

A Practical Formulation for an Anisotropic and Nonstationary Matérn Class Correlation Operator

Timothy A. Smith^{1,2,3}

¹Cooperative Institute for Research in Environmental Sciences (CIRES) at the University of Colorado Boulder, Boulder, CO, USA

²Physical Sciences Laboratory (PSL), National Oceanic and Atmospheric Administration (NOAA), Boulder, CO, USA

³Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, TX, USA

Correspondence

Timothy A. Smith, David Skaggs Research Center, Boulder, CO, 80305, USA
Email: tim.smith@noaa.gov

Funding information

This work was funded in part by NASA NNH19ZDA001N-SLCST and a JPL/Caltech Subcontract (ECCO Consortium)

A key component of data assimilation methods is the specification of univariate spatial correlations, which appear in the background-error covariance. For realistic problems in meteorology and oceanography, correlation length scales are nonstationary (variable in space) and anisotropic (variable in each direction). Variational approaches typically use an operator to enforce correlation length scales, and thus the operator must be designed to capture desired levels of nonstationarity and anisotropy. For systems with complex boundaries, such as the ocean, it is natural to use a filtering approach based on the application of an elliptic, Laplacian-like operator. Here we show how an elliptic operator can be formulated to capture a general Matérn-type correlation structure. We show how nonstationarity and anisotropy can be encoded into the operator via a simple change of variables based on user-defined normalization length scales. The change of variables defines a mapping between the computational domain and a space where the analytical Matérn correlation function applies. In addition to the mapping, two other hyperparameters *separately* control the correlation length scale (i.e. range) and shape. As a practical use-case, we apply the operator to a global ocean model. We show that when the normalizing length scales correspond to the local grid scale, the range parameter has an intuitive interpretation as the number of neighboring grid cells at which correlation drops to 0.14. Finally, the correlation model is shown to be computationally efficient in two regards. First, the necessary linear solve can be performed with a high tolerance

($\sim 10^{-3}$) while still achieving the correct statistics, requiring few iterations to converge. Secondly, the operator's exponent, which controls the correlation shape, is linearly related to the diagonal elements of its matrix representation. As a result, using an exponent greater than one can improve convergence properties. Thus, the framework provides flexibility in controlling correlation shape.

KEYWORDS

correlation operators, covariance modelling, background error, ocean data assimilation, variational assimilation, stochastic PDEs

1 | INTRODUCTION

High dimensional geophysical inverse problems, such as numerical weather prediction and oceanographic state estimation, are typically ill-posed due to the sparsity of data relative to the size of the control vector. A classical method for handling this ill-posedness is to prescribe some type of regularization in order to “spread” information to the uninformed regions and variables in the control vector (e.g., Wunsch, 2006). In the Bayesian interpretation of the inverse problem, this regularization is defined so that it represents the prior uncertainty or background-state error (e.g., Bui-Thanh et al., 2013). Ideally, this uncertainty captures the true error in the background state, but for all practical applications the background error must be approximated. Moreover, to make the problem computationally tractable, it is often assumed that the background error is governed by Gaussian statistics, such that the uncertainty is fully described by a covariance matrix.

For realistic data assimilation (DA) problems in meteorology and oceanography, a well formed background error will contain covariance relationships between different variables (e.g., between temperature and velocity components), it will have spatially dependent length scales of covariation (i.e. nonstationarity or inhomogeneity), and it will respect the system's anisotropy, such that length scales of covariance differ appropriately in each direction (e.g., Bannister, 2008a). Within variational DA systems, the background error covariance is usually represented as an operator so that it can be applied efficiently during an iterative optimization algorithm. Thus, for an operator-based covariance model to be useful in this context, it must be able to respect multivariate, nonstationary, and anisotropic features that are necessary for the given problem setting.

Typically, the background state error covariance is decomposed into two operators following Derber and Bouttier (1999). The first, “balance” operator captures multivariate (i.e. cross-variable) covariance information while the second captures “unbalanced” (i.e. univariate) covariance information. Our focus is on the latter operator, but we note the review by Bannister (2008b) which outlines balance operators used in atmospheric DA, and e.g., Weaver et al. (2005); Moore et al. (2011) for oceanographic examples.

Univariate covariance operators are further decomposed into two stages, where spatial correlations are specified first, and then scaled to the appropriate amplitude. In atmospheric DA, it is common to transform the control vector into a wavelet or spherical harmonic basis in order to specify the background error (e.g., Bannister, 2008b). However, for systems with complex boundaries like the ocean, it is far more straightforward to formulate the correlation operator in terms of the original, physical domain. Considering methods that operate in the physical or grid space, there are generally two classes of correlation models that are commonly used. The first class is encapsulated by explicit functional forms, including, for example, the functions developed by Gaspari and Cohn (1999); Gneiting (1999); Gaspari et al. (2006), which have the benefit of providing compact support. The second class of correlation models can generally be described as a filtering approach. Purser et al. (2003b,a) show

correlation functions based on recursive filters, and Dobricic and Pinaridi (2008) extend this to be used with complex boundaries. More recently, Purser et al. (2022) show a “beta filter” approach which enables compact support and a highly generalizable correlation shape via an efficient multigrid approach. Alternatively, within this class of models are those based on the solution to a differential equation.

Correlation models that are based on the solution to differential equations have several advantages. Most importantly, these operators handle complex boundaries naturally and the infrastructure required to obtain their solution typically exist within the underlying numerical model. Within oceanographic state estimation, a widely used framework is based on the solution to the diffusion equation (e.g., Nguyen et al., 2021; Forget et al., 2015; Blockley et al., 2014; Moore et al., 2011; Daget et al., 2009; Muccino et al., 2008; Di Lorenzo et al., 2007; Weaver et al., 2003). The diffusion-based framework relies on either an explicit, pseudo-time stepping method (Weaver and Courtier, 2001) or an implicit solution (Mirouze and Weaver, 2010; Carrier and Ngodock, 2010; Weaver and Mirouze, 2013), where the correlation structure underlying the solution corresponds to either a Gaussian or more general auto-regressive function, respectively. Alternatively, Lindgren et al. (2011) (hereafter L11) show an explicit link between the numerical solution to a stochastic partial differential equation (SPDE) and a Matérn-type covariance model. As of yet, however, it has remained unclear how this model could be used to specify univariate correlations with appropriate nonstationarity and anisotropy for an operational variational DA system.

Here, we extend the work by L11 to show how the framework can be used within variational DA. Our emphasis is on oceanographic applications, although the methodology is more general. We show how the mapping method suggested by L11 can be used to formulate a correlation operator that respects anisotropy and nonstationarity in a way that is relevant to many ocean general circulation models. In essence, the mapping method is simply a change of variables between the computational domain and a space where correlations can be described by the analytical Matérn correlation function for isotropic and stationary fields. Here we make practical suggestions on how to define the mapping via normalization length scales. With these normalizing length scales are defined, we show how two hyperparameters *separately* control the “range” and shape of the correlation model. As an intuitive example, when the normalizing length scales are defined relative to the model grid, a “range parameter” can be interpreted as the “number of neighboring grid cells” at which correlation drops to an expected value. We show that this interpretation holds no matter how many times the operator is applied, such that the overall shape can change while the correlation length scale remains fixed. It is therefore straightforward to use this approach to achieve desirable anisotropic and nonstationary statistics by simply tuning these two parameters.

The paper is laid out as follows. In Section 2 we give some context for how univariate correlation models are used in variational DA. We provide a review of diffusion based correlation operators, and then review the Matérn type covariance developments in L11 which we build on. In Section 3 we show how the Matérn model can be mapped from its isotropic, stationary form into a more complex computational domain. We then provide suggestions for parameterizing the model so that it can intuitively capture anisotropy and nonstationarity. In Section 4 we show numerical results of this correlation model applied to the global ocean, using the “Lat-Lon-Cap” grid introduced by Forget et al. (2015). Finally, in Section 5 we provide some discussion on the advantages of this model, and a general comparison to the widely used diffusion based models.

2 | MATHEMATICAL CONTEXT AND BACKGROUND

In order to provide some mathematical context for our correlation model developments, we first outline the generic inverse problem that is central to a variety of applications such as numerical weather prediction and state estimation. Our notation closely follows Ide et al. (1997), and we note that matrix notation is used to describe the problem, but these matrices are never formed explicitly. Rather, all matrices presented are described by operators that can be applied scalably in high dimensional inverse problems. We then review correlation models that are based on the application of a diffusion operator (Weaver and Courtier,

2001; Mirouze and Weaver, 2010), which are commonly used for large scale geophysical inverse problems (e.g., Forget et al., 2015; Moore et al., 2011). We then discuss the developments from L11, outlining the connection between the solution to a SPDE and a Gaussian random field with Matérn type covariance. We finish by showing a comparison of auto-regressive and the more general Matérn type correlation functions with a classical Gaussian. These comparisons motivate our development of an anisotropic, nonstationary correlation model, which is presented in Section 3.

2.1 | Data assimilation setting

We consider the general problem of finding the optimal control vector, \mathbf{x} , which minimizes the regularized model-data misfit cost function

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2} \|H(\mathbf{x}) - \mathbf{y}^o\|_{\mathbf{R}^{-1}}^2 + \frac{1}{2} \|\mathbf{x} - \mathbf{x}^b\|_{\mathbf{B}^{-1}}^2 .$$

Here $\|\mathbf{v}\|_A = \sqrt{\mathbf{v}^T \mathbf{A} \mathbf{v}}$ is a weighted Euclidean norm. The solution to this inverse problem, \mathbf{x}^a , arises from a tradeoff between fitting the observational data, \mathbf{y}^o , via the observation operator, $H(\cdot)$, and minimizing deviations from the background-state \mathbf{x}^b . This tradeoff is governed by the two error covariances, \mathbf{R} and \mathbf{B} , which dictate how much deviation is acceptable in either term. On the one hand, the observational error covariance matrix \mathbf{R} represents our uncertainty in the observational data, together with our confidence in the model's ability to represent the observed values. On the other hand, the background-state covariance matrix \mathbf{B} represents our uncertainty in the prior estimate or background state, \mathbf{x}^b .

To be explicit, we focus on how our correlation model fits into the background-state error covariance, \mathbf{B} . However, we note that the formulation could be used for specifying correlations between observations in a similar fashion to Guillet et al. (2019), but we leave this for future investigation. In the general case, the control vector \mathbf{x} could be multivariate, including initial conditions of the system state, uncertain boundary conditions, or uncertain parameter fields. Here we employ the decomposition proposed by Derber and Bouttier (1999) in order to separate the multivariate (i.e. cross-variable) covariance relationships from the univariate (i.e. assumed independent) covariance relationships. Specifically, the background-state covariance is decomposed as follows

$$\mathbf{B} := \mathbf{K}_b \mathbf{B}_u \mathbf{K}_b^T ,$$

where \mathbf{K}_b is a balance operator that deals with the cross-variable correlations. The matrix \mathbf{B}_u describes the covariance for the unbalanced variables and has a block-diagonal structure, such that each covariance is described independently. The unbalanced covariance is further factored as

$$\mathbf{B}_u := \Sigma C \Sigma$$

where Σ is a diagonal scaling matrix, containing the desired pointwise standard deviation values and C is a block diagonal correlation matrix, describing each variable's independent correlation structure. To be concrete, this can be viewed as

$$\mathbf{B}_u = \begin{pmatrix} \Sigma_\alpha \mathbf{C}_\alpha \Sigma_\alpha & 0 & \dots & 0 \\ 0 & \Sigma_\beta \mathbf{C}_\beta \Sigma_\beta & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & \Sigma_\gamma \mathbf{C}_\gamma \Sigma_\gamma \end{pmatrix}$$

where α, β, γ are placeholders for unique variables, and each $\Sigma_\alpha, \mathbf{C}_\alpha$ pair describes the covariance for its particular variable. In this formulation, \mathbf{C}_α describes the spatial correlation structure for the variable α , and is scaled to a proper covariance given the diagonal matrix Σ_α .

The goal of this paper is to formulate a generic operator for \mathbf{C} that could be used in a variational DA system to specify correlations for any variable, e.g., $\mathbf{C}_\alpha, \mathbf{C}_\beta$, etc. As with diffusion based correlation models, we specify \mathbf{C} through its square root $\mathbf{C}^{1/2}$, such that $\mathbf{C} = \mathbf{C}^{1/2} \mathbf{C}^{T/2}$. To provide some additional context for this development, we first review past work that has achieved this with a generalized diffusion-based operator.

2.2 | Diffusion based correlation models

A common method for specifying the correlation structure, \mathbf{C} , in variational DA systems, especially in oceanographic applications, is through the solution of a generalized diffusion equation,

$$\frac{\partial \theta}{\partial t} = \nabla \cdot \kappa \nabla \theta . \quad (1)$$

Here κ is a diffusion tensor that controls anisotropy and nonstationarity, t is a ‘‘pseudo-time’’ coordinate, and the solution to this equation has a Gaussian, or ‘‘Gaussian-like’’, covariance.

Building on work from Derber and Rosati (1989); Egbert et al. (1994); Bennett et al. (1996), Weaver and Courtier (2001) showed that an explicit, forward Euler solution to Equation (1) provides a scalable approach to defining correlations in complex domains. That is, the solution $\theta(T)$ is given by forward pseudo-time stepping

$$\theta(T) = (I + \nabla \cdot \kappa \nabla)^p \theta(t_0) , \quad (2)$$

which requires p applications of the operator

$$\mathcal{A}_{\text{ED}} := (I + \nabla \cdot \kappa \nabla) ,$$

where p is chosen in order to achieve numerical stability (see Weaver and Courtier, 2001, for details regarding the discretized form of this operator, and extensions of the model briefly shown here). The solution $\theta(T)$ is shown to have an approximately Gaussian covariance structure. A correlation model is thus defined by estimating the pointwise variance of $\mathcal{A}_{\text{ED}}^p \cdot \hat{\sigma}^2(\mathbf{s})$, in order to define the normalization matrix

$$\Lambda := \text{diag} \{ 1 / \hat{\sigma}_n \}_{n=1}^{N_\theta} ,$$

where we use n to refer to each grid cell, such that the correlation matrix is defined through its square root as

$$\mathbf{C}_{\text{ED}}^{1/2} := \Lambda \mathcal{A}_{\text{ED}}^p . \quad (3)$$

The explicit diffusion-based correlation model as briefly summarized here directly approximates a Gaussian structure. However, a limitation of this approach is that in some cases, many iterations are required to keep the scheme stable (i.e. a large value of p is required in Equation (2)). As such, Mirouze and Weaver (2010) and Carrier and Ngodock (2010) developed

correlation models based on the *implicit* solution of Equation (1):

$$(I - \nabla \cdot \kappa \nabla)^M \theta(T) = \theta(t_0), \quad (4)$$

which is unconditionally stable. In this case, a square-root correlation operator is defined through the “implicit diffusion operator”

$$\mathcal{A}_{\text{ID}} := (I - \nabla \cdot \kappa \nabla) \quad (5)$$

as

$$\mathbf{C}_{\text{ID}}^{1/2} := \Lambda \mathcal{A}_{\text{ID}}^{-M}, \quad (6)$$

with Λ defined based on the operations that precede it.

Mirouze and Weaver (2010) provide the theoretical underpinnings for the implicit diffusion approach, and show details regarding its implementation in general curvilinear coordinates. Additionally, the authors show that this correlation model corresponds with an M th order auto-regressive (AR) function, which is a subclass of Matérn type correlation models (see Weaver and Mirouze, 2013, for more description of the parameters controlling this model). Thus, the correlation model has a more general shape, which in the limit $M \rightarrow \infty$, approaches a classical Gaussian structure (see Section 2.4).

In this work, we formulate a correlation operator in a similar fashion by directly parameterizing the elliptic PDE that corresponds to the more generic Matérn type correlation (or covariance) structure. In the following subsections, we provide some background on this general correlation structure to give our developments some context and motivation.

2.3 | Review of the Matérn correlation structure

The Matérn covariance function between two points, $\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2 \in \hat{\Omega} = \mathbb{R}^{N_d}$ can be expressed as:

$$c(\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2) = \frac{\hat{\sigma}^2}{2^{\epsilon-1} \mathcal{G}(\epsilon)} \left(\sqrt{\hat{\delta}} \|\hat{\mathbf{s}}_2 - \hat{\mathbf{s}}_1\| \right)^{\epsilon} \mathcal{B}_{\epsilon} \left(\sqrt{\hat{\delta}} \|\hat{\mathbf{s}}_2 - \hat{\mathbf{s}}_1\| \right). \quad (7)$$

Here \mathcal{G} is the Gamma function, \mathcal{B}_{ϵ} is the modified Bessel function of the second kind and order ϵ , $\hat{\delta} > 0$ is a scaling parameter, and $\epsilon > 0$ controls the mean-square differentiability of the underlying statistical process described by the Matérn covariance. We note that in the limit of $\epsilon \rightarrow \infty$, the Matérn-type covariance becomes a classical Gaussian. Throughout, we refer to a “Matérn field” as any random field that has covariance described by the Matérn covariance function, Equation (7). Additionally, we note that this analytical function only pertains to fields where the assumptions of stationarity and isotropy hold, and we use notation with “hats” ($\hat{\cdot}$) when this is the case. The marginal variance of such a Matérn field has the analytical form

$$\hat{\sigma} = \frac{\mathcal{G}(\epsilon)}{\mathcal{G}(\epsilon + N_d/2) \hat{\delta}^{\epsilon} (4\pi)^{N_d/2}}, \quad (8)$$

where N_d is the dimensionality of the field being described. Thus, the Matérn covariance can be described via a correlation function, $r(\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2)$, that is scaled by the marginal variance:

$$c(\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2) = \hat{\sigma} r(\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2).$$

The key relationship discussed in L11 is that any discrete solution to the elliptic SPDE,

$$\left(\hat{\delta} - \hat{\nabla} \cdot \hat{\nabla}\right)^M \hat{\theta}(\hat{\mathbf{s}}) = \mathcal{W}(\hat{\mathbf{s}}), \quad (9)$$

is a Matérn field, for any triangulation or rectangular lattice of \mathbb{R}^{N_d} . Here \mathcal{W} is a white noise process defined on the space $\hat{\Omega}$ and $M \geq 1$ is the total number of times that the inverse of the operator

$$\hat{\mathcal{A}} := (\hat{\delta} - \hat{\nabla} \cdot \hat{\nabla})$$

is applied to obtain the solution. This number is related to the mean differentiability and dimensionality of the field via the relation

$$M = \epsilon/2 + N_d/4. \quad (10)$$

For our purposes, we consider M as a parameter which can be tuned to obtain the desired correlation shape (see Section 2.4). In order to simplify the solution process, we consider positive integer values of M , although the theory underlying more general operators for which $M = q/2$ with $q \in \mathbb{Z}^+$ is covered in L11.

Throughout this paper we use the empirical relation discussed by L11 to specify $\hat{\delta}$ via

$$\hat{\rho} = \sqrt{\frac{8\epsilon}{\hat{\delta}}}. \quad (11)$$

The so-called ‘‘range parameter’’, $\hat{\rho}$, defines the distance between two points at which correlation drops to 0.14. We note that the range parameter used here is larger by a factor of 2 than what is sometimes used (e.g., Rasmussen and Williams, 2006), but we prefer this definition because of its intuitive interpretation. With M and $\hat{\rho}$ defined, we provide an explicit form of the Matérn correlation function which will be used throughout:

$$r(d; \hat{\rho}, M) = r(d; \hat{\rho}, \epsilon(M)) = \frac{1}{2^{\epsilon-1} \mathcal{G}(\epsilon)} \left(\sqrt{8\epsilon} \frac{d}{\hat{\rho}}\right)^{\epsilon} \mathcal{B}_{\epsilon} \left(\sqrt{8\epsilon} \frac{d}{\hat{\rho}}\right), \quad (12)$$

where d is the distance between two points and $\epsilon = \epsilon(M)$ is defined via Equation (10).

In Section 3, we suggest to define a correlation operator through the elliptic operator in Equation (9). For an isotropic and stationary case, the operator would be defined as

$$\mathbf{C}_{\text{iso,stat}}^{1/2} = \mathbf{\Lambda} \left(\hat{\delta} - \hat{\nabla} \cdot \hat{\nabla}\right)^{-M}, \quad (13)$$

with $\mathbf{\Lambda}$ appropriately defined, but we extend this to a more general case in Section 3.

2.4 | Comparing the Matérn and auto-regressive functions

To provide some motivation behind our developments of the Matérn correlation function, we compare its structure to auto-regressive (AR) (corresponding to the operator $\mathcal{A}_{\text{ID}}^{-M}$; (Mirouze and Weaver, 2010)) and Gaussian functions. Figure 1(a) shows a comparison of these 1D correlation structures as a function of distance, d , where each use a representative length scale of 4. The Matérn function and Gaussian consistently reach an approximate correlation of 0.14 when distance is equal to $\hat{\rho}$. The consistency here makes using the Matérn correlation structure intuitive in practice: no matter what value of M is used (or equivalently no

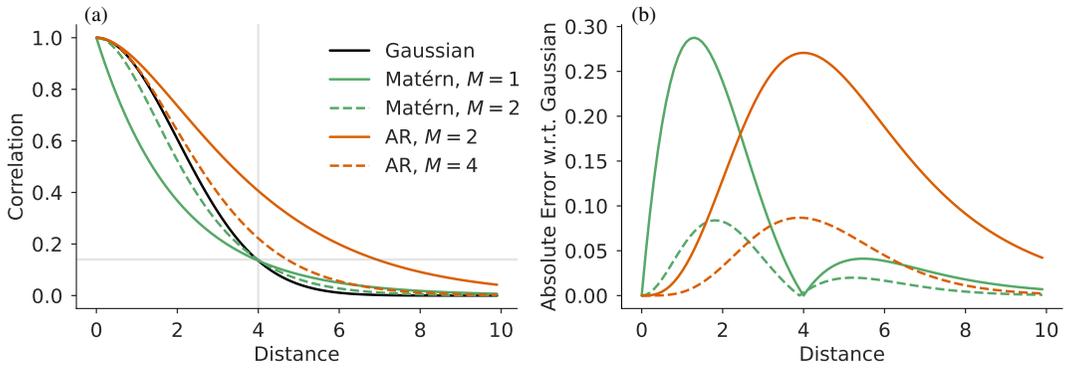


FIGURE 1 (a) Correlation as a function of distance given by Matérn (green; Equation (13)), auto-regressive (AR; orange), and Gaussian (black; $r_g(d) = \exp(-2(d/4)^2)$) functions. For each, we choose a length scale of $\hat{\rho} = 4$, noting for comparison purposes that this corresponds to $L = 2$ using the notation in Mirouze and Weaver (2010) for the auto-regressive function. Finally, we note that we use $N_d = 3$ to compute $\epsilon(M)$ (see Equation (10)) since this corresponds to our numerical experiments in Section 4. (b) Absolute error considering the difference between the Matérn and AR functions with respect to a Gaussian with the same length scale.

187 matter what shape is desired), we can expect this approximate relationship to hold.

188 Additionally, Figure 1(b) shows the error in each Matérn and AR functions relative to a Gaussian. We note that error for the
 189 Matérn function is comparable to the AR function, although M is halved. This has practical implications, since in both cases M
 190 corresponds to the number of inverse elliptic operator applications required for achieving a correlation structure via the implicit
 191 diffusion approach via $\mathcal{A}_{\text{ID}}^{-M}$ (AR) or the approach proposed here (Matérn). Thus, if the goal is to approximate a Gaussian, then
 192 using the direct Matérn shape results in half the required inverse elliptic solves. However, we also stress that the desired shape of
 193 the correlation function will be application dependent, and so we present a comparison from which practitioners can take their
 194 pick.

195 3 | A NONSTATIONARY AND ANISOTROPIC MATÉRN CORRELATION OP- 196 ERATOR

197 Here we propose to use the SPDE operator described by L11 as a means to describe an anisotropic, nonstationary correlation
 198 model in a similar manner to the diffusion-based methods described in Section 2.2. To do so, we employ the “mapping method”
 199 described by L11 which we show for the general M th order SPDE in Section 3.1. In summary, the basic idea we present is to use
 200 the mapping method to nondimensionalize or re-scale the elliptic operator. We provide a simple scaling argument for why this is
 201 a good idea in Section 3.2, and discuss practical choices for the nondimensionalization based on the grid scale in Section 3.3.

202 3.1 | Mapping method or change of variables

203 In L11 it is suggested that solving the SPDE in a transformed coordinate system allows one to readily incorporate anisotropy and
 204 nonstationarity into a Matérn covariance model. In this mapping method, we consider solutions to the isotropic, stationary SPDE
 205 (9) to be defined in a transformed, or “deformed” (Sampson and Guttorp, 1992), space $\hat{\Omega}$. Then, assume that we have a mapping

φ that maps between this transformed space and our computational domain, Ω :

$$\varphi : \hat{\Omega} \ni \hat{\mathbf{s}} \rightarrow \mathbf{s} \in \Omega .$$

With this mapping, we can employ a change of variables (Smith, 1934) to rewrite the SPDE in the computational domain as:

$$\left(\frac{\hat{\delta}}{\det(\Phi(\mathbf{s}))} - \nabla \cdot \frac{\Phi(\mathbf{s})\Phi(\mathbf{s})^T}{\det(\Phi(\mathbf{s}))} \nabla \right)^M \theta(\mathbf{s}) = \det(\Phi(\mathbf{s}))^{-1/2} \mathcal{W}(\mathbf{s}) .$$

Here we have defined the Jacobian as

$$\Phi(\mathbf{s}_0) := \frac{\partial \varphi}{\partial \hat{\mathbf{s}}} \Big|_{\varphi^{-1}(\mathbf{s}_0)} ,$$

and for now we assume that $\varphi^{-1}(\mathbf{s}_0)$ is well defined. For our purposes, this turns out to be the case, but this becomes clear when Φ is defined in Section 3.2. With the following definitions:

$$K(\mathbf{s}) := \frac{\Phi(\mathbf{s})\Phi(\mathbf{s})^T}{\det(\Phi(\mathbf{s}))} \quad \delta(\mathbf{s}) := \frac{\hat{\delta}}{\det(\Phi(\mathbf{s}))} \quad \mathcal{A} := (\delta(\mathbf{s}) - \nabla \cdot K(\mathbf{s})\nabla) \quad (14)$$

the SPDE in Equation (9) can be written in the computational domain's coordinate system as

$$\mathcal{A}^M \theta(\mathbf{s}) = \det(\Phi(\mathbf{s}))^{-1/2} \mathcal{W}(\mathbf{s}) , \quad (15)$$

where zero flux, Neumann boundary conditions are applied at the boundaries (see Appendix A for details).

Here, we propose to use this generic form to define a square root of the correlation matrix in a similar fashion to Weaver and Courtier (2001); Mirouze and Weaver (2010); Carrier and Ngodock (2010) as follows,

$$\mathbf{C}^{1/2} := \Lambda \mathcal{A}^{-M} \det(\Phi(\mathbf{s}))^{-1/2} , \quad (16)$$

where Λ is once again a variance-preserving normalization matrix defined by the operations that precede it. In this model, anisotropy and nonstationarity are controlled by $\Phi(\mathbf{s})$, and in the following sections we discuss how this can be assigned for practical applications in geophysical inverse problems. We note that in this discussion we loosely mix the use of finite dimensional matrices and infinite dimensional operators in order to ease the presentation, but we provide a more careful derivation of their discretized forms relevant to our numerical experiments in Appendix A.

3.2 | Scaling the Laplacian term for anisotropy

Here we focus on parameterizing $\Phi(\mathbf{s})$ in order to achieve an anisotropic correlation model that is relevant for variational DA. We illustrate our choice with a scaling argument focusing on how $K(\mathbf{s})$ influences correlation length scales.

Consider a 3D field $\theta(\mathbf{s}) \sim \Theta$ that exhibits spatial variability at the length scales, L_x , L_y , and L_z in the direction of longitude, latitude, and height, respectively, where $L_x, L_y \gg L_z$, such that the field exhibits highly anisotropic fluctuations. This is a common situation in large scale geophysical fluid dynamics, where fields (e.g., temperature, velocity) exhibit length scales of variability that are much greater in either horizontal dimension compared to the vertical (e.g., Vallis, 2006). Without any

233 rescaling, i.e. without K , the Laplacian term in \mathcal{A} is unbalanced

$$234 \quad \nabla^2 \theta(\mathbf{s}) \sim \frac{\Theta}{L_x^2} + \frac{\Theta}{L_y^2} + \frac{\Theta}{L_z^2} \quad (17)$$

$$235 \quad \simeq \frac{\Theta}{L_z^2}.$$

236 As a result, the correlation model will have unrealistically large (small) correlations in the vertical (horizontal). Our goal is
237 therefore to define the elements of K such that each term is of the same order of magnitude.

To achieve this balance between Laplacian terms, we suggest a straightforward, perhaps obvious, specification of Φ :

$$238 \quad \Phi = \begin{pmatrix} L_x & 0 & 0 \\ 0 & L_y & 0 \\ 0 & 0 & L_z \end{pmatrix},$$

239 where we simply ignore the off-diagonal elements of Φ . The determinant in this case is $\det(\Phi) = L_x L_y L_z$ and according to the
240 definitions in Equation (14):

$$241 \quad K = \begin{pmatrix} 1/L_z & 0 & 0 \\ 0 & 1/L_z & 0 \\ 0 & 0 & L_z/(L_x L_y) \end{pmatrix},$$

242 so that

$$243 \quad \nabla \cdot K \nabla \theta(\mathbf{s}) \sim \frac{3}{L_x L_y L_z} \Theta.$$

244 The key is that K scales each term in the Laplacian so that they are approximately the same order of magnitude, and the operator
245 is balanced in all directions. In the case of nonstationarity, we simply require this balance to apply locally and allow the length
246 scales L_x , L_y , and L_z to vary in space.

247 3.3 | Harnessing the grid scale for nonstationarity

248 At this point, we must prescribe values for the normalizing length scales $L_x(\mathbf{s})$, $L_y(\mathbf{s})$, and $L_z(\mathbf{s})$ in order to fill $\Phi(\mathbf{s})$.
249 Considering the scaling analysis of the Laplacian in Section 3.2, a simple yet reliable choice for these is to use the underlying
250 grid-scale of the numerical model (e.g., the ocean or atmosphere general circulation model).

251 We consider using the length scale of the grid elements to be reasonable because a baseline level anisotropy and nonstation-
252 arity is usually encoded into the grid. A prime example of this nonstationarity is represented by the vertical axis of ocean model
253 grids, which are designed to capture a variety of behavior in a computationally efficient manner (Griffies, 2004). Toward the
254 surface, the ocean is tightly coupled to the atmosphere, sea ice, and rivers, and ocean models use a finely resolved vertical grid
255 to capture the ocean component of these coupled processes. In the interior ocean, well below the mixed layer, the ocean acts
256 more like stacked layers, and variation in properties like temperature and salinity occurs over much larger distances than at the
257 surface (Talley, 2011). Vertical grids are correspondingly much coarser at depth than near the surface. As a concrete example,
258 the height-based vertical grid we use in Section 4 varies from $\sim 5 - 10$ m near the surface, and spacing increases to $O(100)$ m
259 below 1,000 m (Figure 3(f)). By using the grid elements directly, our correlation model can capture the nonstationarity motivated

by the physical processes which influence the model grid’s development.

Another justification for using the grid elements to specify $\Phi(\mathbf{s})$ is that this provides a practical and intuitive nondimensionalization. With this definition, the correlation model exhibits isotropic and stationary behavior within the nondimensional space defined by the computational grid. As such, the idealized correlation function (Equation (13)) applies relative to the grid spacing, and one can view $\hat{\rho}$ as an intuitive, nondimensional parameter controlling the “number of neighboring grid cells” at which correlation decays to 0.14. Our numerical experiments in Section 4 show that this is a good approximation in the case of a realistic global ocean model grid.

4 | APPLICATION TO THE GLOBAL OCEAN

Here we show results from a numerical implementation of the correlation model described in Section 3. For this application we use the “Lat-Lon-Cap” (LLC) grid used by the ECCOv4 state estimate (see their Section 2 of Forget et al., 2015, for a complete description of the grid). The overall goal with these numerical experiments is to show that even in this relatively complicated global grid, the correlation model generally follows the expected Matérn correlation structure (Section 4.1), while maintaining anisotropy and nonstationarity that is relevant to the physical system (Section 4.2). Additionally, we show how Neumann boundary conditions to the differential operator affect the solution in Section 4.3. We finish by showing that the model can be applied efficiently with a relatively imprecise solver tolerance (Section 4.4) and that it is relatively cheap even when the number of applications, M , is greater than one (Section 4.5).

For all experiments, we compute statistical quantities from 1,000 samples (see Appendix A and specifically Equations (21) and (22) for sampling details). We use a block-Successive Over Relaxation (SOR) method (Appendix B) to find numerical solutions to the elliptic SPDE, using a tolerance of 10^{-3} for all results except where specified in Section 4.4. Additionally, we use the following normalizing length scales $L_x(i, j) = \Delta x_g(i, j)$, $L_y(i, j) = \Delta y_g(i, j)$, and $L_z(k) = \Delta r_f(k)$ (Figure 7) where we have switched from the spatial coordinate $\mathbf{s} \in \Omega$ to the computational grid indices i, j, k . All experiments are performed with a 3D field, which could represent an ocean state property like temperature or salinity.

4.1 | Correspondence with theoretical correlation structure

We first show that the sample correlation structure computed on the LLC grid corresponds with the analytical Matérn-type correlation function (Equation (13)). For this comparison we compute the correlation field in the transformed space $\hat{\Omega}$, where it can be considered isotropic and stationary. Because we use the grid spacing to define this mapping, the correlation distances are computed simply by counting the number of neighboring grid cells in each direction from the point in consideration.

Figure 2 shows the comparison between the theoretically expected correlation structure (black) and the numerically computed sample correlation structure for $\hat{\rho} = \{5, 10, 15, 20\}$ (color) using $M = \{1, 2, 4, 8\}$ (panels). Recall that as M increases, the correlation structure approaches a Gaussian shape, such that there is little visible difference between e.g., $M = 4$ and $M = 8$. The correlation is computed in the direction of longitude, indicated by δi . The shading indicates the spread between the first and ninth deciles of the sample correlation, computed at all depth levels and latitudes from 70°S to 37°N at 127.5°W - a subset chosen simply to ease the calculation. Similar plots showing correlation in the meridional and vertical directions are shown in the Supplemental Material.

Generally speaking, the colored curves match the analytical expression well, and each colored curve intersects the horizontal gray line, indicating a correlation value of 0.14, where $\hat{\rho} = \delta i$. We note that the largest spread in the computed correlation structure occurs when $M = 1$, especially for larger values of $\hat{\rho}$. Considering an analogy to Laplacian versus biharmonic damping in ocean models (e.g., Holland, 1978; Griffies and Hallberg, 2000), we suggest that there is more spread when $M = 1$ because

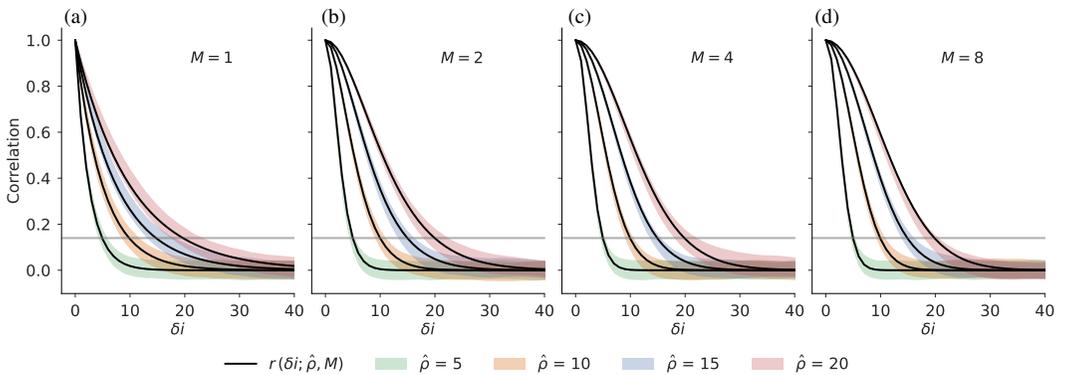


FIGURE 2 Correlation structure computed from the theoretical Matérn correlation function (black; Equation (13)) and from 1,000 samples using a subset of the “Lat-Lon-Cap” grid within the Pacific Ocean (shaded coloring). The sample correlation is computed in the zonal direction, δi , indicating the number of neighboring grid cells from 127.5°W . The shading indicates the spread between the first and ninth deciles, based on sample correlations at all depth levels and latitudes from 70°S and 37°N . Similar plots showing correlation as a function of meridional and vertical distance are provided in the Supplemental Material. Recall that as $M \rightarrow \infty$, the Matérn correlation function approaches a Gaussian (cf. Figure 1).

298 the operator \mathcal{A}^{-1} contains only a Laplacian term. Compared to the cases when $M > 1$, which contain biharmonic and higher
 299 order Laplacian terms in \mathcal{A}^{-M} , the Laplacian case is less scale-selective. That is, the operator does not cutoff higher frequency
 300 variability sharply, allowing noise to pollute the sample statistics (see Griffies and Hallberg, 2000, Section 2 for a quantitative
 301 description of this cutoff in frequency space). We note, however, that the spread still allows for a reasonable interpretation that
 302 the operator \mathcal{A}^{-1} captures the behavior of the analytical Matérn correlation function.

303 4.2 | Sample correlation maps

304 Maps of the sample correlation field on the LLC grid are shown in Figure 3, at $(0.2^\circ\text{N}, 127.5^\circ\text{W}, 722\text{ m depth})$ and $(10.5^\circ\text{N},$
 305 $87.5^\circ\text{W}, 5\text{ m depth})$ in panels (a & d) and (b & e), respectively. These two locations are chosen to highlight local anisotropy in
 306 the latitude-longitude plane and nonstationarity in the vertical axis. For these calculations we use a perhaps unrealistically large
 307 correlation length scale defined by $\hat{\rho} = 20$ for illustrative purposes.

308 Comparing panels (a) and (b) of Figure 3, we see that near the equator, the correlation structure is stretched zonally, while
 309 poleward of $\sim 10 - 15^\circ$ the structure is closer to being isotropic. Elongated zonal correlation length scales at the equator are
 310 consistent with observations (Meyers et al., 1991), so we consider this anisotropic behavior to be desirable. In our case, we
 311 achieve this anisotropy via local refinements in the meridional grid scale near the equator, which are specifically designed to
 312 capture tropical zonal currents (Forget et al., 2015). Specifically, Figure 3(c) shows how the meridional grid spacing (Δy) refines
 313 near the equator, while the longitudinal grid spacing (Δx) slightly increases near the equator. The result of this grid refinement
 314 is that near the equator, a range of $\hat{\rho}$ grid cells covers a shorter distance meridionally than it does at e.g., 15°N . Overall, this
 315 example highlights how details in the model grid can be harnessed to achieve physically relevant correlation structures. However,
 316 we note that if this correlation shape is desired but would not immediately occur due to a different model grid definition, then the
 317 normalizing length scales could easily be modified with a weighting function to achieve the desired structure.

318 Panels (d) and (e) of Figure 3 illustrate nonstationary correlation structures obtained along the vertical axis. As discussed in
 319 Section 3.3, it would be reasonable to expect shorter correlation lengths near the ocean surface due to localized processes there,
 320 and relatively longer correlation length scales in the interior ocean. This general behavior is shown in panel (e), where correlation

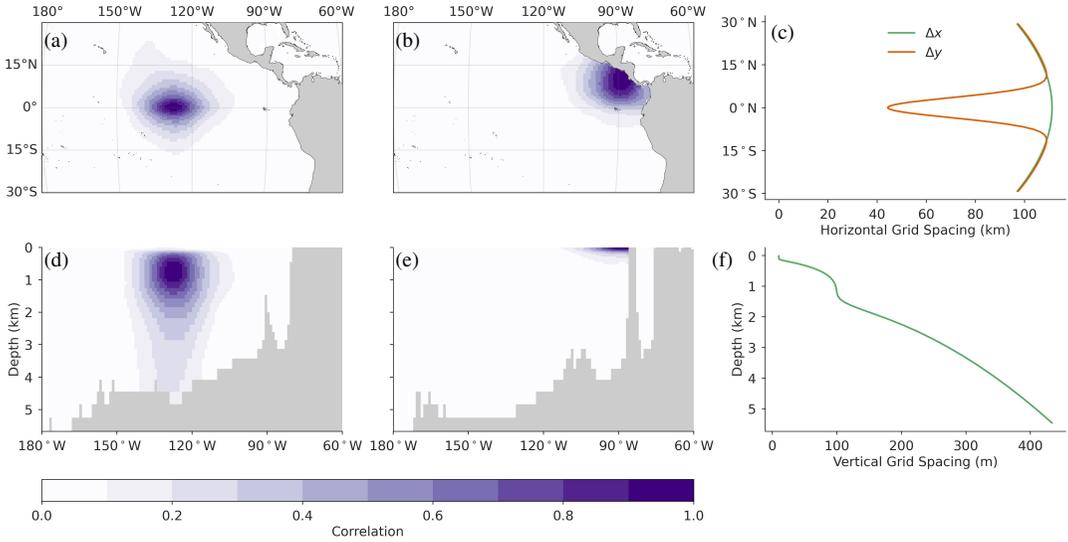


FIGURE 3 Two sample correlation fields and a depiction of the computational grid. (a & b) Sample correlation field in the latitude-longitude plane at (0.2°N, 127.5°W, 722 m depth) and (10.5°N, 87.5°W, 5 m depth), respectively. (c & d) The same sample correlation fields as above, shown in the depth-longitude plane. (e & f) The local horizontal and vertical grid spacing, respectively. The correlation fields are computed with $\hat{\rho} = 20$ and $M = 2$.

321 > 0.1 is confined to the upper hundred meters of the ocean, while in panel (d), correlation > 0.1 spans the full depth of the ocean.
 322 Once again, the correlation structures shown here correspond directly to the vertical grid spacing, which is shown in Figure 3(f).

323 4.3 | Pointwise sample standard deviation

324 The pointwise, sample standard deviation is shown in Figure 4, where it is represented as a ratio with respect to the “expected”
 325 value from Equation (8) for an isotropic, stationary Matérn field. Throughout most of the domain, the sample standard deviation is
 326 approximately equal to the theoretical value. Near continental boundaries, however, the standard deviation is inflated, especially
 327 for large values of $\hat{\rho}$ and in regions of tightly confined topographic boundaries such as in the Caribbean Sea. This deviation near
 328 the boundaries is expected for correlation models based on the solution of differential equations (e.g., Weaver and Courtier, 2001;
 329 Lindgren et al., 2011), as a result of the zero flux, Neumann boundary conditions used to find the solution. Thus, the theoretical
 330 value cannot be used directly and it is necessary to calculate or estimate the true variance of the operator in order to formulate
 331 the normalization matrix Λ . Throughout this work, we have used the estimated standard deviation shown in Figure 4 to fill Λ ,
 332 based on 1,000 random samples. Given that the numerically computed correlation structure compares well to the theoretical
 333 value (Section 4.1), this normalization method appears to be reliable. We also note this method is convenient because it is
 334 embarrassingly parallel, and scales well to arbitrarily high dimensional fields. However, for cases when the correlation operator
 335 is to be updated repeatedly, e.g., in cycled DA, other normalization methods could be explored for use with this operator as in
 336 Weaver et al. (2021).

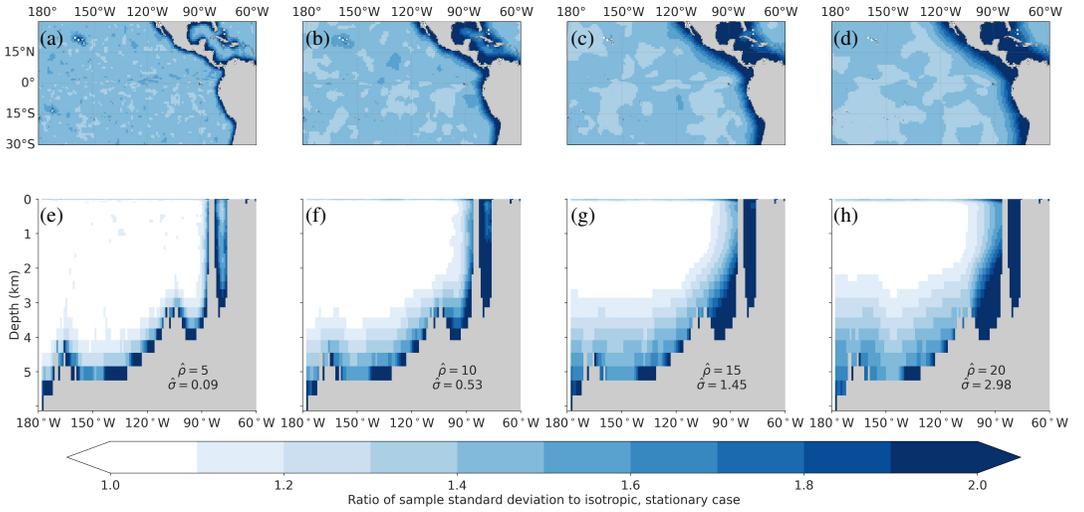


FIGURE 4 The ratio of the sample standard deviation to the value computed from Equation (8) for an isotropic, stationary Matérn field (indicated in the bottom row by $\hat{\sigma}$). Panels (a-d) show the ratio in the latitude-longitude plane at the surface for $\hat{\rho} = \{5, 10, 15, 20\}$, respectively. Panels (e-h) show the corresponding fields in the depth-longitude plane along 10.5°N. The largest deviations from the theoretical value are near the boundaries, as expected. All fields use $M = 2$.

4.4 | High efficiency with low precision

The numerical results shown in this section have employed the iterative algorithm in Appendix B to obtain approximate correlation statistics. As with any iterative algorithm, one must specify a tolerance that can be used to determine when the algorithm has converged to an approximate solution. Within this framework, one can always set a tolerance based on the numerical precision being used to be confident that the solver has converged. However, in this section we show that this is likely to be unnecessarily ambitious.

To be specific, Figure 5(a) shows the relative error in the approximation that correlation is equal to 0.14 when $\hat{\rho} = \delta i$ for $\hat{\rho} = \{5, 10, 15, 20\}$ (i.e. corresponding to the curves in Figure 2). The error in the approximation is shown for a range of solver tolerances, where 10^{-15} is chosen as an approximate lower bound tolerance for double precision. For tolerances at 10^{-3} and smaller, the error converges to roughly the same value, indicating that the desired statistics of the correlation model are obtained even with a relatively imprecise solve. We note that Carrier and Ngodock (2010) describe similar findings with the implicit diffusion correlation model.

The motivation for using a high tolerance is indicated by Figure 5(b), which shows how the number of iterations required to converge increases with the specified tolerance. Solving to a tolerance of 10^{-15} requires a factor of 6-13 more iterations than are required with a tolerance of 10^{-3} . Of course, the specific computational savings obtained will depend on the iterative method that is being used, but we provide this as a concrete example to highlight that an imprecise solve is both valid and advantageous.

4.5 | Rapid convergence for $M > 1$

For applications where a Gaussian correlation structure is desired, the correlation model presented here requires $M > 1$ to approach the Gaussian structure (Figure 1). Moreover, it could also be desirable to use \mathcal{A}^{-M} with $M > 1$, given that $M = 1$ produces larger spread in the correlation structure (Figure 2(a)), and because the correlation structure drops off much more

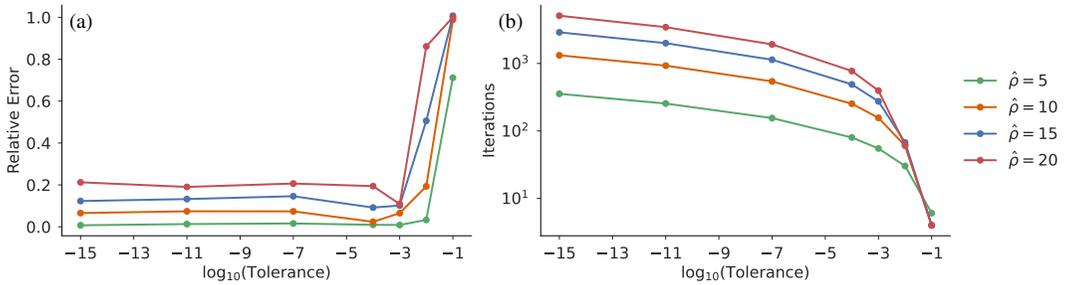


FIGURE 5 (a) The relative error in the approximation that correlation equals 0.14 when $\hat{\rho} = \delta i$, as a function of the tolerance used for the iterative block-SOR method described in Appendix B. Each curve is computed as the error between the theoretical (black) curve and the average of each shaded curve shown in Figure 2. (b) The number of iterations required for the block-SOR method to converge to the specified tolerance. Averages are computed from 1,000 samples using $M = 1$.

357 rapidly for neighboring points. For these cases, it may be natural to assume that using this model with $M > 1$ would be less
 358 efficient than when $M = 1$ because it requires multiple applications of an inverse elliptic operator. However, here we show that
 359 this is not necessarily true and it can be *more* efficient to use $M > 1$ than $M = 1$.

360 Figure 6(a) shows the total number of iterations required to find a solution to Equation (16), for a variety of combinations of
 361 $\hat{\rho}$ and M . Here, “total iterations” refers to all block-SOR iterations required by the algorithm in Appendix B, summing over
 362 all applications of \mathcal{A}^{-M} . Evidently, using $M > 1$ actually requires *fewer* total iterations to converge to a solution than when
 363 $M = 1$. The reason for this is as follows. For any value of $\hat{\rho}$, δ increases linearly with M , which increases the amplitude of
 364 the diagonal elements of the matrix representation of \mathcal{A} . In each case, the off-diagonal matrix elements, determined by the
 365 Laplacian operator, remain fixed. Thus, the matrix becomes more diagonally dominant: the amplitude of the diagonal elements
 366 increases relative to the sum total of off-diagonals. The degree of diagonal dominance is an important property for determining
 367 the convergence of our SOR-based elliptic solver, where a more diagonally dominant matrix tends to converge faster (Golub and
 368 Van Loan, 2013). Evidence of this behavior can be seen in Figure 6(b), which shows the number of iterations required for each
 369 individual application of \mathcal{A}^{-1} to converge. Here we see that each application gets cheaper as M , and therefore δ , increases. The
 370 improvement per iteration is evidently enough to reduce the total iterations, shown in panel (a). The exception to this behavior is
 371 when $\hat{\rho} = 5$, where for $M \geq 4$ the total number of iterations overtakes the case of $M = 1$ due to the repeated solves.

372 We note that Jacobi, Gauss-Seidel, and SOR methods are rarely used for modern applications (an exception being the
 373 line-SOR method in the sea-ice solver of the MITgcm (Losch et al., 2010)). For example, the main source code of the MITgcm
 374 (Marshall et al., 1997; Campin et al., 2021) uses a conjugate gradient method for the pressure solve at each time step. However,
 375 given the performance benefits noted here and the simplicity of implementing the SOR scheme, it could be used as an efficient
 376 preconditioner in the event that a different solution method becomes overly expensive for $M > 1$.

377 5 | DISCUSSION

378 In this work we have shown a general methodology which can be used to achieve nonstationary and anisotropic Matérn type
 379 correlation structures within a domain with complex boundaries. To summarize, the general procedure is as follows. First, one
 380 chooses a normalization length scale for each dimension, thereby defining (the Jacobian of) a mapping between a space where
 381 correlation is isotropic and stationary, and the more complex domain. These normalizing length scales are essential because
 382 they determine the local anisotropy and nonstationarity of the correlation operator. Next, one must choose a range parameter,

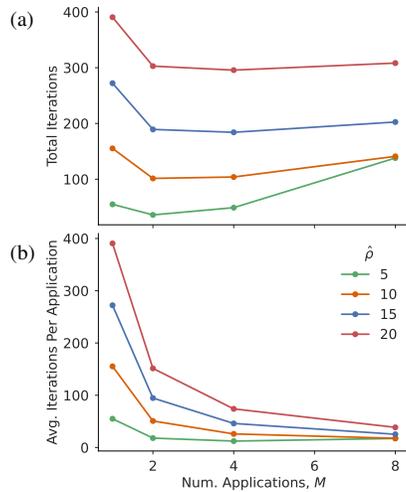


FIGURE 6 (a) Total number of iterations required to compute $\theta = \mathcal{A}^{-M} \mathbf{z}$, for a standard normally distributed vector $\mathbf{z} \in \mathbb{R}^{N_d}$. “Total iterations” is represented as the average number of total iterations from 1,000 random samples. (b) The average number of iterations per application of \mathcal{A}^{-M} , as a function of M . For all $\hat{\rho}$ and M combinations, each application gets cheaper as M increases. To compute the “average iterations per application”, we take the average iteration per application of \mathcal{A}^{-M} , and compute the sample average of this quantity from 1,000 random samples.

determining the distance relative to the normalizing length scales at which correlation drops to 0.14. Finally, one selects the shape of the correlation structure, which also sets the number of times the elliptic PDE must be solved.

Our presentation has focused on the practical application of this correlation operator within an ocean general circulation model. As such, we set the normalizing length scales based on the local grid scale. With this setting, the range parameter is shown to be a highly intuitive dial, controlling correlation length scales as a simple function of the number of neighboring grid cells. Using this definition was further shown to be beneficial at the equator, for example, because grid scale refinements there result in correlation length scales that are longer zonally than meridionally, coinciding with observed autocorrelation structures (Meyers et al., 1991). However, we recognize that there could be features that are desirable to capture in a correlation model that are not represented in the definition of the underlying model grid. In this case, the normalizing length scales could be further tuned with local factors or functions to achieve these desired features. Alternatively, these length scales could be set entirely independently of the grid, for instance as a function of a phenomenological length scale such as the local Rossby radius of deformation.

A key feature of the correlation model shown here is that the range parameter, $\hat{\rho}$, and the number of inverse elliptic operator applications, M , control the correlation length scale and shape *separately*. We consider this to be an attractive feature when compared to the implicit diffusion approach. Even when the “length scale” is fixed in the implicit diffusion model, changing M modifies the shape of the correlation function in such a way that there is no consistent characteristic distance, for instance at which correlation would drop below a threshold value (Figure 1, and also Figs. 1 and 2 from Guillet et al. (2019)). We note that in the Matérn correlation model presented here that the parameter $\hat{\delta}$ changes with M while in the implicit diffusion approach, $\delta(\mathbf{s}) \rightarrow I$. Apparently the simple variation in this parameter is enough to balance the multiple applications of \mathcal{A}^{-1} , such that the resulting correlation structure maintains a consistently identifiable length scale via $\hat{\rho}$.

As noted in Mirouze and Weaver (2010); Carrier and Ngodock (2010), a drawback to the explicit diffusion approach from Weaver and Courtier (2001) is that it requires many iterations to satisfy numerical stability. In our experimentation with this

model on the LLC grid as implemented in the MITgcm (Campin et al., 2021), we have found the number of iterations required for numerical stability to be roughly a factor of three larger than the necessary (but insufficient) lower bound for numerical stability. We therefore find approaches based on the implicit solution of a PDE to be more straightforward, as it is more intuitive to specify a solution tolerance rather than guess the number of iterations required for convergence. Moreover, our numerical experiments indicate that an imprecise solution (to a tolerance of $\sim 10^{-3}$) is sufficient for capturing the desired statistical behavior of the model, and therefore its implementation is highly efficient. Finally, because the correlation model shown here is formulated through an inverse elliptic operator, we have access to the inverse correlation operator, which could be used directly as regularization while solving an inverse problem (e.g., Bui-Thanh et al., 2013), or for specifying spatial correlations within the inverse observation error covariance as in Guillet et al. (2019).

For some applications it could be desirable to specify oscillating or ‘‘lobed’’ correlation models, which can be achieved with the explicit or implicit diffusion models (Weaver and Courtier, 2001; Weaver and Mirouze, 2013). We suggest that such extensions are possible for the Matérn type correlation operator shown here, based on results shown by L11 in the complex plane with a tunable oscillation parameter. These more general shapes could be explored in future work for the case of multi-dimensional fields as shown here.

A | DISCRETIZATION OF THE MATÉRN SPDE

In the following analysis we consider a 3D field $\theta(\mathbf{s})$, $\mathbf{s} \in \Omega$ and show the discretized form for \mathcal{A}^{-M} . We begin by showing the form for $M = 1$, and then simply illustrate how the operator can be applied iteratively for $M > 1$. We carry out the discretization on a structured, nonuniform, Arakawa C grid (Arakawa and Lamb, 1977) according to the finite volume method - as is the general setting in the MITgcm. We note that our development is similar to Fuglstad et al. (2014), who show a differential operator for a 2D field on a uniform grid.

Figure 7 shows the general structure of the grid, and defines the various grid cell distances used in the derivation. We begin by integrating Equation (15),

$$\begin{aligned} \int_{\Omega} \delta(\mathbf{s}) \theta(\mathbf{s}) \, d\mathbf{s} - \int_{\Omega} \nabla \cdot K(\mathbf{s}) \nabla \theta(\mathbf{s}) \, d\mathbf{s} &= \int_{\Omega} \mathcal{W}(\mathbf{s}) \det(\Phi(\mathbf{s}))^{-1/2} \, d\mathbf{s} \\ \sum_{i,j,k} \int_{E_{i,j,k}} \delta(\mathbf{s}) \theta(\mathbf{s}) \, d\mathbf{s} - \sum_{i,j,k} \int_{E_{i,j,k}} \nabla \cdot K(\mathbf{s}) \nabla \theta(\mathbf{s}) \, d\mathbf{s} &= \sum_{i,j,k} \int_{E_{i,j,k}} \mathcal{W}(\mathbf{s}) \det(\Phi(\mathbf{s}))^{-1/2} \, d\mathbf{s}, \end{aligned} \quad (18)$$

where in second line we distribute the integral across each grid cell $E_{i,j,k} \subset \Omega$ and i, j, k indicates indices in longitude, latitude, and depth, respectively.

Starting with the first term,

$$\begin{aligned} \delta_{i,j,k} &:= \frac{1}{V_{i,j,k}} \int_{E_{i,j,k}} \delta(\mathbf{s}) \, d\mathbf{s} \\ &= \frac{\hat{\delta}}{V_{i,j,k}} \int_{E_{i,j,k}} \frac{1}{\det(\Phi_{i,j,k})} \, d\mathbf{s} \end{aligned}$$

so that

$$\int_{E_{i,j,k}} \delta(\mathbf{s}) \theta(\mathbf{s}) \, d\mathbf{s} = \hat{\delta} \frac{V_{i,j,k}}{\det(\Phi_{i,j,k})} \theta_{i,j,k}$$

where $V_{i,j,k} = \Delta x_g^{i,j} \Delta y_g^{i,j} \Delta r_f^k$ is the grid cell volume, indicated by Figure 7. We note that a cartesian style notation is used to

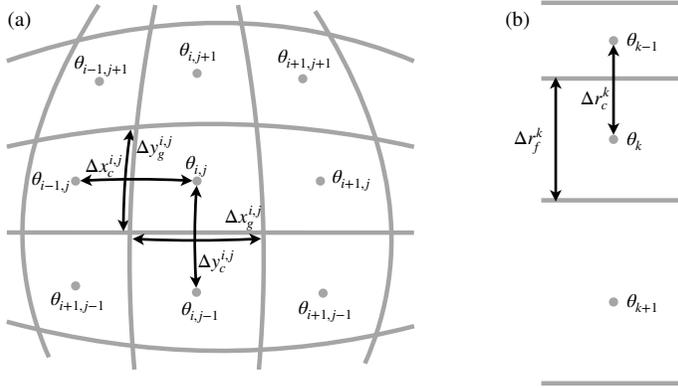


FIGURE 7 The structured finite volume grid used in the MITgcm. (a) The horizontal (latitude-longitude) plane viewed from above. (b) The vertical grid. The circles denote the tracer location at the grid cell center, where fields like temperature and salinity are located.

435 ease the presentation, but a more general curvilinear coordinate system is used in the computation in order to support the various
 436 sectors of the Lat-Lon-Cap (LLC) grid (see Section 2 and Appendix A of Forget et al., 2015, for more details on the LLC grid).

437 The third term is, from the definition of a white noise process (Adler and Taylor, 2007),

$$438 \int_{E_{i,j,k}} \det(\Phi(\mathbf{s}))^{-1/2} \mathcal{W}(\mathbf{s}) \, d\mathbf{s} = \sqrt{\frac{V_{i,j,k}}{\det(\Phi_{i,j,k})}} z_{i,j,k}$$

439 where $z_{i,j,k}$ is an uncorrelated (independent) standard Gaussian at each grid cell center i, j, k , and we used:

$$440 \det(\Phi_{i,j,k}) := \frac{1}{V_{i,j,k}} \int_{E_{i,j,k}} \det(\Phi(\mathbf{s})) \, d\mathbf{s}.$$

441 The second term, containing the Laplacian is handled as follows

$$442 \int_{E_{i,j,k}} \nabla \cdot \mathcal{K}(\mathbf{s}) \nabla \theta(\mathbf{s}) \, d\mathbf{s} = \int_{\partial E_{i,j,k}} \mathcal{K}(\mathbf{s}) \nabla \theta(\mathbf{s}) \cdot \hat{\mathbf{n}} \, d\mathbf{s}$$

443 where $\hat{\mathbf{n}}$ is an outward normal to the cell boundary $\partial E_{i,j,k}$. Throughout this work, we assume the tensor $\mathcal{K}(\mathbf{s})$ to be represented
 444 as the diagonal matrix:

$$445 \mathcal{K}(\mathbf{s}) = \begin{pmatrix} \kappa^{ux}(\mathbf{s}) & 0 & 0 \\ 0 & \kappa^{vy}(\mathbf{s}) & 0 \\ 0 & 0 & \kappa^{wz}(\mathbf{s}) \end{pmatrix},$$

446 which could be generalized in future work. We represent the discretized form of this tensor as

$$447 \quad K_{i,j,k} = \begin{pmatrix} \kappa_{i,j,k}^{ux} & 0 & 0 \\ 0 & \kappa_{i,j,k}^{vy} & 0 \\ 0 & 0 & \kappa_{i,j,k}^{wz} \end{pmatrix},$$

448 where the elements $\kappa_{i,j,k}^{ux}$, $\kappa_{i,j,k}^{vy}$, and $\kappa_{i,j,k}^{wz}$ determine the flux in and out of the grid cell:

$$449 \quad \kappa_{i,j,k}^{ux} := \frac{1}{\Delta y_g^{ij} \Delta r_f^k} \int_{\partial E_{i,j,k}^W} \kappa^{ux}(\mathbf{s}) \, ds$$

$$451 \quad \kappa_{i,j,k}^{vy} := \frac{1}{\Delta x_g^{ij} \Delta r_f^k} \int_{\partial E_{i,j,k}^S} \kappa^{vy}(\mathbf{s}) \, ds$$

$$453 \quad \kappa_{i,j,k}^{wz} := \frac{1}{\Delta x_g^{ij} \Delta y_g^{ij}} \int_{\partial E_{i,j,k}^B} \kappa^{wz}(\mathbf{s}) \, ds$$

454 where $\partial E_{i,j,k}^W$, $\partial E_{i,j,k}^S$, and $\partial E_{i,j,k}^B$ are the western, southern, and bottom boundaries of the grid cell. The discretized gradient is
455 approximated via the finite difference directional derivative at each cell face:

$$\begin{aligned} \frac{\partial \theta}{\partial x}(\mathbf{s}_{i,j,k}^W) &\approx \frac{\theta_{i,j,k} - \theta_{i-1,j,k}}{\Delta x_c^{ij}} & \frac{\partial \theta}{\partial x}(\mathbf{s}_{i,j,k}^E) &\approx \frac{\theta_{i+1,j,k} - \theta_{i,j,k}}{\Delta x_c^{i+1,j}} \\ \frac{\partial \theta}{\partial y}(\mathbf{s}_{i,j,k}^S) &\approx \frac{\theta_{i,j,k} - \theta_{i,j-1,k}}{\Delta y_c^{ij}} & \frac{\partial \theta}{\partial y}(\mathbf{s}_{i,j,k}^N) &\approx \frac{\theta_{j+1,k} - \theta_{i,j,k}}{\Delta y_c^{i,j+1}} \\ \frac{\partial \theta}{\partial z}(\mathbf{s}_{i,j,k}^T) &\approx \frac{\theta_{i,j,k-1} - \theta_{i,j,k}}{\Delta r_c^k} & \frac{\partial \theta}{\partial z}(\mathbf{s}_{i,j,k}^B) &\approx \frac{\theta_{i,j,k} - \theta_{i,j,k+1}}{\Delta r_c^{k+1}} \end{aligned}$$

457 for the west, east, south, north, top, and bottom cell faces, respectively. Note that the vertical coordinate is somewhat “flipped”
458 when compared to the other dimensions, where the index $k = 1$ at the surface and $k = N_k$ at depth. Putting these definitions
459 together,

$$\begin{aligned} \int_{\partial E_{i,j,k}} \mathcal{K}(\mathbf{s}) \nabla \theta(\mathbf{s}) \cdot \hat{\mathbf{n}} \, d\tau := & \\ & \left[\left(\frac{\kappa^{ux} \Delta y_g \Delta r_f}{\Delta x_c} \right)_{i+1,j,k} (\theta_{i+1,j,k} - \theta_{i,j,k}) - \left(\frac{\kappa^{ux} \Delta y_g \Delta r_f}{\Delta x_c} \right)_{i,j,k} (\theta_{i,j,k} - \theta_{i-1,j,k}) \right] + \\ & \left[\left(\frac{\kappa^{vy} \Delta x_g \Delta r_f}{\Delta y_c} \right)_{i,j+1,k} (\theta_{i,j+1,k} - \theta_{i,j,k}) - \left(\frac{\kappa^{vy} \Delta x_g \Delta r_f}{\Delta y_c} \right)_{i,j,k} (\theta_{i,j,k} - \theta_{i,j-1,k}) \right] + \\ & \left[\left(\frac{\kappa^{wz} \Delta x_g \Delta y_g}{\Delta r_c} \right)_{i,j,k} (\theta_{i,j,k-1} - \theta_{i,j,k}) - \left(\frac{\kappa^{wz} \Delta x_g \Delta y_g}{\Delta r_c} \right)_{i,j,k+1} (\theta_{i,j,k} - \theta_{i,j,k+1}) \right]. \end{aligned} \quad (19)$$

461 The discretized form of this operator is generated by defining the coefficients that form the seven point numerical stencil:

$$\begin{aligned}
 c_{\text{east}}^{i,j,k} &:= \left(\frac{\kappa^{ux} \Delta y_g \Delta r_f}{\Delta x_c} \right)_{i+1,j,k} & c_{\text{west}}^{i,j,k} &:= \left(\frac{\kappa^{ux} \Delta y_g \Delta r_f}{\Delta x_c} \right)_{i,j,k} \\
 c_{\text{north}}^{i,j,k} &:= \left(\frac{\kappa^{vy} \Delta x_g \Delta r_f}{\Delta y_c} \right)_{i,j+1,k} & c_{\text{south}}^{i,j,k} &:= \left(\frac{\kappa^{vy} \Delta x_g \Delta r_f}{\Delta y_c} \right)_{i,j,k} \\
 c_{\text{top}}^{i,j,k} &:= \left(\frac{\kappa^{wz} \Delta x_g \Delta y_g}{\Delta r_c} \right)_{i,j,k} & c_{\text{bottom}}^{i,j,k} &:= \left(\frac{\kappa^{wz} \Delta x_g \Delta y_g}{\Delta r_c} \right)_{i,j,k+1}
 \end{aligned}$$

$$c_{\text{center}}^{i,j,k} := -c_{\text{east}}^{i,j,k} - c_{\text{west}}^{i,j,k} - c_{\text{north}}^{i,j,k} - c_{\text{south}}^{i,j,k} - c_{\text{top}}^{i,j,k} - c_{\text{bottom}}^{i,j,k} .$$

465 With these definitions we define the matrix \mathbf{L} such that $\mathbf{L}\theta \simeq \int_{\Omega} \nabla \cdot \mathbf{K}(\mathbf{s}) \nabla \theta(\mathbf{s}) \, d\mathbf{s}$ where the row corresponding to the index
 466 i, j, k

$$\left[c_{\text{top}} \cdots c_{\text{south}} \cdots c_{\text{west}} c_{\text{center}} c_{\text{east}} \cdots c_{\text{north}} \cdots c_{\text{bottom}} \right]_{i,j,k}$$

468 With each term in Equation (18) defined above, we have the system of equations in matrix form:

$$\begin{aligned}
 (\mathbf{D}_\delta - \mathbf{L})\theta &= \mathbf{D}_z \mathbf{z} \\
 \mathbf{A}\theta &= \mathbf{D}_z \mathbf{z}
 \end{aligned} \tag{20}$$

470 where

$$\mathbf{D}_\delta := \text{diag} \left\{ \hat{\delta} \frac{V_n}{\det(\Phi_n)} \right\}_{n=1}^{N_\theta} \quad \mathbf{D}_z := \text{diag} \left\{ \sqrt{\frac{V_n}{\det(\Phi_n)}} \right\}_{n=1}^{N_\theta} \quad \mathbf{A} := (\mathbf{D}_\delta - \mathbf{L})$$

472 where for notational simplicity we index each grid cell with n . Note that when we prescribe Φ as in Section 3, then $\det(\Phi_n) = V_n$
 473 so that $\mathbf{D}_\delta = \hat{\delta} \mathbf{1}$ and $\mathbf{D}_z = \mathbf{I}$. The solution, θ , is the discretized form of a Matérn field with covariance $\mathbf{D}_z \mathbf{A}^{-2} \mathbf{D}_z$. The solution
 474 for cases when $M > 1$ is simply obtained through an iterative application of \mathbf{A}^{-1} (see also Theorem 4 in Appendix C.4 of L11),
 475 i.e. for $M > 1$:

$$\mathbf{A}^{M-1} \theta_M = \theta$$

477 A.1 | Neumann boundary conditions

478 Throughout this work we use zero flux, Neumann boundary conditions as follows:

$$\mathbf{K}(\mathbf{s}) \nabla \theta(\mathbf{s}) \cdot \hat{\mathbf{n}} = 0 \quad \mathbf{s} \in \partial \Omega_{\text{solid}} .$$

480 These boundary conditions are implemented by zeroing out the flux terms in the definition of \mathbf{L} along the boundary. More
 481 specifically, these are defined using a “mask” field which has values of one in the ocean and 0 on land. The mask is used to

modify the coefficients of \mathbf{L} as follows:

$$\begin{aligned}
 c_{\text{east}}^{i,j,k} &:= c_{\text{east}}^{i,j,k} m_{i,j,k} m_{i+1,j,k} & c_{\text{west}}^{i,j,k} &:= c_{\text{west}}^{i,j,k} m_{i,j,k} m_{i-1,j,k} \\
 c_{\text{north}}^{i,j,k} &:= c_{\text{north}}^{i,j,k} m_{i,j,k} m_{i,j+1,k} & c_{\text{south}}^{i,j,k} &:= c_{\text{south}}^{i,j,k} m_{i,j,k} m_{i,j-1,k} \\
 c_{\text{top}}^{i,j,k} &:= c_{\text{top}}^{i,j,k} m_{i,j,k} m_{i,j,k+1} & c_{\text{bottom}}^{i,j,k} &:= c_{\text{bottom}}^{i,j,k} m_{i,j,k} m_{i,j,k-1}
 \end{aligned}$$

where $m_{i,j,k}$ is used to denote the mask value at each grid cell.

A.2 | Correlation operator

The discretized form of the correlation operator used in this work is more formally defined as

$$\mathbf{C} := \mathbf{A}\mathbf{A}^{-M}\mathbf{D}_z. \quad (21)$$

To estimate the sample standard deviation used to fill \mathbf{A} , we draw 1,000 independent standard normal vectors $\mathbf{z}_l \in \mathbb{R}^{N_\theta}$ $l \in \{1, 2, \dots, 1000\}$, solve

$$\mathbf{A}^M \theta_l = \mathbf{D}_z \mathbf{z}_l \quad l \in \{1, 2, \dots, 1000\}, \quad (22)$$

and compute the pointwise standard deviation from $\{\theta_l\}_{l=1}^{1000}$.

A.3 | Code availability

This study used the MITgcm (Campin et al., 2021), with modifications that can be found at github.com/timothyas/MITgcm/tree/matern-correlation, for solving the presented linear system on the LLC grid. The grid files that make up the LLC grid can be found with the ECCOV4 data, which were accessed at (ECCO Consortium et al., 2020). All code used to generate, postprocess, and analyze the results in this paper can be found at Smith (2022). The python packages `xmitgcm` (Abernathey et al., 2021), `xarray` (Hoyer and Hamman, 2017), `dask` (Dask Development Team, 2016), and `zarr` (Miles et al., 2020) were essential for these tasks.

B | A BLOCK SUCCESSIVE OVER RELAXATION METHOD

Applying the correlation operator described in this paper, $\mathbf{C}^{1/2}$, requires the solution to an elliptic equation. Here we describe the block-Successive Over Relaxation (SOR) method that was implemented to solve this problem. We note that both preconditioned conjugate gradient algorithms exist in the MITgcm for 2D (in the latitude longitude plane) and 3D fields. However, the preconditioner for these solvers is designed specifically for the pressure solve at each time step (Marshall et al., 1997), and so may not be generally applicable to our problem. As such, we opted to implement the SOR algorithm outlined here because it was simple to do so, and efficient enough for our purposes.

The SOR method is an iterative method for solving $\mathbf{A}\mathbf{x} = \mathbf{b}$. At iteration k , the elements of \mathbf{x} are x_i^k , and we seek the update:

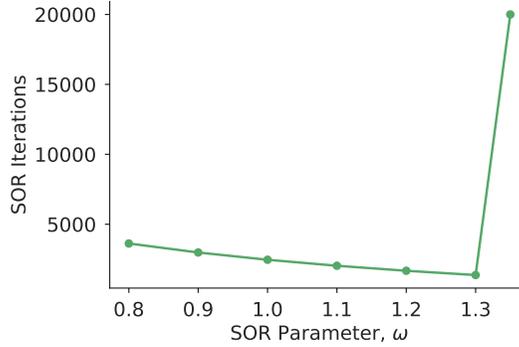


FIGURE 8 Performance of the SOR algorithm for various settings of ω using 10 samples, $\hat{\rho} = 10$, tolerance of 10^{-15} , and $M = 1$.

507

$$\bar{x}_i^{k+1} = (1 - \omega)x_i^k + \frac{\omega}{a_{ii}} \left(b_i - \sum_{j < i} a_{ij} \bar{x}_j^{k+1} - \sum_{j > i} a_{ij} x_j^k \right), \quad i = 1, 2, \dots, N, \quad (23)$$

508

509 where the elements of \mathbf{x} are x_i^k (and similarly for \mathbf{A} and \mathbf{b}), and ω is the SOR parameter. Here the notation \bar{x}_i^{k+1} refers to the
 510 fact this is a local update, i.e. there is no communication between the processes assigned to each portion of the computational
 511 domain. The only areas where this local update causes the algorithm to deviate from a standard SOR method is when neighboring
 512 elements \bar{x}_j^{k+1} , $i \neq j$ are in the “halo” regions of a process’s subdomain which are only updated at the end of each iteration. In
 513 this case, these neighboring elements take on the value from the previous iteration: $\bar{x}_j^{k+1} = x_j^k$.

514

We find this simple implementation to be an effective method for solving the linear system. Figure 8 shows the number
 515 of iterations required for convergence as a function of the parameter ω , with an optimal value of approximately $\omega^* = 1.3$. We
 516 note that the number of iterations required to converge at this optimal value is about half that of $\omega = 1$, which coincides with
 517 a (block) Gauss-Seidel algorithm (again differing from its true form due to the halo updates). Of course, Figure 8 shows the
 518 major drawback of the SOR algorithm: near the “optimal” value for the SOR parameter, efficiency is highly sensitive. For our
 519 implementation, at $\omega = 1.35$, we see that the algorithm does not fully converge and runs until set limit of 20,000 iterations.

520 ACKNOWLEDGEMENTS

521 I am grateful to Vikram Garg early discussions on the work by Lindgren et al. (2011), and to Jemima Tabeart for introducing me
 522 to the implicit diffusion operator developments by Mirouze and Weaver (2010). I thank Patrick Heimbach, Michael Diamond,
 523 Michael MacFerrin, and Fangfang Yao for comments which improved the manuscript.

524 REFERENCES

525 Ryan Abernathey, Raphael Dussin, Timothy Smith, Ian Fenty, Pascal Bourgault, Spencer Jones, Ed Doddridge, Fraser Goldsworth,
 526 Martin Losch, Mattia Almansi, Takaya Uchida, Andrea Cimattoribus, Aaron David Schneider, Adam Leskis, Antonio Quintana,
 527 Aurélien Ponte, Brian Rose, Dhruv Balwada, Guillaume Sérazin, and Liam Brannigan. Mitgcm/xmitgcm: v0.5.2, July 2021. URL
 528 <https://doi.org/10.5281/zenodo.5139886>.

- 529 R. J. Adler and Jonathan E. Taylor. *Random Fields and Geometry*. Springer Monographs in Mathematics. Springer-Verlag, New York, 2007.
530 ISBN 978-0-387-48112-8. doi: 10.1007/978-0-387-48116-6. URL <https://www.springer.com/gp/book/9780387481128>.
- 531 Akio Arakawa and Vivian R. Lamb. Computational Design of the Basic Dynamical Processes of the UCLA General Circulation Model.
532 In JULIUS Chang, editor, *Methods in Computational Physics: Advances in Research and Applications*, volume 17 of *General*
533 *Circulation Models of the Atmosphere*, pages 173–265. Elsevier, January 1977. doi: 10.1016/B978-0-12-460817-7.50009-4. URL
534 <https://www.sciencedirect.com/science/article/pii/B9780124608177500094>.
- 535 R. N. Bannister. A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics
536 and measurements of forecast error covariances. *Quarterly Journal of the Royal Meteorological Society*, 134(637):1951–1970,
537 2008a. ISSN 1477-870X. doi: 10.1002/qj.339. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.339>. _eprint:
538 <https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.339>.
- 539 R. N. Bannister. A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the
540 forecast error covariance statistics. *Quarterly Journal of the Royal Meteorological Society*, 134(637):1971–1996, 2008b.
541 ISSN 1477-870X. doi: 10.1002/qj.340. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.340>. _eprint:
542 <https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.340>.
- 543 A. F. Bennett, B. S. Chua, and L. M. Leslie. Generalized inversion of a global numerical weather prediction model. *Meteorology*
544 *and Atmospheric Physics*, 60(1-3):165–178, 1996. ISSN 0177-7971, 1436-5065. doi: 10.1007/BF01029793. URL <http://link.springer.com/10.1007/BF01029793>.
- 546 E. W. Blockley, M. J. Martin, A. J. McLaren, A. G. Ryan, J. Waters, D. J. Lea, I. Mirouze, K. A. Peterson, A. Sellar, and D. Storkey.
547 Recent development of the Met Office operational ocean forecasting system: an overview and assessment of the new Global FOAM
548 forecasts. *Geoscientific Model Development*, 7(6):2613–2638, November 2014. ISSN 1991-9603. doi: 10.5194/gmd-7-2613-2014.
549 URL <https://gmd.copernicus.org/articles/7/2613/2014/>.
- 550 Tan. Bui-Thanh, Omar. Ghattas, James. Martin, and Georg. Stadler. A Computational Framework for Infinite-Dimensional Bayesian
551 Inverse Problems Part I: The Linearized Case, with Application to Global Seismic Inversion. *SIAM Journal on Scientific Computing*,
552 35(6):A2494–A2523, January 2013. ISSN 1064-8275. doi: 10.1137/12089586X. URL [https://epubs.siam.org/doi/10.1137/](https://epubs.siam.org/doi/10.1137/12089586X)
553 [12089586X](https://epubs.siam.org/doi/10.1137/12089586X).
- 554 Jean-Michel Campin, Patrick Heimbach, Martin Losch, Gael Forget, edhill3, Alistair Adcroft, amolod, Dimitris Menemenlis, dfer22,
555 Chris Hill, Oliver Jahn, Jeff Scott, stephdut, Matt Mazloff, Baylor Fox-Kemper, antnguyen13, Ed Doddridge, Ian Fenty, Michael Bates,
556 AndrewEichmann-NOAA, Timothy Smith, Torge Martin, Jonathan Lauderdale, Ryan Abernathy, samarkhawi, hongandyan,
557 Bruno Deremble, dngoldberg, Pascal Bourgauf, and raphael dussin. MITgcm/MITgcm: checkpoint67z, June 2021. URL <https://doi.org/10.5281/zenodo.4968496>.
- 559 Matthew J. Carrier and Hans Ngodock. Background-error correlation model based on the implicit solution of a diffusion equation. *Ocean*
560 *Modelling*, 35(1-2):45–53, January 2010. ISSN 14635003. doi: 10.1016/j.ocemod.2010.06.003. URL [https://linkinghub.](https://linkinghub.elsevier.com/retrieve/pii/S1463500310000855)
561 [elsevier.com/retrieve/pii/S1463500310000855](https://linkinghub.elsevier.com/retrieve/pii/S1463500310000855).
- 562 N. Daget, Anthony T. Weaver, and M. A. Balmaseda. Ensemble estimation of background-error variances in a three-dimensional
563 variational data assimilation system for the global ocean. *Quarterly Journal of the Royal Meteorological Society*, 135(641):1071–
564 1094, 2009. ISSN 1477-870X. doi: 10.1002/qj.412. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.412>.
565 _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.412>.
- 566 Dask Development Team. *Dask: Library for dynamic task scheduling*, 2016. URL <https://dask.org>.
- 567 J. Derber and F. Bouttier. A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus A:*
568 *Dynamic Meteorology and Oceanography*, 51(2):195–221, January 1999. ISSN null. doi: 10.3402/tellusa.v51i2.12316. URL <https://doi.org/10.3402/tellusa.v51i2.12316>. Publisher: Taylor & Francis _eprint: <https://doi.org/10.3402/tellusa.v51i2.12316>.
- 570 John Derber and Anthony Rosati. A Global Oceanic Data Assimilation System. *Journal of Physical Oceanography*, 19(9):1333–1347,
571 September 1989. ISSN 0022-3670, 1520-0485. doi: 10.1175/1520-0485(1989)019<1333:AGODAS>2.0.CO;2. URL [http://](http://journals.ametsoc.org/view/journals/phoc/19/9/1520-0485_1989_019_1333_agodas_2_0_co_2.xml)
572 journals.ametsoc.org/view/journals/phoc/19/9/1520-0485_1989_019_1333_agodas_2_0_co_2.xml. Publisher: American
573 Meteorological Society Section: Journal of Physical Oceanography.

- 574 Emanuele Di Lorenzo, Andrew M. Moore, Hernan G. Arango, Bruce D. Cornuelle, Arthur J. Miller, Brian Powell, Boon S. Chua, and
575 Andrew F. Bennett. Weak and strong constraint data assimilation in the inverse Regional Ocean Modeling System (ROMS): Develop-
576 ment and application for a baroclinic coastal upwelling system. *Ocean Modelling*, 16(3):160–187, January 2007. ISSN 1463-5003.
577 doi: 10.1016/j.ocemod.2006.08.002. URL <https://www.sciencedirect.com/science/article/pii/S1463500306000916>.
- 578 Srdjan Dobricic and Nadia Pinardi. An oceanographic three-dimensional variational data assimilation scheme. *Ocean Modelling*, 22
579 (3-4):89–105, January 2008. ISSN 14635003. doi: 10.1016/j.ocemod.2008.01.004. URL [https://linkinghub.elsevier.com/
580 retrieve/pii/S1463500308000176](https://linkinghub.elsevier.com/retrieve/pii/S1463500308000176).
- 581 ECCO Consortium, Ian Fenty, Ou Wang, Ou Forget, Patrick Heimbach, and Rui Ponte. Ecco Central Estimate (Version 4 Release 2), jul
582 2020. URL <https://ecco.jpl.nasa.gov/drive/files>.
- 583 Gary D. Egbert, Andrew F. Bennett, and Michael G. G. Foreman. TOPEX/POSEIDON tides estimated using a
584 global inverse model. *Journal of Geophysical Research: Oceans*, 99(C12):24821–24852, 1994. ISSN 2156-
585 2202. doi: 10.1029/94JC01894. URL <http://onlinelibrary.wiley.com/doi/abs/10.1029/94JC01894>. _eprint:
586 <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/94JC01894>.
- 587 G. Forget, J.-M. Campin, P. Heimbach, C. N. Hill, R. M. Ponte, and C. Wunsch. Ecco version 4: an integrated framework for
588 non-linear inverse modeling and global ocean state estimation. *Geoscientific Model Development*, 8(10):3071–3104, 2015. doi:
589 10.5194/gmd-8-3071-2015. URL <http://www.geosci-model-dev.net/8/3071/2015/>.
- 590 Geir-Arne Fuglstad, Finn Lindgren, Daniel Simpson, and Håvard Rue. Exploring a New Class of Non-stationary Spatial Gaussian
591 Random Fields with Varying Local Anisotropy. *arXiv:1304.6949 [stat]*, April 2014. URL <http://arxiv.org/abs/1304.6949>.
592 arXiv: 1304.6949.
- 593 Gregory Gaspari and Stephen E. Cohn. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal*
594 *Meteorological Society*, 125(554):723–757, 1999. ISSN 1477-870X. doi: 10.1002/qj.49712555417. URL [https://onlinelibrary.
595 wiley.com/doi/abs/10.1002/qj.49712555417](https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.49712555417). _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.49712555417>.
- 596 Gregory Gaspari, Stephen E. Cohn, Jing Guo, and Steven Pawson. Construction and application of covariance functions
597 with variable length-fields. *Quarterly Journal of the Royal Meteorological Society*, 132(619):1815–1838, 2006. ISSN
598 1477-870X. doi: 10.1256/qj.05.08. URL <https://onlinelibrary.wiley.com/doi/abs/10.1256/qj.05.08>. _eprint:
599 <https://onlinelibrary.wiley.com/doi/pdf/10.1256/qj.05.08>.
- 600 Tilmann Gneiting. Correlation functions for atmospheric data analysis. *Quarterly Journal of the Royal Meteorological Society*, 125
601 (559):2449–2464, 1999. ISSN 1477-870X. doi: 10.1002/qj.49712555906. URL [https://onlinelibrary.wiley.com/doi/abs/
602 10.1002/qj.49712555906](https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.49712555906). _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.49712555906>.
- 603 Gene H Golub and Charles F Van Loan. *Matrix computations*. JHU press, 2013.
- 604 Stephen M. Griffies. *Fundamentals of ocean climate models*. Princeton university press, 2004.
- 605 Stephen M. Griffies and Robert W. Hallberg. Biharmonic Friction with a Smagorinsky-Like Viscosity for Use in Large-Scale Eddy-
606 Permitting Ocean Models. *Monthly Weather Review*, 128(8):2935–2946, August 2000. ISSN 1520-0493, 0027-0644. doi: 10.
607 1175/1520-0493(2000)128<2935:BFWASL>2.0.CO;2. URL [https://journals.ametsoc.org/view/journals/mwre/128/
608 8/1520-0493_2000_128_2935_bfwasl_2.0.co_2.xml](https://journals.ametsoc.org/view/journals/mwre/128/8/1520-0493_2000_128_2935_bfwasl_2.0.co_2.xml). Publisher: American Meteorological Society Section: Monthly Weather
609 Review.
- 610 Oliver Guillet, Anthony T. Weaver, Xavier Vasseur, Yann Michel, Serge Gratton, and Selime Gürol. Modelling spatially correlated
611 observation errors in variational data assimilation using a diffusion operator on an unstructured mesh. *Quarterly Journal of the*
612 *Royal Meteorological Society*, 145(722):1947–1967, 2019. ISSN 1477-870X. doi: 10.1002/qj.3537. URL [http://onlinelibrary.
613 wiley.com/doi/abs/10.1002/qj.3537](http://onlinelibrary.wiley.com/doi/abs/10.1002/qj.3537). _eprint: <https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3537>.

- 614 William R. Holland. The Role of Mesoscale Eddies in the General Circulation of the Ocean—Numerical Experiments Using a Wind-
615 Driven Quasi-Geostrophic Model. *Journal of Physical Oceanography*, 8(3):363–392, May 1978. ISSN 0022-3670, 1520-0485. doi:
616 10.1175/1520-0485(1978)008<0363:TROMEI>2.0.CO;2. URL [http://journals.ametsoc.org/view/journals/phoc/8/
617 3/1520-0485_1978_008_0363_tromei_2_0_co_2.xml](http://journals.ametsoc.org/view/journals/phoc/8/3/1520-0485_1978_008_0363_tromei_2_0_co_2.xml). Publisher: American Meteorological Society Section: Journal of Physical
618 Oceanography.
- 619 S. Hoyer and J. Hamman. xarray: N-D labeled arrays and datasets in Python. *Journal of Open Research Software*, 5(1), 2017. doi:
620 10.5334/jors.148. URL <https://doi.org/10.5334/jors.148>.
- 621 Kayo Ide, Philippe Courtier, Michael Ghil, and Andrew C. Lorenc. Unified Notation for Data Assimilation : Operational, Sequential and
622 Variational (gtSpecial IssueltData Assimilation in Meteology and Oceanography: Theory and Practice). *Journal of the Meteorological
623 Society of Japan. Ser. II*, 75(1B):181–189, 1997. doi: 10.2151/jmsj1965.75.1B_181.
- 624 Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between gaussian fields and gaussian markov random fields: the
625 stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73
626 (4):423–498, 2011. ISSN 1467-9868. doi: 10.1111/j.1467-9868.2011.00777.x. URL [http://dx.doi.org/10.1111/j.1467-
627 9868.2011.00777.x](http://dx.doi.org/10.1111/j.1467-9868.2011.00777.x).
- 628 Martin Losch, Dimitris Menemenlis, Jean-Michel Campin, Patick Heimbach, and Chris Hill. On the formulation of sea-ice models. Part 1:
629 Effects of different solver implementations and parameterizations. *Ocean Modelling*, 33(1):129–144, January 2010. ISSN 1463-5003.
630 doi: 10.1016/j.ocemod.2009.12.008. URL <http://www.sciencedirect.com/science/article/pii/S1463500309002418>.
- 631 John Marshall, Alistair Adcroft, Chris Hill, Lev Perelman, and Curt Heisey. A finite-volume, incompressible Navier Stokes model for
632 studies of the ocean on parallel computers. *Journal of Geophysical Research: Oceans*, 102(C3):5753–5766, 1997. ISSN 2156-2202.
633 doi: 10.1029/96JC02775. URL <http://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/96JC02775>.
- 634 Gary Meyers, Helen Phillips, Neville Smith, and Janet Sprintall. Space and time scales for optimal interpolation of temperature — Tropical
635 Pacific Ocean. *Progress in Oceanography*, 28(3):189–218, January 1991. ISSN 0079-6611. doi: 10.1016/0079-6611(91)90008-A.
636 URL <https://www.sciencedirect.com/science/article/pii/007966119190008A>.
- 637 Alistair Miles, John Kirkham, Martin Durant, James Bourbeau, Tarik Onalan, Joe Hamman, Zain Patel, shikharsg, Matthew Rocklin,
638 raphael dussin, Vincent Schut, Elliott Sales de Andrade, Ryan Abernathy, Charles Noyes, sbalmer, pyup.io bot, Tommy Tran,
639 Stephan Saalfeld, Justin Swaney, Josh Moore, Joe Jevnik, Jerome Kelleher, Jan Funke, George Sakkis, Chris Barnes, and Anderson
640 Banihirwe. zarr-developers/zarr-python: v2.4.0, January 2020. URL <https://doi.org/10.5281/zenodo.3773450>.
- 641 I. Mirouze and Anthony T. Weaver. Representation of correlation functions in variational assimilation using an im-
642 plicit diffusion operator. *Quarterly Journal of the Royal Meteorological Society*, 136(651):1421–1443, 2010. ISSN
643 1477-870X. doi: 10.1002/qj.643. URL <http://onlinelibrary.wiley.com/doi/abs/10.1002/qj.643>.
644 [_eprint:
https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.643](https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.643).
- 645 Andrew M. Moore, Hernan G. Arango, Gregoire Broquet, Brian S. Powell, Anthony T. Weaver, and Javier Zavala-Garay. The
646 Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems: Part I – System overview and
647 formulation. *Progress in Oceanography*, 91(1):34–49, October 2011. ISSN 0079-6611. doi: 10.1016/j.pocan.2011.05.004. URL
648 <http://www.sciencedirect.com/science/article/pii/S0079661111000516>.
- 649 J. C. Muccino, H. Luo, H. G. Arango, D. Haidvogel, J. C. Levin, A. F. Bennett, B. S. Chua, G. D. Egbert, B. D. Cornuelle, A. J. Miller,
650 E. Di Lorenzo, A. M. Moore, and E. D. Zaron. The Inverse Ocean Modeling System. Part II: Applications. *Journal of Atmospheric
651 and Oceanic Technology*, 25(9):1623–1637, September 2008. ISSN 0739-0572, 1520-0426. doi: 10.1175/2008JTECHO522.1.
652 URL http://journals.ametsoc.org/view/journals/atot/25/9/2008jtecho522_1.xml. Publisher: American Meteorological
653 Society Section: Journal of Atmospheric and Oceanic Technology.
- 654 An T. Nguyen, Helen Pillar, Victor Ocaña, Arash Bigdeli, Timothy A. Smith, and Patrick Heimbach. The Arctic Subpolar
655 Gyre sTate Estimate: Description and Assessment of a Data-Constrained, Dynamically Consistent Ocean-Sea Ice Estimate for
656 2002–2017. *Journal of Advances in Modeling Earth Systems*, 13(5):e2020MS002398, 2021. ISSN 1942-2466. doi: <https://doi.org/10.1029/2020MS002398>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002398>.
657 [_eprint:
658 https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020MS002398](https://doi.org/10.1029/2020MS002398).

- 659 R. James Purser, Wan-Shu Wu, David F. Parrish, and Nigel M. Roberts. Numerical Aspects of the Application of Recursive Filters
660 to Variational Statistical Analysis. Part I: Spatially Homogeneous and Isotropic Gaussian Covariances. *Monthly Weather Review*,
661 131(8):1524–1535, August 2003a. ISSN 1520-0493, 0027-0644. doi: 10.1175//1520-0493(2003)131<1524:NAOTAO>2.
662 O.CO;2. URL [http://journals.ametsoc.org/view/journals/mwre/131/8/_1520-0493_2003_131_1524_naotao_2.0.co_2.](http://journals.ametsoc.org/view/journals/mwre/131/8/_1520-0493_2003_131_1524_naotao_2.0.co_2.xml)
663 [xml](http://journals.ametsoc.org/view/journals/mwre/131/8/_1520-0493_2003_131_1524_naotao_2.0.co_2.xml). Publisher: American Meteorological Society Section: Monthly Weather Review.
- 664 R. James Purser, Wan-Shu Wu, David F. Parrish, and Nigel M. Roberts. Numerical Aspects of the Application of Recursive Filters to
665 Variational Statistical Analysis. Part II: Spatially Inhomogeneous and Anisotropic General Covariances. *Monthly Weather Review*,
666 131(8):1536–1548, August 2003b. ISSN 1520-0493, 0027-0644. doi: 10.1175//2543.1. URL [http://journals.ametsoc.org/](http://journals.ametsoc.org/view/journals/mwre/131/8/_2543.1.xml)
667 [view/journals/mwre/131/8/_2543.1.xml](http://journals/mwre/131/8/_2543.1.xml). Publisher: American Meteorological Society Section: Monthly Weather Review.
- 668 R. James Purser, Miodrag Rancic, and Manuel S. F. V. De Ponca. The Multigrid Beta Function Approach for Modeling of Background
669 Error Covariance in the Real-Time Mesoscale Analysis (RTMA). *Monthly Weather Review*, 150(4):715–732, April 2022. ISSN
670 1520-0493, 0027-0644. doi: 10.1175/MWR-D-20-0405.1. URL [https://journals-ametsoc-org.colorado.idm.oclc.org/](https://journals-ametsoc-org.colorado.idm.oclc.org/view/journals/mwre/150/4/MWR-D-20-0405.1.xml)
671 [view/journals/mwre/150/4/MWR-D-20-0405.1.xml](https://journals-ametsoc-org.colorado.idm.oclc.org/view/journals/mwre/150/4/MWR-D-20-0405.1.xml). Publisher: American Meteorological Society Section: Monthly Weather
672 Review.
- 673 Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. Adaptive computation and machine
674 learning. MIT Press, Cambridge, Mass, 2006. ISBN 978-0-262-18253-9. OCLC: ocm61285753.
- 675 Paul D. Sampson and Peter Guttorp. Nonparametric Estimation of Nonstationary Spatial Covariance Structure. *Journal of the American*
676 *Statistical Association*, 87(417):108–119, 1992. ISSN 0162-1459. doi: 10.2307/2290458. URL [https://www.jstor.org/stable/](https://www.jstor.org/stable/2290458)
677 [2290458](https://www.jstor.org/stable/2290458). Publisher: [American Statistical Association, Taylor & Francis, Ltd.].
- 678 T. Smith. Change of variables in Laplace’s and other second-order differential equations. *Proceedings of the Physical Society*, 46
679 (3):344–349, May 1934. ISSN 0959-5309. doi: 10.1088/0959-5309/46/3/307. URL [https://doi.org/10.1088/0959-](https://doi.org/10.1088/0959-5309/46/3/307)
680 [5309/46/3/307](https://doi.org/10.1088/0959-5309/46/3/307). Publisher: IOP Publishing.
- 681 Timothy Smith. timothyas/generic-matern-covariance: Initial Submission, July 2022. URL [https://doi.org/10.5281/zenodo.](https://doi.org/10.5281/zenodo.6878930)
682 [6878930](https://doi.org/10.5281/zenodo.6878930).
- 683 Lynne D Talley. *Descriptive physical oceanography: an introduction*. Academic press, 2011.
- 684 Geoffrey K Vallis. *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-scale Circulation*. Cambridge Univ. Press, 2006.
- 685 Anthony T. Weaver and Philippe Courtier. Correlation modelling on the sphere using a generalized diffusion equation. *Quarterly*
686 *Journal of the Royal Meteorological Society*, 127(575):1815–1846, 2001. ISSN 1477-870X. doi: 10.1002/qj.4971275718. URL
687 <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.4971275718>.
- 688 Anthony T. Weaver and I. Mirouze. On the diffusion equation and its application to isotropic and anisotropic correlation
689 modelling in variational assimilation. *Quarterly Journal of the Royal Meteorological Society*, 139(670):242–260, 2013.
690 ISSN 1477-870X. doi: 10.1002/qj.1955. URL <http://onlinelibrary.wiley.com/doi/abs/10.1002/qj.1955>.
691 [_eprint:
https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.1955](https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.1955).
- 692 Anthony T. Weaver, J. Vialard, and D. L. T. Anderson. Three- and Four-Dimensional Variational Assimilation with a General Circulation
693 Model of the Tropical Pacific Ocean. Part I: Formulation, Internal Diagnostics, and Consistency Checks. *Monthly Weather Review*, 131
694 (7):1360–1378, July 2003. ISSN 1520-0493, 0027-0644. doi: 10.1175/1520-0493(2003)131<1360:TAFVAW>2.0.CO;2. URL
695 http://journals.ametsoc.org/view/journals/mwre/131/7/1520-0493_2003_131_1360_tafvaw_2.0.co_2.xml. Publisher:
696 American Meteorological Society Section: Monthly Weather Review.
- 697 Anthony T. Weaver, C. Deltel, E. Machu, S. Ricci, and N. Daget. A multivariate balance operator for variational ocean
698 data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3605–3625, 2005. ISSN 1477-
699 870X. doi: 10.1256/qj.05.119. URL <https://onlinelibrary.wiley.com/doi/abs/10.1256/qj.05.119>.
700 [_eprint:
https://onlinelibrary.wiley.com/doi/pdf/10.1256/qj.05.119](https://onlinelibrary.wiley.com/doi/pdf/10.1256/qj.05.119).

- 701 Anthony T. Weaver, Marcin Chrust, Benjamin Ménétrier, and Andrea Piacentini. An evaluation of methods for normalizing diffusion-
702 based covariance operators in variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 147(734):289–320,
703 2021. ISSN 1477-870X. doi: 10.1002/qj.3918. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.3918>. _eprint:
704 <https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3918>.
- 705 C. Wunsch. *Discrete Inverse and State Estimation Problems. With Geophysical Fluid Applications*. Cambridge Un. Press, Cambridge,
706 2006.