Supporting Information for "Leveraging Uncertainty Quantification to Design Ocean Climate Observing Systems"

Nora Loose[1], Patrick Heimbach[1,2,3]

[1]Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, TX, USA

[2]Jackson School of Geosciences, The University of Texas at Austin, Austin, TX, USA

[3]Institute for Geophysics, The University of Texas at Austin, Austin, TX, USA

Contents of this file

Text S1. Inverse Uncertainty Propagation. The Bayesian approach states the deterministic inverse problem (eq. (1)) as one of Bayesian inference over the space of unknown control variables, which are to be inferred from the observations and the ocean GCM dynamics. The solution is given by the posterior probability density function $\pi_{\text{post}}(\mathbf{u}|\mathbf{y}) \propto e^{-J(\mathbf{u})}$. Hence, the deterministic and Bayesian formulation of the inverse problem are interconnected by the fact that the deterministic least squares cost function J is the negative log-posterior in the Bayesian interpretation. Furthermore, the deterministic solution $\mathbf{u}_{\text{min}}$ is the Maximum a Posteriori (MAP) point, i.e., the most likely

solution, in the Bayesian framework. For more details, the reader is referred to the books by Tarantola (2005); Law, Stuart, and Zygalakis (2015).

To make the computation of the posterior probability density function $\pi_{\text{post}}(\mathbf{u}|\mathbf{y})$ computationally tractable, a linearization of the observation operator about the MAP Point, $\mathbf{u}_{\text{min}}$, is necessary (e.g., Bui-Thanh et al., 2012). This yields

$$\mathbf{Obs}(\mathbf{u}) \approx \mathbf{Obs}(\mathbf{u}_{\text{min}}) + \mathbf{A}(\mathbf{u} - \mathbf{u}_{\text{min}}), \tag{S.1}$$

where $\mathbf{A} = \frac{\partial(\mathbf{Obs})}{\partial \mathbf{u}}_{|\mathbf{u}_{\text{min}}}$ is the Jacobian matrix of the observation operator $\mathbf{u} \mapsto \mathbf{Obs}(\mathbf{u})$, evaluated at $\mathbf{u}_{\text{min}}$. The posterior distribution $\pi_{\text{post}}(\mathbf{u}|\mathbf{y}) \propto e^{-J(\mathbf{u})}$ then becomes

$$\pi_{\text{post}}(\mathbf{u}|\mathbf{y}) \propto C \cdot \exp\left(-\frac{1}{2}\left[(\mathbf{u} - \mathbf{u}_{\text{min}})^T (\mathbf{A}^T\,\mathbf{R}^{-1}\,\mathbf{A} + \mathbf{B}^{-1}\,)^{-1} (\mathbf{u} - \mathbf{u}_{\text{min}})\right]\right), \tag{S.2}$$

where C is a constant factor, given by

$$C = \exp\left(-\frac{1}{2}\Big[(\mathbf{y} - \mathbf{Obs}(\mathbf{u}_{\text{min}}))^T\,\mathbf{R}^{-1}\,(\mathbf{y} - \mathbf{Obs}(\mathbf{u}_{\text{min}})) + (\mathbf{u}_{\text{min}} - \mathbf{u}_0)^T\,\mathbf{B}^{-1}\,(\mathbf{u}_{\text{min}} - \mathbf{u}_0)\Big]\right),$$

and can therefore be absorbed by the proportionality $\propto$. The right hand side of eq. (S.2) describes a Gaussian $\mathcal{N}(\mathbf{u}_{\text{min}}, \mathbf{P})$ with mean $\mathbf{u}_{\text{min}}$ and covariance matrix

$$\mathbf{P} = (\mathbf{A}^T\,\mathbf{R}^{-1}\,\mathbf{A} + \mathbf{B}^{-1})^{-1}. \tag{S.3}$$

The covariance matrix $\mathbf{P}$ (eq. (S.3)) is equal to $\mathbf{H}_J^{-1}$, the inverse of the linearized Hessian matrix of the cost function J (eq. (1)) at $\mathbf{u}_{\text{min}}$ (see also Thacker, 1989).

The linearized Hessian $\mathbf{H}_J$, in turn, is the sum of two matrices: $\mathbf{A}^T\,\mathbf{R}^{-1}\,\mathbf{A}$, which is the linearized Hessian of the model-data misfit term J_{misfit} (eq. (1)), and $\mathbf{B}^{-1}$, which is the Hessian of the regularization term J_{prior} (eq. (1)). It is the first matrix, $\mathbf{A}^T\,\mathbf{R}^{-1}\,\mathbf{A}$,

that characterizes the observational constraints on the control variables. Therefore, as a next step, we perform an eigen-decomposition of the misfit Hessian. This will give further insights into the model input components that are best determined by the observing system under consideration.

As a preparatory step, we rescale the model-data misfit term, J_{misfit}, through a change of variables $\tilde{\mathbf{u}} = \mathbf{B}^{-1/2}\,\mathbf{u}$. Here, $\mathbf{B}^{1/2}$ is an invertible square root of $\mathbf{B}$, i.e., satisfies $\mathbf{B}^{1/2}\,\mathbf{B}^{T/2} = \mathbf{B}$. The rescaling can be thought of as nondimensionalization if $\mathbf{B}$ was diagonal. The rescaling is necessary in order to treat all control variables equally, since they represent different physical variables, characteristic of different orders of magnitudes. In the new coordinates, the Hessian of $J_{\text{misfit}}(\tilde{\mathbf{u}})$ is given by the rescaled misfit Hessian (also referred to as the *prior-preconditioned* misfit Hessian, e.g., Bui-Thanh et al., 2012), equal to the $N \times N$ matrix

$$\mathbf{H}_{\text{misfit}} = \mathbf{B}^{T/2}\,\mathbf{A}^T\,\mathbf{R}^{-1}\,\mathbf{A}\,\mathbf{B}^{1/2}. \tag{S.4}$$

For simpicity, we will hereafter drop the term 'rescaled', and refer to $\mathbf{H}_{\text{misfit}}$ solely as the misfit Hessian. The misfit Hessian can be rewritten in terms of its eigen-decomposition:

$$\mathbf{H}_{\text{misfit}} = \sum_{i=1}^{M'} \lambda_i \mathbf{v}_i \mathbf{v}_i^T, \tag{S.5}$$

with an orthonormal set of eigenvectors $\{\mathbf{v}_i\}_{i=1}^{M'}$ and corresponsing eigenvalues $\lambda_i > 0$. M' is defined as the number of strictly positive eigenvalues, while all remaining eigenvalues λ_i, $i > M'$, are equal to zero.

The Woodbury Formula (e.g., Section 2.7.3 in Press et al., 2007) states that, given any $N \times N$ matrix $\mathbf{M}$ and any $N \times M'$ matrix $\mathbf{V}$, the following identity holds true:

$$(\mathbf{M} + \mathbf{V}\mathbf{V}^T)^{-1} = \mathbf{M}^{-1} - \mathbf{M}^{-1}\mathbf{V}\left(\mathbf{1}_{M'\times M'} + \mathbf{V}^T\mathbf{M}^{-1}\mathbf{V}\right)^{-1}\mathbf{V}^T\mathbf{M}^{-1}. \tag{S.6}$$

Here, $\mathbf{1}_{M'\times M'}$ is the $M' \times M'$ identity matrix. Assuming $\mathbf{M}^{-1}$ is already known, the formula provides an efficient way to compute the inverse of the sum of $\mathbf{M}$ and a low-rank matrix $\mathbf{V}\mathbf{V}^T$ (if $M' \ll N$). Using eqs. (S.3), (S.4), and (S.5), we have

$$\begin{aligned}\mathbf{P} &= \mathbf{B}^{1/2}\left(\mathbf{H}_{\text{misfit}} + \mathbf{1}_{N\times N}\right)^{-1}\mathbf{B}^{T/2} \\ &= \mathbf{B}^{1/2}\left(\sum_{i=1}^{M'}\lambda_i\mathbf{v}_i\mathbf{v}_i^T + \mathbf{1}_{N\times N}\right)^{-1}\mathbf{B}^{T/2}.\end{aligned} \tag{S.7}$$

The Woodbury Formula (eq. (S.6)) can now be applied to the inner piece in eq. (S.7), with $\mathbf{M} = \mathbf{1}_{N\times N}$ and $\mathbf{V}$ defined as the matrix formed by columns of $\sqrt{\lambda_i}\cdot\mathbf{v}_i$:

$$\mathbf{V} = \left[\sqrt{\lambda_1}\cdot\mathbf{v}_1\middle|\cdots\middle|\sqrt{\lambda_{M'}}\cdot\mathbf{v}_{M'}\right].$$

This yields

$$\mathbf{P} = \mathbf{B}^{1/2}\left(\mathbf{1}_{N\times N} - \sum_{i=1}^{M'}\frac{\lambda_i}{\lambda_i+1}\mathbf{v}_i\mathbf{v}_i^T\right)\mathbf{B}^{T/2} = \mathbf{B} - \sum_{i=1}^{M'}\frac{\lambda_i}{\lambda_i+1}\left(\mathbf{B}^{1/2}\mathbf{v}_i\right)\left(\mathbf{B}^{1/2}\mathbf{v}_i\right)^T,$$

using the fact that $\{\mathbf{v}_i\}_{i=1}^{M'}$ is a set of orthonormal vectors.

Text S2. Forward Uncertainty Propagation. Consistent with the linearization of the observation operator (eq. (S.1)), the function $\mathbf{u} \mapsto \text{QoI}(\mathbf{u})$ is linearized about $\mathbf{u}_{\text{min}}$:

$$\text{QoI}(\mathbf{u}) \approx \text{QoI}(\mathbf{u}_{\text{min}}) + \frac{\partial(\text{QoI})}{\partial\mathbf{u}}_{|\mathbf{u}_{\text{min}}}(\mathbf{u} - \mathbf{u}_{\text{min}}). \tag{S.8}$$

The posterior distribution of the Bayesian solution of the inverse problem, $\pi_{\text{post}}(\mathbf{u}|\mathbf{y})$, is approximately Gaussian, given by $\mathcal{N}(\mathbf{u}_{\text{min}}, \mathbf{P})$, with $\mathbf{P}$ given by eq. (2). A forward

propagation of the posterior uncertainty (dotted black arrow, (UQ2), Fig. 1) leads to a posterior Gaussian distribution for QoI($\mathbf{u}$), since eq. (S.8) describes an affine transformation. The distribution is given by $\mathcal{N}(\text{QoI}(\mathbf{u}_{\min}), (\sigma^{\mathbf{P}}_{\text{QoI}})^2)$, where $(\sigma^{\mathbf{P}}_{\text{QoI}})^2$ is the (scalar) variance, given by the projection

$$(\sigma^{\mathbf{P}}_{\text{QoI}})^2 = (\nabla_{\mathbf{u}}\text{QoI})^T \, \mathbf{P} \, (\nabla_{\mathbf{u}}\text{QoI}) \, .$$

Similarly, the prior distribution of QoI($\mathbf{u}$) is obtained by a forward uncertainty propagation of the Gaussian prior $\mathcal{N}(\mathbf{u}_0, \mathbf{B})$ (dotted green arrow, (UQ2), Fig. 1). This leads to a prior Gaussian distribution for QoI($\mathbf{u}$), given by $\mathcal{N}(\text{QoI}(\mathbf{u}_0), (\sigma^{\mathbf{B}}_{\text{QoI}})^2)$ with

$$(\sigma^{\mathbf{B}}_{\text{QoI}})^2 = (\nabla_{\mathbf{u}}\text{QoI})^T \, \mathbf{B} \, (\nabla_{\mathbf{u}}\text{QoI}) \, .$$

References

Bui-Thanh, T., Burstedde, C., Ghattas, O., Martin, J., Stadler, G., & Wilcox, L. C. (2012). Extreme-scale UQ for Bayesian Inverse Problems Governed by PDEs. In *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis* (pp. 1–11).

Law, K., Stuart, A., & Zygalakis, K. (2015). Data Assimilation - A Mathematical Introduction. *Springer*.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2007). *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press.

Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics.

Thacker, W. C. (1989). The role of the Hessian matrix in fitting models to measurements. *Journal of Geophysical Research: Oceans*, *94*(C5), 6177–6196. doi: 10.1029/JC094iC05p06177

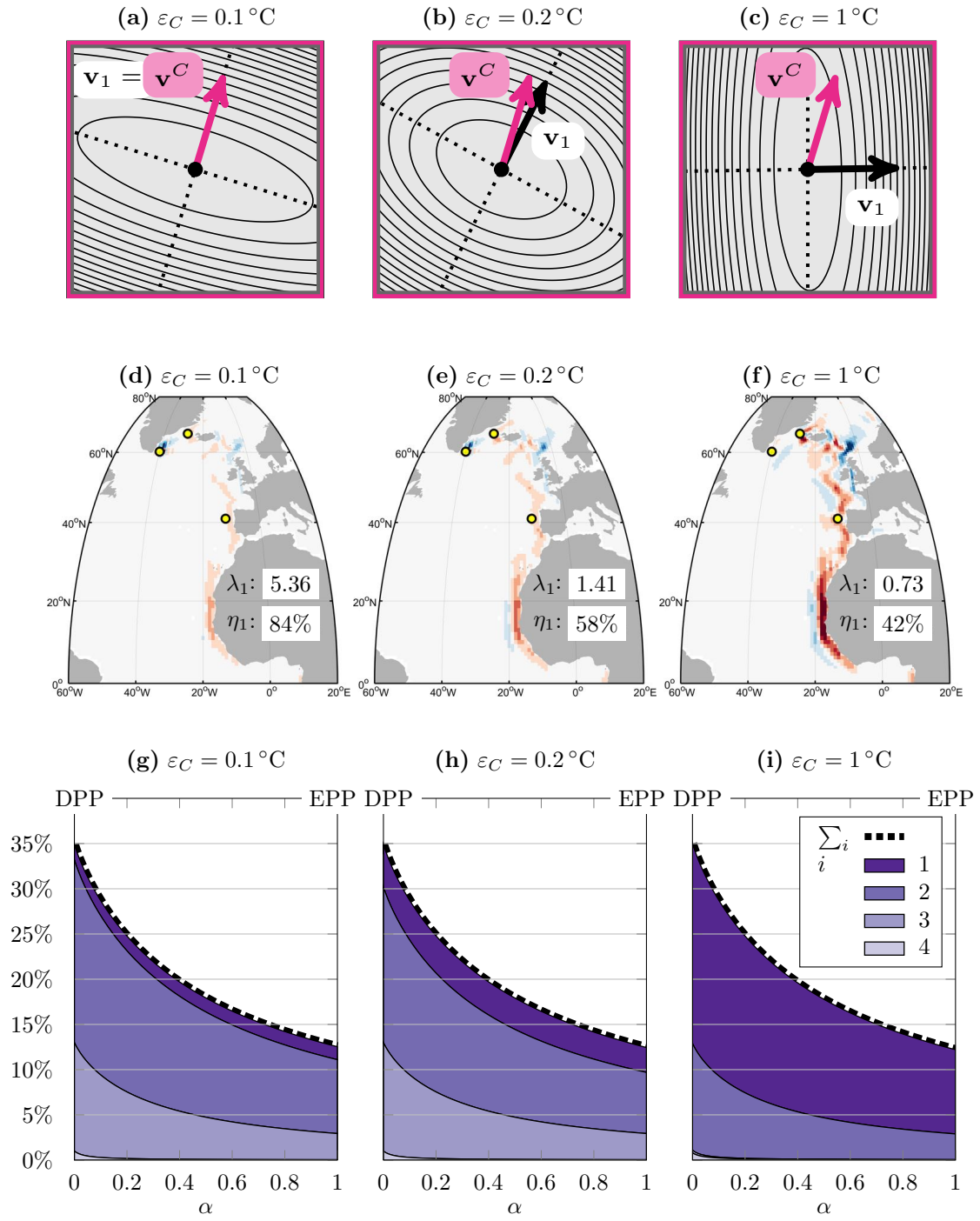

Figure S1. Figure caption on following page.

Figure S1. The panels in the middle column, i.e., (b),(e),(h), coincide with Figs. 8(a),(c), and Fig. 9(d). The left and the right column are as the middle column, but for different choices of ε_C. (a)-(c): Orientation of the first eigenvector, $\mathbf{v}_1$ (black vector) within the $\{\theta^A, \theta^C\}$-informed subspace of the control space. The ellipses show the contour lines of $J_{\text{misfit}}(\tilde{\mathbf{u}})$. The larger ε_C, the more $\mathbf{v}_1$ deviates from $\mathbf{v}^C$, from $\mathbf{v}_1 = \mathbf{v}^C$ in (a) to $\mathbf{v}_1$ being almost orthogonal to $\mathbf{v}^C$ in (c). (d)-(f): τ_y component of $\mathbf{v}_1$. The inlets show the corresponding eigenvalue λ_1, as well as the associated effectiveness $\eta_1 = \lambda_1/(\lambda_1 + 1)$. The larger ε_C, the more $\mathbf{v}_1$ reflects the characteristic sensitivity patterns of the subsurface observations, concentrated along the eastern and northern boundary of the North Atlantic (cf. Fig. 4), but the lower λ_1 and η_1. (g)-(i): Decay in proxy potential for the QoI, HT_{ISR}, as a function of α. For all three cases, the DPP ($\alpha = 0$) is equal to 35.0%. From left to right, the EPP ($\alpha = 1$) decreases slightly (although almost negligibly) from 12.6% in (g) to 12.2% in (i). From left to right, the main contribution to proxy potential shifts from the second to the first eigenvector.