4DEnVar

n.douglas, t.quaife, r.bannister

February 2023

These notes on 4DEnVar follow the derivations given in [1] and [2]. Some basic prior knowledge of data assimilation and linear algebra is assumed. We begin by reviewing 4DVar for the required nomenclature and proceed by detailing the 4DEnVar method including the generation of the posterior ensemble.

1 4DVar

1.1 State Vector

In 4DVar we consider the dynamical nonlinear discretised system

$$\mathbf{x}_{t+1} = \mathbf{m}_t\left(\mathbf{x}_t\right)$$

for $t = 0, 1, 2, \ldots, N$, where $\mathbf{x}_t \in \mathbb{R}^n$ is the state vector (which may represent state, parameters or a concatenation of both) and $\mathbf{m}_t$ maps the state vector at time t to the state vector at time $t+1$. Here we are assuming strong constraints, hence the absence of a process error term.

1.2 Observations

The vector $\mathbf{y}_t \in \mathbb{R}^{r_t}$ represents the available observations at time t. State space is mapped to observation space via the equation

$$\mathbf{y}_t = \mathbf{h}_t\left(\mathbf{x}_t\right) + \epsilon_t$$

where $\mathbf{h}_t : \mathbb{R}^n \to \mathbb{R}^{r_t}$ is the nonlinear observation operator and $\epsilon_t \in \mathbb{R}^{r_t}$ denotes the observation errors. Errors are assumed to be unbiased, Gaussian and uncorrelated in time with known covariance matrices $\mathbf{R}_t$:

$$\bar{\epsilon}_t \sim N\left(\mathbf{0}, \mathbf{R}_t\right)$$

1.3 Prior State

In 4DVar we require a prior distribution for the state vector at time 0 denoted by $\mathbf{x}_\mathrm{b}$ (‘b’ for ‘background’ synonymous with ‘prior’) along with a known covariance matrix $\mathbf{B}$. This prior distribution is assumed to be Gaussian.

1.4 Cost Function

The aim of 4DVar is to find an optimised initial state vector, $\mathbf{x}_0$, that optimises a cost function that sums the distance to the prior estimate with the distance of the simulated observations to the actual observations weighted by the inverses of their respective covariance matrices:

$$\begin{aligned} J(\mathbf{x}_0) &= \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_\mathrm{b})^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_\mathrm{b}) + \frac{1}{2}\sum_{t=0}^{N}(\mathbf{h}_t(\mathbf{x}_t) - \mathbf{y}_t)^T \mathbf{R}_t^{-1}(\mathbf{h}_t(\mathbf{x}_t) - \mathbf{y}_t) \\ &= \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_\mathrm{b})^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_\mathrm{b}) + \frac{1}{2}\sum_{t=0}^{N}(\mathbf{h}_t(\mathbf{m}_{0\to t}(\mathbf{x}_0)) - \mathbf{y}_t)^T \mathbf{R}_t^{-1}(\mathbf{h}_t(\mathbf{m}_{0\to t}(\mathbf{x}_0)) - \mathbf{y}_t) \\ &= \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_\mathrm{b})^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_\mathrm{b}) + \frac{1}{2}\left(\hat{\mathbf{h}}(\mathbf{x}_0) - \hat{\mathbf{y}}\right)^T \hat{\mathbf{R}}^{-1}\left(\hat{\mathbf{h}}(\mathbf{x}_0) - \hat{\mathbf{y}}\right) \end{aligned} \tag{1}$$

where $\mathbf{m}_{0\to t}(\mathbf{x}_0) = \mathbf{m}_{t-1}(\mathbf{m}_{t-2}...(\mathbf{m}_0(\mathbf{x}_0)))$ maps the state vector at time 0 to time t, where

$$\hat{\mathbf{y}} = \begin{pmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{pmatrix}, \quad \hat{\mathbf{h}}(\mathbf{x}_0) = \begin{pmatrix} \mathbf{h}_0(\mathbf{x}_0) \\ \mathbf{h}_1(\mathbf{m}_0(\mathbf{x}_0)) \\ \vdots \\ \mathbf{h}_N(\mathbf{m}_{0\to N}(\mathbf{x}_0)) \end{pmatrix},$$

and

$$\hat{\mathbf{R}} = \begin{pmatrix} \mathbf{R}_0 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_1 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{R}_N \end{pmatrix}$$

The matrix $\hat{\mathbf{R}}$ is a diagonal matrix - the off-diagonal elements are 0 when the observations are uncorrelated. The state that minimizes the cost function is often called the analysis or posterior estimate. The number of observations available is often much smaller than the dimension of the state vector. In these cases, the optimisation problem is ill-posed and including a prior term in the cost function regularizes the problem and ensures a locally unique solution.

1.5 Optimisation

The posterior estimate can be found by inputting the cost function, prior estimate and the gradient of the cost function into a gradient-based descent algorithm. The gradient of the cost function is given by

$$\nabla J(\mathbf{x}_0) = \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_\mathrm{b}) + \mathbb{J}\left(\hat{\mathbf{h}}(\mathbf{x}_0) - \hat{\mathbf{y}}\right)^T \hat{\mathbf{R}}^{-1}\left(\hat{\mathbf{h}}(\mathbf{x}_0) - \hat{\mathbf{y}}\right)$$

where we have used $\nabla\left(\mathbf{x}^T A\mathbf{x}\right) = 2A\mathbf{x}$ and $\nabla\left(\mathbf{f}(\mathbf{x})^T A\mathbf{f}(\mathbf{x})\right) = 2\mathbb{J}(\mathbf{f})^T A\mathbf{f}(\mathbf{x})$ for symmetric A where $\mathbb{J}$ is the Jacobian:

$$\mathbb{J}\left(\hat{\mathbf{h}}(\mathbf{x}_0) - \hat{\mathbf{y}}\right) = \begin{pmatrix} \mathbb{J}(\mathbf{h}_0(\mathbf{x}_0)) \\ \mathbb{J}(\mathbf{h}_1(\mathbf{m}_0(\mathbf{x}_0))) \\ \vdots \\ \mathbb{J}(\mathbf{h}_N(\mathbf{m}_{0\to N}(\mathbf{x}_0))) \end{pmatrix} = \begin{pmatrix} \mathbb{J}(\mathbf{h}_0) \\ \mathbb{J}(\mathbf{h}_1)\,\mathbb{J}(\mathbf{m}_0) \\ \vdots \\ \mathbb{J}(\mathbf{h}_N)\,\mathbb{J}(\mathbf{m}_{N-1})...\mathbb{J}(\mathbf{m}_0) \end{pmatrix}$$

(see accompanying notes for Taylor Expansion - Eq. 1). Note that the Jacobians are evaluated at different states: $\mathbb{J}(\mathbf{h}_i)$ and $\mathbb{J}(\mathbf{m}_i)$ are evaluated at $\mathbf{x}_i$. We define $\mathbb{J}\left(\mathbf{h}_i\right) \doteq \mathbf{H}_i$ (linearised observation operator) and $\mathbb{J}\left(\mathbf{m}_i\right) \doteq \mathbf{M}_i$ (tangent linear model). It follows that

$$\nabla J\left(\mathbf{x}_0\right) = \mathbf{B}^{-1}\left(\mathbf{x}_0 - \mathbf{x}_\mathrm{b}\right) + \hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1}\left(\hat{\mathbf{h}}\left(\mathbf{x}_0\right) - \hat{\mathbf{y}}\right)$$

where

$$\hat{\mathbf{H}} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1\mathbf{M}_0 \\ \vdots \\ \mathbf{H}_N\mathbf{M}_{N,0} \end{pmatrix}$$

with $\mathbf{M}_{t,0} = \mathbf{M}_{t-1}\mathbf{M}_{t-2}\cdots\mathbf{M}_0$.

1.6 Problems

Both the linearised observation operator and the tangent linear model can be difficult to compute. In addition, issues may arise when the prior error covariance matrix $\mathbf{B}$ becomes large, ill-conditioned and difficult to invert. As a result, finding the minimising state for the cost function in Eq. (1) can be slow. To ensure the 4DVar cost function converges as efficiently as possible and to avoid the explicit computation of the matrix $\mathbf{B}$, the problem is often preconditioned using a control variable transform.

1.7 Pre-Conditioning

We define the preconditioning matrix $\mathbf{U}$ by

$$\mathbf{B} = \mathbf{U}\mathbf{U}^T \tag{2}$$

and

$$\mathbf{x}_0 = \mathbf{x}_\mathrm{b} + \mathbf{U}\mathbf{w} \tag{3}$$

so that $\mathbf{w} = \mathbf{U}^{-1}\left(\mathbf{x}_0 - \mathbf{x}_\mathrm{b}\right) \in \mathbb{R}^m$. Substituting Eq. (2) and (3) into Eq. (1) we find the new cost function:

$$J(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \frac{1}{2}\left(\hat{\mathbf{h}}\left(\mathbf{x}_\mathrm{b} + \mathbf{U}\mathbf{w}\right) - \hat{\mathbf{y}}\right)^T \hat{\mathbf{R}}^{-1}\left(\hat{\mathbf{h}}\left(\mathbf{x}_\mathrm{b} + \mathbf{U}\mathbf{w}\right) - \hat{\mathbf{y}}\right).$$

where we have used

$$\left(\mathbf{x}_0 - \mathbf{x}_\mathrm{b}\right)^T \mathbf{B}^{-1}\left(\mathbf{x}_0 - \mathbf{x}_\mathrm{b}\right) = \left(\mathbf{U}\mathbf{w}\right)^T\left(\mathbf{U}\mathbf{U}^T\right)^{-1}\mathbf{U}\mathbf{w} = \mathbf{w}^T\mathbf{U}^T\left(\mathbf{U}^T\right)^{-1}\mathbf{U}^{-1}\mathbf{U}\mathbf{w} = \mathbf{w}^T\mathbf{w}$$

Note that we can take a first order Taylor expansion of the term $\hat{\mathbf{h}}(\mathbf{x}_b + \mathbf{U}\mathbf{w})$ in the cost function:

$$\hat{\mathbf{h}}\left(\mathbf{x}_b + \mathbf{U}\mathbf{w}\right) = \begin{pmatrix} \mathbf{h}_0\left(\mathbf{x}_b + \mathbf{U}\mathbf{w}\right) \\ \mathbf{h}_1\left(\mathbf{m}_0\left(\mathbf{x}_b + \mathbf{U}\mathbf{w}\right)\right) \\ \vdots \\ \mathbf{h}_N\left(\mathbf{m}_{0\to N}\left(\mathbf{x}_b + \mathbf{U}\mathbf{w}\right)\right) \end{pmatrix}$$

$$
\begin{aligned}
&\approx \begin{pmatrix} \mathbf{h}_0\left(\mathbf{x}_b\right)+\mathbb{J}(\mathbf{h}_0)\mathbf{U}\mathbf{w} \\ \mathbf{h}_1\left(\mathbf{m}_0\left(\mathbf{x}_b\right)\right)+\mathbb{J}(\mathbf{h}_1)\mathbb{J}(\mathbf{m}_0)\mathbf{U}\mathbf{w} \\ \vdots \\ \mathbf{h}_N\left(\mathbf{m}_{0\to N}\left(\mathbf{x}_b\right)\right)+\mathbb{J}(\mathbf{h}_N)\mathbb{J}(\mathbf{m}_{N-1})...\mathbb{J}(\mathbf{m}_0)\mathbf{U}\mathbf{w} \end{pmatrix} \\
&= \mathbf{h}(\mathbf{x}_b)+\hat{\mathbf{H}}\mathbf{U}\mathbf{w}
\end{aligned} \tag{4}
$$

(see accompanying notes for Taylor Expansion). Hence, we can approximate the cost function as

$$
J(\mathbf{w})=\frac{1}{2}\mathbf{w}^T\mathbf{w}+\frac{1}{2}\left(\hat{\mathbf{h}}\left(\mathbf{x}_{\mathrm{b}}\right)+\hat{\mathbf{H}}\mathbf{U}\mathbf{w}-\hat{\mathbf{y}}\right)^T\hat{\mathbf{R}}^{-1}\left(\hat{\mathbf{h}}\left(\mathbf{x}_{\mathrm{b}}\right)+\hat{\mathbf{H}}\mathbf{U}\mathbf{w}-\hat{\mathbf{y}}\right) \tag{5}
$$

with the gradient of the cost function given as

$$
\nabla J(\mathbf{w})=\mathbf{w}+\mathbf{U}^T\hat{\mathbf{H}}^T\hat{\mathbf{R}}^{-1}\left(\hat{\mathbf{h}}\left(\mathbf{x}_{\mathrm{b}}\right)+\hat{\mathbf{H}}\mathbf{U}\mathbf{w}-\hat{\mathbf{y}}\right) \tag{6}
$$

As the square root of a matrix is not unique there will be multiple choices for the preconditioning matrix $\mathbf{U}$.

2 4DEnVar

Given an ensemble of m joint state parameter vectors, we can define the perturbation matrix

$$
\mathbf{X}'_{\mathrm{b}}=\frac{1}{\sqrt{m-1}}\left(\mathbf{X}_{\mathrm{b},1}-\overline{\mathbf{x}}_{\mathrm{b}},\mathbf{X}_{\mathrm{b},2}-\overline{\mathbf{x}}_{\mathrm{b}},\ldots,\mathbf{X}_{\mathrm{b},m}-\overline{\mathbf{x}}_{\mathrm{b}}\right) \tag{7}
$$

Here the m ensemble members $\mathbf{X}_{b,i}$ can come from a previous forecast (in which case $\overline{\mathbf{x}}_{\mathrm{b}}$ is the mean of the m ensemble members) or from a known distribution $\mathcal{N}\left(\mathbf{x}_{\mathrm{b}},\mathbf{B}\right)$ such that $\overline{\mathbf{x}}_{\mathrm{b}}=\mathbf{x}_{\mathrm{b}}$. Using $\mathbf{X}'_{\mathrm{b}}$ we can approximate the background or prior error covariance matrix by

$$
\mathbf{B}\approx\mathbf{X}'_{\mathrm{b}}\left(\mathbf{X}'_{\mathrm{b}}\right)^T
$$

We can then transform to ensemble space using the matrix $\mathbf{X}'_{\mathrm{b}}$ as our preconditioning matrix by defining

$$
\mathbf{x}_0=\overline{\mathbf{x}}_{\mathrm{b}}+\mathbf{X}'_{\mathrm{b}}\mathbf{w} \tag{8}
$$

where $\mathbf{w}$ is a vector of length m. Defining $\mathbf{x}_0$ in this way reduces the problem in cases where the state or parameter vector is much larger than the ensemble size and also regularises the problem in cases where the state or parameter vector contains elements of contrasting orders of magnitude. Setting $\mathbf{U}=\mathbf{X}'_b$, Equations (5) and (6) become

$$
J(\mathbf{w})=\frac{1}{2}\mathbf{w}^T\mathbf{w}+\frac{1}{2}\left(\hat{\mathbf{H}}\mathbf{X}'_{\mathrm{b}}\mathbf{w}+\hat{\mathbf{h}}\left(\overline{\mathbf{x}}_{\mathrm{b}}\right)-\hat{\mathbf{y}}\right)^T\hat{\mathbf{R}}^{-1}\left(\hat{\mathbf{H}}\mathbf{X}'_{\mathrm{b}}\mathbf{w}+\hat{\mathbf{h}}\left(\overline{\mathbf{x}}_{\mathrm{b}}\right)-\hat{\mathbf{y}}\right)
$$

with gradient

$$
\begin{aligned}
\nabla J(\mathbf{w}) &= \mathbf{w}+\mathbf{X}_{\mathrm{b}}^T\hat{\mathbf{H}}^T\hat{\mathbf{R}}^{-1}\left(\hat{\mathbf{H}}\mathbf{X}'_{\mathrm{b}}\mathbf{w}+\hat{\mathbf{h}}\left(\overline{\mathbf{x}}_{\mathrm{b}}\right)+\hat{\mathbf{y}}\right) \\
&= \mathbf{w}+\left(\hat{\mathbf{H}}\mathbf{X}'_{\mathrm{b}}\right)^T\hat{\mathbf{R}}^{-1}\left(\hat{\mathbf{H}}\mathbf{X}'_{\mathrm{b}}\mathbf{w}+\hat{\mathbf{h}}\left(\overline{\mathbf{x}}_{\mathrm{b}}\right)-\hat{\mathbf{y}}\right)
\end{aligned}
$$

We can see that the tangent linear model and adjoint are still present in these equations within $\hat{\mathbf{H}}$. However, we can approximate

$$\hat{\mathbf{H}}\mathbf{X}'_{\mathrm{b}} \approx \mathbf{Y}'_b \doteq \frac{1}{\sqrt{m-1}}\left(\hat{\mathbf{h}}\left(\mathbf{X}_{\mathrm{b},1}\right)-\hat{\mathbf{h}}\left(\overline{\mathbf{x}}_{\mathrm{b}}\right), \hat{\mathbf{h}}\left(\mathbf{X}_{\mathrm{b},2}\right)-\hat{\mathbf{h}}\left(\overline{\mathbf{x}}_{\mathrm{b}}\right), \ldots, \hat{\mathbf{h}}\left(\mathbf{X}_{\mathrm{b},m}\right)-\hat{\mathbf{h}}\left(\overline{\mathbf{x}}_{\mathrm{b}}\right)\right) \tag{9}$$

to get

$$J(\mathbf{w})=\frac{1}{2}\mathbf{w}^T\mathbf{w}+\frac{1}{2}\left(\mathbf{Y}'_b\mathbf{w}+\hat{\mathbf{h}}\left(\overline{\mathbf{x}}_{\mathrm{b}}\right)-\hat{\mathbf{y}}\right)^T\hat{\mathbf{R}}^{-1}\left(\mathbf{Y}'_b\mathbf{w}+\hat{\mathbf{h}}\left(\overline{\mathbf{x}}_{\mathrm{b}}\right)-\hat{\mathbf{y}}\right) \tag{10}$$

and

$$\nabla J(\mathbf{w})=\mathbf{w}+\left(\mathbf{Y}'_b\right)^T\hat{\mathbf{R}}^{-1}\left(\mathbf{Y}'_b\mathbf{w}+\hat{\mathbf{h}}\left(\overline{\mathbf{x}}_{\mathrm{b}}\right)-\hat{\mathbf{y}}\right) \tag{11}$$

where we have used the Taylor expansion $\hat{\mathbf{h}}\left(\mathbf{x}+\delta\mathbf{x}\right)\approx\hat{\mathbf{h}}(\mathbf{x})+\hat{\mathbf{H}}\delta\mathbf{x}$ with $\delta\mathbf{x}=\sqrt{m-1}\mathbf{X}'_{b,i}$ to get

$$\begin{aligned}\sqrt{m-1}\hat{\mathbf{H}}\mathbf{X}'_{b,i} &\approx \hat{\mathbf{h}}\left(\overline{\mathbf{x}}_b+\sqrt{m-1}\mathbf{X}'_{b,i}\right)-\hat{\mathbf{h}}\left(\overline{\mathbf{x}}_b\right)=\hat{\mathbf{h}}\left(\overline{\mathbf{x}}_b+\mathbf{X}_{b,i}-\overline{\mathbf{x}}_b\right)-\hat{\mathbf{h}}\left(\overline{\mathbf{x}}_b\right)\\ &= \hat{\mathbf{h}}\left(\mathbf{X}_{b,i}\right)-\hat{\mathbf{h}}\left(\overline{\mathbf{x}}_b\right)\end{aligned} \tag{12}$$

(see accompanying notes for Taylor Expansion). Equations (10) to (12) use the full nonlinear model and nonlinear observation operator and thus avoid the computation of the tangent linear and adjoint models. Computing the minimum of the cost function Eq. (10) using gradient Eq. (11) yields the maximum a posteriori estimate $\mathbf{w}_a$, which inserted into Eq. (8) gives us the maximum a posteriori estimate of the state vector $\mathbf{x}_a$.

2.1 The Analytical Solution

For relatively small matrices, it is simple enough to set the gradient function to $\mathbf{0}$ and solve for $\mathbf{w}$:

$$\begin{aligned}\nabla J(\mathbf{w}) &= \mathbf{0}\\ \Longrightarrow \mathbf{w}+\left(\mathbf{Y}'_b\right)^T\hat{\mathbf{R}}^{-1}\mathbf{Y}'_b\mathbf{w} &= -\left(\mathbf{Y}'_b\right)^T\hat{\mathbf{R}}^{-1}\left(\hat{\mathbf{h}}\left(\overline{\mathbf{x}}_{\mathrm{b}}\right)-\hat{\mathbf{y}}\right)\\ \Longrightarrow \mathbf{w} &= -\left(\mathbf{I}+\left(\mathbf{Y}'_b\right)^T\hat{\mathbf{R}}^{-1}\mathbf{Y}'_b\right)^{-1}\left(\mathbf{Y}'_b\right)^T\hat{\mathbf{R}}^{-1}\left(\hat{\mathbf{h}}\left(\overline{\mathbf{x}}_{\mathrm{b}}\right)-\hat{\mathbf{y}}\right)\\ &= -\left(\mathbf{Y}'_b\right)^T\left(\hat{\mathbf{R}}+\mathbf{Y}'_b\left(\mathbf{Y}'_b\right)^T\right)^{-1}\left(\hat{\mathbf{h}}\left(\overline{\mathbf{x}}_{\mathrm{b}}\right)-\hat{\mathbf{y}}\right)\end{aligned}$$

using the Sherman-Morrison-Woodbury Formula.

2.2 Calculating the posterior ensemble

The analysis error covariance matrix $\mathbf{P}_a$ is given by

$$\mathbf{P}_a=\left(\mathbf{I}-\mathbf{K}\hat{\mathbf{H}}\right)\mathbf{P}_b \quad\Longrightarrow\quad \mathbf{X}'_a\mathbf{X}'^T_a\approx\left(\mathbf{I}-\mathbf{K}\hat{\mathbf{H}}\right)\mathbf{X}'_b\mathbf{X}'^T_b$$

where $\mathbf{K}$ is the Kalman gain matrix

$$\mathbf{K}=\mathbf{P}_b\hat{\mathbf{H}}^T\left(\hat{\mathbf{R}}+\hat{\mathbf{H}}\mathbf{P}_b\hat{\mathbf{H}}^T\right)^{-1}\approx\mathbf{X}'_b\mathbf{X}'^T_b\hat{\mathbf{H}}^T\left(\hat{\mathbf{R}}+\hat{\mathbf{H}}\mathbf{X}'_b\mathbf{X}'^T_b\hat{\mathbf{H}}^T\right)^{-1}=\mathbf{X}'_b\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]^T\left(\hat{\mathbf{R}}+\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]^T\right)^{-1}$$

In all that follows $\hat{\mathbf{H}}\mathbf{X}'_b$ can be approximated with $\mathbf{Y}'_\mathbf{b}$ to avoid the computation of the tangent linear model and linearised observation operator. The analysis error covariance matrix becomes:

$$
\begin{aligned}
\mathbf{X}'_a\mathbf{X}'^T_a &= \left(\mathbf{I}-\mathbf{X}'_b\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]^T\left(\hat{\mathbf{R}}+\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]^T\right)^{-1}\hat{\mathbf{H}}\right)\mathbf{X}'_b\mathbf{X}'^T_b \\
&= \mathbf{X}'_b\left(\mathbf{I}-\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]^T\left(\hat{\mathbf{R}}+\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]^T\right)^{-1}\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]\right)\mathbf{X}'^T_b \\
&= \mathbf{X}'_b\left(\mathbf{I}+\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]^T\hat{\mathbf{R}}^{-1}\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]\right)^{-1}\mathbf{X}'^T_b \qquad (13)
\end{aligned}
$$

where we have used the following identity (from the Sherman-Morrison-Woodbury formula):

$$
\mathbf{A}^T\left(\hat{\mathbf{R}}+\mathbf{A}\mathbf{A}^T\right)^{-1}=\left(\mathbf{I}+\mathbf{A}^T\hat{\mathbf{R}}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^T\hat{\mathbf{R}}^{-1}
$$

with $\mathbf{A}=\hat{\mathbf{H}}\mathbf{X}'_b$ to rewrite

$$
\begin{aligned}
&\mathbf{I}-\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]^T\left(\hat{\mathbf{R}}+\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]^T\right)^{-1}\left[\hat{\mathbf{H}}\mathbf{X}'_b\right] \\
=\ &\mathbf{I}-\left(\mathbf{I}+\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]^T\hat{\mathbf{R}}^{-1}\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]\right)^{-1}\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]^T\hat{\mathbf{R}}^{-1}\left[\hat{\mathbf{H}}\mathbf{X}'_b\right] \\
=\ &\left(\mathbf{I}+\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]^T\hat{\mathbf{R}}^{-1}\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]\right)^{-1}\left(\mathbf{I}+\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]^T\hat{\mathbf{R}}^{-1}\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]\right) \\
&\qquad-\left(\mathbf{I}+\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]^T\hat{\mathbf{R}}^{-1}\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]\right)^{-1}\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]^T\hat{\mathbf{R}}^{-1}\left[\hat{\mathbf{H}}\mathbf{X}'_b\right] \\
=\ &\left(\mathbf{I}+\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]^T\hat{\mathbf{R}}^{-1}\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]\right)^{-1}\left\{\left(\mathbf{I}+\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]^T\hat{\mathbf{R}}^{-1}\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]\right)-\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]^T\hat{\mathbf{R}}^{-1}\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]\right\} \\
=\ &\left(\mathbf{I}+\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]^T\hat{\mathbf{R}}^{-1}\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]\right)^{-1}
\end{aligned}
$$

It follows that

$$
\mathbf{X}'_a=\mathbf{X}'_b\left(\mathbf{I}+\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]^T\hat{\mathbf{R}}^{-1}\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]\right)^{-1/2} \qquad (14)
$$

Provided that the real matrix $\mathbf{I}+\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]^T\hat{\mathbf{R}}^{-1}\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]$ is symmetric and has positive eigenvalues, a unique square root matrix can be found from a Cholesky decomposition. Inverting the result will provide the matrix of weights to apply to the prior ensemble perturbation matrix:

$$
\mathbf{W}_a=\left(\mathbf{I}+\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]^T\hat{\mathbf{R}}^{-1}\left[\hat{\mathbf{H}}\mathbf{X}'_b\right]\right)^{-1/2} \qquad (15)
$$

Note that very small negative eigenvalues can often appear due to bad conditioning. They can be removed using a 'ridge regression', which is a very small positive increment in the eigenvalues to make them all positive.

Finally, note that the posterior perturbation matrix is defined as

$$\mathbf{X}'_a = \frac{1}{\sqrt{m-1}} \left(\mathbf{X}_{a,1} - \overline{\mathbf{x}}_a, \mathbf{X}_{a,2} - \overline{\mathbf{x}}_a, \ldots, \mathbf{X}_{a,m} - \overline{\mathbf{x}}_a\right)$$

and so the posterior ensemble can be obtained via

$$\mathbf{X}_{\mathrm{a}} = \left(\mathbf{x}_{\mathrm{a}} + \sqrt{m-1}\mathbf{X}'_{a,1}, \mathbf{x}_{\mathrm{a}} + \sqrt{m-1}\mathbf{X}'_{a,2}, \ldots, \mathbf{x}_{\mathrm{a}} + \sqrt{m-1}\mathbf{X}'_{a,m}\right) \tag{16}$$

As well as avoiding the calculation of the linearised observation operator and the tangent linear model, 4DEnVar also provides posterior uncertainty information.

3 References

[1] Ewan Pinnington, Tristan Quaife, Amos Lawless, Karina Williams, Tim Arkebauer, and Dave Scoby. The Land Variational Ensemble Data Assimilation Framework: LAVENDAR v1.0.0. Geoscientific Model Development, 13: 55–69, 2020.
[2] Ewan Pinnington, Javier Amezcua, Elizabeth Cooper, Simon Dadson, Rich Ellis, Jian Peng, Emma Robinson, Ross Morrison, Simon Osborne, and Tristan Quaife. Improving soil moisture prediction of a high-resolution land surface model by parameterising pedotransfer functions through assimilation of SMAP satellite data. Hydrology and Earth System Sciences, 25(3):1617-1641, 2021.