
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/curso/>

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

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

Inferencia Bayesiana secuencial

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:

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

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

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

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

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

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 Ol 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.

$$\begin{aligned}\mathbf{x}_k^a &= \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k^f) \\ \mathbf{P}_k^a &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^f \\ \mathbf{K}_k &= \mathbf{P}_k^f\mathbf{H}_k^\top(\mathbf{R}_k + \mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^\top)^{-1}\end{aligned}$$

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.

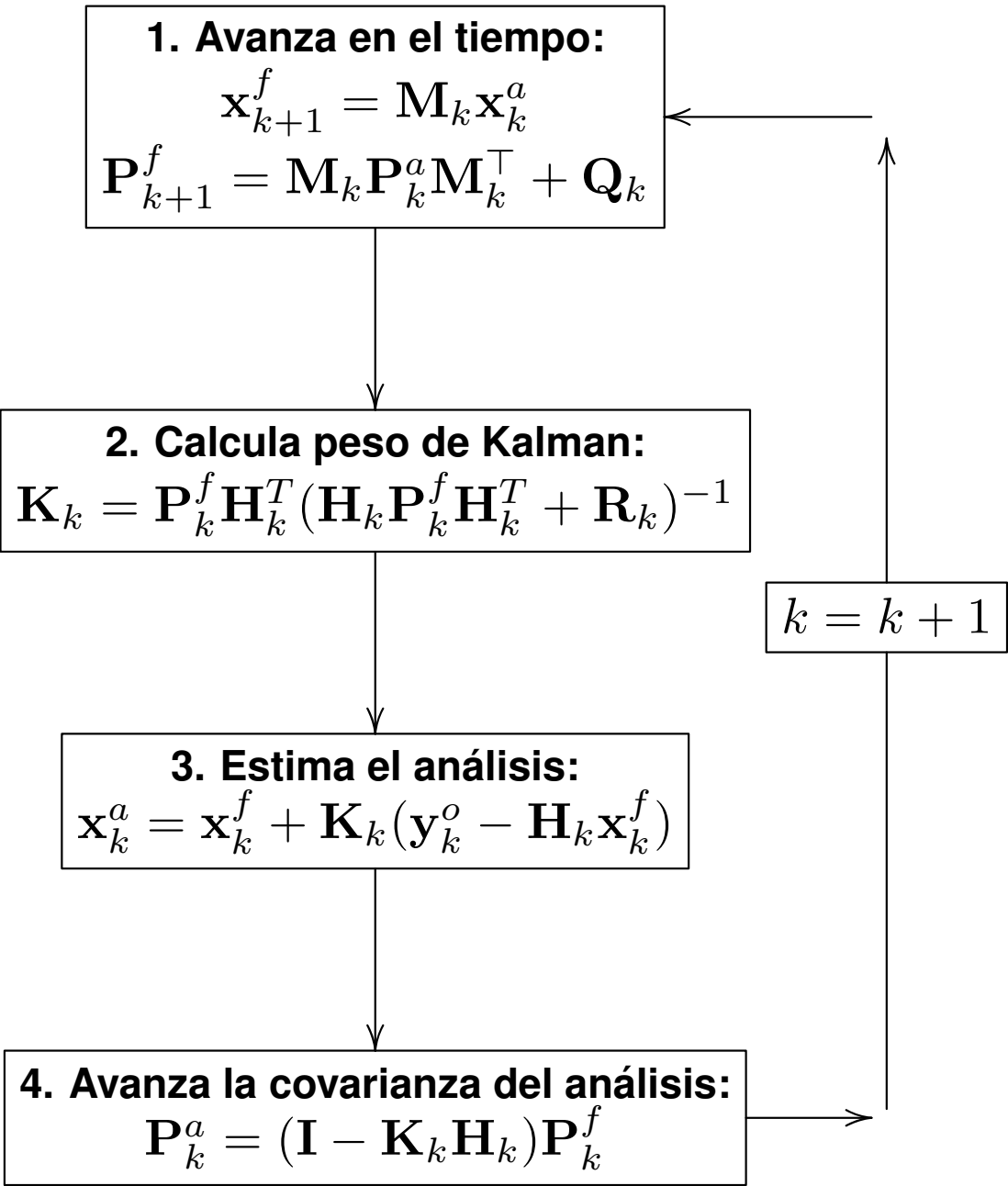
Si el modelo es insesgado pero el análisis tiene sesgo (\mathbf{R}_k , \mathbf{Q}_k incorrectas) el pronóstico será sesgado.

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
Datos x_0^a y P_0^a



Evolución de la covarianza del pronóstico

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

El error de pronóstico esta compuesto 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

Hasta el momento hemos definido la evolución de las variables sin embargo aun no hemos determinado cual es el peso que deben tener las observaciones con respecto al peso que se le da a los pronósticos. Es decir debemos determinar el peso de la innovación en:

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

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

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

6. Asimilación variacional de 4 dimensiones (4DVar)

Asimilación variacional de 4 dimensiones

Pensemos en un problema donde **las observaciones estan distribuidas no solo espacialmente pero también temporalmente**. Si el modelo y las observaciones son perfectas lo que buscaríamos entonces es que la trayectoria en el espacio-tiempo sea lo mas cercana posible a las observaciones.

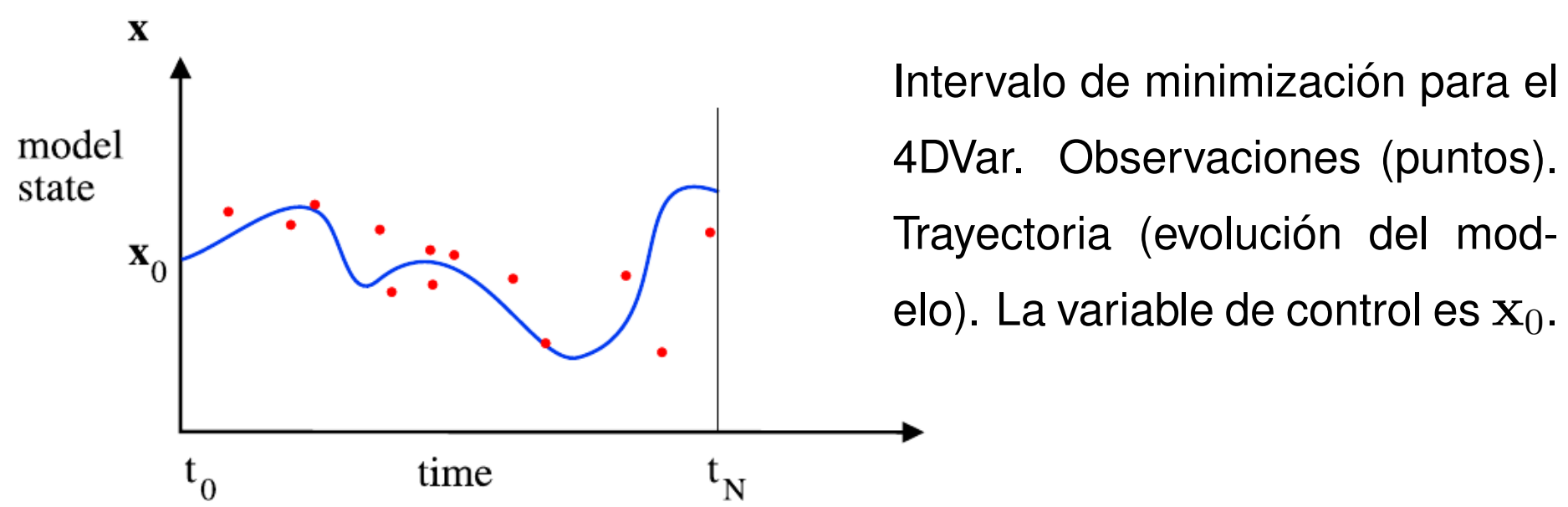
La condición inicial que nos permita evolucionar al modelo lo mas cerca posible de las observaciones es **la condición inicial óptima**.

Pongamos esto en fórmulas, sea una función de costo dada por

$$J(\mathbf{x}_0) = \frac{1}{2} \int_0^T \langle (\mathbf{x}(t, \mathbf{x}_0) - \mathbf{y}(t)) | (\mathbf{x}(t, \mathbf{x}_0) - \mathbf{y}(t)) \rangle dt$$

$\mathbf{y}(t)$ son las observaciones continuas en el tiempo, realizadas en el/los mismos puntos del estado. $\langle | \rangle$ representa un producto interno en algun espacio.

Modelo adjunto



En realidad la función de costo es una función de las condiciones iniciales, los parámetros del modelo y las condiciones de contorno. Por el momento, asumimos conocidos los parámetros y las condiciones de contorno.

Pulido and Thuburn (2005) asumen conocida la condición inicial y condiciones de contorno y estiman los parámetros del modelo.

Modelo adjunto

Para minimizar la J la información esencial es la sensibilidad de J a perturbaciones de \mathbf{x}_0 . Si se realiza un cambio en la condición inicial el cambio concomitante en la J viene dado por la derivada direccional

$$\delta J = \langle \Delta_{\mathbf{x}_0} J | \delta \mathbf{x}_0 \rangle$$

La información que necesitamos para cuantificar la sensibilidad es el **gradiente de la función de costo en función de la condición inicial**.

La forma mas trivial de calcular el gradiente es a través de diferencias finitas esto es

$$\frac{\partial J}{\partial x_{0i}} = \frac{J(\mathbf{x}_0 + \alpha \mathbf{e}_i) - J(\mathbf{x}_0)}{\alpha}$$

donde \mathbf{e}_i es una de las n direcciones que forman el espacio del estado. Es decir, que necesitamos correr $n + 1$ veces el modelo para poder determinar a J $n = 10^7$. Totalmente prohibitivo.

La alternativa es el modelo adjunto cuyo costo es en teoría de 2 veces el modelo forward (en la práctica puede llegar a unas 5 veces). [No se debe confundir con el modelo inverso el cual no existe para los procesos atmosféricos/ya que no son reversibles.].

Modelo adjunto

LeDimet and Talagrand, 1986. Minimización de J . La variación de la función de costo

$$\delta J = \frac{1}{2} \int_0^T (\langle \bar{\mathbf{x}} + \delta \mathbf{x} - \mathbf{y} | \bar{\mathbf{x}} + \delta \mathbf{x} - \mathbf{y} \rangle - \langle \bar{\mathbf{x}} - \mathbf{y} | \bar{\mathbf{x}} - \mathbf{y} \rangle) dt$$

$$\delta J = \int_0^T \langle \delta \mathbf{x} | \bar{\mathbf{x}} - \mathbf{y} \rangle dt$$

donde se asume que $\mathcal{E}\{|\cdot|\}$ es conmutativo.

Supongamos un vector arbitrario multiplicado por la evolución de la perturbación

$$\int_0^T \left\langle \frac{d}{dt} \delta \mathbf{x} | \mathbf{p} \right\rangle dt = \langle \delta \mathbf{x}(T) | \mathbf{p}(T) \rangle - \langle \delta \mathbf{x}(0) | \mathbf{p}(0) \rangle - \int_0^T \left\langle \delta \mathbf{x} | \frac{d}{dt} \mathbf{p} \right\rangle dt$$

donde se uso integración por partes. Tomemos $\mathbf{p}(T) = 0$ y usando la aproximación lineal

$$\int_0^T \langle \mathbf{M} \delta \mathbf{x} | \mathbf{p} \rangle dt = - \langle \delta \mathbf{x}(0) | \mathbf{p}(0) \rangle - \int_0^T \left\langle \delta \mathbf{x} | \frac{d}{dt} \mathbf{p} \right\rangle dt$$

Si denotamos el operador adjunto por \mathbf{M}^* tenemos que: $\langle \mathbf{x} | \mathbf{M} \mathbf{y} \rangle = \langle \mathbf{M}^* \mathbf{x} | \mathbf{y} \rangle$

$$- \langle \delta \mathbf{x}(0) | \mathbf{p}(0) \rangle = \int_0^T \left\langle \delta \mathbf{x} | \mathbf{M}^* \mathbf{p} + \frac{d}{dt} \mathbf{p} \right\rangle dt$$

Modelo adjunto

Si exigimos que se cumpla la ecuación

$$\mathbf{M}^* \mathbf{p} + \frac{d}{dt} \mathbf{p} = \bar{\mathbf{x}} - \mathbf{x}$$

junto con la condición final $p(T) = 0$ entonces se tiene que

$$\delta J = - \langle \delta \mathbf{x}(0) | \mathbf{p}(0) \rangle$$

Del teorema del gradiente sabemos que la variaciones de J pueden ser expresadas por

$$\delta J = -\mathcal{E}\{\nabla_{\mathbf{x}_0} J | \delta \mathbf{x}_0\}$$

Luego se tiene que

$$\nabla_{\mathbf{x}_0} J = -\mathbf{p}(0)$$

Es decir que el gradiente de la función de costo viene dado por la solución de la ecuación

$$\mathbf{M}^* \mathbf{p} + \frac{d}{dt} \mathbf{p} = \bar{\mathbf{x}} - \mathbf{x}$$

integrada desde $t = T$ donde $p(T) = 0$ hasta $t = 0$.

Esta es la ecuación del modelo adjunto.

Detalles del modelo adjunto

La derivación del modelo adjunto \mathbf{M}^* , viene de tomar el TLM del modelo no-lineal y luego a esta matriz se calcula su adjunta.

La variable \mathbf{p} que es integrada desde $t = T$ el tiempo final al tiempo inicial $t = 0$ es denominada variable/s adjunta/s.

Notar que para el cálculo del adjunto también necesitamos el cálculo del modelo no-lineal hacia adelante ya que

$$\frac{d}{dt}\mathbf{x} = \mathbf{M}(\mathbf{x})$$

Para obtener el mínimo se debe cumplir que $\nabla_{\mathbf{x}_0} J = 0$ en este caso se da que $p(0) = 0$.

Precaución: mas alla de que la integración del modelo adjunto es hacia atrás, no se debe confundir el modelo adjunto con el modelo inverso.

Algoritmo de 4DVar

Volviendo a la asimilación de $n = 10^7$, discretizamos espacio y tiempo, y asumamos ahora que existen errores tanto en el background como en las observaciones, luego la función de costo viene dada por dos términos:

$$J(\mathbf{x}_0) = \frac{1}{2}[\mathbf{x}_0 - \mathbf{x}_0^f]^T (\mathbf{P}^f)^{-1} [\mathbf{x}_0 - \mathbf{x}_0^f] + \frac{1}{2} \sum_i [\mathbf{y}_i - H(\mathbf{x}_i)]^T \mathbf{R}^{-1} [\mathbf{y}_i - H(\mathbf{x}_i)]$$

donde x_0 es la condición inicial y variable de control.

La evolución del modelo viene dada por

$$\mathbf{x}_i = M(\mathbf{x}_{i-1}).$$

En una comparación a los algoritmos que hemos introducido, notar que 4DVar permite introducir mediciones a distintos tiempos y éstas se pueden utilizar con restricciones “dinámicas” (por el modelo) para determinar el estado \mathbf{x}_0 .

Algoritmo de 4DVar

Calculemos el gradiente de J desde $k = 0$ a $k = N$

$$\delta J = \sum_{k=0}^N \langle \nabla_{\mathbf{x}_k} J | \delta \mathbf{x}_k \rangle$$

Aplicando sistemáticamente el tangente lineal

$$\delta J = \sum_{k=0}^N \langle \nabla_{\mathbf{x}_k} J | \mathbf{M}_{k-1} \cdots \mathbf{M}_0 \delta \mathbf{x}_0 \rangle$$

donde por notación tomamos

$$\mathbf{M}_k = \left. \frac{\partial M}{\partial \mathbf{x}} \right|_{x_k}$$

donde \mathbf{M}_k es el tangente lineal, evaluado en \mathbf{x}_k . El modelo es no lineal, en primer orden se cumple que

$$\delta \mathbf{x}_{k+1} = \mathbf{M}_k \delta \mathbf{x}_k$$

Aplicando el modelo adjunto

$$\delta J = \sum_{k=0}^N \langle \mathbf{M}_0^* \mathbf{M}_1^* \cdots \mathbf{M}_{k-1}^* \nabla_{\mathbf{x}_k} J | \delta \mathbf{x}_0 \rangle$$

Algoritmo de 4DVar

Teniendo en cuenta nuevamente que

$$\delta J = \langle \nabla_{\mathbf{x}_0} J | \delta \mathbf{x}_0 \rangle$$

Luego se tiene que el gradiente viene dado por

$$\nabla_{\mathbf{x}_0} J = \sum_{k=0}^N \mathbf{M}_0^* \cdots \mathbf{M}_{k-1}^* \nabla_{\mathbf{x}_k} J$$

En principio entonces habría que evaluar primero al adjunto \mathbf{M}_{k-1}^* luego aplicar $k - 2$ etc. Sin embargo todo esto lo podemos realizar usando una sola integración del adjunto.

Notar que “*” es el transpuesto complejo conjugado, por lo que si en la práctica trabajamos con reales es T .

Algoritmo de 4DVar

Para cada iteración/paso de la minimización se requiere entonces evaluar a $J(\mathbf{x}_0)$ y a $\nabla_{\mathbf{x}_0} J$, es decir se requiere de una integración del modelo (no lineal) hacia adelante y luego una integración del adjunto.

Algoritmo iterativo hacia adelante

1. Se integra el modelo desde $k = 0$ hasta N es decir: $\mathbf{x}_k = M(\mathbf{x}_{k-1})$.
2. Se calculan las innovaciones normalizadas

$$\mathbf{d}_k = \mathbf{R}_k^{-1}(\mathbf{y}_k - H_k(\mathbf{x}_k))$$

3. Se calcula la fnes de costo para cada tiempo $J_{0k} = [\mathbf{y}_k - H_k(\mathbf{x}_k)]^* \mathbf{d}_k$
4. Se suman las funciones de costo para cada tiempo $J_0 = \sum_{k=0}^N J_{0k}(\mathbf{x})$

Algoritmo de 4DVar

Para calcular el gradiente de J , la parte del background es equivalente a 3DVar.

$$J_b(\mathbf{x}_0) = (\mathbf{P}^f)^{-1}(\mathbf{x}_0 - \mathbf{x}_0^f)$$

Con respecto a la parte de las observaciones de la función de costo comenzamos derivando la ecuación directamente $\nabla_{\mathbf{x}_k} J = 2\mathbf{H}_k^* \mathbf{d}_k$ luego evolucionamos hacia atrás hasta 0

$$\nabla_{\mathbf{x}_0} J = \sum_{k=0}^N \mathbf{M}_0^* \cdots \mathbf{M}_{k-1}^* (2\mathbf{H}_k^* \mathbf{d}_k)$$

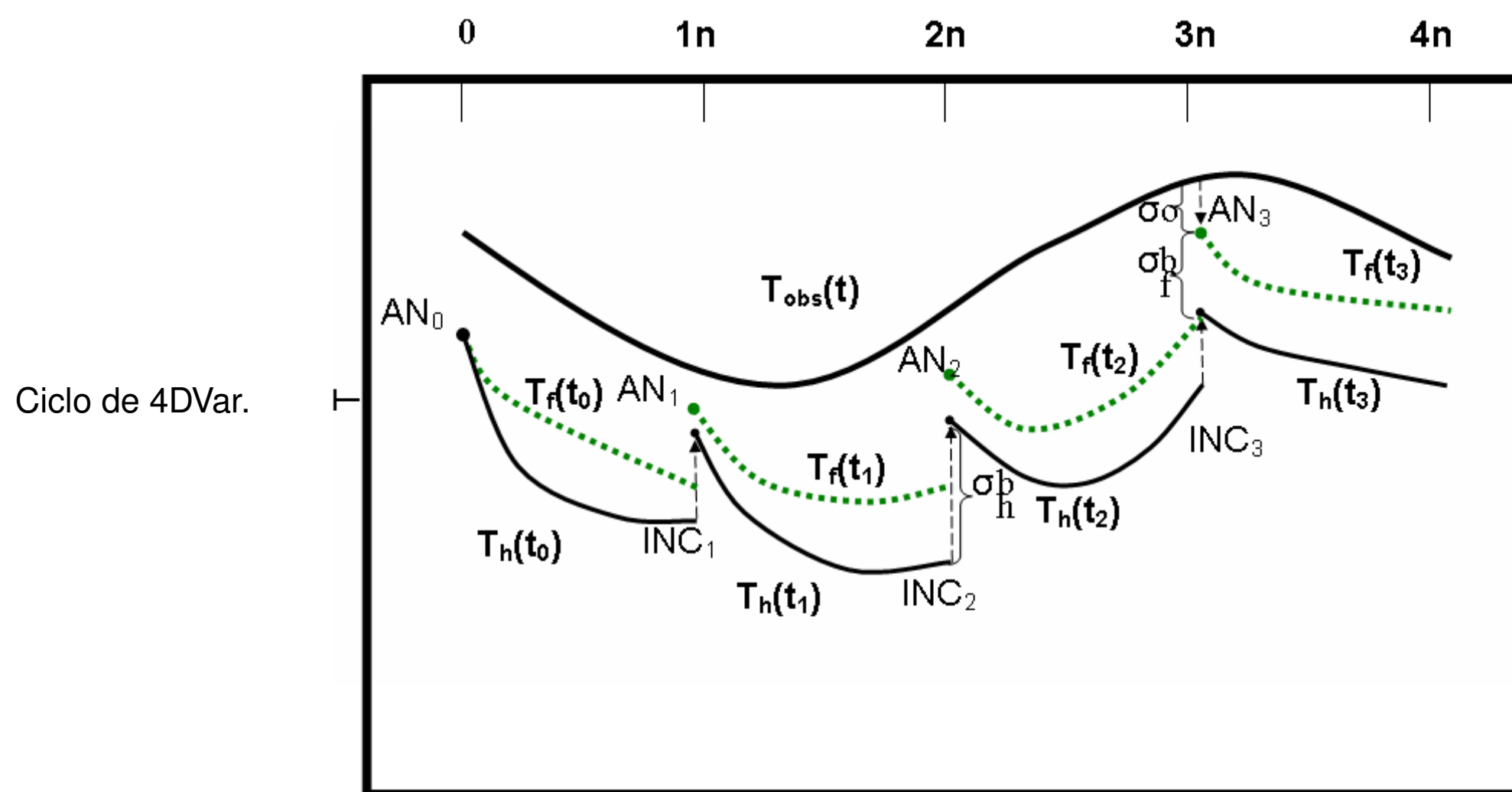
Expandiendo la suma

$$\nabla J_o = \mathbf{H}_0^T \mathbf{d}_0 + \mathbf{M}_1^T [\mathbf{H}_1^T \mathbf{d}_1 + \mathbf{M}_2^T [\mathbf{H}_2^T \mathbf{d}_2 + \cdots + \mathbf{M}_N^T 2\mathbf{H}_N^T d_N]]$$

Algoritmo iterativo hacia atrás

1. $\tilde{\mathbf{x}}_N = 0$ variable adjunta inicial
2. $\tilde{\mathbf{x}}_{k-1} = \mathbf{M}_k^T (\tilde{\mathbf{x}}_k + \mathbf{H}_k^T \mathbf{d}_k)$ (esta es la integración del modelo desde N hasta 0).
3. Al final de la recurrencia lo que tenemos es: $\tilde{\mathbf{x}}_0 = -\frac{1}{2} \nabla J_o(\mathbf{x}_0)$

Ciclo de asimilación 4DVar



T_h estimación standard sin tratamiento de error de modelo.

T_f estimación con tratamiento de error de modelo (Pulido and Thuburn, QJ 2005, QJ 2006, JC 2008)

7. Ensemble Kalman filter

Extended Kalman filter

El filtro de Kalman esta definido para un modelo lineal y operador de observacion lineal, ¿como podemos extenderlo a casos no lineales? Aproximando en series de Taylor

$$\begin{aligned}\mathbf{x}_k &= \mathcal{M}(\bar{\mathbf{x}}_{k-1}^a) + \boldsymbol{\eta}_k \approx \mathcal{M}(\bar{\mathbf{x}}_{k-1}^a) + \mathbf{M}_k(\mathbf{x}_{k-1} - \bar{\mathbf{x}}_{k-1}^a) + \boldsymbol{\eta}_k \\ &\approx \bar{\mathbf{x}}_k^f + \mathbf{M}_k(\mathbf{x}_{k-1} - \bar{\mathbf{x}}_{k-1}^a) + \boldsymbol{\eta}_k\end{aligned}$$

donde $\mathbf{M}_k = \partial \mathcal{M} / \partial \mathbf{x} |_{\bar{\mathbf{x}}_{k-1}^a}$ **asumimos** $\bar{\mathbf{x}}_k^f = \mathcal{M}(\bar{\mathbf{x}}_{k-1}^a)$.

$$\mathcal{H}(\mathbf{x}_k) \approx \mathcal{H}(\bar{\mathbf{x}}_k^a) + \mathbf{H}_k(\mathbf{x}_k - \bar{\mathbf{x}}_k^f)$$

Entonces en el análisis medio no hay ninguna diferencia,

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

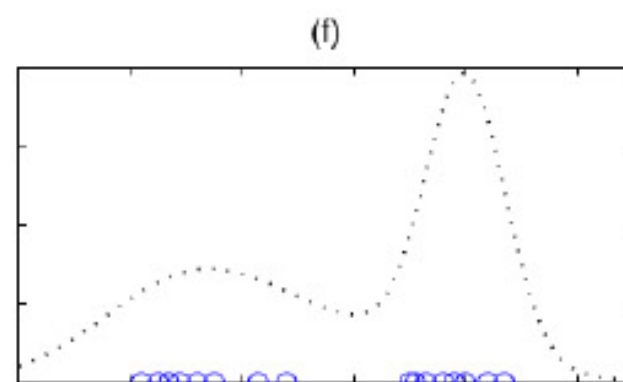
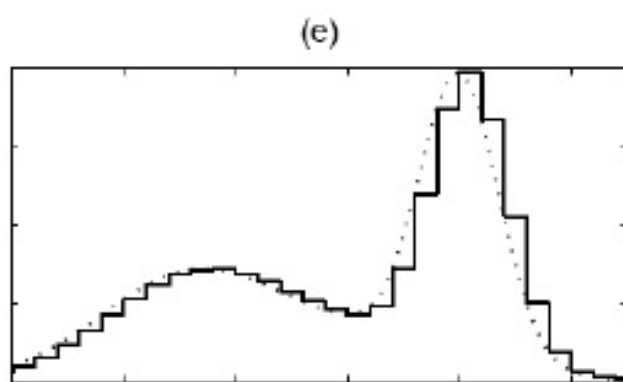
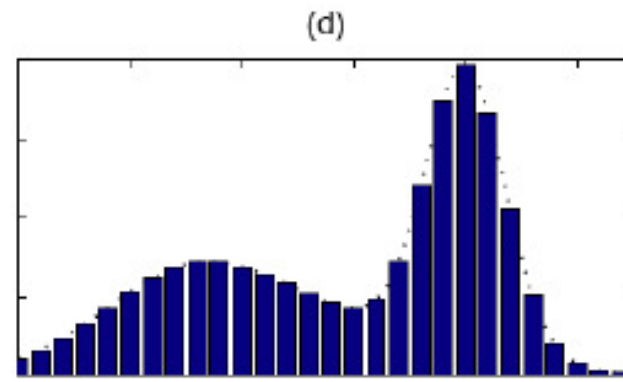
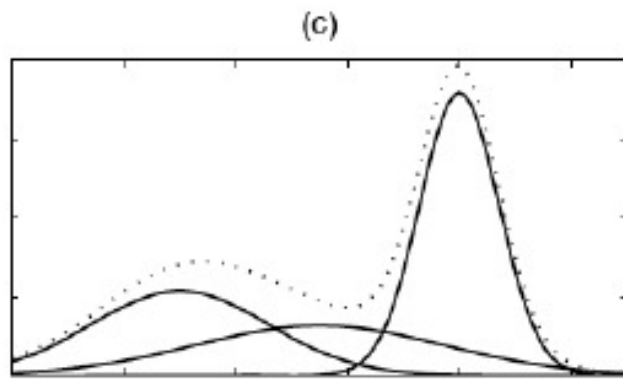
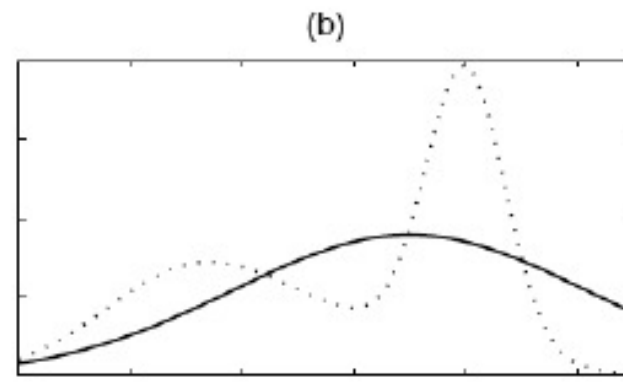
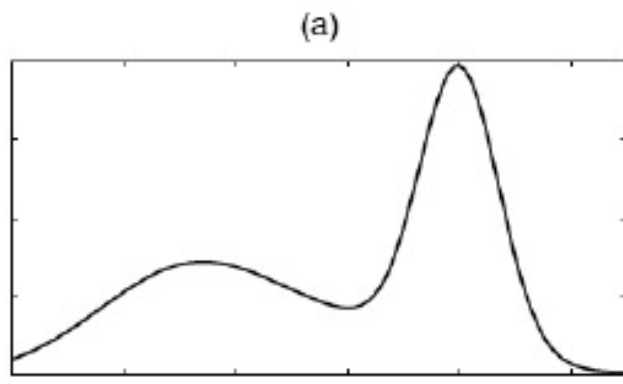
En la covarianza consideramos la dinámica lineal:

$$\mathbf{P}_k^f = \mathbf{M}_k \mathbf{P}_{k-1}^a \mathbf{M}_k^\top + \mathbf{Q}_k$$

El gran impedimento del ExtKF es que en la práctica es imposible considerar la evolucion de \mathbf{P}_k^f , es $10^7 \times 10^7$.

Muestreo de Monte Carlo

Formas de representar a una densidad de probabilidad.



Muestreo de Monte Carlo

En general uno puede representar una densidad de probabilidad por un conjunto muestras aleatorias iid (un ensamble). Todas las cantidades derivadas a partir de la densidad:

$$\mathcal{E}(f(\mathbf{x})) = \int f(\mathbf{x})p(\mathbf{x})d\mathbf{x} \approx \sum_{j=1}^{N_e} f(\mathbf{x}_j)$$

esto es lo que se denomina una integración de Monte Carlo. El error en la estimación decae como $N_e^{-1/2}$.

En particular podemos aproximar la media por

$$\hat{\mathbf{x}}_k = \frac{1}{N_e} \sum_j^{N_e} \mathbf{x}_k^j$$

Y la covarianza viene dada por

$$\hat{\mathbf{P}}_k = \frac{1}{N_e} \sum_j^{N_e} (\mathbf{x}_k^j - \hat{\mathbf{x}}_k)(\mathbf{x}_k^j - \hat{\mathbf{x}}_k)^\top$$

conocida por sample covariance, o sample estimates (las cuales convergen para $N_e \rightarrow \infty$).

Tenemos una forma alternativa de calcular las covarianzas del forecast y el análisis!!!!

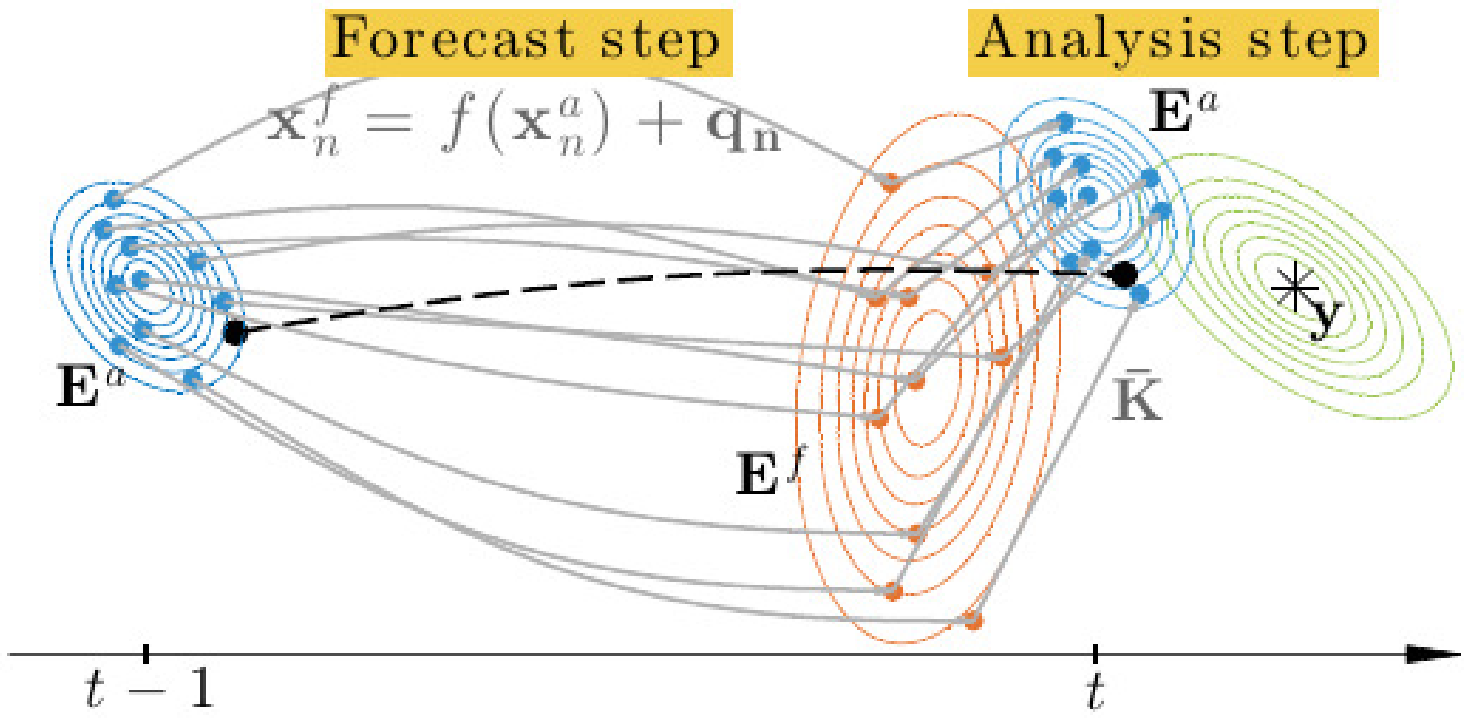
Muestras de la densidad predictiva

Supongamos un ensemble $\mathbf{x}_0^{1:N_e}$ de estados que son muestras de $p(\mathbf{x}_0)$. Para obtener la predicción density $p(\mathbf{x}_1)$ lo que hacemos es evolucionar a las muestras

$$\mathbf{x}_1^j = \mathcal{M}(\mathbf{x}_0^j) + \boldsymbol{\eta}_1^j$$

donde $\boldsymbol{\eta}_1^j$ son realizaciones del model error. Esto nos da en forma exacta (a lo Monte Carlo) la prediction density.

Filtro de Kalman por ensambles



Filtro de Kalman por ensambles estocástico

A cada miembro del ensamble lo asimilamos como en un filtro de Kalman,

$$\mathbf{x}_k^{a,j} = \mathbf{x}_k^{f,j} + \mathbf{K}_k(\mathbf{y}_k - \mathcal{H}(\mathbf{x}_k^{f,j}) - \boldsymbol{\nu}_k^j)$$

Para el caso lineal con $N_e \rightarrow \infty$ la solución converge al filtro de Kalman $\hat{\mathbf{x}}^a = \bar{\mathbf{x}}^a$ y $\hat{\mathbf{P}}^a = \bar{\mathbf{P}}^a$.

La Kalman gain viene dada por $\hat{\mathbf{K}}_k = \hat{\mathbf{P}}_k^f \mathbf{H}^\top (\mathbf{H} \hat{\mathbf{P}}_k^f \mathbf{H}^\top + \mathbf{R}_k)$

donde se esta usando la sample forecast covariance.

En el caso de \mathcal{H} no lineal. Escribiendo a las perturbaciones como columnas de una matriz:

$$\mathbf{X}^f = (N_e - 1)^{-1/2} [(\mathbf{x}_k^{f,j} - \hat{\mathbf{x}}_k^f)]_{j=1, N_e}$$

$$\mathbf{Y} = (N_e - 1)^{-1/2} [(\mathcal{H}(\mathbf{x}_k^{f,j}) + \nu_k^j - \overline{\mathcal{H}_k^{(f,m)}})]_{j=1, N_e}$$

Luego la Kalman gain es $\hat{\mathbf{K}} = \mathbf{X}^f \mathbf{Y}^\top (\mathbf{Y} \mathbf{Y}^\top)^{-1}$.

Estamos agregando en cada miembro el model error y el observacional error en forma estocastica (de alli el nombre).

Notar que en realidad para hacer el update Bayesiano estamos asumiendo que la densidad apriori es Gaussiana, pero sin embargo las muestras provienen de una densidad no-Gausiana en el caso de que el modelo sea no lineal.

Filtro de Kalman por ensambles serial

Si las observaciones no estan correlacionadas entre si, i.e. \mathbf{R} es diagonal, entonces en lugar de asimilar todos las observaciones a la vez podemos hacerlo de a una (para un tiempo fijo).

Como se vio en el frame Bayesiano, todo lo que hacemos es usar la posterior de asimilar la observacion anterior como prior density de la proxima observacion,

$$p(\mathbf{x}_k | y_k^1, \dots, y_k^{N_y}) = p(y_k^{N_y} | \mathbf{x}_k) p(\mathbf{x}_k | y_k^1, \dots, y_k^{N_y-1})$$

$$p(\mathbf{x}_k | y_k^1, \dots, y_k^{N_y-1}) = p(y_k^{N_y-1} | \mathbf{x}_k) p(\mathbf{x}_k | y_k^1, \dots, y_k^{N_y-2})$$

Lo que hacemos entonces es corregir

$$\mathbf{x}_k^{j,i} = \mathbf{x}_k^{j,i-1} + \mathbf{k}(y_k^i - h^i(\mathbf{x}_k^{j,i-1}) - \nu_k^{j,i})$$

donde $\mathbf{k} = \mathbf{P}_{xy}^{i-1} / P_{yy}^{i-1}$ con $\mathbf{P}_{xy}^{i-1} = \frac{1}{N_e-1} \sum_j (\mathbf{x}_k^{j,i-1} - \hat{\mathbf{x}}_k^{i-1})(h^i(\mathbf{x}_k^{j,i-1}) - \hat{h}^{i-1})$ y $P_{yy}^{i-1} = \sigma_R^2 + \frac{1}{N_e-1} \sum_j (h^i(\mathbf{x}_k^{j,i-1}) - \hat{h}^{i-1})^2$.

EnKFs determinísticos. ETKF

Filtros de raíz cuadrada. Satisfacen la media y covarianza del análisis sin términos estocásticos.

Hunt 2007. Transforma al espacio de las perturbaciones.

$$\hat{\mathbf{x}}^a = \hat{\mathbf{x}}^f + \mathbf{X}^f \bar{\mathbf{w}}^a, \quad \hat{\mathbf{P}}^a = \mathbf{X}^f \tilde{\mathbf{P}}^a (\mathbf{X}^f)^\top$$

En el espacio de las perturbaciones, la media y covarianza vienen dadas por

$$\begin{aligned} \bar{\mathbf{w}}^a &= \tilde{\mathbf{P}}^a (\mathbf{Y}^f)^\top \mathbf{R}^{-1} [\mathbf{y} - \bar{\mathbf{y}}^f], \\ \tilde{\mathbf{P}}^a &= [(N_e - 1)\mathbf{I} + (\mathbf{Y}^f)^\top \mathbf{R}^{-1} \mathbf{Y}^f]^{-1}. \end{aligned}$$

Cada uno de los miembros en el espacio físico viene dado por

$$\mathbf{x}^{a,j} = \hat{\mathbf{x}}^f + \mathbf{X}^f \mathbf{w}^{a,j}.$$

con $\mathbf{w}^{a,j}$ de la matriz

$$\mathbf{W}^a = (N_e - 1)^{1/2} \left(\tilde{\mathbf{P}}^a \right)^{1/2}$$

Divergencia del filtro

Existen dos fuertes restricciones en el EnKF que hacen que \mathbf{P}^a no sea correcto y finalmente llevan a la divergencia del filtro.

Divergencia del filtro: Si se subestima el \mathbf{P}^f esto lleva luego a una subestimación de \mathbf{P}^a la cual afectará al ciclo siguiente. Debido a este efecto recursivo en algún momento la varianza del forecast se vuelve menor que la de la observación. El filtro cree por demás en los pronósticos y no cree en las observaciones. De esta forma divergiendo totalmente.

En general el EnKF se usa con $\mathbf{Q} = 0$ por desconocerse el modelo de error. Este es uno de los problemas que lleva a la divergencia del filtro.

Nuestro grupo ha dedicado importantes esfuerzos para el desarrollo de técnicas que permitan estimar \mathbf{Q} y \mathbf{R} a partir de observaciones. Algoritmo EM (Tandeo et al QJ 2015; Dreano et al QJ 2017), filtros de Kalman jerárquicos (Scheffler et al 2018), estimación por ML (Pulido et al Tellus 2018) etc.

Deficiencia de rango:

Sampling errors.

Factor de inflación

Una forma que se utiliza para evitar la divergencia del filtro es inflar la covarianza del pronóstico.

Suponiendo $\alpha > 1$, las perturbaciones del pronóstico las aumentamos

$$\mathbf{X}^f = \alpha^{1/2} \mathbf{X}^f$$

antes de asimilar. De esta manera $\mathbf{P}^f = \alpha \mathbf{X}^f (\mathbf{X}^f)^\top$

Se puede tomar un factor constante, 1.05 a 1.40 depende la severidad del model error y del undersampling.

También existen métodos adaptativos que van calculando el factor de inflación en cada tiempo (e.g. Miyoshi, MWR, 2011).

Localización

Las correlaciones espaciales de la \mathbf{P}^f deberían decaer exponencialmente a 0 con la distancia. Sin embargo debido al sampling error aparecen correlaciones espurias y no decae a 0.

Producto de Hadamard entre una función de correlación suave ρ y la sample covariance $\hat{\mathbf{P}}^f$,

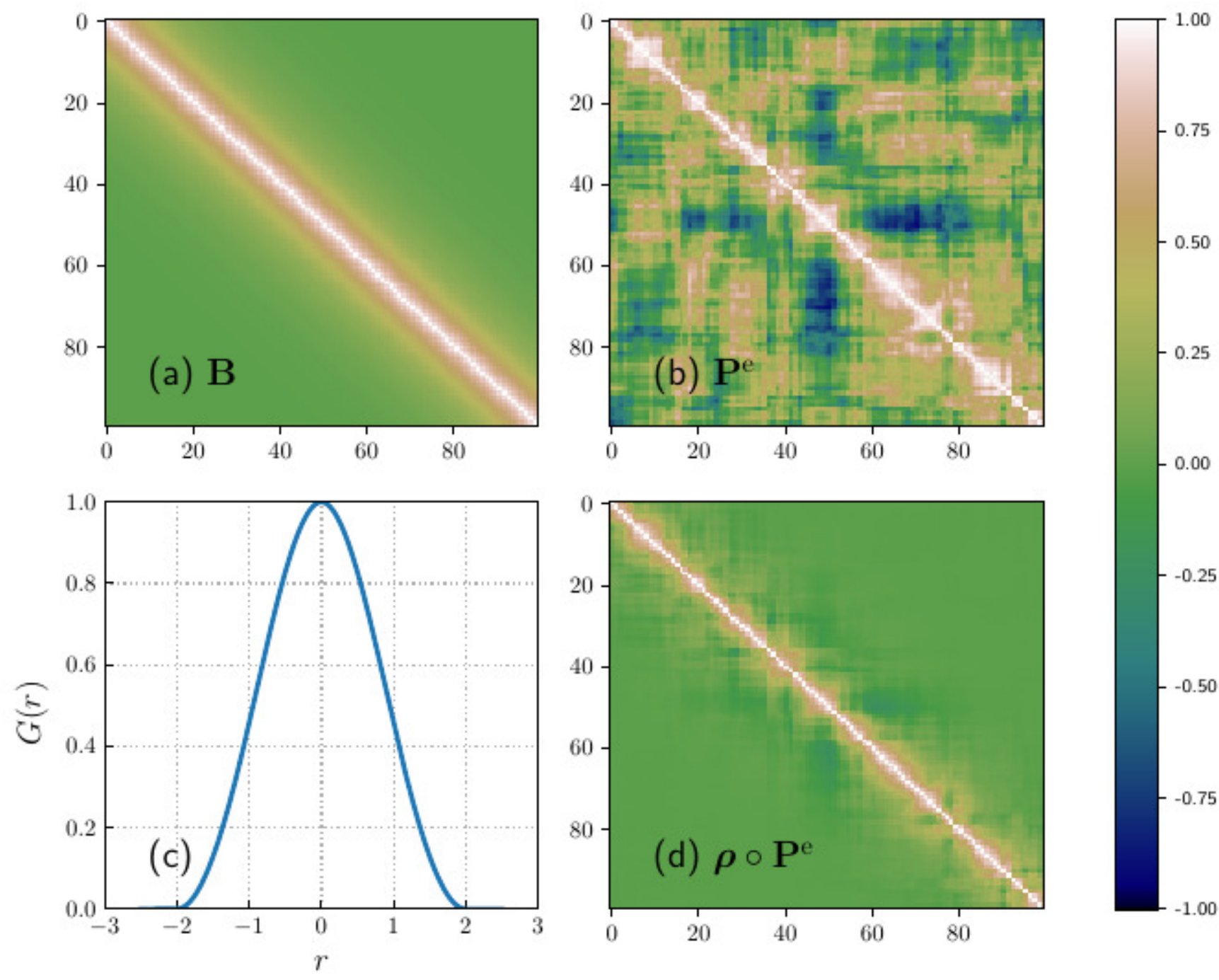
$$\tilde{\mathbf{P}}^f = [\rho \circ \hat{\mathbf{P}}^f]_{ij} = \rho_{ij} \hat{\mathbf{P}}^f_{ij}$$

Esto hace que la covarianza decaiga como lo hace la ρ .

En general se suele elegir la función de Gaspari-Cohn MWR (1999), es una función que decae exponencialmente con una escala de longitud c pero que tiene un radio de corte para distancias grandes (con respecto a la escala $r > 2c$).

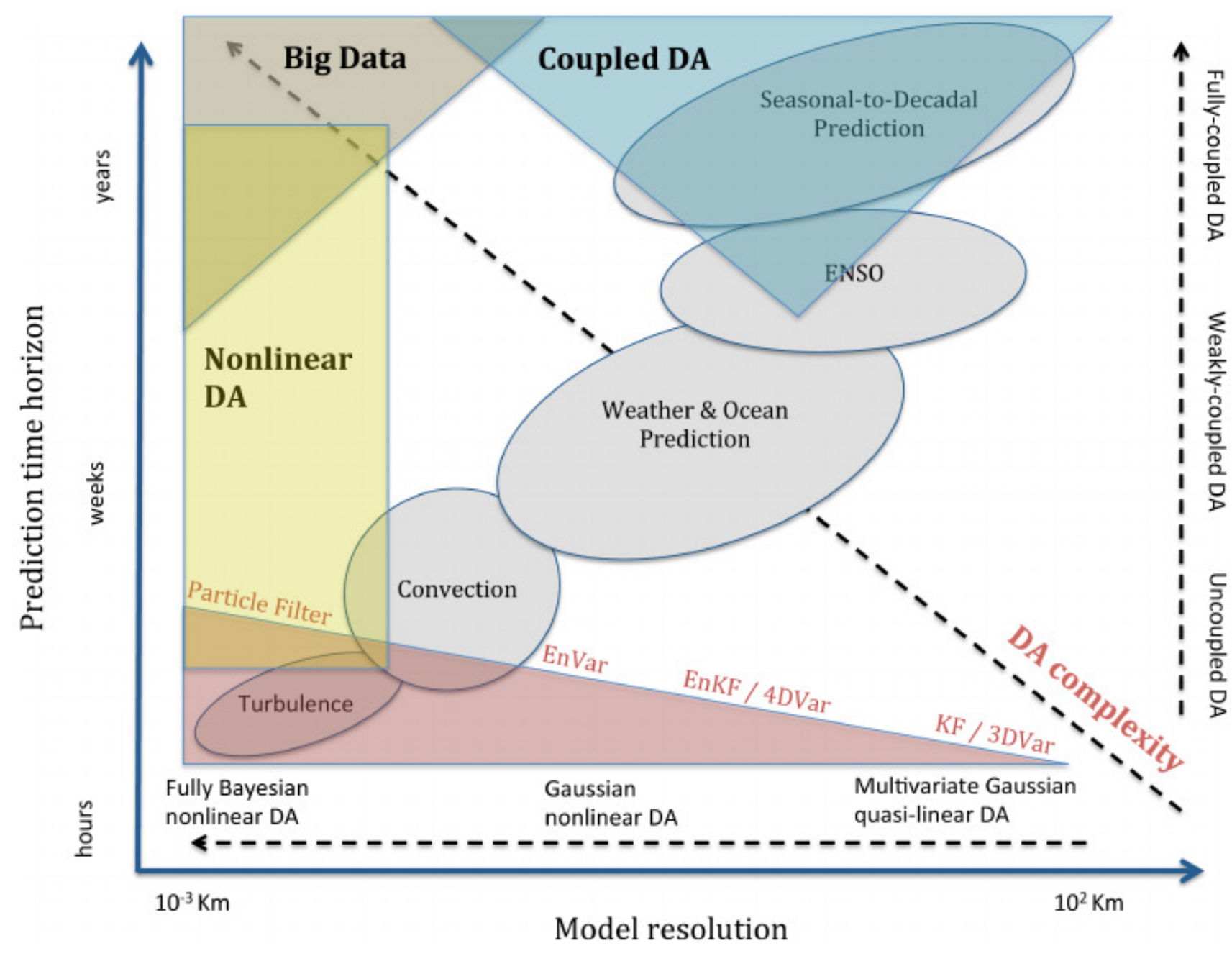
Localización del dominio. Se asimila en cada punto y se toman observaciones en la cercanía del punto.

Localización



Localización de la covarianza con Gaspari-Cohn.

Hacia donde vamos?



Bayes con Monte Carlo sampling

Las técnicas de Monte Carlo consisten en producir un conjunto de muestras independientes de una dada distribución (iid).

Supongamos muestras o partículas de la densidad posterior $p(\mathbf{x}_{k-1} | \mathbf{x}_{1:k-1})$,

$x_{k-1}^{(1:N_p)}$ con sus correspondientes pesos $w_{k-1}^{(1:N_p)}$.

Para determinar a la densidad de predicción se evoluciona con el modelo $\mathbf{x}_k^{(j)} = \mathcal{M}(\mathbf{x}_{k-1}^{(j)}, \boldsymbol{\nu}_k^j)$

Para la transición se toma una sola realización. Terminamos con N_p muestras: $x_k^{1:N_p}$

La densidad de predicción es dada por:

$$p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) \approx \sum_{j=1}^{N_p} w_k^j \delta(\mathbf{x} - \mathbf{x}_k^{(j)})$$

Filtros de partículas

Usando principios de importance sampling, la densidad de predicción se toma como **densidad proposal**.

Luego usamos Bayes con esa proposal:

$$p(\mathbf{x}_k | \mathbf{y}_{1:k}) \propto p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) = \sum_{j=1}^{N_p} w_{k-1}^{(j)} p(\mathbf{y}_k | \mathbf{x}_k^{(j)}) \delta(\mathbf{x}_k - \mathbf{x}_k^{(j)})$$

entonces tenemos una formula recursiva para los pesos de cada partícula.

$$\tilde{w}_k^{(j)} = \tilde{w}_{k-1}^{(j)} p(\mathbf{y}_k | \mathbf{x}_k^{(j)})$$

luego tenemos que normalizarlos:

$$w_k^{(j)} = \tilde{w}_k^{(j)} / \sum_l \tilde{w}_k^{(l)}.$$

Filtros de partículas

Remuestreo

El metodo no funciona por un largo tiempo, en unos pocos ciclos el SIS termina con una sola particula con peso 1 y el resto 0 (degeneración del filtro).

Para solucionar este problema:

Filtros de partículas

Las partículas se remuestran con reemplazo considerando los pesos. Se generan copias con una probabilidad dada por los pesos.

A partir de partículas con pesos altos, mientras partículas con pesos bajos son eliminadas (probabilísticamente).

El remuestreo trae consigo otro problema: empobrecimiento del muestreo.

