Data Assimilation and Climate
AIMS Rwanda, April, 2020

Final Quiz

1. Let $\mathbf{Y} = \mathbf{H}\mathbf{x} + \boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_\epsilon)$. Assume that $\mathbf{Y}$ is stochastic and that $\mathbf{x}$ is not stochastic. If we make the estimate $\hat{\mathbf{X}} = (\mathbf{H}^T\mathbf{C}_\epsilon^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}_\epsilon^{-1}\mathbf{Y}$, what is the expected value $\mathbb{E}[\hat{\mathbf{X}}]$?

2. Assume we want to estimate a scalar state variable X, which could represent a temperature or a wind velocity component. Consider the data model

$$Y = X + \epsilon$$

where the conditional distribution of the data given the state is $Y|X \sim \mathcal{N}(X, \sigma_y^2)$.

   (a) Given a single observation y_1 of x, what is the conditional probability density function $f_{Y|X}(y_1|x)$?

   (b) Assume that we are given m independent observations $\mathbf{y} = (y_1, y_2, \ldots, y_m)^T$ of the scalar variable x

      i. Find the conditional probability distribution function $f_{\mathbf{Y}|X}(y_1, y_2, \ldots, y_m|x)$

      ii. Assume X has prior distribution $X \sim \mathcal{N}(\mu, \sigma_x^2)$. Use Bayes law to find a proportional ($\propto$) posterior probability distribution $f_{X|\mathbf{Y}}(x|y_1, y_2, \ldots, y_m)$.

      iii. Completing the square in the posterior distribution found in **??** gives

$$f_{X|\mathbf{Y}}(x|y_1, y_2, \ldots, y_m) \propto \exp\left\{-\frac{1}{2}\left(x - (n/\sigma_y^2 + 1/\sigma_x^2)^{-1}\left(\sum_{i=1}^{m} y_i/\sigma_y^2 + \mu/\sigma_x^2\right)\right)^2 (n/\sigma_y^2 + 1/\sigma_x^2)\right\}$$

      Find constants c_1 and c_2 so that that the mean of the posterior distribution is

$$c_1\overline{y} + c_2\mu,$$

      where $\overline{y} = \frac{1}{m}\sum_{i=1}^{m} y_i$.

      iv. Discuss how the weights combine the prior mean and average of the data to form the posterior estimate. Also what happens when (a) there is large uncertainty in the prior, (b) a large number of observations, and (c) a small number of observations.

3. Consider the process and data models

$$\begin{aligned}
\mathbf{X}(i) &= \mathbf{M}_i\mathbf{X}(i-1) + \boldsymbol{\delta}_i \quad i = 1, \ldots, N \\
\mathbf{Y}(i) &= \mathbf{H}_i\mathbf{X}(i) + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, N
\end{aligned}$$

with $\mathbf{X}(0) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_i)$, $\boldsymbol{\delta}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_i)$, and recall the Kalman filter

$$\begin{aligned}
\boldsymbol{\mu}_{i|i-1} &= \mathbf{M}_i\boldsymbol{\mu}_{i-1|i-1} \\
\boldsymbol{\Sigma}_{i|i-1} &= \mathbf{Q}_i + \mathbf{M}_i\boldsymbol{\Sigma}_{i-1|i-1}\mathbf{M}_i^T \\
\mathbf{K}_i &= \boldsymbol{\Sigma}_{i|i-1}\mathbf{H}_i^T\left(\mathbf{H}_i^T\boldsymbol{\Sigma}_{i|i-1}\mathbf{H}_i + \mathbf{R}_i\right)^{-1} \\
\boldsymbol{\mu}_{i|i} &= \boldsymbol{\mu}_{i|i-1} + \mathbf{K}_i\left(\mathbf{y}(i) - \mathbf{H}_i\boldsymbol{\mu}_{i|i-1}\right) \\
\boldsymbol{\Sigma}_{i|i} &= (\mathbf{I} - \mathbf{K}_i\mathbf{H}_i)\,\boldsymbol{\Sigma}_{i|i-1}
\end{aligned}$$

for $i = 1, \ldots, N$. Explain what happens the forecast and and filtered estimates, and their uncertainties, when you apply the Kalman filter without any data, i.e. $\mathbf{H}_i = \mathbf{0}$. Also discuss the forecast, for example, when you run a climate model where observations are obtained every 6 hours and you run the assimilation for 6 hours without any data.

4. Programming
   Recall the 3DVar cost function

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(\mathbf{H}\mathbf{x} - \mathbf{y})^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y}),$$

   with state estimate $\hat{\mathbf{x}} = \mathbf{x}^b + \mathbf{K}\left(\mathbf{y} - \mathbf{H}\mathbf{x}^b\right)$, $\mathbf{K} = \mathbf{B}\mathbf{H}^T\left(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}\right)^{-1}$.

   (a) We seek two temperatures, x_1 and x_2, in London and Paris. The climatologist gives us an initial guess (based on climate records) of $\mathbf{x}_b = (10, 5)^T$, with background error

$$\mathbf{B} = \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}.$$

   We observe $\mathbf{y} = (4°)$ in Paris with an error variance $\mathbf{R} = (0.25)$, but do not have an observation for London. This means that $\mathbf{H} = (0, 1)$. Use 3DVar to find an optimal estimate of the temperatures in London and Paris, given the background and observation with errors.

   (b) We seek monthly rainfall estimates in Kigali. The climatologist gives us initial guesses based on climate records of average monthly rainfall in Table 1 below. Table 1 also contains observations of average monthly rainfall recorded at the airport in 2019.

      i. Assume that the background error variance is $0.1mm$ for all months, and the observation error variance is $0.25mm$ for all months. Use 3DVar to find optimal estimates of monthly rainfall in 2019. Plot your optimal estimates, together with the data and background as points. Use a legend for your plot and describe how the estimates compare with the background and data.

      ii. Assume that the background error variance is $0.25mm$ for all months, and the observation error variance remains at $0.25mm$ for all months. Plot optimal estimates of monthly rainfall in 2019, together with the data and background. Describe how the estimates compare with the background and data.

      iii. Assume that there are no observations for July and August, and that the background and observation error variances are still $0.25mm$ for all months. Adjust $\mathbf{H}$ to reflect that there are no observations in July and August. Plot optimal estimates of monthly rainfall in 2019, together with the data and background. Describe how the estimates compare with the background and data.

Month	Background	Observations
January	76.9	80
February	91	90
March	114.2	110
April	154.2	160
May	88.1	80
June	18.6	20
July	11.4	10
August	31.1	40
September	69.6	70
October	105.7	100
November	112.7	110
December	77.4	80

Table 1: Average monthly rainfall in Kigali (mm)