

# Model error covariance estimation in weak-constraint 4DVar

Peter Jan van Leeuwen, Anne Pein

July 28, 2022

## Abstract

TO DO: In this work, we present a method to estimate the model error covariance matrix based on from the data assimilation Using a weak-constraint 4D-Var implementation, we study different toy examples with an additive model error with statistics that is constant over the assimilation window. In particular, we analyse the sensitivity of the analysis to the model error covariance and explore possibilities to derive information about the true model error from the analysis that will allow an iterative tuning of the covariance matrix. In the spirit of previous approaches, we rely here on statistics based on differences between the analysis, background and observations.

## Contents

### 1 Introduction

### 2 Methodology

- 2.1 Problem formulation . . . . .
- 2.2 Error covariance estimation basics . . . . .
- 2.3 Q estimation in 4DVar based on the forcing formulation . . . . .

### 3 Numerical experiments

- 3.1 Experimental setup . . . . .
- 3.2 Results . . . . .

### 4 Conclusion

### 5 Convergence proof

## 1 Introduction

The analysis state in weak-constraint variational data assimilation is computed as the minimiser of a suitable cost function that penalises deviations from the observations, from a background state and from the dynamical model itself, each weighted by their respective error covariance matrices. Allowing for errors in the dynamical model is of importance in many geoscience applications where the full model might not be available e.g. due to unresolved processes. Higher accuracy and reduced sensitivity to initial condition compared to the strong-constraint formulation have been observed. In particular, via an efficient implementation of the so-called representer method in observation space, the weak-constraint 4DVar is suitable for large operational systems. Focusing on this specific setting, we revisit one of the major challenges in data assimilation: estimating model error covariance matrices. Compared to observation and background error covariances the influence and estimation of model error covariances is much less studied. We estimate the model error covariance matrix in a 4DVar data assimilation experiment, assuming that the background error covariance and observation error covariance are correctly specified in the scheme. The idea is based on residual statistics that have been used for similar purposes before, see (Hollingsworth and Lonnberg, 1989; Dee and da Silva, 1999; Desroziers et al., 2005; Chapnik et al., 2006). Note that this approach does not resolve any of the issues addressed in [5] regarding the separate estimation of several unknown error covariances.

TO DO

## 2 Methodology

### 2.1 Problem formulation

We assume that our model state evolves according to

$$x^{n+1} = \mathcal{M}(x^n) + \varepsilon_{\mathbf{Q}}^{n+1}, \quad n = 0, \dots, N_t, \quad (2.1)$$

where  $x^n \in \mathbb{R}^{N_x}$  and  $\mathcal{M}$  is a possibly non-linear, deterministic model. The time dependent errors  $\varepsilon_{\mathbf{Q}}^n$  account for the missing physics in the model and numerical approximations, we assume that  $\varepsilon_{\mathbf{Q}}^n \sim \mathcal{N}(0, \mathbf{Q}_n)$  and  $\mathbf{Q}_n \in \mathbb{R}^{N_x \times N_x}$  is the model error covariance matrix at time  $n$ .

Furthermore, observations of certain state variables at certain time steps are available

$$y^k = \mathcal{H}^k(x^k) + \varepsilon_{\mathbf{R}}^k, \quad (2.2)$$

where  $k = t_1, \dots, t_{m_t}$  and  $\mathcal{H}^k : \mathbb{R}^{N_x} \rightarrow \mathbb{R}^{m_x}$  denotes the possibly non-linear measurement operator at time  $k$ , i.e. the operator that maps a state vector into the observation space. The uncertainty associated with the measurement is assumed to be distributed as  $\varepsilon_{\mathbf{R}}^k \sim \mathcal{N}(0, \mathbf{R}^k)$ , where  $\mathbf{R}^k \in \mathbb{R}^{m_x \times m_x}$  is the observation error covariance matrix at time  $k$ . Let us also introduce the notations

$$\begin{aligned} x &= (x^0, \dots, x^{N_t}) \in \mathbb{R}^{(N_t+1) \times N_x}, \\ y &= (y^{t_1}, \dots, y^{t_{m_t}}) \in \mathbb{R}^{m_t \times m_x}, \\ \mathcal{H}(x) &= (\mathcal{H}^{t_1}(x^{t_1}), \dots, \mathcal{H}^{t_{m_t}}(x^{t_{m_t}})). \end{aligned}$$

Finally, we have are given a background state

$$x_b^0 = x^0 + \varepsilon_{\mathbf{B}}, \quad (2.3)$$

with  $\varepsilon_{\mathbf{B}} \sim \mathcal{N}(0, \mathbf{B})$  and background error covariance matrix  $\mathbf{B} \in \mathbb{R}^{N_x \times N_x}$ .

Via Maximum Likelihood considerations one can derive a cost function whose minimiser is the desired analysis state of the weak-constraint problem. In the state space formulation it is given as follows (see [])

$$J(x) = \frac{1}{2}(x_b^0 - x^0)^\top \mathbf{B}^{-1}(x_b^0 - x^0) + \frac{1}{2}(y - \mathcal{H}(x))^\top \mathbf{R}^{-1}(y - \mathcal{H}(x)) \quad (2.4)$$

$$+ \frac{1}{2} \sum_{n=1}^{N_t} (x^n - \mathcal{M}(x^{n-1}))^\top \mathbf{Q}_n^{-1} (x^n - \mathcal{M}(x^{n-1})), \quad (2.5)$$

where we used the notation  $\mathbf{R} = \text{diag}(\mathbf{R}^k)_{k=t_1, \dots, t_{m_t}}$ .

Now, in real applications the covariance matrices  $\mathbf{B}, \mathbf{R}, \mathbf{Q}_n$  are unknown and need to be inferred. Thus for the cost function (2.5) typically only approximations of the true error covariance matrices are available. In our work, we assume that the background error covariance matrix and observation error covariance matrix are known and correctly specified in (2.5). However, the model error covariance matrix is assumed to be unknown and the objective is to iteratively improve the estimate of this matrix. Before going into details about our method, we will briefly recall error covariance estimation theory based on consistency diagnostics in the next subsection.

### 2.2 Error covariance estimation basics

Let us highlight some ideas on error covariance estimation in data assimilation. Assume we compute an analysis state  $x_a$  based on the following scheme

$$x_a = x_b + \tilde{\mathbf{K}}(y - \mathbf{H}(x_b)), \quad (2.6)$$

where  $x_b$  is the background state,  $y$  are the observations and  $\mathbf{H}$  is the observation operator.  $\tilde{\mathbf{K}}$  denotes the gain matrix

$$\tilde{\mathbf{K}} = \tilde{\mathbf{B}}\mathbf{H}^\top(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^\top + \tilde{\mathbf{R}})^{-1},$$

based on estimates  $\tilde{\mathbf{B}}$  and  $\tilde{\mathbf{R}}$  of the true background and observation error covariance matrices  $\mathbf{B}_t$  and  $\mathbf{R}_t$ . In [1] Desroziers et al introduced consistency diagnostics based on the following residuals  $d_b = y - \mathbf{H}(x_b)$ ,  $d_a = y - \mathbf{H}(x_a)$  and  $d_{ab} = \mathbf{H}(x_a) - \mathbf{H}(x_b)$ . Namely, it holds

$$\begin{aligned} \langle d_a d_b^\top \rangle &= \tilde{\mathbf{R}}(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H} + \tilde{\mathbf{R}})^{-1} \langle d_b d_b^\top \rangle = \tilde{\mathbf{R}}(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H} + \tilde{\mathbf{R}})^{-1}(\mathbf{H}\mathbf{B}_t\mathbf{H} + \mathbf{R}_t), \\ \langle d_{ab} d_b^\top \rangle &= \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H} + \tilde{\mathbf{R}})^{-1} \langle d_b d_b^\top \rangle = \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H} + \tilde{\mathbf{R}})^{-1}(\mathbf{H}\mathbf{B}_t\mathbf{H} + \mathbf{R}_t), \end{aligned}$$

If the true background and observation error covariance matrices are specified in the gain matrix, i.e.  $\tilde{\mathbf{K}} = \mathbf{K} = \mathbf{B}_t \mathbf{H}^\top (\mathbf{H} \mathbf{B}_t \mathbf{H}^\top + \mathbf{R}_t)^{-1}$ , then it holds

$$\begin{aligned}\mathbf{R}_t &= \langle d_a d_b^\top \rangle, \\ \mathbf{H} \mathbf{B}_t \mathbf{H}^\top &= \langle d_{ab} d_b^\top \rangle,\end{aligned}$$

where  $\langle \cdot \rangle$  denotes the mathematical expectation operator. Furthermore, they suggest to use these relations to iteratively estimate these matrices in case they are not correctly specified in the data assimilation scheme via

$$\begin{aligned}\tilde{\mathbf{R}}_{i+1} &= \langle d_{a_i} d_b^\top \rangle \\ \mathbf{H} \tilde{\mathbf{B}}_{i+1} \mathbf{H}^\top &= \langle d_{a_i b} d_b^\top \rangle,\end{aligned}$$

where  $a_i$  denotes the analysis of a previous iteration, generated with  $\tilde{\mathbf{K}}_i$  in (2.6). See also [4] and [3].

### 2.3 Q estimation in 4DVar based on the forcing formulation

We consider the weak constraint 4DVar data assimilation scheme in the forcing formulation (see [2] for more information). That is, we derive an analysis state  $z_a$  by minimising the cost function

$$J(z) = \frac{1}{2}(z - z_b)^\top \mathbf{D}^{-1}(z - z_b) + \frac{1}{2}(y - \mathbf{H}_z z) \mathbf{R}^{-1}(y - \mathbf{H}_z z). \quad (2.7)$$

Here  $\mathbf{D}$  is the prior covariance matrix over the whole time window

$$\mathbf{D} = \begin{pmatrix} \mathbf{B} & 0 & \cdots & \cdots & 0 \\ 0 & \mathbf{Q} & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \mathbf{Q} \end{pmatrix}. \quad (2.8)$$

Note that we assume  $\mathbf{Q}$  to be time independent and uncorrelated over time, this causes the block diagonal structure in  $\mathbf{D}$ .

Let  $x_a = (x_a^1, \dots, x_a^{N_t})$  with  $x_a^n = (x_1^n, \dots, x_{N_x}^n)$ , for  $n = 1, \dots, N_t$ , denote the analysis of the dynamical model in state space, which evolves according to

$$x_a^n = m(x_a^{n-1}) + \beta_a^n. \quad (2.9)$$

Now the analysis  $z_a$  in the forcing formulation is given as

$$z_a = \begin{pmatrix} x_a^0 \\ \beta_a^1 \\ \vdots \\ \beta_a^{N_t} \end{pmatrix}. \quad (2.10)$$

In particular, the vector  $z_a$  is related to the analysis in state space via  $z_a = \mathbf{P} x_a$ , in which  $\mathbf{P}$  is given by

$$\mathbf{P} = \begin{pmatrix} \text{Id} & 0 & \cdots & \cdots & \cdots & 0 \\ -\mathbf{M}_{0 \rightarrow 1} & \text{Id} & 0 & \cdots & \cdots & \vdots \\ 0 & -\mathbf{M}_{1 \rightarrow 2} & \text{Id} & 0 & \cdots & \vdots \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & \cdots & 0 & -\mathbf{M}_{N_t-2 \rightarrow N_t-1} & \text{Id} & 0 \\ 0 & \cdots & \cdots & 0 & -\mathbf{M}_{N_t-1 \rightarrow N_t} & \text{Id} \end{pmatrix}, \quad (2.11)$$

where  $\mathbf{M}_{i-1 \rightarrow i}$  is the linearised deterministic model from time  $i-1$  to time  $i$ . With this construction we can define the observation operator in 'z-space' as

$$\mathbf{H}_z z_a = \mathbf{H}_z \mathbf{P} x_a = \mathbf{H}_x x_a, \quad (2.12)$$

such that  $\mathbf{H}_z = \mathbf{H}_x \mathbf{P}^{-1}$ .

The analysis in  $z$  space derived as the minimum of (2.7) is given by

$$z_a - z_b = \left( \mathbf{D}^{-1} + \mathbf{H}_z^\top \mathbf{R}^{-1} \mathbf{H}_z \right)^{-1} \mathbf{H}_z^\top \mathbf{R}^{-1} (y - \mathbf{H}_z z_b) \quad (2.13)$$

$$= \mathbf{D} \mathbf{H}_z^\top (\mathbf{H}_z \mathbf{D} \mathbf{H}_z^\top + \mathbf{R})^{-1} (y - \mathbf{H}_z z_b). \quad (2.14)$$

Note the structural similarity to (2.6). We can now form the analogous diagnostics as in the Desroziers setup, i.e. we form  $d_{ab} = \mathbf{H}_z z_a - \mathbf{H}_z z_b = \mathbf{H}_x x_a - \mathbf{H}_x x_b$  and  $d_b = y - \mathbf{H}_z z_b = y - \mathbf{H}_x x_b$ . Then just like before we have

$$\begin{aligned} \langle d_{ab} d_b^\top \rangle &= \mathbf{H}_z \mathbf{D} \mathbf{H}_z^\top (\mathbf{H}_z \mathbf{D} \mathbf{H}_z^\top + \mathbf{R})^{-1} \langle d_b d_b^\top \rangle \\ &= \mathbf{H}_z \mathbf{D} \mathbf{H}_z^\top (\mathbf{H}_z \mathbf{D} \mathbf{H}_z^\top + \mathbf{R})^{-1} (\mathbf{H}_z \mathbf{D}_t \mathbf{H}_z^\top + \mathbf{R}) \end{aligned}$$

Just like suggested in the Desroziers setup, we can now iteratively estimate  $\mathbf{D}$  via

$$\mathbf{H}_z \mathbf{D}_{i+1} \mathbf{H}_z^\top = \langle d_{a_i b_i} d_{b_i}^\top \rangle. \quad (2.15)$$

Note however that

$$\mathbf{H}_z \mathbf{D}_{i+1} \mathbf{H}_z^\top = \mathbf{H}_x \mathbf{P}^{-1} \mathbf{D} \mathbf{P}^{-\top} \mathbf{H}_x^\top$$

Now note that

$$\mathbf{P}^{-\top} = \begin{pmatrix} \text{Id} & \mathbf{M}_{0 \rightarrow 1}^\top & \mathbf{M}_{0 \rightarrow 2}^\top & \cdots & \mathbf{M}_{0 \rightarrow N_t}^\top \\ 0 & \text{Id} & \mathbf{M}_{1 \rightarrow 2}^\top & \cdots & \mathbf{M}_{1 \rightarrow N_t}^\top \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & \cdots & 0 & \text{Id} & \mathbf{M}_{N_t-1 \rightarrow N_t}^\top \\ 0 & \cdots & \cdots & 0 & \text{Id} \end{pmatrix}, \quad (2.16)$$

that is  $\mathbf{P}^{-\top}$  is an upper triangular matrix with identities on the diagonal and  $\mathbf{P}^{-1}$  is a lower triangular matrix with identities on the diagonal. This means however that  $\mathbf{P}^{-1} \mathbf{D} \mathbf{P}^{-\top}$  has not the diagonal elements of  $\mathbf{D}$  on its diagonal, i.e. the information gets mixed up. However, if we multiply equation (2.15) by  $\mathbf{H}_z^{-1}$  (pseudoinverse) from the left and by  $\mathbf{H}_x$  from the right, we end up with

$$\begin{aligned} \mathbf{D}_{i+1} \mathbf{H}_z^\top \mathbf{H}_x &= \mathbf{H}_z^{-1} \langle (\mathbf{H}_z z_a^i - \mathbf{H}_z z_b^i) (y - \mathbf{H}_x x_b^i)^\top \rangle \mathbf{H}_x \\ &= \langle (z_a^i - z_b^i) (y - \mathbf{H}_x x_b^i)^\top \mathbf{H}_x \rangle, \end{aligned}$$

which is exactly what we derived further below as well. **This is a bit weird, because probably  $\mathbf{H}_z$  is not invertible? Maybe the old derivation is better, but the analogy to the stuff from section 1 is nice. Also why do we need to multiply  $\mathbf{H}_x$  from the right? otherwise dimensions go not out right?**

**old derivation** If we now multiply (2.13) from the left by  $(\mathbf{D}^{-1} + \mathbf{H}_z^\top \mathbf{R}^{-1} \mathbf{H}_z)$  we find

$$\begin{aligned} \left( \mathbf{D}^{-1} + \mathbf{H}_z^\top \mathbf{R}^{-1} \mathbf{H}_z \right) (z_a - z_b) &= \mathbf{H}_z^\top \mathbf{R}^{-1} (y - \mathbf{H}_z z_b) \\ \Leftrightarrow \mathbf{D}^{-1} (z_a - z_b) &= \mathbf{H}_z^\top \mathbf{R}^{-1} (y - \mathbf{H}_z z_b) - \mathbf{H}_z^\top \mathbf{R}^{-1} \mathbf{H}_z (z_a - z_b), \end{aligned}$$

leading to

$$z_d := z_a - z_b = \mathbf{D} \mathbf{H}_z^\top \mathbf{R}^{-1} (y - \mathbf{H}_x x_a). \quad (2.17)$$

Let us define

$$c := \mathbf{H}_x^\top (y - \mathbf{H}_x x_b). \quad (2.18)$$

Now we multiply eq. (2.17) with  $c^\top$  from the right and apply the statistical expectation operator, leading to

$$\langle z_d c^\top \rangle = \mathbf{D} \mathbf{H}_z^\top \mathbf{R}^{-1} \langle (y - \mathbf{H}_x x_a) (y - \mathbf{H}_x x_b)^\top \rangle \mathbf{H}_x. \quad (2.19)$$

Now note that

$$\begin{aligned} \langle (y - \mathbf{H}_x x_a) (y - \mathbf{H}_x x_b)^\top \rangle &= (\text{Id} - \mathbf{H}_z \mathbf{K}) \langle (y - \mathbf{H}_z z_b) (y - \mathbf{H}_x x_b)^\top \rangle \\ &= \mathbf{R} \left( \mathbf{H}_z \mathbf{D} \mathbf{H}_z^\top + \mathbf{R} \right)^{-1} (\mathbf{H}_z \mathbf{D}_t \mathbf{H}_z + \mathbf{R}), \end{aligned}$$

where  $\mathbf{D}_t = \text{diag}(\mathbf{B}, \mathbf{Q}_t, \dots, \mathbf{Q}_t)$ . Hence, if we assign the true model error covariance matrix  $\mathbf{Q}_t$  to the data assimilation scheme, i.e. if  $\mathbf{Q} = \mathbf{Q}_t$ , then

$$\langle (y - \mathbf{H}_x x_a) (y - \mathbf{H}_x x_b)^\top \rangle = \mathbf{R}. \quad (2.20)$$

Thus, we force our new estimate  $\hat{\mathbf{D}}$  to fulfil

$$\langle z_d c^\top \rangle = \hat{\mathbf{D}} \mathbf{H}_z^\top \mathbf{R}^{-1} \mathbf{R} \mathbf{H}_x = \hat{\mathbf{D}} \mathbf{H}_z^\top \mathbf{H}_x = \hat{\mathbf{D}} \mathbf{P}^{-\top} \mathbf{H}_x^\top \mathbf{H}_x. \quad (2.21)$$

Now note that

$$\mathbf{P}^{-\top} = \begin{pmatrix} \text{Id} & \mathbf{M}_{0 \rightarrow 1}^\top & \mathbf{M}_{0 \rightarrow 2}^\top & \cdots & \mathbf{M}_{0 \rightarrow N_t}^\top \\ 0 & \text{Id} & \mathbf{M}_{1 \rightarrow 2}^\top & \cdots & \mathbf{M}_{1 \rightarrow N_t}^\top \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & \cdots & 0 & \text{Id} & \mathbf{M}_{N_t-1 \rightarrow N_t}^\top \\ 0 & \cdots & \cdots & 0 & \text{Id} \end{pmatrix}, \quad (2.22)$$

and thus

$$\hat{\mathbf{D}} \mathbf{P}^{-\top} \mathbf{H}_x^\top \mathbf{H}_x = \begin{pmatrix} \mathbf{B} & \mathbf{B} \mathbf{M}_{0 \rightarrow 1}^\top & \mathbf{B} \mathbf{M}_{0 \rightarrow 2}^\top & \cdots & \mathbf{B} \mathbf{M}_{0 \rightarrow N_t}^\top \\ 0 & \hat{\mathbf{Q}} & \hat{\mathbf{Q}} \mathbf{M}_{1 \rightarrow 2}^\top & \cdots & \hat{\mathbf{Q}} \mathbf{M}_{1 \rightarrow N_t}^\top \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & \cdots & 0 & \hat{\mathbf{Q}} & \hat{\mathbf{Q}} \mathbf{M}_{N_t-1 \rightarrow N_t}^\top \\ 0 & \cdots & \cdots & 0 & \hat{\mathbf{Q}} \end{pmatrix} \mathbf{H}_x^\top \mathbf{H}_x \quad (2.23)$$

Now let us define  $\mathbf{W} := \langle z_d c^\top \rangle$  with  $\mathbf{W} = (\mathbf{W}_{ij})_{i,j=0,\dots,N_t}$ , where  $\mathbf{W}_{ij} \in \mathbb{R}^{N_x \times N_x}$ , for  $i, j \in \{1, \dots, N_t\}$ . Then we have the following relations for our new estimate of the model error covariance matrix

$$\mathbf{W}_{ii} = \hat{\mathbf{Q}} \mathbf{H}_x^\top \mathbf{H}_x, \quad \text{for } i = 1, \dots, N_t. \quad (2.24)$$

That is, we need to estimate the matrices  $\mathbf{W}_{nn} = \langle z_d^n (c^n)^\top \rangle$ ,  $n = 1, \dots, N_t$ . For each  $n \in \{1, \dots, N_t\}$  we have only one realisation of  $z_d^n$  and  $c^n$ . To compute sample cross covariance matrices we therefore make use of a priori knowledge of the model error covariance matrix. Namely, we assume that every gridpoint has the same statistics, i.e. we assume that  $\mathbf{W}_{nn}$  is of the following circulant structure

$$\mathbf{W}_{nn} = \begin{pmatrix} w_1^n & w_2^n & \cdots & w_{N_x}^n \\ w_{N_x}^n & w_1^n & \cdots & w_{N_x-1}^n \\ \vdots & \vdots & \ddots & \vdots \\ w_2^n & w_3^n & \cdots & w_1^n \end{pmatrix}. \quad (2.25)$$

Even though this assumption might not be fully fulfilled in applications, it is often the case that gridpoints in close proximity share similar statistical behaviour. They can be grouped together accordingly and spatial statistics within each of these groups can be taken into account.

In case of periodic boundary conditions, we can assume even more, namely, that  $w_2^n = w_{N_x}^n$ ,  $w_3^n = w_{N_x-1}^n, \dots, w_{N_x/2}^n = w_{N_x/2+2}^n$ . Note that we assume here  $N_x$  to be even for notational simplicity. If a gridpoint is not observed the corresponding column in  $\mathbf{W}_{nn}$  will consist of zeros, and likewise on the right hand side of (2.24) the term  $\mathbf{H}_x^\top \mathbf{H}_x$  will set the corresponding columns in  $\hat{\mathbf{Q}}$  to zero, i.e. we do not have any estimate for these entries of the model error covariance matrix. However, we can still extract important characteristics like the variance and the correlation length.

Let us denote the total number of observed gridpoints as  $m_x$  and we assume that observations are equidistantly spaced with stepsize  $\Delta m_x$ , where the first variable is always observed. An estimate  $\hat{w}_i^n$  for  $w_i^n$  when taking periodic boundary conditions into account can be derived as

$$\hat{w}_i^n = 1/2 \left( \frac{1}{m_x} \sum_{k=0}^{m_x-1} z_d^n (k \Delta m_x + 1) c^n (1 + ((k \Delta m_x + i) \bmod N_x)) \right. \\ \left. + \frac{1}{m_x} \sum_{k=0}^{m_x-1} z_d^n (1 + ((k \Delta m_x + i) \bmod N_x)) c^n (1 + k \Delta m_x) \right),$$

and  $\hat{w}_i^n = 0$  if the  $i$ th variable is unobserved. An estimate  $\hat{\mathbf{W}}_{nn}$  of (2.25) is then given by the circulant matrix generated by  $(\hat{w}_i^n)_{i=1,\dots,N_x}$  where the columns corresponding to unobserved variables are set equal to zero.

Finally, we average over time and arrive at  $\hat{w} = \frac{1}{N_t} \sum_{k=1}^{N_t} \hat{w}^n$ . We then derive the new estimate of the variance of  $\mathbf{Q}_t$  as  $\hat{\sigma}_t^2 = \hat{w}_1$ . To recover the correlation length scale of the estimated covariance matrix we fit a function  $\exp(-x^2/(2\ell^2))$  through the data points  $(k \Delta x_m, \frac{1}{w_1} w_{k \Delta x_m + 1})_{k=0,\dots,m_x-1}$  and the fitting parameter  $\ell$  gives the new estimate for the correlation length, i.e.  $\hat{\ell}_t = \ell$ . Assuming that the correlation is close to zero for gridpoints far away from each other, we might only fit the function through the first  $p$  data points in order to reduce noise.

### 3 Numerical experiments

#### 3.1 Experimental setup

In the following we will outline the setup for the twin experiments that we use to test the above estimation approach.

We consider the Lorenz 96 model, which is given as follows, for  $i = 1, \dots, N_x$ ,

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F =: f_i(x), \quad (3.1)$$

where  $x = (x_1, \dots, x_n)$  denotes the model state at positions  $1, \dots, N_x$ . The equation is equipped with periodic 'boundary conditions', i.e.  $x_{-1} = x_{N-1}$ ,  $x_0 = x_N$  and  $x_{N+1} = x_1$ . We set the forcing constant  $F$  as  $F = 8$  in the following. The Lorenz 96 model is widely used in data assimilation experiments as a model for the dynamics of an atmospheric quantity along a circle of constant latitude of the earth, see Figure 1.

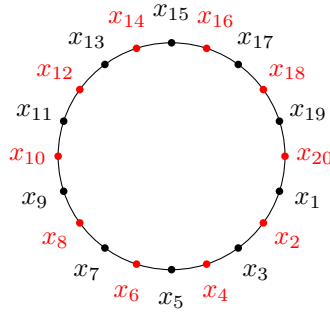


Figure 1: Gridpoints of the Lorenz 96 model with periodic boundaries and  $N_x = 20$ . We always observe the first variable  $x_1$  and then every  $\Delta m_x$ -th gridpoint. In the figure above  $\Delta m_x = 2$ , thus only the black variables are observed, while the red ones are unobserved.

Let us define a temporal grid with  $t_n = n\Delta t$  and we set  $x(t_n) =: x^n$ . Furthermore, let  $m$  denote some numerical integration scheme, such that (approximately), for  $n \in [1, N_t]$  and some initial conditions  $x^0$

$$x^n = m(x^{n-1}).$$

In our experiments we use the Runge-Kutta numerical integration scheme with a time step of 0.01.

We generate the truth via the 'imperfect model'

$$\hat{x}^n = m(\hat{x}^{n-1}) + \varepsilon_m^n, \quad (3.2)$$

where  $\varepsilon_m^n$  are samples from  $\mathcal{N}(0, \mathbf{Q}_t)$  and  $\mathbf{Q}_t = (q_{i,j})_{i,j}$  is the true model error covariance matrix. We choose a Gaussian correlation between the state variables, more precisely we set for  $j = 0, \dots, N/2$

$$q_{i, (i+j) \bmod N} = \sigma_t^2 \exp\left(-\frac{j^2}{2\ell_t^2}\right),$$

$$q_{i, i-j} = \sigma_t^2 \exp\left(-\frac{j^2}{2\ell_t^2}\right),$$

where  $\sigma_t^2$  is the true model error variance and  $\ell_t$  the true model error correlation length. Note that we are taking the periodic boundary conditions of the model into account.

We generate artificial observations at certain observations times  $\tau_j$ ,  $j \in \{1, \dots, m_t\}$ , and of certain state variables  $r_i, i \in \{1, \dots, m_x\}$ , via

$$y^j = H(\hat{x}^{\tau_j}) + \varepsilon_o^{i,j} = \hat{x}^{\tau_j} + \varepsilon_o^j \quad (3.3)$$

where  $\varepsilon_o^j$  is a sample from  $\mathcal{N}(0, \mathbf{R})$ . Here,  $\mathbf{R} \in \mathbb{R}^{m_x}$  is the observation error covariance matrix and  $m = m_x \times m_t$  the total number of observations in the given assimilation window. We choose  $\mathbf{R} = \text{diag}(\sigma_o^2)$ . Note that for simplicity we assume that we directly observe the state variables and observations are available exactly at the temporal grid points of the discretised problem. Thus, we do not need to perform any interpolation at this point.

We generate the initial condition by perturbing the equilibrium state  $x_i = F$ ,  $i = 1, \dots, N_x$ , slightly, namely we set  $x_0 = F + 0.01$  and then we forwards evolve this perturbed state for 1000 time steps. We denote the resulting state vector as  $\bar{x}^0$ . The initial true state is then generated as

$$\hat{x}^0 = \bar{x}^0 + \varepsilon_b \quad (3.4)$$

where  $\varepsilon_b$  is a sample from  $\mathcal{N}(0, \mathbf{B})$ .

Furthermore, we generate the background state  $\bar{x}$  given by the forecast of the initial state via Runge-Kutta integration of the perfect model  $m$ . That is starting with  $\bar{x}^0$  we compute

$$\bar{x}^{n+1} = m(\bar{x}^n). \quad (3.5)$$

Finally, as mentioned above, we assume that the true values of  $\mathbf{B}$  and  $\mathbf{R}$  are known and will be assigned to the data assimilation scheme. The true model error covariance matrix is unknown and we start with an initial guess given by  $\hat{\mathbf{Q}}_0 = \mathbf{B}$ . For the first assimilation window we perform the 4DVar data assimilation and derive a new estimate  $\hat{\mathbf{Q}}_1$  via the method described in section 2.3. This matrix is subsequently assigned to the data assimilation scheme for the second assimilation window. This procedure is iterated for a number of  $N$  assimilation windows.

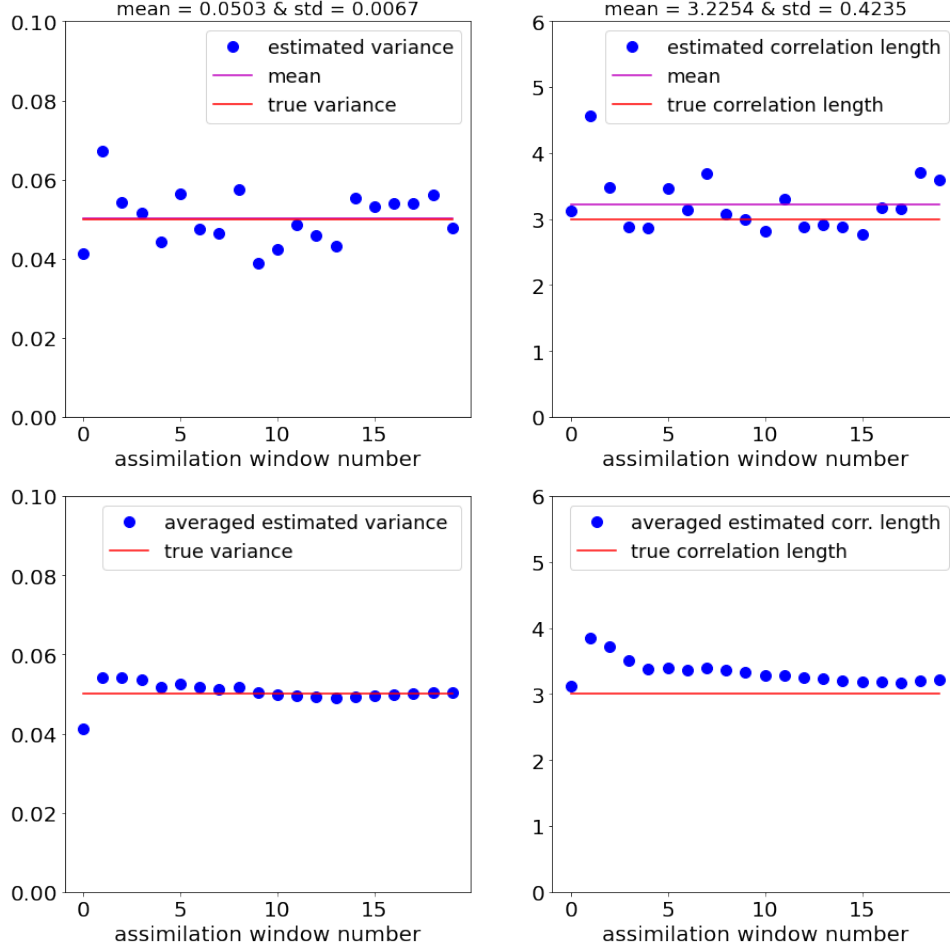
## 3.2 Results

### Experiment 1

We choose  $\mathbf{B} = \text{diag}(\sigma_b^2)$  (need to consider more complex matrices here). Parameters are summarised in the following table:

$N$	$N_t$	$N_x$	$m_t$	$m_x$	$\sigma_b^2$	$\sigma_o^2$	$\sigma_t^2$	$\ell_t$	$p$
20	5	200	5	200	0.01	0.01	0.05	3	10

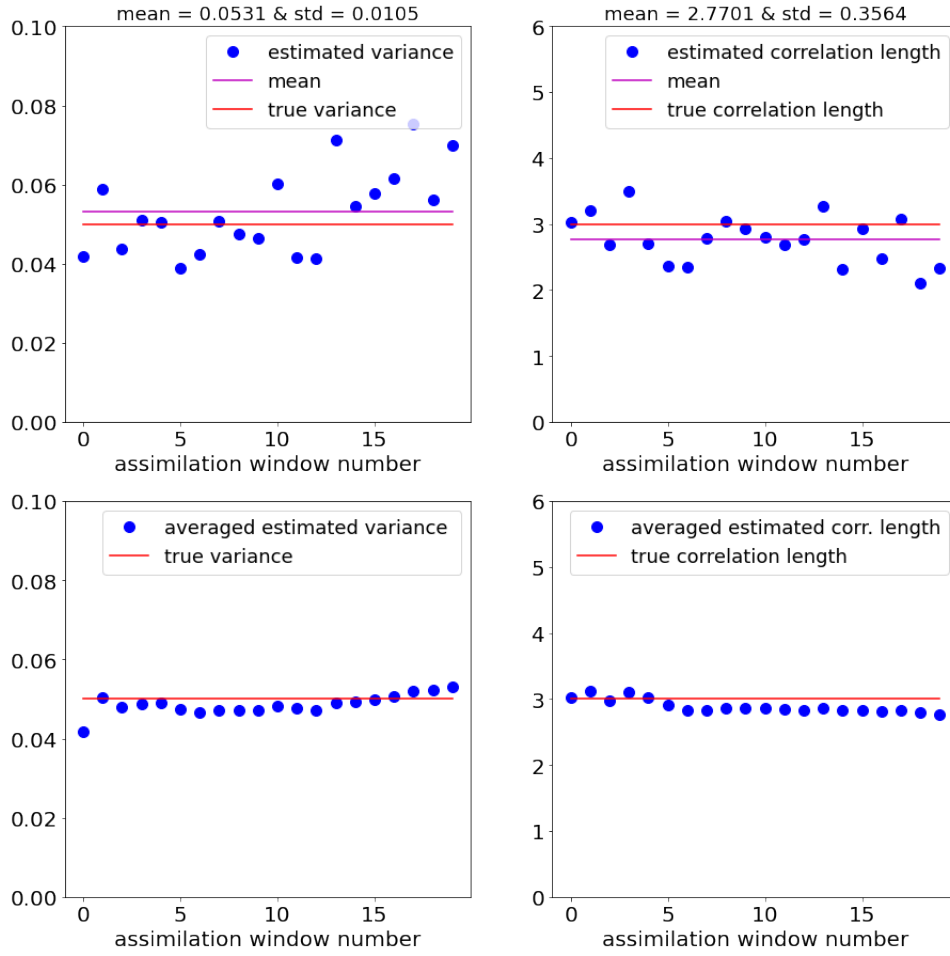
In particular every variable is observed at every time step.





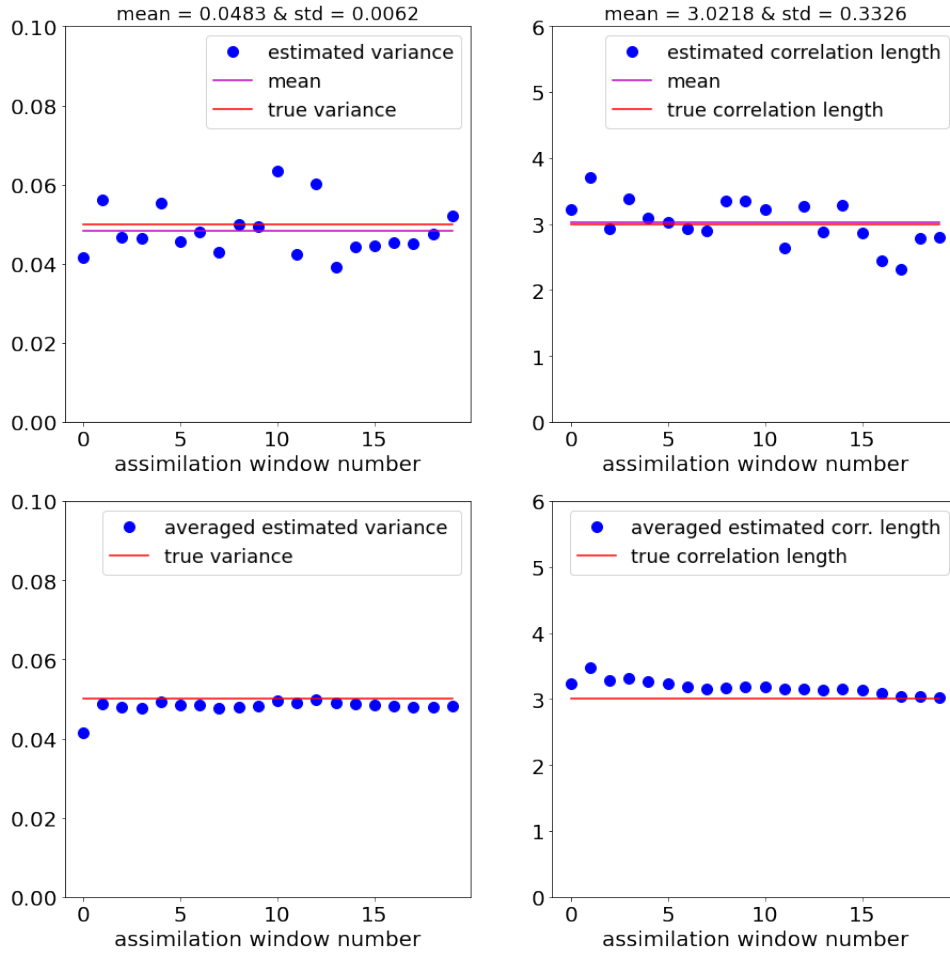
## Experiment 2 - partial observation in space

Same parameters as in Experiment 1, the only difference is that we observe only every second state variable, i.e.  $m_x = 100$  and  $\Delta m_x = 2$ .



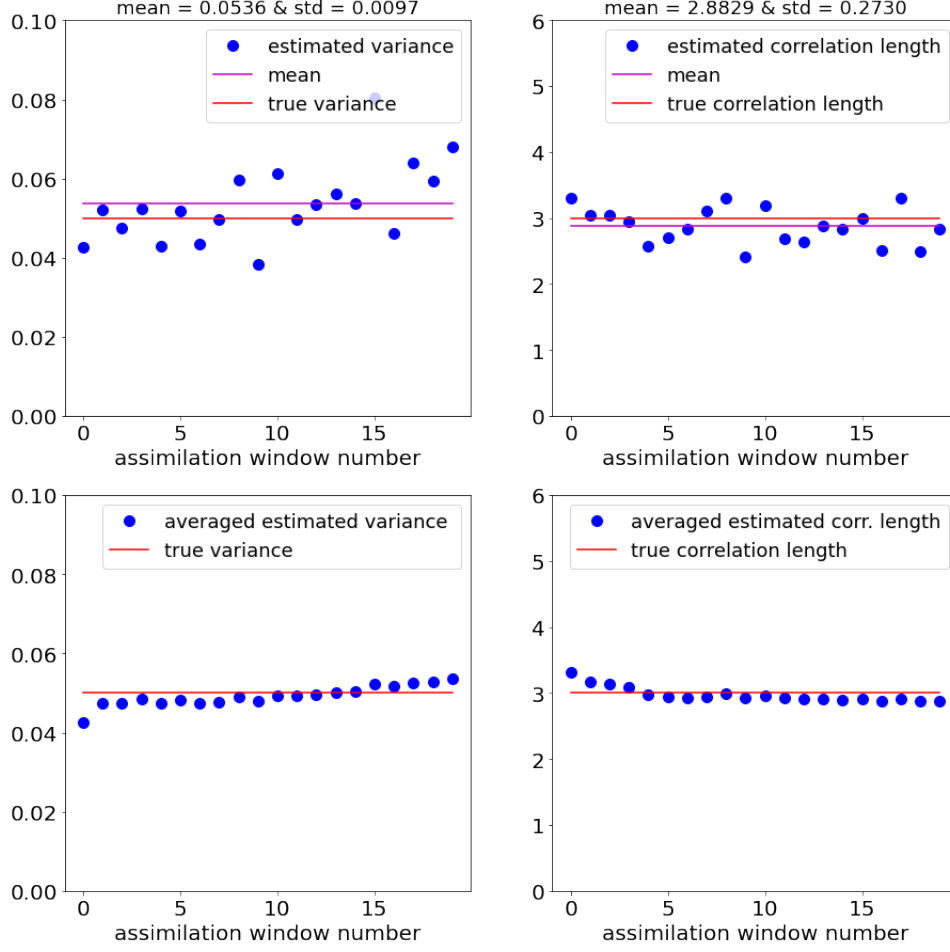
### Experiment 3 - partial observation in time

Same parameters as in Experiment 1, the only difference is that we observe only every second time step, i.e.  $m_t = 3$  and  $\Delta m_t = 2$ .



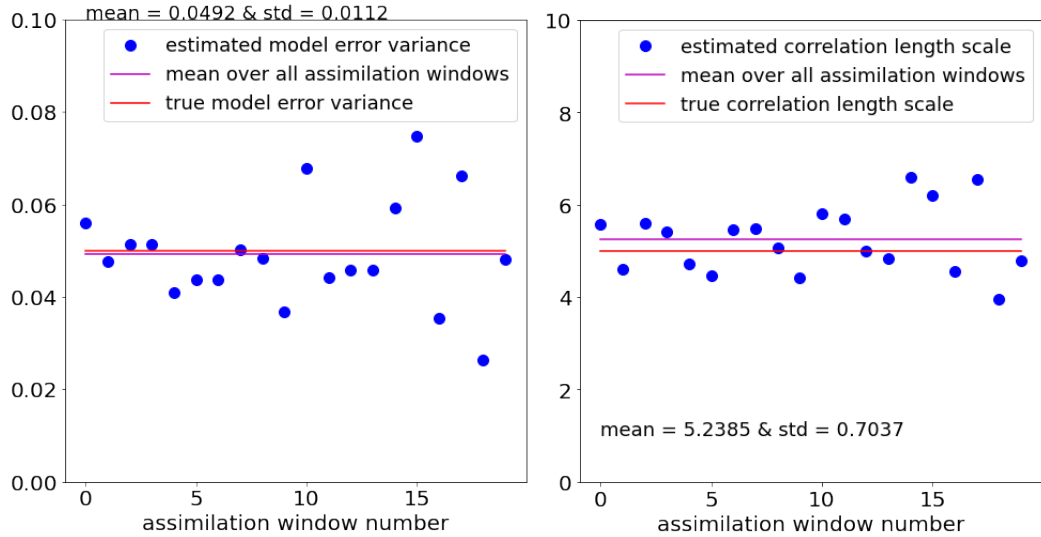
#### Experiment 4 - partial observation in time and space

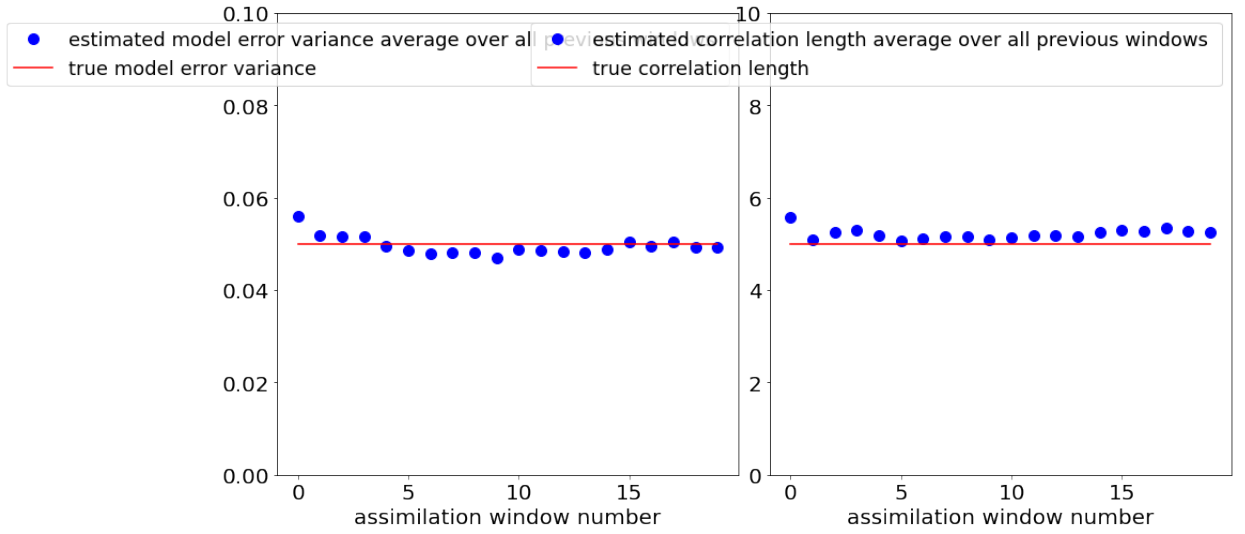
Same parameters as in Experiment 1, the only difference is that we observe only every second time step and every second time step, i.e.  $m_t = 3$  and  $\Delta m_t = 2$  and  $m_x = 100$  and  $\Delta m_x = 2$ .



#### Experiment 5 - partial observation in time and space, complicated $B$

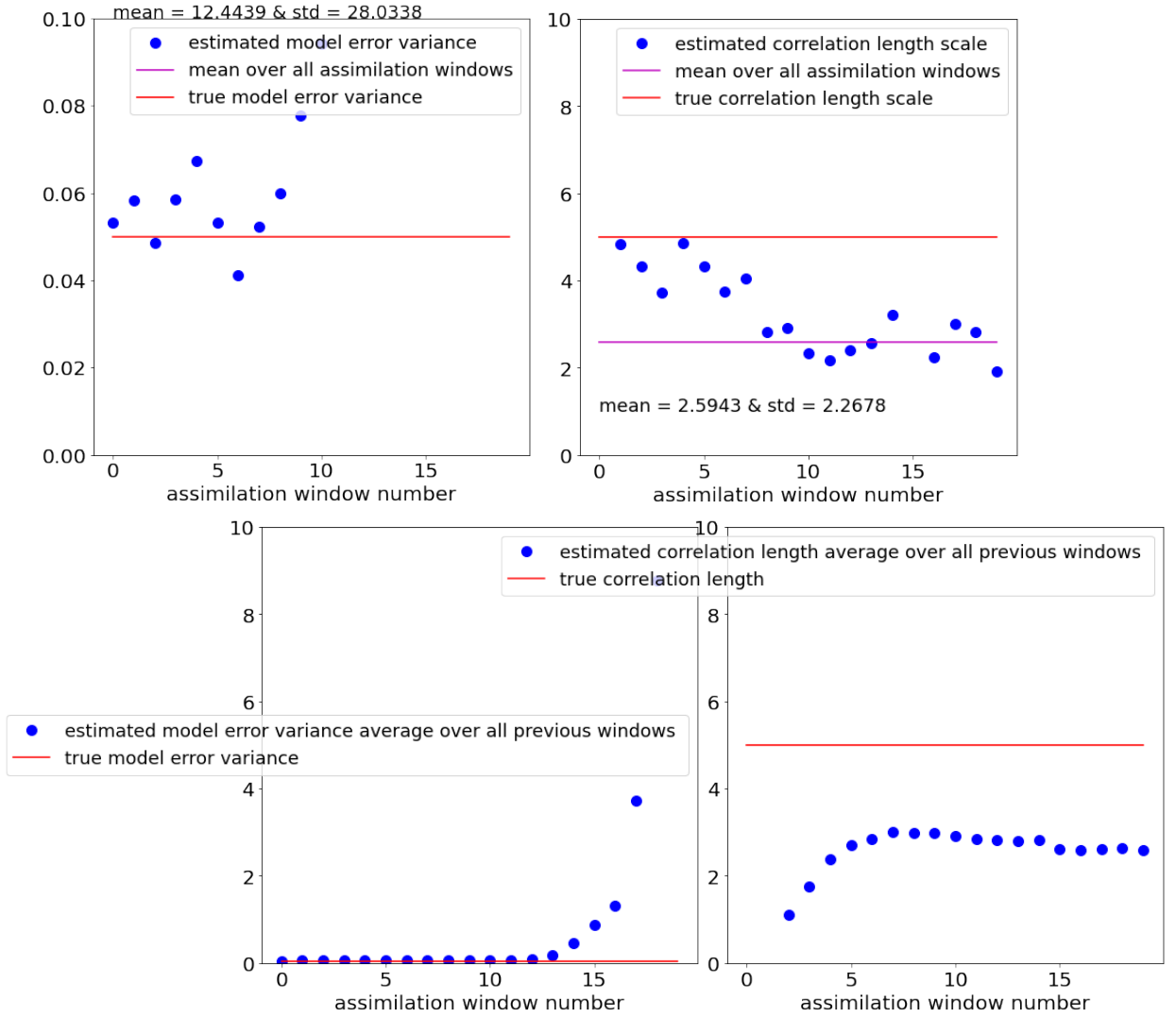
Same parameters as in Experiment 1, the only difference is that we observe only every second time step and every second time step, i.e.  $m_t = 3$  and  $\Delta m_t = 2$  and  $m_x = 100$  and  $\Delta m_x = 2$ . Moreover,  $\ell_t = 5$  and we assume  $B$  has variance 0.01 and correlation length 3 and we use this as our initial guess for  $Q$ .





### Experiment 6 - partial observation in time and space, complicated $B$

Same as experiment 5, only difference is that only every fourth gridpoint is observed.



## 4 Conclusion

## 5 Convergence proof

need to think about this

It is surprisingly easy to proof convergence. Start with writing the prior estimate for  $Q$  as

$$Q_0 = Q_t + E_0 \quad (5.1)$$

where we assume  $\|E\| \ll \|Q_t\|$ . Let's now write the Kalman gain used in the data assimilation as:

$$\begin{aligned} K_0 &= D_0 H_z^T (H_z D_0 H_z^T + R)^{-1} \\ &= (D_t + E_0) H_z^T (H_z D_0 H_z^T + R + H_z E_0 H_z^T)^{-1} \\ &= (D_t + E_0) H_z^T (H_z D_0 H_z^T + R)^{-1} \left( I + H_z E_0 H_z^T (H_z D_0 H_z^T + R)^{-1} \right)^{-1} \\ &= D_t H_z^T (H_z D_0 H_z^T + R)^{-1} \left( I - H_z E_0 H_z^T (H_z D_0 H_z^T + R)^{-1} \right) + E_0 H_z^T (H_z D_0 H_z^T + R)^{-1} + O(\|E_0\|^2) \\ &= K_t - K_t H_z E_0 H_z^T (H_z D_0 H_z^T + R)^{-1} + E_0 H_z^T (H_z D_0 H_z^T + R)^{-1} + O(\|E_0\|^2) \\ &= K_t + (I - K_t H_z) K_E \end{aligned} \quad (5.2)$$

in which

$$K_E = E_0 H_z^T (H_z D_0 H_z^T + R)^{-1} \quad (5.3)$$

Let us now form

$$\begin{aligned} \langle z_d c^T \rangle &= \langle K(y - H_x x_b)(y - H_x x_b)^T H_x \rangle \\ &= K(H_z D H_z^T + R) H_x^T \\ &= D_t H_z^T H_x + (I - K_t H_z) E_0 H_z^T H_x \end{aligned} \quad (5.4)$$

This has to be equal the the right-hand-side in our expression above, leading to:

$$D_t H_z^T H_x + (I - K_t H_z) E_0 H_z^T H_x = (D_t + E_1) H_z^T H_x \quad (5.5)$$

such that

$$E_1 H_z^T H_x = (I - K_t H_z) E_0 H_z^T H_x \quad (5.6)$$

Since

$$\begin{aligned} \|(I - K_t H_z)\| &= \|(I - (D^{-1} + H_z^T R^{-1} H_z) H_z^T R^{-1} H_z)\| \\ &= \|((D^{-1} + H_z^T R^{-1} H_z)(D^{-1} + H_z^T R^{-1} H_z - H_z^T R^{-1} H_z))\| \\ &= \|((D^{-1} + H_z^T R^{-1} H_z) D^{-1})\| < 1 \end{aligned} \quad (5.7)$$

we see that  $E_n$  convergence to zero when we iterate the scheme.

However, it is important to note that the  $H_z^T H_x$  term projects on the observation space, so it might be that convergence is faster in the observed part than in the unobserved part of  $D$ , and hence of  $Q$ . Since we assume  $Q$  to be the same in observed and unobserved parts of the domain we can smooth out this difference in convergence by averaging over observed and unobserved gridpoints simultaneously. But it might be that we cannot estimate  $B$  with any accuracy (although  $H_z^T = P^{-T} H_x$  might spread information to time zero via  $P$ ).

## References

- [1] Gérald Desroziers, Loic Berre, Bernard Chapnik, and Paul Poli. Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 131(613):3385–3396, 2005.
- [2] Geir Evensen, Femke C Vossepoel, and Peter Jan van Leeuwen. *Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem*. Springer Nature, 2022.
- [3] Pierre Gauthier, Ping Du, Sylvain Heilliette, and Louis Garand. Convergence issues in the estimation of interchannel correlated observation errors in infrared radiance data. *Monthly Weather Review*, 146(10):3227–3239, 2018.
- [4] Richard Ménard. Error covariance estimation methods based on analysis residuals: Theoretical foundation and convergence properties derived from simplified observation networks. *Quarterly Journal of the Royal Meteorological Society*, 142(694):257–273, 2016.
- [5] Ricardo Todling. A complementary note to ‘a lag-1 smoother approach to system-error estimation’: the intrinsic limitations of residual diagnostics. *Quarterly Journal of the Royal Meteorological Society*, 141(692):2917–2922, 2015.