A HESSIAN-BASED METHOD FOR UNCERTAINTY QUANTIFICATION IN GLOBAL OCEAN STATE ESTIMATION[*]

ALEXANDER G. KALMIKOV[†] AND PATRICK HEIMBACH[†]

Abstract. Derivative-based methods are developed for uncertainty quantification (UQ) in large-scale ocean state estimation. The estimation system is based on the adjoint method for solving a least-squares optimization problem, whereby the state-of-the-art MIT general circulation model (MITgcm) is fit to observations. The UQ framework is applied to quantify Drake Passage transport uncertainties in a global idealized barotropic configuration of the MITgcm. Large error covariance matrices are evaluated by inverting the Hessian of the misfit function using matrix-free numerical linear algebra algorithms. The covariances are projected onto target output quantities of the model (here Drake Passage transport) by Jacobian transformations. First and second derivative codes of the MITgcm are generated by means of algorithmic differentiation (AD). Transpose of the chain rule product of Jacobians of elementary forward model operations implements a computationally efficient adjoint code. Computational complexity of the Hessian code is reduced via forward-over-reverse mode AD, which preserves the efficiency of adjoint checkpointing schemes in the second derivative calculation. A Lanczos algorithm is applied to extract the leading eigenvectors and eigenvalues of the Hessian matrix, representing the constrained uncertainty patterns and the inverse of the corresponding uncertainties. The dimensionality of the misfit Hessian inversion is reduced by omitting its nullspace (as an alternative to suppressing it by regularization), excluding from the computation the uncertainty subspace unconstrained by the observations. Inverse and forward uncertainty propagation schemes are designed for assimilating observation and control variable uncertainties and for projecting these uncertainties onto oceanographic target quantities.

1. Introduction. The ocean is an essential component of the global climate system. Its circulation exhibits significant variability on all time scales (e.g., [40]), from seconds (surface waves) to decades (Rossby wave propagation) to millennia (slow diffusive processes). Various field programs, including the World Ocean Circulation Experiment (WOCE) or the Argo profiling float program, as well as satellite missions, such as TOPEX/POSEIDON and its follow-up altimetric missions or the Gravity Recovery And Climate Experiment (GRACE) gravimetric mission, have provided very large, disparate data sets to probe specific elements of the circulation (e.g., [47]). Yet, because of its opaqueness to electromagnetic radiation, remote sensing methods to comprehensively monitor the ocean's evolution in space and time are limited, for the most part, to surface properties, thus rendering oceanographic research an overall "data-starved" field.

Oceanographers have therefore turned to formal estimation methods to synthesize sparse and heterogeneous data streams into a coherent dynamical framework through the use of ocean general circulation models [42]. One such effort is the Estimating

---

[†]Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139 (kalex@alum.mit.edu, heimbach@mit.edu).

the Circulation and Climate of the Oceans (ECCO) consortium [36, 45]. Over the last decade it has developed a dynamically consistent framework, in which the MIT general circulation model (MITgcm) [24, 1] is fit in a least-squares sense to most available observations that have been collected over the last two decades. ECCO has resorted to the adjoint, or Lagrange multiplier, method for computing the gradient of a least-squares cost function with respect to a very high-dimensional (typically 10^8) input or control vector [44]. That vector captures the major uncertain variables and their spatiotemporal characteristics, which serve as inputs to the ocean simulations, consisting of three-dimensional fields of initial conditions and model parameters (e.g., mixing coefficients), as well as time-varying two-dimensional surface boundary conditions used to force the ocean (winds, surface air temperature, precipitation, radiation). The use of the adjoint permits the simultaneous adjustment of these inputs, in the following referred to as *controls*, in the course of the gradient-based nonlinear optimization, and the inference of an optimal estimate of the time-evolving ocean state, which is being used in a variety of ocean and climate science studies (see [46, 47]for recent overviews). Those studies mandate, instead of filter methods like the Kalman filter, the use of smoother methods such as the adjoint. They produce state estimates in which all properties (heat, freshwater) are properly accounted for through time, following known conservation laws and equations of motions exactly [47].

A serious limitation and outstanding challenge in ECCO has remained the neglect of formal methods that take into account uncertainties in the observations and in the a priori knowledge on the controls to infer uncertainties in the estimated state and derived output quantities of interest (QoI) that are of relevance for ocean and climate research. Here, we present a computationally feasible derivative-based strategy, which enables such uncertainty quantification within the ECCO framework. The strategy is based on a two-step approach: in a first step posterior control uncertainties are inferred from observation and a priori control uncertainties as part of the least-squares optimization ("inverse uncertainty propagation"); in a second step the prior and posterior control uncertainties are mapped to uncertainties in the QoI via the Jacobian of the forward model ("forward uncertainty propagation"). Comparison between forward propagation of prior and posterior control uncertainties enables the assessment of impact of observations in reducing uncertainties in the QoI.

Key ingredients to this strategy are the availability of first (tangent linear and adjoint) and second (Hessian) order derivatives of the general circulation model. In particular, the Hessian is a measure of the uncertainty of the least-squares problem through its relationship to the inverse error covariance (e.g., [38]). In the context of ECCO, we take advantage of the ability to flexibly generate first and second order derivative codes of the MITgcm via algorithmic differentiation [12, 10]. Calculation of the full Hessian matrix, however, is not feasible for high-dimensional uncertainty spaces, such as considered in ECCO. Another goal, then, is the compression of the Hessian by removal of its effective nullspace and the implementation of matrix-free numerical linear algebra approaches which enable calculation of leading spatial (and temporal) uncertainty reduction patterns.

To demonstrate our framework in the context of the MITgcm we present an application to quantifying uncertainties in the zonal (eastward) vertically integrated mass or volume transport in the Southern Ocean. In view of the challenges associated with estimating this quantity from a complex flow, practical considerations have led oceanographers to monitor this transport at the location of Drake Passage, the narrowest constriction of the Antarctic Circumpolar Current (ACC) between South

America and the West Antarctic Peninsula [33, 29]. The ACC represents one of the dynamically most important circulation features of the world ocean (e.g., [32]) which significantly impacts the global carbon cycle [16], thus giving Drake Passage transport a prominent role among a number of ocean-related climate indices (e.g., [41]). Regional state estimation efforts have combined available observations of the Southern Ocean to constrain model-based estimates of its flow over recent years [28] and infer its sensitivity to various forcing perturbations [27]. However, uncertainties provided along with estimated Drake Passage transport of 153 ± 5 Sv have been based solely on 10-day temporal variability and have not captured contributions from observation, forcing, or initial condition uncertainties.

For the present purpose of proof-of-concept, we restrict ourselves to the consideration of vertically integrated barotropic motions of the flow, thus excluding initially important baroclinic contributions. Simulations presented here are from an idealized configuration at $2° \times 2°$ horizontal resolution and with simulated rather than real observations.

The nonlinear least-squares problem is solved with the adjoint method by iterative gradient-based minimization of the cost function. The uncertainty of the nonlinear solution is quantified with the *Hessian method*, based on the second-order expansion of the cost function at the solution point [42]. Our method is related to earlier efforts on error analysis in the context of ocean data assimilation. Specifically, Moore et al. [30] have developed a four-dimensional variational (4DVar) data assimilation and posterior error covariance modeling framework in the context of the Regional Ocean Modeling System (ROMS). Our framework extends these efforts in a number of ways: (1) ECCO-type configurations typically target global rather than regional simulations and time scales ranging from years to decades; (2) derivative code may be flexibly generated using algorithmic differentiation for different control problems and model configurations, e.g., using as control variables model parameters, passive tracers, sea ice, or domain geometry (see references in [47]); and (3) we compute the complete nonlinear Hessian, which is necessary for nonlinear large-residual problems [35]. The Hessian inversion algorithm implemented preserves the original prior-independent Hessian eigenvector structure necessary to maintain the dynamical consistency of the assimilated uncertainty fields, separating their signal from a numerically motivated prior. In this regard our method differs from Bayesian UQ schemes based on simultaneous diagonalization of the likelihood Hessian and the prior [26] or diagonalization of the prior-preconditioned linearized Hessian [5].

While our perspective is deterministic, close links exist to Bayesian methods for UQ, which offer exciting extensions but are not explored further here (e.g., [2, 25, 4]). The derivative codes obtained via AD retain the nonlinearity of the ocean dynamics through derivative evaluation with respect to the exact time-evolving state, but provide only a linearized leading order description of uncertainty propagation. Although in principle this is inferior to non-Gaussian nonlinear UQ techniques [18, 34], in practice those are susceptible to the "curse of dimensionality" [31] and are, unlike the adjoint method, not scalable to big data problems that involve high-dimensional uncertainty spaces. The geophysical origin of the high dimensionality lies with the spatiotemporal nonuniform uncertainty patterns and their spread across different physical variables.

The remainder of the paper is organized as follows: we provide a brief but self-consistent conceptual framework of the proposed UQ method within the ECCO infrastructure (section 2). We then describe the developed numerical machinery for

computing and inverting the Hessian via matrix-free and order reduction methods (section 3), with technical details and verification tests reported in the appendix. An application to quantify uncertainties of Drake Passage transport in a barotropic configuration of the MITgcm (section 4) demonstrates the ability of the framework to reduce uncertainties in the transport based on propagation of observation and prior control uncertainties to posterior uncertainties in the optimal control estimates. A discussion section provides a perspective of the work presented for global ocean state estimation in support for climate research.

2. Uncertainty quantification methodology.

2.1. Inverse modeling framework for ocean state estimation. The proposed UQ methodology for global ocean state estimation is developed in the least-squares (or minimum variance) estimation framework of ECCO [44]. An optimal fit of a multivariate nonlinear forward model $\boldsymbol{M}(\mathbf{x})$ to observations $\mathbf{y}$ is sought through adjustment of its initial and boundary conditions, forcing fields, and geometrical and physical parameters, arranged together in the control vector $\mathbf{x}$. In the presence of noise $\mathbf{n}$ the model $\boldsymbol{M}$ is connected to observations via

$$(1) \qquad \mathbf{y} = \boldsymbol{M}(\mathbf{x}) + \mathbf{n}.$$

Because it is, in general, not invertible, the optimal value for $\mathbf{x}$ is sought through minimization of a least-squares cost function, whose general form is

$$(2) \qquad J(\mathbf{x}) = \frac{1}{2} \left[(\boldsymbol{M}(\mathbf{x}) - \mathbf{y})^T \mathbf{R}^{-1} (\boldsymbol{M}(\mathbf{x}) - \mathbf{y}) + (\mathbf{x} - \mathbf{x}_0)^T \mathbf{P}_0^{-1} (\mathbf{x} - \mathbf{x}_0) \right].$$

The first term represents the model-data misfit weighted by the uncertainty of observations, given by the second moment matrix

$$(3) \qquad \mathbf{R} = \left\langle \mathbf{n}\,\mathbf{n}^T \right\rangle,$$

which, for Gaussian noise with zero mean, is equal to the covariance matrix of noise. The second term has two interpretations. First, in the context of inversion the term regularizes an in general ill-posed problem, suppressing its nullspace when observations are insufficient to fully constrain the solution. Second, to the extent that prior information (mean and dispersion) exists on the controls, the regularization term penalizes deviations of the controls from the prior $\mathbf{x_0}$. The quadratic penalty is weighted by the a priori assumed uncertainty of the controls, given by their prior covariance if $\mathbf{x_0}$ is the prior mean of the unknown controls $\mathbf{x}$:

$$(4) \qquad \mathbf{P}_0 = \left\langle (\mathbf{x}_0 - \mathbf{x})\,(\mathbf{x}_0 - \mathbf{x})^T \right\rangle.$$

The solution to the inverse problem can be found by noting that at the minimum of the cost function its gradient[1] is zero:

$$(5) \qquad \nabla J(\hat{\mathbf{x}}) = \left(\frac{\partial \boldsymbol{M}}{\partial \mathbf{x}^T} \right)^T \mathbf{R}^{-1} \left(\boldsymbol{M}(\hat{\mathbf{x}}) - \mathbf{y} \right) + \mathbf{P}_0^{-1} \left(\hat{\mathbf{x}} - \mathbf{x}_0 \right) = 0.$$

---

[1]Adopting the standard vector derivatives notation (e.g., [22, 35]), the gradient of a multivariate function is defined as transpose of its Jacobian matrix: $\nabla \boldsymbol{M}(\mathbf{x}) \equiv \partial \boldsymbol{M}/\partial \mathbf{x} = (\partial \boldsymbol{M}/\partial \mathbf{x}^T)^T$.

If the forward model $\boldsymbol{M}$ is only weakly nonlinear and its first order Taylor expansion is a valid approximation throughout the optimization domain, the model can be expressed in terms of model Jacobian (or tangent linear model) $\partial \boldsymbol{M}/\partial \mathbf{x}^T$

$$(6) \qquad \boldsymbol{M}(\mathbf{x}) = \boldsymbol{M}(\mathbf{x}_0) + \left.\frac{\partial \boldsymbol{M}}{\partial \mathbf{x}^T}\right|_{\mathbf{x}_0} (\mathbf{x} - \mathbf{x}_0) + O(\|\mathbf{x} - \mathbf{x}_0\|^2).$$

Within the tangent linear region the solution to the inverse problem can be found by a linear transformation of the a priori misfit, given by

$$(7) \quad \hat{\mathbf{x}}_\mathrm{L} = \mathbf{x}_0 + \left[\left(\frac{\partial \boldsymbol{M}}{\partial \mathbf{x}^T}\right)^T \mathbf{R}^{-1} \left(\frac{\partial \boldsymbol{M}}{\partial \mathbf{x}^T}\right) + \mathbf{P}_0^{-1}\right]^{-1} \left(\frac{\partial \boldsymbol{M}}{\partial \mathbf{x}^T}\right)^T \mathbf{R}^{-1} (\mathbf{y} - \boldsymbol{M}(\mathbf{x}_0)).$$

The uncertainty of the solution $\hat{\mathbf{x}}$ is given in the approximate linearized domain by

$$(8) \qquad \mathbf{P}_\mathrm{L} = \left[\left(\frac{\partial \boldsymbol{M}}{\partial \mathbf{x}^T}\right)^T \mathbf{R}^{-1} \left(\frac{\partial \boldsymbol{M}}{\partial \mathbf{x}^T}\right) + \mathbf{P}_0^{-1}\right]^{-1}$$

(for details see, e.g., [17] or [11]).

Inasmuch that the quadratic regularization term in (2) implies two distinct effects on the solution of the inverse problem, the matrix $\mathbf{P_0}$ has two separate roles in calculating the solution uncertainty. The form of the linearized solution uncertainty (8) highlights the regularization role of the prior covariance in conditioning the matrix inversion. If no prior is specified or if prior uncertainty is large, the uncertainty of the solution is given by matrix inversion of the product of inverse observation uncertainty with model Jacobian and its transpose. This *solution information matrix* is equal to the observation information $\mathbf{R^{-1}}$ projected backward[2] by the adjoint of the forward model Jacobian. For ill-posed inverse problems the solution information matrix is singular and cannot be inverted. In practice, the required information matrix inversion may be ill-conditioned also for formally invertible models. Such information singularity and poor conditioning arise if some components of the solution vector $\mathbf{x}$ are poorly determined by the data or there is high disparity between levels of uncertainty in different components. The standard remedy is to add a small regularization matrix, here $\mathbf{P_0^{-1}}$, to ensure well-posedness of the matrix inversion.

The second role of $\mathbf{P_0}$ stems from the Bayesian estimation framework [5] where, under the Gaussian statistics assumption, the regularization term in (2) has a formal interpretation as the negative logarithm of the prior probability of the controls $\mathbf{x}$. Moreover, the nonregularized assimilated uncertainty in (8) quantifies the error dispersion of maximum likelihood estimates, while the regularized version calculates the *posterior covariance* of the solution. Therefore, the inversion of the regularized information matrix quantifies the prior-to-posterior transformation of uncertainty, which for positive-definite covariance matrices is guaranteed to result in prior-to-posterior *uncertainty reduction*. This is seen by applying the matrix inversion lemma[3] to (8):

$$(9) \qquad \mathbf{P}_\mathrm{L} = \mathbf{P}_0 - \mathbf{P}_0 \left(\frac{\partial \boldsymbol{M}}{\partial \mathbf{x}^T}\right)^T \left[\mathbf{R} + \left(\frac{\partial \boldsymbol{M}}{\partial \mathbf{x}^T}\right) \mathbf{P}_0 \left(\frac{\partial \boldsymbol{M}}{\partial \mathbf{x}^T}\right)^T\right]^{-1} \left(\frac{\partial \boldsymbol{M}}{\partial \mathbf{x}^T}\right) \mathbf{P}_0.$$

---

[2]Compare to forward uncertainty projection by matrix congruence with model Jacobian (17).

[3]The matrix inversion lemma, also known as the Woodbury matrix identity or the Sherman–Morrison–Woodbury formula, is the matrix identity $(A + UCV)^{-1} = A^{-1} - A^{-1}U\left(C^{-1} + VA^{-1}U\right)^{-1} VA^{-1}$.

The second term on the right-hand side is the positive-definite difference between the prior and the posterior covariances.

2.2. The Hessian method for inverse uncertainty propagation. When the forward model $\boldsymbol{M}$ is linear, the cost function (2) is quadratic, as captured by its *Hessian matrix*, i.e., the matrix of second derivatives

$$\mathbf{H}_J(\mathbf{x}) \equiv \frac{\partial^2 J}{\partial \mathbf{x}\, \partial \mathbf{x}^T}. \tag{10}$$

For quadratic cost functions the Hessian matrix is independent of the control vector $\mathbf{x}$ where the derivative is calculated. For a nonlinear model $\boldsymbol{M}$ the cost function is not globally quadratic. However, for sufficiently smooth functions there exists a range around any $\mathbf{x}$ where J is approximately quadratic and its curvature is described locally by the Hessian. In the vicinity of its minimum $\hat{\mathbf{x}}$ the cost function can be approximated, to second order, by the quadratic form

$$J(\mathbf{x}) \approx \tfrac{1}{2}(\mathbf{x} - \hat{\mathbf{x}})^T \, \mathbf{H}_J(\hat{\mathbf{x}})\, (\mathbf{x} - \hat{\mathbf{x}}), \tag{11}$$

where $\mathbf{H}_J(\hat{\mathbf{x}})$ is the Hessian matrix evaluated at $\hat{\mathbf{x}}$ and the gradient of J vanishes (5). The analytical expressions for the Hessians of model-data misfit and of the regularized cost function (2) at $\hat{\mathbf{x}}$ are, respectively,

$$\mathbf{H} = \left(\frac{\partial \boldsymbol{M}}{\partial \mathbf{x}^T}\right)^T \mathbf{R}^{-1} \left(\frac{\partial \boldsymbol{M}}{\partial \mathbf{x}^T}\right) + \left(\frac{\partial^2 \boldsymbol{M}}{\partial \mathbf{x}\, \partial \mathbf{x}^T}\right)^T \mathbf{R}^{-1} \left(\boldsymbol{M}(\hat{\mathbf{x}}) - \mathbf{y}\right), \tag{12}$$

$$\mathbf{H}_J(\hat{\mathbf{x}}) = \mathbf{H} + \mathbf{P}_0^{-1}. \tag{13}$$

The first term in (12) is the back-projected observation uncertainty, equal to the linearized solution information matrix discussed above (8). This first term is commonly referred to as the *linearized Hessian* and was calculated in the previous Hessian-based uncertainty studies [38, 11, 30]. The second term[4] is the product of the second derivative tensor of the multivariate model $\boldsymbol{M}$, i.e., the collection of the Hessian matrices of each of its scalar components [38], with the residual misfits vector normalized by the observation covariance. Known as the nonlinear term [11], it vanishes unless the model $\boldsymbol{M}$ is nonlinear and there remains a residual misfit at the solution point. For the regularized cost function (2) a third term appears in the complete Hessian (13), adding the contribution of the quadratic regularization term and equal to the inverse of the prior covariance matrix. In a Bayesian framework the data misfit Hessian $\mathbf{H}$ (12) is also known as the likelihood Hessian [26], which suggests regarding (13) as the posterior Hessian.

In ocean state estimation problems forward models are nonlinear, data are noisy, and residual misfits do not vanish. Moreover, the regularization introduces bias and

---

[4] Explicitly, in index notation,

$$\left[\left(\frac{\partial^2 \boldsymbol{M}}{\partial \mathbf{x}\, \partial \mathbf{x}^T}\right)^T \mathbf{R}^{-1} \left(\boldsymbol{M}(\hat{\mathbf{x}}) - \mathbf{y}\right)\right]_{i,j} = \sum_k \frac{\partial^2 M_k}{\partial x_i \partial x_j} \sum_l \left(R^{-1}\right)_{kl} \left(M_l(\hat{\mathbf{x}}) - y_l\right),$$

where $\boldsymbol{M}(\mathbf{x}) = (M_1(\mathbf{x})...M_n(\mathbf{x}))^T$ is a multivariate function representing the model, $M_k \colon \mathbf{R}^m \to \mathbf{R}$, and $\frac{\partial^2 M_k}{\partial x_i \partial x_j}$ is the Hessian of $M_k(\mathbf{x})$.

residual misfit even in perfectly invertible problems, as is seen in the optimality condition (5). In nonlinear, large-residual problems, inversion of the complete Hessian

$$(14) \qquad \mathbf{P} \approx \mathbf{H}_J(\hat{\mathbf{x}})^{-1}$$

provides a better estimate of solution uncertainty and fully specifies the so-called *exact confidence regions* [35] within the range of validity of (11).

2.3. Forward uncertainty propagation. When the modeling goal is to estimate target QoI which are not among the estimated model control variables $\mathbf{x}$, then another nonlinear forward model $\boldsymbol{N}$, either prognostic or diagnostic, is used to map the control variables to the desired target QoI vector $\mathbf{z}$,

$$(15) \qquad \mathbf{z} = \boldsymbol{N}(\mathbf{x}) = \boldsymbol{N}\left(\boldsymbol{M}^{-1}(\mathbf{y})\right).$$

That combination of models leads to a two-step state estimation procedure. First apply the inverse model to assimilate observations in the dynamical ocean model to find optimal model controls associated with the optimal state estimate. Then, the optimized model is used to evaluate the QoI, $\mathbf{z}$.

Our purpose is to extend this process in two important ways: (1) to provide formal uncertainty estimates of the inferred optimal controls, i.e., provide a posteriori estimates of their a priori assumed uncertainties by assimilation of observation uncertainty in the model; and (2) to provide formal uncertainty of the QoI through propagation of the a posteriori as well as the prior uncertainties of the controls. The difference between the uncertainty in $\mathbf{z}$ inferred from propagating the prior and the posterior uncertainties on $\mathbf{x}$ provides a measure of the degree to which the observations are able to constrain the models $\boldsymbol{M}$ and $\boldsymbol{N}$ and reduce the prior uncertainties.

For nonlinear models, when the input uncertainty is not too large relative to the range of validity of their tangent linear expansion, the uncertainty can be propagated forward by the tangent linear model. Consider the nonlinear target model $\boldsymbol{N}(\mathbf{x})$ expanded about the estimated model controls $\hat{\mathbf{x}}$:

$$(16) \qquad \boldsymbol{N}(\mathbf{x}) = \boldsymbol{N}(\hat{\mathbf{x}}) + \frac{\partial \boldsymbol{N}}{\partial \mathbf{x}^T}(\mathbf{x} - \hat{\mathbf{x}}) + O(\|\mathbf{x} - \hat{\mathbf{x}}\|^2).$$

Substituting into the definition of the target variable uncertainty and neglecting higher than first order terms results in the linearized forward uncertainty projection by the Jacobian matrix $\partial \boldsymbol{N}/\partial \mathbf{x}^T$:

$$(17) \qquad \mathbf{P}_{\mathbf{zz}} = \left\langle \left(\boldsymbol{N}(\hat{\mathbf{x}}) - \boldsymbol{N}(\mathbf{x})\right)\left(\boldsymbol{N}(\hat{\mathbf{x}}) - \boldsymbol{N}(\mathbf{x})\right)^T \right\rangle = \left(\frac{\partial \boldsymbol{N}}{\partial \mathbf{x}^T}\right)\mathbf{P}_{\mathbf{xx}}\left(\frac{\partial \boldsymbol{N}}{\partial \mathbf{x}^T}\right)^T.$$

Either prior or posterior versions of $\mathbf{P}_{\mathbf{xx}}$ can be propagated, allowing us to compare the prior and posterior uncertainties of the target variable $\mathbf{z}$. Explicitly, restricting to only a scalar target variable z, the prior estimate of the target and its uncertainty are given by

$$(18) \qquad z_0 = N(\mathbf{x}_0), \qquad \Delta z_0^2 = \left(\frac{\partial N}{\partial \mathbf{x}^T}\right)\mathbf{P}_0\left(\frac{\partial N}{\partial \mathbf{x}^T}\right)^T.$$

The posterior target estimate and its uncertainty are obtained by substituting $\hat{\mathbf{x}}$ for $\mathbf{x}_0$ and $\mathbf{P}$ for $\mathbf{P}_0$ in (18). The posterior uncertainty of the target variable is always

smaller than its prior uncertainty due to the positive-definiteness of covariances in (9).

Further insight on forward uncertainty propagation is provided by rewriting the projection (18) explicitly in scalar form:

$$(19) \qquad \Delta z^2 = \sum_i \left(\frac{\partial N}{\partial x_i}\right)^2 \Delta x_i^2 + \sum_i \sum_{j \neq i} \frac{\partial N}{\partial x_i}\frac{\partial N}{\partial x_j}\Delta x_i \Delta x_j \rho_{ij}.$$

Here Δx_i are the standard deviations of the controls, and ρ_{ij} are the correlation coefficients. The first sum includes only the contributions of the diagonal terms of the covariance of controls, representing propagation of the uncorrelated uncertainties. The second term—the double summation of the off-diagonal covariance terms—is the contribution of the correlations in the uncertainties of controls. If the correlation uncertainty terms are zero, then the uncertainty propagation formula (19) reduces to the *addition in quadrature*[5] rule for independent uncertainties [37].

2.4. Proposed UQ scheme. Summarizing the proposed UQ scheme to accompany the inverse-predictive solution (15), the workflow of the algorithm is given by (20). Observation data uncertainty $\mathbf{P_{yy}}$ is assimilated into the model by the inverse uncertainty propagation procedure (14). Then, the assimilated uncertainty $\mathbf{P_{xx}}$ is projected forward onto the desired target variables (17), shown here for a single target QoI scalar z.

$$(20) \qquad \begin{array}{ccccc} \text{Data uncertainty} & \rightarrow & \text{Controls uncertainty} & \rightarrow & \text{Target uncertainty} \\ \mathbf{P_{yy}} & \rightarrow & \mathbf{P_{xx}} & \rightarrow & \Delta z^2 \end{array}$$

The implementation of this scheme is straightforward for invertible models by inversion of the Hessian of the nonregularized version of the data-model misfit function J and its forward projection (18), here given by the target gradient[6] vector $\mathbf{g}$.

$$(21) \qquad \mathbf{R} \;\longrightarrow\; \mathbf{P_{xx}} = (\mathbf{H})^{-1} \;\longrightarrow\; \Delta z^2 = \mathbf{g}^T \mathbf{P_{xx}}\mathbf{g}$$

$$(22) \qquad \begin{array}{c} \mathbf{P}_0 \qquad\qquad\quad \longrightarrow \quad \Delta z^2 = \mathbf{g}^T \mathbf{P}_0 \mathbf{g} \\ \downarrow \\ \mathbf{R} \;\longrightarrow\; \mathbf{P} = \left(\mathbf{H} + \mathbf{P}_0^{-1}\right)^{-1} \;\longrightarrow\; \Delta z^2 = \mathbf{g}^T \mathbf{P}\mathbf{g} \end{array}$$

This uncertainty assimilation procedure can be performed independently of a priori assumed uncertainty of controls in which case it is regarded as "pure" propagation of observations' uncertainty $\mathbf{R} = \mathbf{P_{yy}}$ to prior-independent uncertainty $\mathbf{P_{xx}}$ (21). For the ill-posed inverse problems in oceanographic applications such prior-independent uncertainty assimilation schemes lead to unbounded estimation uncertainties because not all combinations of model controls are constrained by the observations. In these cases we resort instead to the prior-to-posterior uncertainty reduction scheme (22), which blends the assimilated observations' uncertainty $\mathbf{P_{yy}}$ with prior uncertainty $\mathbf{P}_0$ of the controls to produce the posterior controls uncertainty $\mathbf{P}$. This melding [20] can be computed by inverting the Hessian of the regularized data-model misfit function (2) or, alternatively, with the reduced-order method (29) developed below. Both

---

[5] Not to be confused with "quadrature rule" for numerical integration.
[6] In vector derivatives notation, $\mathbf{g} = \nabla N(\mathbf{x}) = \partial N/\partial \mathbf{x} = (\partial N/\partial \mathbf{x}^T)^T$.
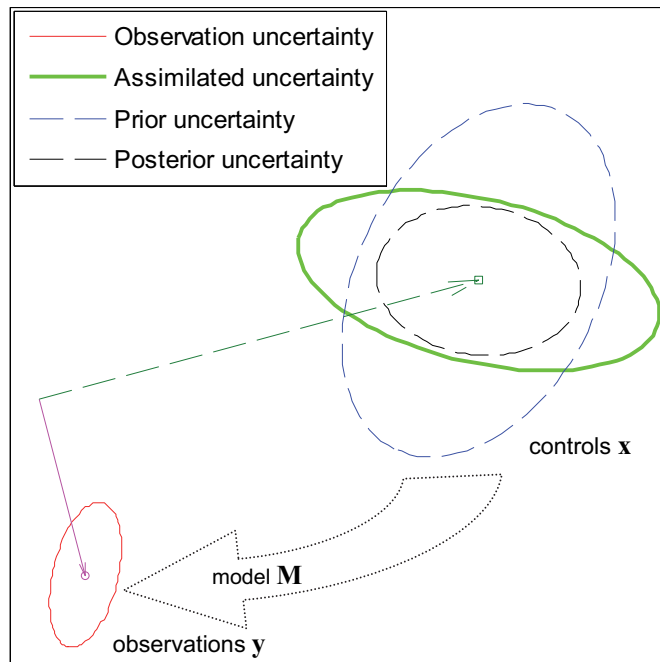
FIG. 1. *Visualization of inverse uncertainty propagation in two dimensions for a linear invertible forward model. Pure (green) assimilated uncertainty is compared to prior (blue) and posterior (black) uncertainties of the controls.*
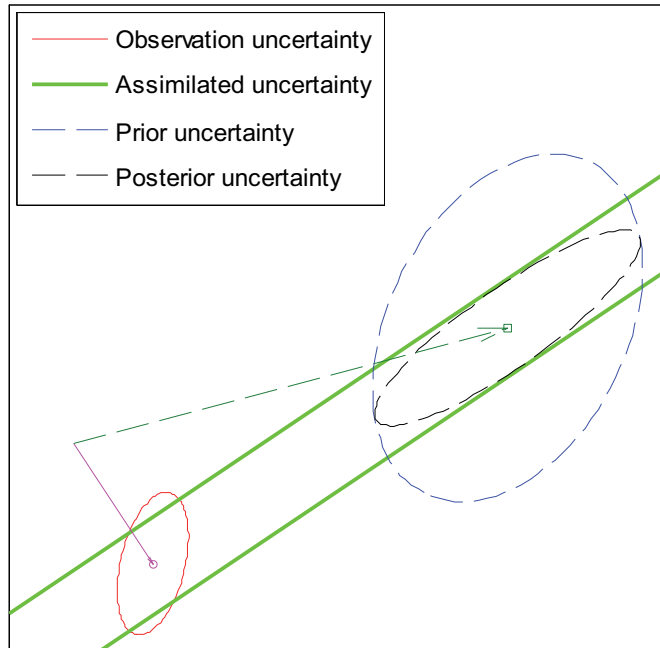

FIG. 2. *Same as Figure 1, but for ill-posed noninvertible linear forward model.*

the prior and the posterior uncertainties of controls can be forward projected onto the QoI space, where they can be compared quantifying the prior-to-posterior target uncertainty reduction.

Consider an invertible linear model in two dimensions, mapping controls $\mathbf{x}$ to observations $\mathbf{y}$, both shown as two-dimensional vectors in Figure 1. Uncertainty covariances are shown by constant-χ^2 ellipses corresponding to $\mathbf{x}^T \mathbf{P}^{-1} \mathbf{x} = \chi^2$. The uncertainty of observations (red) is mapped backward to model controls, according to (21), resulting in the assimilated uncertainty (green). It is contrasted with the other control uncertainties, calculated according to scheme (22) by combining the assimilated uncertainty with the prior (blue) to produce the posterior (black) uncertainty. Both schemes are illustrated in Figure 2 for a noninvertible linear model, when the inversion of the Hessian in (21) is ill-posed and only the prior-to-posterior reduction scheme (22) can be fully implemented. In this case observations cannot constrain certain linear combinations of controls, leaving infinite uncertainty in the corresponding directions. These directions, known as the nullspace of the Hessian, are illustrated by the unbounded elongated ellipse of the assimilated uncertainty. In directions orthogonal to the nullspace the assimilated uncertainty is finite, and controls are constrained by observations. This data-supported subdomain of controls is known as the range space of the Hessian.

3. Numerical machinery.

3.1. Algorithmic differentiation.
The developed UQ methodology relies heavily on derivatives of the nonlinear forward model, computed by stand-alone derivative codes. The *Jacobian* (tangent linear), *adjoint*, and *Hessian* codes of the MITgcm were generated by means of algorithmic (or automatic) differentiation (AD) using the source-to-source transformation tool Transformation of Algorithms in Fortran (TAF) [7]. The tool parses the original code, analyzes it as a sequence of elementary operations (23) for which elementary derivatives are known analytically, and combines the elementary Jacobian matrices according to the chain rule to produce the derivative code. Formally,

$$(23) \qquad \mathbf{\Psi}(\mathbf{x}_0) = \mathbf{\Psi}_n \circ \mathbf{\Psi}_{n-1} \circ \cdots \circ \mathbf{\Psi}_1 \circ \mathbf{x}_0,$$

$$(24) \qquad d\mathbf{x}_n = \frac{\partial \mathbf{\Psi}}{\partial \mathbf{x}_0^T} d\mathbf{x}_0 = \left(\frac{\partial \mathbf{\Psi}_n}{\partial \mathbf{x}_{n-1}^T} \frac{\partial \mathbf{\Psi}_{n-1}}{\partial \mathbf{x}_{n-2}^T} \cdots \frac{\partial \mathbf{\Psi}_1}{\partial \mathbf{x}_0^T} \right) d\mathbf{x}_0.$$

The order of multiplication of local Jacobians in (24) is the same as in the original code, resulting in "forward mode" derivative code known as the *tangent linear model*. Due to the associativity of matrix multiplication (25), that order can also be reversed, requiring transposition of the multiplied Jacobians and producing the so-called *reverse mode* derivative code, or *adjoint model* (see, e.g., [6, 23, 13]):

$$\frac{\partial \mathbf{\Psi}}{\partial \mathbf{x}_0^T} = \left(\frac{\partial \mathbf{\Psi}_n}{\partial \mathbf{x}_{n-1}^T} \left(\frac{\partial \mathbf{\Psi}_{n-1}}{\partial \mathbf{x}_{n-2}^T} \cdots \left(\frac{\partial \mathbf{\Psi}_1}{\partial \mathbf{x}_0^T} \right) \right) \right)$$

$$= \left(\left(\left(\frac{\partial \mathbf{\Psi}_n}{\partial \mathbf{x}_{n-1}^T} \right) \frac{\partial \mathbf{\Psi}_{n-1}}{\partial \mathbf{x}_{n-2}^T} \right) \cdots \frac{\partial \mathbf{\Psi}_1}{\partial \mathbf{x}_0^T} \right) = \left(\frac{\partial \mathbf{\Psi}_1}{\partial \mathbf{x}_0^T}^T \cdots \frac{\partial \mathbf{\Psi}_{n-1}}{\partial \mathbf{x}_{n-2}^T}^T \frac{\partial \mathbf{\Psi}_n}{\partial \mathbf{x}_{n-1}^T}^T \right)^T.$$

(25)

For multivariate functions with the number of independent variables much larger than the number of dependent variables, $\dim(\mathbf{\Psi}) \ll \dim(\mathbf{x})$, the reverse order multiplication of adjoint Jacobians results in significant computational savings. The adjoint

procedure is particularly efficient for scalar functions, i.e., $\dim(\boldsymbol{\Psi}) = 1$, reducing matrix multiplication to a sequence of matrix-vector products.

Details of application of AD machinery to complex codes such as the MITgcm are rather involved [12], requiring in practice the combination of manual adjoint programming with AD-generated code. Algorithmically, the central novelty of this study is the successful generation of the second derivative (Hessian) code of the MITgcm. An outline of the technical work is deferred to the appendix.

3.2. Hessian inversion. Calculating the assimilated data uncertainty covariance or the posterior uncertainty covariance requires the inversion of the Hessian (12) or (13), respectively. Below we describe the numerical procedure for both calculations.

3.2.1. Prior-independent assimilated data uncertainty. For small to medium dimensional problems (up to $O(10^3 - 10^4)$), the nondegenerate Hessian matrix can be inverted by dense matrix algebra methods. Our implementation is based on LU factorization using a sequence of LAPACK library procedures DGETRF and DGETRI. This direct approach cannot be applied when the Hessian matrix is singular or ill-conditioned. Due to limitations of finite precision numerics, machine round-off errors make dense algebra calculations of ill-conditioned inverse useless even if the inverse formally exists. Nonetheless, the inverse can be constructed using the pseudo-inverse matrix form (27) based on the orthogonal spectral expansion (26) that always exists for symmetric matrices:

$$(26) \qquad \mathbf{H} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T = \sum_i \lambda_i \mathbf{v}_i \mathbf{v}_i^T,$$

$$(27) \qquad \mathbf{H}^+ = \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{V}^T = \sum_i \frac{1}{\lambda_i} \mathbf{v}_i \mathbf{v}_i^T.$$

Here λ_i and $\mathbf{v}_i$ are Hessian eigenvalues and eigenvectors, $\boldsymbol{\Lambda}$ is the diagonal eigenvalue matrix, and $\mathbf{V}$ is the orthogonal matrix of eigenvectors (its columns). Each outer product of eigenvectors, $\mathbf{v}_i \mathbf{v}_i^T$, is a rank-1 matrix. The pseudoinverse is the linear combination of these rank-1 matrices weighted by the reciprocals of the Hessian eigenvalues (see, e.g., [38]). For singular or nearly singular matrices it is possible to construct a reduced-rank pseudoinverse by omitting zero or small eigenvalues of the Hessian and corresponding eigenvectors from (27), which is equivalent to just setting the corresponding diagonal terms of $\boldsymbol{\Lambda}^{-1}$ to zero. These omitted eigenvectors correspond to control variable combinations that are poorly constrained by the data or lie in the nullspace of the problem, their eigenvalues representing the small or zero information gained. Since the reciprocals of Hessian eigenvalues represent the uncertainty variances of the constrained control variable combinations, the uncertainty of the omitted eigenvectors is very large. The interpretation of the reduced-rank pseudoinverse then is omitting the nullspace and contracting the assimilated uncertainty covariance only to the data-supported subdomain in the control parameters space.

Adding prior information to the UQ problem has the effect of regularizing the solution uncertainty, i.e., suppressing the nullspace without omitting it as in the case above. For small problems, regularization is advantageous for technical reasons, allowing inversion of the Hessian. The unfavorable consequences of regularization, however, include introduction of bias and melding of the prior covariance structure

with the assimilated uncertainty signal. In cases when selection of the prior is numerically motivated, its arbitrariness from a physics point of view may be detrimental to the dynamically consistent estimation process and mask the physics of the system. Moreover, for large inverse problems, the regularization is counterproductive also numerically, as neither direct inversion nor full eigendecomposition are feasible for the resulting regularized Hessian, while the sparsity of the original Hessian allows compression of its inversion computation. Therefore, the pure prior-independent Hessian inversion is preferred to regularization and is sufficient for analyzing the uncertainty in the controls domain. Nonetheless, for forward uncertainty projection to the target quantity space in our inverse-predictive state estimation framework, the partial Hessian inversion is generally not applicable. The reduced-rank Hessian pseudoinverse cannot be propagated forward with Jacobian products (17) because the omitted unconstrained control parameter combinations belonging to the nullspace may project forward unbounded uncertainty, rendering the target variable completely uncertain. Therefore, for compatibility with forward uncertainty propagation an alternative method for large Hessian inversion, involving regularization, is developed below.

3.2.2. Posterior uncertainty reduction. When prior uncertainty is specified for large-scale ill-posed inverse problems, regularizing the posterior and suppressing its spectral decay, inversion of the regularized Hessian (13) is not feasible computationally either by dense matrix algebra or in the pseudoinverse form. Nonetheless, the compactness of the original low-rank Hessian of misfits (12) can be utilized to compress the inversion. Our method is motivated by the traditional Gauss–Markov solution form (9) but is based on application of the matrix inversion lemma to the Hessian diagonalization product $\mathbf{H} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathbf{T}}$ instead of the solution information matrix product in (8). Rewriting (13), the posterior covariance is computed by

$$(28) \qquad \mathbf{P}^{-1} = \mathbf{P}_0^{-1} + \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T,$$

$$(29) \qquad \mathbf{P} = \mathbf{P}_0 - \mathbf{P}_0\mathbf{V}\left(\mathbf{\Lambda}^{-1} + \mathbf{V}^T\mathbf{P}_0\mathbf{V}\right)^{-1}\mathbf{V}^T\mathbf{P}_0.$$

It significantly reduces the dimensionality of computation for low-rank Hessians, reducing the number of columns in the eigenvector matrix $\mathbf{V}$ and the size of eigenvalue matrix $\mathbf{\Lambda}$ to the number of the nonzero Hessian eigenvalues. This lossless computational compression is a major advantage of the proposed approach: while the traditional solution (9) reduces the required matrix inversion computations from the dimensionality of the control space to that of the observations space, (29) further reduces the inversion computations to dimensionality of only the independent observations which add nonzero information to uncertainty constraints in the control space, equal to the rank of the Hessian

$$(30) \qquad \dim(\mathbf{P}) \geq \dim(\mathbf{R}) \geq \dim(\mathbf{\Lambda}) = \text{rank}(\mathbf{H}).$$

The inversion of the transformed uncertainty matrix $\mathbf{\Lambda}^{-1} + \mathbf{V}^T\mathbf{P}_0\mathbf{V}$ is implemented with dense matrix algebra, and the remaining calculations, which involve only matrix multiplications, can be implemented with matrix-free techniques. The downside of this approach, as opposed to the reduced-rank pseudoinverse above, is that the results depend on the structure of the prior covariance matrix.

Rearranging (29) allows for matrix-free implementation for large prior matrices that are available only through matrix-vector products:

$$\mathbf{P} = \left[\mathbf{I} - \mathbf{P}_0\mathbf{V}\left(\mathbf{\Lambda}^{-1}+\mathbf{V}^T\mathbf{P}_0\mathbf{V}\right)^{-1}\mathbf{V}^T\right]\mathbf{P}_0 = \left[\mathbf{P}_0\left[\mathbf{I} - \mathbf{P}_0\mathbf{V}\left(\mathbf{\Lambda}^{-1}+\mathbf{V}^T\mathbf{P}_0\mathbf{V}\right)^{-1}\mathbf{V}^T\right]^T\right]^T.$$
(31)

A convenient interpretation is obtained by rearranging (31),

$$(32) \qquad\qquad \mathbf{P}\mathbf{P}_0^{-1} = \mathbf{I} - \mathbf{P}_0\mathbf{V}\left(\mathbf{\Lambda}^{-1}+\mathbf{V}^T\mathbf{P}_0\mathbf{V}\right)^{-1}\mathbf{V}^T,$$

with the right-hand side of (32) seen as an uncertainty reduction operator. For diagonal $\mathbf{P}_0$ the diagonal of (32) is the rate of uncertainty variance reduction. Another measure of uncertainty reduction, valid for any $\mathbf{P}_0$, is comparison of prior and posterior diagonals,

$$(33) \qquad\qquad 1 - \sqrt{\frac{diag(\mathbf{P})_i}{diag(\mathbf{P}_0)_i}},$$

which is the same as the *posterior relative error reduction* in [21].

3.3. Forward uncertainty propagation. In the developed UQ framework two options are possible for forward uncertainty propagation. One can choose to restrict the analysis only to the subdomain supported by the observations, where the "posterior" uncertainty covariance is given by (27). The forward uncertainty propagation is then given by

$$(34) \qquad\qquad \mathbf{T}\mathbf{H}^+\mathbf{T}^T = \mathbf{T}\mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{V}^T\mathbf{T}^T = \sum_i \frac{1}{\lambda_i}\mathbf{T}\mathbf{v}_i\mathbf{v}_i^T\mathbf{T}^T.$$

Here, for clarity, we shorten the notation by defining the target transformation $\mathbf{T}$ equal to the Jacobian matrix $\partial\boldsymbol{N}/\partial\mathbf{x}^T$. Note that (34) does not represent the formal error covariance of the target variable when the data-supported subdomain of controls, while excluding the nullspace of the observation operator, also excludes components of the target transformation row space, that is, if the excluded combinations of control variables, which are not constrained by the observations, are not also in the nullspace of the target transformation. In that case (34) would map their unconstrained uncertainty to the target variable range, and the reduced-order sum would not include all the uncertainty.

Alternatively, for the regularized posterior uncertainty of controls (29), the posterior target uncertainty is calculated by

$$(35) \qquad \mathbf{T}\mathbf{P}\mathbf{T}^T = \mathbf{T}\mathbf{P}_0\mathbf{T}^T - \mathbf{T}\mathbf{P}_0\mathbf{V}\left(\mathbf{\Lambda}^{-1}+\mathbf{V}^T\mathbf{P}_0\mathbf{V}\right)^{-1}\mathbf{V}^T\mathbf{P}_0\mathbf{T}^T$$

implemented in a matrix-free form by

$$(36) \qquad \mathbf{T}\mathbf{P}\mathbf{T}^T = \mathbf{T}\left[\mathbf{T}\mathbf{P}_0\right]^T - \mathbf{T}\mathbf{P}_0\mathbf{V}\left[\mathbf{T}\mathbf{P}_0\mathbf{V}\left(\mathbf{\Lambda}^{-1}+\mathbf{V}^T\mathbf{P}_0\mathbf{V}\right)^{-T}\right]^T.$$

Remember that the prior target uncertainty is given by $\mathbf{T}\mathbf{P}_0\mathbf{T}^T$, and only the second right-hand-side term needs to be calculated for target uncertainty reduction. If the target model is not linear but (17) applies, the Jacobian derivative code needs to be implemented. For a single scalar target variable (18) the implementation is straightforward with the standard adjoint code of the target model, as the target transformation $\mathbf{T}$ is replaced by the single row vector $\partial N/\partial\mathbf{x}^T = \mathbf{g}^T$, i.e., the transposed gradient of the target variable.

4. Application to uncertainty of Drake Passage volume transport.

4.1. Drake Passage transport in a global barotropic model. The developed UQ method is implemented in global barotropic configuration of MITgcm and applied to analyze the uncertainty in estimating the vertically integrated volume transport through the Drake Passage. The model numerically solves the nonlinear barotropic momentum and continuity equations (37). Here u, v are vertically integrated (barotropic) velocities, η is the sea surface elevation, H is the ocean depth, f and g are Coriolis and gravity parameters, and ρ is the density. The flow is forced by surface wind stress τ^x, τ^y and balanced by the bottom friction, parameterized with a linear bottom drag coefficient r.

$$
(37) \qquad
\begin{aligned}
\frac{\partial u}{\partial t} + uu_x + vu_y &= fv - g\eta_x + \frac{\tau^x}{\rho H} - \frac{ru}{H}, \\
\frac{\partial v}{\partial t} + uv_x + vv_y &= -fu - g\eta_y + \frac{\tau^y}{\rho H} - \frac{rv}{H}, \\
\frac{\partial \eta}{\partial t} &= -\left(u(H+\eta)\right)_x - \left(v(H+\eta)\right)_y.
\end{aligned}
$$

To leading order the dynamics of this system represent the adjustment to a geostrophic forcing-friction equilibrium. The transient adjustment is governed by barotropic mechanisms such as inertia-gravity (Poincaré) and Kelvin waves [3]. The geostrophic steady state flow is parallel to sea surface elevation contours, forced by wind stress, and balanced by bottom friction. An idealized barotropic circumpolar current is modeled by forcing the Southern Ocean by a homogeneous zonal jet of steady westerly winds (Figure 3(a)). The model is spun up from rest to the geostrophic steady state (Figure 3(b)). In the adjustment process the flow is accelerated eastward by the wind and is deflected equatorward by the Coriolis force. This leads to a transient northward transport redistributing the water mass across the jet and generating a northward
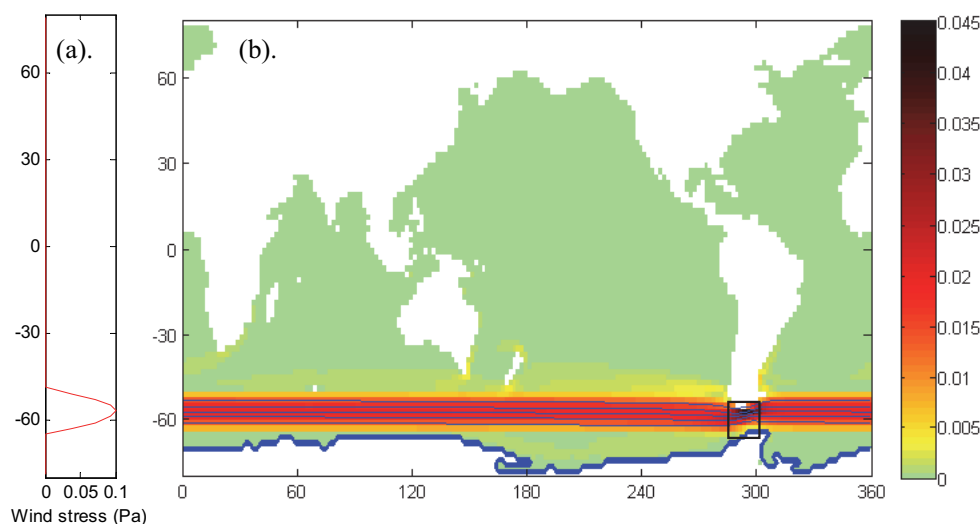

FIG. 3. *Barotropic MITgcm configuration of ACC forced by idealized zonal wind stress profile (a). The steady state circumpolar circulation (b) is visualized by the magnitude of barotropic flow (color shades, m/s) overlaid with sea surface height contours (blue, every 5 cm). The area of assimilated sea surface height observations at the Drake Passage is shown with the black rectangle.*

TABLE 1
Schematic arrangement in the six-panel figures of the six two-dimensional control variables of the composite control vector.

Boundary conditions	Initial conditions
zonal wind stress τ^x	zonal flow $u_{t=0}$
meridional wind stress τ^y	meridional flow $v_{t=0}$
bottom drag coefficient r	surface elevation $\eta_{t=0}$

gradient of sea surface elevation, which balances the Coriolis force in the equilibrium. The resulting circumpolar flow and sea surface height contours are aligned with the wind forcing, except in the narrowing of the Drake Passage, where the contours are tighter and the flow is stronger, especially on the northern side of the Passage. The flow gives rise to a volume transport of roughly 100 Sv (1 Sverdrup = 10^6 m^3/s).

The dynamical coupling between barotropic transport and sea surface elevation patterns is used to estimate the strength of the circumpolar circulation in a global ocean model. While direct measurements of subsurface oceanic flow are feasible only at very low resolution, the surface signature of the flow allows its indirect inference utilizing detailed ocean surface observations [43]. The circumpolar transport is estimated at the Drake Passage since it is the narrowest natural constriction on the path of the circulation, where its spatial extent is more compact and the dynamical signals are stronger. Surface elevation observations over a rectangular area at the Drake Passage (Figure 3(b)) are assimilated in the barotropic model to constrain the estimated transport. Assimilation of the observation uncertainty quantifies the scope of this constraint. Uncertainties of the observations, the model, and the estimated transport are linked by the barotropic dynamics. The mechanisms of these linkages are quantified and analyzed with the developed UQ machinery below.
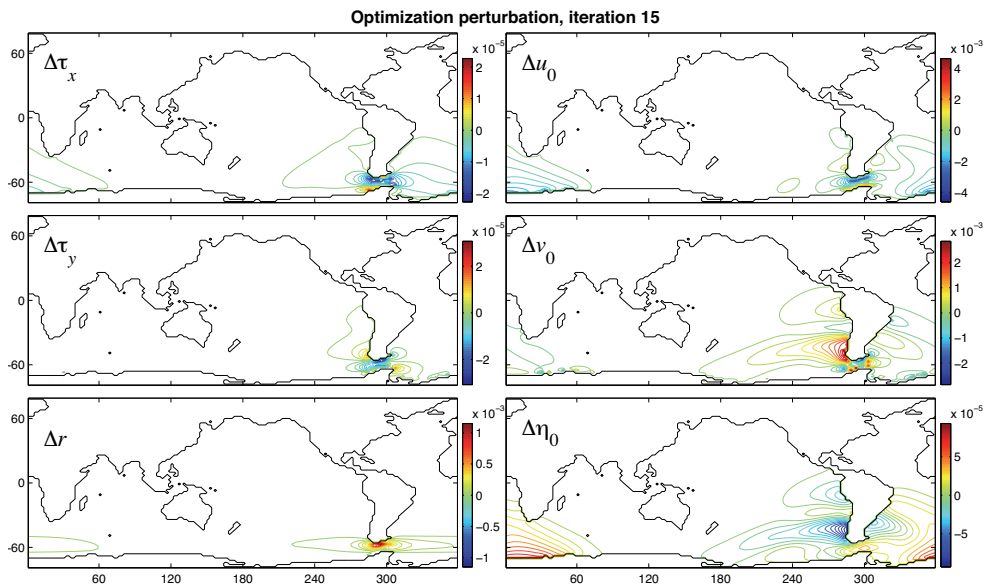

FIG. 4. *Adjusted control variables as a result of the optimization after* 15 *iterations.*
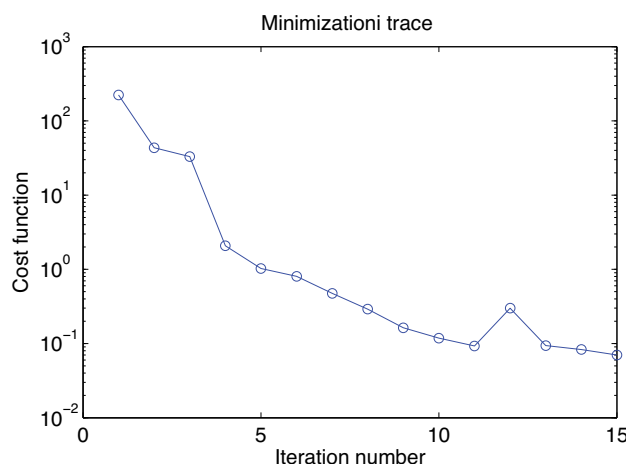
FIG. 5. *Cost function reduction as a result of gradient-based optimization.*

4.2. Solving the least-squares inverse problem. The idealized ocean state estimation problem is solved by minimizing the nonlinear misfit function (2) between surface elevation fields of a reference steady state flow and the synthetic observations obtained from a perturbed steady state simulation. The resulting nonlinear optimization problem is solved with a gradient-based method iteratively adjusting the control variables via a quasi-Newton variable storage algorithm [8]. The adjoint model is used for efficiently computing the gradient of the cost function with respect to the uncertain control variables. The controls are six two-dimensional fields of model initial and boundary conditions (Table 1), each of 80 by 180 scalar elements. The optimal adjustments of these controls are shown in Figure 4, where the six fields are arranged corresponding to the table.

The least-squares solution is shown for a time-dependent problem, where surface elevation observations are assimilated six hours after the model initialization. The spatial patterns of the optimal controls show the required adjustments to the initial and boundary condition fields to minimize the misfit between the reference and assimilated surface elevation six hours into the model run. The assimilated ACC fields are from a weaker flow simulation produced with weaker wind forcing than in the reference simulation. Correspondingly, the resulting control adjustments exhibit negative zonal wind forcing and positive bottom drag. The patterns of the initial conditions' adjustments correspond to transient oscillations of barotropic waves propagating along continents around the Drake Passage.

After 15 iterations of optimization, the misfit cost function is reduced by three orders of magnitude (Figure 5). Since the optimum fit is imperfect, a nonvanishing contribution from the nonlinear Hessian term (12) driven by the residual model-data misfit will be retained, highlighting the need to resolve the complete nonlinear Hessian.

4.3. Forward uncertainty propagation. Forward uncertainty propagation via (18) maps uncertainties of model controls to those of the target variables. The controls are six two-dimensional fields of model initial and boundary conditions, estimated in the least-squares solution (Figure 4). The target is the integrated zonal (west-to-east) volume transport through the Drake Passage. Here we analyze the uncertainty of the steady state transport in simulations initialized with steady state

Transport uncertainty, quadrature (diagonal) contributions, 6 hours.

FIG. 6. *Quadrature contributions (in m^3/s) to uncertainty of Drake Passage transport in forward propagation of prior uncertainty of controls. The six panels correspond to the initial and boundary condition control fields arranged according to Table* 1. *Results shown for six hour forward uncertainty propagation simulation. Sources of Drake Passage transport uncertainty include local effects of zonal wind forcing and bottom friction, upstream meridional wind forcing, and teleconnected effects of initial conditions carried by the adjoint Kelvin waves.*

initial conditions. Note that although the forward model is in steady state at all times, perturbations about this equilibrium are not steady and result in evolving uncertainties due to barotropic adjustment processes. Detailed time-resolving analysis of uncertainty evolution is beyond the scope of this paper and will be reported elsewhere (see also [17]). According to (22), either the prior or the posterior uncertainty of controls can be propagated forward. In this subsection we discuss the propagation of the prior uncertainty, which is assumed to be uncorrelated between different control vector components and homogeneous for each of the six different fields of the control vector. In other words, the prior covariance matrix is diagonal, and the diagonal is given by the squares of the standard errors [$\Delta\tau^x$, $\Delta\tau^y$, Δr, $\Delta u_{t=0}$, $\Delta v_{t=0}$, $\Delta\eta_{t=0}$] = [0.1 Pa, 0.1 Pa, 5.0e-3 m/s, 0.01 m/s, 0.01 m/s, 0.1 m] corresponding to the fields in Table 1.

The prior uncertainty of controls is propagated forward by (18) using the adjoint MITgcm model. For the diagonal prior covariance the scalar form of the forward projection (19) reduces to the quadrature sum

$$(38) \qquad \Delta z^2 = \sum_i \left(\frac{\partial N}{\partial x_i}\right)^2 \Delta x_i^2.$$

The square roots of the right-hand-side sum terms are the quadrature contributions to transport uncertainty due to uncorrelated control uncertainties, shown in Figure 6 for each scalar component of the control vector. Analysis of the patterns and their evolution dynamics allows identification and understanding of the physical mechanisms of forward uncertainty propagation. The displayed patterns correspond to adjoint

Kelvin waves propagating in the opposite direction from actual Kelvin waves along the Antarctic and South American coasts. These uncertainty waves are radiated from the Drake Passage area and propagate backward, carrying the information on the sources of transport uncertainty of the forward evolving circumpolar circulation. Different adjoint mechanisms are active on different time scales. A detailed discussion of adjoint uncertainty dynamics is beyond the scope of this paper and is found in [17]. Adding up all the quadrature uncertainty contributions results in the prior estimate of uncertainty of transport. For the time shown in Figure 6 the standard error is 4.7 Sv. The magnitude of the simulated zonal transport is 100 Sv; therefore its uncertainty is $\pm 4.7\%$.

4.4. Inverse uncertainty propagation. Inverse uncertainty propagation assimilates observation uncertainty into model controls via the Hessian of model-observation misfit. The assimilated observations are sea surface heights over a rectangular area at the Drake Passage (Figure 3(b)), specified with uncorrelated and homogeneous (homoscedastic) uncertainties with standard error $\Delta \eta = 0.01$ m. The developed UQ machinery resolves the dynamical variations of the patterns of assimilated uncertainty for different inverse propagation times, illustrated below for six hour inverse propagation.

Since the assimilated observations are not sufficient to fully constrain the control fields, the Hessian of misfits is not invertible, the covariance of the assimilated uncertainty cannot be explicitly computed, and the prior-independent UQ scheme (21) is ill-posed. Nonetheless, the assimilated uncertainty can be analyzed by extracting the leading eigenvectors of the Hessian. These eigenvectors represent the principal uncertainty patterns of model controls constrained by the assimilated observations' uncertainty. The standard errors of these patterns are given by the inverse square roots
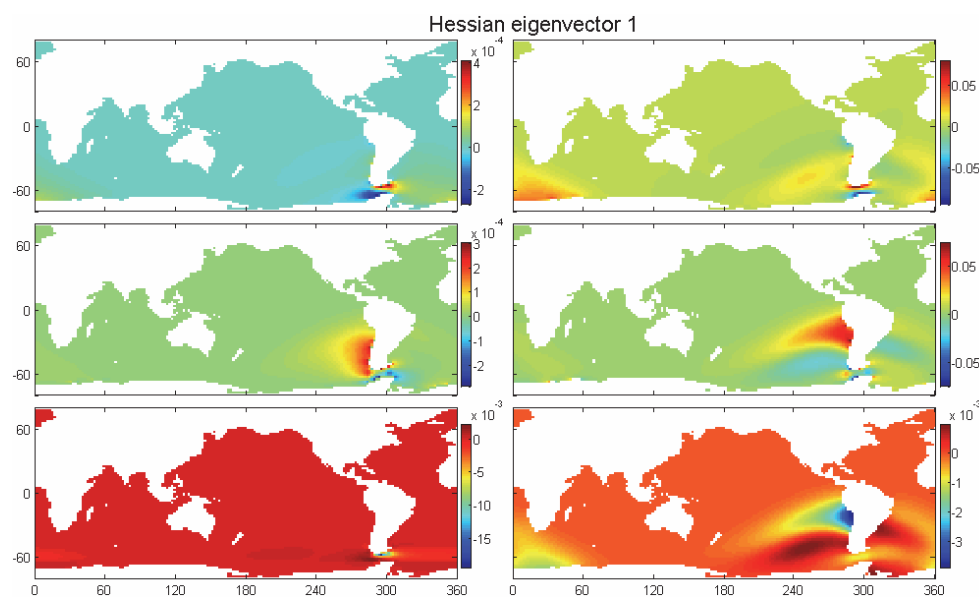

FIG. 7. *The leading eigenvector of the Hessian for assimilated sea surface height observations at Drake Passage, shown for the six hour inverse uncertainty propagation simulation. The configuration of control vector fields is shown in Table* 1.
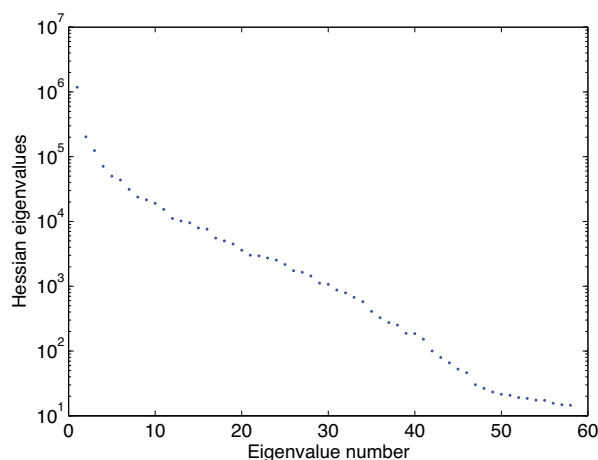
Fig. 8. *The Hessian eigenvalue spectrum for assimilated sea surface height observations at Drake Passage, shown for the six hour inverse uncertainty propagation simulation.*

of the corresponding eigenvalues (26). The leading eigenvector is shown in Figure 7; its patterns reflect local and teleconnected features most constrained by assimilation of uncertainty of observed sea surface height. The strongest signal is that of the adjoint Kelvin waves propagating in the opposite direction from actual Kelvin waves along continental coasts in the initial conditions fields. Dipoles are seen across the Drake Passage in zonal forcing and the initial zonal flow uncertainties. The striking similarity of the Hessian patterns and the quadrature uncertainty fields (Figure 6) is explained by noting that both are constructed by Jacobian fields, although the former by the outer product of gradients of surface elevation (12) and the latter by the inner product of the gradient of transport (18). Gradients of various model outputs are also known as the sensitivity fields (e.g., [14]). Thus, our analysis explains different uncertainty concepts in terms of sensitivities, formalizing the common intuition that uncertainty and sensitivity are closely related. That is, the controls that transport is very sensitive to are those contributing most to its uncertainty, and the controls that modeled observations are most sensitive to are the dominant features of the principal patterns constrained by assimilation of these observations.

The eigenvalues of the Hessian represent the information gained by assimilating uncertain observations. The spectrum of the eigenvalues (Figure 8) decays exponentially, meaning that the information gained is concentrated in few orthogonal modes of uncertainty and justifies the reduced-rank Hessian algorithm (30). These modes are constrained by the observations and correspond to the range space of the estimation problem. The rest of the orthogonal modes are not constrained and correspond to the nullspace of the estimation. Formally, if no prior constraint of uncertainty of the nullspace modes is assumed, that uncertainty remains unbounded in the state estimation process and no finite error covariance matrix exists. An alternative approach, suppressing the null space and ensuring existence of finite covariance, is illustrated below.

4.5. Posterior uncertainty reduction. Inverse uncertainty propagation is more straightforward to analyze if prior uncertainty information is available. In practice, a plausible assumption is made about controls' prior uncertainties in order to regularize the estimation problem by suppressing its nullspace. The resulting posterior

uncertainty is smaller than (or equal to) the prior for any error vector combination, and it can always be expressed by the posterior covariance matrix. It is a challenge to fully analyze and even display the complete structure of the posterior covariance matrix. While the prior matrix is assumed to be diagonal (see section 4.2) and can be visualized by a single control space vector, the posterior is not; see (19). In principle, the covariance may be visualized column by column, but it is unfeasible for the large dimensionality of the ocean state estimation problem. However, leaving the correlation of uncertainties aside, the marginal uncertainties of controls are fully quantified by the diagonal. The marginal prior-to-posterior uncertainty reduction (39) in control space is shown in Figure 9.

$$(39) \qquad\qquad \frac{\Delta_{prior} - \Delta_{post}}{\Delta_{prior}} \cdot 100.$$

The results are very insightful. The uncertainty is reduced in all control fields—not only in the field where observations were assimilated. This indicates the dynamical cross-coupling between the different control fields and across the control vector. Moreover, the reduction is not limited to the rectangular area of the assimilated observations but extends beyond it, indicating the spatial cross-coupling within each separate control field. The detailed patterns of uncertainty reduction spread convey the physical mechanisms of uncertainty propagation through the model. In particular, note that for the observed field (the surface elevation), relatively negligible uncertainty reduction is achieved in the geographical area of data assimilation, and most of the effect propagates away carried by the adjoint uncertainty waves radiating from the Drake Passage. For the other control fields, in contrast, most of uncertainty reduction occurs in the Drake Passage region. Weak reduction is seen for zonal flow carried
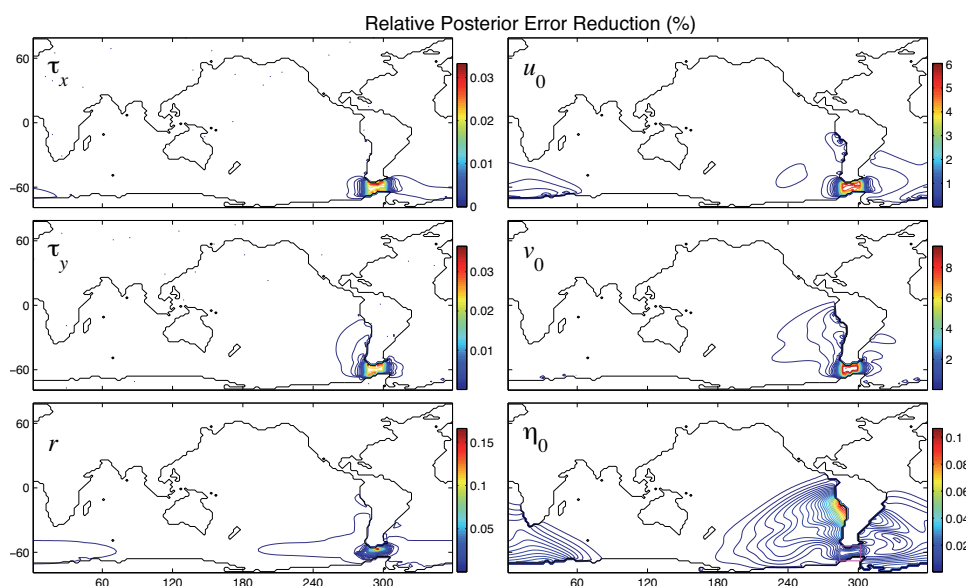

Fig. 9. *Relative posterior uncertainty reduction (shown in %) of six control fields for assimilated sea surface observations at Drake Passage (shown with magenta rectangle on right bottom panel) for six hour inverse uncertainty propagation simulation. Control vector fields configuration is according to Table 1. Only the diagonal covariance matrix terms can be shown in a single six panel figure; a display of the complete posterior uncertainty covariance would require $O(10^5)$ figures like this.*

by the adjoint Kelvin waves along the Antarctic coast and for meridional flow along the South American continent. Uncertainty reduction of the forcing fields is mostly concentrated in the data assimilation area, although the southward wind stress uncertainty is slightly reduced upstream of the Drake Passage as wind forces the flow around the continent and the bottom drag uncertainty reduction follows the circulation zonally around the globe.

Combining inverse and forward uncertainty propagation to project the posterior uncertainty of controls to the target variable space and comparing to the prior forward projection (section 4.3) quantifies the posterior uncertainty reduction of Drake Passage transport. After six hours from the steady state initialization, the posterior uncertainty of transport is 2.5 Sv. Compared to the prior transport uncertainty of 4.7 Sv, it is a relative uncertainty reduction of 47%. This number evolves dynamically as posterior uncertainties of controls adjust to the steady state, with both the forward and the inverse uncertainty propagations explicitly time dependent.

5. Summary and discussion. This paper develops and illustrates the implementation of a formal UQ methodology for dynamically consistent large-scale ocean state estimation in the context of a nonlinear time-dependent state-of-the-art global ocean general circulation model. Uncertainty is quantified by covariance matrices, which are propagated between observation, control, and target variable domains by first and second derivative codes. These codes were constructed from the MITgcm via algorithmic differentiation (AD). The main technical novelty of this work is successful generation of the complete nonlinear Hessian code of the MITgcm. Its fidelity was verified via symmetry, finite difference, and Taylor remainder tests. The ability to utilize the complete nonlinear Hessian information is necessary for nonlinear large-residual inverse problems [35] and distinguishes our method from the linearized Hessian-based analysis of error estimates in 4DVAR (e.g., [30]) and Bayesian UQ in large-scale linear inverse problems (e.g., [5]). Our UQ method is designed for global ocean state estimation with Hessians with $O(10^{16})$ elements and is demonstrated here for an $O(10^{10})$ element Hessian. Lossless compression of Hessian matrix inversion is achieved by excluding the unconstrained principal uncertainty patterns belonging to the nullspace of the inverse problem. Unlike the methods of [26] or [5], the Hessian inversion algorithm implemented preserves the original prior-independent Hessian eigenvector structure necessary to maintain the dynamical consistency of the assimilated uncertainty fields, separating their signal from the numerically motivated prior.

Inverse and forward uncertainty propagation schemes were designed for the inverse-predictive ocean state estimation framework of ECCO. The major extension consists in assimilating observation uncertainties into those of the prior control variable uncertainties and projecting these posterior uncertainties onto key oceanographic target QoI for ocean and climate research. Two versions of schemes are presented: one does not require prior assumptions and allows "pure" uncertainty analysis unmasked by numerical regularization, and the other evaluates posterior reduction of prior uncertainties and allows forward uncertainty propagation in regularized ill-posed inverse problems.

The developed Hessian method is applied to the quantification of Drake Passage transport uncertainty in a global idealized barotropic (vertically integrated) configuration of the MITgcm. Observational constraints were provided in terms of simulated satellite altimetry. Uncertainties enter the problem through errors in the initial conditions (velocity and surface elevation fields), the surface boundary conditions (spatial pattern of wind stress forcing), and model parameters (spatially varying bottom drag).

The estimated error bounds of the transport are of magnitude similar to that of the previously reported temporal variability-based standard deviation. The simulated UQ experiment with realistically scaled assimilated observation uncertainties demonstrates a significant posterior transport uncertainty reduction. The patterns of posterior uncertainty constraint in the control variables domain demonstrate cross-coupling (covariability) between different control fields, here initial and boundary conditions, as well as the nonlocal effects of the assimilated observation uncertainty. The nondiagonal nature of posterior covariances illustrates how observations of one variable may contribute to the reduction in uncertainties of other variables that are dynamically connected through the model. The transparency of the method enabled the identification of physical mechanisms underlying the uncertainty dynamics. Forward and inverse propagation of uncertainty is explained in terms of barotropic mechanisms such as Kelvin wave propagation. These barotropic uncertainty waves are identified as the agent of global uncertainty teleconnection in the ocean. They spread the assimilated information via the dynamical model across the control variable and model state space and are associated with transient sensitivity patterns, as clarified by the structure of our UQ framework. Both the uncertainty and the sensitivity fields are described by derivatives of the dynamical ocean model. They resolve the mathematical links between perturbations of the model's inputs and outputs and therefore describe the uncertainty and sensitivity effects of one on the other.

Looking forward, our near-term goal is to integrate the developed machinery in the full realistic configuration of the ECCO ocean state estimation system. This will require the extension from synthetic to real observations, taking into account the heterogeneous data streams and associated spatiotemporal sampling characteristics, including the full three-dimensional baroclinic ocean dynamics. The developed UQ method can be applied for model calibration to guide selection of physical parameterizations, boundary conditions, and numerical parameters. The uncertainty assimilation methodology is applicable to new observation systems design by quantifying the expected information gains and optimizing for the specific observation goals. Future research directions include extending the formal quantitative framework of the current uncertainty bounds study to the field of statistical inference to formally quantify confidence regions, goodness of fit, and uncertainty partition between resolved and unresolved uncertainty sources. Improving our understanding and quantification of uncertainty dynamics of the real ocean constitutes a pressing contribution of computational science to the grand challenge problem of ocean and climate observability and predictability.

Appendix. Matrix-free Hessian computation and verification.

A.1. Hessian computation via AD. Algorithmic differentiation (AD) of complex codes requires fine control of how the automatic AD algorithms are applied. Generation of efficient derivative code with TAF involves the addition of flow directives to guide the parsing of the source code, to bypass external subroutines and precompiled libraries, and to plug in manually differentiated subroutines such as multiprocessor parallel input-output, communication, and arithmetic primitives subroutines. Flow directives are also used to reuse self-adjoint parts of the code, simplifying the adjoint and improving its computational efficiency. Time-stepping loops need to be restructured and combined with TAF storage directives to enable automatic generation of nested checkpointing schemes, which balance recomputation and storage complexity in reverse mode. Differentiation of active variables (including control variables) stored

in files, a unique feature of the MITgcm code needed to reduce memory footprint (MITgcm User Manual, Ch. 5, http://mitgcm.org/), is implemented by scalar-valued dummy pointer variables and the related auxiliary input-output subroutines. First derivative (adjoint and tangent linear) codes of the MITgcm were successfully used in previous studies (for a recent overview, see [46, 47]).

The repeated differentiation of the AD-generated first derivative code to generate second derivatives involves some additional algorithmic challenges that needed to be solved. This code includes auxiliary subroutine calls introduced by the AD compiler and has some required variables removed by the first differentiation. The former are part of input-output and memory storage constructs needed for implementation of the checkpointing scheme. These constructs require special handling in the second differentiation, which is needed to preserve the computational efficiency of adjoint checkpointing in the Hessian code. The latter are the dummy variable pointers needed for the second differentiation, which are correctly removed in the first differentiation process (because they are symbolically constant, i.e., inactive, with respect to the independent variables in the first differentiation). These dummy variables must be manually restored in the code before the second differentiation, which was implemented with additional preprocessing utilities.

A.2. Matrix-free Hessian code. Application of TAF for generation of the Hessian code allows several options. Direct differentiation of the adjoint code with TAF allows one to choose between calculation of the full matrix, its few columns, or its left or right product with an arbitrary column or row vector. For multivariate functions with only a scalar-valued dependent variable the most computationally efficient strategy for evaluation of the second derivative is the so-called forward-over-reverse mode [7]. Therefore, we first apply TAF to the full nonlinear MITgcm (parent model) in reverse mode to obtain the adjoint code, and then TAF is applied in forward mode to generate the Hessian code of the cost function. This procedure reduces the computational complexity of the resulting Hessian code because tangent linear differentiation of the adjoint code preserves the efficiency of its checkpointing schemes.

The second application of TAF can be done either in *vector mode*, to generate the entire matrix in a single run, or in *scalar mode*, which for each run produces a single Hessian-vector right product. The latter either allows column-by-column evaluation of the full Hessian matrix or can be used in power method-based iterative eigenanalysis [39], implemented here with a Lanczos algorithm.[7] Such a matrix-free method allows diagonalization of the Hessian without computing or storing the full matrix, both of which can be prohibitively large. For example, in the present application to a coarse-resolution ($2°$) global ocean model of domain size nx·ny=180·80 and with six two-dimensional control variable fields, the size of our 80·180·6=86400 element long control vector is 675 KB in double precision, while the size of the Hessian matrix is 56 GB. Moreover, if the run time for a single Hessian-vector product operation is $O(100$ seconds) (see Table 2), it would require several months of computer run time for a single evaluation of the whole Hessian matrix. The goal of our UQ machinery development is application to systems with even larger dimensionality, e.g., $O(10^8)$ in global ECCO configurations [44]. Explicit computing or storing of Hessians with $O(10^{16})$ elements, as well as application of dense matrix algebra numerics, is infeasible with present-day computers. Therefore, we build a Hessian-vector product second deriva-

---

[7]The Lanczos algorithm is implemented with the implicitly restarted Arnoldi method library ARPACK [19], using its symmetric matrix version known as the Lanczos iteration.

<div align="center">TABLE 2</div>

Code execution duration (user CPU time) for different simulated time periods (90 and 30 days) and computational complexity factor ratios (number of timesteps scaling) relative to the forward (FWD) and the adjoint (AD) codes. The codes were executed on Dual Core AMD Opteron(tm) Processor 875 (2210 MHz, cache size: 1024 KB).

	FWD	AD	HESS
CPU time (in sec) 90-day run	37	$168 - 213$	$375 - 426$
CPU time (in sec) 30-day run	12.5	$55 - 71$	$129 - 146$
CPU ratio (to FWD)	1	$4.4 - 5.8$	$10.1 - 11.6$
CPU ratio (to AD)		1	$2.0 - 2.3$

tive code and develop matrix-free reduced-rank Hessian UQ algorithms described in the previous subsections.

A.3. AD code performance. Typical run time requirements and simulation length scalability of the first and the second derivative codes of the MITgcm are listed in Table 2. The execution time scales linearly with simulation length of the forward model and the derivative codes. The computational complexity is $O(N)$, where N is the number of simulation time steps. Linear scaling factors of the derivative codes are shown as ratios relative to the forward and the adjoint codes execution times. These complexity factors are consistent with performance ratios reported by [7] of 5.5 and 11 for the adjoint and the Hessian models, respectively, relative to the forward model.

A.4. Verification of the Hessian code. To verify the derivative codes, several tests were designed and described in the following.

A.4.1. Symmetry test. A basic theorem in calculus requires that for sufficiently smooth functions the Hessian matrix must be symmetric. This symmetry is confirmed by examination of selected subsections of the Hessian matrix extracted column-by-column by application of the Hessian-vector product code to natural basis vectors $\hat{e}_i = (0, \ldots, 1, \ldots, 0)^T$. Relative symmetry errors of Hessian matrix elements h_{ij}, quantified by

$$(40) \qquad 2\frac{h_{ij} - h_{ji}}{h_{ij} + h_{ji}},$$

are shown in Figure 10(a). The symmetry is confirmed up to 6 to 7 significant digits, with the error larger for elements which are farther from the diagonal. This larger relative error is explained by the small magnitude of these remote off-diagonal Hessian elements Figure 10(b).

A.4.2. Finite difference test. Consistency of the AD-generated Hessian can be checked versus its finite difference (FD) approximation. Figure 11 shows a selected Hessian matrix element computed with forward finite difference scheme

$$(41) \qquad \frac{\partial^2 J}{\partial x_i \, \partial x_j} = \frac{J(x + \Delta_i \hat{e}_i + \Delta_j \hat{e}_j) - J(x + \Delta_i \hat{e}_i) - J(x + \Delta_j \hat{e}_j) + J(x)}{\Delta_i \Delta_j}.$$

Here Δ_i, Δ_j represent the magnitudes of finite perturbations of the independent variables, and the natural basis vectors $\hat{e}_i$, $\hat{e}_j$ are the perturbed components. The figure shows good agreement between AD and FD estimates over a range of six decades. For extremely small perturbations FD derivatives diverge from the AD estimate, and the error is due to a limited floating point machine precision in calculation of FD

derivatives. The relative AD versus FD error is shown on logarithmic scale in Figure 12. The results are qualitatively similar to the first derivative test of [9]. For sufficiently large FD perturbations the error grows linearly with perturbation magnitude, i.e., $O(\Delta)$, as expected for first order accuracy of the forward FD scheme. This verifies the consistency of the AD calculation and highlights its advantage over the FD scheme.

A.4.3. Taylor remainder test. Another verification is to measure the magnitude of the remainder term in second order Taylor series expansion (Patrick Farrell, personal communication). For sufficiently smooth functions the quadratic Taylor polynomial is

$$
(42) \quad
\begin{aligned}
J(\mathbf{x}) \;&=\; J(\mathbf{x}_0) \;+\; \left.\frac{\partial J}{\partial \mathbf{x}^T}\right|_{\mathbf{x}_0} (\mathbf{x} - \mathbf{x}_0) \\
&\quad + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T \left.\frac{\partial J^2}{\partial \mathbf{x}\, \partial \mathbf{x}^T}\right|_{\mathbf{x}_0} (\mathbf{x} - \mathbf{x}_0) \;+\; o(\|\mathbf{x} - \mathbf{x}_0\|^2).
\end{aligned}
$$

The magnitude of the remainder term is tested by computing the normalized remainder coefficient

$$
(43) \quad
\frac{\left\| J(\mathbf{x}) - J(\mathbf{x}_0) - \dfrac{\partial J}{\partial \mathbf{x}^T}(\mathbf{x} - \mathbf{x}_0) - \dfrac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T \dfrac{\partial J^2}{\partial \mathbf{x}\, \partial \mathbf{x}^T}(\mathbf{x} - \mathbf{x}_0) \right\|}{\|\mathbf{x} - \mathbf{x}_0\|^2}.
$$

Figure 13 shows it to decrease linearly, consistent with third order differentiability of the cost function and the scaling of the next order Taylor term. For very small perturbations of the controls, the consistency breaks, which is again explained by the limited machine precision.
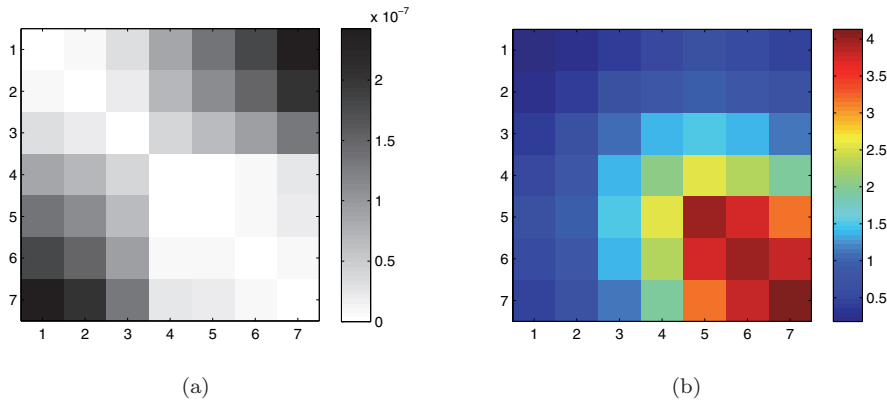

(a)

(b)

FIG. 10. *Relative symmetry errors of computed Hessian matrix shown for 7 by 7 Hessian matrix subsection* (a) *and the actual values of the Hessian subsection* (b).
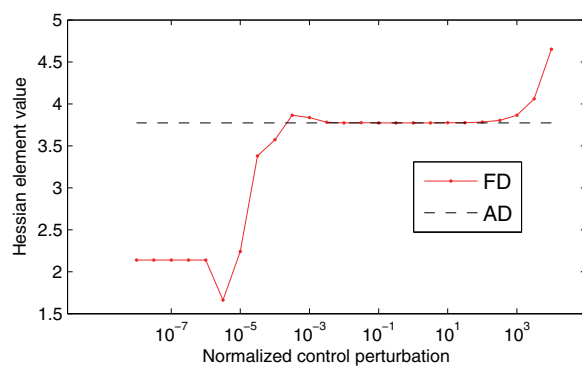
FIG. 11. *Comparison of the finite difference approximations of an off-diagonal element of the 9 Hessian matrix with the AD calculated value, shown for a range of finite difference perturbations.*
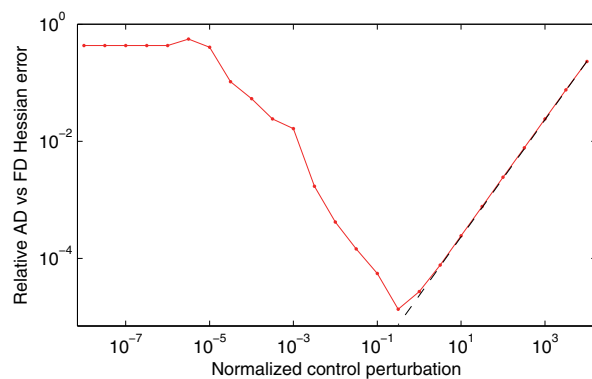

FIG. 12. *Relative AD versus FD error on logarithmic scale. The black dashed line is a linear fit to the error confirming first order FD error scaling.*
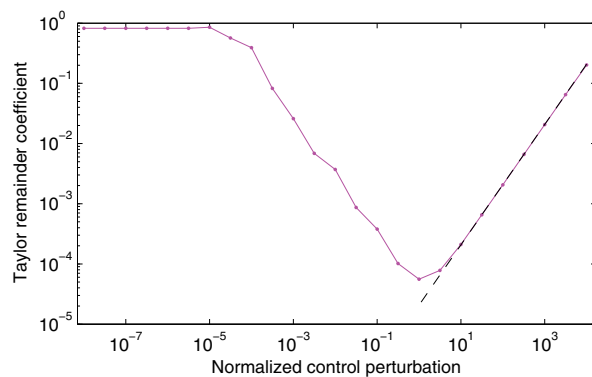

FIG. 13. *Normalized Taylor remainder coefficient for second order Taylor series expansion with AD derivatives. The linear fit (black dashed line) over the range of remainder convergence confirms the third order scaling of the next Taylor expansion term.*

REFERENCES

[1] A. ADCROFT, J. CAMPIN, C. HILL, AND J. MARSHALL, *Implementation of an atmosphere-ocean general circulation model on the expanded spherical cube*, Mon. Weather Rev., 132 (2004), pp. 2845–2863.

[2] L. BIEGLER, G. BIROS, O. GHATTAS, M. HEINKENSCHLOSS, D. KEYES, B. MALLICK, L. TENORIO, B. VAN BLOEMEN WAANDERS, K. WILLCOX, AND Y. MARZOUK, *Large-Scale Inverse Problems and Quantification of Uncertainty*, Wiley Series in Computational Statistics 712, Wiley, New York, 2010.

[3] B. CUSHMAN-ROISIN AND J. M. BECKERS, *Introduction to Geophysical Fluid Dynamics: Physical and Numerical Aspects*, Int. Geophys. Ser. 101, Academic Press, New York, 2011.

[4] T. A. EL MOSELHY AND Y. M. MARZOUK, *Bayesian inference with optimal maps*, J. Comput. Phys., 231 (2012), pp. 7815–7850.

[5] H. P. FLATH, L. C. WILCOX, V. AKÇELIK, J. HILL, B. VAN BLOEMEN WAANDERS, AND O. GHATTAS, *Fast algorithms for Bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial Hessian approximations*, SIAM J. Sci. Comput., 33 (2011), pp. 407–432.

[6] R. GIERING AND T. KAMINSKI, *Recipes for adjoint code construction*, ACM Trans. Math. Software, 24 (1998), pp. 437–474.

[7] R. GIERING, T. KAMINSKI, AND T. SLAWIG, *Generating efficient derivative code with TAF: Adjoint and tangent linear Euler flow around an airfoil*, Future Generation Comput. Syst., 21 (2005), pp. 1345–1355.

[8] J. C. GILBERT AND C. LEMARÉCHAL, *Some numerical experiments with variable-storage quasi-Newton algorithms*, Math. Program., 45 (1989), pp. 407–435.

[9] M. GILES, D. GHATE, AND M. DUTA, *Using Automatic Differentiation for Adjoint CFD Code Development*, Report NA-05/25, Oxford University Computing Laboratory, Oxford, UK, 2005.

[10] A. GRIEWANK AND A. WALTHER, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, 2nd ed., SIAM, Philadelphia, 2008.

[11] J. R. GUNSON AND P. MALANOTTE-RIZZOLI, *Assimilation studies of open-ocean flows 2. Error measures with strongly nonlinear dynamics*, J. Geophys. Res., 101 (1996), 28473–28488.

[12] P. HEIMBACH, C. HILL, AND R. GIERING, *An efficient exact adjoint of the parallel MIT general circulation model, generated via automatic differentiation*, Future Generation Comput. Syst., 21 (2005), pp. 1356–1371.

[13] P. HEIMBACH AND V. BUGNION, *Greenland ice sheet volume sensitivity to basal, surface, and initial conditions, derived from an adjoint model*, Ann. Glaciology, 50 (2009), pp. 67–80.

[14] P. HEIMBACH, C. WUNSCH, R. M. PONTE, G. FORGET, C. HILL, AND J. UTKE, *Timescales and regions of the sensitivity of Atlantic meridional volume and heat transport: Toward observing system design*, Deep Sea Research Part II, 58 (2011), pp. 1858–1879.

[15] P. HEIMBACH AND C. WUNSCH, *Decadal ocean (and ice) state estimation for climate research: What are the needs?*, Oberwolfach Reports, 9 (2012), pp. 3451–3454.

[16] T. ITO, M. WOLOSZYN, AND M. MAZLOFF, *Anthropogenic carbon dioxide transport in the Southern Ocean driven by Ekman flow*, Nature, 463 (2010), pp. 80–83.

[17] A. KALMIKOV, *Uncertainty Quantification in Ocean State Estimation*, Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA, and Woods Hole Oceanographic Institution, Woods Hole, MA, 2013.

[18] O. KNIO AND O. LE MAITRE, *Uncertainty propagation in CFD using polynomial chaos decomposition*, Fluid Dyn. Res., 38 (2006), pp. 616–640.

[19] R. B. LEHOUCQ, D. C. SORENSEN, AND C. YANG, *ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*, Software Environ. Tools 6, SIAM, Philadelphia, 1998.

[20] P. F. J. LERMUSIAUX AND A. ROBINSON, *Data assimilation via error subspace statistical estimation. Part* I: *Theory and schemes*, Mon. Weather Rev., 127 (1999), pp. 1385–1407.

[21] M. LOSCH AND C. WUNSCH, *Bottom topography as a control variable in an ocean model*, J. Atmos. Ocean. Technol., 20 (2003), pp. 1685–1696.

[22] J. R. MAGNUS AND H. NEUDECKER, *Matrix differential calculus with applications in statistics and econometrics*, John Wiley & Sons, New York, 1988.

[23] J. MAROTZKE, R. GIERING, K. Q. ZHANG, D. STAMMER, C. HILL, AND T. LEE, *Construction of the adjoint MIT ocean general circulation model and application to Atlantic heat transport sensitivity*, J. Geophys. Res., 104 (1999), pp. 529–548.

[24] J. MARSHALL, A. ADCROFT, C. HILL, L. PERELMAN, AND C. HEISEY, *A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers*, J. Geophys. Res., 102 (1997), pp. 5753–5766.

[25] J. MARTIN, L. C. WILCOX, C. BURSTEDDE, AND O. GHATTAS, *A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion*, SIAM J. Sci. Comput., 34 (2012), pp. A1460–A1487.

[26] Y. MARZOUK, *Bayesian inference for inverse problems: Foundations and computational methods*, in Large-scale Inverse Problems and Quantification of Uncertainty Workshop, Institute for Mathematics and Its Applications (IMA), University of Minnesota, Minneapolis, MN, 2011.

[27] M. R. MAZLOFF, *On the sensitivity of the Drake Passage transport to air-sea momentum flux*, J. Clim., 25 (2012), pp. 2279–2290.

[28] M. R. MAZLOFF, P. HEIMBACH, AND C. WUNSCH, *An eddy-permitting Southern Ocean state estimate*, J. Phys. Oceanogr., 40 (2010), pp. 880–899.

[29] M. P. MEREDITH ET AL., *Sustained monitoring of the Southern Ocean at Drake Passage: Past achievements and future priorities*, Rev. Geophys., 49 (2011), RG4005.

[30] A. M. MOORE, H. G. ARANGO, G. BROQUET, B. S. POWELL, A. T. WEAVER, AND J. ZAVALA-GARAY, *The Regional Ocean Modeling System (ROMS)* 4-*dimensional variational data assimilation systems Part* I: *System overview and formulation*, Prog. Oceanogr., 91 (2011), pp. 34–49.

[31] J. T. ODEN, R. MOSER, AND O. GHATTAS, *Computer predictions with quantified uncertainty, Part* II, SIAM News, 43 (2010), pp. 1–4.

[32] D. OLBERS, D. BOROWSKI, C. VOLKER, AND J. WOLFF, *The dynamical balance, transport and circulation of the Antarctic Circumpolar Current*, Antarct. Sci., 16 (2004), pp. 439–470.

[33] S. R. RINTOUL, C. W HUGHES, AND D. OLBERS, *The Antarctic circumpolar current system*, International Geophysics, 77 (2001), pp. 271–302.

[34] T. P. SAPSIS AND P. F. LERMUSIAUX, *Dynamically orthogonal field equations for continuous stochastic dynamical systems*, Phys. D, 238 (2009), pp. 2347–2360.

[35] G. A. F. SEBER AND C. J. WILD, *Nonlinear Regression*, Wiley Ser. Probab. Statist. 503. Wiley, New York, 2003.

[36] D. STAMMER, C. WUNSCH, R. GIERING, C. ECKERT, P. HEIMBACH, J. MAROTZKE, A. ADCROFT, C. HILL, AND J. MARSHALL, *The global ocean circulation during* 1992–1997*, estimated from ocean observations and a general circulation model*, J. Geophys. Res., 107 (2002), p. 3118.

[37] J. R. TAYLOR, *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*, University Science Books, Herndon, VA, 1997.

[38] W. C. THACKER, *The role of the Hessian matrix in fitting models to measurements*, J. Geophys. Res., 94 (1989), pp. 6177–6196.

[39] L. N. TREFETHEN AND D. BAU, III, *Numerical Linear Algebra*, SIAM, Philadelphia, 1997.

[40] G. K. VALLIS, *Atmospheric and Oceanic Fluid Dynamics*, Cambridge University Press, Cambridge, UK, 2006.

[41] W. WEIJER, B. M. SLOYAN, M. E. MALTRUD, N. JEFFERY, M. W. HECHT, C. A. HARTIN, E. VAN SEBILLE, I. WAINER, AND L. LANDRUM, *The Southern Ocean and its climate in CCSM*4, J. Clim., 25 (2012), pp. 2652–2675.

[42] C. WUNSCH, *Discrete Inverse and State Estimation Problems: With Geophysical Fluid Applications*, Cambridge University Press, Cambridge, UK, 2006.

[43] C. WUNSCH AND D. STAMMER, *Satellite altimetry, the marine geoid, and the oceanic general circulation*, Ann. Rev. Earth Planet. Sci., 26 (1998), pp. 219–253.

[44] C. WUNSCH AND P. HEIMBACH, *Practical global oceanic state estimation*, Phys. D, 230 (2007), pp. 197–208.

[45] C. WUNSCH, P. HEIMBACH, R. M. PONTE, AND I. FUKUMORI, *The Global General Circulation of the Ocean Estimated by the ECCO-Consortium*, Oceanography, 22 (2009), pp. 88–103.

[46] C. WUNSCH AND P. HEIMBACH, *Two decades of the Atlantic meridional overturning circulation: Anatomy, variations, extremes, prediction, and overcoming its limitations*, J. Clim., 26 (2013), pp. 7167–7186.

[47] C. WUNSCH AND P. HEIMBACH, *Dynamically and kinematically consistent global ocean circulation and ice state estimates*, in Ocean Circulation and Climate: A 21st Century Perspective, G. Siedler, J. Church, J. Gould, and S. Griffies, eds., Elsevier, New York, 2013, pp. 553–579.