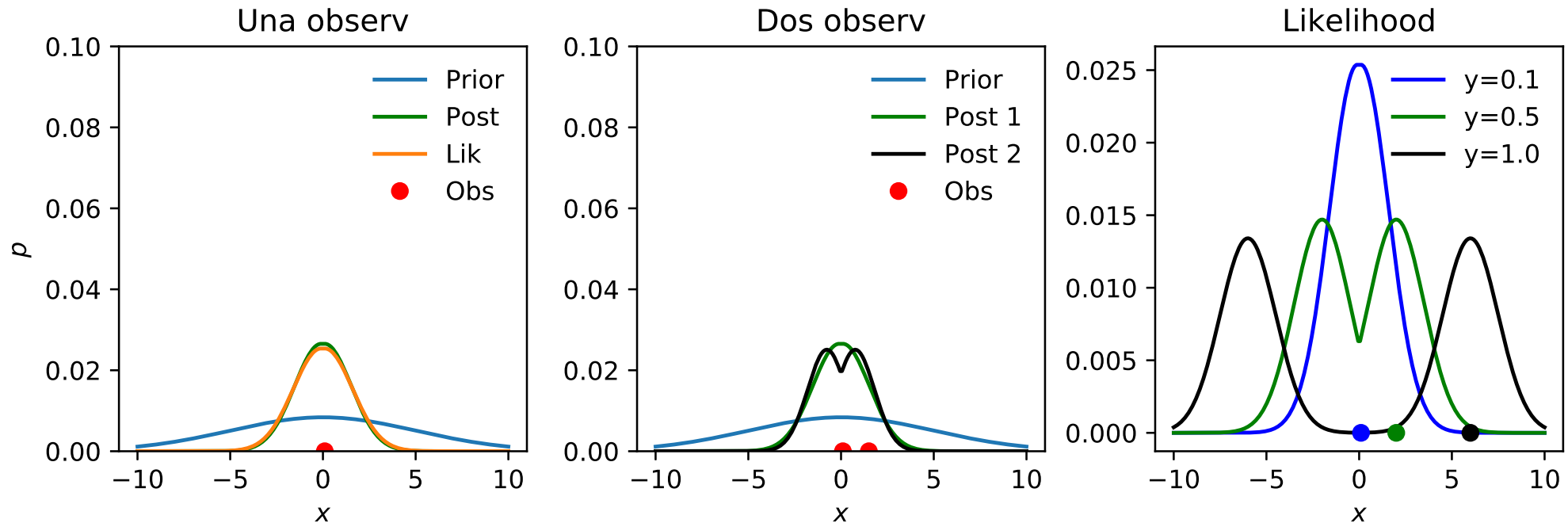
Error observacional: densidad **Gamma**



Ejemplo: Mediciones del módulo de la velocidad

Que pasa si lo ignoramos y seguimos pensando en un mundo Gaussiano.

Error observacional: densidad **Gaussiano**



Conclusión: \mathcal{H} no lineal (no-inyectivo) puede resultar en:

- La moda y la media difieren.
- Pueden existir densidades bimodales.

¿Que es lo que da OI? ¿Que estimación e incerteza resulta en estos casos?

Marco Bayesiano

Asumiendo no linealidades suaves y \mathcal{H} inyectiva en el dominio de interés, aproximamos linealmente alrededor de \mathbf{x}^b :

$$\mathbf{y} - \mathcal{H}(\mathbf{x}) = \mathbf{y} - \mathcal{H}(\mathbf{x}) \approx \mathbf{y} - \mathcal{H}(\mathbf{x}^b) + \mathbf{H}(\mathbf{x} - \mathbf{x}^b)$$

donde \mathbf{H} es el **tangente lineal** definido por

$$\mathbf{H} = \left. \frac{d\mathcal{H}}{dx} \right|_{\mathbf{x}_b}$$

Dado que ahora tenemos transformaciones lineales, la posterior será una Gaussiana. completando cuadrados en el argumento de la exponencial se obtiene:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}[\mathbf{y} - \mathcal{H}(\mathbf{x}^b)]$$

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^b$$

Error del análisis

El error del análisis viene dado por

$$\mathbf{e}^a = \mathbf{e}^b + \mathbf{K}[\mathbf{y} - \mathcal{H}(\mathbf{x}^t) + \mathcal{H}(\mathbf{x}^t) - \mathcal{H}(\mathbf{x}^b)]$$

$$\mathbf{e}^a = \mathbf{e}^b + \mathbf{K}[\mathbf{v} + \mathcal{H}(\mathbf{x}^b + \mathbf{x}^t - \mathbf{x}^b) - \mathcal{H}(\mathbf{x}^b)]$$

El problema aqui es que \mathcal{H} no solo representa interpolación espacial (transformación lineal) sino también transformaciones de variables (nolineal) por lo que se debe aproximar:

$$\mathcal{H}(\mathbf{x}^b - \mathbf{e}^b) \approx \mathcal{H}(\mathbf{x}^b) - \mathbf{H}\mathbf{e}^b$$

El error del análisis es aproximado por

$$\mathbf{e}^a = \mathbf{e}^b + \mathbf{K}[\mathbf{v} - \mathbf{H}\mathbf{e}^b]$$

Varianza mínima

Calculemos la covarianza del análisis

$$\mathcal{E}\{\mathbf{e}^a(\mathbf{e}^a)^T\} = \mathcal{E}\{(\mathbf{e}^b + \mathbf{K}[\mathbf{v} - \mathbf{H}\mathbf{e}^b])(\mathbf{e}^b + \mathbf{K}[\mathbf{v} - \mathbf{H}\mathbf{e}^b])^T\}$$

$$\mathbf{P}^a = \mathbf{P}^b + \mathbf{K}(\mathbf{R} + \mathbf{H}\mathbf{P}^b\mathbf{H}^T)\mathbf{K}^T - \mathbf{K}\mathbf{H}\mathbf{P}^b - \mathbf{P}^b\mathbf{H}^T\mathbf{K}^T$$

Si se pide varianza mínima

$$\frac{dTr(\mathbf{P}^a)}{d\mathbf{K}} = 0$$

El algoritmo de interpolación óptima viene dado por:

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

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}[\mathbf{y} - H(\mathbf{x}^b)]$$

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^b$$

Esto define el ciclo de interpolación óptima. Sin embargo necesitamos conocer las covarianzas, i.e. \mathbf{P}^b , \mathbf{R} .

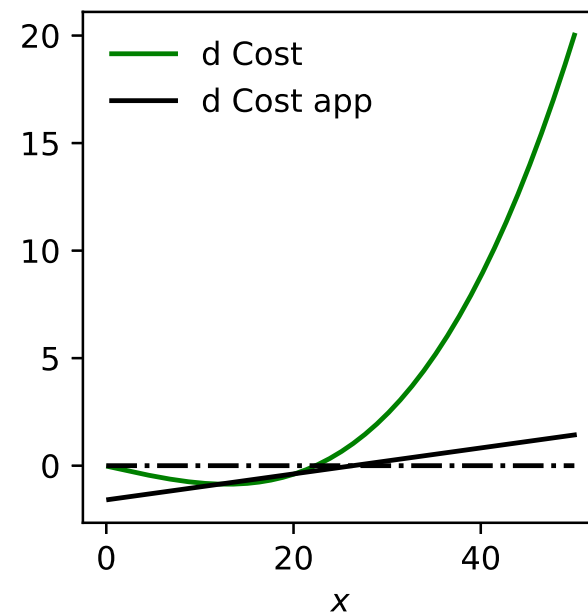
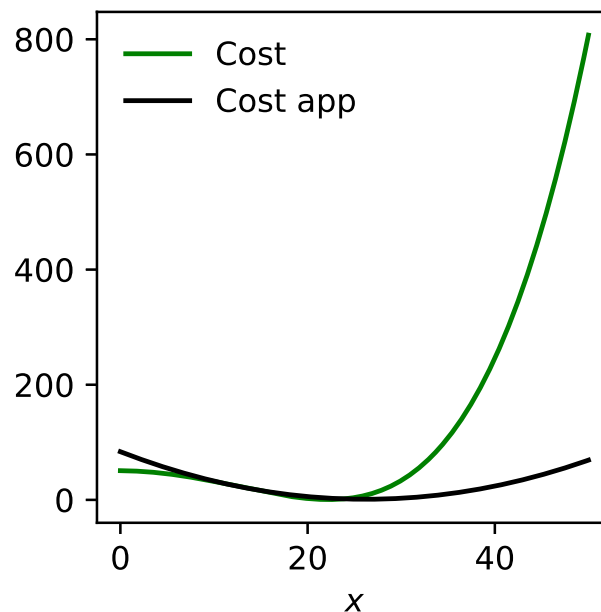
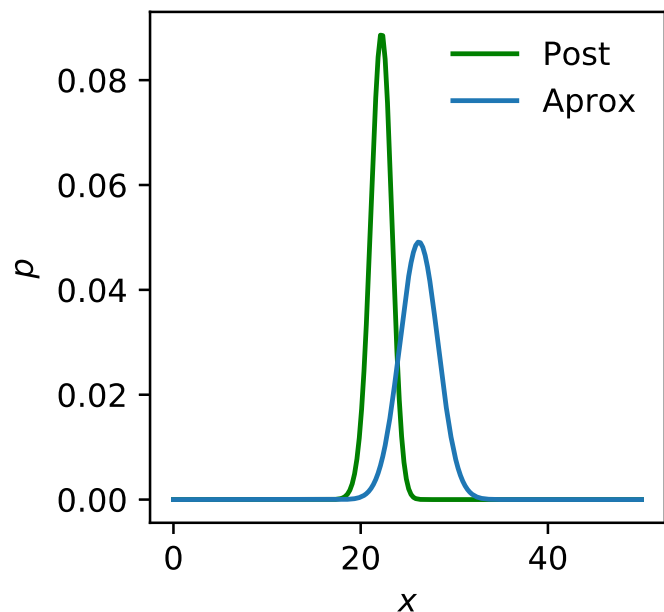
El algoritmo suele ser llamado interpolación estadística (Lorenc 1981, 1986) por las simplificaciones que se realizan en la práctica con las covarianzas.

Es la solución de interpolación óptima, óptima?

La aproximación que estamos realizando es:

$$\begin{aligned}\log p(\mathbf{x}|\mathbf{y}) &= -\frac{1}{2} \left(\|\mathbf{x} - \mathbf{x}^b\|_{\mathbf{P}^b} + \|\mathbf{y} - \mathcal{H}(\mathbf{x})\|_{\mathbf{R}} \right) \\ &\approx -\frac{1}{2} \left(\|\mathbf{x} - \mathbf{x}^b\|_{\mathbf{P}^b} + \|\mathbf{y} - \mathcal{H}(\mathbf{x}^b) + \mathbf{H}(\mathbf{x} - \mathbf{x}^b)\|_{\mathbf{R}} \right)\end{aligned}$$

Caso: $\mathcal{H}(x) = 0.1x^2$



$J = -\log p(x|y)$. Parámetros: $\sigma_b = 10$, $\sigma_r = 5$, $x^b = 12$, $y = 50$

5. Asimilación variacional de 3 dimensiones (3DVar)

Asimilación variacional de 3 dimensiones (3DVar)

La asimilación de datos variacional de 3 dimensiones ha sido una de las técnicas mas populares en la producción de análisis en los centros meteorológicos mas importantes del mundo:

- Parrish and Derber, 1992 NMC/NCEP
- Anderson et al 1998, Ravier et al. 1998, Courtier et al. 1998 ECMWF
- Gauthier et al. 1999 CMC - Environment Canada
- Lorenc et al. 2000 Met Office

El mayor problema de 3DVar es la hipótesis de covarianza del background \mathbf{P}^b estacionaria (no hay una ley para la evolución de la covarianza del background).

Actualmente sigue siendo muy popular, pero la mayoría de los centros usa 4DVar y EnKF.

Ejemplo 1D con 3DVar

En el ejemplo 1D con interpolación óptima vimos que para el caso de errores no-sesgados:

$$x^a = x^b + K(x^o - x^b), \quad K = \frac{(\sigma^b)^2}{(\sigma^o)^2 + (\sigma^b)^2}, \quad (\sigma^a)^{-2} = (\sigma^o)^{-2} + (\sigma^b)^{-2}$$

Supongamos que en lugar de buscar el estimador de varianza mínima (MV) queremos el estimador MAP. Asumamos además que los errores son gaussianos.

La PDF conjunta de errores de observaciones y modelos (la $p_y(\mathbf{y})$ es irrelevante) es:

$$p(x^t) = \frac{1}{2\pi\sigma^b\sigma^o} \exp \left[-\frac{(x^o - x^t)^2}{2(\sigma^o)^2} - \frac{(x^b - x^t)^2}{2(\sigma^b)^2} \right]$$

Luego el valor mas probable es el que maximiza la pdf, es decir el que minimiza su argumento:

$$J(x) = \left[\frac{(x^o - x^t)^2}{2(\sigma^o)^2} + \frac{(x^b - x^t)^2}{2(\sigma^b)^2} \right]$$

La solución es:

$$x^a = x^b + \frac{(\sigma^b)^2}{(\sigma^o)^2 + (\sigma^b)^2} (x^o - x^b)$$

Conclusión: Si el operador observacional es lineal y las distribuciones Gaussianas, el estimador MAP y el MV son equivalentes.

Asimilación variacional de 3 dimensiones

Hipótesis

1. De las observaciones $\mathbf{y} = \mathcal{H}(\mathbf{x}) + \mathbf{v}$ se asume que la variable aleatoria \mathbf{v} tiene una distribución normal $\mathbf{v} \propto \mathcal{N}(0, \mathbf{R})$.
2. Se asume que el estado tiene una distribución normal con media \mathbf{x}^b y matriz de covarianza \mathbf{P}^b , tal que $\mathbf{x} \propto \mathcal{N}(\mathbf{x}^b, \mathbf{P}^b)$.
3. Se asume que las observaciones son insesgadas $\mathcal{E}\{\mathbf{v}\} = 0$.

Densidad apriori y verosimilitud

La verosimilitud de las observaciones tiene la forma $\mathbf{y} \propto \mathcal{N}(\mathcal{H}(\mathbf{x}), \mathbf{R})$, es decir

$$p_{y|x}(\mathbf{y}|\mathbf{x}) \propto \exp\left[-\frac{1}{2}(\mathbf{y} - \mathcal{H}(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - \mathcal{H}(\mathbf{x}))\right]$$

Los errores del background son tales que $\mathcal{E}\{\mathbf{x} - \mathbf{x}^b\} = 0$ y $\mathbf{P}^b = \mathcal{E}\{(\mathbf{x} - \mathbf{x}^b)(\mathbf{x} - \mathbf{x}^b)^T\}$

$$p_x(\mathbf{x}) \propto \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T (\mathbf{P}^b)^{-1}(\mathbf{x} - \mathbf{x}^b)\right]$$

De la regla de Bayes

$$p_{x|y} = \frac{p_{y|x}(\mathbf{y}|\mathbf{x})p_x(\mathbf{x})}{p_y(\mathbf{y})}$$

dado que queremos determinar a \mathbf{x} , $p_y(\mathbf{y})$ no influye en la estimación.

Asimilación variacional de 3 dimensiones

Luego la probabilidad condicional del estado dada las observaciones es

$$p_{x|y}(\mathbf{x}|\mathbf{y}) \propto \exp\left\{-\frac{1}{2}\left[(\mathbf{x}-\mathbf{x}^b)^T(\mathbf{P}^b)^{-1}(\mathbf{x}-\mathbf{x}^b) + (\mathbf{y}-\mathcal{H}(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-\mathcal{H}(\mathbf{x}))\right]\right\}$$

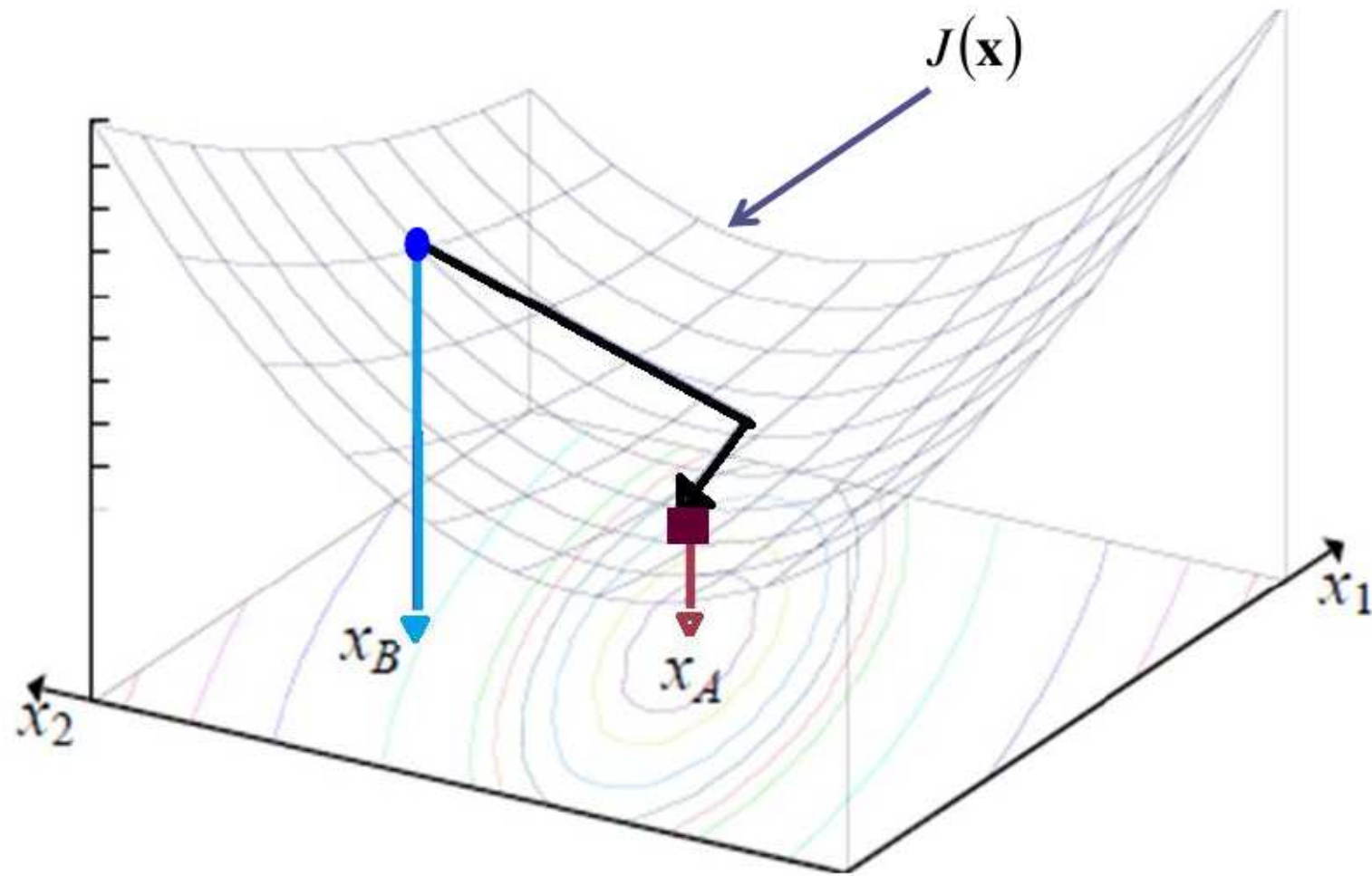
La idea es determinar el máximo de la probabilidad condicional a priori, el modo de la pdf, entonces tenemos que encontrar el máximo del argumento o el mínimo de:

$$J(\mathbf{x}) = \frac{1}{2}\left[(\mathbf{x} - \mathbf{x}^b)^T(\mathbf{P}^b)^{-1}(\mathbf{x} - \mathbf{x}^b) + (\mathbf{y} - \mathcal{H}(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y} - \mathcal{H}(\mathbf{x}))\right]$$

Entonces 3dVar es una estimación MAP donde se asume estadística gaussiana.

Notar que en ningún momento se asume que \mathcal{H} sea lineal.

Optimización en 3dVar



Minimización de la función de costo

Los algoritmos de minimización mas populares que se utilizan son **gradientes conjugados** o **quasi-Newton**.

Para ambos algoritmos se necesita el gradiente de la función de costo.

Esto es de por si un requerimiento evidente ya que estamos en un espacio de 10^7 dimensiones por lo que evaluando solo la función de costo es absolutamente inviable (requeriríamos del orden de 10^7 evaluaciones en cada iteración).

El gradiente de la función de costo viene dado por

$$\nabla J = -\mathbf{H}^\top \mathbf{R}^{-1}(\mathbf{y} - \mathcal{H}(\mathbf{x})) + (\mathbf{P}^b)^{-1}(\mathbf{x} - \mathbf{x}^b)$$

donde $\mathbf{H}^{TL} = \frac{\partial \mathcal{H}}{\partial \mathbf{x}} \Big|_{\mathbf{x}^b}$ y $\mathbf{H}^T = (\mathbf{H}^{TL})^T$.

Es decir que requerimos del tangente lineal y el adjunto del operador \mathcal{H}

3DVar vs OI

- OI: asume de antemano que H es lineal y estadística Gaussiana.
- 3DVAR: solo asumimos que la pdf debe ser unimodal.
- Si el problema tiene estadística gaussiana y operador observacional lineal en ese caso **3DVar y OI son equivalentes** (Lorenc, 1986).

Cuestiones numéricas:

- OI requiere de la inversión de la matriz $\mathbf{HP}^b\mathbf{H}^T + \mathbf{R}$ de $m \times m$ ($m \approx 10^5$).
- 3DVar minimiza una función en un espacio de n dimensiones ($n \approx 10^7$).

Simplificaciones en la práctica:

- OI selecciona un conjunto de observaciones y luego realiza una inversión “regional” ($m = 100$).
- 3DVar la minimización es realizada con unas pocas iteraciones del orden de $100 \ll n$.

Precondicionamiento

Número de condición: El número de condición de una matriz es la razón entre el autovalor mas grande y el mas pequeño, viz es siempre mayor a 1. En un problema de minimización el número de condición se refiere al Hessiano de J .

En dos dimensiones estamos pensando en la razón entre el eje mayor y el eje menor de la elipse. Entonces mientras mas cercano a 1 sea el nro de condición tendremos una circunferencia.

Si el nro de condición es muy grande decimos que el problema esta mal-condicionado (ill-conditioned), necesitamos muchas iteraciones de minimización para obtener el mínimo.

10^7 iteraciones de minimización??? **Una estimación mal-condicionada conlleva al fracaso.**

Espacio de control

Escribiendo el problema en la forma incremental, $\delta \mathbf{x} = \mathbf{x} - \mathbf{x}_b$ luego la función de costo es:

$$J(\delta \mathbf{x}) = \frac{1}{2} [(\delta \mathbf{x})^T (\mathbf{P}^b)^{-1} (\delta \mathbf{x}) + (\mathbf{d} - \mathbf{H}(\delta \mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{d} - \mathbf{H}(\delta \mathbf{x}))]$$

definamos una nueva variable $\boldsymbol{\chi}$

$$\delta \mathbf{x} = (\mathbf{P}^b)^{1/2} \boldsymbol{\chi}$$

Transformando la función de costo a esta nueva variable,

$$J(\boldsymbol{\chi}) = \frac{1}{2} \boldsymbol{\chi}^T \boldsymbol{\chi} + \frac{1}{2} (\mathbf{d} - \mathbf{H}(\mathbf{P}^b)^{1/2} \boldsymbol{\chi})^T \mathbf{R}^{-1} (\mathbf{d} - \mathbf{H}(\mathbf{P}^b)^{1/2} \boldsymbol{\chi})$$

donde el incremento es $\mathbf{d} = \mathbf{y} - H(\mathbf{x}_b)$.

Luego el gradiente viene dado por

$$\nabla J = [\mathbf{I} + (\mathbf{P}^b)^{1/2}(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})(\mathbf{P}^b)^{1/2}] \boldsymbol{\chi} - (\mathbf{P}^b)^{1/2} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{d}$$

y el Hessiano es

$$\nabla^2 J(\boldsymbol{\chi}) = [\mathbf{I} + (\mathbf{P}^b)^{1/2}(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})(\mathbf{P}^b)^{1/2}]$$

En el caso de \mathcal{H} fuertemente no lineales, el preconditionamiento en un entorno no implica preconditionamiento en todo el dominio (mas relevante para 4DVar), es decir en otro lugar puede seguir estando mal-condicionado (e.g. Pulido et al. 2012).

Incrementos filtrados: Outer loop/Inner loop

Minimización en el espacio observacional PSAS

De la solución de OI:

$$\hat{\mathbf{x}} = \mathbf{x}^b + \mathbf{P}^b \mathbf{H}^T (\mathbf{R} + \mathbf{H} \mathbf{P}^b \mathbf{H}^T)^{-1} (\mathbf{y} - \mathbf{H} \mathbf{x}^b)$$

Proponemos resolver el sistema en 2 pasos con una nueva variable \mathbf{z} tal que

$$\delta \mathbf{x} = \mathbf{P}^b \mathbf{H}^T \mathbf{z}$$

$$(\mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R}) \mathbf{z} = \mathbf{y} - \mathbf{H} \mathbf{x}^b$$

en lugar de invertir la matriz proponemos que es el gradiente de una función de costo que se anula:

$$\nabla_{\mathbf{z}} J = (\mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R}) \mathbf{z} - (\mathbf{y} - \mathbf{H} \mathbf{x}^b)$$

Luego la función de costo viene dada por

$$J_{PSAS}(\mathbf{z}) = \mathbf{z}^T (\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R})\mathbf{z} - \mathbf{z}^T (\mathbf{y} - \mathbf{H}\mathbf{x}^b)$$

esta es la metodología denominada PSAS (Physical space analysis system), ha sido aplicada en la NASA y en el NRL. Notar que \mathbf{z} es un N_y -vector es decir que yace en el espacio de las observaciones.

Una vez obtenido el valor óptimo $\hat{\mathbf{z}}$ hay que volver al espacio del estado, con la fórmula:

$$\delta\hat{\mathbf{x}} = \mathbf{P}^b\mathbf{H}^T\hat{\mathbf{z}}^T$$

Hemos tenido que asumir de entrada la hipótesis de linealidad en \mathcal{H} , esta es esencial para trabajar con la metodología PSAS.

Aplicación secuencial de OI - 3DVar

El problema de pronosticos es inherentemente secuencial.

El \mathbf{x}^b en un sistema de asimilación se obtiene a partir de evolucionar el modelo de la estimación previa:

$$\mathbf{x}_k^b = \mathcal{M}(\mathbf{x}_{k-1}^a)$$

Sin embargo cuando tenemos que hacer el análisis en el siguiente paso

$$p(\mathbf{x}_k | \mathbf{y}_k) \propto \exp\left\{-\frac{1}{2}\left[(\mathbf{x} - \mathbf{x}^b)^T (\mathbf{P}^b)^{-1} (\mathbf{x} - \mathbf{x}^b) + (\mathbf{y} - \mathcal{H}(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - \mathcal{H}(\mathbf{x}))\right]\right\}$$

No disminuye la incerteza del apriori cuando tenemos múltiples observaciones distribuidas en el tiempo. \mathbf{P}^b esta fijo.

Pero como hemos visto con observaciones multiples la incerteza decrece!!!

Inferencia Bayesiana secuencial

Supongamos que hacemos a partir de alguna condición inicial $p(\mathbf{x}_0)$ hacemos una primera asimilación en $k = 1$,

$$p(\mathbf{x}_1|\mathbf{y}_1) \propto p(\mathbf{y}_1|\mathbf{x}_1)p(\mathbf{x}_1)$$

donde $p(\mathbf{x}_1) = \int p(\mathbf{x}_1|\mathbf{x}_0)p(\mathbf{x}_0)d\mathbf{x}_0$.

La covarianza de $p(\mathbf{x}_1|\mathbf{y}_1)$ es menor a la de $p(\mathbf{x}_1)$. Por lo que para realizar la próxima asimilación en $k = 2$. lo que nos conviene es realizar una evolución a partir de $p(\mathbf{x}_1|\mathbf{y}_1)$,

$$p(\mathbf{x}_2|\mathbf{y}_1) = \int p(\mathbf{x}_2|\mathbf{x}_1)p(\mathbf{x}_1|\mathbf{y}_1)d\mathbf{x}_1$$

Luego asimilamos haciendo Bayes:

$$p(\mathbf{x}_2|\mathbf{y}_2, \mathbf{y}_1) \propto p(\mathbf{y}_2|\mathbf{x}_2)p(\mathbf{x}_2|\mathbf{y}_1)$$

De esta forma estamos usando como densidad a priori en el segundo tiempo

Inferencia Bayesiana secuencial

una que incorporó **toda** la información de la observación en el tiempo $k = 1$.

En forma recursiva se obtiene para el paso k -ésimo

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

donde $p(\mathbf{x}_k | \mathbf{y}_{1:k-1})$ es la denominada forecast/prediction density:

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

De esta manera obtenemos la **solución completa del modelo de Markov escondido**.

En la práctica determinar la densidad de predicción es muy complicado.

Pero veamos algunas simplificaciones del esquema secuencial que nos permitan ir mas alla del OI/3dVar.

Modelo lineal

Consideremos el sistema dinámico lineal estocástico y operador observacional lineal

$$\mathbf{x}_{k+1} = \mathbf{M}_k \mathbf{x}_k + \boldsymbol{\eta}_k,$$

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \boldsymbol{\nu}_k,$$

- \mathbf{x}_k es el estado del modelo en el tiempo t_k , N_x -dim.
- $\boldsymbol{\eta}_k$ es el error del modelo $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$.
- \mathbf{y}_k es N_y -vector de la medición.
- $\boldsymbol{\nu}_k$ es m-vector del error de la medición $\boldsymbol{\nu} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$.
- \mathbf{M}_k es la $N_x \times N_x$ matriz de transición, i.e. **modelo lineal**.
- \mathbf{H}_k es el operador **lineal** de observación, $N_y \times N_x$.

Hipótesis sobre las variables

Las variables aleatorias, errores del modelo y la observación, son Gaussianos de media 0, es decir no-sesgados (unbiased)

$$\mathcal{E}\{\boldsymbol{\eta}_k\} = 0$$

$$\mathcal{E}\{\boldsymbol{\nu}_k\} = 0$$

Por otro lado asumimos son independientes entre sí y no-correlacionadas temporalmente:

$$\mathcal{E}\{\boldsymbol{\nu}_k(\boldsymbol{\eta}_l)^\top\} = 0$$

$$\mathcal{E}\{\boldsymbol{\eta}_k(\boldsymbol{\eta}_l)^T\} = \mathbf{Q}_k \delta_{lk}$$

$$\mathcal{E}\{\boldsymbol{\nu}_k(\boldsymbol{\nu}_l)^T\} = \mathbf{R}_k \delta_{lk}^k$$

Evolución modelo lineal

Supongamos $p(\mathbf{x}_0) \propto \exp(-\frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}^b\|_{\mathbf{P}_0^b})$, la densidad en $k = 1$ será

$$p(\mathbf{x}_1) = \int p(\mathbf{x}_1|\mathbf{x}_0)p(\mathbf{x}_0)\mathrm{d}x_0$$

como la matriz de transición es lineal, $p(\mathbf{x}_1)$ es Gaussiana,

$$p(\mathbf{x}_1) \propto \exp(-\frac{1}{2}\|\mathbf{x}_1 - M_1\mathbf{x}^b\|_{\mathbf{P}_1^f})$$

donde

$$\bar{\mathbf{x}}_1^f = \mathbf{M}_1\mathbf{x}^b \quad \text{Media de las evoluciones!}$$

$$\mathbf{P}_1^f = \mathbf{M}_1^\top \mathbf{P}_0^b \mathbf{M}_1 + \mathbf{Q}_1 \quad \text{Incerteza del pronóstico.}$$

NOTA: $k = 0, 1$ se refiere al número de ciclo y no tiene nada que ver con los pasos de tiempo del modelo. M_1 representa la evolución con el modelo del tiempo t_0 a t_1 .

Asimilación de la prediction density en $k = 1$

Conociendo la densidad de predicción en $k = 1$ podemos asimilarla con la observaciones en ese tiempo:

$$p(\mathbf{x}_1|\mathbf{y}_1) \propto p(\mathbf{y}_1|\mathbf{x}_1)p(\mathbf{x}_1)$$

Esto es lo mismo que hicimos en OI pero ahora en lugar de la background density tenemos la prediction density. El resultado es:

$$p(\mathbf{x}_1|\mathbf{y}_1) \propto \exp\left(-\frac{1}{2}\|\mathbf{x}_1 - \mathbf{x}_1^a\|_{\mathbf{P}_1^a}\right)$$

donde $\mathbf{x}_1^a = \mathbf{x}_1^f + \mathbf{K}_1(\mathbf{y}_1 - \mathbf{H}_1\mathbf{x}_1^f)$,

$$\mathbf{P}_1^a = (\mathbf{I} - \mathbf{K}_1\mathbf{H}_1)\mathbf{P}_1^f$$

$$\mathbf{K}_1 = \mathbf{P}_1^f\mathbf{H}_1^\top(\mathbf{R}_1 + \mathbf{H}_1\mathbf{P}_1^f\mathbf{H}_1^\top)$$

Asimilación en k

Para pasos siguientes lo único que se debe hacer es usar la media y covarianza estimadas en el paso anterior.

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^f)$$

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f$$

$$\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^\top (\mathbf{R}_k + \mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^\top)^{-1}$$

Hemos obtenido el filtro de Kalman recursivo.

Asumimos que no hay sesgo en el modelo, ni tampoco en el análisis entonces el pronóstico es insesgado.

Si el modelo es insesgado pero el análisis tiene sesgo (\mathbf{R}_k , \mathbf{Q}_k incorrectas) el pronóstico será sesgado.

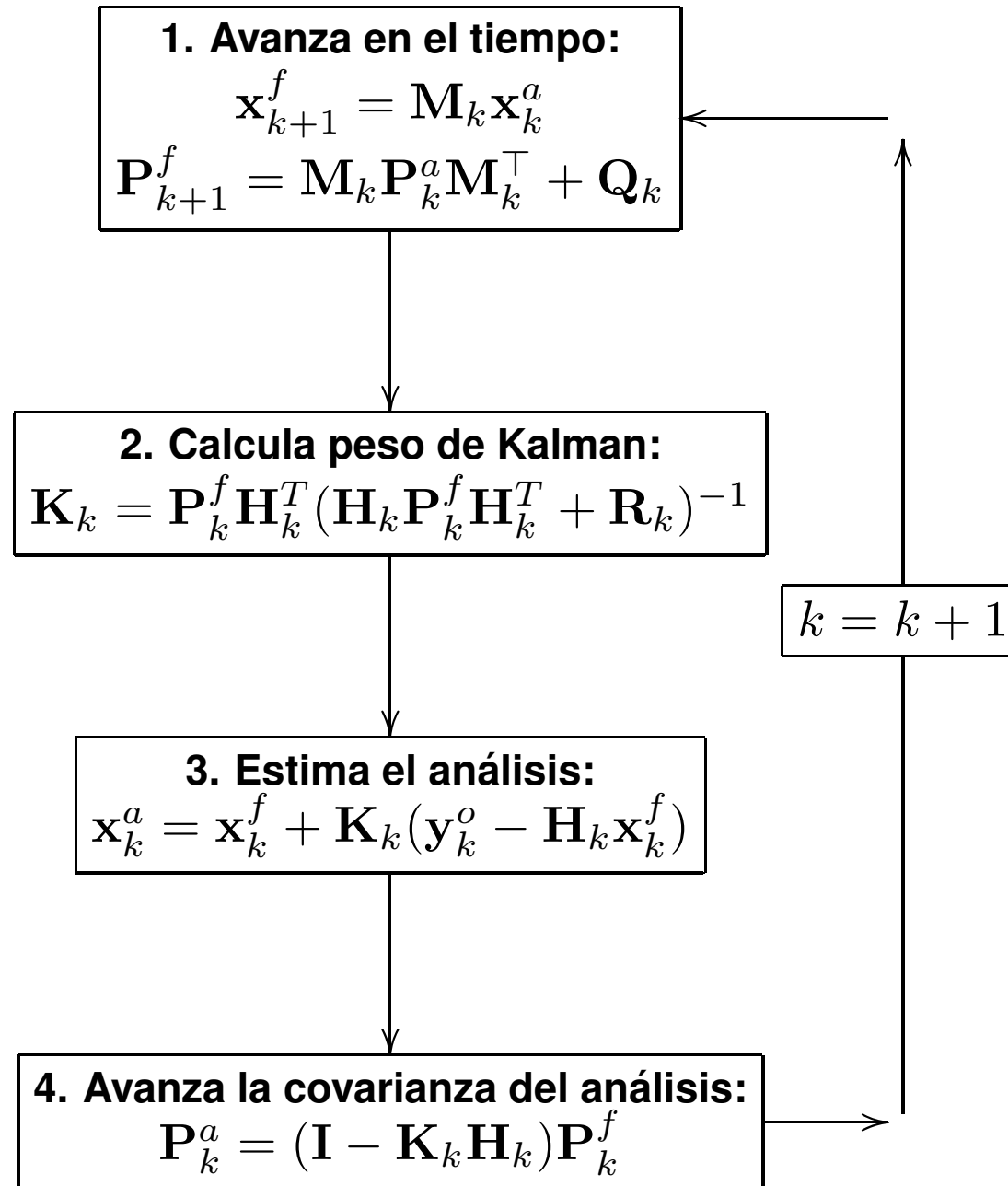
Las covarianzas del error observacional y del error de modelo tienen un rol fundamental en la asimilación, es esencial que sean precisas. En la realidad, es difícil conocerlas.

6. Filtro de Kalman lineal

Pasos del filtro

Ciclo del filtro de Kalman

Dados x_0^a y P_0^a



Evolución de la covarianza del pronóstico

$$\begin{aligned}\mathbf{P}_{k+1}^f &= \mathcal{E}\{\mathbf{e}_{k+1}^f (\mathbf{e}_{k+1}^f)^T\} \\ &= \mathbf{M}\mathbf{P}_k^a\mathbf{M}_k^T + \mathbf{Q}_k\end{aligned}$$

La incerteza del pronóstico esta compuesta por dos términos, el error propio del modelo \mathbf{Q}_k y la amplificación del error del análisis (del paso de tiempo previo) a través de la dinámica del modelo $\mathbf{M}\mathbf{P}_k^a\mathbf{M}_k^T$.

Esto es el error de predictibilidad; si la dinámica del modelo tiene modos inestables, entonces tendremos un crecimiento exponencial de los errores del análisis, sin embargo las observaciones son las encargadas de acotar el crecimiento de estos errores.

Estos modos inestables son esenciales para el funcionamiento de los filtros de Kalman por ensambles en muchas dimensiones.

Matriz de peso de varianza mínima

Tenemos que determinar cual es el peso que deben tener las observaciones con respecto al peso que se le da a los pronósticos:

$$\hat{\mathbf{x}}_k^a = \hat{\mathbf{x}}_k^f + \tilde{\mathbf{K}}_k (\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_k^f)$$

Si definimos la varianza del error del análisis

$$\begin{aligned} J_k^a &= \mathcal{E} \{ \mathbf{e}_k^a \mathbf{S}_k (\mathbf{e}_k^a)^T \} \\ &= \mathcal{E} \{ \text{Tr}(\mathbf{S}_k \mathbf{e}_k^a (\mathbf{e}_k^a)^T) \} \\ &= \text{Tr}(\mathbf{S}_k \mathbf{P}_k^a) \end{aligned}$$

donde \mathbf{S}_k es una matriz de scaling positiva definida.

Lo que queremos entonces es determinar el $\tilde{\mathbf{K}}_k$ que produce la mínima función de costo, leáse mínima varianza del análisis.

$$\frac{dJ}{d\tilde{\mathbf{K}}_k} = \mathbf{S}_k \left[-2\mathbf{H}_k \mathbf{P}_k^f (\mathbf{I} - \tilde{\mathbf{K}}_k \mathbf{H}_k)^T + 2\mathbf{R}_k \tilde{\mathbf{K}}_k^T \right]$$

Matriz de peso de varianza mínima

Pidiendo que la derivada sea 0, se obtiene que la matriz de peso óptima viene dada por

$$\tilde{\mathbf{K}}_k = \mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1}$$

Para el $\tilde{\mathbf{K}}_k$ óptimo se tiene que

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f$$