New Methods for Estimating Ocean Eddy Heat Transport Using Satellite Altimetry

SHANE R. KEATING, ANDREW J. MAJDA, AND K. SHAFER SMITH

Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, New York

ABSTRACT

Attempts to monitor ocean eddy heat transport are strongly limited by the sparseness of available observations and the fact that heat transport is a quadratic, sign-indefinite quantity that is particularly sensitive to unresolved scales. In this article, a suite of stochastic filtering strategies for estimating eddy heat transport are tested in idealized two-layer simulations of mesoscale oceanic turbulence at high and low latitudes under a range of observation scenarios. A novel feature of these filtering strategies is the use of computationally inexpensive stochastic models to forecast the underlying nonlinear dynamics. The stochastic model parameters can be estimated by regression fitting to climatological energy spectra and correlation times or by adaptively learning these parameters "on-the-fly" from the observations themselves.

The authors show that, by extracting high-wavenumber information that has been aliased into the low wavenumber band, "stochastically super-resolved" velocity fields with a nominal resolution increase of a factor of 2 or more can be derived. Observations of the upper-layer streamfunction are projected onto an empirical orthogonal function basis for the vertical structure to produce filtered estimates for both upper- and lower-layer streamfunctions and hence net heat transport. The resulting time-mean poleward eddy heat transport is significantly closer to the true value when compared with standard estimates based upon optimal interpolation. By contrast, the temporal variability of the heat transport is underestimated because of poor temporal resolution. Implications for estimating poleward eddy heat transport using current and next-generation altimeters are discussed.

1. Introduction

The poleward redistribution of heat by the ocean is a primary constituent of the global heat balance, comparable in magnitude to that in the atmosphere, yet it remains one of the most poorly constrained features of the ocean general circulation. The principal uncertainty is the contribution from ocean eddies, which can drive a rectified flux of heat due to temporal correlations between buoyancy fluctuations and the eddy velocity. Hall and Bryden (1982) estimated that the eddy-driven flux contributes as much as 25% of the time-mean transport across 24°N and was the largest source of uncertainty in their study. Likewise, eddy-permitting ocean models (Jayne and Marotzke 2002) and data-assimilation products (Volkov et al. 2008) indicate a significant eddy contribution to the time-mean transport in the Antarctic Circumpolar Current, the Kuroshio, and the tropics, and up to one-third of the global heat transport variability. However, efforts to constrain poleward eddy heat transport are limited by the difficulty in obtaining observations of the full water column with sufficient spatial and temporal detail to resolve the eddy field (Wunsch 1999; Ganachaud and Wunsch 2000), while estimates of the time-mean heat transport based upon one-time hydrographic sections are valid only where eddy variability is small compared to the total heat transport variability (Volkov et al. 2008).

A picture of the *surface* ocean eddy field is provided by satellite measurements of sea surface height anomalies relative to the geoid. Under the assumption of geostrophic balance, the current $\mathbf{u}$ at the surface is given by

$$f\mathbf{u} = g\hat{z} \times \nabla\eta, \tag{1.1}$$

where f is the Coriolis parameter, g is gravitational acceleration, and η is the sea surface height anomaly. As shown in the appendices, however, the *subsurface* geostrophic

Corresponding author address: Shane R. Keating, Courant Institute of Mathematical Sciences, New York University, 251 Mercer St., New York, NY 10012.
E-mail: skeating@cims.nyu.edu

current has additional contributions from the horizontal gradient of the hydrostatic density field. Thus, in the absence of a model for the interior pressure field, the full three-dimensional eddy velocity cannot be determined from observations of the surface geostrophic current alone (Wunsch and Stammer 1998). Consequently, recent attempts to calculate net heat transport from satellite altimetry have incorporated estimates of the vertical structure provided by ocean sections or profiling floats (Roemmich and Gilson 2001; Qiu and Chen 2005). Even so, current-generation altimetry maps are capable of resolving only the largest eddy scales and can depend sensitively on the kinetic energy spectrum of the unresolved flow (Bennett 1984; Poje et al. 2010; Keating et al. 2011). In particular, turbulent mixing in regions with an energetic submesoscale, such as the subtropics (Tulloch et al. 2011), is strongly resolution dependent, and so estimates of eddy heat transport based upon altimetry in these regions should be viewed with some caution.

In this article, we examine novel methods for estimating eddy heat transport using sparse, noisy observations of the upper-ocean velocity field. Our "truth signal" is provided by numerical simulations of the two-layer Phillips model under parameter regimes typifying baroclinic turbulence at low and high latitudes. We study a range of observation scenarios, allowing us to explore the interplay of eddy length and time scales with the spatiotemporal resolution capability of satellite observations. Although the Phillips model is a highly idealized representation of oceanic turbulence, these scenarios nonetheless provide stringent test cases incorporating several key challenges to any filtering strategy, in particular: *sparseness in vertical space*, due to the fact that observations of the upper-layer velocity field alone are not sufficient to calculate the heat transport; *sparseness in horizontal space*, since much of the ocean's eddy kinetic energy resides at scales below the resolution of current-generation altimetry products; and *sparseness in time*, because the satellite orbit configuration generally limits the temporal resolution capability of altimeters to no more than an eddy turnover time. Finally, even in this minimal model, there is an enormous range of interacting scales as well as stiff dynamics due to small Rossby radius, a regime that poses severe practical difficulties for contemporary ensemble-based filtering approaches (Harlim and Majda 2010).

As we shall show, these issues pose a severe challenge to attempts to estimate the turbulent eddy field in the Phillips model, and, by extension, the real ocean. The difficulty is compounded by the fact that eddy heat transport is a sign-indefinite quadratic quantity, and so is particularly sensitive to model error at unresolved scales. In this study, we filter sparse observations of the upper ocean through a dynamical forecast model to obtain an estimate of the eddy transport. Our approach is similar to that of Harlim and Majda (2010), who tested a suite of turbulent filtering algorithms in quasigeostrophic simulations mimicking baroclinic turbulence in the atmosphere and the ocean, with the latter scenario providing a particularly rigorous test case with small Rossby radius and stiff model dynamics. A powerful feature of these filtering strategies is the use of cheap stochastic models to forecast the underlying nonlinear dynamics. This "judicious" model error avoids the so-called curse of dimensionality while preserving the essential structure of the turbulent signal through the use of a *mean stochastic model* that is regression fitted to reproduce climatological energy spectra and correlation times. Harlim and Majda (2010) also tested an alternative scheme, the *stochastic parameterized extended Kalman filter*, in which the forecast model parameters are learned adaptively from the observations by incorporating them into the state vector and obtaining estimates using the usual filtering methodology. Both of these simple stochastic filtering strategies exhibited high skill in the atmospheric and oceanic regimes compared with contemporary ensemble-based data assimilation schemes, and have been extensively validated in a number of stringent idealized test problems (Gershgorin et al. 2010a,b).

The present study focuses on the difficult but realistic ocean case where much of the eddy kinetic energy lies below the scale of resolution of the observations. In contrast to Harlim and Majda (2010), who sought to reconstruct turbulent spectra and streamfunctions from observations of the barotropic velocity field, we will principally be interested in the poleward heat transport in baroclinic flows—a more challenging quantity to constrain, even in the time mean. We utilize an algorithm developed by Majda and Grote (2007) and Harlim and Majda (2008b) that actually exploits the spatial sparseness of the observations by utilizing the fact that a coarse observation network will alias high-wavenumber information into the resolved wave band. By appropriately filtering sparse observations to extract this information, one can derive *stochastically super-resolved* velocity fields with a nominal resolution increase of a factor of 2 or more, leading to considerable improvement in the eddy transport estimate. Likewise, the vertical sparseness of satellite observations is addressed by modeling the vertical structure of the flow using empirical orthogonal functions (EOFs), which project the layer streamfunctions onto a basis set aligned with the maximum and minimum energy modes (e.g., De Mey and Robinson 1987; Gavart and De Mey 1997). Observations of the upper-layer streamfunction are then

projected onto each of the EOFs to produce filtered estimates for both upper- and lower-layer streamfunctions, and, hence, the net heat transport.

In section 2 the Phillips model is discussed in the context of two parameter regimes corresponding to ocean turbulence at low and high latitudes. The poleward eddy heat transport is calculated under a range of spatiotemporally sparse sampling scenarios and compared with the "true" heat transport obtained from the fully resolved model simulations. In section 3 we describe a suite of filtering strategies for vertically, horizontally, and temporally sparse satellite observations. The results of these filtering strategies are presented in section 4. Finally, in section 5 we summarize our conclusions and discuss potential issues associated with the application of these techniques to the real ocean.

2. Poleward eddy heat transport in the Phillips model

The Phillips (1954) model consists of an advection-diffusion equation for the potential vorticity q_i in each of two immiscible layers with fractional layer thicknesses $d_1 = H_1/H_0$ and $d_2 = H_2/H_0$, densities $\rho_2 > \rho_1$, and mean zonal velocities $U_1 > U_2$:

$$\partial_t q_i + J(\psi_i, q_i) + U_i \partial_x q_i + \Pi_i \partial_x \psi_i = -\delta_{i2} r \nabla_H^2 \psi_i, \quad i = 1, 2, \tag{2.1}$$

and an inversion relation that relates the potential vorticity (PV) to the streamfunction ψ_i:

$$q_i = \nabla^2 \psi_i + (R^2 d_i)^{-1}(\psi_{3-i} - \psi_i), \quad R = \sqrt{g' H_0}/f_0, \tag{2.2}$$

where $k_D = R^{-1}(d_1 d_2)^{-1/2}$ is the baroclinic deformation wavenumber (or inverse deformation radius), $f = f_0 + \beta y$ is the Coriolis parameter, $g' = g(1 - \rho_1/\rho_2)$ is the reduced gravity of the interface, and r is the bottom drag, which only acts on the lower layer. The mean meridional PV gradient is $\Pi_i = \beta + k_D^2 d_{3-i} U_0$, where $U_0 = U_1 - U_2$ is the mean shear.

The numerical model is periodic in the horizontal, uses a dealiased, pseudospectral method to compute the nonlinear terms, and dissipates enstrophy with an exponential cutoff filter [see Smith et al. (2002); Smith and Vallis (2002) for details]. We set the fractional layer thicknesses to be $d_1 = 0.2$, $d_2 = 0.8$, consistent with a typical pycnocline depth of 1 km and total ocean depth of 5 km, although in the context of this crude two-layer model the correspondence should not be interpreted too broadly [for a more in-depth discussion of typical

pycnocline depths in the ocean, see, e.g., Gnanadesikan (1999)]. The deformation radius in the ocean varies from about 10 km at the highest latitudes to a few hundred kilometers in the subtropics. A reasonable scale separation between the deformation scale and the domain scale is desired in order to avoid domain-scale interactions in our simulations and allow for some inverse cascade. To that end we allow the domain scale to vary with the deformation radius, but hold their ratio fixed so that the deformation wavelength $\lambda = 2\pi/k_D$ is $^1/_{10}$ times the size of the domain.

The remaining physical quantities in the model can be expressed in terms of two nondimensional parameters: the supercriticality parameter $\tilde{\beta} = \beta/d_1 U_0 k_D^2$, which determines the strength of the baroclinic instability, and the nondimensionalized bottom drag $\tilde{r} = r/k_D U_0$. We consider two contrasting parameter regimes for $\tilde{\beta}$ and $\tilde{r}$:

- Low latitudes: Observations and numerical models (e.g., Richards et al. 2006) show nearly supercritical flow at low latitudes, forming zonal jets with widths slightly larger than the deformation scale. In this case, the numerical model requires only a small drag ($\tilde{r} \ll 1$) to absorb the energy injected by baroclinic instability, since the eddy scale is effectively set by β. Specifically we choose $\tilde{\beta} = 0.9$ and $\tilde{r} = 0.3$, and take $\lambda = 320$ km, so that the domain size is 3200 km.
- High latitudes: Linear stability analysis (e.g., Smith 2007) indicates strong supercriticality at high latitudes and a significant inverse cascade. However, observations show high-latitude eddies at most an order of magnitude larger than the deformation scale (e.g., Stammer 1998; Tulloch et al. 2011) and not particularly anisotropic, suggesting that the cascade is arrested by some mechanism other than β. Here, we use the drag to set the eddy scale to be near the observed scale. Specifically, we set $\tilde{\beta} = 0.2$ and $\tilde{r} = 0.9$, and choose $\lambda = 80$ km, so that the domain scale is 800 km.

Time-averaged streamfunctions for each case are plotted in Figs. 1 and 2. The model parameters for these two cases are summarized in Table 1.

The spatial and temporal resolution capability of altimetric observations depends upon the number of satellites used, their orbital configurations, and the subsequent smoothing and interpolation of the raw sea level measurements. Determining the effective resolution of the resulting processed velocity fields—by which we mean the range of frequencies and wavelengths that can be distinguished on a given spatiotemporal mesh (i.e., the Nyquist wave band)—is a nontrivial problem (see, e.g., Wunsch 1989; Greenslade et al. 1997; Ducet et al. 2000; Chelton et al. 2001). Tai (2004) demonstrates that the "midpoint grid" (the approximately rectilinear grid of
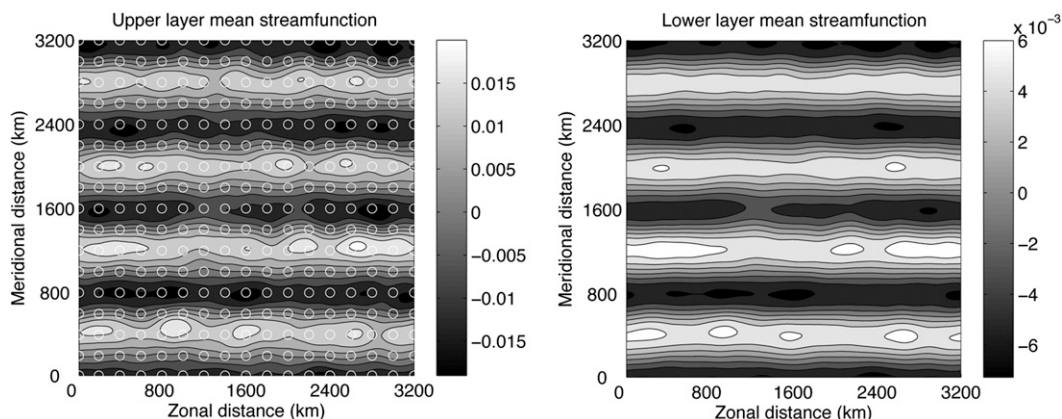
FIG. 1. Upper- and lower-layer streamfunctions for the low-latitude case. White circles show approximate current-generation satellite spatial resolution capability. The streamfunctions are normalized by $\mathcal{U}L_{\text{dom}}$, where $\mathcal{U}$ is the rms velocity and L_{dom} is the domain scale.

points located midway between ground-track crossover points) has the same resolution as the full sample set. Thus, for simplicity of exposition, we adopt the expedient of simultaneously sampling the velocity field on a regular $2N \times 2N$ spatial grid once every observation time, where N is the maximum resolved (Nyquist) wavenumber. The observation grid size and time step are chosen to coincide with the approximate correlation length and time scales of the satellite observations.

The spatial resolution capabilities in the meridional and zonal directions are optimal at high latitudes, which is also where the deformation radius and observed eddy scale are smallest. As a reasonable estimate, we take the grid spacing of observations to be approximately 100 km at high latitudes and 200 km at low latitudes. This corresponds to $N = 4$ and $N = 8$ at low and high latitudes, respectively. Since the deformation wavenumber is $k_D = 10$ in each parameter regime, these cases provide difficult, but

plausible observation scenarios. This is illustrated in Fig. 3, which shows the spectral ranges corresponding to $N = 4$ and $N = 8$ observational networks, superimposed upon the normalized energy spectrum at low and high latitudes. In both high- and low-latitude cases, the $N = 8$ network captures most of the energy, whereas the $N = 4$ network fails to resolve the most energetic wavenumbers. We will also consider networks with $N = 16$ as an indication of the possible gains that might be expected from future altimetry products such as the proposed Surface Water Ocean Topography mission (e.g., Durand et al. 2010).

The decorrelation time scales of altimetric maps are set to be 15 days outside a latitudinal band that extends 10° either side of the equator (Ducet et al. 2000). This roughly corresponds to the turnover time for mesoscale eddies at these latitudes (Stammer 1997). Thus, we set both the eddy turnover time (defined as $T_{\text{eddy}} = 2\pi Z^{-1/2}$, where Z is the time-averaged total enstrophy) and the
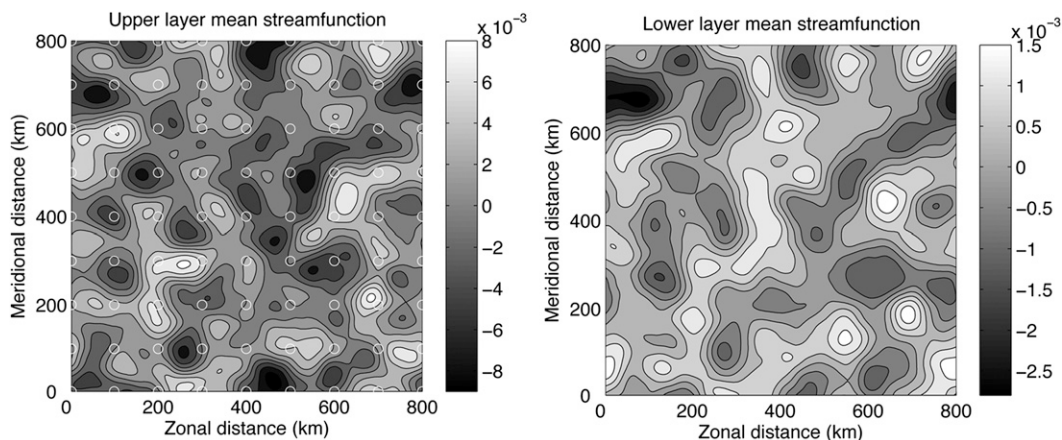

FIG. 2. As in Fig. 1, but for the high-latitude case.

Physical parameter	Low latitude	High latitude
Supercriticality parameter	0.9	0.2
Bottom drag	0.3	0.9
Domain scale (km)	3200	800
Deformation wavelength (km)	320	80
Observation resolution (km)	200	100
Deformation wavenumber	10	10
Nyquist wavenumber	8	4
Eddy turnover time (day)	15	15
Observation time (day)	15	15

observational time step to be 15 days in each of our parameter regimes. Figure 3 shows the decorrelation time for each horizontal Fourier mode $\hat{\psi} = \hat{\psi}_{kl}(t)$:

$$T = \langle|\hat{\psi}|^2\rangle_t^{-1} \mathrm{Re} \int_0^\infty \langle\hat{\psi}(t)\hat{\psi}*(t+\tau)\rangle_t \, d\tau. \qquad (2.3)$$

The thick contour shows the observation time scale, so that modes outside this contour decay rapidly compared with the observation time. As can be seen, the spectral range corresponding to the $N = 8$ observation network captures most of the modes that remain correlated between observation times. By contrast, the $N = 4$ observation network resolves only the most slowly evolving modes.

Neglecting the effects of salinity and assuming a linear equation of state, temperature in the ocean is proportional to buoyancy. Moreover, in the two-layer quasigeostrophic model, buoyancy is proportional to the difference of the upper and lower streamfunctions $\psi_1 - \psi_2$ (see e.g., Vallis 2006), and its poleward flux is directly related to the flux of potential vorticity, as can be seen by making use of (2.2) and introducing the baroclinic streamfunction $\tau = \sqrt{d_1 d_2}(\psi_1 - \psi_2)$:

$$\langle v_1 \tau\rangle = -\left(\frac{d_1}{d_2}\right)^{1/2} k_D^{-2} \langle v_1 q_1\rangle, \qquad (2.4)$$

where $\langle\cdot\rangle$ denotes a two-dimensional spatial average, and the relative vorticity flux $\langle v_1 \nabla_H^2 \psi_1\rangle$ vanishes in a horizontally periodic domain. The heat flux spectra (normalized by the time average) at low and high latitudes are shown in Fig. 3. The thick contour in the low-latitude case indicates where the heat flux contribution changes from negative at low wavenumbers to positive at high wavenumbers (there is no such sign change in the high-latitude case). An $N = 4$ observational network, which only resolves low wavenumbers, will underestimate the eddy heat flux in both scenarios: in the low-latitude case, this will actually produce the wrong sign. Conversely, we expect the $N = 8$ network to perform

significantly better, as it is capable of resolving much of the heat flux spectrum, and the $N = 16$ case to perform better still, as it resolves the full inverse cascade range.

Because the buoyancy is proportional to the difference of the upper and lower streamfunctions, knowledge of the surface field alone is insufficient to determine the heat flux. The use of satellite altimetry to estimate eddy fluxes therefore requires an estimate of the vertical structure of the flow; in the present case, this means estimating ψ_2 given (sparse) observations of ψ_1. The simplest possible estimate (which we will show to be insufficient) uses optimal interpolation (Daley 1991). In this approach the vertical structure of the flow is modeled by projecting the Fourier transform of the layer streamfunctions onto vertical mode empirical orthogonal functions:

$$\begin{pmatrix} \tilde{\chi}_{kl}^+ \\ \tilde{\chi}_{kl}^- \end{pmatrix} = \mathbf{V}(k, l) \begin{pmatrix} \tilde{\psi}_{kl}^1 \\ \tilde{\psi}_{kl}^2 \end{pmatrix}, \qquad (2.5)$$

where $\mathbf{V}(k, l)$ diagonalizes the time-averaged covariance matrix for the energy-weighted vertical modes (see the appendices for details). The EOFs are ordered so that $\hat{\chi}^+$ has the maximum energy or variance. If we assume that only the most energetic EOF contributes to the vertical structure of each horizontal wavenumber (i.e., $\hat{\chi}^- \approx 0$) then both $\tilde{\psi}^1$ and $\tilde{\psi}^2$ can be expressed solely in terms of $\hat{\chi}^+$. In that case, (2.5) provides a relationship between the upper and lower streamfunctions:

$$\tilde{\psi}_{kl}^2 \approx -\frac{V_{21}(k, l)}{V_{22}(k, l)}\tilde{\psi}_{kl}^1, \qquad (2.6)$$

which then can be used to estimate the eddy flux using observations of the upper-layer velocity field alone. Time series of the true and optimally interpolated eddy heat flux (normalized by the true time-averaged value) are shown in Fig. 4 for the high-latitude and low-latitude cases. Observational scenarios corresponding to $N = 4$, 8, and 16 networks are considered. The true heat fluxes were calculated using 2.5-day snapshots of the velocity field, that is, 6 times faster than the observation time scale used for the optimally interpolated estimates. As expected, the $N = 4$ case drastically underestimates the eddy heat flux in both scenarios, and in the low-latitude case, the estimate produces the wrong sign due to dominant negative contributions from low wavenumber Rossby modes. The estimate is better in the $N = 8$ and $N = 16$ cases, which are capable of resolving the most energetic eddies. Also shown in Fig. 4 are the normalized energy spectra in the zonal and meridional directions estimated using optimal interpolation. At high
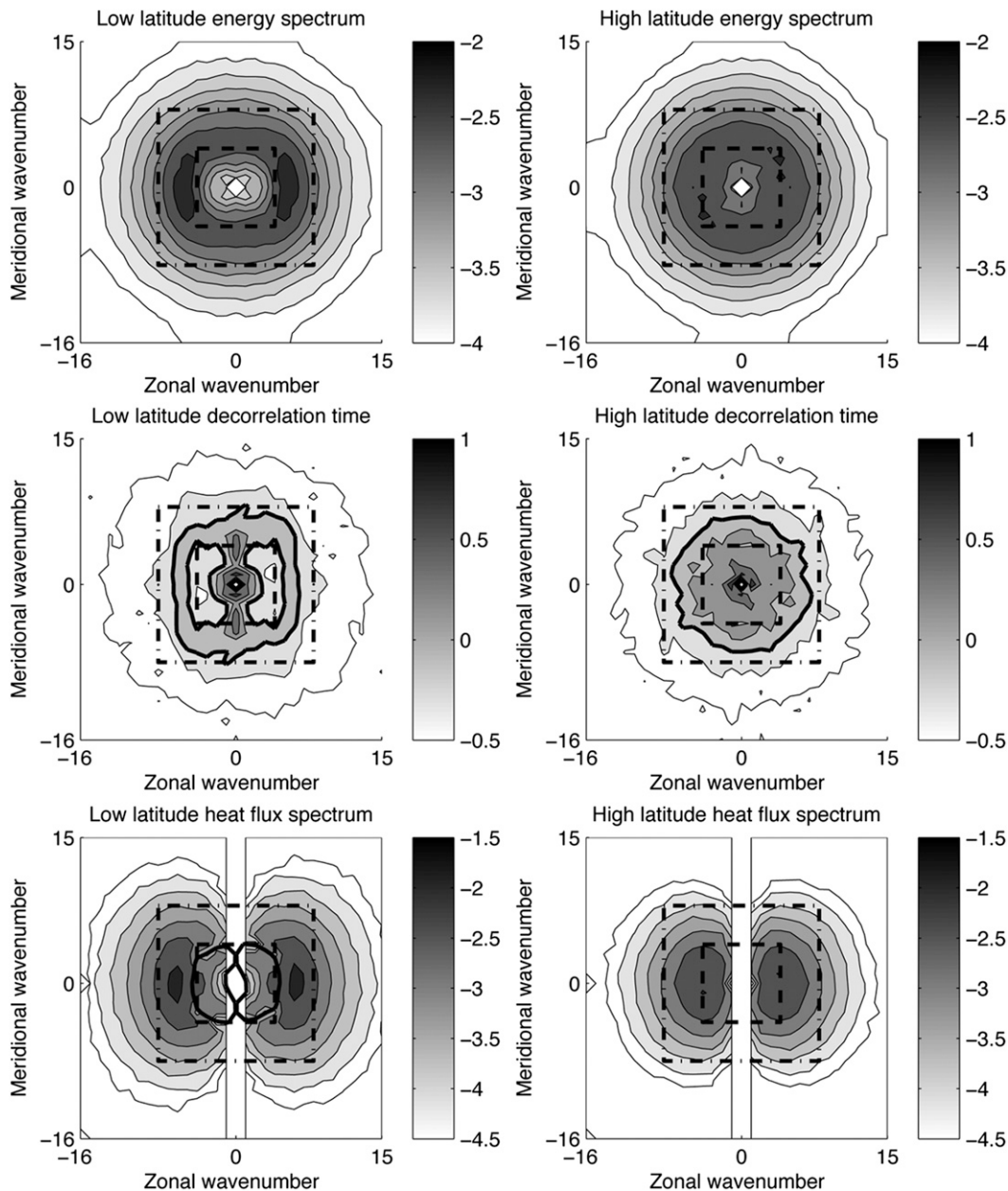
FIG. 3. (top) Log of the normalized energy spectrum for (left) low- and (right) high-latitude cases. The spectral range shown corresponds to the wavenumbers resolved by an $N = 16$ observation network; the dashed–dotted and dashed regions correspond to $N = 8$ and $N = 4$ networks, respectively. (middle) Log of the correlation time expressed in units of the eddy turnover time. Only the correlation time of the largest EOF for each horizontal wavenumber is shown. The thick contour shows the observation time scale (equal to one turnover time) so that modes outside this contour decay rapidly compared with the observation time. (bottom) Log of the normalized heat flux spectrum. The thick contour indicates where the contribution to the heat flux switches from negative (at low wavenumbers) to positive (at high wavenumbers). There is no sign change in the high-latitude case.

spatial resolution, optimal interpolation accurately represents the energy spectra, but at $N = 4$ the coarse observations alias higher wavenumbers into the resolved wave band and so optimal interpolation overestimates the energy spectra.

We conclude that a reliable estimate of the poleward eddy heat flux requires detailed information about the vertical and horizontal structure of the turbulent flow. Since satellite observations are restricted to the upper-ocean velocity field and are coarsely resolved with respect
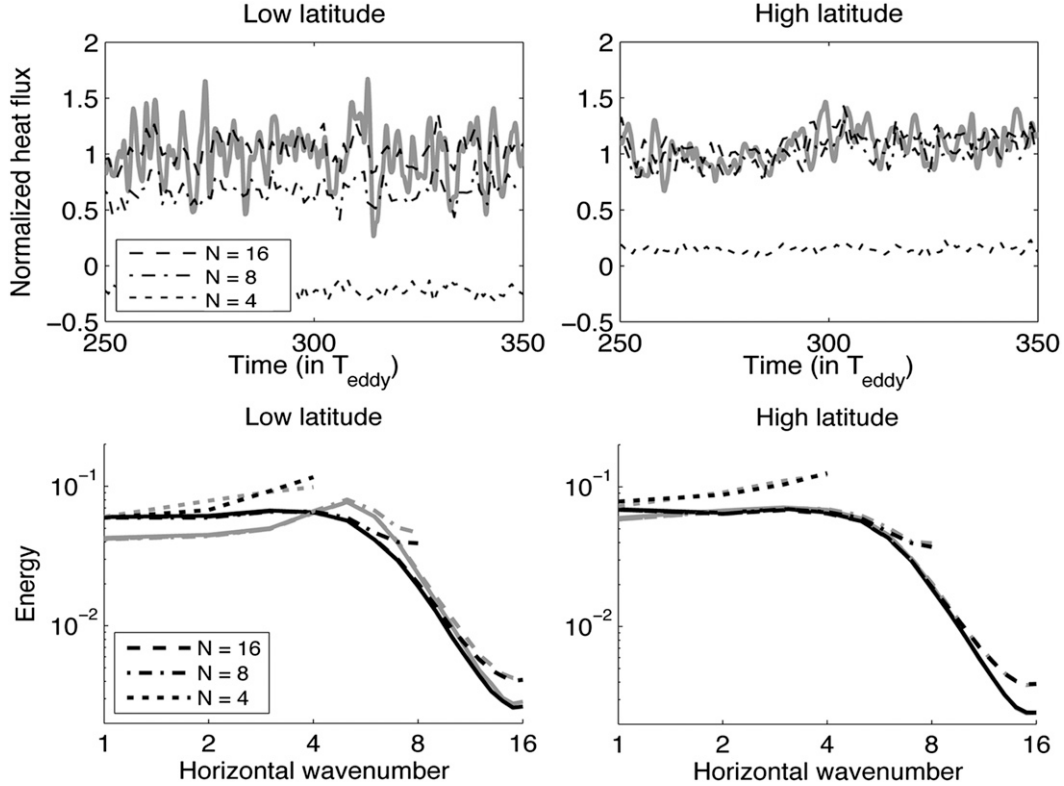
FIG. 4. (top) Time series of the poleward eddy heat transport at (left) low latitudes and (right) high latitudes. The true eddy heat flux is indicated by a solid gray line. Also shown are OI estimates of the heat flux from coarse observations with $N = 16$ (dashed line), $N = 8$ (dashed–dotted line), and $N = 4$ (dotted line). In each case, the heat flux has been normalized by the respective time-averaged true heat flux. (bottom) Normalized energy spectra in the zonal (black) and meridional (gray) directions, estimated using OI.

to the scales of the most energetic eddies (especially at high latitudes), it is natural to seek a turbulence filtering or data assimilation approach, in which a model for the full three-dimensional streamfunction is combined with observations of the upper layer to produce a best estimate of the true state of the full system. Standard data assimilation approaches, such as three-dimensional variational data assimilation (3DVAR), four-dimensional variational data assimilation (4DVAR), and ensemble Kalman filters, involve a considerable computational overhead and significant model error, particularly in the ocean (Harlim and Majda 2010). A suite of cheap, skillful alternative filtering methodologies is described in the next section.

3. Filtering methodology for sparse observations

Filtering (or data assimilation) seeks to obtain an estimate of the true system state $\mathbf{z}_s$ at time $t = t_s$ by combining incomplete, noisy observations $\mathbf{y}_s = \mathcal{G}(\mathbf{z}_s)$ with an internal forecast model $\mathbf{x}_s = \mathcal{F}(\mathbf{x}_{s-1})$. An important body of research in applied mathematics, engineering, and the geosciences is concerned with filtering real-world

turbulent flows with many interacting degrees of freedom. In these applications, the forecast model (which can be a general circulation model in the case of atmospheric or oceanic flows) often has significant model errors due to unresolved dynamical processes, and the resulting high-dimensional inverse problem must be solved approximately. Here, we adopt an alternative strategy: rather than seeking to represent the full nonlinear dynamics as accurately as possible, we utilize a forecast model incorporating judicious model error that dramatically increases the speed and stability of the filter while preserving the essential structure of the turbulent system (Majda et al. 2010). Specifically, for each element of the state vector $\mathbf{x}_s = \{\hat{\chi}_\alpha(t_s)\}$, we replace the nonlinear governing equation with a linear stochastic model of the following form:

$$\partial_t \hat{\chi}_\alpha = -(\gamma_\alpha - \mathbf{i}\omega_\alpha)\hat{\chi}_\alpha(t) + \sigma_\alpha \dot{W}(t), \quad \alpha = \{\pm, k, l\},$$

(3.1)

where γ_α and ω_α are linear damping rates and frequencies, σ_α is a stochastic noise strength, and $\dot{W}$ is circularly symmetric complex white noise forcing. While

there is no formal justification for the form of the sto-chastic forecast model in (3.1), it is perhaps the simplest possible stochastic parameterization of the turbulent exchange of energy from and to a background "sea" of eddies of different scales, and has found wide applica-tion and validation in turbulence modeling (see, e.g., DelSole 2004; Harlim and Majda 2010).

The linear stochastic forecast model in (3.1) offers two key advantages for turbulence filtering: first, it is ex-tremely cheap to implement, and second, the inverse problem can be solved *exactly* using the Kalman (1960) filter algorithm, which provides the optimal filtered (or analysis) estimate when the observation and forecast model take the following form:

$$\mathbf{y}_s = \mathbf{G}\mathbf{z}_s + \boldsymbol{\sigma}_s^{\text{obs}}, \quad \mathbf{x}_s = \mathbf{F}\mathbf{x}_{s-1} + \boldsymbol{\sigma}_s^{\text{sys}}, \quad (3.2)$$

where $\mathbf{F}, \mathbf{G}$ in (3.2) are, respectively, the linear dynamics operator and the observation operator (which need not sample the entire state space), and $\boldsymbol{\sigma}_s^{\text{sys}}, \boldsymbol{\sigma}_s^{\text{obs}}$ are mean-zero Gaussian random variables representing system noise and observation noise with covariance tensors $\mathbf{R}^{\text{sys}}$ and $\mathbf{R}^{\text{obs}}$. Under these conditions, the forecast f and analysis a states are Gaussian random variables with mean and covariance satisfying the following:

$$\langle \mathbf{x}_s^f \rangle = \mathbf{F}\langle \mathbf{x}_{s-1}^a \rangle, \quad \mathbf{R}_s^f = \mathbf{F}\mathbf{R}_{s-1}^a \mathbf{F}^* + \mathbf{R}^{\text{sys}}, \quad (3.3)$$

$$\langle \mathbf{x}_s^a \rangle = (1 - \mathbf{K}_s\mathbf{G})\langle \mathbf{x}_s^f \rangle + \mathbf{K}_s\mathbf{y}_s, \quad \mathbf{R}_s^a = (1 - \mathbf{K}_s\mathbf{G})\mathbf{R}_s^f, \quad (3.4)$$

where $\mathbf{K}_s = \mathbf{R}_s^f \mathbf{G}^*(\mathbf{G}\mathbf{R}_s^f\mathbf{G}^* + \mathbf{R}^{\text{obs}})^{-1}$ is the Kalman gain matrix and * denotes the conjugate transpose. For more details on the Kalman filter, see Chui and Chen (1999) or Wunsch (2006), for example.

From (3.1), it is straightforward to show that

$$\mathbf{F} = \text{diag}[e^{-(\gamma_\alpha - \mathbf{i}\omega_\alpha)\Delta t}],$$

$$\mathbf{R}^{\text{sys}} = \text{diag}[\gamma_\alpha^{-1}|\sigma_\alpha|^2(1 - e^{-2\gamma_\alpha\Delta t})/2], \quad (3.5)$$

with integration time step Δt. The stochastic parameters $\gamma_\alpha, \omega_\alpha, \sigma_\alpha$ can be determined either offline, via long-time averages of the turbulent flow, or adaptively, by as-suming that the parameters are themselves governed by a stochastic differential equation. Strategies for offline and adaptive parameter estimation are described below.

a. Super-resolution of sparse observations

Majda and Grote (2007) and Harlim and Majda (2008b) describe a novel turbulence filtering algorithm for sparse observations that produces a "super re-solved" estimate of the state with a higher nominal

resolution than the observing network. This is achieved by exploiting the fact that sparse observations preserve information about high-wavenumber modes due to the effect of aliasing. To see this, consider the discrete Fourier transform of the field $f_{mn} = f(m\tilde{h}, n\tilde{h})$ evaluated on a coarse grid with spacing $\tilde{h} = \pi/\tilde{N}$:

$$\tilde{f}_{kl} = \frac{1}{(2\tilde{N})^2} \sum_{m,n=0}^{2\tilde{N}-1} f_{mn} e^{\mathbf{i}\tilde{h}(km+ln)},$$

$$f_{mn} = \text{Re} \sum_{k,l=-\tilde{N}}^{\tilde{N}-1} \tilde{f}_{kl} e^{-\mathbf{i}\tilde{h}(km+ln)}. \quad (3.6)$$

The Fourier coefficients $\tilde{f}_{kl}$ will differ from the corre-sponding coefficients $\hat{f}_{kl}$ defined on a finescale grid with spacing $h = \pi/N \le \tilde{h}$, because the coarse Fourier trans-form samples, with equal weight, all Fourier modes that reside in the *aliasing set* of each horizontal wavenumber:

$$\tilde{f}_{kl} = \sum_{i,j=1}^{P} \hat{f}_{k_i l_j}, \quad k_i \bmod \tilde{N} = k, \quad l_j \bmod \tilde{N} = l, \quad (3.7)$$

where $P = N/\tilde{N}$. Thus, an observation of the mode (k, l) on the coarse network can be thought of as a superpo-sition of aliased wavenumbers (k_i, l_j) plus some obser-vational noise:

$$\tilde{f}_{kl}^{\text{obs}}(t_m) = \mathbf{G}\mathbf{z}_{kl}(t_m) + \tilde{\sigma}_{kl}^{\text{obs}}(t_m), \quad (3.8)$$

where $\mathbf{G} = [1 \dots 1]$ is the observation operator and $\mathbf{z}_{kl}(t_m) = [\hat{f}_{k_1 l_1}(t_m) \cdots \hat{f}_{k_p l_p}(t_m)]^{\text{T}}$ is the P^2-dimensional state vector. The Kalman filter then provides an esti-mate for both resolved and unresolved wavenumbers by combining the incomplete observation in (3.8) with the stochastic forecast model (3.5). Since the aliasing sets for each primary mode (k, l) are disjoint, the full N^2-dimensional system reduces to independent P^2-dimensional filtering problems for each of the $\tilde{N}^2$ pri-mary modes.

We emphasize that the Kalman filter does not create new information; rather, it produces an estimate for the unresolved modes based upon the information avail-able. Because the nominal grid resolution of the result-ing state estimate is a factor P times higher than the original observations, we refer to this technique as *sto-chastic super-resolution*. Figure 5 demonstrates how stochastic super-resolution can be used to double the nominal resolution of a coarse observational network by filtering four aliased modes per observed wavenumber.

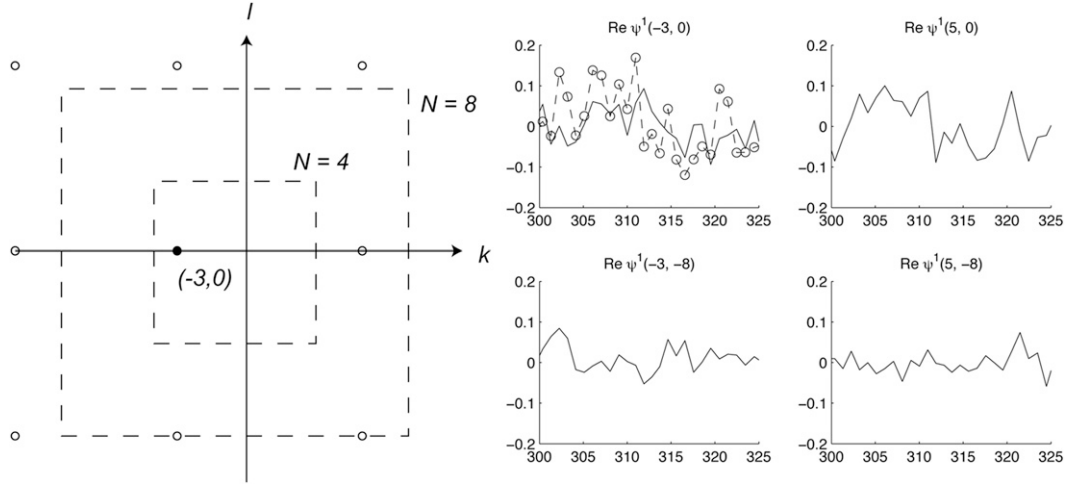In the present application, we sample the upper layer on a coarse grid so that

FIG. 5. (left) Stochastic super-resolution of an 8×8 observation network. The resolved wavenumbers all lie inside the inner dashed box $|k|, |l| \leq N = 4$. The wavenumbers indicated by the white circles show the first eight horizontal wavenumbers in the aliasing set of the horizontal wavenumber $(-3, 0)$, indicated by the filled circle. To obtain a nominal resolution of 16×16 (corresponding to the outer dashed box $|k|, |l| \leq 8$), one must filter the primary wavenumber $(-3, 0)$ plus the three aliased wavenumbers closest to the origin: $(5, 0)$, $(-3, -8)$, and $(5, -8)$. (right) The aliasing set of the mode $(-3, 0)$. The observation (dashed line with circles) samples, with equal weight, each member of the aliasing set. Note that the largest contribution actually comes from aliased mode $(5, 0)$.

$$\tilde{\psi}_{kl}^{\text{obs}} = [\,1 \quad 0\,]\begin{pmatrix} \tilde{\psi}_{kl}^1 \\ \tilde{\psi}_{kl}^2 \end{pmatrix} + \tilde{\sigma}_{kl}^{\text{obs}}, \qquad (3.9)$$

where observations of the (k, l) mode sample all modes in the aliasing set:

$$\begin{pmatrix} \tilde{\psi}_{kl}^1 \\ \tilde{\psi}_{kl}^2 \end{pmatrix} = [\mathcal{I}_2 \cdots \mathcal{I}_2]\begin{pmatrix} \tilde{\psi}_{k_1 l_1}^1 \\ \tilde{\psi}_{k_1 l_1}^2 \\ \vdots \\ \tilde{\psi}_{k_P l_P}^1 \\ \tilde{\psi}_{k_P l_P}^2 \end{pmatrix}. \qquad (3.10)$$

Note that the vector $[\cdots]$ contains P^2 copies of the 2×2 identity matrix $\mathcal{I}_2$ and so is a matrix of size $2 \times 2P^2$. In terms of EOFs, the state vector is $\mathbf{z}_{kl} = [\hat{\chi}_{k_1 l_1}^+, \hat{\chi}_{k_1 l_1}^-, \cdots \hat{\chi}_{k_P l_P}^+, \hat{\chi}_{k_P l_P}^-]^{\mathrm{T}}$ and

$$\begin{pmatrix} \tilde{\psi}_{kl}^1 \\ \tilde{\psi}_{kl}^2 \end{pmatrix} = [\mathbf{V}^{-1}(k_1, l_1) \cdots \mathbf{V}^{-1}(k_P, l_P)]\begin{pmatrix} \tilde{\chi}_{k_1 l_1}^+ \\ \tilde{\chi}_{k_1 l_1}^- \\ \vdots \\ \tilde{\chi}_{k_P l_P}^+ \\ \tilde{\chi}_{k_P l_P}^- \end{pmatrix}, \qquad (3.11)$$

where $\mathbf{V}(k, l)$ is the transformation matrix from layer streamfunctions to EOFs and again the vector $[\cdots]$ is a

matrix of size $2 \times 2P^2$. Thus, the observed streamfunction is

$$\hat{\psi}_{kl}^{\text{obs}} = \mathbf{G}(k, l)\mathbf{z}_{kl} + \tilde{\sigma}_{kl}^{\text{obs}},$$
$$\mathbf{G}(k, l) = [\,1 \quad 0\,] \times [\mathbf{V}^{-1}(k_1, l_1) \cdots \mathbf{V}^{-1}(k_P, l_P)], \qquad (3.12)$$

where the observation operator $\mathbf{G}(k, l)$ is a $1 \times 2P^2$ matrix (i.e., there are $2P^2$ filtered modes for each observed wavenumber).

b. Offline and adaptive parameter estimation

Harlim and Majda (2008a) outline a simple procedure for uniquely determining the stochastic parameters in the forecast model in (3.5) from the long-time statistics of the turbulent flow. In the mean stochastic model (MSM), the parameter set $(\gamma_\alpha, \omega_\alpha, \sigma_\alpha)$ is chosen to reproduce the correct energy spectrum and decorrelation time for each EOF $\hat{\chi}_\alpha$:

$$e_\alpha = \langle |\hat{\chi}_\alpha|^2 \rangle_t, \quad T_\alpha + \mathbf{i}\Theta_\alpha = e_\alpha^{-1}\int_0^\infty \langle \hat{\chi}_\alpha(t)\hat{\chi}_\alpha^*(t + T) \rangle_t \, dT. \qquad (3.13)$$

Substituting the stochastic model (3.1) into (3.13) yields expressions for the stochastic parameters $(\gamma_\alpha, \omega_\alpha, \sigma_\alpha)$ in terms of the turbulent statistics $(e_\alpha, T_\alpha, \Theta_\alpha)$:

$$\gamma_\alpha = \frac{T_\alpha}{T_\alpha^2 + \Theta_\alpha^2}, \quad \omega_\alpha = -\frac{\Theta_\alpha}{T_\alpha^2 + \Theta_\alpha^2}, \quad \sigma_\alpha = \sqrt{2\gamma_\alpha e_\alpha}. \qquad (3.14)$$
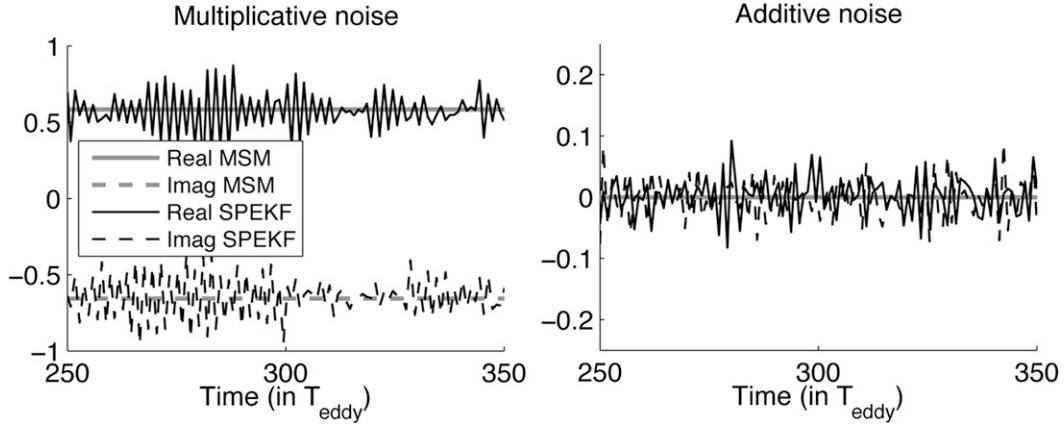
FIG. 6. Time series of the real and imaginary components of the multiplicative and additive noise terms for mode $(-3, 0, +)$ in the low-latitude case. Superimposed are the MSM values for these terms, for comparison.

A limitation of the MSM and other simple autoregressive stochastic processes for filtering turbulent flows is model error. This arises in two principal ways. First, obtaining the stochastic model parameters requires knowledge of the long-time statistics of the full state vector, including the lower layer and unresolved high-wavenumber modes. In practice, these statistics will be only partially or imperfectly known. Second, the stochastic parameters are estimated offline and held fixed throughout the filtering process, and thus are unable to adjust to temporal intermittency in the flow.

A standard approach to deal with model error is to estimate the model parameters "on the fly" by allowing them to evolve in time and augmenting the system state to filter them as if they were unobserved state variables (e.g., Dee and Da Silva 1998; Baek et al. 2006; Dowd 2006; Anderson 2007; Harlim and Majda 2010). Typically, the resulting augmented system will involve nonlinear dependencies between the original state variables and the stochastic model parameters, so a linear tangent approximation is often used. A particular form for the nonlinear augmented system with *exactly* solvable statistics was recently studied by Gershgorin et al. (2010a,b):

$$\partial_t \hat{\chi} = -m(t)\hat{\chi}(t) + a(t) + \sigma \dot{W}(t), \quad (3.15a)$$

$$\partial_t m = -\lambda_m [m(t) - \overline{m}] + \sigma_m \dot{W}_m(t), \quad (3.15b)$$

$$\partial_t a = -\lambda_a [a(t) - \overline{a}] + \sigma_a \dot{W}_a(t). \quad (3.15c)$$

Here $m(t)$ and $a(t)$ are complex multiplicative and additive bias corrections and $\dot{W}_m(t)$ and $\dot{W}_a(t)$ are circularly symmetric complex white noise forcings. The augmented state vector $\mathbf{z}_{kl} = [\hat{\chi}^+_{k_1 l_1}, \hat{\chi}^-_{k_1 l_1}, \cdots \hat{\chi}^+_{k_p l_p}, \hat{\chi}^-_{k_p l_p}, m^+_{k_1 l_1}, m^-_{k_1 l_1}, \cdots a^+_{k_p l_p}, a^-_{k_p l_p}]^{\mathrm{T}}$ is then filtered using the modified observation operator:

$$\mathbf{G}(k,l) = [1 \quad 0] \times [\mathbf{V}^{-1}(k_1, l_1) \cdots \mathbf{V}^{-1}(k_P, l_P) \mathcal{O}_2 \cdots \mathcal{O}_2], \quad (3.16)$$

where there are $2P^2$ copies of the 2×2 zero matrix $\mathcal{O}_2$, reflecting the fact that the bias correction terms cannot be observed directly. The modified observation operator is a matrix of size $1 \times 6P^2$ (i.e., there are $6P^2$ filtered variables per observed wavenumber). The resulting stochastic parameterized extended Kalman filter (SPEKF) algorithm provides an estimate of the mean and covariance of the full state vector (including unobserved high-wavenumber and lower-layer variables) at time $t = t_s$ and an on-the-fly estimate of the stochastic parameters $m(t)$ and $a(t)$ for each mode.[1]

Following Harlim and Majda (2010), we set the equilibrium mean multiplicative and additive bias corrections and the energy spectrum to their MSM values:

$$\overline{m} = \gamma_{\mathrm{MSM}} - \mathbf{i}\omega_{\mathrm{MSM}}, \quad \overline{a} = 0, \quad \sigma = \sigma_{\mathrm{MSM}}, \quad (3.17)$$

so that SPEKF acts as an adaptive learning algorithm for the MSM parameterization (see Fig. 6). Because the bias terms $m(t)$ and $a(t)$ are filtered from the observations themselves, SPEKF can correct for significant model error in the original MSM parameterizations. The remaining augmented model parameters $\lambda_m = \gamma_m + \mathbf{i}\omega_m$, $\lambda_a = \gamma_a + \mathbf{i}\omega_a$, σ_m and σ_a do not have a direct interpretation in terms of the turbulent statistics of the physical system; rather, they can be tuned to mimic a

---

[1] We note that the nonlinearity in the forecast model in (3.15) permits nontrivial correlations between real and imaginary components of the state vector and these should be treated separately in (3.16) and the Kalman filter solution in (3.4). For the sake of expedience we will not write this out explicitly.

range of turbulent behaviors including strong intermittency, transient instability, and laminar flow (Branicki et al. 2011). Crucially, there exists a robust regime of parameter values for which SPEKF has high filtering skill when compared with MSM, as was demonstrated by Gershgorin et al. (2010a,b) in comprehensive numerical studies. A more limited study for the flows under consideration here yielded the following values, which are held constant for all wavenumbers:

$$\sigma_m = 5\sigma_{MSM}, \quad \sigma_a = \sigma_{MSM}, \quad \gamma_m = \gamma_a = 0.1\gamma_{MSM},$$
$$\omega_m = \omega_a = 5\omega_{MSM}. \tag{3.18}$$

As we have indicated, a significant advantage of the form of (3.15) is that exact nonlinear expressions for the mean and covariance of the non-Gaussian forecast distribution can be derived using the mathematical properties of conditionally Gaussian processes (Gershgorin and Majda 2008; Gershgorin et al. 2010b). Branicki et al. (2011) have also developed a simpler quasi-Gaussian closure approximation to the SPEKF algorithm with similarly high skill for a wide range of parameter values. In this article, we adopt a hybrid quasi-Gaussian scheme in which the forecast mean values of $\hat{\chi}, m$, and a are calculated using the exact nonlinear statistics of the augmented model, while the forecast covariance tensor is obtained from a Monte Carlo estimate using the pathwise solutions of (3.15):

$$\hat{\chi}(t_s) = \hat{\chi}(t_{s-1})e^{-j(t_{s-1}, t_s)} + \int_{t_{s-1}}^{t_s} dt[a(t) + \sigma \dot{W}(t)]e^{-j(t, t_s)}, \tag{3.19a}$$

$$m(t_s) = \overline{m} + [m(t_{s-1}) - \overline{m}]e^{-\lambda_m(t_s - t_{s-1})}$$
$$+ \sigma_m \int_{t_{s-1}}^{t_s} dt \, \dot{W}_m(t)e^{-\lambda_m(t_s - t)}, \tag{3.19b}$$

$$a(t_s) = \overline{a} + [a(t_{s-1}) - \overline{a}]e^{-\lambda_a(t_s - t_{s-1})}$$
$$+ \sigma_a \int_{t_{s-1}}^{t_s} dt \, \dot{W}_a(t)e^{-\lambda_a(t_s - t)}, \tag{3.19c}$$

with $j(t_0, t_1) = \int_{t_0}^{t_1} dt \, m(t)$. The forecast covariance is then estimated from an ensemble of solutions of (3.19) generated using 100 normally distributed initial values $\hat{\chi}(t_{s-1})$, $m(t_{s-1})$, $a(t_{s-1})$ with mean and covariance tensor equal to that obtained at the previous analysis step:

$$\mathbf{x}_{s-1}^a = \langle \mathbf{x}_{s-1}^a \rangle + \mathbf{P}_{s-1}\mathbf{w}_{s-1}, \quad \mathbf{P}_{s-1}\mathbf{P}_{s-1}^T = \mathbf{R}_{s-1}^a, \tag{3.20}$$

where $\mathbf{P}_{s-1}$ is the Cholesky decompostion of the analysis covariance tensor and $\mathbf{w}_{s-1}$ is a vector of Gaussian unbiased uncorrelated random numbers with mean zero and unit variance.

Figure 6 shows filtered estimates of the multiplicative and additive noise terms $m(t)$ and $a(t)$ for mode $(-3, 0, +)$. For comparison, the MSM values for these terms in (3.17) have been superimposed on top of these estimates. As can be seen, the SPEKF noise parameters fluctuate about the equilibrium values, and, in the case of the multiplicative noise, display temporally intermittent bursting behavior (Branicki et al. 2011). This temporal variability is particularly important in the multiplicative term because it drives a rectified contribution to the forecast mean of $\hat{\chi}$ due to the nonlinearity in (3.15a), and hence directly influences the eddy heat transport.

4. Results

The filtering strategies described in the previous section were tested by calculating the poleward eddy heat transport $\langle v_1 \tau \rangle$ from the filtered estimates for the upper and lower streamfunctions, and comparing this with the same quantity calculated from the true three-dimensional streamfunction. The filtered streamfunctions were estimated from observational networks with $N_{obs} = 4$, $N_{obs} = 8$, and $N_{obs} = 16$ using the MSM and SPEKF described in the previous section. The MSM stochastic model parameters were calculated from long, fully resolved time series of the upper and lower streamfunctions, while the SPEKF model parameters were assigned according to the rules in (3.17) and (3.18). The filters were run for 400 assimilation steps ($\approx$400 eddy turnover times) after a transient adjustment period of 100 assimilation steps. Each final streamfunction estimate was averaged over 10 independent filter runs.

Filtered streamfunctions were calculated with nominal resolution N_{filt} equal to that of the observation network N_{obs}. These were compared with stochastically super-resolved (SSR) streamfunctions with nominal resolutions of up to $N_{filt} = 16$ by filtering either one or three aliased modes per primary wavenumber in each direction (see Fig. 5). Thus, to obtain a nominal resolution of $N_{filt} = 8$ from an $N_{obs} = 4$ observational network, a total of eight modes (four in each layer) were filtered for every observed mode. To obtain a nominal resolution of $N_{filt} = 16$ from the same network, 18 modes (9 in each layer) were filtered. In all cases the observational noise covariance was set at 5% of the total kinetic energy and distributed evenly among all observed Fourier modes.

Figures 7 and 8 depict time series of the estimated eddy heat transport after the initial transient adjustment
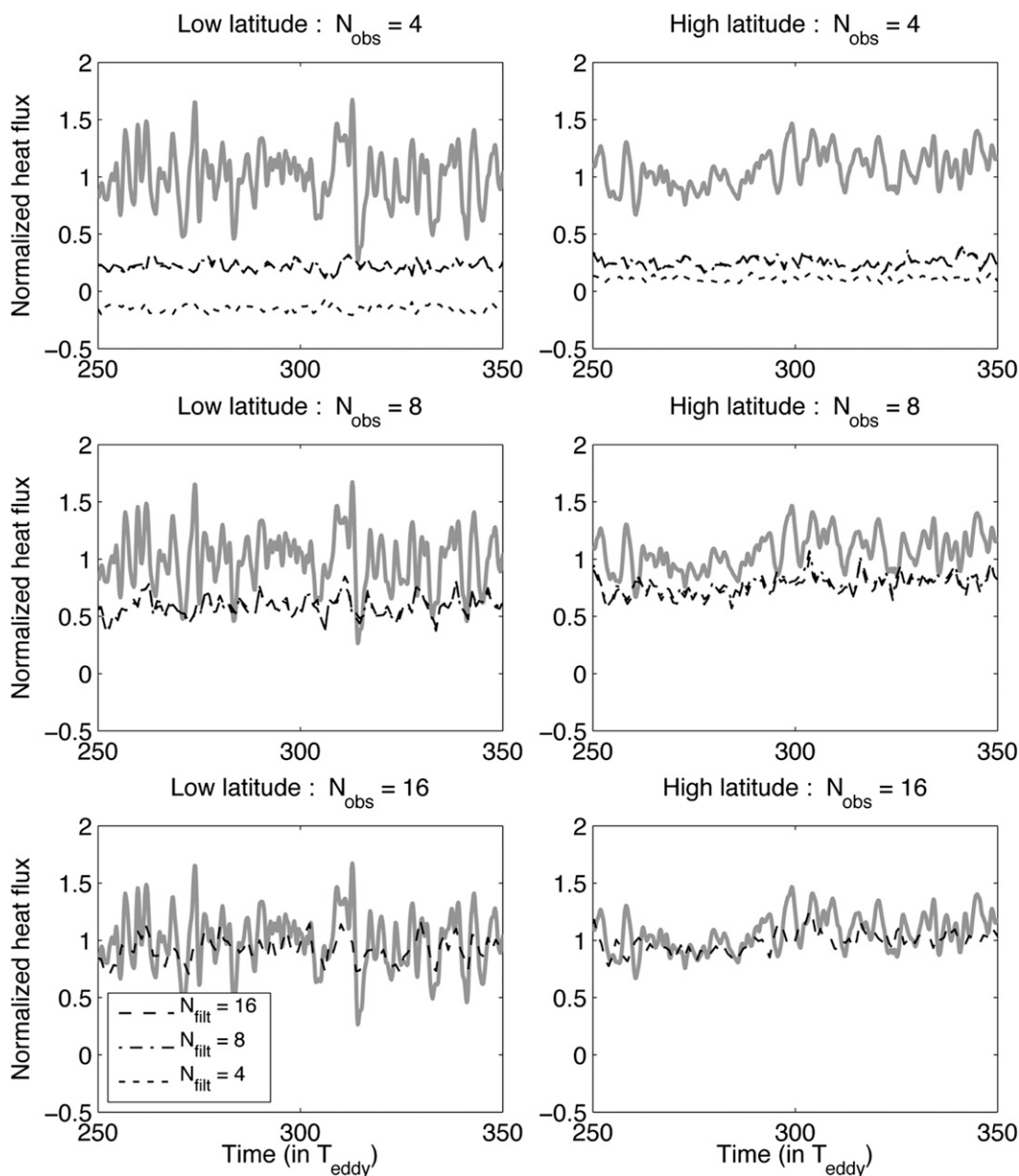
FIG. 7. Time series of the poleward eddy heat transport estimated using the MSM at (left) low latitudes and (right) high latitudes. The true eddy heat flux is indicated by a solid gray line. Also shown is the heat flux calculated using filtered velocity fields with nominal resolutions of N_{filt} = 4, 8, and 16. In each case, the heat flux has been normalized by its respective true heat flux average.

phase. As with the optimal interpolation (OI) estimates, the MSM algorithm tends to underestimate the eddy heat flux, and predicts the incorrect sign at low latitudes with N_{obs} = 4 due to the dominance of low-wavenumber zonal modes there. The inclusion of SSR modes improves the estimate at both high and low latitudes and captures about 20%–25% of the time-mean eddy heat flux. When the baseline observation resolution is increased to N_{obs} = 8, over half of the time-mean eddy

heat flux is captured, again a result that could have been anticipated from the OI estimates at this resolution. Likewise, when the observation resolution is increased further to N_{obs} = 16, over 90% of the eddy heat flux is obtained. Note that, for observation resolutions of both N_{obs} = 4 and N_{obs} = 8, little further improvement is seen when stochastic super-resolution increases the nominal resolution from N_{filt} = 8 to N_{filt} = 16. This is consistent with the intuition that wavenumber modes
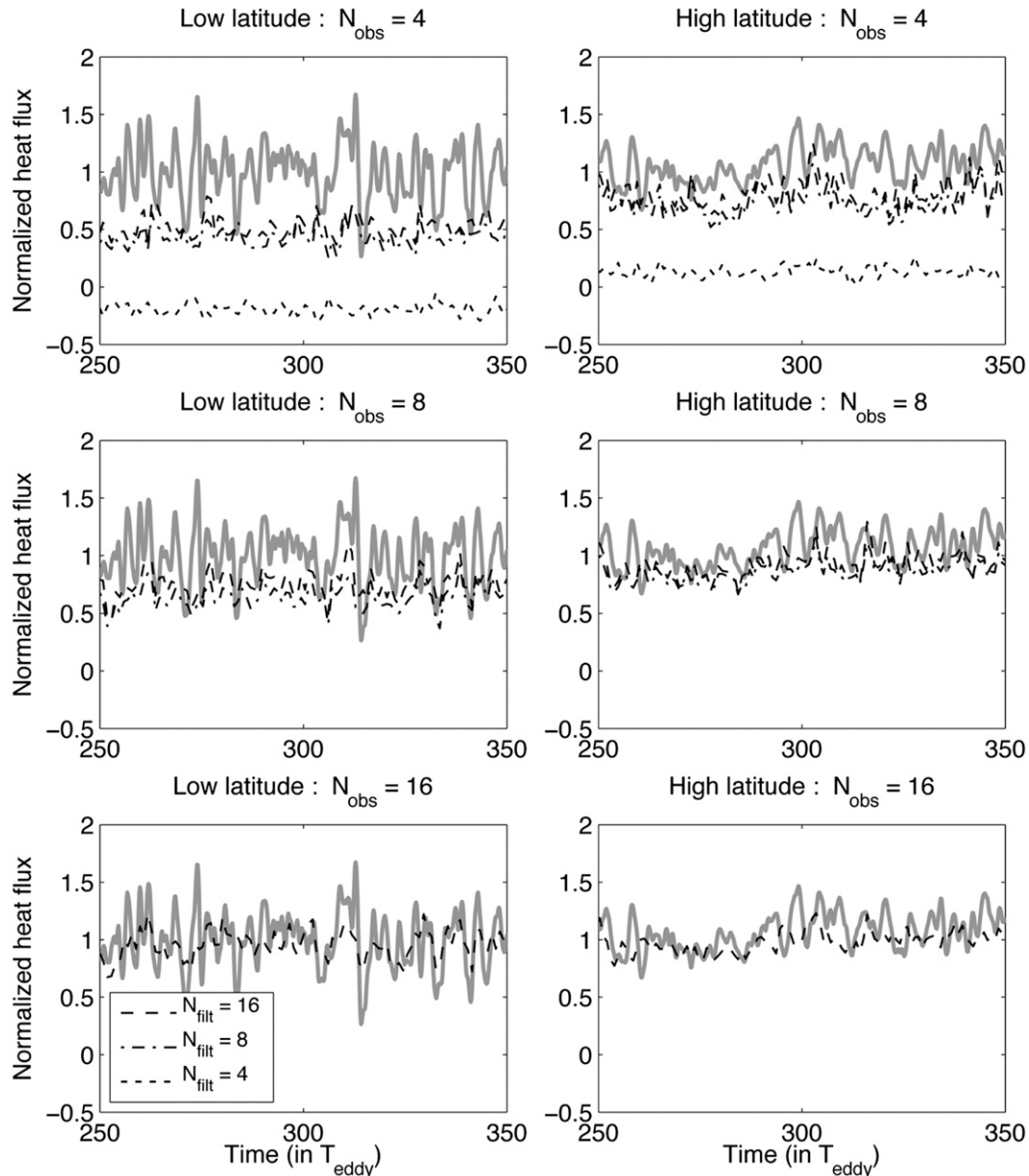
FIG. 8. As in Fig. 7, but using the SPEKF.

with $4 \leq |k|, |l| \leq 8$ dominate the total eddy heat flux. In all cases, the temporal variability of the filtered output is considerably less than the truth signal because the long observation time step is unable to resolve rapid variations in the true heat flux.

In the case of SPEKF, the addition of SSR modes to the $N_{obs} = 4$ observational network leads to considerable improvement in the time-mean eddy heat transport estimate, up to 50% at low latitudes and 75% at high latitudes. This is likely because the inclusion of aliased modes means the filters are better able to adapt to intermittent bursts of instability near the deformation

radius (i.e., $4 \leq |k|, |l| \leq 8$), which would otherwise be unresolved. While it is unreasonable to expect an exact reconstruction of the temporal variations of the signal with such a long observational time step, the temporal variability of the eddy heat flux estimated from SPEKF are at least *statistically* closer to the truth. Once more, the $N_{obs} = 8$ and $N_{obs} = 16$ observational networks gain little by filtering aliased modes, because in these cases the bulk of the eddy field is already captured by the observations. An important observation to make from both Figs. 7 and 8 is that SSR tends to improve the estimate of the time-mean heat flux (particularly in the
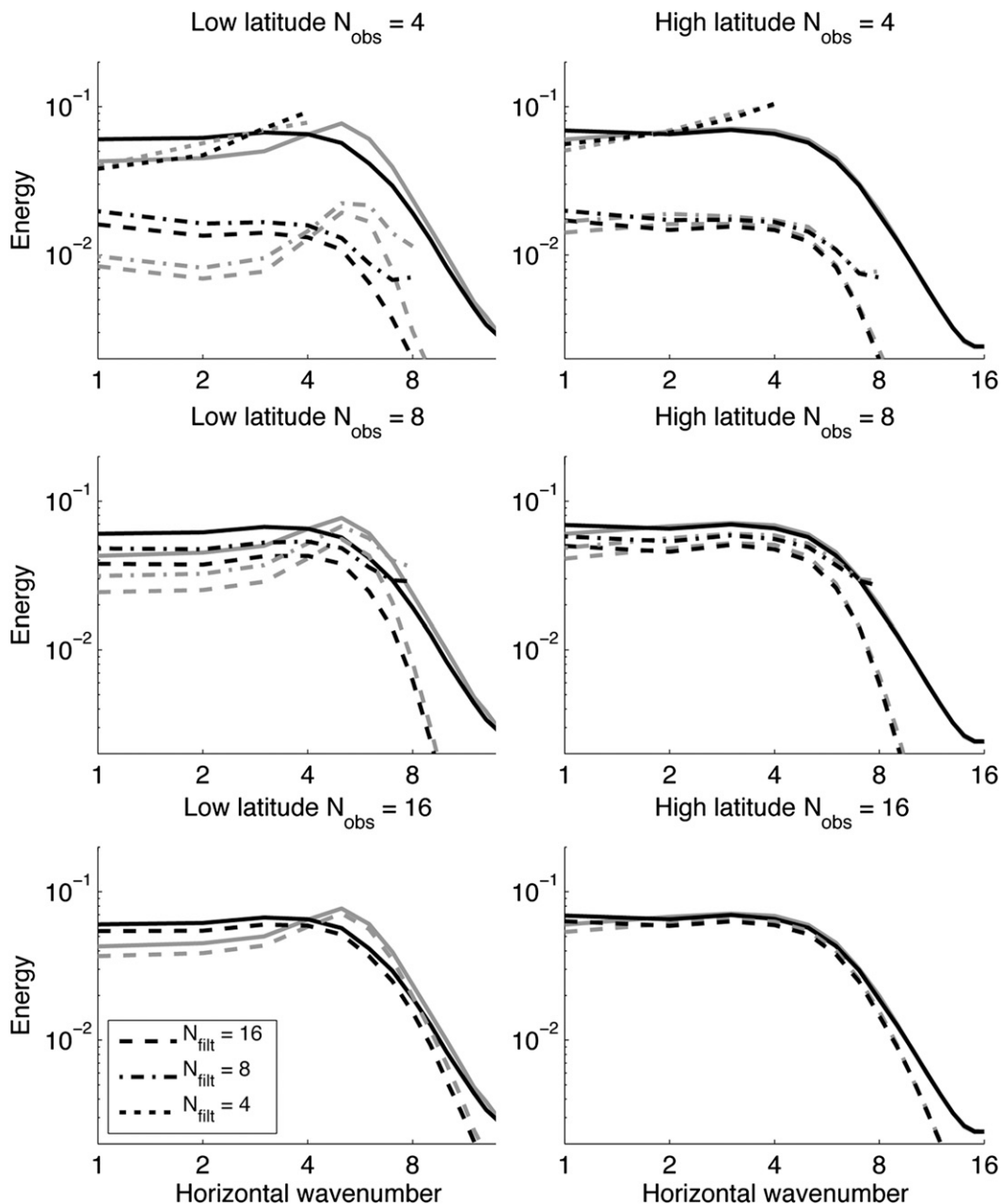
FIG. 9. Normalized energy spectra in the zonal (black) and meridional (gray) directions, estimated using MSM at (left) low latitudes and (right) high latitudes. Also shown is the energy spectrum calculated using filtered velocity fields with N_{filt} = 4, 8, and 16. In each case, the energy spectrum has been normalized by the true total energy.

case of SPEKF), but it is no substitute for increased baseline observational resolution.

Depth-averaged meridional and zonal kinetic energy spectra estimated using MSM and SPEKF are shown in Figs. 9 and 10. In both low- and high-latitude cases, the N_{obs} = 8 and N_{obs} = 16 observational networks capture most of the kinetic energy, while the N_{obs} = 4 network

overestimates the kinetic energy at low wavenumbers because the coarse observations alias high-wavenumber energy into the resolved wave band. The SSR algorithm redistributes this kinetic energy to higher wavenumber aliased modes and successfully reconstructs the peak of the spectra. Note that the total energy estimated by the filter tends to be less than the true total energy, as would
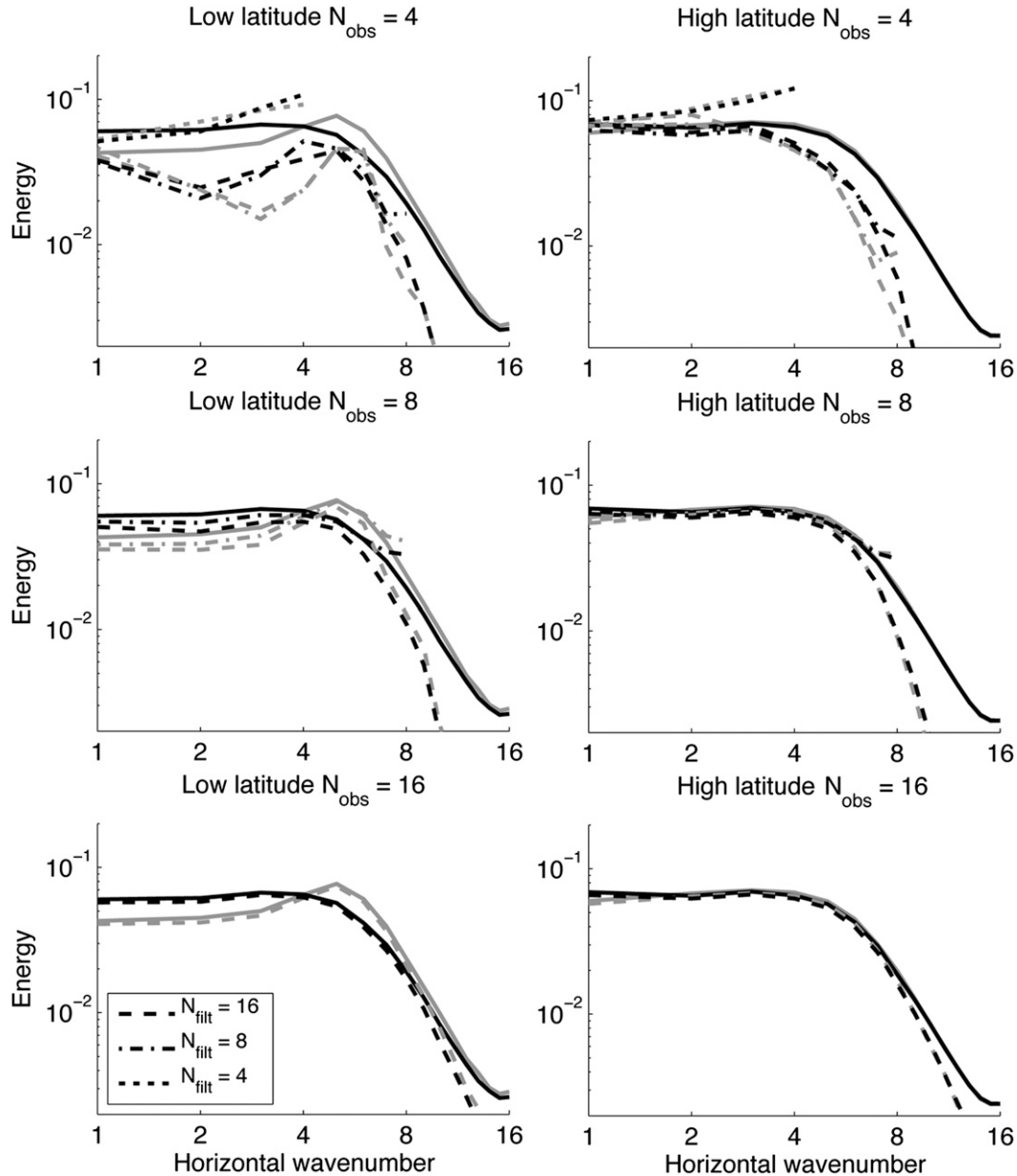
FIG. 10. As in Fig. 9, but using the SPEKF.

be expected from low-resolution observations. SPEKF does a better job of reconstructing the energy spectrum from coarse observations, particularly at low wavenumbers.

Figure 11 summarizes the time-mean heat transport estimates for the low- and high-latitude cases with $N_{obs} = 4$, $N_{obs} = 8$, and $N_{obs} = 16$. For each of the filtering strategies considered, the equilibrium heat flux, normalized by the true time-mean heat flux, was averaged over 400 eddy turnover times and plotted against the maximum resolved (Nyquist) wavenumber N_{obs}. We

find that SSR gives significant improvement over the OI estimate of the time-mean heat transport at the lowest spatial resolution considered ($N_{obs} = 4$). This is because the unresolved modes are better represented in the SSR velocity fields. Of the filtering strategies considered at this resolution, the SPEKF algorithm with maximum stochastic super-resolution performs best, capturing over 75% of the time-mean signal at high latitudes. This result should be compared with $\approx 13\%$ captured without filtering (i.e., with optimal interpolation) at spatial resolutions corresponding to current satellite capabilities.
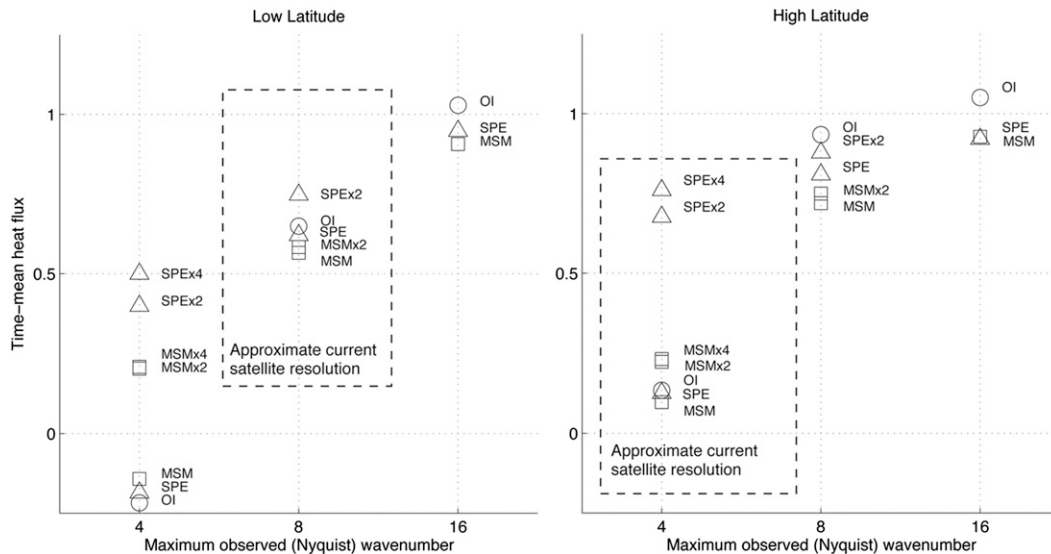
FIG. 11. Time mean of the heat flux in the low- and high-latitude case (normalized with respect to the truth) vs maximum observable (Nyquist) wavenumber: OI estimate (circles), MSM (squares), and SPEKF (circles). Also shown are stochastically super-resolved filter estimates with a nominal resolution double (×2) or quadruple (×4) that of the original observing network. The approximate current satellite resolutions at these latitudes are indicated by dashed boxes.

Likewise, using approximate current satellite resolutions, almost 75% of the time-mean heat flux at low latitudes is captured by the SPEKF algorithm with maximum stochastic super-resolution. While this is encouraging, it is not as significant as at high latitudes because the observational network in this case successfully captures the bulk of the mesoscale eddy field, and so gains little by the inclusion of SSR modes.

In summary, we find that the combination of SSR and SPEKF provides a simple, computationally inexpensive method of estimating the three-dimensional streamfunction from spatiotemporally sparse observations of the upper-ocean velocity field. The time-mean eddy heat transport calculated from the filtered streamfunction shows significant improvement over both optimal interpolation and mean stochastic model estimates in the scenarios we have investigated. With such a long observational time step, we do not expect to reconstruct the exact temporal variations of the heat flux, however, and the filtered estimates still underrepresent the temporal variability of the heat transport. Even so, the skill of the SPEKF algorithm is particularly promising, because it is capable of learning the appropriate stochastic model parameters on the fly, effectively learning the appropriate subgrid-scale parameterization from the observations themselves. This lends the algorithm significant flexibility in dealing with model error in the forecast model, for instance by using the incorrect energy spectrum for the linear stochastic model in (3.1).

5. Conclusions

We have tested the performance of a suite of novel filtering strategies for estimating poleward eddy heat transport from spatially and temporally sparse observations of the upper-ocean velocity field. Baroclinically unstable, eddy-rich flow fields were generated using the Phillips (1954) model in parameter regimes corresponding to oceanic mesoscale turbulence at high and low latitudes. Net heat transport was calculated by projecting observations of the upper-layer streamfunction onto vertical mode empirical orthogonal functions (EOFs) calculated from climatological data. Plausible space–time sampling scenarios were considered, representing current-generation satellite resolution capability. These scenarios constitute stringent test cases for the filtering strategies considered, particularly at high latitudes where contemporary satellite observations are barely capable of resolving the most energetic eddies. In addition, we studied higher-resolution observation scenarios with the goal of anticipating the possible gains that might be expected from the proposed Surface Water Ocean Topography altimetry mission (e.g., Durand et al. 2010).

The filtering strategies considered incorporate a number of features that improve their performance when compared with standard Kalman, extended Kalman, and ensemble data-assimilation techniques (Harlim and Majda 2010). Nonlinear interaction terms in the forward model

are replaced by a simple autoregressive stochastic model for each mode, so that the dynamical operator is diagonalized and the inverse problem can be solved exactly using the usual Kalman (1960) algorithm. Thus, by embracing "judicious" model error, the speed and stability of the filters is greatly increased without compromising the essential structure of the turbulent interactions among modes. The parameters appearing in the stochastic forecast model are determined either offline using the mean stochastic model (MSM), in which the parameters are regression fitted to reproduce the climatological energy spectrum and correlation times of the turbulent flow, or adaptively using the stochastic parameterized extended Kalman filter (SPEKF), which augments the system state to produce an "on the fly" estimate of the stochastic model parameters from the observations themselves. In addition, the filtering strategies exploit the fact that a coarse observation network will preserve information about unresolved high wavenumbers due to the effect of aliasing. Hence, by extracting high-wavenumber information aliased into the resolved band, one can obtain stochastic super-resolution (SSR) estimates for unresolved modes, boosting the nominal resolution of the observation network by a factor of 2 or more and improving the eddy transport estimated from sparse satellite observations.

The filtering strategies were tested by comparing the poleward eddy heat flux calculated using the filtered streamfunctions and the true, fully resolved flow. In addition, we calculated the heat transport one would obtain from optimal interpolation (OI) by projecting sparse observations of the upper layer onto the most energetic EOF to estimate the lower-layer streamfunction, thus allowing us to quantify the gain provided by filtering sparse observations. We found that SSR gave significant improvement over the OI estimate of the time-mean heat transport at the lowest spatial resolution considered ($N = 4$). This is because the unresolved modes are better represented in the SSR velocity fields, a result of particular relevance at high latitudes, where the Nyquist wave band excludes the peak of the kinetic energy spectrum. Particularly encouraging is the performance of SPEKF, since this algorithm is considerably more adaptable to large-scale instabilities and seasonal modulations in the turbulent flow. An intriguing by-product of the SPEKF algorithm is a real-time stochastic parameterization of the unresolved scales, which might prove valuable for developing more skillful eddy parameterization schemes for ocean climate models.

The results at low latitudes are also positive, although the improvement over OI is not as large because the observational network in this case successfully captures the bulk of the mesoscale eddy field. However, there are two reasons to believe that SSR and SPEKF might give better results with altimetric velocity fields at low latitudes than is obtained here. First, in this preliminary study, we have repeatedly used the perfect model parameters for the vertical structure, and these values will, in practice, be subject to some uncertainty, if they are known at all. The full SPEKF filter can overcome this important model error and learn these parameters directly from the data (Harlim and Majda 2010). Second, it is believed that surface-trapped modes might play a more significant role in the subtropics, which would lead to a shallower submesoscale energy spectrum than is represented by the Phillips model (Tulloch et al. 2011). Recently, Keating et al. (2011) demonstrated that eddy transports in flows with a strong surface component are much more sensitive to the spatial and temporal sampling resolution, so results obtained using the Phillips model might not be appropriate for this region. We plan to explicitly examine this issue in the context of quasigeostrophic model simulations incorporating nontrivial stratification and surface intensification in the near future.

It should be emphasized that the stochastic filtering strategies studied here should not be considered as a substitute for existing variational and ensemble-based data assimilation systems such as 3DVAR, 4DVAR, or ensemble Kalman filter techniques. Rather, they should be viewed as an inexpensive, complementary approach to dealing with sparse satellite observations that could easily be "nested" within contemporary assimilation frameworks, a subject we will take up in a later study. A number of additional hurdles remain before practical filtering algorithms incorporating these techniques can be developed for use with real altimetry data, however. In particular, the use of a Cartesian grid in this study meant that the aliasing condition assumed a particularly simple form. Aliasing conditions for irregular (non Cartesian) satellite orbits are also known (Tai 2004), but they are more complicated and it is yet to be determined if they can be exploited to produce "super resolved" altimetry maps. Moreover, irregularly spaced observations also lead to subtleties in the construction of a reduced forecast model with a diagonal covariance matrix as in (3.5) (Harlim 2011); this needs to be further studied in the context of repeat ground-track satellite orbits. Finally, for expediency we assumed that the observations were made simultaneously at all grid points once every observational time step. In reality, the sampling pattern has a complex spatiotemporal evolution with shorter subcycles within each exact repeat period, and so the temporal resolution of satellite observations is more subtle than we have represented here (Chelton et al.

2001). These caveats notwithstanding, the results of our study suggest that cheap, simple filtering techniques such as SSR and SPEKF can enhance the ability of current-generation altimetry products to constrain the role of ocean eddies in the global heat budget, and can provide a valuable measure of the "return on investment" that might be expected from advances in satellite altimetry in the coming decades.

APPENDIX A

Subsurface Geostrophic Current

The geostrophic current at arbitrary depth z is related to the horizontal gradient of pressure via

$$f\mathbf{u}(x,y,z) = \rho^{-1}\hat{\mathbf{z}} \times \nabla p. \qquad (A.1)$$

Under the assumption of hydrostatic balance (a good approximation in the ocean away from regions of strong convection near the surface), the density $\rho(x, y, z)$ and pressure $p(x, y, z)$ are related by

$$p(x,y,z) = P + g\int_z^{\eta(x,y)} \rho(x,y,z')\,dz', \qquad (A.2)$$

where P is the pressure at the sea surface $z = \eta(x, y)$. The subsurface geostrophic current is, by Leibniz's rule,

$$f\mathbf{u}(x,y,z) \approx \rho^{-1}(x,y,z)\{g\rho_0\hat{\mathbf{z}} \times \nabla\eta$$
$$+ g\int_z^\eta \hat{\mathbf{z}} \times \nabla\rho(x,y,z')\,dz'\}, \qquad (A.3)$$

where we have used $\rho(x, y, \eta) = \rho_0(x, y)$. Sufficiently close to the surface, the last term disappears and we obtain

$$f\mathbf{u}(x,y,0) \approx g\hat{\mathbf{z}} \times \nabla\eta. \qquad (A.4)$$

This surface current contribution is present throughout the water column, although it is slightly compensated by changes in the density field through a multiplicative factor $\rho_0(x, y)/\rho(x, y, z)$. At depth, the last term in (A.3) become increasingly important, ultimately dominating (and potentially reversing) the subsurface current at depths comparable with the depth of the pycnocline (Wunsch and Stammer 1998). In the absence of a model for the pressure field throughout the water column, therefore, observations of sea surface height anomalies do not constrain the subsurface geostrophic current.

APPENDIX B

Empirical Orthogonal Functions

The empirical orthogonal functions (EOFs) at each horizontal wavenumber (k, l) are constructed so that the variance of the first EOF $\tilde{\chi}_{kl}^+$ is maximal and contains the majority of the energy at that wavenumber. We first construct the energy-weighted barotropic streamfunction ψ and baroclinic streamfunction τ at each wavenumber:

$$\begin{pmatrix}\tilde{\psi}_{kl} \\ \tilde{\tau}_{kl}\end{pmatrix} = \mathbf{M}\begin{pmatrix}\tilde{\psi}_{kl}^1 \\ \tilde{\psi}_{kl}^2\end{pmatrix}, \quad \mathbf{M}(k,l) = \begin{pmatrix}\sqrt{k^2+l^2} & 0 \\ 0 & \sqrt{k^2+l^2+\lambda^2}\end{pmatrix}\begin{pmatrix}\dfrac{d_1}{\sqrt{d_1 d_2}} & \dfrac{d_2}{\sqrt{d_1 d_2}} \\ \sqrt{d_1 d_2} & -\sqrt{d_1 d_2}\end{pmatrix}. \qquad (B.1)$$

Writing $\tilde{\psi}_{kl} \mapsto \psi$, etc., and assuming that $\langle\psi\rangle = 0 = \langle\tau\rangle$, we now seek eigenvectors of the covariance matrix of the energy-weighted vertical modes:

$$\mathbf{C} = \mathrm{cov}\begin{pmatrix}\psi \\ \tau\end{pmatrix} = \begin{pmatrix}\langle|\psi|^2\rangle & \langle\psi\tau*\rangle \\ \langle\psi*\tau\rangle & \langle|\tau|^2\rangle\end{pmatrix}. \qquad (B.2)$$

Angle brackets $\langle\cdot\rangle$ represent a long-time average. The matrix of eigenvectors $\mathbf{N} = \mathbf{N}(k,l)$ diagonalizes the covariance matrix so that

$$\mathbf{D} = \mathbf{N}*\mathbf{C}\mathbf{N} = \mathrm{diag}(e_+, e_-). \qquad (B.3)$$

Here * denotes the conjugate transpose. The EOFs are obtained by projecting the vertical modes onto the eigenspace defined by $\mathbf{N}$:

$$\begin{pmatrix}\chi^+ \\ \chi^-\end{pmatrix} = \mathbf{N}*(k,l)\begin{pmatrix}\psi \\ \tau\end{pmatrix}, \quad \langle|\chi^\pm|^2\rangle = e_\pm. \qquad (B.4)$$

Finally, the EOFs are expressed in terms of the layer streamfunctions via

$$\begin{pmatrix} \chi^+ \\ \chi^- \end{pmatrix} = \mathbf{V}(k,l) \begin{pmatrix} \psi^1 \\ \psi^2 \end{pmatrix}, \quad \mathbf{V}(k,l) = \mathbf{N}^*(k,l)\mathbf{M}(k,l).$$

$$\text{(B.5)}$$

APPENDIX C

Forecast Mean Solutions to the Stochastic Parameterized Extended Kalman Filter

The forecast means of the multiplicative and additive noise biases at time $t = t_1$ are straightforwardly obtained from the pathwise solutions (3.19b,c),

$$\langle m_1 \rangle = \overline{m} + (\langle m_0 \rangle - \overline{m})e^{-\lambda_m(t_1-t_0)},$$
$$\langle a_1 \rangle = \langle a_0 \rangle e^{-\lambda_a(t_1-t_0)}, \qquad \text{(C.1)}$$

where $\langle m_0 \rangle$, $\langle a_0 \rangle$ are the initial (analysis) means at time $t = t_0$. The forecast mean of $\hat{\chi}$ is

$$\langle \hat{\chi}_1 \rangle = \langle \hat{\chi}_0 e^{-j(t_0,t_1)} \rangle + \int_{t_0}^{t_1} dt \, \langle a_0 e^{-j(t,t_1)} \rangle e^{-\lambda_a(t-t_0)}.$$

$$\text{(C.2)}$$

The quantities in angle brackets can be calculated using the statistics of conditionally Gaussian processes: in particular, for real s, t, and real, mean-zero, Gaussian random variables x, y, z it can be shown (Gershgorin et al. 2010b) that

$$\langle x \exp(sy + \mathbf{i}tz) \rangle$$
$$= (s\langle xy \rangle + \mathbf{i}t\langle xz \rangle) \exp\left(\frac{s^2}{2}\langle y^2 \rangle - \frac{t^2}{2}\langle z^2 \rangle + \mathbf{i}st\langle yz \rangle\right).$$

$$\text{(C.3)}$$

Using $m_0' = m_0 - \langle m_0 \rangle$, $a_0' = a_0 - \langle a_0 \rangle$, and $\tau = \lambda_m^{-1}[e^{-\lambda_m(t-t_0)} - e^{-\lambda_m(t_1-t_0)}]$, we find

$$\langle \hat{\chi}_0 e^{-j(t_0,t_1)} \rangle = \langle \hat{\chi}_0 \rangle e^{-\overline{m}(t_1-t_0)}, \qquad \text{(C.4)}$$

$$\langle a_0 e^{-j(t,t_1)} \rangle = (\langle a_0 \rangle - \langle m_0'a_0' \rangle \tau)e^{-\langle m_0 \rangle \tau - \overline{m}(t_1-t-\tau)+\langle (a_0')^2 \rangle \tau^2/2}.$$

$$\text{(C.5)}$$

REFERENCES

Anderson, J., 2007: An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus,* **59A,** 210–224.

Baek, S.-J., B. Hunt, E. Kalnay, E. Ott, and I. Szunyogh, 2006: Local ensemble Kalman filtering in the presence of model bias. *Tellus,* **58A,** 293–306.

Bennett, A., 1984: Relative dispersion: Local and nonlocal dynamics. *J. Atmos. Sci.,* **41,** 1881–1886.

Branicki, M., B. Gershgorin, and A. Majda, 2011: Filtering skill for turbulent signals for a suite of nonlinear and linear extended Kalman filters. *J. Comput. Phys.,* **231,** 1462–1498.

Chelton, D. B., J. C. Ries, B. J. Haines, L.-L. Fu, and P. S. Callahan, 2001: Satellite altimetry. *Satellite Altimetry and Earth Sciences—A Handbook of Techniques and Applications,* L.-L. Fu and A. Cazenave, Eds., International Geophysical Series, Vol. 69, Academic Press, 1–131.

Chui, C., and G. Chen, 1999: *Kalman Filtering: With Real-Time Applications.* Springer, 243 pp.

Daley, R., 1991: *Atmospheric Data Analysis.* Cambridge University Press, 457 pp.

Dee, D., and A. Da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.,* **124,** 269–295.

DelSole, T., 2004: Stochastic models of quasigeostrophic turbulence. *Surv. Geophys.,* **25,** 107–149.

De Mey, P., and A. Robinson, 1987: Assimilation of altimeter eddy fields in a limited-area quasi-geostrophic model. *J. Phys. Oceanogr.,* **17,** 2280–2293.

Dowd, M., 2006: A sequential Monte Carlo approach for marine ecological prediction. *Environmetrics,* **17,** 435–455.

Ducet, N., P. Le Traon, and G. Reverdin, 2000: Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and -2. *J. Geophys. Res.,* **105** (C8), 19 477–19 498.

Durand, M., L.-L. Fu, D. Lettenmaier, D. Alsdorf, E. Rodriguez, and D. Esteban-Fernandez, 2010: The Surface Water and Ocean Topography mission: Observing terrestrial surface water and oceanic submesoscale eddies. *Proc. IEEE,* **98** (5), 766–779.

Ganachaud, A., and C. Wunsch, 2000: Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data. *Nature,* **408,** 453–457.

Gavart, M., and P. De Mey, 1997: Isopycnal EOFs in the Azores Current region: A statistical tool for dynamical analysis and data assimilation. *J. Phys. Oceanogr.,* **27,** 2146–2157.

Gershgorin, B., and A. Majda, 2008: A nonlinear test model for filtering slow-fast systems. *Comm. Math. Sci.,* **6** (3), 611–649.

——, J. Harlim, and A. Majda, 2010a: Improving filtering and predictions of spatially extended turbulent systems with model errors through stochastic parameter estimation. *J. Comput. Phys.,* **229,** 32–57.

——, ——, and ——, 2010b: Test models for improving filtering with model errors through stochastic parameter estimation. *J. Comput. Phys.,* **229,** 1–31.

Gnanadesikan, A., 1999: A simple predictive model for the structure of the oceanic pycnocline. *Science,* **283,** 2077–2079.

Greenslade, D., D. Chelton, and M. Schlax, 1997: The midlatitude resolution capability of sea level fields constructed from single and multiple satellite altimeter datasets. *J. Atmos. Oceanic Technol.,* **14,** 849–870.

Hall, M., and H. Bryden, 1982: Direct estimates and mechanisms of ocean heat transpor. *Deep-Sea Res.,* **29,** 339–359.

Harlim, J., 2011: Interpolating irregularly spaced observations for filtering turbulent complex systems. *SIAM J. Sci. Comput.,* **33** (5), 2620–2640.

——, and A. Majda, 2008a: Filtering nonlinear dynamical systems with linear stochastic models. *Nonlinearity,* **21,** 1281–1306.

——, and ——, 2008b: Mathematical strategies for filtering complex systems: Regularly spaced sparse observations. *J. Comput. Phys.,* **227,** 5304–5341.

——, and ——, 2010: Filtering turbulent sparsely observed geophysical flows. *Mon. Wea. Rev.,* **138,** 1050–1083.

Jayne, S. R., and J. Marotzke, 2002: The oceanic eddy heat transport. *J. Phys. Oceanogr.,* **32,** 3328–3345.

Kalman, R., 1960: A new approach to linear filtering and prediction problems. *J. Basic Eng. ASME,* **82,** 35–45.

Keating, S. R., K. S. Smith, and P. R. Kramer, 2011: Diagnosing lateral mixing in the upper ocean with virtual tracers: Spatial and temporal resolution dependence. *J. Phys. Oceanogr.,* **41,** 1512–1534.

Majda, A., and M. Grote, 2007: Explicit off-line criteria for stable accurate time filtering of strongly unstable spatially extended systems. *Proc. Natl. Acad. Sci. USA,* **104,** 1124–1129.

——, J. Harlim, and B. Gershgorin, 2010: Mathematical strategies for filtering turbulent dynamical systems. *Disc. Contin. Dyn. Syst.,* **27** (2), 441–486.

Phillips, N., 1954: Energy transformations and meridional circulations associated with simple baroclinic waves in a two-level, quasigeostrophic model. *Tellus,* **6,** 273–286.

Poje, A., A. Haza, T. Özgökmen, M. Magaldi, and Z. Garraffo, 2010: Resolution dependence of relative dispersion statistics in a hierarchy of ocean models. *Ocean Modell.,* **31** (1–2), 36–50.

Qiu, B., and S. Chen, 2005: Eddy-induced heat transport in the subtropical North Pacific from Argo, TMI, and altimetry measurements. *J. Phys. Oceanogr.,* **35,** 458–473.

Richards, K. J., N. A. Maximenko, F. O. Bryan, and H. Sasaki, 2006: Zonal jets in the Pacific Ocean. *Geophys. Res. Lett.,* **33,** L03605, doi:10.1029/2005GL024645.

Roemmich, D., and J. Gilson, 2001: Eddy transport of heat and thermocline water in the North Pacific: A key to interannual/decadal climate variability? *J. Phys. Oceanogr.,* **31,** 675–687.

Smith, K. S., 2007: The geography of linear baroclinic instability in Earth's oceans. *J. Mar. Res.,* **65,** 655–683.

——, and G. Vallis, 2002: The scales and equilibration of midocean eddies: Forced-dissipative flow. *J. Phys. Oceanogr.,* **32,** 1699–1720.

——, G. Boccaletti, C. Henning, I. Marinov, C. Tam, I. Held, and G. Vallis, 2002: Turbulent diffusion in the geostrophic inverse cascade. *J. Fluid Mech.,* **469,** 13–48.

Stammer, D., 1997: Global characteristics of ocean variability estimated from regional TOPEX/POSEIDEN altimeter measurements. *J. Phys. Oceanogr.,* **27,** 1743–1769.

——, 1998: On eddy characteristics, eddy transports, and mean flow properties. *J. Phys. Oceanogr.,* **28,** 727–739.

Tai, C.-K., 2004: The resolving power of a single exact-repeat altimetric satellite or a coordinated constellation of satellites. *J. Atmos. Oceanic Technol.,* **21,** 810–818.

Tulloch, R., J. Marshall, C. Hill, and K. Smith, 2011: Scales, growth rates, and spectral fluxes of baroclinic instability in the ocean. *J. Phys. Oceanogr.,* **41,** 1057–1076.

Vallis, G., 2006: *Atmospheric and Ocean Fluid Dynamics: Fundamentals and Large-Scale Circulation.* Cambridge University Press, 745 pp.

Volkov, D., T. Lee, and L.-L. Fu, 2008: Eddy-induced meridional heat transport in the ocean. *Geophys. Res. Lett.,* **35,** L20601, doi:10.1029/2008GL035490.

Wunsch, C., 1989: Sampling characteristics of satellite orbits. *J. Atmos. Oceanic Technol.,* **6,** 891–907.

——, 1999: Where do ocean eddy heat fluxes matter? *J. Geophys. Res.,* **104,** 13 235–13 249.

——, 2006: *Discrete Inverse and State Estimation Problems: With Geophysical Fluid Applications.* Cambridge University Press, 384 pp.

——, and D. Stammer, 1998: Satellite altimetry, the marine geoid, and the oceanic general circulation. *Annu. Rev. Earth Planet. Sci.,* **26,** 219–253.