Parametric Estimation of the Stochastic Dynamics of Sea Surface Winds

WILLIAM F. THOMPSON

Institute of Applied Mathematics, University of British Columbia, Vancouver, British Columbia, Canada

ADAM H. MONAHAN

School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia, Canada

DAAN CROMMELIN

Centrum voor Wiskunde & Informatica, Amsterdam, Netherlands

ABSTRACT

In this study, the parameters of a stochastic–dynamical model of sea surface winds are estimated from long time series of sea surface wind observational data. The model was introduced by A. H. Monahan, who developed an idealized model from a highly simplified representation of the momentum budget of a surface atmospheric layer of fixed depth. Such estimation of model parameters is challenging, in particular for a multivariate model with nonlinear terms as is considered here. The authors use a method developed recently by Crommelin and Vanden-Eijnden, which approaches the estimation problem variationally, finding the spectrally "best fit" stochastic differential equation to a time series of observations.

While the estimation procedure assumes forcing that is white in time, observed time series are generally better approximated as forced by red noise. Using a red-noise-forced linear system, the authors first show that the estimation procedure can still be used to estimate model parameters. Because the assumption of white noise is violated, these estimates lead to model autocorrelation functions that differ from the observed time series. Application of the estimation procedure to the wind data is further complicated by the fact that the boundary layer model is inconsistent with certain observed features of the wind. When these mismatches between the model and observations are accounted for, the estimation procedure generally results in parameter estimates consistent with the climatological features of the associated meteorological fields. Important exceptions to this result are the layer thickness and layer-top eddy diffusivity, which are poorly estimated where the vector winds are close to Gaussian.

1. Introduction

In Monahan (2006a), an idealized model was developed for the stochastic dynamics of sea surface winds, based on a highly simplified representation of the momentum budget of a surface atmospheric layer of fixed depth. This model results in an analytic expression for the probability distribution of surface winds in terms of physically meaningful parameters. The focus of this earlier study was on this probability distribution, which is of interest in the context of air–sea interactions (e.g.,

Jones and Toba 2001; Donelan et al. 2002; Fairall et al. 2003), wind power (e.g., Liu et al. 2008; Capps and Zender 2009), and wind extremes (Sampe and Xie 2007). No attention was paid to the temporal structure of the simulated winds. Neither was there an effort made to estimate model parameters from observations. In fact, this cannot be done using the probability distribution alone as it does not uniquely determine the model parameter set. In the present study, the model parameters are estimated from long time series of sea surface wind data.

The model from Monahan (2006a) consists of two coupled stochastic differential equations (SDEs). Parameter estimation for SDEs is a challenging task in general, and the fact that the SDE considered here is multivariate (two dimensional) and contains nonlinear terms adds to

Corresponding author address: William F. Thompson, Institute of Applied Mathematics, University of British Columbia, 121-1984 Mathematics Rd., Vancouver BC V6T 1Z2, Canada.
E-mail: wft@math.ubc.ca

the difficulty. Furthermore, the observational data may not exactly satisfy an SDE. For example, the data may not be Markov. Recently, Crommelin and Vanden-Eijnden have introduced a variational approach which avoids the restriction that the available data are exactly described by an SDE (Crommelin 2012; Crommelin and Vanden-Eijnden 2006b, 2011). In this approach, the data are fit with the "closest" SDE, in spectral terms. The method is computationally cheap, and it can handle two-dimensional SDEs and nonlinear terms.

Alternative estimation methods for this purpose—for example, Markov chain Monte Carlo methods [see Sørensen (2004) for a survey]—are often computationally very demanding, making them less suitable to process time series for many different spatial locations as will be done here. Moreover, they usually rely on model properties (e.g., a diagonal diffusion matrix) that can prove too restrictive for the model under consideration here.

The analysis in this study will demonstrate that the method of Crommelin and Vanden-Eijnden is applicable in situations when the data are not exactly described by the model to which they are fit. In particular, we will consider the effects of fitting a model driven by Gaussian white noise to data for which the driving variability has an autocorrelation time (ACT) that is not short relative to the time scale of the resolved dynamics. In earlier studies, the model was used as a tool to investigate the influence of large-scale and boundary layer processes on the probability distribution of surface winds. Obtaining estimates of model parameters is the first step in an assessment of the quantitative utility of the model. Furthermore, as we will show, this analysis can be used to identify model features that are irreconcilable with the data.

In section 2, we offer a brief description of the parameter estimation method, demonstrate the results of its application to simulated data from a simple SDE model, and discuss strategies used to improve parameter estimates. In section 3, we analyze the stochastic model for sea surface wind dynamics and consider the application of the estimation method to data generated by this model. Last, we estimate model parameters from long time series of sea surface winds in section 4. A discussion and conclusions are presented in section 5.

2. Method

A basic outline of the method of Crommelin and Vanden-Eijnden (2006b, 2011) for the parametric estimation of diffusions follows. Consider a stochastic process, $\mathbf{X}_t \in \mathbb{R}^d$, that is described by the SDE

$$d\mathbf{X}_t = \mathbf{b}(\mathbf{X}_t)dt + \mathbf{\Sigma}(\mathbf{X}_t)d\mathbf{W}_t, \tag{1}$$

where $\mathbf{b}$ is a vector function of length d and $\mathbf{\Sigma}$ is a $d \times d$ matrix function. This equation is interpreted in the sense of an Ito SDE (Gardiner 1985; Øksendal 2003). Governed by these dynamics, $\mathbf{X}_t$ has a probability density function (pdf) whose evolution can be expressed as

$$p(\mathbf{x}, t + \delta t) = \mathcal{P}_{\delta t}^\dagger p(\mathbf{x}, t) \tag{2}$$

(where $\mathcal{Q}^\dagger$ denotes the formal adjoint of the operator $\mathcal{Q}$). The operator $\mathcal{P}_{\delta t}$ is given by

$$\mathcal{P}_{\delta t} f(\mathbf{x}) = \mathbb{E}[f(\mathbf{X}_{t+\delta t}) \mid \mathbf{X}_t = \mathbf{x}] \tag{3}$$

and is known as the conditional expectation operator (Gobet et al. 2004). The infinitesimal generator $\mathcal{G}$ for the diffusion process is given by

$$\mathcal{G} = \sum_{i=1}^d \mathbf{b}_i \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d (\mathbf{\Sigma}\mathbf{\Sigma}^T)_{ij} \frac{\partial^2}{\partial x_i \partial x_j}, \tag{4}$$

where $\mathbf{\Sigma}^T$ indicates the transpose of the matrix $\mathbf{\Sigma}$. It is a standard result that the infinitesimal generator (referred to as "the generator") and the conditional expectation operator satisfy the following operator equation:

$$\mathcal{P}_t = \exp(t\mathcal{G}). \tag{5}$$

Formally, Eq. (2) is equivalent to the Fokker–Planck equation. When the dynamics admit a stationary process, the leading eigenvalue for the operator $\mathcal{G}$ will be zero, while all others will be strictly negative.

A particular set of observations that one desires to model as an SDE may not be generated by dynamics of the form of Eq. (1). For example, the data may be non-Markovian because only a projection of the full state space is sampled. Rather than require the data come exactly from an SDE, the approach of Crommelin and Vanden-Eijnden finds the closest SDE to the data in terms of the eigenstructure of the generators of the model and the data. In particular, this approach minimizes the residual of the eigenproblem $\|\mathcal{G}\hat{\phi} - \hat{\lambda}\hat{\phi}\|_{2,w}$ (the subscript "2, w" denotes a weighted Euclidean norm that will be described soon), where $\mathcal{G}$ is the generator of the model and $\hat{\phi}$ and $\hat{\lambda}$ represent the eigenfunctions and eigenvalues of the generator, estimated from the data (throughout the text, estimated quantities will be denoted by a caret). In this sense, the approach finds the closest SDE model to the data.

The method of Crommelin and Vanden-Eijnden requires the estimation of the eigenstructure of the generator from data. From Eq. (5), there is a direct relationship

between the eigenstructures of $\mathcal{P}_t$ and $\mathcal{G}$: they have the same eigenfunctions and the eigenvalues can be related via a simple transformation:

$$\mathcal{P}_t\hat{\phi}_k = \hat{\Lambda}_k\hat{\phi}_k \leftrightarrow \mathcal{G}\hat{\phi}_k = \hat{\lambda}_k\hat{\phi}_k, \quad \text{where} \quad \hat{\lambda}_k = \frac{1}{t}\log(\hat{\Lambda}_k). \tag{6}$$

We can exploit this relationship between eigenstructures to obtain the eigenstructure of $\mathcal{G}$ from that of $\mathcal{P}_t$, as the latter is relatively easy to estimate from time series data. We can also define the adjoint eigenproblem: $\mathcal{P}_t^\dagger(\hat{\xi}_l\rho) = \overline{\hat{\Lambda}_l}\hat{\xi}_l\rho$, where ρ denotes the stationary distribution of Eq. (1). The eigenstructure of the conditional expectation operator is estimated by using a truncated Galerkin approximation which solves a modified version of the eigenproblem in Eq. (6). By approximating the eigenfunction $\hat{\phi}_k(\mathbf{x})$ by an expansion over n basis functions $\{\beta_i\}$ [i.e., $\hat{\phi}_k(\mathbf{x}) \approx \sum_{i=1}^n v_{ki}\beta_i(\mathbf{x})$] chosen such that they are square integrable with respect to the invariant measure (the stationary probability density function for $\mathbf{X}_t$), we obtain the following eigenproblem:

$$\sum_{i=1}^n v_{ki}\langle\mathcal{P}_{\delta t}\beta_i(\mathbf{X}_t),\beta_j(\mathbf{X}_t)\rangle = \Lambda_k\left[\sum_{i=1}^n v_{ki}\langle\beta_i(\mathbf{X}_t),\beta_j(\mathbf{X}_t)\rangle\right],$$
$$1 \le j, \quad k \le n, \tag{7}$$

where v_{ki} is the ith entry in the kth eigenvector and Λ_k is the corresponding eigenvalue. For the adjoint eigenproblem, we can define a similar weak form by expanding $\hat{\xi}_l$ onto the same basis [i.e., $\hat{\xi}_l(\mathbf{x}) \approx \sum_{j=1}^n w_{lj}\beta_j(\mathbf{x})$]. By the properties of the conditional expectation operator and our definition of the inner product,

$$\langle\mathcal{P}_{\delta t}\beta_i(\mathbf{X}_t),\beta_j(\mathbf{X}_t)\rangle = \mathbb{E}[\beta_i(\mathbf{X}_{t+\delta t})\beta_j(\mathbf{X}_t)], \tag{8}$$

$$\langle\beta_i(\mathbf{X}_t),\beta_j(\mathbf{X}_t)\rangle = \mathbb{E}[\beta_i(\mathbf{X}_t)\beta_j(\mathbf{X}_t)]. \tag{9}$$

It is not possible in general to perform a complete decomposition of ϕ_k into a finite number of basis functions, but this is not required for this procedure. Using the sample mean as approximations in Eqs. (8) and (9), we create the linear system governing the eigenvalue problem [Eq. (7)] and solve it numerically to get estimates $\{\hat{\Lambda}_k\}$ for the eigenvalues of the conditional expectation operator. Furthermore, we obtain estimates of the projections of the eigenfunctions $\hat{\phi}_k$ onto the basis β_i.

With a set of estimated eigenvalues and projected eigenfunctions, we use Eq. (6) to obtain estimates of the eigenvalues for the generator $\{\hat{\lambda}_k\}$. Assuming a particular

form for the model SDE, its generator can be constructed and the objective function is given by

$$\|\mathcal{G}\hat{\phi} - \hat{\lambda}\hat{\phi}\|_{2,w} = \sum_{k,l=1}^n m_{kl}|\langle\mathcal{G}(\{b_i\},\{a_j\})\hat{\phi}_k,\hat{\xi}_l\rangle$$
$$- \hat{\lambda}_k\langle\hat{\phi}_k,\hat{\xi}_l\rangle|^2, \tag{10}$$

where, for the models under consideration, the drift and diffusion of the generator can be expressed in the following form:

$$\mathbf{b}(\mathbf{x}) = \sum_{i=1}^{N_b} b_i\mathbf{g}_i(\mathbf{x}), \quad \mathbf{A}(\mathbf{x}) = \mathbf{\Sigma\Sigma}^\mathrm{T}(\mathbf{x}) = \sum_{i=1}^{N_a} a_i\mathbf{H}_i(\mathbf{x}) \tag{11}$$

and m_{kl} are weight coefficients (to be discussed later). The objective function [Eq. (10)] proposed in Crommelin and Vanden-Eijnden (2011) is defined in this work as E^g and other choices for objective functions are offered. The minimization of Eq. (10) is performed with respect to the parameters $\{b_i\}$, $\{a_i\}$ and can be done using a least squares or quadratic programming minimization (Crommelin and Vanden-Eijnden 2006a). If $\mathbf{b}(\mathbf{x})$ and $\mathbf{A}(\mathbf{x})$ cannot be expressed in the form given by Eq. (11), the estimation procedure can still be applied but a different minimization technique must be used. We will now demonstrate the application of the estimation procedure to a simple stochastic process, for which analytical results are available.

a. The Ornstein–Uhlenbeck process

As an illustration of the application of the parameter estimation method, we consider the univariate Ornstein–Uhlenbeck process (OUp) $x = x(t)$, where,

$$dx = \mu x dt + \sigma dW, \quad x(0) = x_0, \tag{12}$$

where $W = W(t)$ is a standard Brownian motion. The constant μ is chosen to be less than zero to ensure that the process possesses stationary solutions. The univariate OUp is one of the simplest stationary continuous time stochastic processes and is easily studied analytically. The generator for the OUp is

$$\mathcal{G} = \mu x\frac{\partial}{\partial x} + \frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2}. \tag{13}$$

It can easily be shown that the eigenvalues of the generator for this process are the nonnegative integer multiples of μ. Simulations of Eq. (12) with $\mu = -1$ and $\sigma = 1$ were generated, from which the eigenvalues of the conditional expectation operator $\mathcal{P}_{\delta t}$ were estimated (where δt is the time step between data points). To assess sampling variability, 100 independent realizations of
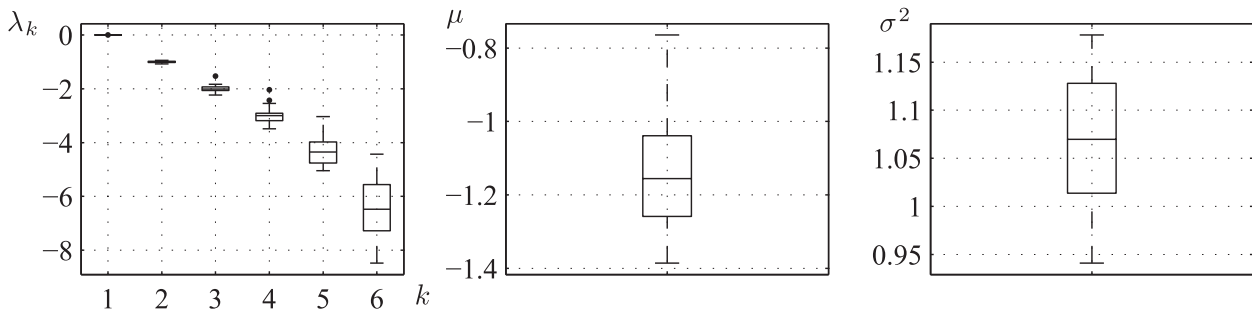
FIG. 1. Estimates of the eigenvalues and the coefficients of the generator of the OUp with $\mu = -1$ and $\sigma = 1$. The time series used to estimate the eigenvalues contains 30 000 points with $\delta t = 0.1$. Six eigenvalues are considered to obtain parameter estimates and no weighting is applied when minimizing the objective function. Sampling variability was assessed by estimating these parameters from 50 independent realizations of the random process. The bottom and top of the boxes span the lower to upper quartiles with the median drawn in between. The whiskers extend to a maximum of 1.5 times the interquartile range and the black dots indicate values lying outside of this range.

$x(t)$ were generated. From these, the corresponding estimates for the eigenvalues of the generator were estimated using Eq. (6) (Fig. 1). Not surprisingly, there is some error in the estimates of the eigenvalues, resulting from truncation of the spectrum and sampling variability. Consistent with the results of previous studies (Crommelin and Vanden-Eijnden 2011), eigenvalues of higher index tend to have greater sampling variability and bias, while the first few eigenvalues (associated with the stationary distribution and the slowest decaying modes) are the most robustly estimated.

Having estimated the eigenvalues of the generator, we now focus on parametric estimates of the generator itself. We assume that the simulated data satisfies an OUp as given by Eq. (12) and minimize the norm $\|\mathcal{G}(\{a_i\}, \{b_i\})\hat{\phi} - \hat{\lambda}\hat{\phi}\|_{2,w}$ with respect to $\{a_i\}$, $\{b_i\}$ such that the drift and diffusion coefficients of the generator of $x(t)$ are given by

$$b(x) = \mu x, \quad A(x) = \sigma^2. \tag{14}$$

The estimates $\hat{\mu}$ and $\widehat{\sigma^2}$ are shown in Fig. 1. Although the estimated values are on average close to the true values, the estimates are clearly biased. Strategies to reduce this bias will be presented in section 2b.

When estimating the parameters of $\mathcal{G}$ from a specified SDE model, it is possible that the generator may be overspecified, in that there are terms within the expressions for the drift and diffusion with coefficients that are identically equal to zero in the dynamics that generated the data. To assess the robustness of the approach to model overspecification for the case of data drawn from OUp, we fit the data to the generator for which

$$b(x) = \sum_{i=0}^{3} b_i x^i, \quad A(x) = \sum_{i=0}^{3} a_i x^i. \tag{15}$$

Ideally, the method should return values of $\{a_i\}$, $\{b_i\}$ that are zero except for $a_0 = \sigma^2$ and $b_1 = \mu$. As expected, these estimates are found to be distributed around the values we predict, although there are some biases (Fig. 2).

b. Weighting of eigenvalues

As was illustrated in Fig. 1, the sampling variability of eigenvalues increases with eigenvalue order. Depending on the problem under consideration, it is important that a sufficient number of eigenvalues be estimated so as to avoid possible degeneracies in the generator (i.e., having an identical pdf for different parameter sets) and to capture the temporal structure of the stochastic process. For example, to estimate the parameters of an Ornstein–Uhlenbeck process, we require at least two eigenvalues as the pdf is determined by the ratio μ/σ^2. This requirement must be balanced against the tendency of estimates of higher-order eigenvalues to be biased.

To retain higher-order eigenvalues in estimates of the values of the coefficients $\{a_i\}$, $\{b_i\}$ while reducing their contribution to the objective function in order to account for their greater uncertainty, we use a weighted least squares method. We choose to use a weighting scheme having weights m_{ij} where

$$m_{ij} = w^{-\lambda_i/\lambda_2}, \quad w > 1, \quad i = 1, 2, \ldots, \quad (\lambda_1 = 0). \tag{16}$$

This weighting scheme assigns the first eigenvalue a weight of 1 and subsequent weights $1 \geq w_2 = w^{-1} \geq w_3 \geq w_4 \cdots$ (since all $\lambda_i \leq \lambda_2 < 0$ for $i \geq 2$ by stationarity). We choose this weighting scheme so that eigenvalues of similar magnitude are penalized by a similar amount and eigenvalues with greater magnitude are penalized more than smaller ones. In principle, one could assign a weight of zero to eigenvalues beyond a certain index, but in
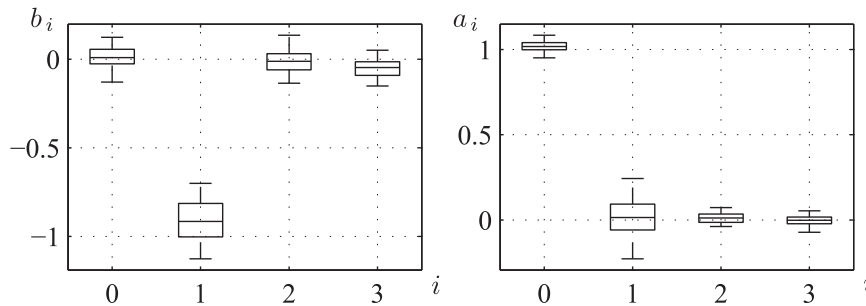
FIG. 2. Estimates for the coefficients of the overspecified model given by Eq. (15) for data generated from the OUp with $\mu = -1$ and $\sigma = 1$, illustrated as in Fig. 1. The time series used to estimate the eigenvalues are 30 000 points long with $\delta t = 0.1$. Note that we have estimated six eigenvalues, and the parameter estimates are slightly biased.

cases in which groups of eigenvalues are close or occur as complex conjugate pairs it is undesirable to give a weight of zero to one eigenvalue and a nonzero weight to another.

Applying this weighting scheme to the variational estimates of the OUp parameters, we find that there is an improvement in the parameter estimates, such that both bias and sampling variance are reduced (Fig. 3). Note that in this approach, w should not be allowed to become too large as that effectively puts all the weight on the first eigenmode, which can lead to degenerate parameter estimates if it depends on a combination of parameters (as is the case for the OUp).

c. The effect of correlated noise

A potential source of mismatch between the data and the model to which they are fit is the assumption that the data are Markov and described by a diffusion process. The Crommelin–Vanden-Eijnden method assumes that data are Markov and that the model to which they are fit is an SDE driven by Gaussian white

noise. Real-world processes are often better modeled by forcing that is correlated in time (e.g., red noise) and so there is a potential discrepancy between the data and the white-noise-driven model to which they are fit. If the driving red-noise process is an Ornstein–Uhlenback process and is directly observed, then the dynamics can be expressed as an SDE in an extended state space with white-noise forcing. However, if the red-noise process cannot be directly observed and its dynamics not accounted for in the generator estimation method, it is still possible to estimate the stochastic dynamics of the data (albeit resulting in biased parameter estimates) provided that the data are subsampled with a coarse-enough sampling interval so that the red-noise effects can effectively be "whitened." This issue is considered in Crommelin and Vanden-Eijnden (2011) for the asymptotic limit in which the ACT of the red-noise forcing (modeled as an OUp) approaches zero. Here, we consider ACT scales that are not small relative to the characteristic time scales of the resolved dynamics.
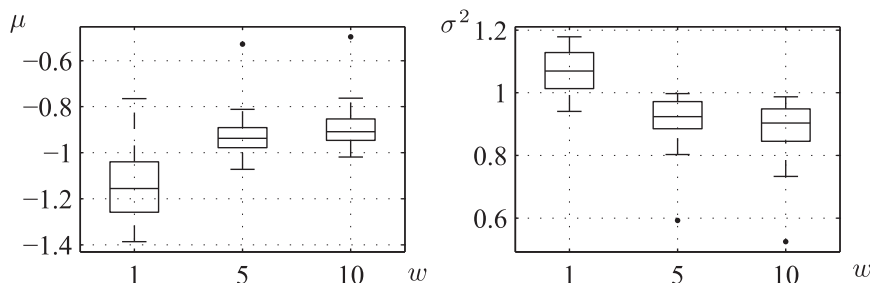

FIG. 3. Estimates of μ and σ^2 when weighting is applied to the eigenvalues in the objective function. The value of w is indicated on the horizontal axis. Because of degeneracies in the generator for the OUp, the estimates of μ and σ^2 are biased to values closer to 0 when w is large. Since the pdf of the OUp depends only on the quantity μ/σ^2, this quantity is well estimated when the lowest eigenvalue is heavily weighted—although μ and σ^2 are often not well estimated themselves when higher-order eigenvalues are suppressed. The box plots are drawn as described in Fig. 1.

We consider a linearly damped process driven by an OUp:

$$dx = -\frac{1}{\tau_x}x\,dt + \frac{1}{\sqrt{\tau_y}}y\,dt, \quad dy = -\frac{1}{\tau_y}y\,dt + \frac{\sigma}{\sqrt{\tau_y}}dW.$$
(17)

It can be shown that in the limit $\tau_y \to 0$ the dynamics of x in Eq. (17) approach those of the univariate OUp given by Eq. (12). However, when τ_y and τ_x are of approximately the same order then the red-noise forcing will cause noticeable differences in the dynamics of x from those of the univariate OUp. Specifically, it can be shown via the Wiener–Khinchin theorem (Gardiner 1985) that the autocovariance function for the variable x in Eq. (17) is given by

$$C_{xx}(T) = \frac{\tau_x^2\sigma^2}{2(\tau_x^2 - \tau_y^2)}\left[\tau_x \exp\left(-\frac{|T|}{\tau_x}\right) - \tau_y \exp\left(-\frac{|T|}{\tau_y}\right)\right].$$
(18)

Having generated a realization of $x(t)$ from Eq. (17) we will fit it to a univariate (white-noise driven) OUp and interpret these results in the context of a univariate OUp that is statistically similar to $x(t)$ in the sense that the stationary variance and ACT scales are equal. Using the autocovariance function above, we can determine the ACT and variance for x:

$$\mathrm{ACT}[x] = \int_0^\infty \frac{C_{xx}(T)}{C_{xx}(0)}dT = \tau_x + \tau_y,$$

$$\mathrm{Var}[x] = C_{xx}(0) = \frac{\sigma^2\tau_x^2}{2(\tau_x + \tau_y)}.$$
(19)

Hence, the equivalent univariate OUp, $\tilde{x}$, satisfies the following SDE:

$$d\tilde{x} = -\frac{1}{\tau_x + \tau_y}\tilde{x}\,dt + \frac{\sigma\tau_x}{\tau_x + \tau_y}d\tilde{W}.$$
(20)

We note that in the $\tau_y \to 0$ limit, Eq. (20) is identical to the result obtained from stochastic homogenization theory (Pavliotis and Stuart 2007). For the fit of $x(t)$ to a univariate time series to be meaningful, it is expected that the parameter estimates should yield estimates consistent with Eq. (20). In fact, applying the estimation procedure to x does not generally yield results that are consistent with Eq. (20) (Fig. 4). When the resolution time step δ_t is of the same order as τ_y, then the noise is not sufficiently decorrelated between time steps to be approximated as "white." Despite the fact that the separation of time scales between $x(t)$ and $y(t)$ may be large, the correlated nature of the noise is still apparent as the autocorrelation function (ACF) for $x(t)$ displays the concave-downward structure at the origin that is characteristic of a non-Markovian forcing (DelSole 2000). Thus, $x(t)$ cannot be well approximated as Markov on the time scale of the resolution of the time series. The Markov assumption in the method results in fitting the data to match short time correlation behavior, causing significant overestimation of the ACT scale. To address this issue, we simply increase the resolution time scale.

This increase in resolution time scale is achieved through data subsampling. Before the eigenstructure of the conditional expectation operator is estimated, we reorder the data such that the time interval between successive points is increased. For example, subsampling with stepsize $2\delta t$ yields the reordering:

$$x_1, x_2, x_3, x_4, \ldots, x_n \to x_1, x_3, x_5, \ldots, x_n, x_2, x_4, \ldots x_{n-1}$$
$$(\text{if } n \text{ is odd}).$$
(21)

By concatenating the subsets of subsampled data, no data are thrown away. Subsampling increases the separation of the red-noise ACT scale, τ_y from that of the resolution time step δt of the data from which the conditional expectation operator is constructed, effectively "whitening" the correlated forcing. It should be noted that minor errors are introduced through the process of concatenating the data subsets (e.g., where x_n is followed by x_2). These errors could be corrected by ignoring the data around the transition points, but given that the number of such points in the reordered time series is much smaller than the number of points in the series itself, the error introduced is negligible.

Applying various degrees of subsampling to an Ornstein–Uhlenbeck process where the forcing is a red-noise process, the estimates of the parameters converge to the expected values of the equivalent white-noise process (Fig. 4). Note that while subsampling results in mean parameter estimates that are closer to those expected for an equivalent OUp [Eq. (20)], the sampling variance of the estimated parameters increases with the degree of subsampling, although the effect is marginal for low degrees of subsampling. Although not shown in Fig. 4, the variance in the parameter estimates increases dramatically if the subsampling degree is increased beyond 10. Finally, when coarsening the resolution of the time series, it is important that the degree of subsampling is not so large that all information about serial dependence of $x(t)$ itself is lost. Loss of this information
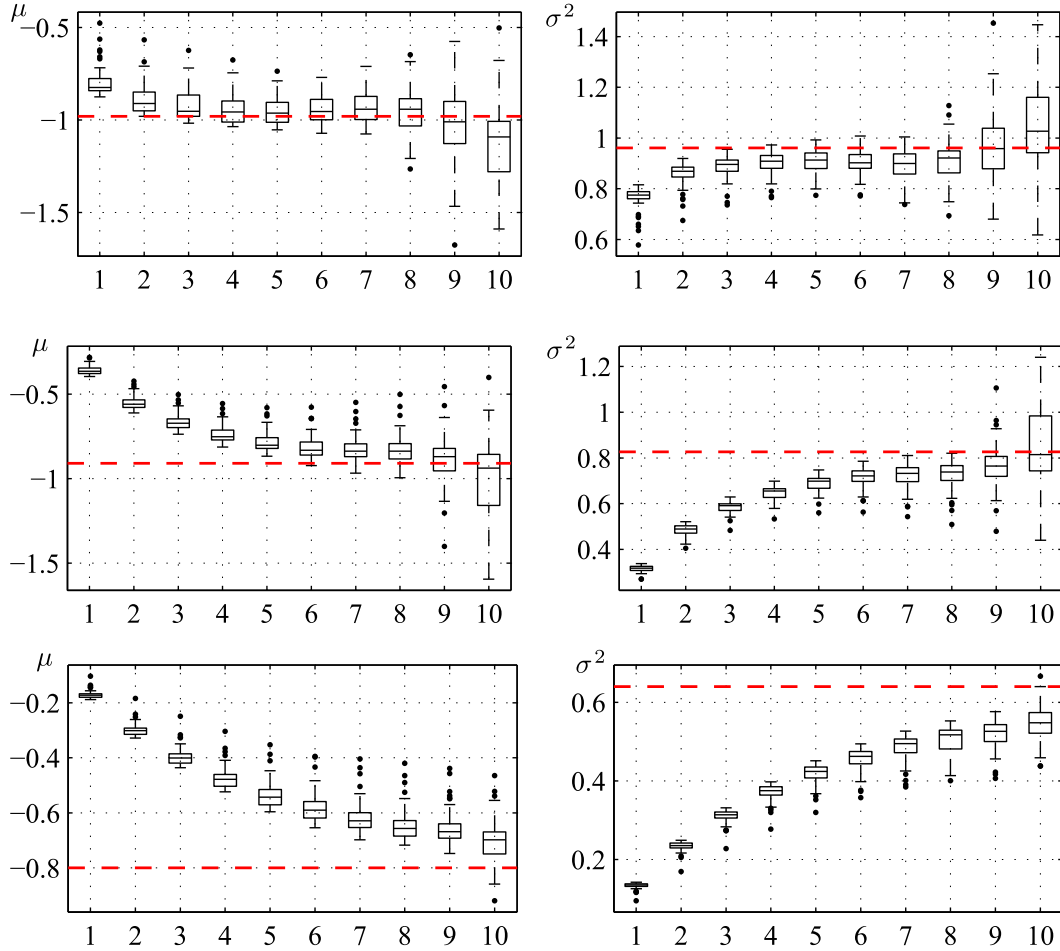
FIG. 4. Estimates for the coefficients μ and σ^2 of the generator when the data are generated by the system Eq. (17). In both figures, $\tau_x = 1$. The ACT scales for y are respectively $\tau_y =$ (top) 0.02, (middle) 0.1, and (bottom) 0.25. Ensembles were computed from 50 time series, each of length 50 000 points with $\delta t = 0.1$. The horizontal axis indicates the degree of subsampling used in the estimations (1 = no subsampling), and the red dashed line indicates the values of the coefficients for the equivalent OUp as determined by Eq. (20). Weighting is applied to offset biased estimates of the higher-order eigenvalues ($w = 2$).

will prevent accurate estimation of eigenmodes beyond the first, and therefore corrupt all parameter estimates.

3. The sea surface wind model

The sea surface wind model that we will consider is a slab model of the lower-atmospheric momentum budget introduced by Monahan (2004, 2006a). It is assumed in this model that vector wind tendencies result from imbalances between surface drag, downward mixing of momentum from above the slab layer, and "large-scale ageostrophic forcing" (the sum of pressure gradient and Coriolis forces), which is decomposed into a mean and fluctuations which are modeled as white noise. With u and v respectively the

zonal and meridional components of the wind vector, the model is given by the nonlinear SDE,

$$du = \left(\langle \Pi_u \rangle - \frac{K}{h^2}u - \frac{c_d}{h}u\sqrt{u^2 + v^2}\right)dt$$
$$+ \sigma_{11}dW_1 + \sigma_{12}dW_2, \qquad (22)$$

$$dv = \left(\langle \Pi_v \rangle - \frac{K}{h^2}v - \frac{c_d}{h}v\sqrt{u^2 + v^2}\right)dt$$
$$+ \sigma_{21}dW_1 + \sigma_{22}dW_2, \qquad (23)$$

where W_1 and W_2 represent (mutually uncorrelated) standard Brownian motions. The parameters $\langle \Pi_u \rangle$ and $\langle \Pi_v \rangle$ represent the mean large-scale driving for their respective components, while the Brownian motion

terms represent stochastic fluctuations of this driving. The turbulent exchange of momentum between the slab layer of depth h and the atmosphere above is represented as a coarsely finite-differenced diffusion with eddy viscosity K. The surface turbulent momentum flux is parameterized with a standard bulk drag law with drag coefficient c_d. The coefficients of the noise terms can be expressed as components of a matrix $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}. \tag{24}$$

The model [Eqs. (22) and (23)] is of the form given by Eq. (1) and has a generator given by

$$\mathcal{G}f = \left(\langle \Pi_u \rangle - \frac{K}{h^2} u - \frac{c_d}{h} u \sqrt{u^2 + v^2} \right) \frac{\partial}{\partial u} f + \left(\langle \Pi_v \rangle - \frac{K}{h^2} v - \frac{c_d}{h} v \sqrt{u^2 + v^2} \right) \frac{\partial}{\partial v} f$$
$$+ \frac{1}{2} \left[(\sigma_{11}^2 + \sigma_{12}^2) \frac{\partial^2}{\partial u^2} + 2(\sigma_{11}\sigma_{12} + \sigma_{21}\sigma_{22}) \frac{\partial^2}{\partial u \partial v} + (\sigma_{21}^2 + \sigma_{22}^2) \frac{\partial^2}{\partial v^2} \right] f \tag{25}$$

so the model parameters $\langle \Pi_u \rangle$, $\langle \Pi_v \rangle$, K, h, and σ_{ij} can be estimated from surface wind data using the method under consideration. We first cast the SDE in the form Eq. (11) by defining

$$\mathbf{b}(u, v) = b_0 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b_1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + b_2 \begin{pmatrix} u \\ v \end{pmatrix}$$
$$+ b_3 \begin{pmatrix} u\sqrt{u^2 + v^2} \\ v\sqrt{u^2 + v^2} \end{pmatrix}, \tag{26}$$

$$\mathbf{A}(u, v) = a_0 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + a_1 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + a_2 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$
$$(\mathbf{A} = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathrm{T}}). \tag{27}$$

The estimation routine will be used to determine the parameters $\{b_i\}, i = 0, 1, 2, 3$ and $\{a_j\}, j = 0, 1, 2$. Throughout these calculations, we assume that the parameter c_d is fixed at 1.3×10^{-3}. In fact, the drag coefficient is a function of wind speed and surface stratification (e.g., Jones and Toba 2001). This dependence is neglected in order to simplify the calculations. Note that because of the form that we have assumed for the wind model, we cannot directly estimate the values of the coefficients σ_{ij} owing to nonuniqueness of the square root of a matrix, but estimating $\{a_j\}$ gives us estimates for the entries of $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathrm{T}}$ ($a_0 = \sigma_{11}^2 + \sigma_{12}^2$, $a_1 = \sigma_{11}\sigma_{21} + \sigma_{21}\sigma_{22}$, $a_2 = \sigma_{21}^2 + \sigma_{22}^2$). We use polynomial basis functions in u and v up to and including degree 2 in the estimation procedure: $\{1, u, v, u^2, uv, v^2\}$. This choice of basis is motivated by the relative simplicity of the functions and the infinite domain: $(u, v) \in \mathbb{R}^2$.

a. Properties of the wind model

Before considering parameter estimates for this model, we first review some of its features. We will then investigate the ability of the estimation procedure to recover model parameters from time series generated by the model itself.

1) REVERSIBILITY

A stochastic process is defined to be reversible if the equations governing its behavior satisfy detailed balance conditions (Risken 1989). Sufficient conditions for reversibility are that the diffusion matrix is a constant proportional to the identity matrix and that the drift can be expressed as the gradient of a potential. Reversibility of the process is of practical utility in the present context as it regularizes calculations in the parameter estimation (Crommelin and Vanden-Eijnden 2011). Results from previous studies (e.g., Monahan 2006a, 2007) suggest that $\boldsymbol{\Sigma}$ is diagonal to a good approximation. This indicates that the process [Eqs. (22) and (23)] is close to reversible although possibly not exactly so. In estimating parameters, we will make the approximation that the system is reversible.

2) STATIONARY PROBABILITY DENSITY FUNCTION

When the noise intensity matrix $\boldsymbol{\Sigma}$ is diagonal ($\sigma_{ij} = \zeta\delta_{ij}$), we can obtain an explicit expression for the stationary pdf for u, v, denoted p_{uv}:

$$p_{uv}(u, v) = \mathcal{N} \exp \left\{ \frac{2}{\zeta^2} \left[\langle \Pi_u \rangle u + \langle \Pi_u \rangle v - \frac{K}{2h^2}(u^2 + v^2) - \frac{c_d}{3h}(u^2 + v^2)^{3/2} \right] \right\}, \tag{28}$$

where $\mathcal{N}$ is the normalization constant (Monahan 2006a). We immediately see that for the probability density function to be bounded, we must have $h > 0$. This constraint is of course physically necessary given

the interpretation of h as an atmospheric-layer thickness. From inspection of this stationary pdf, we see that there is a degeneracy with respect to the parameters such that different sets of the parameters ($\langle\Pi_u\rangle$, $\langle\Pi_v\rangle$, K, c_d, h, σ) will yield the same probability density function. In particular, the pdf is determined by the following four quantities:

$$\left\{ \frac{\langle\Pi_u\rangle}{\zeta^2}, \frac{\langle\Pi_v\rangle}{\zeta^2}, \frac{c_d}{h\zeta^2}, \frac{K}{h^2\zeta^2} \right\}. \tag{29}$$

Under parameter changes in which these quantities are conserved the probability density function will remain unchanged. While we do not have an expression for the pdf for the anisotropic or correlated noise model ($\sigma_{11} \neq \sigma_{22}$

TABLE 1. The base-case parameters used to generate realizations from the wind model [Eqs. (21) and (22)].

Parameter	Value
$\langle\Pi_u\rangle$	$1.6\,\mathrm{km\,h^{-2}}$
$\langle\Pi_v\rangle$	$0.8\,\mathrm{km\,h^{-2}}$
c_d	1.3×10^{-3}
K	$4.5 \times 10^{-2}\,\mathrm{km^2\,h^{-1}}$
h	$1\,\mathrm{km}$
σ_{11}, σ_{22}	$12.2\,\mathrm{km\,h^{-3/2}}$
σ_{12}, σ_{21}	0

or σ_{21}, $\sigma_{12} \neq 0$), inspection of the Fokker–Planck equation shows that the pdf is invariant so long as the following quantities are conserved:

$$\left\{ \frac{\langle\Pi_u\rangle}{\sigma_{11}^2 + \sigma_{12}^2}, \frac{\langle\Pi_v\rangle}{\sigma_{11}^2 + \sigma_{12}^2}, \frac{c_d}{h(\sigma_{11}^2 + \sigma_{12}^2)}, \frac{K}{h^2(\sigma_{11}^2 + \sigma_{12}^2)}, \frac{\sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22}}{\sigma_{11}^2 + \sigma_{12}^2}, \frac{\sigma_{21}^2 + \sigma_{22}^2}{\sigma_{11}^2 + \sigma_{12}^2} \right\}. \tag{30}$$

By rescaling these parameters in a particular way, we can modify the time scale of the process without modifying the pdf. As we show in section 4, this is a useful tool for dealing with model mismatch in this particular problem.

b. Estimating model parameters from simulated data

To evaluate the estimation of model parameters in a "perfect model" framework, we will now consider applying the estimation method to time series simulated from the stochastic wind model. For these simulations, we take parameter values that result in wind statistics within the range of the observed statistics and simulate time series with the 6-h resolution of the surface wind data that we will consider in section 4. The parameters that we use for our simulations are given in Table 1. A forward Euler scheme (Kloeden and Platen 1992) is used to integrate the model with simulation time step $dt = 0.001$ h, for a duration of 45 years. The model is resolved every 6 h (yielding a time series with a length of 65 700 data points).

1) WEIGHTING OF THE EIGENVALUES IN THE OBJECTIVE FUNCTION

To assess the effect that weighting the eigenvalues has on the quality of the estimation, we apply weights (as described in section 2) to the estimation procedure. The results of these calculations are shown in Fig. 5. For all values of the weight parameter, we see that model parameters are estimated well by the method. We note that an increasing weight tends to improve the estimates of all parameters in that the median of the estimated values is closer to the true value. For some parameters, the sample variance

decreases with increased weighting, while in other cases it increases. These results reinforce the result that weighting generally improves the parameter estimates, although in the present case there is only a modest dependence of the recovered parameters on the weight value.

2) THE EFFECT OF RED NOISE

The assumption of white-noise forcing of the atmospheric-layer momentum budget is a useful approximation, but it is physically unrealistic. In fact, we expect that fluctuations in the "large-scale forcing" should occur on similar time scales as fluctuations in the surface winds themselves. Replacing the white-noise forcing terms dW_1 and dW_2 of Eqs. (22) and (23) (with $\sigma_{12} = \sigma_{21} = 0$) with red-noise forcing terms $\eta_1 dt$ and $\eta_2 dt$ such that

$$d\eta_i = -\frac{1}{\tau_i}\eta_i dt + \frac{\sigma}{\tau_i}dW_{i+2}, \quad \eta_i = \eta_{i,0} \quad \text{at}$$
$$t = 0, \quad i = 1, 2 \tag{31}$$

results in changes to the dynamics that the generator estimation scheme cannot account for (when the procedure is applied to time series of u and v alone). The influence of red-noise forcing on model parameter estimates was tested using a range of forcing ACT scales τ_i (Fig. 6). We see that when the red-noise forcing is close to white (i.e., τ_i is close to zero), the eigenvalue and parameter estimates are close to the true values, as was the case with white-noise forcing. In contrast, eigenvalue and parameter estimates where the ACT is on the order of the resolution of the data show significant deviations from the true values.
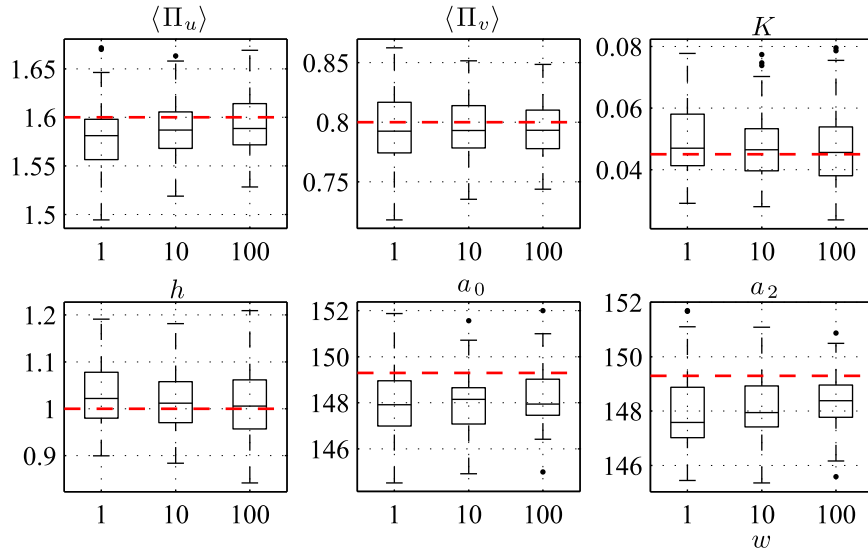
FIG. 5. The influence of w on estimates of various parameters from the wind model using simulated data from Eqs. (22) and (23). The values of w are shown on the horizontal axis for each parameter boxplot. The true value of each parameter is indicated by the dashed red line. No subsampling of the data was carried out.

When the system is forced with red noise, the ACT scale of the measured trajectory is increased. If time t is rescaled to $t = \alpha t^*$, then the initial parameter estimates can be rescaled as

$$h = \alpha h^*, \quad \langle \Pi_i \rangle = \frac{1}{\alpha} \langle \Pi_i \rangle^*,$$

$$K = \alpha K^*, \quad \sigma_{ij} = \frac{1}{\sqrt{\alpha}} \sigma_{ij}^*. \tag{32}$$

As the ACT scale of the white-noise forced model is directly scaled by the value of h, the estimate of h increases to match the ACT scale of the data. To conserve the pdf the values of the square of the diffusion matrix (a_0, a_1, a_2) must decrease to h increases. The values of $\langle \Pi_u \rangle$ and $\langle \Pi_v \rangle$ decrease by a factor close to that of the decrease in a_0 to maintain the correct values of mean(u) and mean(v) with reduced σ_i, in accordance with the conserved quantities given in Eq. (30). Thus, while the presence of red-noise forcing results in biased parameter estimates, these biases can be understood in terms of the model dynamics.

c. Resimulation of winds using the reconstructed model

As a final analysis of the accuracy of the reconstructed models, we will compare the statistics of simulations they generate with those from the data to which they were fit. In particular, we will investigate how well the means, standard deviations, and skewness of the resimulated data match those of the original time series. As a demonstration of the accuracy of the reconstructed model parameters, the results of this analysis for data from the model driven by white noise are displayed in Fig. 7. Also displayed is the ACF of u for both the original data and resimulated data. Noting the small relative error in the computed statistics, we see that the reconstructed model is able to accurately reconstruct the statistics of time series produced by these dynamics.

We also considered the ability of the reconstructed model to capture the vector wind statistics when the time series are produced from the model with red-noise driving. In this case, while the parameter estimates are expected to be biased relative to their true values, the reconstructed model should be able to capture the moments of the time series (cf. section 2c). In fact, the first three moments of the PDF are recovered to a good accuracy (Fig. 8). However, when the parameter estimation is carried out without subsampling of the data, the estimated parameters give resimulated data with an autocovariance function that matches only up to the resolution time step of the data. This bias is consistent with the fact that the estimation routine is predicated on the assumption that the data is Markovian (DelSole 2000).

As was discussed in section 2, this bias can be addressed by subsampling the data to a sufficiently large degree that the memory of the driving process is suppressed. This can be accomplished in practice by first performing a preliminary analysis of the ACF of the
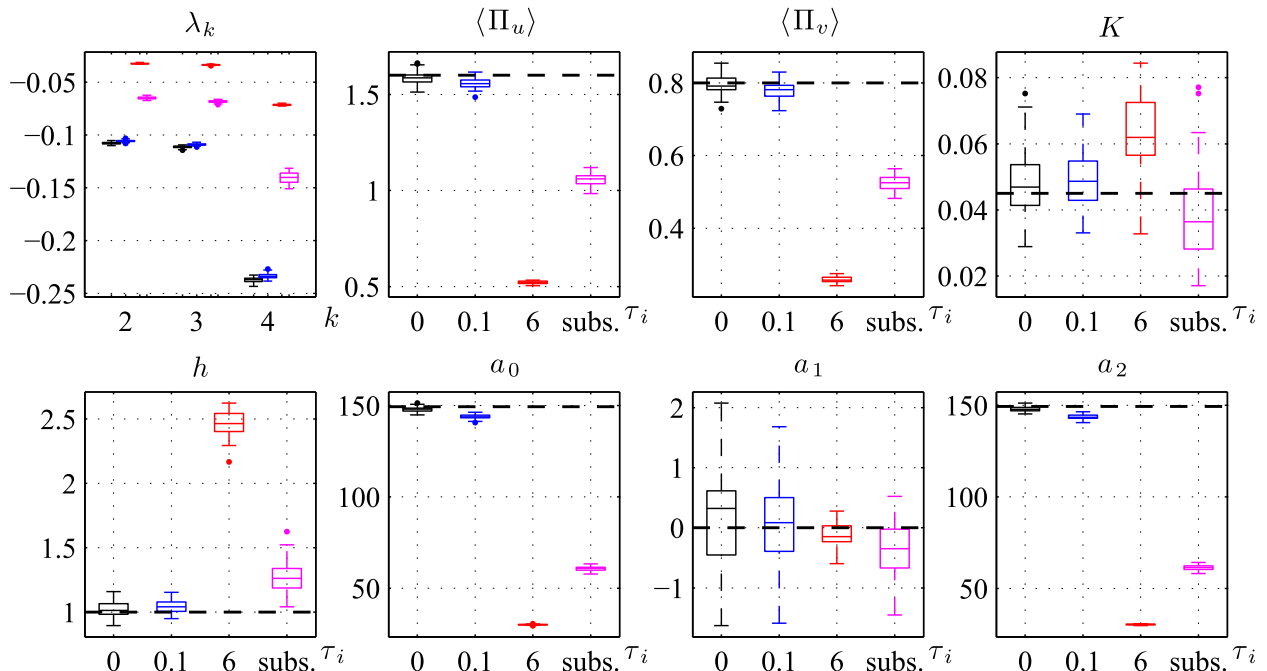
FIG. 6. (top left) Boxplots of estimates for the first three nonzero eigenvalues for the wind model with varying ACTs in the forcing. The black boxplots indicate the eigenvalue estimates for simulated time series with white-noise forcing, while the blue and red boxplots indicate the estimates when red noise having short ($\tau_i = 0.1$ h) and long ($\tau_i = 6$ h) ACTs is used. The pink boxplots indicate the parameter estimates from the time series with $\tau_i = 6$ h with subsampling of degree 4 applied. The data are resolved at $\delta t = 6$ h. The other panels display the estimates for the parameters of the SDEs. The true values for the white-noise case are indicated with a black dashed line. In each case, an ensemble of parameter estimates from 50 independent realizations was computed.

data to estimate the ACT of the data, and then subsampling the data such that the time step is on the same scale as the ACT of the data. In the present case, subsampling the data by taking every fourth point results in an evident improvement in the simulation of the autocorrelation function (Fig. 8), without significantly altering the estimates of the other statistics. As we have mentioned, this technique will only work when the ACT scale of the red-noise driving is sufficiently short compared to that of the dynamics of the observed variable such that the subsampling eliminates the effect of memory in the forcing without destroying the autocorrelation structure of the resolved dynamics. When fitting the model to observed winds, we will combine data subsampling with a parameter adjustment, such that the modeled time series autocorrelation approximates that of the observations as closely as possible.

4. Estimation of model parameters from reanalysis wind data

Having considered the application of the estimation method in a perfect model setting, we now consider the reconstruction of wind model parameters from a global sea surface wind dataset. For this analysis, we will use

the 6-hourly 10-m winds from the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) data, available on a 2.5° × 2.5° grid from 1 September 1957 to 31 August 2002 (downloaded from http://apps.ecmwf.int/datasets/). Reanalysis products provide a three-dimensional representation of the atmosphere on a regular grid by assimilating observations into a fixed forecast model. As such, reanalysis winds are not direct observations but instead represent a balance between observations and the predictions of a global, comprehensive model of atmospheric physics. These data were used for the reconstruction rather than direct, remotely sensed observations [such as from the SeaWinds scatterometer on the Quick Scatterometer (QuikSCAT) satellite] because of their relatively fine resolution in time and long duration. In fact, there is little difference between the statistical features of remotely sensed surface winds and those from a range of different reanalysis products (e.g., Monahan 2006b, 2012).

We will first present the results of the application of the estimation procedure to data from three representative locations. Following this, parameter estimates will be obtained across the global ocean between 60°S and 60°N (avoiding regions with sea ice for which the surface
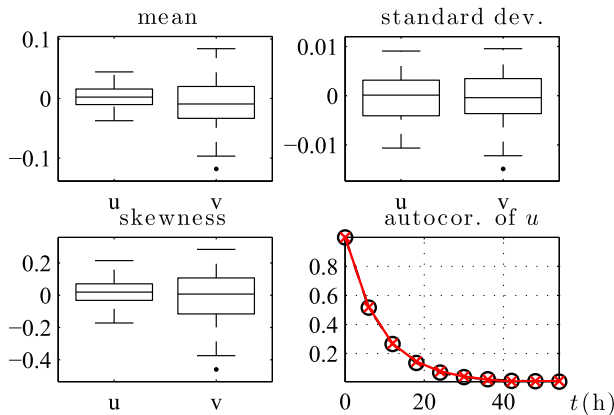
FIG. 7. (top),(bottom left) Relative error of simulated statistics relative to original statistics (mean, standard deviation, and skewness) for simulations from models fit to time series produced by the wind model with white-noise forcing. For each of these, the relative error of a quantity z is defined as $(z_{original} - z_{reconstructed})/z_{original}$. (bottom right) The computed ACF of u from the original time series (black, circles) and from the resimulated time series (red, crosses). The estimates of the parameters were obtained without subsampling and with weight $w = 1000$. In each time series, $\delta t = 6$ h and 30 000 data points are used.
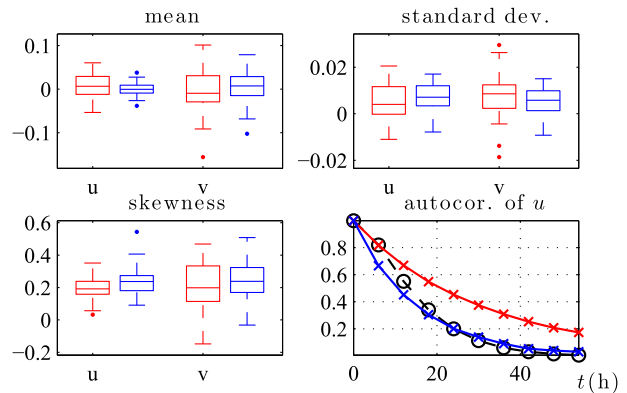

FIG. 8. As in Fig. 7, but for the wind model [Eqs. (22) and (23)] fit to time series generated with red-noise forcing with ACT scales similar to the resolution of the time series ($\tau_i = \delta t = 6$ h). Red symbols denote results obtained using parameter estimates without data subsampling, while the blue symbols denote the results following subsampling of degree 4. These calculations were carried out with an ensemble of 50 independent realizations.

wind model is not appropriate). To ensure that the estimated parameters are physically meaningful (despite potential mismatches between observations and the wind model), we impose constraints on the optimization. First, we require that the layer thickness h be bounded in between 1 m and 100 km. This range is clearly well outside of the physically meaningful range; the most important requirement here is that h be nonnegative. Also based on physical requirements, K is constrained to be nonnegative. Without these constraints, the estimation method sometimes estimates unphysical values of K and h that are negative. Negative values of h are particularly problematic, as these are inconsistent with stationary solutions of the model. That negative values of h can potentially occur without this constraint can be explained by the fact that the algorithm estimates the parameter c_d/h, which is often near zero, rather than h itself.

a. Limitations of the model

Natural processes are always more complicated than any model chosen to study them. As such, we expect that there are aspects of observed wind variability that will not be captured by the model and that may influence the parameter estimates. As discussed above, an important difference between the wind data and the model is that the real data are almost certainly non-Markovian in nature, while the model solutions are, by construction, Markov processes. While it would be more accurate to fit the data to a model in which the variations in the "large scale" forcing are modeled as red-noise processes, we do

not have these forcing time series from observations and as such cannot include them in the parameter estimation process. In addition to the challenges posed by the "red noise" nature of the data, there is a potential problem posed by a difference of ACT scales between the zonal and meridional components of the wind data (Monahan 2012). In many locations the meridional component experiences a much quicker rate of decorrelation than the zonal component. In the model, the single parameter h scales the ACT scale of both components. As the white-noise processes driving u and v have the same (infinitely short) memory, the model cannot account for this anisotropy in autocorrelation structure. The process of estimating model parameters from observations will have to accommodate this fact.

One of the predictions of the wind model is that the mean and skewness of the vector winds are spatially anticorrelated. In particular, the component of the wind in the direction of the time-mean wind is predicted to be negatively skewed (Monahan 2004). While this is broadly consistent with observations, in some locations the observed skewness of the along-mean wind component is weakly positive; in such locations, there will be a mismatch between the modeled and observed pdfs. Furthermore, while the relationship between the mean and skewness of the vector winds is captured qualitatively by the model, it underestimates the magnitude of the skewness (Monahan 2006a). Thus, it is not to be expected that the statistics of the reconstructed model will exactly match those of the observed winds.

Finally, for the sake of simplicity and to be able to make use of the largest amount of data in our reconstructions, in the present analysis we have neglected
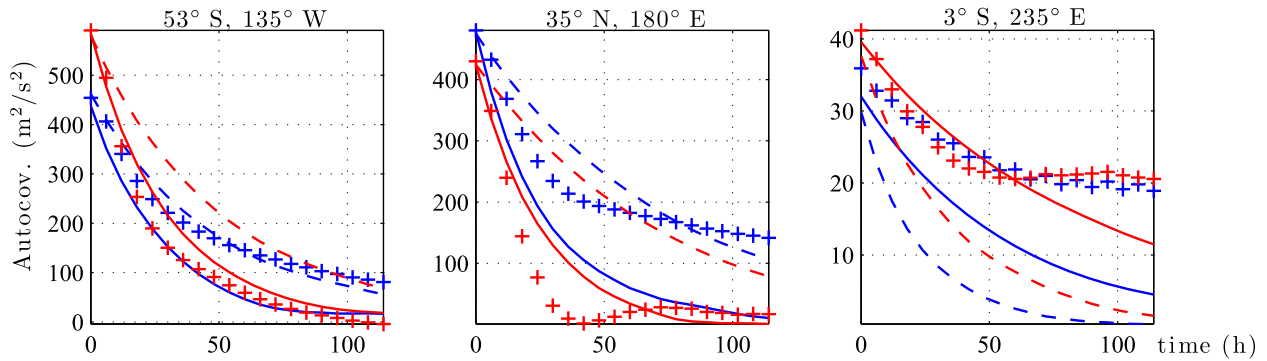
FIG. 9. The autocovariance functions for the zonal and meridional wind directions (blue and red, respectively). Crosses: observed estimates. Dashed lines: simulations based on parameter estimates without a rescaling of h. Solid lines: simulations using parameter estimates that include a rescaling of h. The rescaling is defined to match the absolute geometric-mean autocovariance at a lag of 18 h.

nonstationarities associated with the seasonal and diurnal cycles in the winds. What effect these non-stationarities may have on the reconstructed parameters is unclear, although seasonal and diurnal variability in the winds is generally considerably smaller than the internally generated "weather" variability over the open ocean (Dai and Deser 1999; Monahan 2006b).

b. Parameter estimates at representative points

We will now consider the estimation of parameters at three different locations selected to be representative of the statistics of large regions over the ocean. These three points are in the Pacific sector of the Southern Ocean (53°S, 135°W), in the midlatitude North Pacific near Japan (35°N, 180°), and in the equatorial Pacific (3°S, 125°W). These points are respectively representative of three broad oceanic provinces. The southernmost point is characterized by relatively large mean vector wind and skewness (see Table 3) with an autocorrelation function that decays on a time scale on the order of a day (Fig. 9). The northernmost point has relatively small mean vector wind and skewness with a strongly anisotropic autocorrelation function that also decays on a time scale on the order of a day. The equatorial point is characterized by large mean vector winds and skewness but a much more slowly decaying autocorrelation function.

The parameter estimates obtained from direct application of the estimation procedure to the reanalysis winds with modified weighting $w = 1000$ and a degree of subsampling of 4 are presented in Table 2. Observed and simulated statistics are given in Table 3. As discussed in section 4a, the model is unable to account for the auto-correlation anisotropy that is evident at the locations considered. This model bias results in a range of consequences; in some cases, the model ACF using the raw estimates is substantially different than either of the

vector wind ACFs (Fig. 9). To offset this effect, we rescale the estimated parameters (as described in section 2) such that the pdf remains unchanged and the ACT scale is changed to match the geometric mean of the ACT scales of u and v.

In the present case, we have used a lag of 18 h for the autocorrelation matching. Rescaled parameter estimates are presented in the second row of Table 2. This rescaling results in modeled ACFs that are closer approximations to those of observations, although significant differences persist (Fig. 9).

A comparison of the observed and modeled statistics at the three points under consideration demonstrates that certain moments of the data are captured better than others (Table 3). The mean wind speeds in the

TABLE 2. The top group of values shows estimates of the parameters in the wind model [Eqs. (22) and (23)] with weighting $w = 1000$ and a subsampling of degree 4. The bottom group of values shows parameter estimates following the rescaling of parameters to improve estimates of the autocorrelation structure as described in section 4b.

Location	53°S, 135°W	35°N, 180°	3°S, 125°W
$\langle \Pi_u \rangle$ (m s^{-2})	3.04×10^{-5}	8.56×10^{-6}	-4.35×10^{-5}
$\langle \Pi_v \rangle$ (m s^{-2})	-3.46×10^{-6}	3.14×10^{-6}	1.28×10^{-5}
K (m^2 s^{-1})	16.7	3.56×10^4	9.51×10^{-6}
h (m)	3.77×10^3	1×10^5	1.21×10^3
a_0 (m^2 s^{-3})	3.98×10^{-4}	2.75×10^{-4}	5.83×10^{-5}
a_1 (m^2 s^{-3})	2.96×10^{-5}	7.63×10^{-6}	-2.18×10^{-5}
a_2 (m^2 s^{-3})	4.97×10^{-4}	2.46×10^{-4}	4.74×10^{-5}
$\langle \Pi_u \rangle$ (m s^{-2})	5.13×10^{-5}	2.47×10^{-5}	-1.76×10^{-5}
$\langle \Pi_v \rangle$ (m s^{-2})	-5.84×10^{-6}	9.07×10^{-6}	5.19×10^{-6}
K (m^2 s^{-1})	9.92	1.23×10^4	2.35×10^{-5}
h (m)	2.23×10^3	3.46×10^4	2.98×10^3
a_0 (m^2 s^{-3})	6.73×10^{-4}	7.94×10^{-4}	2.36×10^{-5}
a_1 (m^2 s^{-3})	5.00×10^{-5}	2.20×10^{-5}	-8.82×10^{-6}
a_2 (m^2 s^{-3})	8.40×10^{-4}	7.10×10^{-4}	1.92×10^{-5}

TABLE 3. The top group of values shows the observed statistics of the ERA-40 data at indicated locations. The bottom group of values shows the computed statistics from the wind model [Eqs. (22) and (23)] with estimated parameters. Estimation of the parameters is carried out using the constraints described in section 4c, weighting $w = 1000$, and a subsampling of degree 4.

Location	53°S, 135°W	35°N, 180°	3°S, 125°W
Mean(u) (m s^{-1})	5.76	2.30	-5.93
Mean(v) (m s^{-1})	-0.61	0.85	1.64
Std dev(u) (m s^{-1})	5.92	6.09	1.66
Std dev(v) (m s^{-1})	6.75	5.76	1.78
Skew(u)	-0.68	0.14	1.78
Skew(v)	-1.4×10^{-2}	6.0×10^{-2}	-0.45
Mean(u) (m s^{-1})	5.30	2.04	-5.94
Mean(v) (m s^{-1})	-0.77	0.45	1.69
Std dev(u) (m s^{-1})	5.97	6.05	1.52
Std dev(v) (m s^{-1})	6.69	5.72	1.70
Skew(u)	-0.31	1.9×10^{-3}	0.31
Skew(v)	1.5×10^{-2}	-4.3×10^{-2}	-0.27

zonal and meridional directions are well captured, as are the standard deviations of those quantities. In contrast, while the sign and relative magnitude of the skewness values are captured by the model, the absolute magnitude is not. As we will see in section 4c, these results are consistent across the global ocean.

Even with rescaling, h takes values significantly greater than 1 km, which is physically unreasonable. As discussed above, this bias in h is consistent with the large-scale ageostrophic forcing having an ACT scale comparable to that of the resolved dynamics. The other model parameters are expected to be correspondingly biased (relative to their true values) so that the model results in reasonable simulations of the vector wind component pdfs.

c. Global parameter estimates

We now reconstruct global fields of the model parameters from the reanalysis surface wind data. The statistics of the simulated winds with estimated parameters are displayed in Fig. 10; the parameter fields are shown in Fig. 11. In both of these plots, results are shown with and without rescaling of h (to bring the observed and modeled autocorrelation structures into closer accord). Note that the restriction on h to keep the estimates bounded within 10^{-3} and 10^{2} km is only applied in the initial parameter estimates and is not applied in the parameter rescaling.

In general, the mean and standard deviation fields of the zonal and meridional winds are reproduced very well. The sign and relative magnitude of the skewness fields are also well reproduced; as noted above, the model is unable to accurately simulate the absolute magnitude of the vector wind skewness.

Considering the estimated parameter fields, we see that the parameter fields are generally less noisy (and more easily interpreted meteorologically) after the rescaling of parameters. The reconstructed $\langle \Pi_u \rangle$ field is strong in the region of midlatitude westerlies and the trade winds, while the reconstructed $\langle \Pi_v \rangle$ field is only strong on the eastward flanks of the subtropical highs. As these parameters determine the mean vector wind, the results are consistent with the mean vector wind climatology. The values of a_0 and a_2 are strongest in the storm tracks of the northwestern Pacific and Atlantic and the Atlantic–Indian Ocean sector of the Southern Ocean, which again is consistent with the interpretation of the stochastic forcing as representing variability in the large-scale driving processes. The a_0 and a_2 maps are also similar which is consistent with the observation that the vector wind standard deviations are generally close to isotropic (Monahan 2006a). That the cross terms a_1 are generally weak is consistent with the observation that the vector winds are, to a first approximation, uncorrelated. These results also provide an a posteriori justification of the assumption that the vector wind dynamics are reversible (section 3).

In contrast, the estimates of h and K are more problematic—particularly where the vector wind skewness is small (Fig. 12). In such regions, it would appear that the estimation routine is unable to distinguish between the linear and nonlinear drag terms in the equations of motion. Skewness in the vector winds results from the nonlinearity of the surface drag in this model. When the vector wind fluctuations are approximately symmetric around the mean, there is a degeneracy between the linear and nonlinear drag terms. Numerical simulations of Eqs. (22) and (23) demonstrate that the modeled wind component ACT is set by both h and K, such that the ACT is unchanged if h and K are increased together in the appropriate way (not shown). When the vector winds are unskewed, h can take arbitrarily large values without substantially changing the shape of $p_{uv}(u, v)$. In such a case, K is determined by the ACT: if h is unreasonably large, so too is K. To improve estimates of h, we will now consider a reinterpretation of the model in which K is set to zero.

d. Improved estimates of h

We consider an alternative interpretation of the wind model in which h is interpreted not as the depth of an arbitrary slab but as the height at which turbulent transport of momentum vanishes. In this interpretation, there is no downward mixing of momentum from above the layer so the parameter K is set to 0 and the only two deterministic forces that act on the wind speeds are the mean "ageostrophic force" and the surface drag.
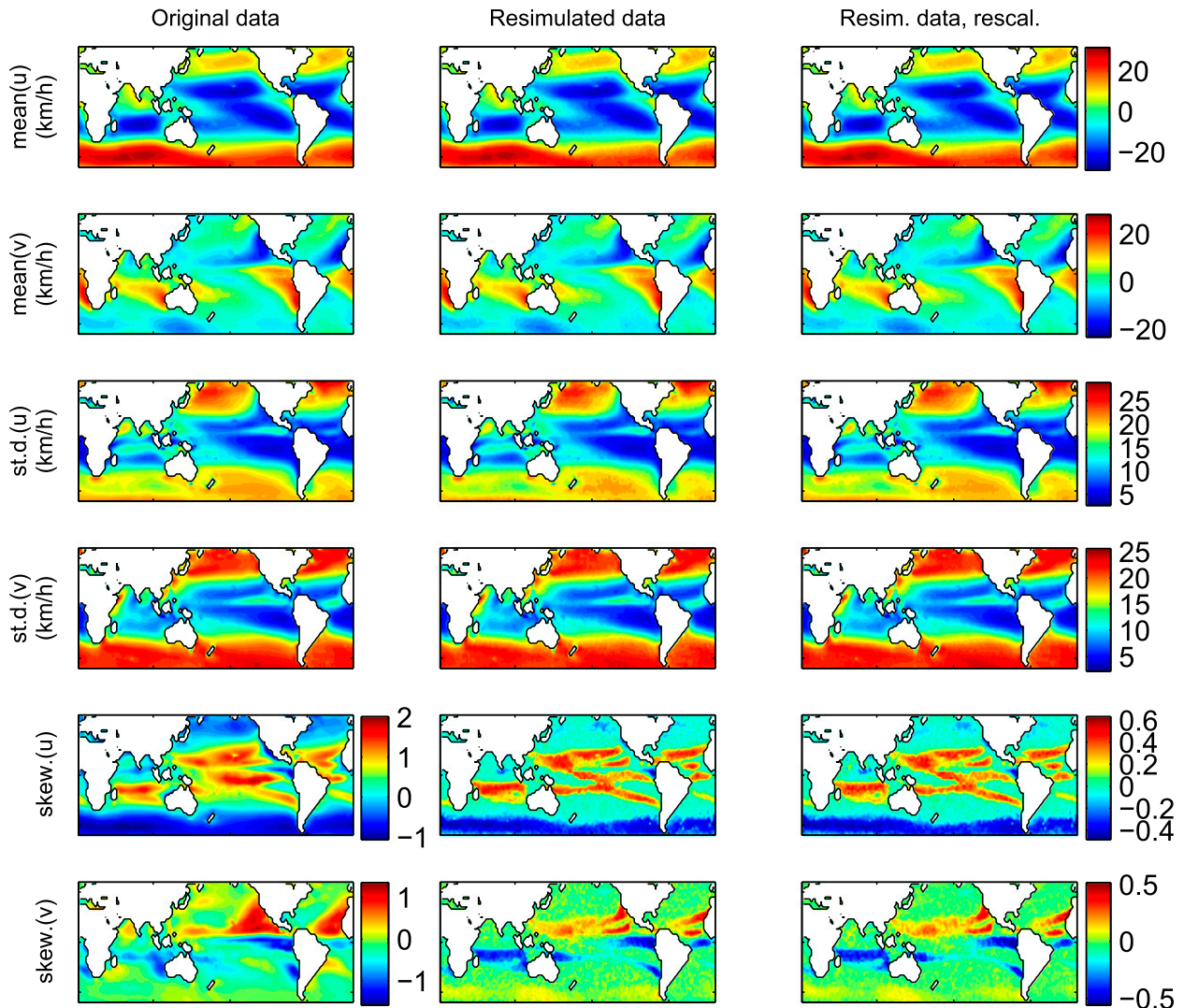
FIG. 10. Statistics of (left) the original data, (middle) the resimulated data with parameters estimated using the C-VE method on the original data, and (right) the resimulated data with parameter estimates that include a rescaling of the parameters so that the ACT scale is more accurately captured. [The parameters were estimated using the weighting scheme of Eq. (16) with $w = 1000$ and subsampling factor of 4.]

Reestimating parameters when we constrain K to be zero, the statistical fields for the mean and standard deviation of the vector wind components do not change (not shown), while the skewness fields are slightly affected (Fig. 13). The linear drag term does influence the shape of the vector wind pdf; in its absence, the flexibility of the model in this context is reduced.

The corresponding fields for the model parameters are not substantially different from the previous results, with the obvious exception of h (Fig. 14). The field of h is markedly smoother than those displayed in Fig. 11. While the estimated values of h are unrealistically large in order to account for the finite ACT scale of the large-scale driving, the values (ranging from a few hundred

meters to a few kilometers) have the correct order of magnitude. The greatest values of h occur in the Arabian Sea, where the winds are observed to have the longest lag ACTs (Monahan 2012). This particularly long ACT likely reflects the monsoonal reversals of the wind in this region, which the model under consideration cannot account for as constructed.

5. Summary and conclusions

The stochastic model of the near-surface atmospheric momentum budget presented in Monahan (2006a) was developed as a tool for the qualitative investigation of physical controls on the variability of sea surface winds.
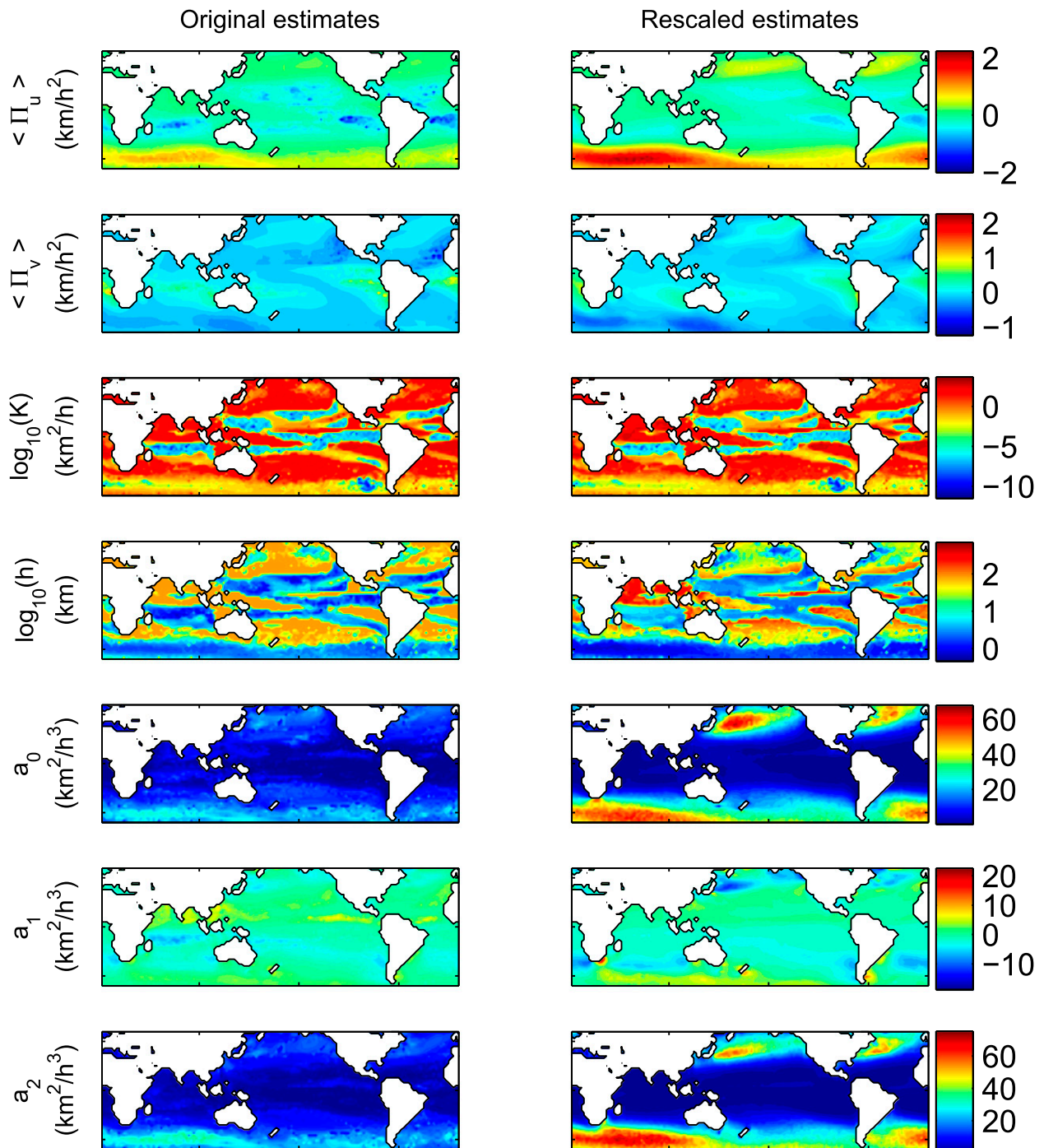
FIG. 11. (left) Estimates of the parameter fields. (right) Parameter field estimates after the rescaling of the parameters so that the overall ACT scale is more accurately captured. [The parameters were estimated using the weighting scheme of Eq. (16) with $w = 1000$ and subsampling factor of 4.] In the initial estimates for h, we have enforced bounds on $h \in [10^{-3}, 10^2]$ km.

An assessment of its utility as a quantitative tool requires observationally based estimates of model parameters. In this study, we have applied the procedure of Crommelin and Vanden-Eijnden (Crommelin 2012; Crommelin and Vanden-Eijnden 2006b, 2011) to estimate

the parameters of a stochastic differential equation describing sea surface wind variability using long time series of 10-m sea surface vector wind components. Although the data include aspects that cannot be accounted for by the model under consideration (diurnal
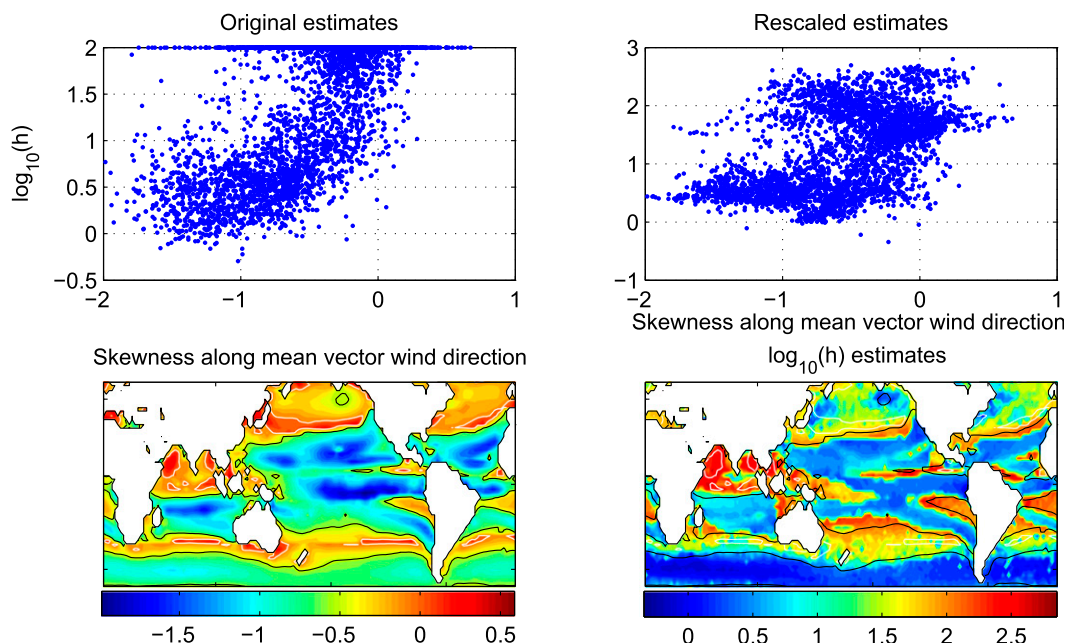
FIG. 12. (top) Scatterplot of the skewness of the wind speed along the mean wind direction against the logarithm of the estimated value of h. Recall that in the original parameter estimates, we apply constraints that include an upper bound on h. (bottom left) Skewness of the wind speeds along the mean wind direction. The white (black) contours indicate the level curves where the skewness is equal to 0 (equal to -0.5). (bottom right) The field of rescaled h estimates with the level curves superimposed.

and annual nonstationarities, anisotropic vector wind autocorrelation function, and positive skewness in the along-mean-wind direction), meaningful estimates of model parameters were obtained. In particular, we have demonstrated that, although the parameter estimates from data obtained from a system with autocorrelated forcing lead to biased autocorrelation functions, these biases can be understood in terms of the dynamics of the system.

An important result of the process of estimating parameters of the stochastic boundary layer momentum budget from sea surface wind observations is a better understanding of the limitations of this model. In particular, it is unable to account for the observed anisotropy in the vector wind autocorrelation structure and results in simulations with realistic ACTs only if unrealistic values of the layer thickness are used. These model limitations can be addressed to some extent by considering a more realistic representation of the large-scale driving processes—particularly coherent structures like extratropical cyclones and equatorial waves (Monahan 2012). Such an extension of the model will be considered in future studies.
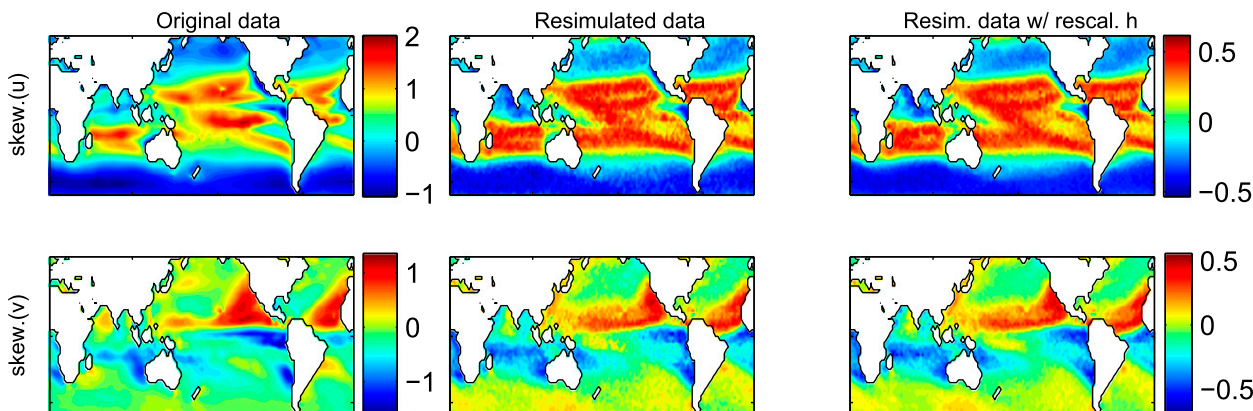

FIG. 13. Skewness fields for the measured and simulated data when K is set to zero.

Original estimates
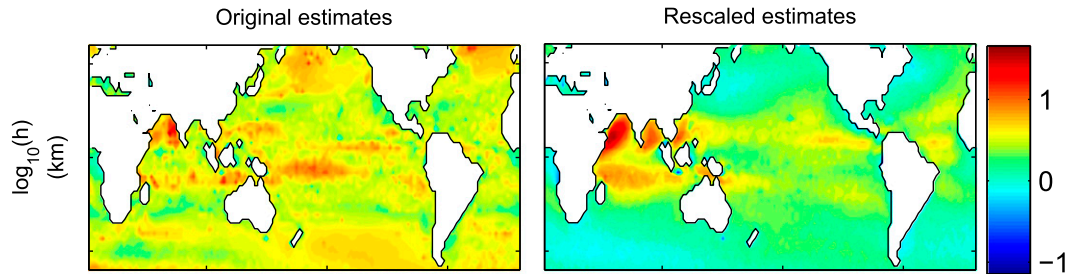
Rescaled estimates

FIG. 14. Estimated $\log_{10}(h)$ fields when K is set to zero.

In this analysis, we have addressed the issue of non-Markov structure in the time series by applying the estimator to a new time series made up of subsamples of the original process concatenated together. An alternative approach is to apply the estimator to each subsample and then average the resulting estimates. A preliminary investigation of this approach indicates that for the time series and model under consideration, it results in estimates of the leading eigenmodes, which are used in the variational analysis that are essentially the same as the estimates from the first approach (with a relative difference of less than 10^{-3}). The benefit of the second of these approaches is that it is more naturally applied to analyses in which the time series are broken down by season or time of day to account for annual or diurnal cycles. A more general comparison of these two approaches to handling non-Markov structure in the time series is an interesting direction of future study.

This analysis demonstrates that the Crommelin and Vanden-Eijnden estimation procedure is a powerful tool for the estimation of model parameters, particularly when the estimation process can be informed by an understanding of the model dynamics. An important outstanding challenge remains the problem of obtaining unbiased parameter estimates when the data are driven by noise that is autocorrelated in time. Consideration of this more general problem is another important direction of future study.

REFERENCES

Capps, S. B., and C. S. Zender, 2009: Global ocean wind power sensitivity to surface layer stability. *Geophys. Res. Lett.,* **36,** L09801, doi:10.1029/2008GL037063.

Crommelin, D., 2012: Estimation of space-dependent diffusions and potential landscapes from non-equilibrium data. *J. Stat. Phys.,* **149,** 220–233, doi:10.1007/s10955-012-0597-4.

——, and E. Vanden-Eijnden, 2006a: Fitting timeseries by continuous-time Markov chains: A quadratic programming approach. *J. Comput. Phys.,* **217,** 782–805, doi:10.1016/j.jcp.2006.01.045.

——, and ——, 2006b: Reconstruction of diffusions using spectral data from timeseries. *Commun. Math. Sci.,* **4,** 651–668, doi:10.4310/CMS.2006.v4.n3.a9.

——, and ——, 2011: Diffusion estimation from multiscale data by operator eigenpairs. *Multiscale Model. Simul.,* **9,** 1588–1623, doi:10.1137/100795917.

Dai, A., and C. Deser, 1999: Diurnal and semidiurnal variations in global surface wind and divergence fields. *J. Geophys. Res.,* **104,** 31 109–31 125, doi:10.1029/1999JD900927.

DelSole, T., 2000: A fundamental limitation of Markov models. *J. Atmos. Sci.,* **57,** 2158–2168, doi:10.1175/1520-0469(2000)057<2158:AFLOMM>2.0.CO;2.

Donelan, M., W. Drennan, E. Saltzman, and R. Wanninkhof, Eds., 2002: *Gas Transfer at Water Surfaces.* Amer. Geophys. Union, 383 pp.

Fairall, C. W., E. F. Bradley, J. E. Hare, A. A. Grachev, and J. B. Edson, 2003: Bulk parameterization of air–sea fluxes: Updates and verification for the COARE algorithm. *J. Climate,* **16,** 571–591, doi:10.1175/1520-0442(2003)016<0571:BPOASF>2.0.CO;2.

Gardiner, C. W., 1985: *Handbook of Stochastic Methods: For Physics, Chemistry and the Natural Sciences.* 2nd ed. Springer-Verlag, 442 pp.

Gobet, E., M. Hoffmann, and M. Reiß, 2004: Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Stat.,* **32,** 2223–2253, doi:10.1214/009053604000000797.

Jones, I. S. F., and Y. Toba, 2001: *Wind Stress over the Ocean.* Cambridge University Press, 307 pp.

Kloeden, P. E., and E. Platen, 1992: *Numerical Solution of Stochastic Differential Equations.* Springer-Verlag, 636 pp.

Liu, W. T., W. Tang, and X. Xie, 2008: Wind power distribution over the ocean. *J. Geophys. Res.,* **35,** L13808, doi:10.1029/2008GL034172.

Monahan, A. H., 2004: A simple model for the skewness of global sea surface winds. *J. Atmos. Sci.,* **61,** 2037–2049, doi:10.1175/1520-0469(2004)061<2037:ASMFTS>2.0.CO;2.

——, 2006a: The probability distribution of sea surface wind speeds. Part I: Theory and SeaWinds observations. *J. Climate,* **19,** 497–520, doi:10.1175/JCLI3640.1.

——, 2006b: The probability distribution of sea surface wind speeds. Part II: Dataset intercomparison and seasonal variability. *J. Climate,* **19,** 521–534, doi:10.1175/JCLI3641.1.

——, 2007: Empirical models of the probability distribution of sea surface wind speeds. *J. Climate,* **20,** 5798–5814, doi:10.1175/2007JCLI1609.1.

——, 2012: The temporal autocorrelation structure of sea surface winds. *J. Climate,* **25,** 6684–6700, doi:10.1175/JCLI-D-11-00698.1.

Øksendal, B., 2003: *Stochastic Differential Equations.* Springer-Verlag, 379 pp.

Pavliotis, G. A., and A. M. Stuart, 2007: *Multiscale Methods: Averaging and Homogenization.* Springer, 310 pp.

Risken, H., 1989: *The Fokker-Planck Equation: Methods of Solution and Applications.* Springer-Verlag, 472 pp.

Sampe, T., and S.-P. Xie, 2007: Mapping high sea winds from space: A global climatology. *Bull. Amer. Meteor. Soc.,* **88,** 1965–1978, doi:10.1175/BAMS-88-12-1965.

Sørensen, H., 2004: Parametric inference for diffusion processes observed at discrete points in time: A survey. *Int. Stat. Rev.,* **72,** 337–354, doi:10.1111/j.1751-5823.2004.tb00241.x.