Minicurso de: Asimilación de datos

Guía de trabajos computacionales

M. Pulido y T. Cocucci

25 de Setiembre, 2018

Problema 1: [1] La correlación espacial de la dinámica del sistema físico es esencial para poder utilizar la información observacional de un punto del espacio en otros puntos. Suponga tenemos un sistema de una dimensión que tiene una correlación espacial del background de forma SOAR (Second order Auto Regression) que es la forma que se tiene en sistemas isotrópicos,

$$\rho(\Delta x) = \left(1 + \frac{|\Delta x|}{L_d}\right) \exp\left(-\frac{|\Delta x|}{L_d}\right) \tag{1}$$

donde $|\Delta x|$ es la distancia entre dos puntos espaciales y L_d es la escala de correlación espacial. Además la varianza del background es σ_b. Suponga que tenemos dos intrumentos de medición A y B con error de medición σ_r tal que la razón σ_r/σ_b viene dada por $\alpha = (\sigma_r/\sigma_b)^2$. El instrumento A esta ubicado en la posición $P_1 = -2L$, el instrumento B esta en una posición P_2 que puede ser cualquier punto de $[-4L, 4L]$ (es decir un instrumento móvil) y requerimos conocer el estado del sistema en $P_0 = 0$. Suponga además se conoce en las tres posiciones el apriori o background y estos vienen dados por x_0^b, x_1^b y x_2^b. Mientras las mediciones de A y B son y_1 e y_2.

Si se asume estadística Gaussiana para los errores el valor esperado del análisis $\mathbf{x}^a$ vendrá dado por la fórmula de interpolación óptima

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^b) \tag{2}$$

donde $\mathbf{x}^b$ el vector del background o apriori (esperado), $\mathbf{y}$ son las observaciones $\mathbf{H}$ es el operador observaciones (el cual asumimos lineal y solo interpola el estado en los lugares de las observaciones, $\mathbf{B}$ es la matriz de covarianza del error del background y $\mathbf{R}$ es la matriz de covarianza del error de la medición y $\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{B} + \mathbf{H}^\top\mathbf{R}\mathbf{H})^{-1}$.

La matriz de covarianza del error del análisis, asumiendo un análisis óptimo (2), viene dada por

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B} \tag{3}$$

(a) Demuestre que la ecuación del análisis en P_0 usando interpolación óptima (2)

$$x_0^a = x_0^b + w_1(y_1 - x_1^b) + w_2(y_2 - x_2^b), \tag{4}$$

[1] Para lectura complementaria del Problema 1 ver Roger Daley, Atmospheric data analysis, Cambridge University Press 1991. Cap. 4.6.

donde los pesos vienen dados por

$$w_1 = \frac{\rho_{10}(1+\alpha) - \rho_{12}\rho_{20}}{(1+\alpha)^2 - \rho_{12}^2}, \tag{5}$$

$$w_2 = \frac{\rho_{20}(1+\alpha) - \rho_{12}\rho_{10}}{(1+\alpha)^2 - \rho_{12}^2}, \tag{6}$$

donde $\rho_{12} = \left(1 + \frac{|P_1-P_2|}{L}\right) \exp\left(-\frac{|P_1-P_2|}{L}\right)$. (Se ha denotado aqui por P, para no confundir el vector de estado $\mathbf{x}$ con la posición espacial, aunque a ésta también se le suele designar con x.) La información observacional de y_1 e y_2 en (4) se transfiere al punto P_0 utilizando la correlación. Usando (3) demuestre que

$$(\sigma_0^a)^2 = (\sigma_0^b)^2 \left[1 - \frac{(1+\alpha)(\rho_{10}^2 + \rho_{20}^2) - 2\rho_{10}\rho_{20}\rho_{12}}{(1+\alpha)^2 - \rho_{12}^2}\right] \tag{7}$$

(b) Utilizando el código python suministrado estudie el impacto de variar α, $\alpha = 0.1,\ 0.25,\ 0.5,\ 1.0$ en el error del análisis. i) ¿Qué pasa cuando crece α con el error del análisis?. ii) ¿Analice en que casos sucede el pantalleo de la observación (peso negativo en una observación)?.

(c) Considere como varían los resultados cuando el instrumento A es movido a: $P/L = -3.0,\ -2.0,\ -1.0,\ 0.0$. Tome $\alpha = 0.25$. i) Cuáles son las ubicaciones del instrumento B que no tienen efecto sobre el análisis? ii) ¿Cúal es la mejor ubicación para el instrumento A?

Problema 2: [2] Suponga realizamos mediciones con un radar, cuya reflectividad denotamos por y, queremos determinar a partir de éstas, por instancia, el contenido de agua líquida que denotamos por x y conocemos al operador observacional que nos relacionan a estas dos variables $y = \mathcal{H}(x)$. Este es un típico problema inverso en el cual no tenemos un modelo de predicción o pronóstico y el apriori viene dado por un conocimiento de la climatología del problema (e.g. la densidad de probabilidad a priori del contenido de agua líquida podria depender del tipo de tormenta). Suponga que el instrumento de medición tiene errores que siguen una distribución Gaussiana con varianza σ_r. Utilice la regla de Bayes para determinar la densidad posterior. Realice los experimentos computacionales y obtenga la densidad posterior en forma gráfica. Extraiga conclusiones sobre el rol del apriori, las observaciones y de la densidad posterior secuencial en el caso de observaciones múltiples. Para los siguientes casos:

(a) El operador observación es la identidad $\mathcal{H} = I$, el apriori Gaussiano de media 1 y desviación estandard 1, $\mathcal{N}(1,1)$. Determine cual es la densidad posterior si una primera medición es $y_1 = 0.5$ con $\sigma_r = 0.6$ y cual es la posterior luego de una segunda medición de $y_2 = 0.9$. Suponga las dos mediciones corresponden al mismo evento.

(b) $\mathcal{H} = I$ y tenemos una densidad apriori uniforme $\mathcal{U}(0,2)$. Determine cual es la densidad posterior si tenemos las mismas mediciones que en (a).

[2] Una referencia clásica de problemas inversos en un marco Bayesiano es el libro de Albert Tarantola, "Inverse Problem Theory and Methods for Model Parameter Estimation", SIAM, 2004.

(c) $\mathcal{H}(x) = \sin(\pi x)$ y tenemos un apriori uniforme $\mathcal{U}(0, 2)$. Determine cual es la densidad posterior si tenemos las mismas mediciones que en (a).

Problema 3: Supongamos el sistema dinámico Lorenz-96 el cual consiste en 40 variables que satisfacen las ecuaciones diferenciales ordinarias

$$\frac{dx_n}{dt} = -x_{n-1}(x_{n-2} - x_{n+1}) - x_n + F, \tag{8}$$

con $n = 1, \cdots, 40$ y condiciones de contorno cíclicas: $x_{41} = x_1$, $x_0 = x_{40}$ y $x_{-1} = x_{39}$. El modelo tiene comportamiento caótico para un forzado externo de $F = 8$ y superiores. Este modelo es un prototipo muy básico de la propagación de ondas en un círculo de latitud.

(a) *Discretizacin del tiempo.* La asimilación se quiere realizar en un periodo temporal desde $t = 0$ hasta $t = T_{max}$. En general, los pasos del tiempo del modelo estan fijados por el propio modelo, asumamos estos son $\delta t = 0.01$. Mientras las observaciones se realizan periodicamente cada una cierta cantidad de pasos de tiempo, $\Delta t = j\delta t$. Ejecute la celda correspondiente del notebook e interprete las variables.

(b) *Especificacin del sistema.* El modelo evolucionar entre los tiempos con observaciones a travs de

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

donde el modelo $\mathcal{M}$ corresponde al Lorenz-96 de 40 variables y $F = 8$, (8). El error de modelo es representado a traves de un ruido aditivo $\boldsymbol{\eta}_k$ dado una distribución gaussiana con media 0 y matriz de covarianzas $Q = 0.3\mathbf{I}$. Consideraremos además que las observaciones son dadas por

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

donde hemos asumido el operador observacional lineal $\mathbf{H}$, y por ahora asumiremos las 40 variables del modelo de Lorenz-96 son directamente observadas, es decir $\mathbf{H}$ es la matriz identidad, $\mathbf{H} = \mathbf{I}$. El error observacional $\boldsymbol{\nu}_k$ ser gaussiano con media 0 y matriz de covarianzas $\mathbf{R} = 0.5\mathbf{I}$. Ejecute la celda correspondiente e interprete las variables. Qu tipo de matrices de covarianzas se han elegido?

(c) *Spin-up del modelo.* Tomando condiciones iniciales aleatorias a partir de una distribución Gaussiana de media 0 y varianza 2 en cada una de las variables del Lorenz-96. Integrar el modelo con pasos de tiempo de $\delta t = 0.01$ hasta $t = 1440$ (que corresponde a 20 años de la atmósfera, es decir un dia corresponde a un $t = 0.2$ de los tiempos del modelo de Lorenz comparando las razones de crecimiento de los modos inestables). Esta primera integración, se realiza para obtener un estado que pertenezca al atractor del modelo de Lorenz-96, esta integración inicial se denomina el "spin-up". Tomaremos el estado final resultado de esta integración spin-up, como la condición inicial $\mathbf{x}_0$ de nuestra "natura".

(d) *Observaciones sintéticas.* Partiendo desde el estado inicial obtenido del spin-up generaremos el estado "verdadero", integrando el modelo hasta un tiempo $T_{max} = 25?$ conservando cada $\Delta t = 0.05$, esta integración es la realidad. El estado escondido que no conocemos e intentaremos estimar con asimilacin de datos. A partir de esta integración de la realidad genere las observaciones sintticas utilizando el operador $\mathbf{H}$ y el ruido observacional.

(e) *Covarianza del background* **B**. Considerando que todas las variables son estadísticamente equivalentes, utilice la integración verdadera del modelo pero con un $t_{max} = 5000$ para determinar las correlaciones con una variable de referencia y estime de esta manera la covarianza climática. Interprete las covarianzas obtenidas de la "climatologa". Compare con la correlación propuesta por un modelo SOAR y con una matriz diagonal. Luego estudiaremos la performance de los algoritmos de asimilacin de datos utilizando distintas matrices de covarianza B.

(f) *Asimilación de datos con interpolación óptima.* Complete los cdigos del algoritmo de interpolacin ptima. Realice experimentos de 500 ciclos de asimilación, t_{max} y para evaluar la performance de los algoritmos utilice el error de raiz cuadrada media (RMS) entre la estimación y el estado verdadero (). Considere dos operadores observacionales uno en el cual se mide en todas las variables y otro en el cual se mide cada 4 variables (10 observaciones). Utilice distintas matrices de covarianza B (diagonal, SOAR, climatologa).

(g) *Asimilación de datos con 3dVar.*

[(i)]

- Repita el experimento anterior usando asimilación variacional. Compare los resultados.
- Genere observaciones sinteticas extras a partir de la misma integracion verdadera con $\mathbf{R} = 0.05\mathbf{I}$ y $\mathbf{R} = 2\mathbf{I}$. Realice un experimento con 3dVar comparando el RMSE de experimentos con $\mathbf{R} = 0.05\mathbf{I}$, $\mathbf{R} = 0.5\mathbf{I}$, y $\mathbf{R} = 2\mathbf{I}$.

(h) *Asimilación de datos con EnKF.*

[(i)]

- Generación de un ensamble inicial. Utilice la simulación de la verdad y elija aleatoriamente N_e estados no correlaciones (cuya separacion entre ellos sea $> 20 * \Delta t = 1$.
- Realice una asimilación de 500 ciclos comparando el RMSE con $N_e = 10, 50, 100$ usando de operador observacional la identidad.
- Realice el mismo experimento que el item anterior pero usando 10 observaciones equidistribuidas.
- Evaluación del impacto de la nolinealidad. Comparar el RMSE del EnKF usando 30 partículas con ciclos de asimilación de $\Delta t = 0.05$, 0.10 y 0.20.
- Para evaluar un caso con error de modelo estructural. Utilice como modelo en el sistema de asimilación un Lorenz-96 en el cual F=9. Compare el RMSE con el caso de modelo "perfecto".