A dynamics-weighted principal components analysis of dominant atmospheric drivers of ocean variability with an application to the North Atlantic subpolar gyre

Daniel E. Amrhein[a] Dafydd Stephenson[a] LuAnne Thompson[b]

[a] *NSF National Center for Atmospheric Research, Boulder, CO*

[b] *University of Washington School of Oceanography, Seattle, WA*

Corresponding author: Dan Amrhein, damrhein@ucar.edu

This work has been accepted to the Journal of Climate. The AMS does not guarantee that the copy provided here is an accurate copy of the Version of Record (VoR).

ABSTRACT: This paper describes a framework for identifying dominant atmospheric drivers of ocean variability. The method combines statistics of atmosphere-ocean fluxes with physics from an ocean general circulation model to derive atmospheric patterns optimized to excite variability in a specified ocean quantity of interest. We first derive the method as a weighted principal components analysis and illustrate its capabilities in a toy problem. Next, we apply our analysis to the problem of interannual upper ocean heat content (HC) variability in the North Atlantic Subpolar Gyre (SPG) using the adjoint of the MITgcm and atmosphere-ocean fluxes from the ECCOv4-r4 state estimate. An unweighted principal components analysis reveals that North Atlantic heat and momentum fluxes in ECCOv4-r4 have a range of spatiotemporal patterns. By contrast, dynamics-weighted principal components analysis collapses the space of these patterns onto a small subset – principally associated with the North Atlantic Oscillation – that dominates interannual SPG HC variance. By perturbing the ECCOv4-r4 state estimate, we illustrate the pathways along which variability propagates from the atmosphere to the ocean in a nonlinear ocean model. This technique is applicable across a range of problems across Earth System components, including in the absence of a model adjoint.

SIGNIFICANCE STATEMENT: While the oceans have absorbed 90% of the excess heat associated with human-forced climate change, the change in the ocean's heat content is not steady, with peaks and troughs superimposed upon a general increase. These fluctuations come from chaotic changes in the atmosphere and ocean, and can be hard to disentangle. We use this case of ocean heat content variability to introduce a new method for determining the patterns of weather and climate in the atmosphere that are most effective at generating fluctuations in the ocean. To do this, we combine the statistics of recent atmospheric activity with output from a state-of-the-art numerical ocean model that reveals physical processes driving changes in ocean quantities including ocean heat content. This approach suggests that the atmospheric patterns that stimulate the most energetic changes in ocean heat content in the northern North Atlantic are not the most energetic patterns present in the atmosphere. We test our findings by preventing these patterns from affecting the ocean in our numerical model, and measure a strong reduction in ocean heat content fluctuations.

1. Introduction

The ocean's distributions of momentum, thermal energy, salt, and other quantities evolve across a range of length and time scales, reflecting contributions from solar radiation, turbulent fluxes of heat and momentum at the air-sea interface, inputs from land and cryosphere, tides, hydrothermal heating, and the internal variability of the turbulent ocean. Identifying processes and pathways by which the ocean changes through time is important for revealing mechanisms and timescales of predictability, fingerprints of anthropogenic changes, and the drivers of ocean variability and change in the past and future. For many scales and processes, variability about an ocean mean state may be usefully approximated as being driven by a stochastic atmosphere (Hasselmann 1976; Frankignoul and Hasselmann 1977; Kushnir et al. 2002), with secondary roles for ocean turbulence, additional external drivers, and ocean-atmosphere feedbacks on time scales longer than those associated with turbulent fluxes. Within these regimes, clarifying dominant pathways of atmospheric influence on the ocean has the potential to provide parsimonious descriptions of variability in a high-dimensional coupled system.

A traditional paradigm for exploring dominant drivers of ocean variability is to identify dynamically important modes of variability in the atmosphere and then to evaluate their impact on the ocean. In the North Atlantic, much of the atmospheric variability on seasonal and longer time scales is associated with large-scale patterns in atmospheric circulation (Deser et al. 2010). In particular, heat fluxes due to the North Atlantic Oscillation (NAO), the dominant mode of winter-time atmospheric variability in the extratropical Northern Hemisphere (Hurrell and Deser 2009), yield a characteristic "tripole" pattern with warmth in mid-latitudes and cooler temperatures in the subpolar region and between the Equator and 30°N (Cayan 1992a,b; Marshall et al. 2001). The NAO has been implicated in a range of ocean and climate variability. Frajka-Williams et al. (2017) find that NAO-related surface heat fluxes likely explain a recent cooling in the subpolar North Atlantic. Ortega et al. (2017) use an eddy permitting multi-century integration of a coupled model to show that 62% of Labrador Sea density variance comes from low-frequency variations of the NAO, with freshwater-driven ocean circulation changes having a larger effect at centennial time scales. Bersch (2002) and Bersch et al. (2007) find that NAO wind anomalies are important for Labrador Sea convection, northward heat transport through the SPG, and SPG structure, and Tesdal et al. (2018) attribute recent freshening in the Labrador Sea to a spin-up in the SPG that may be associated with NAO and Arctic Oscillation winds. Finally, Böning et al. (2006) and Lozier et al. (2008) attribute decadal variations in SPG heat content (HC) and structure to combined influences of NAO winds and buoyancy forcing.

However, there are also several lines of evidence that the NAO is not the only driver of hydrographic change in the SPG. Häkkinen et al. (2011) find that NAO-like patterns of wind stress curl changes are not principally responsible for anomalous northward penetration of warm and saline subtropical waters; instead, a secondary atmospheric circulation mode, resembling the East Atlantic Pattern (EAP), was implicated that modulated NAO and storm track strength and that had a larger projection onto SPG variability. Similarly, Barrier et al. (2014) performed forward sensitivity analyses in a coupled model and found that different patterns of atmospheric variability were associated with different time scales of ocean response. Kim et al. (2016) note that Labrador Sea convection resumed in the winter of 2008/2009 after a hiatus beginning in the mid-1990s despite that year having the same positive sign of NAO as the previous winter, and suggest a possible connection to La Niña as an additional source of variability in deep Atlantic water masses.

While it is natural to evaluate the role of leading atmospheric modes in forcing ocean variability, there is no requirement that a pattern derived to maximize the contribution to atmospheric variability – for instance, through a regional atmospheric empirical orthogonal function / principal component

(EOF-PC; Lorenz (1956)) analysis or by regression of an index of atmospheric variables – will be the dominant driver of variability for specified quantities in the ocean. A second, "bottom-up" approach poses an inverse question: Given an ocean quantity of interest (hereafter QoI), such as the heat content of an ocean volume, what is the hypothetical atmospheric variability that would most efficiently excite it? This problem can be addressed using adjoint sensitivity analyses, which leverage linearized ocean general circulation model dynamics to determine the origins of changes in ocean QoIs. A growing body of literature uses adjoint sensitivities to study ocean hydrography and dynamics (Marotzke et al. 1999; Köhl and Stammer 2004; Bugnion et al. 2006a,b; Czeschel et al. 2010, 2012; Mazloff 2012; Fukumori et al. 2015; Pillar et al. 2016; Jones et al. 2018; Kostov et al. 2019; Stephenson and Sévellec 2021a,b) revealing the adjoint approach as a powerful method for determining pathways of change for ocean processes on climate-relevant scales.

A challenge in interpreting adjoint sensitivities is that their spatiotemporal structure is set by the choice of the QoI and by the dynamics of the ocean model, with no information included about the dynamics or statistics of the atmosphere except indirectly through their impact on simulated ocean circulation. For instance, Stephenson and Sévellec (2021b) use an adjoint approach to show that North Atlantic heat content variability can originate from winds along narrow bands that stimulate Ekman transport and coastal upwelling. Similarly, Jones et al. (2018), following Marotzke et al. (1999), decompose sensitivities of Labrador Sea HC into kinematic (constant circulation) and dynamic (changing circulation) components and argue that HC changes can emerge advectively from upstream source waters as well as via an ocean wave propagation mechanism excited from forcing applied in a narrow band of the West African shelf. The narrow regions implicated by these studies as optimal ocean drivers, with zonal length scales on the order of hundreds of kilometers, reflect the scales of Rossby deformation radii in the ocean and ultimately of ocean model grid boxes, in contrast to dominant length scales of wind variability of thousands of kilometers. A consequence explored in previous literature both in the context of the El Niño-Southern Oscillation (Kleeman and Moore 1997; Moore and Kleeman 1999; Zavala-Garay et al. 2003; Moore et al. 2006; Kleeman 2008) and the Atlantic circulation (Chhak and Moore 2007; Zanna and Tziperman 2008; Chhak et al. 2009) is that it is important to consider the projection of atmospheric variability onto ocean sensitivities, rather than just the sensitivities themselves, in order to understand drivers of ocean QoIs. A corollary is that the leading EOF of atmospheric variability need not be the most important driver for the variability in a particular ocean diagnostic. Similarly, dominant patterns ("stochastic optimals", discussed further below) in ocean sensitivities to hypothetical atmospheric conditions might not indicate important avenues by which the atmosphere drives ocean variability, but might instead languish as "potential" pathways that are never actually activated.

This work combines "top-down" approaches informed by atmospheric statistics and "bottom-up" approaches shaped by ocean dynamics through adjoint sensitivity analysis. As opposed to classical EOF-PC analyses, we develop "empirical–dynamical functions" (EDFs) and "dynamics-weighted principal components" (DPCs) that reflect both model dynamics and observed atmospheric statistics. Our approach parallels model reduction procedures in control engineering, where one seeks to reduce the degrees of freedom in a dynamical system (often to minimize computational burden) while preserving features in both its "controllability" (i.e., where the system can go) and "observability" (any properties are of interest) of the system. Following work by Adamjan et al. (1971), Moore (1981) describes an approach for "balanced truncation" that approximates a system in a new basis informed by both controllability and observability. (See Antoulas (2005) and Brunton and Kutz (2022) for additional introduction; Rowley (2005) shows that balanced truncation can be computed efficiently using the singular value decomposition, which is the approach used here.) The explicit connection to the present work is that atmospheric EOFs are an estimate of the principal directions of controllability in the atmosphere, while stochastic optimals describe the principal directions of observability in the case where we "observe" the atmosphere via its impact on the ocean. Balanced truncation for model reduction has been applied previously in atmosphere-ocean contexts by Farrell and Ioannou (2001), Moore et al. (2022), and Xu et al. (2024). Here, we focus on dominant dynamical connections revealed by low-dimensional descriptions of forced ocean variability.

The remainder of this paper is as follows. First, we present a derivation of the EDF–DPC approach as an optimization problem. Under limiting conditions, EDFs recover EOFs and stochastic optimals. Next, we demonstrate the approach in a simplified stochastic system and show how EDFs bear the imprint of both sensitivities and forcing statistics. We then apply the EDF–DPC decomposition using the adjoint of the MITgcm for the problem of understanding leading contributions by heat fluxes and wind stress to interannual variability in North Atlantic Subpolar Gyre heat content. EDFs outperform EOFs for driving variability in the linearized dynamical framework of the adjoint, and the leading EDFs of both heat flux and wind stress are highly correlated with the NAO. To evaluate the efficacy of EDFs in a nonlinear model, we then rerun the ECCOv4-r4 ocean state estimate with atmospheric fluxes modified to omit EDFs. We find good correspondence between variance in the nonlinear MITgcm and what is predicted by linear (adjoint) dynamics, though in the case of heat fluxes, the removal of the leading NAO–like EDF pattern leads to a long-term cooling trend. Rerunning the ECCOv4-r4 state estimate with additional EDF perturbations illustrates the mechanisms by which atmospheric variability adjusts heat content in the North Atlantic.

2. Theoretical Framework: Dynamically weighted principal components

a. Adjoint sensitivities and ocean variability

As noted in Section 1, adjoint representations of ocean models are powerful tools for evaluating causes of ocean variability. We begin by introducing these concepts. We denote column vectors by bold variables and define the ocean model state vector, $\mathbf{x}(t)$, to be the set of prognostic variables (temperature, salinity, velocity, etc.) at time t at all latitudes, longitudes, and depths. Then the evolution of an ocean general circulation model can be written as

$$\mathbf{x}(t+\Delta t) = F\left[\mathbf{x}(t), \mathbf{u}(t)\right] \tag{1}$$

where F is a nonlinear operator and $\mathbf{u}(t)$ is a vector of time-varying atmospheric fluxes inclusive of all ocean model grid boxes and flux types. Next, we define a scalar, time-varying ocean "quantity of interest" $\mathrm{QoI}(t)$ as a weighted sum over the model state vector,

$$\mathrm{QoI}(t) = \sum_j \boldsymbol{\alpha}^\top(t, t_j)\mathbf{x}(t_j), \tag{2}$$

where the vector $\boldsymbol{\alpha}(t, t_j)$ consists of weights – reflecting, e.g., model grid box volumes and areal and temporal extent – defining the appropriate integral, for instance, to yield annually- and regionally-averaged heat content.

The adjoint sensitivity $s(\tau)$ is given by

$$\mathbf{s}(\tau) = \frac{\partial \mathrm{QoI}(t)}{\partial \mathbf{u}(t-\tau)}, \tag{3}$$

and is a linearized estimate of how $\mathrm{QoI}(t)$ changes in response to small changes in $\mathbf{u}$ at a time lead τ. Here and throughout this paper we make the simplifying stationarity assumption that $\mathbf{s}(\tau)$ is not a function of t. If a finite-amplitude change $\delta\mathbf{u}(\tau)$ is made in the fluxes (e.g., an increase in wind stress over the Northern Hemisphere), then the change $\delta\mathrm{QoI}(t)$ is given by (modifying Fukumori et al. (2015))

$$\delta\mathrm{QoI}(t) \approx \sum_{i=1}^{N_\tau} \mathbf{s}(\tau_i)^\top \delta\mathbf{u}(t-\tau_i), \tag{4}$$

where changes are summed over lags $\tau_1, \tau_2, \ldots \tau_{N_\tau}$ and we obtain equality when the model response to flux adjustments is linear. As described in greater detail in Section 4a, adjoint sensitivities are an output of the state estimation machinery underlying the ECCO state estimate, and can be produced from the MITgcm via automatic differentiation. They can be similarly computed from other models that have adjoint capabilities (the Regional Ocean Modeling System, ROMS, Moore et al. 2004; and Tangent and Adjoint Models for the Nucleus for European Modelling of the Ocean, NEMOTAM, Vidard et al. 2015).

We estimate total QoI variance σ_Σ^2 by assuming a linear response to fluxes and taking the expectation over time of squared QoI anomalies, $\sigma_\Sigma^2 = \left\langle (\delta\mathrm{QoI}(t))^2 \right\rangle$. Substituting Eq. (4), we obtain

$$\sigma_\Sigma^2 = \sum_{i=1}^{N_\tau}\sum_{j=1}^{N_\tau} \mathbf{s}(\tau_i)^\top \left\langle \delta\mathbf{u}(t-\tau_i)\delta\mathbf{u}^\top(t-\tau_j) \right\rangle \mathbf{s}(\tau_j) \tag{5}$$

$$= \sum_{i=1}^{N_\tau}\sum_{j=1}^{N_\tau} \mathbf{s}(\tau_i)^\top \mathbf{C}_{ij} \mathbf{s}(\tau_j) \tag{6}$$

where $\mathbf{C}_{ij}$ is the spatial covariance matrix of $\delta\mathbf{u}$ at time lag $\tau_i - \tau_j$. Covariances of air-sea fluxes can have complex structure in space and time, reflecting, e.g., the propagation of properties through the ocean and atmosphere. Here we discuss three approximations to make the description of this variability more tractable. First, we approximate $\mathbf{C}_{ij}$ as separable in space and time (Hasselmann 1993; Chen et al. 2021),

$$\mathbf{C}_{ij} = d_{ij}\mathbf{C}, \tag{7}$$

which assumes that covariances of atmospheric fluxes at different lags are the same, up to a lag-dependent scaling factor d_{ij}^2. While there are limitations inherent in assuming separability – one cannot, for instance, represent propagating waves – it nonetheless can describe fluxes with non-zero correlations in time (i.e., not just white noise) and non-stationary (time-evolving) covariances. Equation (6) can then be expressed in terms of a matrix trace as

$$\sigma_\Sigma^2 = \mathrm{tr}\,(\mathbf{ZC}) \tag{8}$$

where we define

$$\mathbf{Z} = \sum_{i=1}^{N_\tau}\sum_{j=1}^{N_\tau} d_{ij}\mathbf{s}(\tau_i)\mathbf{s}(\tau_j)^\top. \tag{9}$$

and we have followed Kleeman and Moore (1997) by incorporating information about flux non-stationarity and temporal covariance in $\mathbf{Z}$ via the d_{ij}. This separability assumption underlies the dynamics-weighted principal components approach; the following two additional approximations can be convenient, but are not required. If a white noise assumption adequately represents space-time covariances, reflecting rapid decorrelation times of atmospheric fluxes relative to the ocean circulation (Hasselmann 1976; Frankignoul and Hasselmann 1977), then one can set lag flux correlations to zero by choosing $d_{ij} = \delta_{ij} d_{ij}$ where δ_{ij} is the Kronecker delta. Note that this form still represents changes in the variance of fluxes through time, which can be large over a seasonal cycle. Finally, if fluxes are furthermore assumed to be stationary (constant spatial covariance through time), then $d_{ij} = \delta_{ij}$ and $\mathbf{Z} = \mathbf{SS}^\top$, where the matrix

$$\mathbf{S} = \left[\mathbf{s}(\tau_1), \mathbf{s}(\tau_2), \ldots, \mathbf{s}(\tau_{N_\tau})\right] \tag{10}$$

is formed by concatenating sensitivities across N_τ discrete lags. The model stochastic optimals (Farrell and Ioannou 1996; Kleeman and Moore 1997) are the left singular vectors of $\mathbf{S}$.

b. Optimal atmospheric drivers of ocean variability

Next, our goal is to decompose atmospheric variability into patterns and corresponding time series, analogous to EOF–PC analysis. We do this by combining adjoint sensitivities from an ocean model and atmospheric fluxes to define a matrix the square of whose diagonal elements sum to the QoI variance. The singular vectors of that matrix yield a set of flux patterns ordered by their contributions to ocean variability. Begin by defining a data matrix

$$\mathbf{U} = (N_t - 1)^{-\frac{1}{2}} \left[\mathbf{u}(t_0), \mathbf{u}(t_1), \ldots, \mathbf{u}(t_{N_t})\right] \tag{11}$$

consisting of vectors $\mathbf{u}(t_i)$ of fluxes concatenated column-wise across N_t discrete times. We have scaled $\mathbf{U}$ so that the zero-lag flux covariance can be estimated as

$$\mathbf{C} = \mathbf{U}\mathbf{U}^\top \tag{12}$$

and flux PCs and EOFs are the left and right singular vectors of $\mathbf{U}$, respectively. We express our decomposition as

$$\mathbf{U} = \sum_{k=1}^{N_{DPC}} \mathbf{p}_k \mathbf{t}_k^\top \tag{13}$$

where $\mathbf{p}_k$ denotes the k^{th} "empirical–dynamical function" (EDF) and $\mathbf{t}_k$ the corresponding "dynamics-weighted principal component" (DPC) up to an integer N_{DPC}. For notational convenience we stipulate that $\|\mathbf{t}_k\| = 1$, where $\|\|$ denotes the vector l^2 norm, so that $\|\mathbf{p}_k\|^2$ is the flux variance accounted for by the k^{th} EDF-DPC pair in $\mathbf{U}$. We require that EDFs represent distinct processes insofar as their variability is uncorrelated in time within $\mathbf{U}$, meaning that (like PCs) the $\mathbf{t}_k$ are orthonormal; however, unlike EOFs, the EDFs are not generally orthogonal in space. Right multiplying (13) by $\mathbf{t}_k$ and using orthonormality, we find that the $\mathbf{p}_k$ are given by the projection of $\mathbf{t}_k$ onto $\mathbf{U}$,

$$\mathbf{U}\mathbf{t}_k = \mathbf{p}_k. \tag{14}$$

To find the set of EDF–DPC pairs, we solve an optimization problem. We first substitute (13) into (8) to obtain

$$\sigma_\Sigma^2 = \mathrm{tr}\left(\sum_{i=1}^{N_{DPC}} \sum_{j=1}^{N_{DPC}} \mathbf{Z}\mathbf{p}_i \mathbf{t}_i^\top \mathbf{t}_j \mathbf{p}_j^\top\right). \tag{15}$$

By orthonormality of the $\mathbf{t}_k$ and Eq. (14) we find

$$\sigma_\Sigma^2 = \sum_{k=1}^{N_{DPC}} \operatorname{tr}\left(\mathbf{Z}\mathbf{p}_k\mathbf{p}_k^\top\right) \tag{16}$$

$$= \sum_{k=1}^{N_{DPC}} \|\mathbf{Z}^{\frac{\top}{2}}\mathbf{U}\mathbf{t}_k\|^2. \tag{17}$$

where we have defined a matrix decomposition $\mathbf{Z} = \mathbf{Z}^{\frac{1}{2}}\mathbf{Z}^{\frac{\top}{2}}$ and used the invariance of trace under cyclic permutations. We can now define an optimization problem to find the leading DPC $\mathbf{t}_1$ that maximizes the contribution to QoI variance $\sigma_1^2 = \|\mathbf{Z}^{\frac{\top}{2}}\mathbf{U}\mathbf{t}_1\|^2$,

$$\mathbf{t}_1 = \underset{\mathbf{t}}{\operatorname{argmax}}\, \|\mathbf{Z}^{\frac{\top}{2}}\mathbf{U}\mathbf{t}\|^2. \tag{18}$$

The solution to (18) for $\|\mathbf{t}_1\| = 1$ is given by the leading right singular vector of the matrix $\mathbf{Z}^{\frac{\top}{2}}\mathbf{U}$. Generalizing beyond the leading DPC, if we define the singular vector decomposition as

$$\mathbf{Z}^{\frac{\top}{2}}\mathbf{U} = \mathbf{L}\mathbf{\Sigma}\mathbf{T}^\top = \sum_{k=1}^{N_{DPC}} \mathbf{l}_k\sigma_k\mathbf{t}_k^\top, \tag{19}$$

then the full set of DPCs is given by the columns $\mathbf{t}_k$ of $\mathbf{T}$. DPCs are ordered by their contributions, σ_k^2, to the total QoI variance. The number of meaningful EDF–DPC pairs N_{DPC} is given by the the number of nonzero σ_k, i.e., the rank of the matrix $\mathbf{Z}^{\frac{\top}{2}}\mathbf{U}$, and obeys $N_{DPC} \leq \min\left(\operatorname{rank}\left(\mathbf{Z}^{\frac{\top}{2}}\right), \operatorname{rank}(\mathbf{U})\right)$.

EDFs recover familiar results in limiting cases. First, the case where $\mathbf{C}$ is proportional to the identity matrix corresponds to fluxes that are Gaussian white noise in space. If the fluxes are also stationary Gaussian white noise in time, then EDFs are equivalent to the stochastic optimals of the model. Similarly, if $\mathbf{Z}$ is proportional to the identity matrix, then the model is insensitive to spatial patterns in fluxes, and the EDFs are equivalent to the flux EOFs. At the opposite limit of spatial degrees of freedom, when fluxes are proportional to a single spatial pattern at all times, the (single) EDF is simply that pattern. When adjoint sensitivities are proportional to a single spatial pattern $\mathbf{s}_1$ at all lags, then $\mathbf{t}_1 = \mathbf{U}^\top\mathbf{s}_1/\left(\mathbf{s}_1^\top\mathbf{C}\mathbf{s}_1\right)$ and the single EDF is proportional to the product of the spatial covariance with $\mathbf{s}_1$, $\mathbf{p}_1 = \mathbf{C}\mathbf{s}_1\left(\mathbf{s}_1^\top\mathbf{C}\mathbf{s}_1\right)$.

Finally, we can construct "impact maps" indicating where QoI variance originates in space for each EDF. As noted by Stephenson and Sévellec (2021b), the map of total variance contributed by fluxes is given by

$$\mathbf{v}_\Sigma = \mathrm{diag}\left(\mathbf{Z}\mathbf{C}\right) \tag{20}$$

where diag($\mathbf{A}$) denotes the vector lying along the diagonal of a matrix $\mathbf{A}$. Substituting Eq. (19), we obtain

$$\mathbf{v}_\Sigma = \sum_{k=1}^{N_{DPC}} \mathrm{diag}\left(\mathbf{Z}^{\frac{1}{2}}\mathbf{l}_k \sigma_k \mathbf{t}_k^\top \mathbf{U}^\top\right) \tag{21}$$

from which we can write the map of variance contributions for the k^{th} EDF-DPC pair as

$$\mathbf{v}_k = \sigma_k \left(\mathbf{Z}^{\frac{1}{2}}\mathbf{l}_k\right) \odot \mathbf{p}_k, \tag{22}$$

where $\odot$ denotes the element-wise product. As we will show in Section 4, impact maps are are useful for diagnosing dominant pathways of variance to the QoI.

3. Demonstration in a simple stochastic system

Before applying the EDF method in the context of a full ocean GCM, we illustrate it in an idealized one-dimensional configuration (Figure 1). In this setup, we generate realizations of random fluxes that are correlated in space (mimicking large-scale atmospheric variability) and stationary Gaussian white noise in time (consistent with Hasselmann (1976)). Three realizations of this stochastic process are shown in Figure 1a. The leading EOFs computed from 10,000 realizations of these fluxes (Figure 1b) have length scales comparable to the extent of the domain and are approximately symmetric about its midpoint. Next, we generate sensitivities of a hypothetical QoI to fluxes at 10 lags as scaled delta functions in the leftmost third of the domain (Figure 1c) with randomly chosen scalings. These sensitivities mimic properties of adjoint sensitivities computed in ocean models, which often have shorter length scales than those in the atmospheric variability and are concentrated within a subset of the spatial domain, as described in Section 1. In the case shown where sensitivities across different lags have nonzero values at distinct spatial locations,

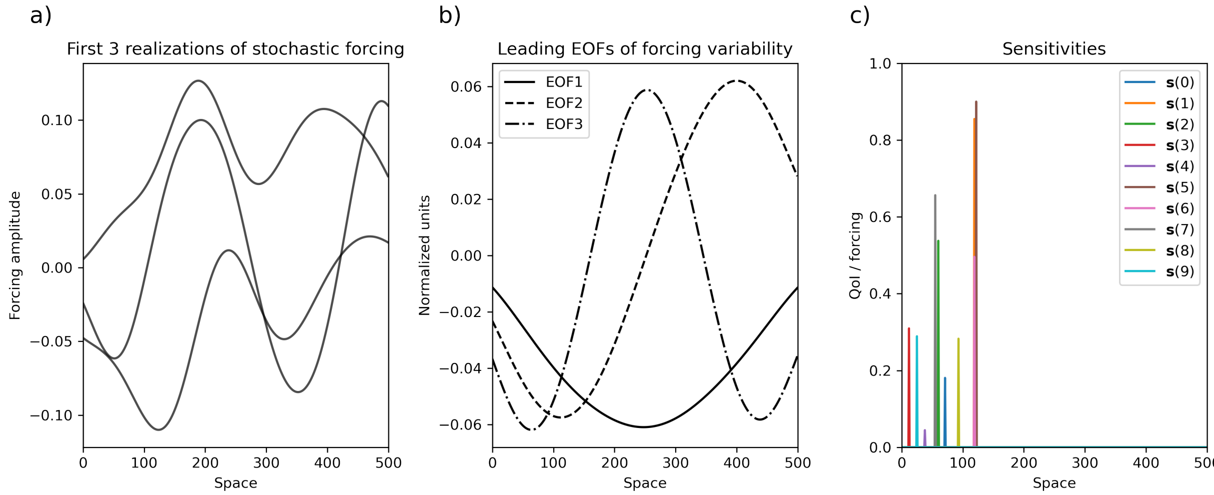

FIG. 1. Setup for EDF–DPC analysis in a simple stochastic system. a) Stochastic forcing in a synthetic one-dimensional system is generated by smoothing Gaussian white noise in space. b) Leading EOFs of this forcing have large spatial scales and are approximately symmetric in space. c) Adjoint sensitivities $\mathbf{s}(\tau)$ are randomly generated with different values across ten time lags.

the stochastic optimals (computed as the left singular vectors of $\mathbf{S}$, not shown) are simply delta functions at those locations, ordered by their magnitudes.

To compute DPCs, we construct $\mathbf{U}$ from Eq. (11) using 10,000 realizations of stochastic forcing and $\mathbf{S}$ from Eq. (10), concatenating across the ten lags. In the stationary white noise case, $\mathbf{Z} = \mathbf{S}\mathbf{S}^\top$ (Section 2a) and we can use $\mathbf{Z}^{\frac{1}{2}} = \mathbf{S}$. Then computing singular vectors of $\mathbf{S}^\top\mathbf{U}$ (Eq. (19)) and EDFs (Eq. (14)) yields ten EDF-DPC pairs with nonzero contributions summing to the total QoI variance. In contrast to the leading EOFs, leading EDF patterns (Figure 2a-c) are asymmetric in space, reflecting the preference imparted by the adjoint sensitivities for the left side of the domain. While the EDF patterns are not orthogonal in space, the corresponding DPC time series (not shown) are orthonormal white noise.

Each EDF–DPC pair's contribution to QoI variance is given by the corresponding squared singular value of $\mathbf{S}^\top\mathbf{U}$ (Figure 3a, circles). For comparison, the QoI variance contribution from the i^{th} EOF $\mathbf{e}_i$ is given by

$$\left(\sigma_i^{\mathrm{EOF}}\right)^2 = \|\lambda_i \mathbf{S}^\top \mathbf{e}_i\|^2 \tag{23}$$

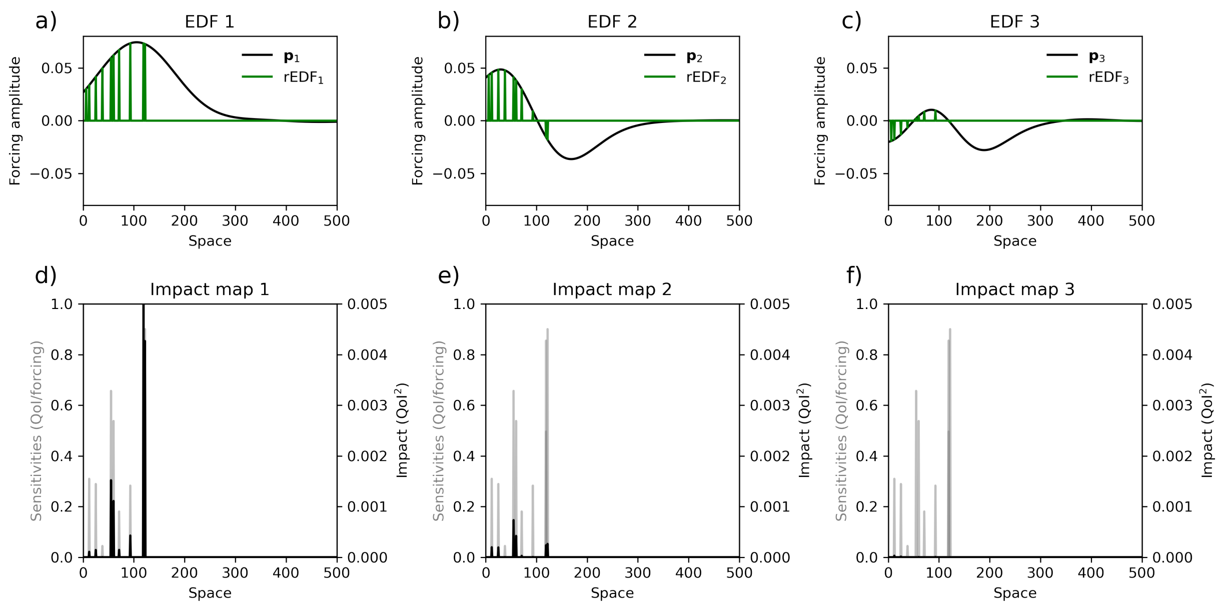

FIG. 2. a-c) The leading three EDFs (spatial patterns) computed in a simple 1-D example. "Reduced" EDFs (rEDF$_k$, green lines) show the subset of each EDF that contributes to QoI variability; other nonzero EDF values arise from spatial forcing covariances. d-f) Impact maps (Equation 22) illustrating contributions to QoI variance across space (black lines) for each EDF. For comparison, these are overlaid on the distribution of adjoint sensitivities across lags (gray lines, also shown in Figure 1c).

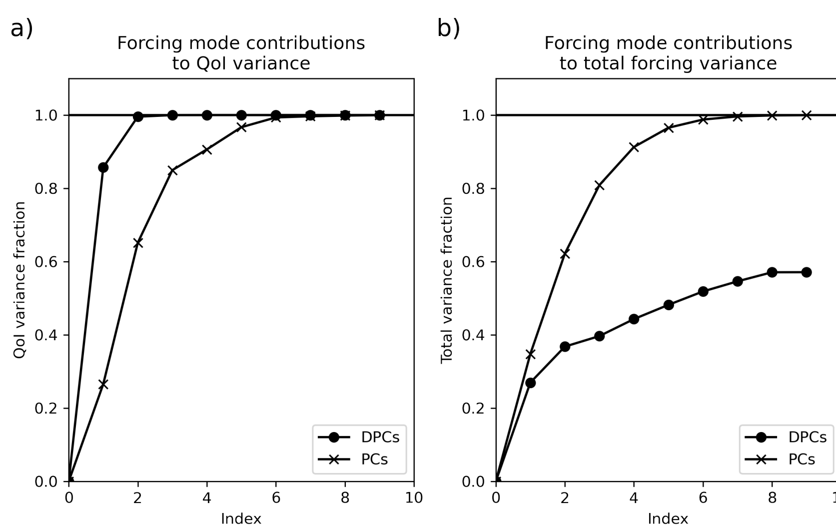

FIG. 3. Comparison of contributions from EDF-DPC and EOF-PC pairs to a) QoI variance and b) total forcing variance in a simple 1-D example.

where λ_i^2 is the contribution of the i^{th} EOF to the total flux variance. As expected, leading DPCs account for a greater fraction of QoI variance than do leading EOFs, with a more rapid convergence of cumulative variance explained (compare circles and X's in Figure 3a). We can perform the equivalent comparison for contributions to the total flux variance by comparing the λ_i^2 to $\|\mathbf{p}_k\|^2$, where the latter describes how much of the variance in $\mathbf{U}$ is explained by the k^{th} EDF–DPC pair, revealing that EOFs maximize contributions to total flux variability more effectively than DPCs (Figure 3b). Thus, EDF-DPC pairs can have an outsize impact on variability in the QoI relative to their contribution to flux variability.

Impact maps (black lines, Figure 2 d-f) are computed using Eq. (22) and indicate the amount of QoI variance contributed by each EDF as a function of space, which is determined by a combination of local sensitivity and EDF amplitudes. By ranking the impact map and selecting locations with leading impacts, we can define a "reduced" EDF (rEDF), plotted in green in Figure 2 (a-c). As we discuss in the next section, the rEDF is useful for clarifying the dominant mechanisms by which the EDF impacts QoI variance. In this simplified case, all of the QoI variance is explained by the subset of locations with nonzero sensitivities. At other locations, EDFs are nonzero because of spatial correlations in the fluxes, and have no impact on the QoI.

4. Leading atmospheric drivers of interannual Subpolar Gyre heat content variability

Next, we examine the EDFs of upper-ocean heat content in the North Atlantic Subpolar Gyre (SPG). This region was chosen for its dynamical importance for AMOC strength across models (Yeager et al. 2021; Oldenburg et al. 2021) including the MITgcm (Kostov et al. 2022) as well as for being a place where ocean dynamics are thought to play an important role in sea surface temperature variability (Buckley et al. 2014, 2015; Wills et al. 2019). This work follows previous studies using the adjoint for investigations of drivers of SPG variability (Jones et al. 2018; Stephenson and Sévellec 2021b).

a. Model setup

The MITgcm (Marshall et al. 1997; Adcroft et al. 2004) simulates ocean circulation under hydrostatic and Boussinesq assumptions. Here we use the nominal 1 degree configuration with 50 vertical levels used for the ECCO version 4, release 4 (ECCOv4-r4) state estimate (Wunsch and

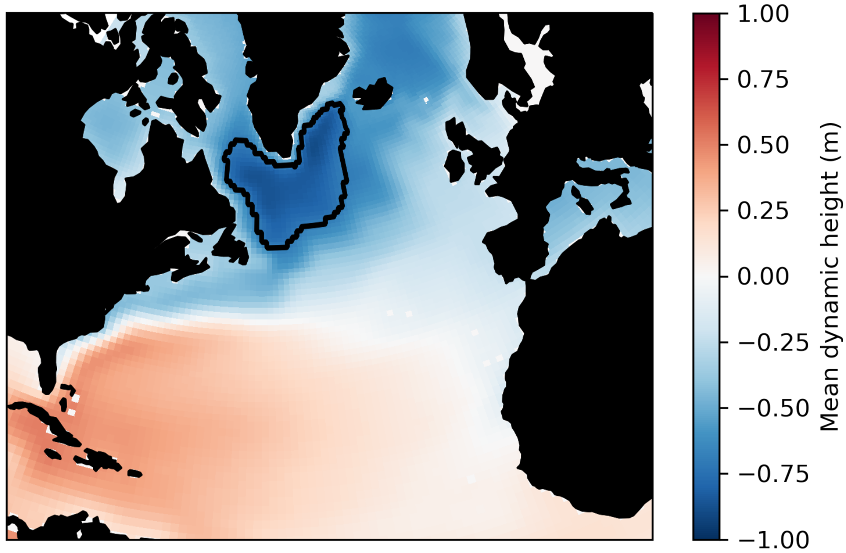

FIG. 4. Definition of the region of interest used for adjoint sensitivities. The black outline demarcates the North Atlantic Subpolar Gyre region, chosen as the largest negative closed contour of time mean dynamic height in the ECCOv4-r4 state estimate (-70 cm) following Foukal and Lozier (2017).

Heimbach 2007; Forget et al. 2015b; Fukumori et al. 2018) with two sets of initial and boundary conditions: one to construct the flux data matrix **U**, and the other to construct adjoint sensitivities. We construct **U** per Equation (11) from fluxes derived for ECCOv4-r4, which assimilates a range of observations to produce a dynamically consistent history of recent ocean variability spanning 1992 to 2017. We construct **U** separately for ECCOv4-r4 heat fluxes and wind stress at 6 hourly resolution, concatenating zonal and meridional wind stress into a single matrix. The statistics of ECCOv4-r4 air-sea fluxes include effects from adjustments of the forcing, initial conditions and mixing parameterizations made to create a product that fits ocean observations; we make no effort to separate this contribution and neglect any possible erroneous impact to large-scale patterns of flux covariance.

The second set of initial and boundary conditions are used to construct the ocean state about which the adjoint is computed. Here we use the initial conditions and forcing of Wolfe et al. (2017), who spun up the MITgcm for 5400 years under CORE Normal Year Forcing (Large and Yeager 2004). Using an annually repeating forcing set for the adjoint ensures that ocean dynamics are not subject to forced interannual variability, such that any variability diagnosed with our method can be attributed to historical fluctuations in **U** over the ECCOv4-r4 period. Following Foukal and Lozier (2017), we define the SPG as the area enclosed by the largest negative closed contour (-70 cm) of dynamic height anomaly in the climatology of the ECCOv4-r4 state estimate (Figure 4). Note that this approach yields an SPG definition with a reduced footprint in the eastern part of the basin relative to Foukal and Lozier (2017) (cf. their Figure 3a). We compute the model adjoint using TAF (Transformation of Algorithms in Fortran; Giering and Kaminski 1998) and compute sensitivities of annual mean heat content (HC) above 700m in the SPG to heat fluxes (HF) and zonal and meridional wind stress (WS) in an Atlantic domain from 35° S to 80° N at lags from 0 hours to 40 years.

b. Atmospheric fluxes and sensitivities in the ECCOv4-r4 state estimate and the MITgcm

Heat flux and wind stress variability in ECCOv4-r4 is summarized by EOF–PC analysis (Figures 5 and 6). In all cases, because we are focused on interannual variability in SPG HC, we compute fluxes as anomalies about a seasonal cycle estimated from the ECCOv4-r4 climatology. The spectrum of squared singular values in the EOF–PC analyses of HF and WS both show a gradual convergence to the total power (the sum of squared singular values; Figures 5g and 6g), indicating that fluxes are composed of a diversity of patterns of roughly equal importance. Leading EOFs of HF (Figure 5 a-c) extend across the North Atlantic, with centers of action reflecting gyre structure and the path of the Gulf Stream. By contrast, leading EOFs of WS (Figure 6 a-c) are centered primarily over the SPG, with only small-amplitude correlated structures in the rest of the domain. Principal components for both HF and WS (Figures 5 and 6, d-f) have variability across a range of timescales, as demonstrated by the low frequency variability of running means computed over annual and 5-year intervals. While we have subtracted the climatological seasonal cycle, the 6-hourly product shows a strong annual cycle in the variance of WS and particularly HF, consistent with a North Atlantic that is stormier and more variable in winter.

Leading stochastic optimals (SOs) and their corresponding lag time series illustrate potential pathways for surface fluxes to change SPG HC in the MITgcm (Figures 7 and 8). We compute SOs and accompanying lag time series as left and right singular vectors of **S**, which is constructed by concatenating snapshots of sensitivities at lag increments of five days spanning 0 to 40 years (Equation (10)). The leading SOs capture large fractions of the spatiotemporal variability in

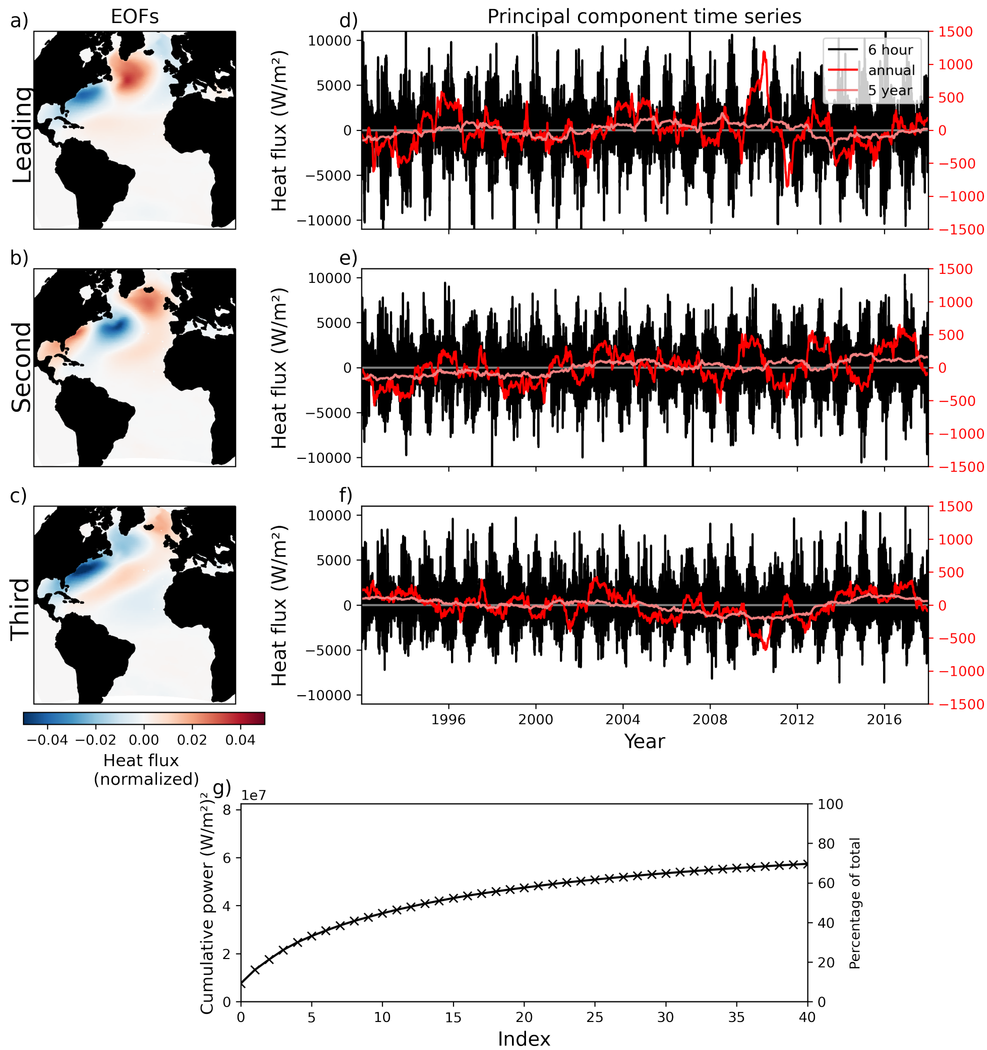

FIG. 5. EOFs (leading spatial patterns; a-c) and PCs (corresponding time series, d-f, shown at 6-hourly resolution and with annual and 5-year moving averages applied), for heat flux anomalies about annual climatology in the ECCOv4-r4 state estimate. Spatial patterns are reported in normalized units. While the estimated seasonal cycle has been removed from heat fluxes, there is a prominent seasonal cycle in the amplitude of variability in all three principal components. g) Variance accounted for by EOF-PC pairs converges gradually to the total variance.

adjoint sensitivities across lags, with the leading SO accounting for roughly 70% and 50% of the structure in the HF and WS cases, respectively (Figures 7g and 8g). HF perturbations that are mostly restricted to the SPG (Figure 7a) lead to HC anomalies that persist for several years, with a strong dependence on the season when the perturbation is applied (Figure 7d). In contrast,

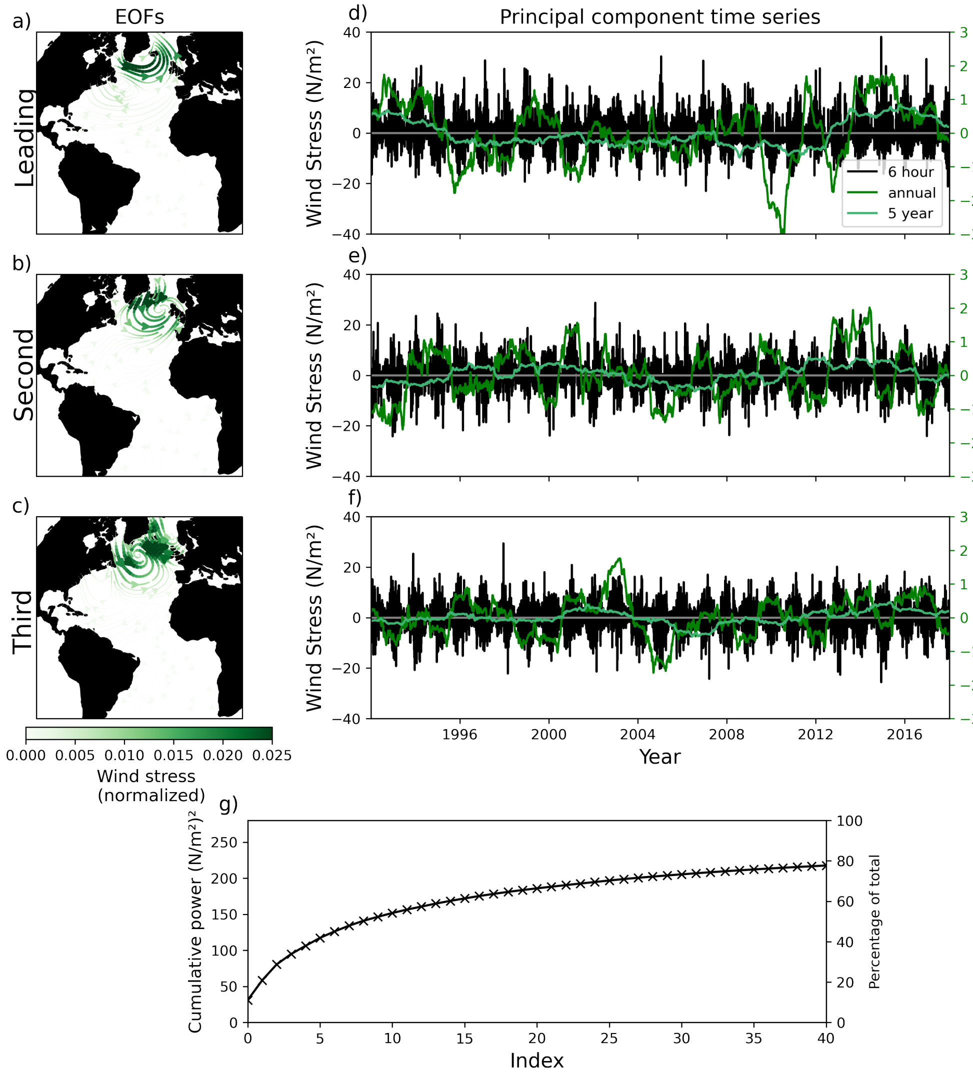

Fig. 6. Same as Figure 5 but for wind stress.

HF perturbations with anomalies extending along the model Gulf Stream (Figure 7c) persist over decadal timescales, again with strong seasonal dependence (Figure 7f). The leading HF SO (Figure 7a) strongly resembles the regional QoI definition in the SPG (cf. 4) and reflects local heating, with modest additional contributions from heat fluxes upstream in the Gulf Stream. The leading WS SO (Figure 8a) is a combination of local effects within the SPG and coastal upwelling mechanisms described by Jones et al. (2018) and Stephenson and Sévellec (2021b), whereby anomalies in HC propagate as Kelvin waves counterclockwise around the North Atlantic towards the Labrador Sea. While both HF and WS sensitivities have a seasonal dependence, it is stronger for HF, with NH wintertime fluxes having up to an order of magnitude greater impact in subsequent years than

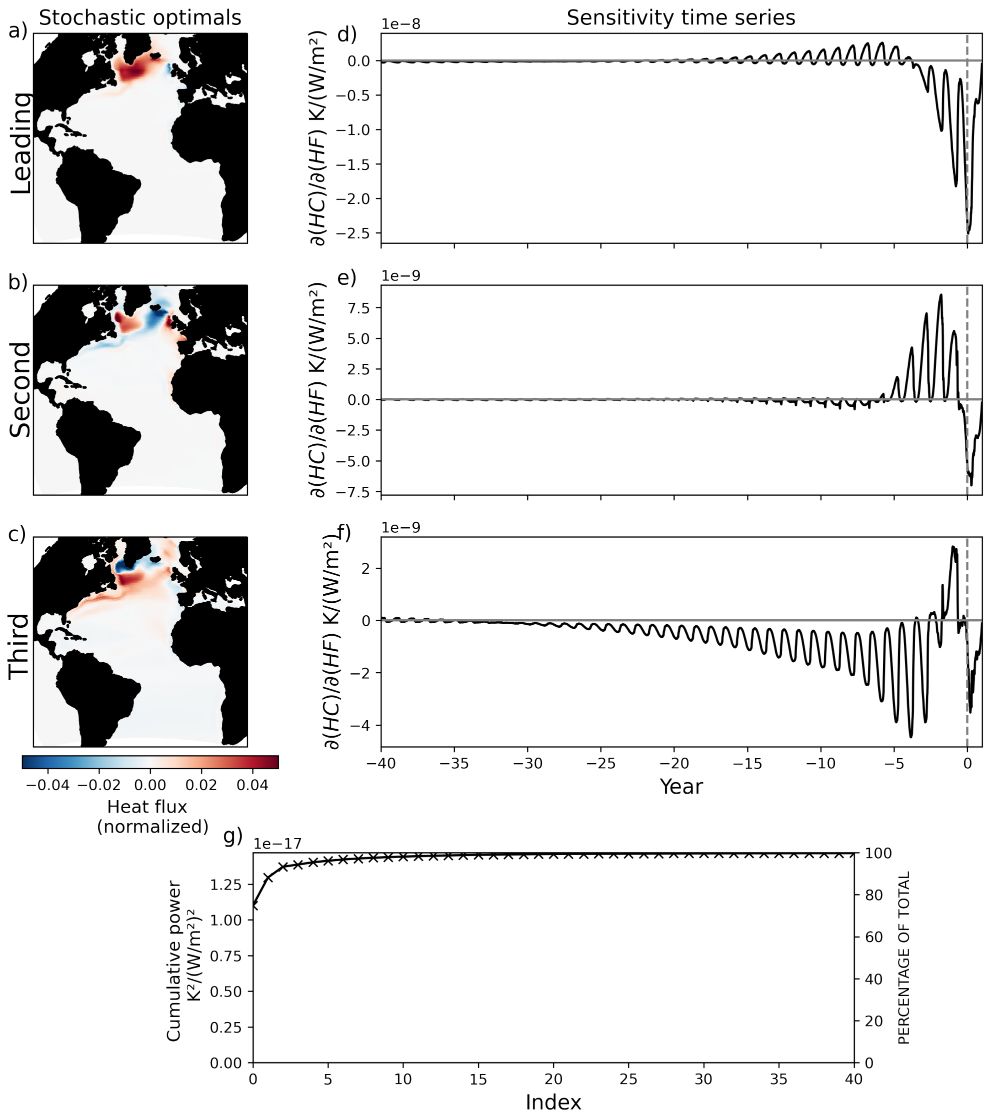

FIG. 7. Stochastic optimals (a-c) and corresponding lag time series (d-f) illustrating the hypothetical most efficient patterns of heat fluxes for driving SPG HC variability. Dotted lines at the zero-lag mark denote the beginning of the one-year period over which SPG HC is averaged to compute the QoI. Cumulative power (g) indicates that roughly 90% of the structure of adjoint sensitivities is accounted for by the leading three SO-lag pairs.

summertime fluxes. This seasonal dependence is consistent with a contrast between a strong, shallow model pycnocline in the summer relative to deeper winter mixed layers that allow greater penetration of thermal anomalies (Stommel 1979; MacGilchrist et al. 2021).

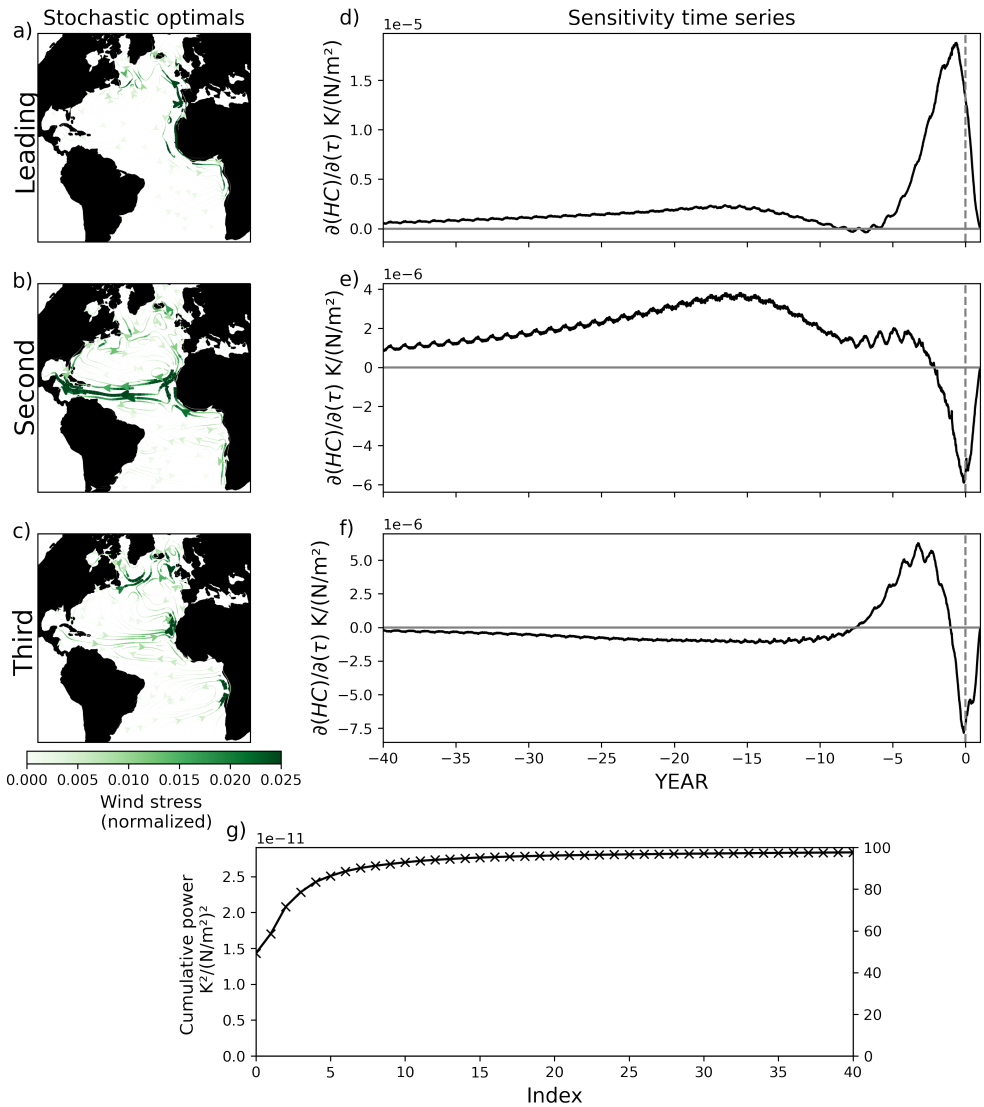

FIG. 8. Same as Figure 7, but for wind stress.

c. Dominant atmospheric drivers of Subpolar Gyre heat content variability

The EOFs (Figures 5 a-c and 6 a-c) and SOs (Figures 7 a-c and 8 a-c) for SPG HC illustrate the dichotomy discussed in Section 1: EOFs are generally large scale and agnostic of the ocean QoI, while SOs specifically reflect SPG properties defined by the QoI, with shorter length scales. Next we compute EDF–DPC pairs to reconcile these perspectives. The typical autocorrelation structure of flux principal components is roughly 1.5 days (not shown), substantially shorter than the interannual time scales of interest, so we compute $\mathbf{Z}$ following (9), consistent with a white

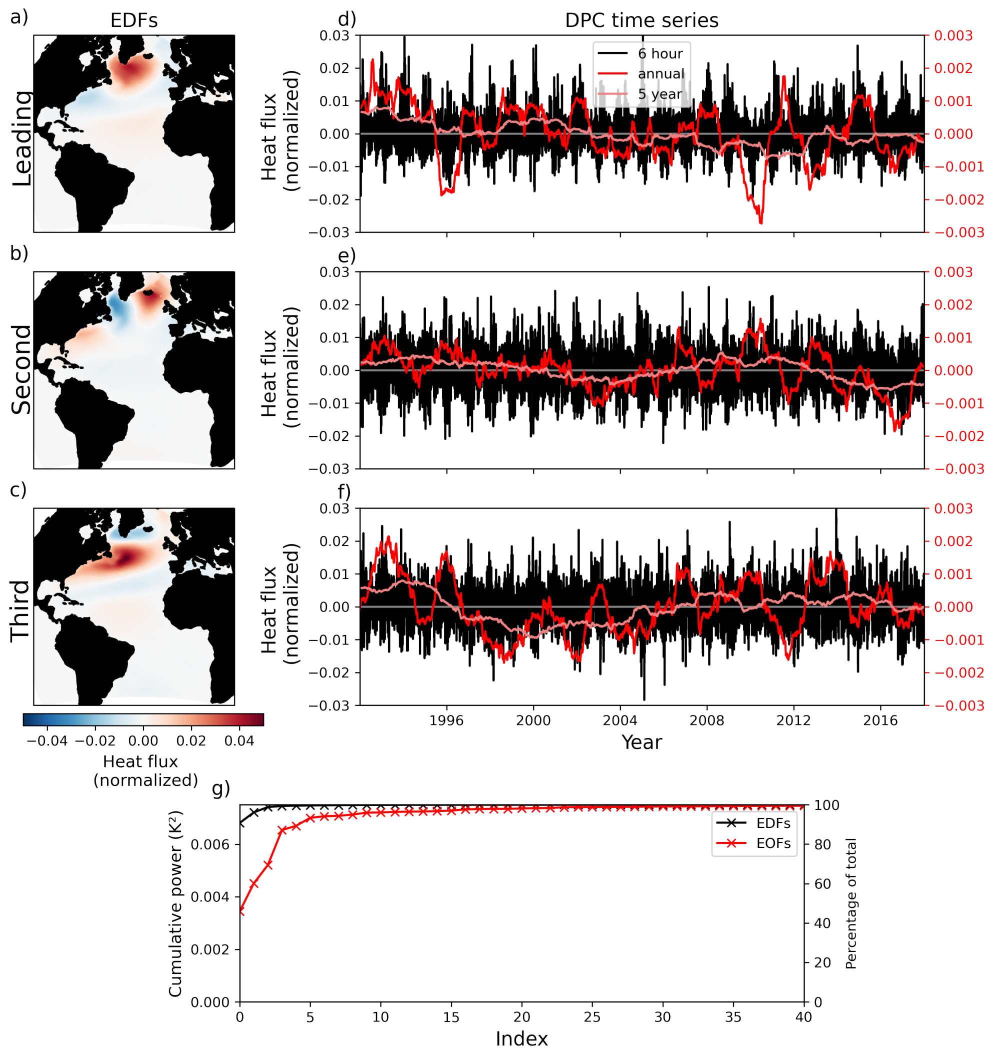

FIG. 9. Pairs of empirical-dynamical functions (a-c) and corresponding dynamics-weighted principal components (d-f) illustrate the statistically most efficient subset of heat fluxes in the ECCOv4-r4 state estimate for driving SPG HC variability in the MITgcm. Both EDFs and DPCs are plotted in normalized units; red DPC curves indicate moving averages of 1 and 5 years. Panel g) compares QoI variance accounted for by leading EDF-DPC (black) and EOF-PC (red) pairs; the leading EDF–DPC pair is expected to account for roughly 90% of the interannual variability in SPG HC caused by HF roughly double that explained by the leading EOF.

noise assumption. We estimate seasonal nonstationarity in variance amplitudes (to specify the d_i in (9)) following Stephenson and Sévellec (2021a) (not shown).

The leading EDF–DPC pairs of HF and WS account for a high fraction of the SPG HC variability driven by those fluxes (black lines in Figures 9g and 10g; note that the total power reported in

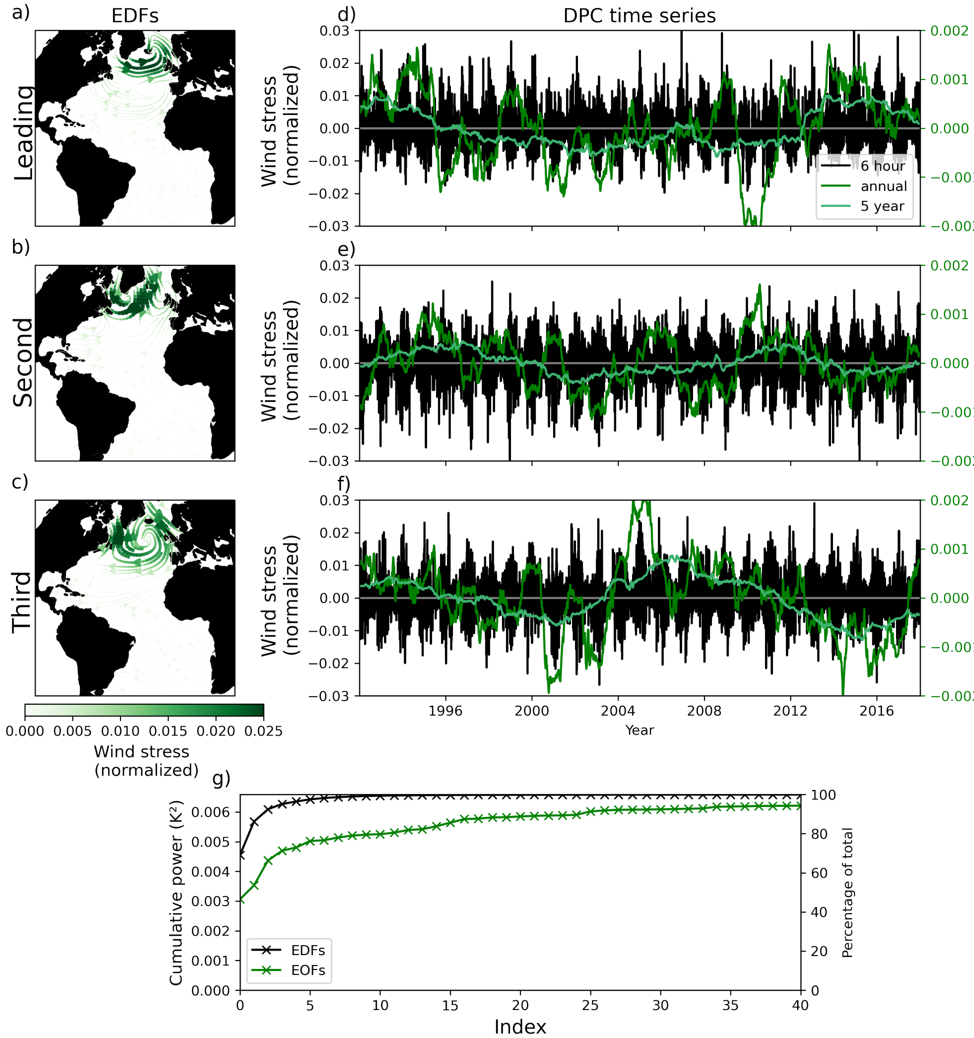

FIG. 10. Same as Figure 9 but for wind stress.

these figures is the total power contributed individually by HF and WS to the QoI). As expected, contributions to QoI variance from leading EDFs are larger than for EOFs (green and red lines in Figures 9g and 10g; cf. simple model results in Figure 3a). The leading HF EDF (Figure 9a) consists primarily of a single center of action centered on the SPG with secondary zonal bands to the south, distinct from the leading EOF (Figure 5a), which has a stronger heat flux minimum over the model Gulf Stream. By contrast, the leading EDF of wind stress (Figure 10a) qualitatively resembles the leading WS EOF (Figure 6a). Similar to PCs, DPCs (Figures 9 d-f and 10 d-f) are approximately white noise in time, with a typical maximum autocorrelation timescale of approximately 1.5 days. The seasonal cycle of variance is less pronounced in HF DPCs than in

PCs (cf. Figure 5d and 9d), possibly owing to separability assumptions made in the derivation of DPCs.

EDF–DPC pairs for HF (Figure 9 a-c) and WS (Figure 10 a-c) reflect a combination of influences from model SOs and atmospheric flux statistics. The EDF patterns in both wind stress and heat flux strongly resemble the corresponding NAO flux patterns obtained by regressing the leading PC of SLP in ECCOv4-r4 onto the ECCOv4-r4 HF and WS fields (94% agreement measured by pattern correlation for both HF and WS, not shown); the NAO tripole pattern (Cayan 1992a) is evident in Figure 9a. While connections between EDFs, EOFs, and SOs can be complex, for heat fluxes we are near one of the limit cases discussed in Section 2b whereby adjoint sensitivities can be represented by a single stochastic optimal (specifically, note that the variance accounted for by the leading SO in Figure 7g is a high fraction of the total). As such, the dominant EDF closely resembles the pattern generated when one multiplies the leading stochastic optimal (Figure 7a) by the spatial covariance of ECCOv4-r4 heat fluxes (not shown).

d. Evaluating EDF–DPC patterns in the ECCOv4-r4 state estimate

Next, we assess how well dominant spatial patterns derived under linearized ocean physics (from the adjoint sensitivities) perform in a nonlinear ocean model constrained to fit data. The ECCO state estimate is derived using a 4DVAR smoother to improve fits to observations over 1992-2017 (Wunsch and Heimbach 2007; Forget et al. 2015a; Fukumori et al. 2017), and the final product is a forward simulation of the MITgcm under adjusted initial conditions, atmospheric conditions (or fluxes), and ocean mixing parameters. We use the flux-forced version of ECCOv4-r4, which permits partitioning drivers of ocean variability into respective contributions without cross terms that can arise, e.g., between winds and surface air temperature when computing bulk fluxes (Fukumori et al. 2021).

As an initial comparison, we convolve adjoint sensitivities with HF and WS from ECCOv4-r4 (Kostov et al. 2021) and find qualitative agreement with annual mean SPG HC in the ECCOv4-r4 state estimate (Figure 11a, duplicated in Figure 12a), suggesting that a linearized system forced by HF and WS can skillfully describe historical variability in the nonlinear state estimate. Next, we subtract EDF–DPC pairs from ECCOv4-r4 fluxes and use these reduced fluxes to re-compute linear reconstructions and re-run the ECCOv4-r4 state estimate. Using Equation (13), we define a

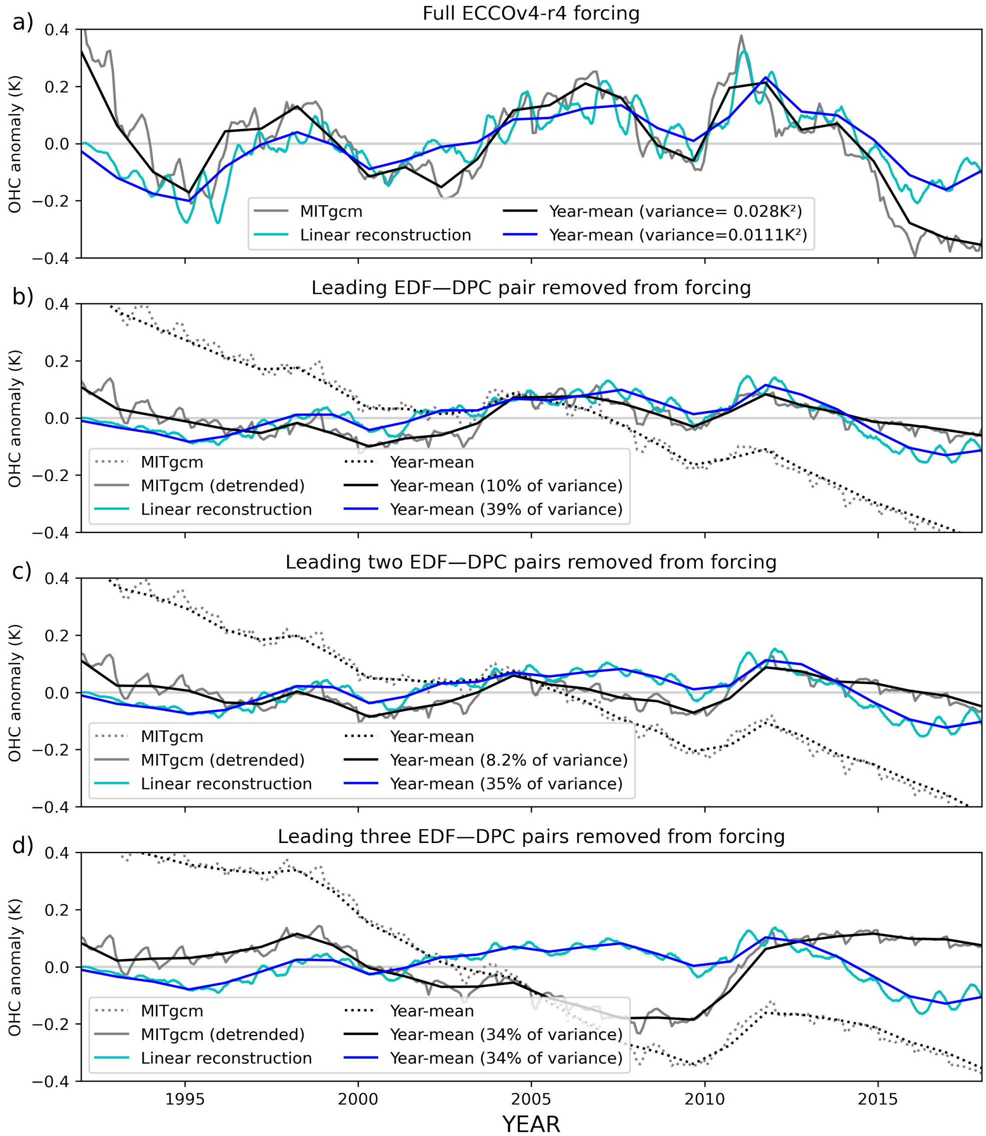

FIG. 11. Consequences of cumulatively removing HF EDF–DPC pairs from the ECCOv4-r4 state estimate. Lines show the evolution of SPG HC over the years of the ECCOv4-r4 reconstruction under full fluxes (a) and after removing the first 1 (b), 2 (c), and 3 (d) HF EDF–DPC pairs. Black and gray lines indicate anomalies computed in the (nonlinear) MITgcm before (dotted lines) and after (solid lines) subtracting a linear trend attributed to a nonlinear response to removing HF EDF-DPC pairs. Blue lines indicate anomalies reconstructed linearly by convolving fluxes with adjoint sensitivities.

set of reduced fluxes by cumulatively removing the g leading EDF-DPC pairs,

$$\mathbf{U}'_g = \mathbf{U} - \sum_{k=1}^{g} \mathbf{p}_k \mathbf{t}_k^\top. \tag{24}$$

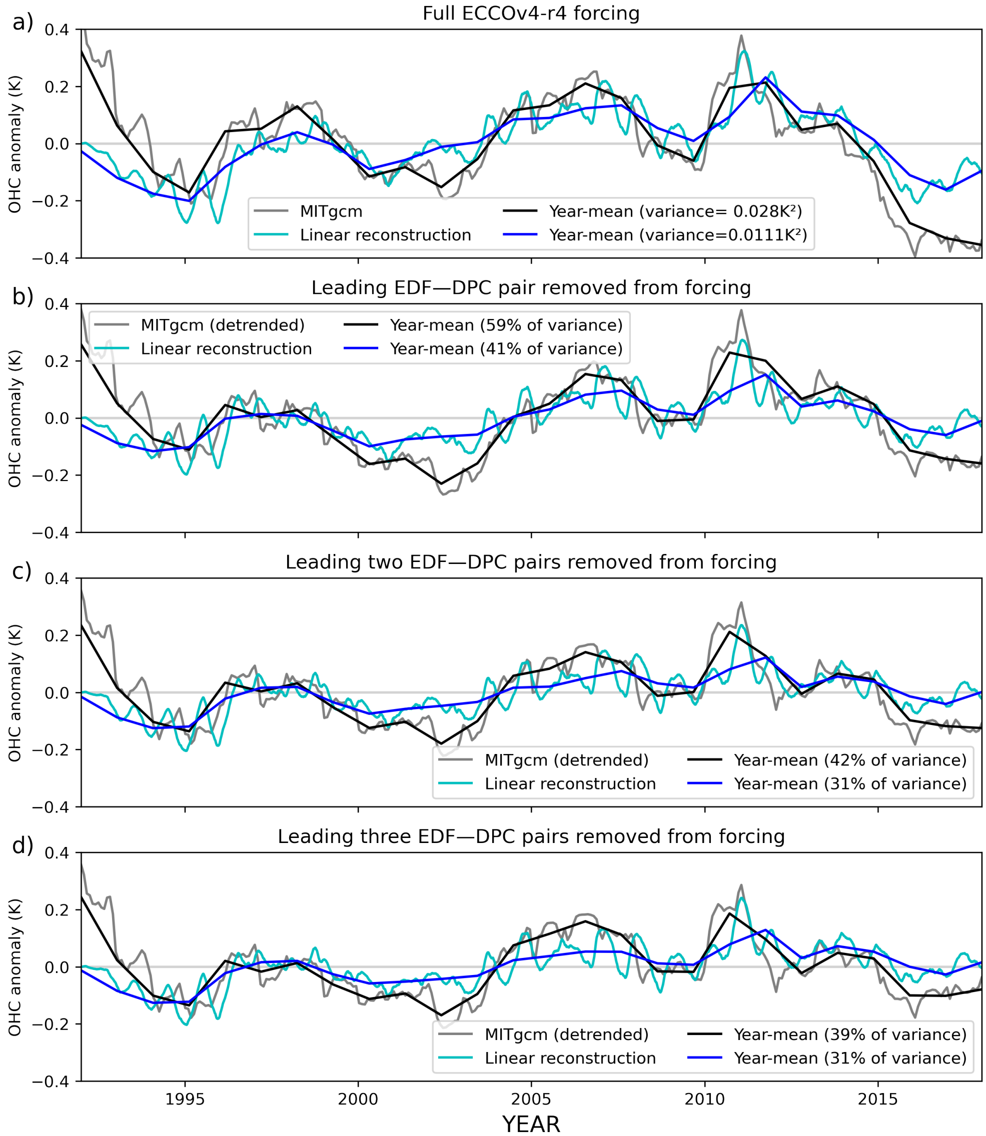

FIG. 12. Same as Figure 11 but for wind stress. Panel a) is the same as Figure 11a and is presented again for comparison.

Removing the leading EDF-DPC pair of HF induces a downward trend in the evolution of SPG HC in the MITgcm (Figure 11b, dotted lines). The absence of this trend in the corresponding linear reconstruction (Figure 11b, blue lines) suggests that it is a nonlinear response of the model to the removal of NAO-like variability, potentially indicating a transition to a different time mean state. Such a drift could arise because the flux-forced configuration of the MITgcm does not adjust heat fluxes with changing upper-ocean temperature. Lohmann et al. (2009) also found a nonlinear response of the circulation to the NAO using modified forcing experiments. While we do not investigate its origins further, if we treat the drift as being a superimposed linear trend and subtract it from the HC response (Figure 11b, solid gray and black lines), we find that subtraction of the first HF EDF–DPC pair results in a 90% reduction in total interannual SPG HC variability in the nonlinear model compared to a roughly 60% reduction in the linear reconstruction. Differences in the effectiveness of the leading EDF-DPC pair in driving variability between linear and nonlinear reconstructions could arise from the trend subtraction and/or additional nonlinearities. The 60% variance reduction in the linear case is also less than the expected reduction of roughly 90% given by σ_1^2 (far left value of left line, Figure 9g); however, some variation about σ_1^2 is expected for variance reductions over finite time intervals such as the ECCOv4-r4 period. Our summary is that removing the leading EDF-DPC pair results in a strong reduction in SPG HC variance in the MITgcm, as also seen in the linearized system, but with an additional trend due to a nonlinear HF response. Additional removal of the second EDF–DPC pair (Figure 11c) leads to a modest additional reduction in QoI variance. While removal of the third pair (Figure 11d) continues to reduce variance in the linear reconstruction, there is roughly a quadrupling of variance in the nonlinear model relative to the case when only two EDF-DPC pairs are removed, suggesting additional nonlinear responses.

For WS (Figure 12), removal of the leading EDF–DPC pair in the nonlinear model simulations also shows qualitative agreement in variance reduction (roughly 40%) with linear reconstructions (roughly 60%), and estimated σ_1^2 (roughly 70%). (Note that the variance contributions attributed to WS and HF when they are removed individually can sum to more than the total variance when there are covariances between those fluxes in time.) Unlike for HF, we do not observe a trend or an increase in variance in the nonlinear model when subtracting one of the leading EDF–DPC pairs. Similar to HF, we conclude that for this quantity of interest, the dominant mechanisms identified under linear assumptions to derive EDF–DPC pairs for WS are effective in the context of a nonlinear ocean GCM.

e. Mechanisms leading to Subpolar Gyre heat content variability

In order to evaluate the mechanisms by which leading EDFs influence the QoI, we make another modification to fluxes in ECCOv4-r4. Rather than removing EDF–DPC pairs, we now add an

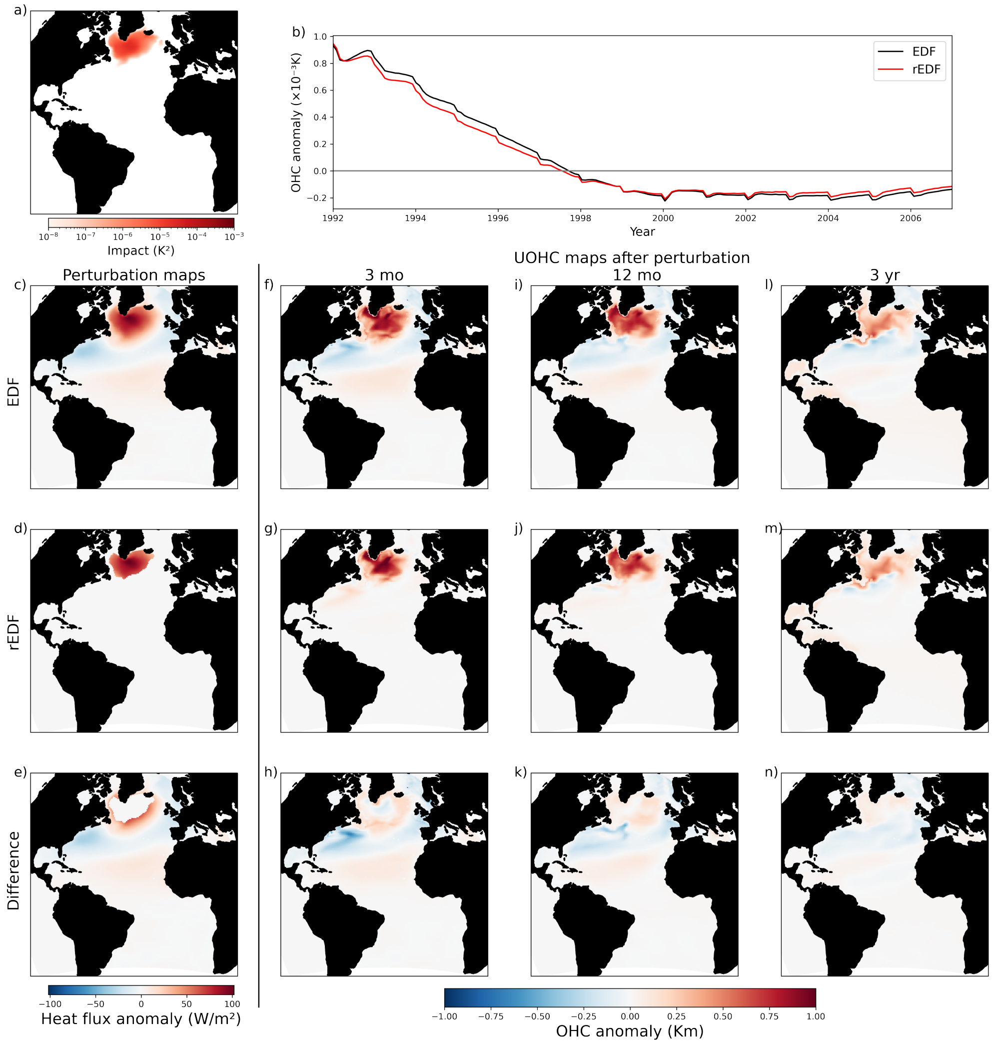

Fig. 13. Forward heat flux perturbation experiments in the MITgcm. Positive heat fluxes correspond to ocean warming. The impact map (a) illustrates the spatial distribution of HF contributions to SPG HC variability under linearized dynamics. Ranking model gridpoints by their impacts allows us to pick a subset of locations within the leading EDF (c) to constitute the leading reduced EDF (rEDF, d), eliminating features (e) that are correlated across atmospheric fluxes but have a small impact on the QoI. Panel (b) shows a high degree of similarity in SPG HC anomaly evolution when perturbed by the leading EDF and leading rEDF. Panels (f,i,l) and (g,j,m) show the evolution of upper ocean heat content anomalies in ECCOv4-r4 after initial 24-hour heat flux perturbations by EDF and rEDF on January 1992; (h,k,n) plot the difference between the two.

initial 24-hour perturbation of fluxes on January 1, 1992 with the spatial pattern of the leading

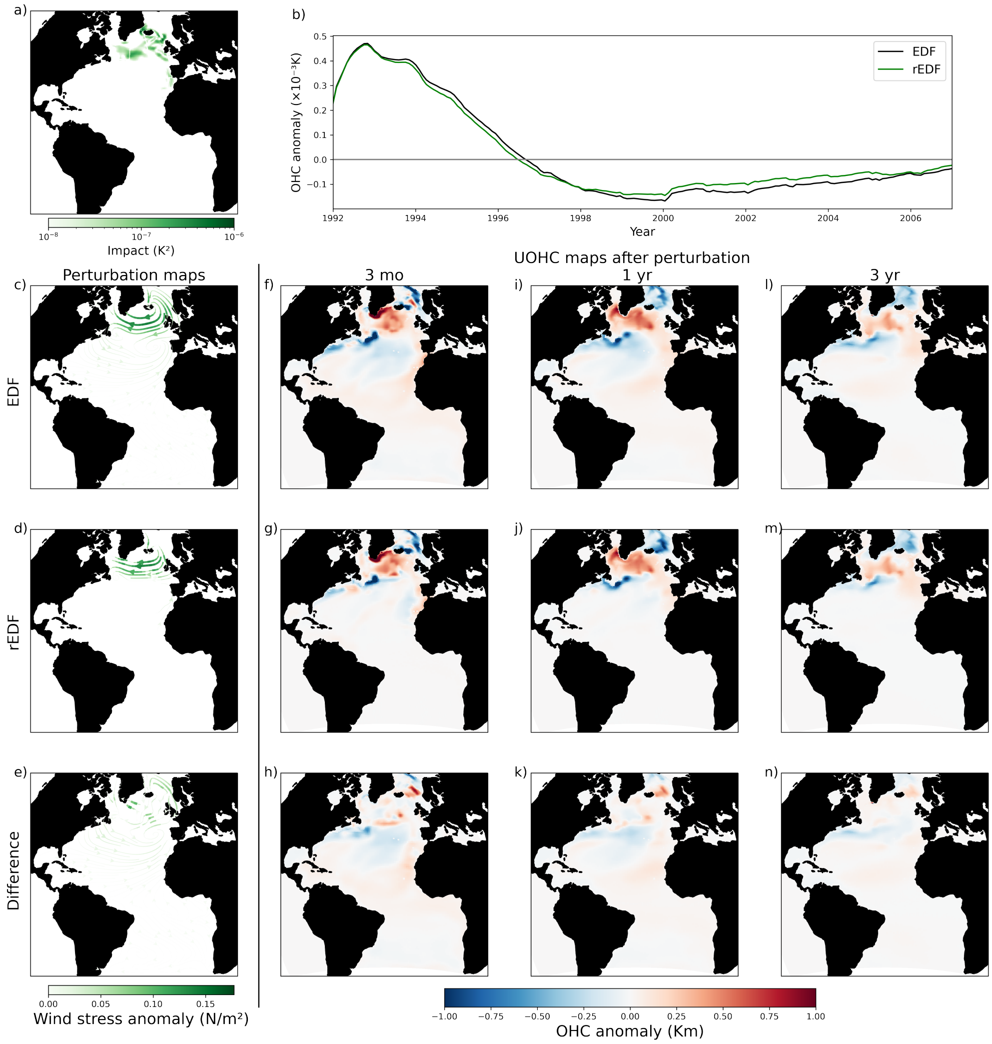

FIG. 14. Same as Figure 13 but for wind stress.

EDF and re-run the state estimate. Anomalies relative to the unperturbed ECCOv4-r4 show how EDF flux perturbations affect the ocean state across space and time.

Not all of the ocean's responses to EDF perturbations necessarily lead to QoI variance (unlike for SOs). This point is illustrated in the simple stochastic system in Section 3, in which EDFs have nonzero values at locations that do not drive QoI variability (specifically, values at locations where sensitivities are zero in the top panels in Figure 2) because fluxes at these locations are correlated with fluxes at other locations that do drive QoI variance. As such, when illustrating pathways of ocean variability, it is helpful to focus on ocean adjustments that cause QoI variance rather than those resulting merely from fluxes correlated with a QoI driver. By ranking surface grid boxes most important for QoI variability using impact maps (Equation (22); Figures 13a and 14a), we define rEDF (reduced EDF) patterns (Figures 13d and 14d) with smaller spatial extents that nevertheless account for 99% of QoI variability. Reduced EDFs are more restricted to the SPG than full EDFs, indicating that contributions within the SPG dominate HC variability there; more distant features in the HF EDF are associated with the tripolar correlation fingerprint of NAO in the North Atlantic (Cayan 1992b,a).

Evolving North Atlantic upper-ocean HC anomalies (integrated over the top 700m) in response to leading EDF and rEDF perturbations illustrate the dominant pathways of fluxes en route to SPG HC variability. As intended, impacts on SPG HC from EDFs and rEDFs are virtually indistinguishable in time (Figures 13b and 14b), but differences between anomalous HC (panels h, k, and n) reveal large-scale evolving patterns in the EDF response, particularly in the subtropical gyre, that do not contribute to SPG HC variance. As such, we focus on upper-ocean heat content anomalies in response to the rEDF (panels g, j, and m).

HC changes due to the HF rEDF perturbation are primarily confined to the SPG over a three year period (Figures 13g, 13j, and 13m), with modest transport into the Labrador Sea and along the tail of the Grand Banks in the Northwest Atlantic. The result (Figure 13, red line) is a warm anomaly in the SPG that decays over several years with small seasonal variations, overshoots to a smaller cooling anomaly, and then decays back to zero. By contrast (Figure 14, red line), SPG HC in response to the WS rEDF perturbation gradually increases, peaking roughly a year after the perturbation, and then (similar to the HF response) decays, overshoots, and decays back to zero. Accompanying this response is cooling northeast and south of the SPG, as well as a rapid initial decrease and gradual recovery in the circulation strength of the SPG (not shown). We note that the WS perturbation acts to oppose time mean patterns of wind stress and wind stress curl over the SPG. These results are consistent with studies attributing 1990s subpolar warming to wind stress changes (Bersch 2002; Lozier et al. 2008; Sarafanov et al. 2008; Häkkinen et al. 2013) and with reductions in the northward penetration of warm subtropical waters under reduced subpolar wind stress curl (Häkkinen et al. 2011; Piecuch et al. 2017) that invoke changes in ocean circulation. We speculate that overshoot behavior in both HF and WS responses results from changes to the density structure and circulation of the SPG and surrounding waters that persist after the dissipation of SPG-averaged HC anomalies, analogous to mechanisms proposed by Desbruyères et al. (2021).

5. Discussion and conclusions

This paper combines constraints from ocean model physics and atmospheric statistics to derive the dominant atmospheric patterns and ocean pathways responsible for driving ocean variability. Leading EDF-DPC pairs maximize ocean variability under assumptions of linear ocean physics and space-time separability of atmosphere-ocean fluxes. These pairs are computed via a dynamics-weighted principal components analysis and recover stochastic optimals and traditional EOFs under limiting conditions; they can thus be seen as a hybrid of "what the ocean wants" to drive variability and "what the ocean gets" from the atmosphere. As expected, these patterns outperform the leading EOFs of atmospheric fluxes for driving ocean variability, even as they account for a smaller fraction of the total flux variance. Applying this approach to the problem of upper-ocean heat content variability in the North Atlantic subpolar gyre, we find that leading EDFs of heat and momentum fluxes (Figures 9 and 10) closely resemble the North Atlantic Oscillation. By re-running the ECCOv4-r4 state estimate, we show that removing leading EDF-DPC pairs is highly effective at reducing SPG HC variability, though a trend in HC response may point to limitations of the linear sensitivity assumption in a flux-forced model. Changes due to heat flux perturbations are consistent with a primarily local, passive ocean response to stochastic variability in the gyre interior, while a delay in the onset of warming due to wind stress fluxes accompanied by nonlocal effects suggests an intermediate role for ocean gyre dynamics.

As noted in Section 1, the NAO has long been established as a source of subpolar gyre heat content variability through both heat fluxes and wind stress (Böning et al. 2006; Lozier et al. 2008; Lohmann et al. 2009; Häkkinen et al. 2011; Zhang and Yan 2017), and our reprisal of its importance may come as no surprise. Nevertheless, we argue that "rediscovering" the NAO serves as a nontrivial proof of concept for the EDF–DPC approach. Just as the center of action of leading EDFs was pulled to the left side of the domain in the simplified 1-D example (Figure 2), we expect that the NAO-like EDF arises from a QoI that coincides geographically with the center of action of the NAO, as well as one that is highly sensitive to wintertime variability. The latter constraint is consistent with the definition of the NAO as the leading mode of atmospheric wintertime variability (Hurrell and Deser 2009). At the same time, we caution that leading modes of sea level pressure are not generally expected to be associated with leading flux EDFs for arbitrary QoIs and regions. It is also instructive to contrast the leading WS EDF (Figure 10a) with the leading WS stochastic optimal

(Figure 8a). The absence of prominent structures along the western coast of Africa suggests that while the Kelvin wave mechanism discussed by Jones et al. (2018) and Stephenson and Sévellec (2021b) is a potential pathway for generating SPG HC variability, it is not a dominant mechanism in practice under recent atmospheric variability.

The EDF-DPC approach can be extended or improved in several ways. We solved for HF and WS EDFs separately and independently found strong correlations with the NAO; however, future approaches could solve for multivariate EDFs across flux types. In addition, using ocean-atmosphere fluxes as boundary conditions may introduce inconsistencies and drifts in perturbed ECCOv4-r4 simulations due to missing turbulent flux feedbacks. An alternative could be instead to compute EDFs for atmospheric variables (air temperature, winds, humidity, etc.), with the caveat that there may be additional covariance relationships among these variables that need to be taken into account. We have made the approximation that the sensitivity is stationary in time, meaning that it depends only on the time lag τ between the QoI and fluxes; while this appears adequate for our purpose, including information about time variations in sensitivities could yield additional information. In this initial implementation, we defined our upper-ocean volume using a uniform depth of 700 m; however, additional insights into the variability of SPG and other water masses might be gained by targeting spatially varying winter mixed layer depths (Buckley et al. 2014, 2015) and/or a QoI defined in isopycnal coordinates. Assuming a fixed 700 m depth also neglects time variations in the depth of SPG mixed layer depth, including across seasons. We hypothesize that defining a QoI based on a density class would further strengthen the preference for atmospheric patterns that dominate in winter time, with a qualitatively similar dominant role for the NAO.

By fusing information from atmospheric statistics and ocean model physics, the EDF–DPC approach inherits potential sources of error from both that we have not attempted to quantify here. Inferring atmospheric statistics from a finite number of samples is a well-studied problem in climate variability and data assimilation (Houtekamer et al. 1998); it may be reasonable to investigate a "rule of thumb" following North et al. (1982) to establish independence criteria for EDFs, or to compute leading EDF–PC pairs in a subset of the time period with available flux data and assess their performance over different intervals. Low-resolution ocean models also have a well-documented host of shortcomings that are inherited through the adjoint sensitivities. The lack of coupling and feedbacks is a limitation of the linearized, forced-ocean perspective: if an atmospheric perturbation changes the ocean state in a way that in turn changes how the ocean responds to future perturbations, then these effects will not be captured by linear sensitivities. The importance of feedbacks might be evaluated, for instance, by applying EDF-like perturbations in a coupled model.

While we have focused on an application for North Atlantic physical oceanography, the EDF–DPC approach is generalizable to a range of applications. Within the framework of forced ocean variability, we expect EDFs to be useful for any QoI whose variability is driven by atmospheric fluxes. Other applications where explicitly recognizing the important role of atmospheric covariances in determining leading drivers of ocean variability include ocean observing system design. For instance, Loose et al. (2020) use adjoint sensitivities as a basis for guiding optimal observations of North Atlantic quantities via a "proxy potential." The work presented here shows that atmospheric conditions most likely to excite ocean stochastic optimals tend to have a large spatial footprint, suggesting that proxy potential might benefit from correlations due to large-scale patterns of variability. Finally, we note that a model adjoint is not required to implement an EDF-DPC approach: while computational costs can be greater, ocean QoI sensitivities can also be estimated via forward perturbation or "Green's function" approaches (e.g., Menemenlis et al. 2005).

Acknowledgments. This work was supported by NASA Physical Oceanography award 80NSSC20K0787. We thank Ichiro Fukumori, Yavor Kostov, and Mark Cane for helpful conversations. Chris Wolfe shared restart files from a long MITgcm integration. Judith Berner stochastically forced the trajectory of this work by providing important insights at unpredictable times. We are grateful for detailed comments from three reviewers that substantially improved the paper. Supercomputing resources were provided by NASA Advanced Supercomputing and Ames Research Center. The adjoint model was generated using TAF (Giering and Kaminski 1998). This material is based upon work supported by the NSF National Center for Atmospheric Research, which is a major facility sponsored by the U.S. National Science Foundation under Cooperative Agreement No. 1852977.

Data availability statement: State estimation output used in our analyses is available from the ECCO project at `www.ecco-group.org/`. The model configuration used for the adjoint model base state is documented with necessary files at `zenodo.org/record/7814839`. Source

code and namelist files for flux-forced ECCOv4-r4 are located in the ECCOv4-r4 directory `/MITgcm/ECCOV4/release4/flux-forced`. Jupyter notebooks necessary to reproduce results from Section 3 are available at `github.com/amrhein/DPCs`. Python code and Jupyter notebooks demonstrating calculation of EDF-DPC pairs and other analyses performed in this paper are available at `github.com/ds4g15/EDF_DPC_paper`. Perturbed ECCOv4-r4 simulations generated for this paper are too large to be retained or publicly archived with available resources; documentation and methods are available from damrhein@ucar.edu.

References

Adamjan, V. M., D. Z. Arov, and M. Kreĭn, 1971: Analytic properties of schmidt pairs for a hankel operator and the generalized schur-takagi problem. *Mathematics of the USSR-Sbornik*, **15 (1)**, 31.

Adcroft, A., C. Hill, J.-M. Campin, J. Marshall, and P. Heimbach, 2004: {O}verview of the formulation and numerics of the {MIT GCM}. *Proceedings of the ECMWF seminar series on Numerical Methods, Recent developments in numerical methods for atmosphere and ocean modelling*, 139–149.

Antoulas, A. C., 2005: *Approximation of large-scale dynamical systems*. SIAM.

Barrier, N., C. Cassou, J. Deshayes, and A.-M. Treguier, 2014: Response of North Atlantic Ocean Circulation to Atmospheric Weather Regimes. *Journal of Physical Oceanography*, **44 (1)**, 179–201, https://doi.org/10.1175/JPO-D-12-0217.1, URL http://journals.ametsoc.org/doi/abs/10.1175/JPO-D-12-0217.1https://journals.ametsoc.org/doi/pdf/10.1175/JPO-D-12-0217.1.

Bersch, M., 2002: North Atlantic Oscillation–induced changes of the upper layer circulation in the northern North Atlantic Ocean. *Journal of Geophysical Research*, **107 (C10)**, 3156, https://doi.org/10.1029/2001JC000901, URL http://doi.wiley.com/10.1029/2001JC000901.

Bersch, M., I. Yashayaev, and K. P. Koltermann, 2007: Recent changes of the thermohaline circulation in the subpolar North Atlantic. *Ocean Dynamics*, **57 (3)**, 223–235, https://doi.org/10.1007/s10236-007-0104-7, arXiv:1011.1669v3.

Böning, C. W., M. Scheinert, J. Dengg, A. Biastoch, and A. Funk, 2006: Decadal variability of subpolar gyre transport and its reverberation in the North Atlantic overturning. *Geophysical*

Research Letters, **33 (21)**, L21S01, https://doi.org/10.1029/2006GL026906, URL http://doi.wiley.com/10.1029/2006GL026906.

Brunton, S. L., and J. N. Kutz, 2022: *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press.

Buckley, M. W., R. M. Ponte, G. Forget, and P. Heimbach, 2014: Low-frequency SST and upper-ocean heat content variability in the North Atlantic. *Journal of Climate*, **27 (13)**, 4996–5018, https://doi.org/10.1175/JCLI-D-13-00316.1.

Buckley, M. W., R. M. Ponte, G. Forget, and P. Heimbach, 2015: Determining the origins of advective heat transport convergence variability in the north atlantic. *Journal of Climate*, **28 (10)**, 3943–3956.

Bugnion, V., C. Hill, and P. H. Stone, 2006a: An Adjoint Analysis of the Meridional Overturning Circulation in a Hybrid Coupled Model. *Journal of Climate*, **19 (15)**, 3751–3767, https://doi.org/10.1175/JCLI3821.1, URL http://journals.ametsoc.org/doi/abs/10.1175/JCLI3821.1.

Bugnion, V., C. Hill, and P. H. Stone, 2006b: An Adjoint Analysis of the Meridional Overturning Circulation in an Ocean Model. *Journal of Climate*, **19 (15)**, 3732–3750, https://doi.org/10.1175/JCLI3787.1, URL http://journals.ametsoc.org/doi/abs/10.1175/JCLI3787.1.

Cayan, D. R., 1992a: Latent and sensible heat flux anomalies over the northern oceans: Driving the sea surface temperature. *Journal of Physical Oceanography*, **22 (8)**, 859–881.

Cayan, D. R., 1992b: Latent and sensible heat flux anomalies over the northern oceans: The connection to monthly atmospheric circulation. *Journal of climate*, **5 (4)**, 354–369.

Chen, W., M. G. Genton, and Y. Sun, 2021: Space-time covariance structures and models. *Annual Review of Statistics and Its Application*, **8**, 191–215.

Chhak, K. C., and A. M. Moore, 2007: The North Atlantic Oscillation as a source of stochastic forcing of the wind-driven ocean circulation. *Dynamics of atmospheres and oceans*, **43 (3-4)**, 151–170.

Chhak, K. C., A. M. Moore, and R. F. Milliff, 2009: Stochastic forcing of ocean variability by the north atlantic oscillation. *Journal of physical oceanography*, **39 (1)**, 162–184.

Czeschel, L., C. Eden, and R. J. Greatbatch, 2012: On the Driving Mechanism of the Annual Cycle of the Florida Current Transport. *Journal of Physical Oceanography*, **42 (5)**, 824–839, https://doi.org/10.1175/JPO-D-11-0109.1, URL http://journals.ametsoc.org/doi/abs/10.1175/JPO-D-11-0109.1.

Czeschel, L., D. P. Marshall, and H. L. Johnson, 2010: Oscillatory sensitivity of Atlantic overturning to high-latitude forcing. *Geophysical Research Letters*, **37 (10)**, n/a–n/a, https://doi.org/10.1029/2010GL043177, URL http://doi.wiley.com/10.1029/2010GL043177, 10/2010GL043177.

Desbruyères, D., L. Chafik, and G. Maze, 2021: A shift in the ocean circulation has warmed the subpolar north atlantic ocean since 2016. *Communications Earth & Environment*, **2 (1)**, 48.

Deser, C., M. A. Alexander, S.-P. Xie, and A. S. Phillips, 2010: Sea surface temperature variability: Patterns and mechanisms. *Annual review of marine science*, **2**, 115–143.

Farrell, B. F., and P. J. Ioannou, 1996: Generalized stability theory. part i: Autonomous operators. *Journal of Atmospheric Sciences*, **53 (14)**, 2025–2040.

Farrell, B. F., and P. J. Ioannou, 2001: Accurate low-dimensional approximation of the linear dynamics of fluid flow. *Journal of the atmospheric sciences*, **58 (18)**, 2771–2789.

Forget, G., J.-M. Campin, P. Heimbach, C. N. Hill, R. M. Ponte, and C. Wunsch, 2015a: {ECCO} version 4: an integrated framework for non-linear inverse modeling and global ocean state estimation.

Forget, G., D. Ferreira, and X. Liang, 2015b: On the observability of turbulent transport rates by Argo: Supporting evidence from an inversion experiment. *Ocean Science*, **11 (5)**, 839–853, https://doi.org/10.5194/os-11-839-2015.

Foukal, N. P., and M. S. Lozier, 2017: Assessing variability in the size and strength of the North Atlantic subpolar gyre. *Journal of Geophysical Research: Oceans*, **122 (8)**, 6295–6308, https://doi.org/10.1002/2017JC012798, URL http://doi.wiley.com/10.1002/2017JC012798.

Frajka-Williams, E., C. Beaulieu, and A. Duchez, 2017: Emerging negative Atlantic Multidecadal Oscillation index in spite of warm subtropics. *Scientific Reports*, **7 (1)**, 1–8, https://doi.org/10.1038/s41598-017-11046-x, URL http://dx.doi.org/10.1038/s41598-017-11046-x.

Frankignoul, C., and K. Hasselmann, 1977: {S}tochastic climate models, part {II} application to sea-surface temperature anomalies and thermocline variability. *Tellus*, **29 (4)**, 289–305.

Fukumori, I., P. Heimbach, R. M. Ponte, C. Wunsch, I. Fukumori, P. Heimbach, R. M. Ponte, and C. Wunsch, 2018: A Dynamically Consistent, Multi-Variable Ocean Climatology. *Bulletin of the American Meteorological Society*, BAMS–D–17–0213.1, https://doi.org/10.1175/BAMS-D-17-0213.1, URL http://journals.ametsoc.org/doi/10.1175/BAMS-D-17-0213.1.

Fukumori, I., O. Wang, and I. Fenty, 2021: Causal mechanisms of sea level and freshwater content change in the beaufort sea. *Journal of Physical Oceanography*, **51 (10)**, 3217–3234.

Fukumori, I., O. Wang, I. Fenty, G. Forget, P. Heimbach, and R. M. Ponte, 2017: ECCO Version 4 Release 3. Tech. rep. URL https://dspace.mit.edu/handle/1721.1/110380.

Fukumori, I., O. Wang, W. Llovel, I. Fenty, and G. Forget, 2015: A near-uniform fluctuation of ocean bottom pressure and sea level across the deep ocean basins of the Arctic Ocean and the Nordic Seas. *Progress in Oceanography*, **134**, 152–172, https://doi.org/10.1016/j.pocean.2015.01.013, URL http://dx.doi.org/10.1016/j.pocean.2015.01.013http://linkinghub.elsevier.com/retrieve/pii/S0079661115000245.

Giering, R., and T. Kaminski, 1998: {R}ecipes for adjoint code construction. *ACM Transactions on Mathematical Software (TOMS)*, **24 (4)**, 437–474.

Häkkinen, S., P. B. Rhines, and D. L. Worthen, 2011: Warm and saline events embedded in the meridional circulation of the northern North Atlantic. *Journal of Geophysical Research: Oceans*, **116 (C3)**.

Häkkinen, S., P. B. Rhines, and D. L. Worthen, 2013: Northern north atlantic sea surface height and ocean heat content variability. *Journal of Geophysical Research: Oceans*, **118 (7)**, 3670–3678.

Hasselmann, K., 1976: {S}tochastic climate models part {I}. {T}heory. *Tellus*, **28 (6)**, 473–485.

Hasselmann, K., 1993: Optimal fingerprints for the detection of time-dependent climate change. *Journal of Climate*, **6 (10)**, 1957–1971.

Houtekamer, P. L., H. L. Mitchell, P. L. Houtekamer, and H. L. Mitchell, 1998: Data Assimilation Using an Ensemble Kalman Filter Technique. *Monthly Weather Re-*

view, **126 (3)**, 796–811, https://doi.org/10.1175/1520-0493(1998)126⟨0796:DAUAEK⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0493{\%}281998{\%}29126{\%}3C0796{\%}3ADAUAEK{\%}3E2.0.CO{\%}3B2.

Hurrell, J. W., and C. Deser, 2009: North Atlantic climate variability: The role of the North Atlantic Oscillation. *Journal of Marine Systems*, **78 (1)**, 28–41, https://doi.org/10.1016/j.jmarsys.2008.11.026, URL http://linkinghub.elsevier.com/retrieve/pii/S0924796309000815.

Jones, D., G. Forget, B. Sinha, S. Josey, E. Boland, A. Meijers, and E. Shuckburgh, 2018: Local and Remote Influences on the Heat Content of the Labrador Sea: an Adjoint Sensitivity Study. *Journal of Geophysical Research: Oceans*, 1–22, https://doi.org/10.17605/OSF.IO/KC2U9, URL https://eartharxiv.org/kc2u9/.

Kim, W. M., S. Yeager, P. Chang, and G. Danabasoglu, 2016: Atmospheric conditions associated with Labrador Sea deep convection: New insights from a case study of the 2006/07 and 2007/08 winters. *Journal of Climate*, **29 (14)**, 5281–5297, https://doi.org/10.1175/JCLI-D-15-0527.1.

Kleeman, R., 2008: Stochastic theories for the irregularity of enso. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **366 (1875)**, 2509–2524.

Kleeman, R., and A. M. Moore, 1997: A theory for the limitation of enso predictability due to stochastic atmospheric transients. *Journal of the atmospheric sciences*, **54 (6)**, 753–767.

Köhl, A., and D. Stammer, 2004: Optimal Observations for Variational Data Assimilation. *Journal of Physical Oceanography*, **34 (3)**, 529–542, https://doi.org/10.1175/2513.1, URL http://journals.ametsoc.org/doi/abs/10.1175/2513.1.

Kostov, Y., H. L. Johnson, and D. P. Marshall, 2019: AMOC sensitivity to surface buoyancy fluxes : the role of air - sea feedback mechanisms. *Climate Dynamics*, **1 (2014)**, https://doi.org/10.1007/s00382-019-04802-4, URL https://doi.org/10.1007/s00382-019-04802-4.

Kostov, Y., M.-J. Messias, H. Mercier, H. L. Johnson, and D. P. Marshall, 2022: Fast mechanisms linking the labrador sea with subtropical atlantic overturning. *Climate Dynamics*, 1–26.

Kostov, Y., and Coauthors, 2021: Distinct sources of interannual subtropical and subpolar atlantic overturning variability. *Nature Geoscience*, **14 (7)**, 491–495.

Kushnir, Y., W. Robinson, I. Bladé, N. Hall, S. Peng, and R. Sutton, 2002: Atmospheric gcm response to extratropical sst anomalies: Synthesis and evaluation. *Journal of Climate*, **15 (16)**, 2233–2256.

Large, W. G., and S. G. Yeager, 2004: *{D}iurnal to decadal global forcing for ocean and sea-ice models: the data sets and flux climatologies*. National Center for Atmospheric Research Boulder.

Lohmann, K., H. Drange, and M. Bentsen, 2009: Response of the North Atlantic subpolar gyre to persistent North Atlantic oscillation like forcing. *Climate dynamics*, **32 (2)**, 273–285.

Loose, N., P. Heimbach, H. Pillar, and K. H. Nisancioglu, 2020: Quantifying dynamical proxy potential through shared adjustment physics in the north atlantic. *Journal of Geophysical Research: Oceans*, **125 (9)**, e2020JC016 112.

Lorenz, E. N., 1956: *Empirical orthogonal functions and statistical weather prediction*, Vol. 1. Massachusetts Institute of Technology, Department of Meteorology Cambridge.

Lozier, M. S., S. Leadbetter, R. G. Williams, V. Roussenov, M. S. C. Reed, and N. J. Moore, 2008: The Spatial Pattern and Mechanisms of Heat-Content Change in the North Atlantic. *Science*, **319 (5864)**, 800–803, https://doi.org/10.1126/science.1146436, URL http://www.sciencemag.org/cgi/doi/10.1126/science.1146436, arXiv:1011.1669v3.

MacGilchrist, G. A., H. L. Johnson, C. Lique, and D. P. Marshall, 2021: Demons in the north atlantic: Variability of deep ocean ventilation. *Geophysical Research Letters*, **48 (9)**, e2020GL092 340.

Marotzke, J., R. Giering, K. Q. Zhang, D. Stammer, C. Hill, and T. Lee, 1999: Construction of the adjoint mit ocean general circulation model and application to atlantic heat transport sensitivity. *Journal of Geophysical Research: Oceans*, **104 (C12)**, 29 529–29 547.

Marshall, J., A. Adcroft, C. Hill, L. Perelman, and C. Heisey, 1997: {A} finite-volume, incompressible {N}avier {S}tokes model for studies of the ocean on parallel computers. *Journal of Geophysical Research: Oceans (1978–2012)*, **102 (C3)**, 5753–5766.

Marshall, J., and Coauthors, 2001: North atlantic climate variability: phenomena, impacts and mechanisms. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, **21 (15)**, 1863–1898.

Mazloff, M. R., 2012: On the sensitivity of the drake passage transport to air-sea momentum flux. *Journal of Climate*, **25 (7)**, 2279–2290, https://doi.org/10.1175/JCLI-D-11-00030.1.

Menemenlis, D., I. Fukumori, and T. Lee, 2005: Using green's functions to calibrate an ocean general circulation model. *Monthly weather review*, **133 (5)**, 1224–1240.

Moore, A. M., H. G. Arango, E. Di Lorenzo, B. D. Cornuelle, A. J. Miller, and D. J. Neilson, 2004: A comprehensive ocean prediction and analysis system based on the tangent linear and adjoint of a regional ocean model. *Ocean Modelling*, **7 (1-2)**, 227–258.

Moore, A. M., J. Fiechter, and C. A. Edwards, 2022: A linear stochastic emulator of the california current system using balanced truncation. *Ocean Modelling*, **174**, 102 023.

Moore, A. M., and R. Kleeman, 1999: Stochastic forcing of enso by the intraseasonal oscillation. *Journal of Climate*, **12 (5)**, 1199–1220.

Moore, A. M., and Coauthors, 2006: Optimal forcing patterns for coupled models of enso. *Journal of climate*, **19 (18)**, 4683–4699.

Moore, B., 1981: Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE transactions on automatic control*, **26 (1)**, 17–32.

North, G. R., T. L. Bell, and R. F. Cahalan, 1982: Sampling Errors in the Estimation of Empirical Orthogonal Funtions. URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0493(1982)110{\%}3C0699:SEITEO{\%}3E2.0.CO;2, 699–706 pp., https://doi.org/10.1175/1520-0493(1982)110⟨0699:SEITEO⟩2.0.CO;2.

Oldenburg, D., R. C. Wills, K. C. Armour, L. Thompson, and L. C. Jackson, 2021: Mechanisms of low-frequency variability in north atlantic ocean heat transport and amoc. *Journal of Climate*, **34 (12)**, 4733–4755.

Ortega, P., J. Robson, R. T. Sutton, and M. B. Andrews, 2017: Mechanisms of decadal variability in the labrador sea and the wider north atlantic in a high-resolution climate model. *Climate Dynamics*, **49**, 2625–2647.

Piecuch, C. G., R. M. Ponte, C. M. Little, M. W. Buckley, and I. Fukumori, 2017: Mechanisms underlying recent decadal changes in subpolar North Atlantic Ocean heat content. *Journal of Geo-*

physical Research: Oceans*, **122 (9)**, 7181–7197, https://doi.org/10.1002/2017JC012845, URL http://doi.wiley.com/10.1002/2016JC012264http://doi.wiley.com/10.1002/2017JC012845.

Pillar, H. R., P. Heimbach, H. L. Johnson, and D. P. Marshall, 2016: Dynamical Attribution of Recent Variability in Atlantic Overturning. *Journal of Climate*, **29 (9)**, 3339–3352, https://doi.org/10.1175/JCLI-D-15-0727.1, URL https://journals.ametsoc.org/doi/pdf/10.1175/JCLI-D-15-0727.1http://journals.ametsoc.org/doi/10.1175/JCLI-D-15-0727.1.

Rowley, C. W., 2005: Model reduction for fluids, using balanced proper orthogonal decomposition. *International Journal of Bifurcation and Chaos*, **15 (03)**, 997–1013.

Sarafanov, A., A. Falina, A. Sokov, and A. Demidov, 2008: Intense warming and salinification of intermediate waters of southern origin in the eastern subpolar north atlantic in the 1990s to mid-2000s. *Journal of Geophysical Research: Oceans*, **113 (C12)**.

Stephenson, D., and F. Sévellec, 2021a: The active and passive roles of the ocean in generating basin-scale heat content variability. *Geophysical Research Letters*, **48 (19)**, e2020GL091 874.

Stephenson, D., and F. Sévellec, 2021b: Dynamical attribution of north atlantic interdecadal predictability to oceanic and atmospheric turbulence under diagnosed and optimal stochastic forcing. *Journal of Climate*, **34 (22)**, 9153–9179.

Stommel, H., 1979: {D}etermination of water mass properties of water pumped down from the {E}kman layer to the geostrophic flow below. *Proceedings of the National Academy of Sciences*, **76 (7)**, 3051–3055.

Tesdal, J.-E., R. P. Abernathey, J. I. Goes, A. L. Gordon, and T. W. N. Haine, 2018: Salinity Trends within the Upper Layers of the Subpolar North Atlantic. *Journal of Climate*, **31 (7)**, 2675–2698, https://doi.org/10.1175/JCLI-D-17-0532.1, URL http://journals.ametsoc.org/doi/10.1175/JCLI-D-17-0532.1.

Vidard, A., P.-A. Bouttier, and F. Vigilant, 2015: Nemotam: tangent and adjoint models for the ocean modelling platform nemo. *Geoscientific Model Development*, **8 (4)**, 1245–1257.

Wills, R. C., K. C. Armour, D. S. Battisti, and D. L. Hartmann, 2019: Ocean–atmosphere dynamical coupling fundamental to the atlantic multidecadal oscillation. *Journal of Climate*, **32 (1)**, 251–272.

Wolfe, C. L., P. Cessi, and B. D. Cornuelle, 2017: An intrinsic mode of interannual variability in the indian ocean. *Journal of Physical Oceanography*, **47 (3)**, 701–719.

Wunsch, C., and P. Heimbach, 2007: {P}ractical global oceanic state estimation. *Physica D: Nonlinear Phenomena*, **230 (1)**, 197–208, https://doi.org/10.1016/j.physd.2006.09.040.

Xu, T., M. Newman, M. A. Alexander, and A. Capotondi, 2024: A forecast test for reducing dynamical dimensionality of model emulators. *Journal of Advances in Modeling Earth Systems*, **16 (1)**.

Yeager, S., and Coauthors, 2021: An outsized role for the labrador sea in the multidecadal variability of the atlantic overturning circulation. *Science Advances*, **7 (41)**, eabh3592.

Zanna, L., and E. Tziperman, 2008: Optimal surface excitation of the thermohaline circulation. *Journal of Physical Oceanography*, **38 (8)**, 1820–1830.

Zavala-Garay, J., A. Moore, C. Perez, and R. Kleeman, 2003: The response of a coupled model of enso to observed estimates of stochastic forcing. *Journal of Climate*, **16 (17)**, 2827–2842.

Zhang, W., and X.-H. Yan, 2017: The subpolar north atlantic ocean heat content variability and its decomposition. *Scientific Reports*, **7 (1)**, 1–8.