Assessing Atmospheric Response to Surface Forcing in the Observations. Part I: Cross Validation of Annual Response Using GEFA, LIM, and FDT

ZHENGYU LIU

School of Atmospheric Sciences, Nanjing University of Information Science and Technology, Nanjing, and Laboratory for Climate and Ocean-Atmosphere Studies, School of Physics, Peking University, Beijing, China, and Center for Climatic Research, and Department of Atmospheric and Oceanic Sciences, University of Wisconsin—Madison, Madison, Wisconsin

NA WEN

School of Atmospheric Sciences, Nanjing University of Information Science and Technology, Nanjing, China

LEI FAN

Laboratory of Physical Oceanography, Ocean University of China, Qingdao, China

ABSTRACT

Three statistic methods [generalized equilibrium feedback analysis (GEFA), linear inverse modeling (LIM), and fluctuation–dissipation theorem (FDT)] are compared for their assessment of the atmospheric response to sea surface temperature variability in the coupled climate system with a sample length comparable with the observations (decades). The comparison is made first in an idealized coupled model and then in the observations. For daily to pentad data, for a linear stochastic system, the simple model study demonstrates that all three methods are able to provide a consistent assessment of the atmospheric response. For monthly data, GEFA is able to produce an assessment comparable with the daily or pentad assessments using the three methods. The consistence of the three methods is further confirmed in the observations for the responses of the atmospheric geopotential height (at 200 hPa) to the tropical ENSO mode and the North Pacific mode. It is found that the three methods produce a consistent response with the overall pattern correlation over 0.95 and the amplitude difference within 10%–20%. The consistent results in both the simple model and the observations suggest that the three statistical methods can be used as a cross validation on the robustness of the assessment of the atmospheric response to surface forcing in the observations.

1. Introduction

One critical issue in climate dynamics is to understand the response of the atmosphere to slowly varying lower boundary forcing, such as the sea surface temperature (SST). From the coupled ocean–atmosphere perspective, the rapid atmosphere variability interacts with the slow oceanic variability constantly: the atmosphere drives the ocean and the ocean, in turn, feeds back on the atmosphere, forming complex coupled feedback loops. Therefore, in the study of ocean–atmosphere interaction, the atmospheric response to SST variability is also sometimes called the feedback response. This feedback response is important because it represents the part of climate variability that is potentially predictable beyond the short atmospheric memory (days to weeks). In the observations, however, this feedback response is usually difficult to assess, because the observed time series of ocean and atmosphere variability are the final product of the coupled feedbacks, which makes it difficult to separate the atmospheric response to SST from its forcing on SST. Ideally, in a model world, the atmospheric feedback response can be assessed using ensemble experiments in which the internal atmospheric variability can be suppressed by the ensemble mean. This ensemble method, however, is not feasible for real-world observation, which only has a single realization. Instead, to assess the atmospheric response to SST variability in a

Corresponding author address: Zhengyu Liu, 1225 W. Dayton St., Madison, WI 53706.
E-mail: zliu3@wisc.edu

single realization of observation, statistical methods are needed.

Traditionally, the simplest method to assess the atmospheric response to a specific set of SST forcing, such as regional SST indices or SST EOFs, is a (multiple) regression of the atmospheric field onto the SST time series. The regression is based on the assumption of a quasi-equilibrium linear[1] atmospheric response, and therefore the atmosphere and SST data are often averaged over a time scale longer than the time scale of the rapid atmospheric response and internal variability (e.g., a month or season). In line with the quasi-equilibrium response, often a simultaneous regression is used. However, a simultaneous regression tends to mix the atmospheric response (to SST) with the atmospheric forcing (on SST) (Frankignoul 1999). Instead, the atmospheric response to the simultaneous SST forcing should be identified with the SST leading the atmosphere, as first proposed by Frankignoul et al. (1998). This SST-led regression is the essence of the so-called generalized equilibrium feedback analysis (GEFA) method (Frankignoul et al. 1998; Liu et al. 2008, hereafter LWL08; see section 2 for more discussion). Another type of method to study the atmospheric response to multiple SST forcings (e.g., a SST field) has been based on the singular value decomposition (SVD; Bretherton et al. 1992) of the covariance matrix between the atmosphere and ocean. This method is also based on the quasi-equilibrium atmospheric response. Again, to isolate the atmospheric response from internal variability, the atmospheric response should be identified with the SST leading the atmosphere [also called maximum covariance analysis (MCA); Czaja and Frankignoul 2002]. The SVD gives pairs of atmosphere–ocean patterns that are most coherent. However, for each pair, the atmospheric pattern is usually not the true response pattern to the corresponding SST pattern. Furthermore, the SVD is not convenient for assessing the atmospheric response to a given set of SST forcing, because the patterns of the SST forcing are the result of SVD and therefore cannot be specified a priori. In other words, the SVD provides a good qualitative analysis of the atmospheric response associated with a SST forcing field but does not give a quantitative assessment of the atmospheric response to a specified set of SST forcings. Here, we will discuss more comprehensive statistic methods that can assess the atmospheric response to a given set of SST forcings systematically.

For statistical assessment, it is highly desirable to develop independent statistical methods for cross validation. There are, at least, three comprehensive statistic methods available for a systematic assessment of the atmospheric response to a given set of SST forcing: GEFA (LWL08; Liu and Wen 2008), linear inverse modeling (LIM; Penland and Sardeshmukh 1995a,b; Newman et al. 2009) and fluctuation–dissipation theorem (FDT; Leith 1975; Bell 1980) (see more discussions on the methods in section 2). In a purely linear system and with sufficiently long sample, in principle, all the three methods should give the same results. In the real world, however, the observed climate anomalies, at various time scales, may not follow linear dynamics. Moreover, the limited climate data (~50 yr) may induce substantial sampling error. It therefore remains unclear if these linear statistical methods can be used to assess the observed atmospheric response consistently.

The purpose of this paper is to show the consistence of GEFA, LIM, and FDT for the assessment of the observed atmospheric response to multiple SST forcing with a sample size comparable with the present observations (decades). We will compare the assessment of the three methods systematically, first in an idealized coupled model and then in the observations. It is found that, with the sample size of several decades, the three methods are able to produce consistent atmospheric response to SST forcing and therefore can be used for cross validation, in the observations. It is interesting and important to note that the robust estimation in LIM and FDT only applies to the atmospheric response to the slow SST forcing (i.e., the so-called feedback matrix), not to the temporal evolution of the coupled system (i.e., the full system matrix). This implies that the three estimates of the feedback response are robust and consistent for the special case of multiple time-scale systems, such as the atmosphere–ocean system here. The paper is arranged as follows: In section 2, we briefly review the three methods. In section 3, the three methods are compared in an idealized coupled model, first in a one-point model and then in a multipoint model. Section 4 compares the three methods in the observations. A summary and some discussions are presented in section 5.

2. GEFA, LIM, and FDT

For pedagogical purposes, we will review GEFA, LIM, and FDT briefly. The review here is to give those readers who are unfamiliar with these methods an introduction of some basic properties and relations of these methods. As a start, we assume climate anomalies in the coupled ocean–atmosphere system, after proper

---

[1] Unless otherwise specified, this paper will be confined to the discussion of linear atmospheric responses.

spatiotemporal average, can be approximated as a stochastically forced linear coupled system,[2]

$$\frac{d\mathbf{x}}{dt} = \mathbf{A_{xx}}\mathbf{x} + \mathbf{A_{xy}}\mathbf{y} + \mathbf{n}_x \quad \text{and} \qquad (1.1a)$$

$$\frac{d\mathbf{y}}{dt} = \mathbf{A_{yx}}\mathbf{x} + \mathbf{A_{yy}}\mathbf{y} + \mathbf{n}_y, \qquad (1.1b)$$

where $\mathbf{x}$ and $\mathbf{y}$ are column vectors representing the atmospheric and oceanic fields, respectively. The $\mathbf{n}_x$ and $\mathbf{n}_y$ are the stochastic noise forcing to the atmosphere and ocean, respectively, which include the anomalies associated with subscale processes and chaotic nonlinear dynamics.

To further focus on the low frequency dynamics of slow oceanic time scale (e.g., months), we will average the atmospheric equation over the slow time scale. Because the slow time scale is much longer than the rapid atmospheric response time, the atmospheric Eq. (1.1a) can be approximated as a quasi-equilibrium response (see appendix A),

$$\mathbf{0} = \mathbf{A_{xx}}\overline{\mathbf{x}} + \mathbf{A_{xy}}\overline{\mathbf{y}} + \overline{\mathbf{n}}_x, \qquad (1.2)$$

where an overbar denotes the time average. Equation (1.2) can be rewritten as

$$\overline{\mathbf{x}}(t) = \mathbf{B}\overline{\mathbf{y}}(t) + \overline{\mathbf{n}}(t), \qquad (1.3)$$

where

$$\mathbf{B} = -\mathbf{A_{xx}}^{-1}\mathbf{A_{xy}} \qquad (1.4)$$

will be called the feedback matrix, which gives the response kernel of a mean atmospheric field $\overline{\mathbf{x}}$ to a SST forcing, with the response to the SST forcing $\overline{\mathbf{y}}$ as $\mathbf{B}\overline{\mathbf{y}}$. It is important to note that, no matter how strong the stochastic atmospheric internal variability $\overline{\mathbf{n}}$ is, it is uncorrelated with the SST of earlier times (earlier than the

---

[2] For the real world, the validity of the linearization depends on the time scale. At very short time scale (e.g., subdaily to daily), atmospheric process can be strongly nonlinear. The linear dynamics becomes more valid for a properly long time average, because the averaged high-frequency chaotic nonlinear dynamics tends to become stochastic noise according to the central limit theorem (Gardiner 1997). Thus, for averages of different time scales, the representation of each term could be different, including the noise term. In our idealized model study of this section and in the next section, however, we will ignore the issues related to the nonlinearity. Instead, we will assume the coupled climate process is determined by a purely linear system (1.1) such that we can focus on the comparison of sampling errors in different assessment methods.

persistence time of the atmospheric variability τ_n). Therefore, we always have

$$\mathbf{C_{ny}}(\tau) = \langle \overline{\mathbf{n}}(t), \overline{\mathbf{y}}(t - \tau) \rangle$$
$$\equiv \overline{\mathbf{n}}(t)\overline{\mathbf{y}}^T(t - \tau) = 0 \quad \text{for} \quad \tau > \tau_n, \quad (1.5)$$

where $\langle\ \rangle$ defines the covariance between the two fields and the superscript "T" represents the matrix transpose. The feedback matrix $\mathbf{B}$ can be obtained in GEFA (Frankignoul et al. 1998; LWL08) with a covariance of Eq. (1.3) with earlier SST $\overline{\mathbf{y}}(t - \tau)$ as

$$\mathbf{B} = \langle \overline{\mathbf{x}}(t), \overline{\mathbf{y}}(t - \tau) \rangle \langle \overline{\mathbf{y}}(t), \overline{\mathbf{y}}(t - \tau) \rangle^{-1}$$
$$= \mathbf{C_{xy}}(\tau)\mathbf{C_{yy}}(\tau)^{-1} \quad \tau > \tau_n. \qquad (1.6)$$

The estimated feedback matrix $\mathbf{B}$ (1.6) is "proportional" but not identical to the SST-led covariance matrix $\mathbf{C_{xy}}(\tau)$. If applied at zero lag, it degenerates to the simultaneous regression matrix $\mathbf{C_{xy}}(0)\mathbf{C_{yy}}(0)^{-1}$ to (1.3) and GEFA becomes simply the simultaneous multiple regression. However, this regression matrix is the true feedback matrix only if the internal atmospheric variability $\overline{\mathbf{n}}(t)$ is not correlated to the SST in (1.5); that is $\mathbf{C_{ny}}(0) = \langle \overline{\mathbf{n}}(t), \overline{\mathbf{y}}(t) \rangle = 0$. This is possible only for two scenarios. First, the internal atmospheric variability is weak (i.e., $\overline{\mathbf{n}}(t) \approx 0$) such as the case in the tropical Pacific (TP) ocean–atmosphere system, where large-scale atmospheric variability $\overline{\mathbf{x}}(t)$ is forced mainly by the SST variability $\overline{\mathbf{y}}(t)$, with little contribution from internal atmospheric variability (e.g., Kumar and Hoerling 1998). Second, the internal variability is not weak, $\overline{\mathbf{n}}(t) \neq 0$, but the region of atmospheric response considered [and, in turn, the internal variability $\overline{\mathbf{n}}(t)$] is not collocated with the SST $\overline{\mathbf{y}}(t)$ or more precisely that the internal variability does not force the SST locally [i.e., $\langle \overline{\mathbf{n}}(t),\ \overline{\mathbf{y}}(t) \rangle = 0$]. Except for these two scenarios, in general, a sufficiently long lag $\tau > \tau_n$ is needed to ensure that the internal variability is not correlated with earlier SST variability as in Eq. (1.5).

The relation between the GEFA feedback matrix (1.6) and regression matrices can also be seen explicitly from the decomposition of (1.6) into two parts as (Liu et al. 2006)

$$\mathbf{B} = \mathbf{C_{xy}}(\tau)\mathbf{C_{yy}}(\tau)^{-1}$$
$$\equiv [\mathbf{C_{xy}}(\tau)\mathbf{C_{yy}}(0)^{-1}][\mathbf{C_{yy}}(0)\mathbf{C_{yy}}(\tau)^{-1}] \quad \text{or}$$

$$[\mathbf{C_{xy}}(\tau)\mathbf{C_{yy}}(\tau)^{-1}][\mathbf{C_{yy}}(\tau)\mathbf{C_{yy}}(0)^{-1}] \equiv [\mathbf{C_{xy}}(\tau)\mathbf{C_{yy}}(0)^{-1}].$$

The latter can be written as

$$\mathbf{B}\mathbf{B}_2(\tau) = \mathbf{B}_1(\tau).$$

Here, $\mathbf{B}_1(\tau) = \mathbf{C}_{xy}(\tau)\mathbf{C}_{yy}(0)^{-1}$ is the regression matrix from the prediction equation $\bar{\mathbf{x}}(t) = \mathbf{B}_1(\tau)\bar{\mathbf{y}}(t-\tau) + \mathbf{e}_1$, and $\mathbf{B}_2(\tau) = \mathbf{C}_{yy}(\tau)\mathbf{C}_{yy}(0)^{-1}$ is the regression matrix of the autoregression prediction equation of SST $\bar{\mathbf{y}}(t) = \mathbf{B}_2(\tau)\bar{\mathbf{y}}(t-\tau) + \mathbf{e}_2$. The equation $\mathbf{B}\mathbf{B}_2(\tau) = \mathbf{B}_1(\tau)$ can therefore be interpreted as follows: the prediction of the atmospheric response $\bar{\mathbf{x}}(t)$ from an earlier SST $\bar{\mathbf{y}}(t-\tau)$ [via $\mathbf{B}_1(\tau)$] equals the rapid (instantaneous) atmospheric feedback response $\bar{\mathbf{x}}(t-\tau)$ to the SST $\bar{\mathbf{y}}(t-\tau)$ of the earlier time $t - \tau$ (via $\mathbf{B}$) following its decaying with the autoregression evolution of the SST itself from $\bar{\mathbf{y}}(t-\tau)$ to $\bar{\mathbf{y}}(t)$ [via $\mathbf{B}_2(\tau)$]. The feedback response therefore represents the part of atmospheric response that is potentially predictable with the SST persistence. This decomposition also suggests that the GEFA feedback matrix $\mathbf{B}$ is proportional to the SST-led regression matrix and therefore is qualitatively similar to the SST-led regression matrix. In other words, the SST-led regression can also be used to describe the atmospheric response qualitatively, as long as the lag is not too large such that the SSTs have not evolved into very different structures.

GEFA or the related regression approach take advantage of the different time scales between the atmosphere and ocean to approximate the atmospheric response to SST as a quasi-equilibrium response. In contrast, LIM and FDT treat the transient atmospheric response explicitly and intend to extract the atmospheric response to SST from the transient evolution of the coupled ocean–atmosphere system. Now, the atmosphere and ocean are treated together in a single coupled system as

$$\frac{d\mathbf{z}}{dt} = \mathbf{A}\mathbf{z} + \tilde{\mathbf{n}}, \qquad (1.7)$$

where

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \quad \text{and} \quad \tilde{\mathbf{n}} = \begin{bmatrix} \mathbf{n}_x \\ \mathbf{n}_y \end{bmatrix}.$$

The coupled system matrix is

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{xx} & \mathbf{A}_{xy} \\ \mathbf{A}_{yx} & \mathbf{A}_{yy} \end{bmatrix}, \qquad (1.8)$$

and it determines the full temporal evolution of the coupled system through the evolution Green's function (or propagator) $\mathbf{G}(\tau) = \exp(\mathbf{A}\tau)$. If the stochastic forcing is Gaussian, the optimal solution of the linear coupled system starting from $\mathbf{z}(t)$ is $\mathbf{z}(t+\tau) = \mathbf{G}(\tau)\mathbf{z}(t)$, with the corresponding optimal system matrix $\mathbf{A}$ estimated in LIM as (Penland 1989)

$$\mathbf{A} = \frac{1}{\tau}\ln[\mathbf{C}_{zz}(\tau)\mathbf{C}_{zz}(0)^{-1}], \qquad (1.9)$$

where $\mathbf{C}_{zz}(\tau) \equiv \langle \mathbf{z}(t), \mathbf{z}(t-\tau) \rangle$ and τ is an arbitrary time lag. Taking the exponential on both sides of (1.9) and integrating over all the lags, we have

$$-\mathbf{A}^{-1} = \int_0^\tau \mathbf{C}_{zz}(\tau')\mathbf{C}_{zz}(0)^{-1}\,d\tau', \quad \tau \to \infty. \qquad (1.10)$$

In the context of linear system here, this is equivalent to the FDT estimator (Leith 1975; Bell 1980; Kirk-Davidoff 2009). Most LIM and FDT studies have been applied to study the temporal evolution of the full system (Cionni et al. 2004), such as ENSO variability (Penland and Magorian 1993; Penland and Sardeshmukh 1995b). However, LIM and FDT can also be used to assess ocean–atmosphere feedback indirectly by using the submatrices $\mathbf{A}_{xy}$ and $\mathbf{A}_{xx}$ as well, as shown in Eq. (1.4) (Newman et al. 2009; Alexander et al. 2008). However, as noted in previous studies (e.g., Gritsun and Branstator 2007; Kirk-Davidoff 2009), an accurate estimation of the system matrix $\mathbf{A}$ in LIM (Penland and Sardeshmukh 1995b) or FDT (e.g., Leith 1975) usually requires the data to resolve all the time scales (modeled in the deterministic part) of a very long sample size; in particular, it requires sufficiently high-frequency sampling of data to resolve the fast atmospheric times scales in the coupled system. Therefore, for our real-world climate data of several decades, it remains unclear if LIM and FDT can be used to assess ocean–atmosphere feedback with sufficient accuracy.[3]

Finally, it is important to point out that our comparison of GEFA, LIM, and FDT estimations of the feedback response is for the special case of multiple time scales: the interaction between a fast atmosphere and a slow ocean. For a general linear stochastic system (1.1), it is clear that the feedback matrix in GEFA estimator (1.6) does not equal to that derived using the submatrices (1.4) from the LIM estimator (1.10) or FDT estimator (1.11). However, for a fast atmosphere and a slow ocean, we will show that the three methods give very consistent estimate in both idealized models (section 3) and in the observations (section 4).

---

[3] The terms "forcing" and "response" here often refer to the slow SST forcing and the rapid atmospheric feedback response to SST, respectively, which are in the convention of ocean–atmosphere interaction studies. They are different from the convention in linear stochastic dynamics such as LIM and FDT, where "forcing" and "response" usually refer to the stochastic noise forcing and the stochastic variables associated with the deterministic terms, respectively.

3. Comparison of GEFA, LIM, and FDT: Idealized model studies

a. An idealized coupled model

We first compare GEFA, LIM, and FDT in an idealized thermally coupled ocean–atmosphere model. The model is an extension of our nonlocal stochastic climate model in LWL08 with the inclusion of the local temporal variability in the atmosphere. It can also be treated as the stochastic climate model of Barsugli and Battisti (1998) with the addition of atmospheric advection. The nondimensional equations are

$$\frac{\partial T_a}{\partial t} + U \frac{\partial T_a}{\partial x} = -aT_a + bT_o + N \quad \text{and} \quad (2.1a)$$

$$M \frac{\partial T_o}{\partial t} = cT_a - dT_o. \quad (2.1b)$$

Here, T_a is the atmospheric temperature, T_o is SST, and N is the stochastic forcing associated with atmospheric internal variability. The coefficients a and d represent damping of the atmosphere and ocean, respectively, and b and c are the coupling coefficients between the atmosphere and ocean. The coefficient $M \gg 1$ represents the large oceanic heat capacity relative to the atmosphere. The atmospheric advection U provides the nonlocal teleconnection in the coupled system, imitating

the dominant atmospheric teleconnection for climate variability at monthly to seasonal time scales (Liu and Alexander 2007).

Dividing the model domain into I intervals of the width $\Delta x = 1/I$ and then using the "upwind" differencing scheme on the advection term in (2.1a), we have the discretized atmospheric equation as

$$\frac{dT_{a,i}}{\partial t} = \left(\frac{U}{\Delta x} - a \right) T_{a,i} - \frac{U}{\Delta x} T_{a,i-1} + bT_{o,i} + N_i(t), \quad (2.2)$$

where i indicates the ith point. The advection term shows that, in addition to the local SST forcing, atmospheric variability at the ith point is also influenced by the advection from the upstream atmosphere at the $(i-1)$th point $[-(U/\Delta x)T_{a,i-1}]$. The discretized coupled equations in (2.2) can be put in the vector form as

$$\frac{d\mathbf{T}_a}{dt} = \mathbf{A_{xx}}\mathbf{T}_a + \mathbf{A_{xy}}\mathbf{T}_o + \mathbf{N} \quad \text{and} \quad (2.3a)$$

$$\frac{d\mathbf{T}_o}{dt} = \mathbf{A_{yx}}\mathbf{T}_a + \mathbf{A_{yy}}\mathbf{T}_o, \quad (2.3b)$$

where

$$\mathbf{A_{xx}} = \begin{bmatrix} U/\Delta x - a & 0 & \ldots & 0 & 0 \\ -U/\Delta x & U/\Delta x - a & 0 & \ldots & 0 \\ 0 & \ldots & \ldots & \ldots & 0 \\ 0 & \ldots & -U/\Delta x & U/\Delta x - a & 0 \\ 0 & \ldots & 0 & -U/\Delta x & U/\Delta x - a \end{bmatrix},$$

$$\mathbf{A_{xy}} = \begin{bmatrix} b & 0 & \ldots & 0 & 0 \\ 0 & b & 0 & \ldots & 0 \\ 0 & \ldots & \ldots & \ldots & 0 \\ 0 & \ldots & 0 & b & 0 \\ 0 & \ldots & 0 & 0 & b \end{bmatrix}, \quad \mathbf{A_{yx}} = \frac{1}{M} \begin{bmatrix} c & 0 & \ldots & 0 & 0 \\ 0 & c & 0 & \ldots & 0 \\ 0 & \ldots & \ldots & \ldots & 0 \\ 0 & \ldots & 0 & c & 0 \\ 0 & \ldots & 0 & 0 & c \end{bmatrix}, \quad \text{and}$$

$$\mathbf{A_{yy}} = \frac{1}{M} \begin{bmatrix} -d & 0 & \ldots & 0 & 0 \\ 0 & -d & 0 & \ldots & 0 \\ 0 & \ldots & \ldots & \ldots & 0 \\ 0 & \ldots & 0 & -d & 0 \\ 0 & \ldots & 0 & 0 & -d \end{bmatrix}. \quad (2.4)$$

The stochastic dynamic system (2.3a) and (2.3b) is integrated with the Euler forward scheme with the random noise generated at each time step using independent Gaussian random number generator. In the rest of this

section, the three methods will be compared first in a one-point model and then in a multipoint model. Because GEFA has been studied extensively in LWL08, we will focus on LIM and FDT and its comparison with GEFA.

b. One-point model study

In the one-point model, we simply set the atmospheric advection to zero ($U = 0$) in (2.1a) and (2.1b) (LWL08). The model now consists of a single variable x for the atmosphere and y for the ocean. Our results are not sensitive to the model parameters as long as the time scale of the ocean is much longer than the atmosphere ($M \gg 1$). Here we only show the results with the model parameters the same as in Barsugli and Battisti (1998) with $a = 1.12$, $b = 0.5$, $c = 1.0$, $d = 1.08$, and $M = 20$. The integration time step is 0.25 (a unit time $t = 1$ here corresponds to a nominal ~4 days). The response time of the atmosphere and SST are ~4 days ($\sim 1/a$) and ~2–3 months ($\sim M/d$), respectively, largely consistent with the observations in the midlatitude. With our application to the observations in mind, the sample size of the model data is taken comparable with the reanalysis of ~50 yr, which is ~20 000 daily data, ~5000 binned 4-day data, and ~650 binned monthly data.

We will first discuss the estimation of the feedback coefficient

$$-A_{xx}^{-1} A_{xy} = b/a$$

in LIM and FDT. Figure 1c shows the daily LIM estimations (i.e., using daily data) at different lag τ in a 100-member ensemble. The ensemble mean estimation (blue solid line) virtually recovers the truth (red solid) at small lags (<30 days), confirming the unbiased estimation of LIM on the feedback coefficient with a sufficiently large sample size (100 member × 50 yr per member = 5000 yr). Furthermore, the ensemble spread (gray bars; ~0.05) is ~10% of the feedback coefficient itself (~0.43). This suggests that, for our realistic sample size, daily LIM can provide a robust estimation of the feedback parameter. As the lag increases, the ensemble mean estimation gradually deviates from the truth and the ensemble spread gradually increases, due to increased sampling error.

The daily FDT estimation also gives good estimation of the feedback parameter, now at large lag of integration τ, consistent with the derivation of Eq. (1.11). Figure 1e shows that the ensemble mean estimation virtually recovers the true feedback at lag $\tau > 30$ days with an ensemble spread of ~0.06, comparable with the best LIM estimation (at small lags) in Fig. 1c. [For too large lag τ, the accuracy of both LIM and FDT estimation decrease because of the lost sampling in the estimation of the lagged covariance (not shown).]

The example above shows that both daily LIM and FDT estimates are able to assess the feedback parameter accurately for a realistic sample size. In contrast, the estimation using monthly LIM (Fig. 1d) and FDT (Fig. 1f) (i.e., using monthly data) is rather poor, because the ensemble mean is significantly biased from the truth even with this large sample size. The failure of monthly LIM and FDT is conceivably because both LIM and FDT, in principle, need to resolve the time scale of atmospheric variability in the atmospheric Eq. (2.1) or (2.3), which, in this case, is a few days (e.g., Penland and Sardeshmukh 1995a). It should be noted here the study here only focus on the statistical issues related to sampling errors. For real-world application, there is still the dynamic issue that the daily average may not be sufficiently long for the validity of the linearization representation of the dynamics of the coupled system (see footnote 2).

It is important to point out that, in daily LIM and FDT, the estimation of the feedback parameter (matrix) **B** is more robust and accurate than that for the system matrix **A**. Figure 2 shows the variation of the A_{xx} and A_{xy} estimated in LIM and FDT at different lags, in the same case as the feedback parameter estimation in Figs. 1c–f. For daily LIM and FDT, Figs. 2a,c show that the ensemble mean estimation of both A_{xx} and A_{xy} deviate from the truth substantially. This is consistent with previous studies (e.g., Gritsun and Branstator 2007) that an accurate estimation of the full system matrix **A** involving the atmospheric response requires a very large sample size. However, it is rather surprising that the feedback parameter can be estimated almost perfectly and the estimation exhibits much less sensitivity to lags (Figs. 1c,e). For example, the daily LIM estimation (at small lags < 10 days) shows the magnitude of both A_{xx} and A_{xy} are greater than the truth by ~20% (Figs. 2a,c), whereas the feedback parameter $-A_{xx}^{-1} A_{xy}$ (Fig. 1c) is estimated almost perfectly. It is indeed interesting (and important) that the feedback parameter $-A_{xx}^{-1} A_{xy}$ can be estimated more accurately than A_{xx} and A_{xy} individually. Similarly, in the FDT estimation (at large lags > 30 days), the ensemble mean estimation of both A_{xx} and A_{xy} are greater than the truth by about 50% (Figs. 2a,c), whereas the feedback parameter deviates from the truth by only ~5% (Fig. 1e). Because the feedback parameter depends on the ratio $-A_{xx}^{-1} A_{xy}$, the accurate estimation of the feedback parameter must be caused by the cancelation of a common sampling error factor in both A_{xx} and A_{xy}. This common factor is likely related to the condition of the covariance matrix $\mathbf{C}_{zz}(0)$ and in turn the inverse. This result has been found highly robust in our numerical experiments not only in the one-point model but also in the multipoint model and, to some extent, in the observations. That is, in general, the sampling error is reduced significantly from the individual submatrices $\mathbf{A}_{xx}$ and $\mathbf{A}_{xy}$ to the feedback matrix $\mathbf{B} = -\mathbf{A}_{xx}^{-1}\mathbf{A}_{xy}$, such that the estimation is more accurate
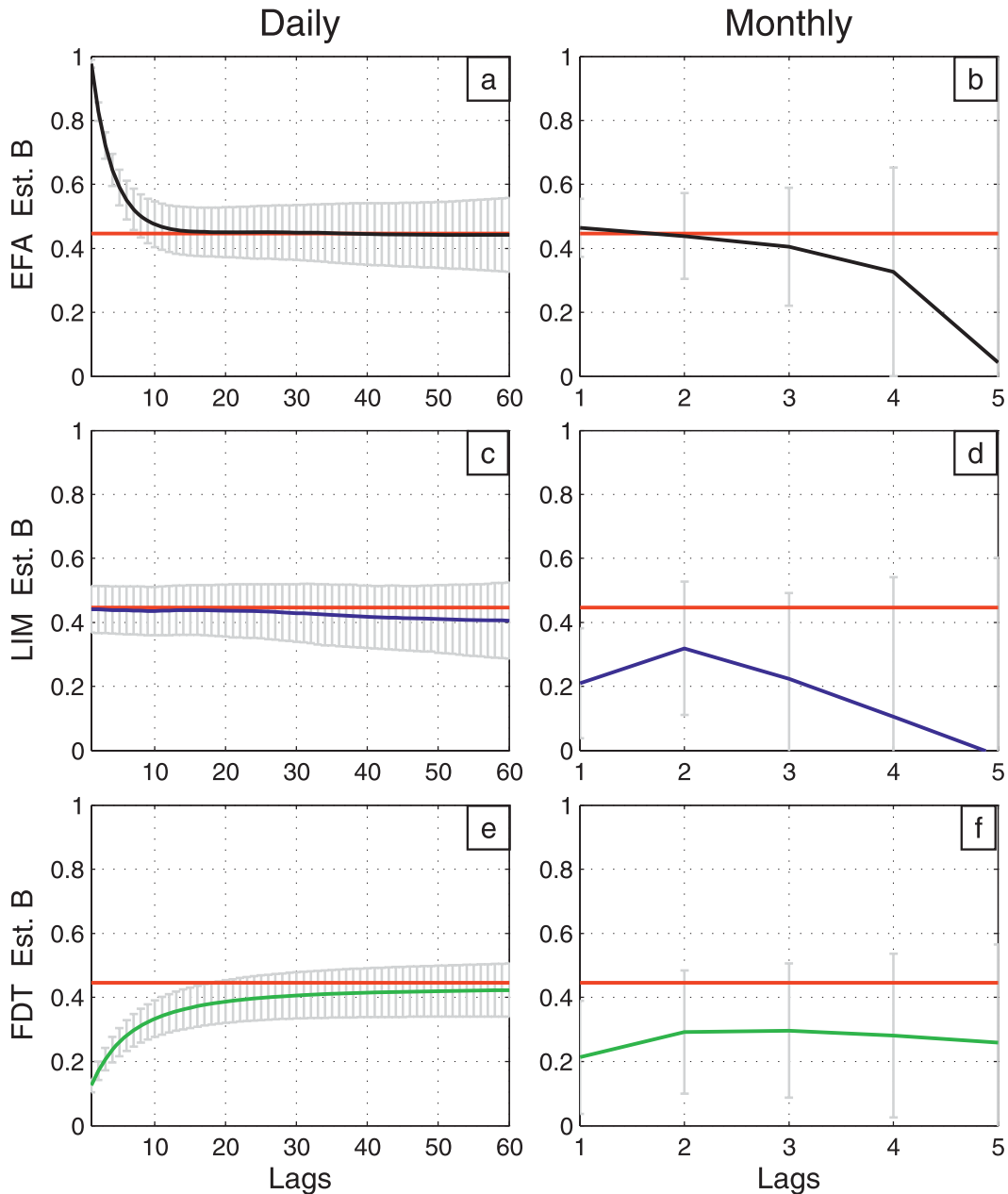
FIG. 1. The univariate GEFA (black line), LIM (blue line), and FDT (green line) estimation of the ocean feedback coefficient, compared with the truth (red line) in the one-point model. The univariate GEFA estimation using (a) daily and (b) monthly data, as a function of lag time τ. A 100-member ensemble is used. The ensemble mean is indicated by the black solid line, and the ensemble spread is indicated by the gray error bar. (c),(d) As in (a),(b), but for LIM estimation. (e),(f) As in (a),(b), but for FDT estimation.

for the feedback matrix **B** than for $\mathbf{A_{xx}}$ and $\mathbf{A_{xy}}$. This also reduces the sensitivity to lag (or the so-called tau test) in **B** than **A** (see appendix B). Our extensive numerical study suggests that the reduced sampling error in $\mathbf{B} = -\mathbf{A_{xx}^{-1}}\mathbf{A_{xy}}$ than in $\mathbf{A_{xx}}$ and $\mathbf{A_{xy}}$ is not a general result for any **x** and **y**; instead, it is caused by the time-scale separation between the fast atmosphere **x** and slow SST **y**.

Therefore, it applies particularly to our assessment of the fast atmospheric feedback response to slow SST forcing.

Now, we turn to GEFA estimation of the feedback parameter and its comparison with LIM and FDT. As discussed in Frankignoul et al. (1998) and LWL08, monthly GEFA can give a good assessment of the feedback parameter for the first few lags. At the best
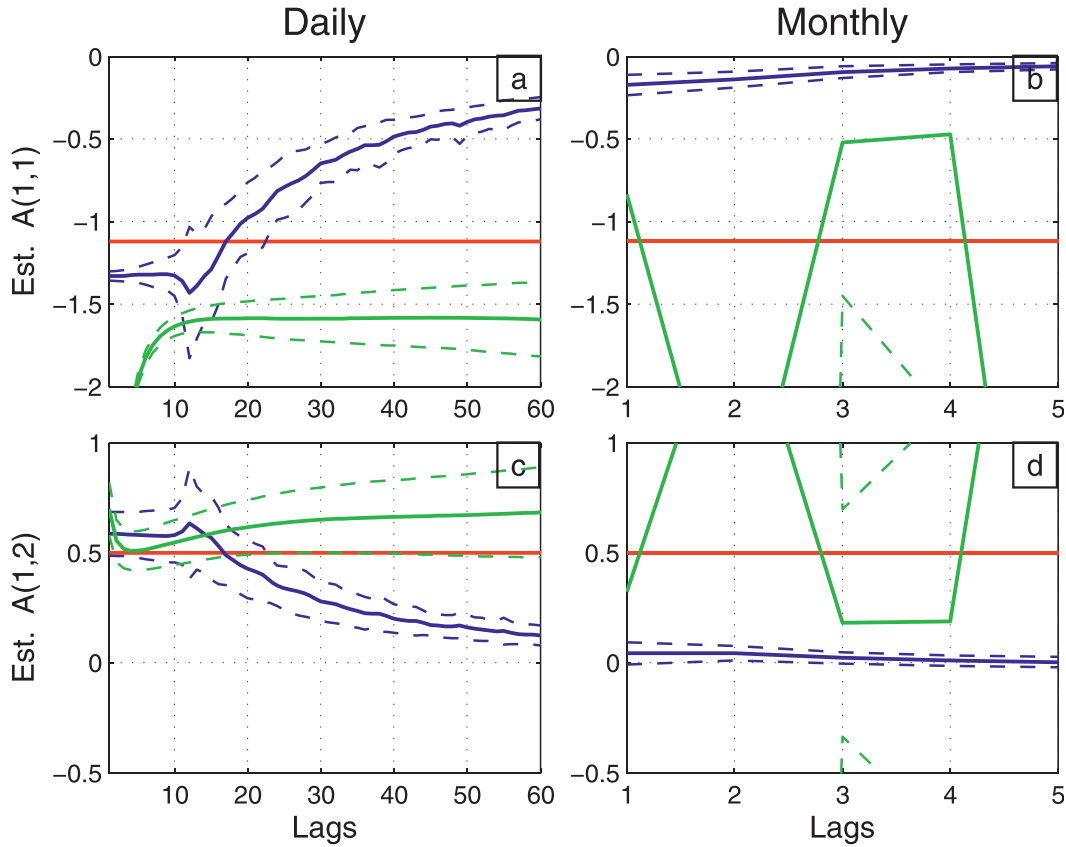
FIG. 2. The system matrix **A** estimated at different lags in the same case of one-point model as in Fig. 1. The element $A(1, 1)$ ($=A_{xx}$) estimated in LIM (blue) and FDT (green) using (a) daily and (b) monthly data, compared with the truth (red). A 100-member ensemble is used, with the ensemble mean in the solid line and the ensemble spread in dash line. (c),(d) As in (a),(b), but for the element $A(1, 2)$ ($=A_{xy}$).

estimation of lag of 1 month (see Fig. 1b), the ensemble mean feedback parameter is ∼0.45 with an ensemble spread of ∼0.05, comparable with the best estimation in daily LIM (Fig. 1c) and FDT (Fig. 1e). This is easy to understand because GEFA is based on the assumption of quasi-equilibrium response (1.2) or (1.3), which is approximately true for monthly data. Here, we further find that GEFA is also able to estimate the feedback parameter using daily data as long as the lag is longer than the atmospheric response time (appendix C). For example, in the daily GEFA estimation (Fig. 1a), the ensemble mean estimation has the best estimation at lags > 15 days, with an accuracy and ensemble spread comparable with the optimal estimation from daily LIM and FDT (Figs. 1c,e).

In summary, our study in the one-point model suggests that, with a realistic sample size, GEFA, LIM, and FDT can all give reasonable assessment of the oceanic feedback parameter. Using daily data, all three methods can give good estimation of the feedback parameter, whereas using monthly data only GEFA seems to give a

good estimation. It is also important to note in daily LIM and FDT that our sample size here, although insufficient for an accurate estimation of the system matrix **A**, is sufficient for a robust estimation of the feedback parameter B.

c. Multipoint model study

We now show that the major conclusions in the one-point model also hold in a six-point model. The model parameters are set the same as in the one-point model, with the addition of the atmospheric advection as $U = -0.08$. The time step is 0.2. Sensitivity studies show that the major conclusions derived from this model remain unchanged in other parameter settings and model resolutions.[4]

---

[4] For models of more grid points, however, a direct estimation of the feedback matrix is subject to larger sampling errors due to the correlation of SST variability among neighboring points (LWL08). Now, the base of the SST forcing needs to be selected carefully. A convenient choice is the leading EOFs (Wen et al. 2010; Fan et al. 2011).
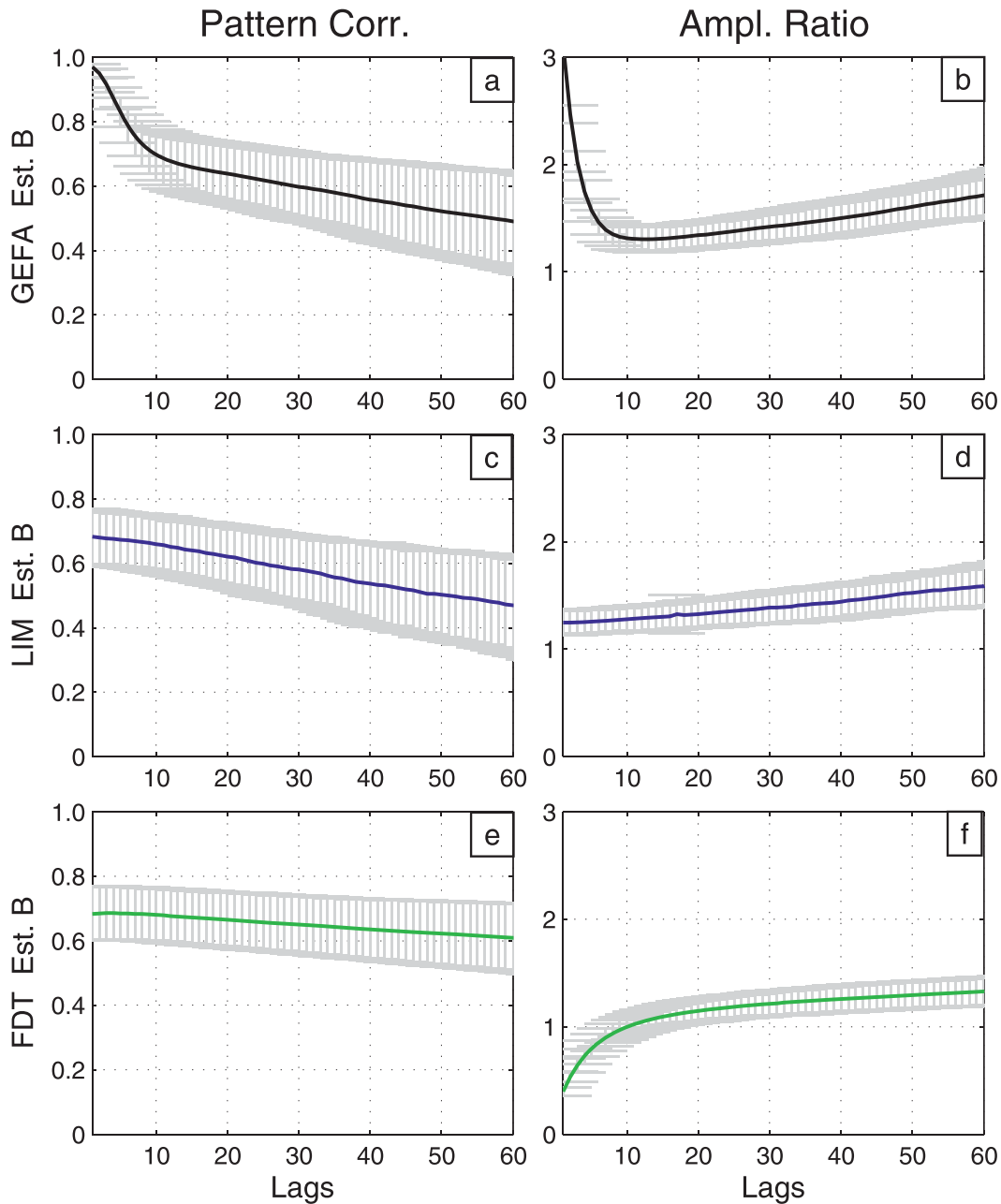
FIG. 3. Daily estimation of the feedback matrix $\mathbf{B}_e$ using GEFA (black line), LIM (blue line), and FDT (green line) at different lags, compared with the truth $\mathbf{B}_t$ in the six-point model. The comparison is made in pattern correlation (corr$\langle \mathbf{B}_e, \mathbf{B}_t \rangle$) and amplitude ratio [std($\mathbf{B}_e$)/std($\mathbf{B}_t$)]. (a) The pattern correlation and (b) the amplitude ratio for GEFA. A 100-member ensemble is used with the ensemble mean in a black solid line and the ensemble spread in gray error bars. (c),(d) As in (a),(b), but for LIM. (e),(f) As in (a),(b), but for FDT.

Figure 3 shows the feedback matrix estimated using daily data. The feedback matrix $\mathbf{B}_e$ is estimated using a 100-ensemble member and is compared with the truth $\mathbf{B}_t$ in pattern correlation corr$\langle \mathbf{B}_e, \mathbf{B}_t \rangle$ and amplitude ratio std($\mathbf{B}_e$)/std($\mathbf{B}_t$). (The pattern correlation and spatial standard deviation are both calculated by lining the

elements of each matrix as one series.) In the GEFA estimation, the pattern correlation of the 100-ensemble mean estimation drops rapidly in the first few lags $\tau < 10$ days and then declines slowly from a correlation of ~0.7 with a narrow spread of ~0.1 (Fig. 3a). Similarly, the amplitude ratio also decreases first rapidly toward

$\tau \sim 10$ days and then increases gradually (Fig. 3b). The high pattern correlation at small lags is due to the persistence of the atmospheric variability, whereas the gradual decrease of the pattern correlation and gradual increase of the amplitude ratio after $\tau \sim 10$ days are due to the increased sampling error.[5] For LIM, the best estimation is achieved at small lags with a pattern correlation of ~ 0.7 and amplitude ratio of ~ 1.3 (Figs. 3c,d). In FDT, the pattern correlation stays around ~ 0.7 while the amplitude ratio increases to slightly over 1 (Figs. 3e,f). Furthermore, the pattern correlation and amplitude ratio of the feedback matrix are not sensitive to lag in LIM (toward the optimal estimation at $\tau \sim 1$ day) and FDT (toward the optimal estimation of $\tau > 10$ days). This insensitivity to lag is due to a cancelation of sampling error between $\mathbf{A}_{xx}$ and $\mathbf{A}_{xy}$ (see Figs. B1, B2) as discussed in the one-point model.

Overall, the three methods in daily estimation give comparable results at their optimal estimation, with a pattern correlation of ~ 0.7 and amplitude ratio of ~ 1.2. In comparison, for monthly estimations (Fig. 4), only GEFA at lag of 1 month gives a good estimation, with a pattern correlation over ~ 0.6 and amplitude ratio of ~ 1.4. The GEFA estimation deteriorates with lag (in months) due to the decrease in the SST autocovariance and in turn increased sampling error (Liu et al. 2006). The monthly estimation of the feedback matrix is poor in LIM or FDT, similar to the one-point model. Therefore, for LIM and FDT, high-frequency sampling of data is desirable for the estimation of the feedback matrix.

In Figs. 3 and 4, the optimal lag for estimating $\mathbf{B}$ is selected with the truth $\mathbf{B}_t$ known. This selection strategy is not feasible in real-world application where the truth is unknown. Here, we will determine the optimal estimation with an empirical approach using the successive pattern correlation corr$\langle \mathbf{B}_\tau, \mathbf{B}_{\tau-1} \rangle$ (where $\mathbf{B}_\tau$ denotes the $\mathbf{B}$ estimated at lag τ) and amplitude ratio std$(\mathbf{B}_{\tau-1})$/std$(\mathbf{B}_\tau)$, as proposed in LWL08. Figure 5 shows the successive pattern correlation and successive amplitude ratio for daily estimation, with the magenta line in each panel denoting the lag of optimal estimation. The optimal lag is selected as the lag at which the successive pattern correlation and amplitude ratio become stabilized to close to 1. Taking the daily GEFA estimation as an example, the optimal selection is picked at lag $\tau = 9$, where the successive pattern correlation and amplitude ratio begin to stabilize at the correlation of ~ 1 and amplitude ratio of ~ 1 (Fig. 5a). This optimal lag is consistent with

the best estimation obtained through the comparison with the truth in Figs. 3a,b. Similarly, the optimal lag determined from the successive pattern correlation and amplitude ratio in LIM and FDT are also similar to those selected from Fig. 3.

Figure 6 summarizes the optimal feedback matrices, for data binned in daily, 4 days, and monthly, estimated in GEFA, LIM, and FDT. The optimal estimation is determined from the successive pattern correlation and amplitude ratio as discussed in Fig. 5, and the estimated feedback matrix is compared with the truth in pattern correlation (Fig. 6a) and amplitude ratio (Fig. 6b). For the three methods, with daily data, they all produce a good pattern correlation (~ 0.7) and amplitude ratio (~ 1.3). The 4-day estimation remains comparable with the daily estimation. The monthly estimation, however, degenerates significantly except for GEFA. These results are consistent with the discussions on Figs. 3 and 4.

In short, our idealized model study suggests that, for our current climate data of several decades, LIM and FDT are valid for high-frequency data whereas GEFA is valid for the high-frequency as well as low-frequency data. Here, the high-frequency data have a sufficiently high frequency of sampling such that it resolves the fast atmospheric dynamics; the low-frequency data are smoothed with coarse graining such that the high frequency atmospheric dynamics is filtered out. Our idealized model study is not a proof of the consistency of the three methods, because it is a special system with a specific sparse system matrix. Nevertheless, it does indicate the possibility of the consistency of the three methods as long as there is a clear time-scale separation between the atmosphere and ocean. This suggests the possibility that the three methods can potentially be used for cross validation against each other.

4. Comparison of GEFA, LIM, and FDT: Observational study

Here, we further show that GEFA, LIM, and FDT also give consistent estimates in the observations. We will study the response of the atmospheric 200-hPa geopotential height (Z200) to the dominant SST variability modes, as in Wen et al. (2010) using monthly GEFA. Specifically, we will compare the monthly GEFA assessment with the pentad LIM and FDT assessments. The geopotential height data are from the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) global reanalysis data from 1958 to 2007 (http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html) with a resolution of $2.5° \times 2.5°$ (144 longitude $\times$ 73 latitude).

---

[5] This has been demonstrated in the coupled model using experiments with the ocean forcing $\mathbf{A}_{xy}\mathbf{y}$ term surpressed in the atmospheric Eq. (1.1a).
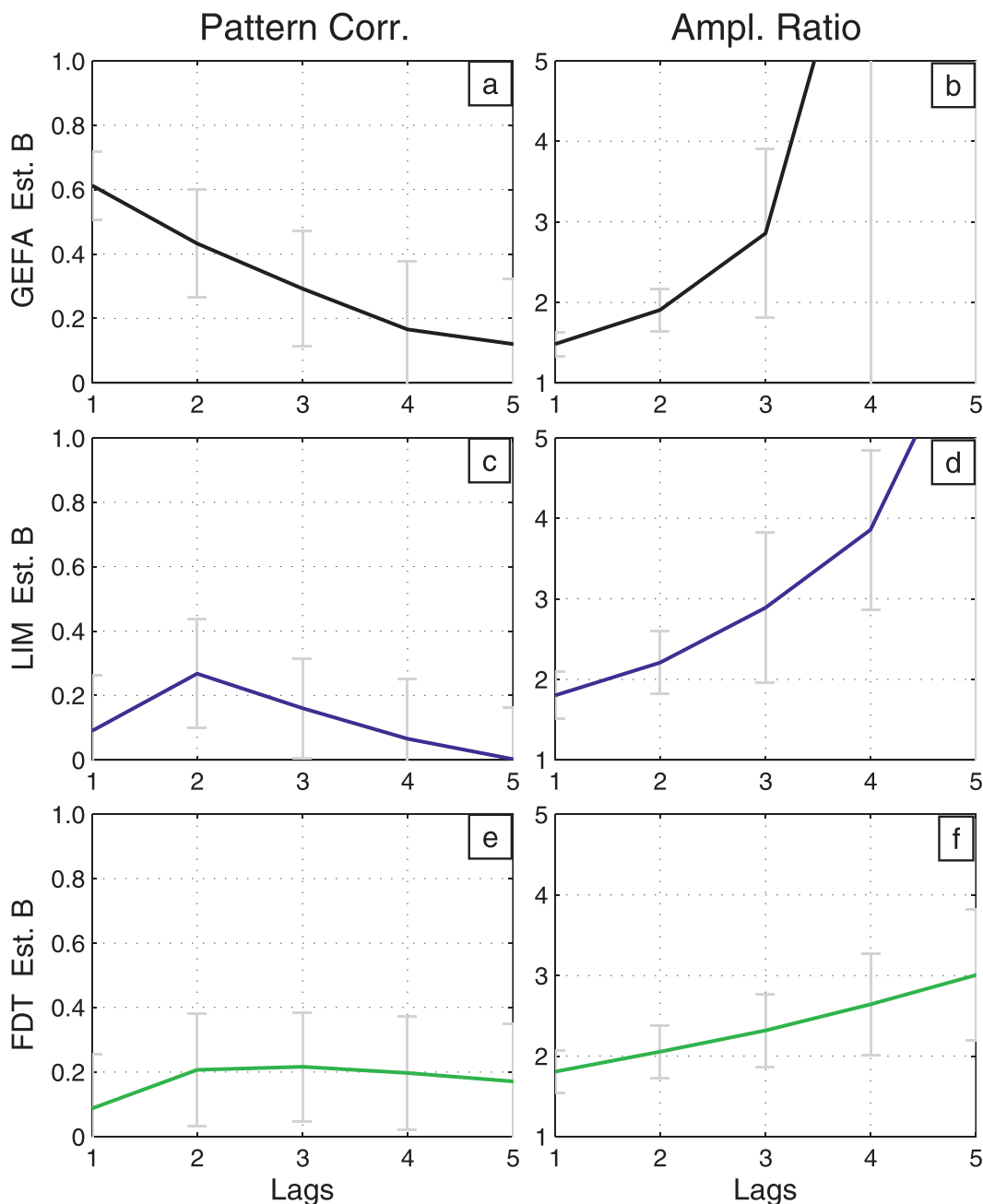
FIG. 4. As in Fig. 3, but for monthly estimation.

The SST data are the skin temperature from NCEP–NCAR reanalysis with the same temporal resolution (daily) as the geopotential height data; the spatial resolution is 1.875° in longitude and a variable Gaussian grid in latitude (192 longitude × 94 latitude). All the data are anomalies from the seasonal cycle and are then detrended with a third-order polynomial filter. Similar to Wen et al. (2010), the set of SST forcing is derived from the two leading EOFs in the TP (20°S–20°N,

120°E–60°W), the tropical Indian Ocean (TI; 20°S–20°N, 35°–120°E), the tropical Atlantic (TA; 20°S–20°N, 65°W–15°E), the North Pacific (NP; 20°–60°N, 120°E–60°W), and the North Atlantic (20°–60°N, 100°W–20°E). The first two EOF modes of the five basins are combined into a set of EOF modes to represent the ocean forcing. These EOFs represent the dominant oceanic variability modes, including the ENSO mode in the tropical Pacific (TP1) and the North Pacific Oscillation (NPO) over the
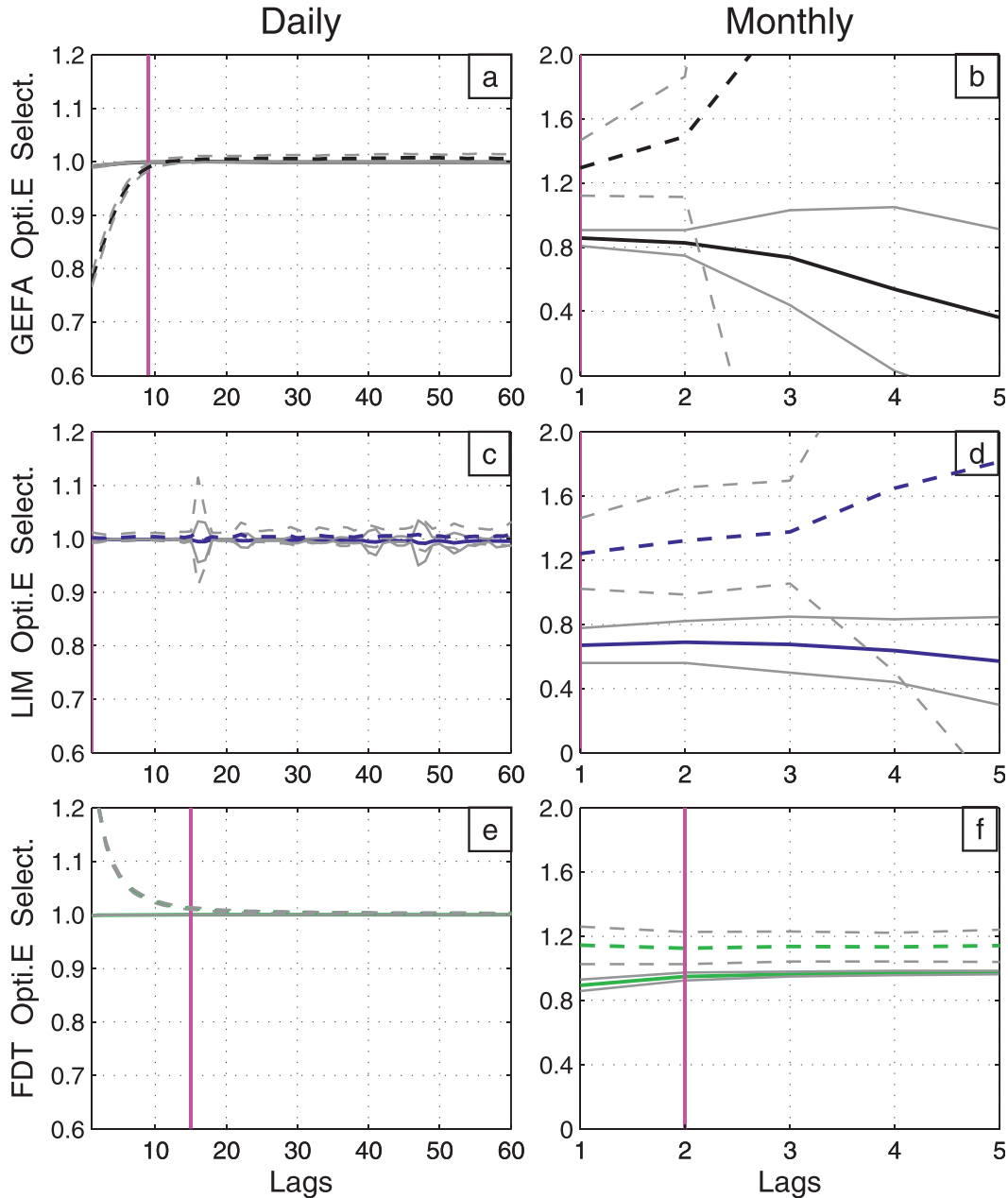
FIG. 5. Optimal estimations (magenta line) for GEFA (black line), LIM (blue line), and FDT (green line) of the feedback matrix **B** in the six-point model. The lag of optimal estimation (magenta line) is selected according to the successive pattern correlation (corr⟨$\mathbf{B}_\tau$, $\mathbf{B}_{\tau-1}$⟩) and amplitude ratio [std($\mathbf{B}_{\tau-1}$)/std($\mathbf{B}_\tau$)]. The successive pattern correlation (solid) and amplitude ratio (dash) for GEFA estimation using (a) daily and (b) monthly data, as a function of lag time τ. A 100-member ensemble is used, with the ensemble mean in black and the ensemble spread in gray. (c),(d) As in (a),(b), but for LIM. (e),(f) As in (a),(b), but for FDT.

North Pacific (NP1) (for their pattern and time series, see Wen et al. 2010). It is important to point out that, as in the idealized model studies, our observational studies in LIM and FDT also show a better tau test for the feedback matrix **B** than the atmospheric system matrix **A** (see Figs. B3, B4). In the following, we only show the response to TP1 and NP1 modes, because the conclusion holds for other modes.

Figures 7a,b show the response of Z200 to TP1 and NP1 using monthly GEFA (Figs. 7a,b). [These monthly responses are reproduced in the pentad GEFA estimation, with an excellent pattern correlation (0.99 for both
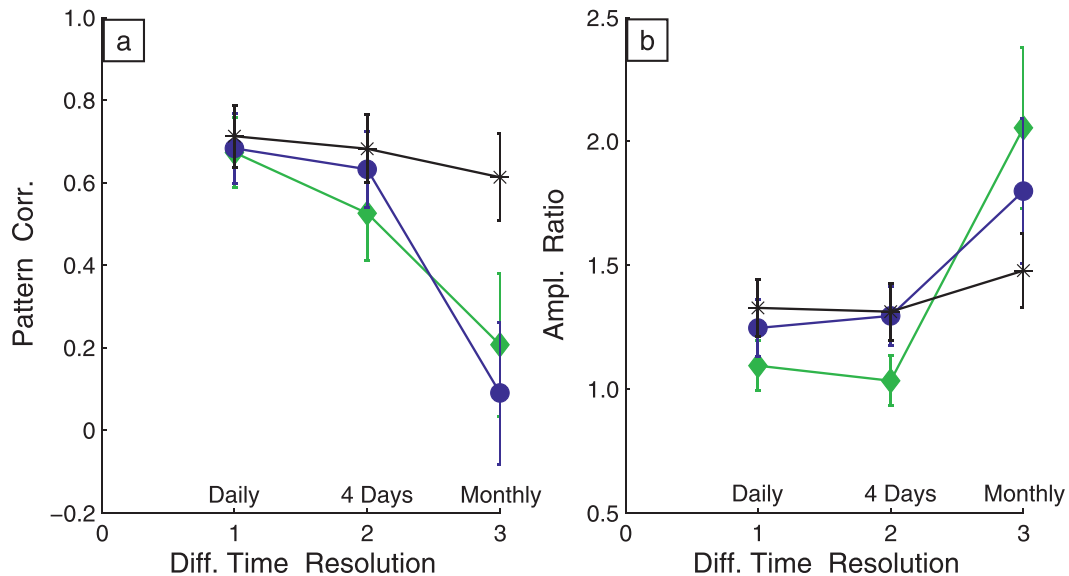
FIG. 6. The comparison of the optimal GEFA (black line with star), LIM (blue line with circle), and FDT (green line with diamond) estimations of the feedback matrix **B** in different time resolution (selected using the successive correlation and amplitude ratio as in Fig. 5). (a) The pattern correlation of the optimal estimation with the true **B**, and (b) the amplitude ratio of the optimal estimation over the true **B**. A 100-member ensemble is used with the ensemble mean in the solid line and the ensemble spread in the error bar. The x axis marks the time resolution for daily, four-day, and monthly estimations.

TP1 and NP1) and amplitude ratio (1.0 for TP1 and 1.1 for NP1). The amplitude of each response pattern is calculated as the mean of the absolute value of the response at each point.] The GEFA responses are now compared with the response in pentad LIM (Figs. 7c,d) and FDT (Figs. 7e,f). The lag of optimal estimation is $\tau = 1$ month for GEFA and $\tau = 3$ and 12 pentads for LIM and FDT, respectively, which is selected using the empirical method of successive pattern correlation and amplitude ratio as discussed in Fig. 5.[6] As in Wen et al. (2010), the statistical significance of the response is tested with a Monte Carlo method in which the atmospheric time series is randomly scrambled 200 times. As in Wen et al. (2010), the monthly GEFA assessment of the atmospheric response to TP1 mode (Fig. 7a) shows a pair of Rossby waves locally over the tropical eastern Pacific, a Pacific North America (PNA) atmospheric teleconnection toward North America, and a Pacific South America (PSA) teleconnection toward the Antarctic. This atmospheric response pattern, at the first order, can be understood as a baroclinic equatorial Rossby wave response to a deep tropical heating associated with the warm equatorial SST (Gill 1980) and the

subsequent barotropic Rossby wave propagation toward the extratropics (Hoskins and Karoly 1981). Furthermore, the GEFA response to TP1 is very similar to the multiple regression patterns derived with lag of +1, 0, and −1 (not shown). This is because TP1 is affected little by internal atmospheric variability such that the covariance between the SST and atmospheric internal variability is always small [i.e., (1.5)]. Thus, the feedback matrix **B** can be derived in (1.7), in principle, at any lag.

These major features of the GEFA atmospheric response are all reproduced in the pentad LIM assessment (Fig. 7c) and FDT assessment (Fig. 7e); the pattern correlations between the GEFA response and LIM and FDT are 0.96 and 0.99, respectively. Furthermore, the amplitudes of the responses in LIM and FDT are also comparable with that of GEFA. For example, the amplitude of the Rossby wave center is ~20 gpm in monthly GEFA, in response to the unit ENSO mode (corresponding to the SST pattern with the maximum anomaly of 1.1°C in the eastern equatorial Pacific); this is comparable with the ~20- and ~15-gpm responses in LIM and FDT, respectively. The overall consistency in the amplitude of the response can also be seen in the amplitude ratio relative to the monthly GEFA estimate: 1.1 for LIM and 1.2 for FDT. The consistent assessment suggest that GEFA, LIM, and FDT can provide a powerful cross validation of the major features of the

---

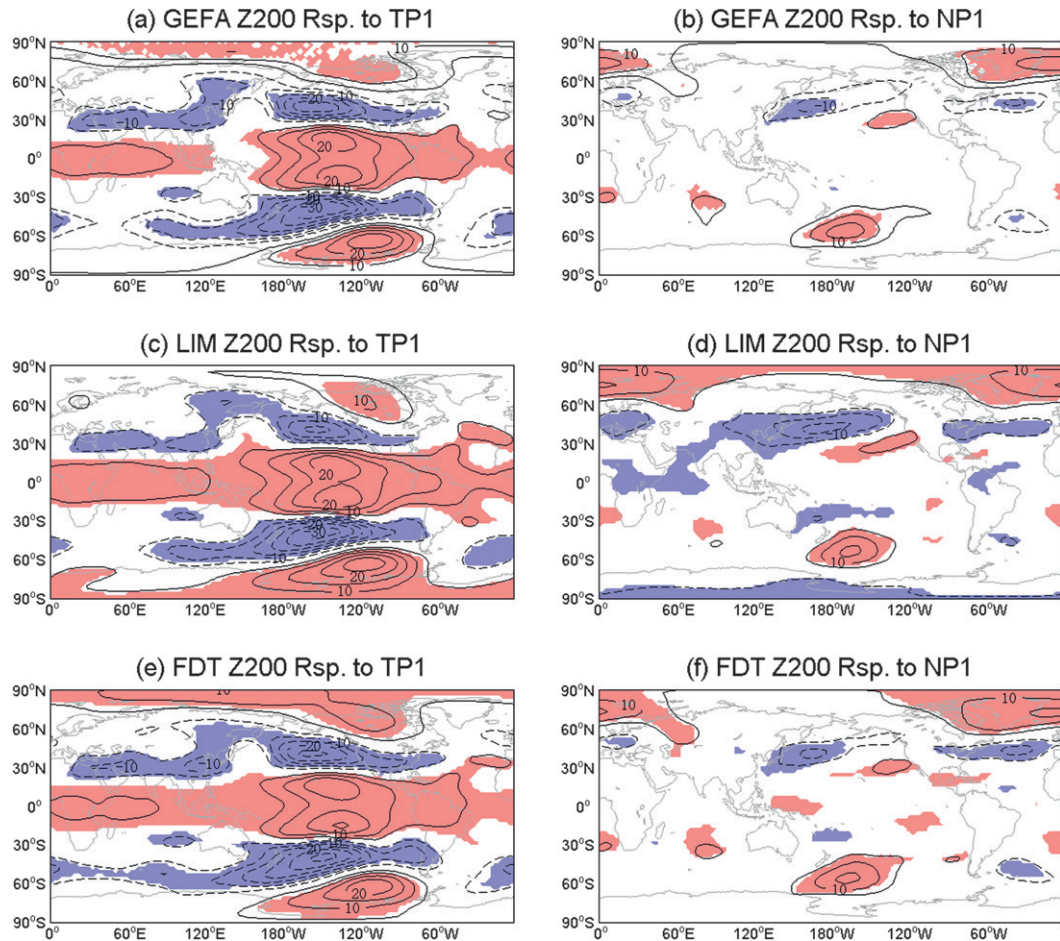[6] The results are similar for optimal LIM and FDT estimations in 2-day data and weekly data.

FIG. 7. Comparison of (a),(b) monthly GEFA response with the pentad (c),(d) LIM and (e),(f) FDT response for the response of the Z200 to the leading EOF mode in the (a),(c),(e) TP (TP1) and (b),(d),(f) NP (NP1). Solid (dashed) contours are for positive (negative), with contour interval of 5 gpm (the zero contour is omitted). The regions above 95% confidence level are shaded in red for positive and in blue for negative.

atmospheric response to TP1, significantly enhancing our confidence in the assessment.

The atmospheric response to the North Pacific SST variability NPO (mode is also found consistent in the three assessments. Figure 7b shows the monthly GEFA estimation of the response of Z200 to NP1 (here, the SST EOF is negative in its maximum loading region in the midlatitude North Pacific). As in Wen et al. (2010), this response is characterized by a local response of warm SST ridge over the Aleutian low and a remote response downstream over the North Atlantic. This warm SST-ridge response locally over the North Pacific is likely to be caused by a dominant winter atmospheric response associated with eddy–mean flow interactions (Peng et al. 1997; Peng and Whitaker 1999). The downstream teleconnection to the North Atlantic resembles closely to the winter atmospheric teleconnection pattern

known as the Aleutian–Icelandic seesaw (Honda et al. 2001, 2005). This teleconnection is initiated by an accumulation of atmospheric wave activity in the Aleutian low region in early to midwinter. Part of the wave activity then propagates across North America in the form of stationary Rossby wave train, forming the stationary anomaly over the North Atlantic. As in the response to TP1, the GEFA response is similar to the corresponding multiple regression pattern with the SST leading atmosphere by one month (not shown). In contrast to the response to TP1, however, this GEFA response pattern differs dramatically from the multiple regression pattern at zero lag and with the atmosphere leading SST by one month, both of which show a much stronger atmospheric activity over the North Pacific (not shown). This is because the North Pacific SST variability is forced predominantly by the atmospheric

internal variability such that the covariance between the atmospheric internal variability and SST is small [i.e., (1.5)] only when the SST leads the atmosphere, as in (1.6).

These major features in GEFA response are in good agreement with the pentad LIM estimation (Fig. 7d) and FDT estimation (Fig. 7f). The pattern correlation between the GEFA response and LIM and FDT are 0.94 and 0.99, respectively. The response amplitudes are also comparable, with the maximum local response ~10 gpm in response to the NP1 EOF mode (with a maximum anomaly of $-0.6°$C in the Kuroshio Extension region). The overall amplitude ratio relative to the monthly GEFA is 0.97 for LIM and 1.13 for FDT. The consistent assessment of the atmospheric response to the NP1 in the three methods should provide a benchmark for modeling studies, which so far have shown diverse results (e.g., Kushnir et al. 2002).

5. Summary and discussion

This paper compares the statistical assessment of the atmospheric response to SST variability with a realistic sample size using three statistical methods: GEFA, LIM, and FDT. The comparison is made in idealized models and in the observations. Our study suggests that the three methods are able to give a consistent assessment of the atmospheric response. Therefore, the three methods can be used together for cross validation of the assessment of the atmospheric response to surface forcing in the observations.

For practical applications, as far as the assessment of the atmospheric response to SST variability is concerned, monthly GEFA is still the most convenient approach. For some applications where only monthly data are available (e.g., for the diagnosis of climate feedback from climate model outputs, which are usually stored in monthly data), monthly GEFA is the only method feasible. However, as long as daily or weekly data are available, the cross validation using LIM and FDT (and high-frequency GEFA) provides important advantages. First, as studied before for monthly GEFA estimations at different lags (LWL08; Wen et al. 2010), the response patterns tend to be robust in all the methods. Here, for the response to the leading SST modes in the observations, the pattern correlation among the three estimates is usually over ~0.95. Therefore, the cross validation enhances our confidence of the response pattern. Second, more important than the response pattern, the cross validation greatly improves our confidence on the response amplitude. The response amplitude has been a problem in monthly GEFA estimation. Because of the rapid decorrelation (in months) of SST forcing, the response amplitude tends to increase

substantially with lag in monthly GEFA (Wen et al. 2010), making it difficult to determine the correct amplitude. Here, the estimated response amplitude of the three methods are consistent within 10%–20% in terms of the overall amplitude ratio, giving us much more confidence of the amplitude estimation. Third, the cross validation may reduce the uncertainty associated with some assumptions on the monthly GEFA. Because of the large lag sensitivity (or the difficult in tau test) of monthly GEFA (appendix B), one problem has been the choice of the lag. For example, because of strong nonlinear eddy–mean flow interactions, the atmospheric response to SST anomaly may take longer than a month (e.g., Ferreira and Frankignoul 2008). This would suggest that one may choose the estimation at lags of 2 months or even longer. However, an increased lag leads to a reduced SST autocovariance and, in turn, an increased sampling error, leaving a difficult choice to balance the errors associated with the dynamics of the atmospheric response and sampling error. The cross validation here helps to reduce this uncertainty, because LIM and FDT (or high-frequency GEFA) do not depend critically on the slow response time scale of the atmospheric response. If the optimal estimation from the three methods show a consistent result, it is highly likely the estimation is truly optimal. However, one should be cautious in using high-frequency data for estimation. This is because the estimation here is accurate only to the extent that the dynamics can be approximated as a linear stochastic processes (see footnotes 1 and 2). Because regional synoptic process can be strongly nonlinear at synoptic time scales, one should be cautious in applying the method to very short data, such as daily. A proper average, such as weekly, may reduces the noise related to nonlinear processes and therefore provides the optimal estimation (e.g., Newman et al. 2009). Indeed, in addition to the implication on sampling error, the consistency between the pentad LIM and FDT (and GEFA) assessment and the monthly GEFA assessment in the observations (Fig. 7) may have an important dynamic implication: it implies that the pentad average seems to be sufficiently long to allow the dynamic process of the global atmospheric response to SST to be approximated as a linear stochastic process as in (1.1a).

One important and robust finding here is that, in LIM and FDT, the feedback matrix **B** is less sensitive to lags than the system matrix **A**. Therefore, the feedback can be estimated more accurately than the dynamics of the full temporal evolution. Physically, this makes sense because the feedback only involves the dynamics of the quasi-equilibrium (or stationary) response of the atmosphere, whereas the temporal evolution involves the full dynamics of the system and therefore requires much more information. However, this finding remains

empirical at this stage, because we have not found a general mathematical proof.

Finally, it is important to point out that this paper is not a general comparison of the three methods. Instead, this paper focuses on the assessment of the response of the fast component (atmosphere) to the slow component (ocean) in a multiple time-scale system. LIM and FDT, in principle, can be used to also study the full temporal evolution of the coupled dynamics, although the study of the full dynamics would require much large sample size. GEFA, however, only has a limited application: it only applies to the study of the response of the fast component to the slow component.

It is encouraging that we now have at least three statistical methods that can be used for cross validation of the atmospheric response in the observations. This cross validation is important, because, in many applications, the atmospheric response to a surface forcing, such as extratropical SST (Kushnir et al. 2002) and land ecosystem (Liu et al. 2006; Notaro et al. 2006), remains highly uncertain in the observations as well as in models. This uncertainty occurs because the feedback signal is overwhelmed by internal atmospheric variability. As discussed before, the three statistical methods here apply to the case where we have specified surface forcings, such as the forcing in different regions, or in different EOF modes, as studied here. These methods are the extension of the simple regression method previously used and are complimentary to another type of statistical methods that identify the optimal response patterns, such as MCA, GEFA–SVD, and maximum response estimation (MRE; Frankignoul et al. 2011). All these statistical assessments are important both in reducing the uncertainty of the assessment and in shedding light on the dynamics of the climate feedback. Finally, these statistical methods can be combined with dynamic model experiments (e.g., Liu and Wu 2004) to provide a comprehensive statistical–dynamical strategy for the assessment and understanding of the atmospheric response to surface boundary forcing.

APPENDIX A

Approximated Equation for the Averaged Variable

Here, we give a heuristic derivation of the averaged atmospheric Eq. (1.2) using a scaling analysis. Our derivation is not a general proof, which is beyond the scope of this paper. Instead, it is intended to give the readers some sense why fundamentally the quasi-equilibrium balance should be valid for the fast atmospheric equation averaged over a long time. We will only consider the original Eq. (1.1a) in its simplest scalar form [the single point model in (2.1a)]

$$\frac{dx}{dt} = -ax + by + n. \tag{A.1}$$

Averaged over the time period from $t - L$ to $t + L$, we have the averaged equation

$$\frac{x(t + L) - x(t - L)}{2L} = -a\overline{x}(t) + b\overline{y}(t) + \overline{n}(t), \tag{A.2}$$

where the overbar denotes the time average as, say, for $x(t)$,

$$\overline{x}(t) \equiv \frac{1}{2L}\int_{t-L}^{t+L} x(t')\, dt'. \tag{A.3}$$

The averaged Eq. (A.2) can be written as a quasi-equilibrium balance similar to (1.2) as

$$0 = -a\overline{x}(t) + b\overline{y}(t) + \overline{n}^*(t), \tag{A.4}$$

where the new noise term consists of the averaged noise and the fluctuation of the averaged local variability (left-hand side term) in (A.2),

$$\overline{n}^*(t) = \overline{n}(t) - \frac{x(t + L) - x(t - L)}{2L}. \tag{A.5}$$

Furthermore, for a sufficiently long average, the left-hand side of (A.2) can be shown to diminish more rapidly than the right-hand side terms, such that the left-hand side becomes negligible in (A.2). Thus, (A.2) is equivalent to (1.2), or (A.5) can be approximated as

$$\overline{n}^*(t) \approx \overline{n}(t). \tag{A.6}$$

To show this, we assume x, y, and n each can be approximated as a red noise process with persistence times of $1/\lambda_x$, $1/\lambda_{yn}$ and $1/\lambda_n$, respectively. The covariance function is therefore, say, for x,

$$\langle x(t + \tau), x(t) \rangle = \sigma_x^2 e^{-\lambda_x |\tau|}. \tag{A.7}$$

The variance of the averaged variable can be derived as

$$\langle \overline{x}(t), \overline{x}(t) \rangle = 2\sigma_x^2 \frac{2\lambda_x L + e^{-2\lambda_x L} - 1}{(2\lambda_x L)^2}. \qquad \text{(A.8)}$$

The maximum variance is achieved at the limit of short average time $\lambda_x L \to 0$ as the original variance $\langle \overline{x}(t), \overline{x}(t) \rangle = \sigma_x^2 = \langle x(t), x(t) \rangle$. As the average time increases, the variance decreases. For an average time longer than the decorrelation time $1/\lambda_x$,

$$\lambda_x L \gg 1, \qquad \text{(A.9)}$$

the variance (A.8) approaches

$$\langle \overline{x}(t), \overline{x}(t) \rangle \to \frac{\sigma_x^2}{\lambda_x L}. \qquad \text{(A.10)}$$

Therefore, the magnitude of the first term on the right-hand side of the averaged Eq. (A.2) decreases as $1/\sqrt{L}$. This $1/\sqrt{L}$ decrease rate is also true for the averaged noise, which can be considered either as a red noise in the form of (A.7) but with a short persistence time $1/\lambda_n \le 1/\lambda_x$ or simply as a white noise. The variance of the averaged y can also be considered similarly as for x except for a longer persistence time $1/\lambda_y \gg 1/\lambda_x$, such that $\langle \overline{y}(t), \overline{y}(t) \rangle$ may not decrease substantially if the average time is longer than the persistence time of the atmosphere [i.e., (A.9)] but not much longer than the persistence time of the SST (e.g., $\lambda_y L \sim 1$). Overall, however, the magnitude of each term on the right-hand side of the averaged Eq. (A.2) decreases no faster than $1/\sqrt{L}$. In comparison, the variance of the left-hand side of (A.2) is on the order of

$$\left\langle \frac{x(t+L) - x(t-L)}{2L}, \frac{x(t+L) - x(t-L)}{2L} \right\rangle = \frac{\sigma_x^2}{2L^2}(1 - e^{-2\lambda_x L}) \to \frac{\sigma_x^2}{2L^2}. \qquad \text{(A.11)}$$

So, its magnitude decreases as $1/L$, much faster than the $1/\sqrt{L}$ rate of the right-hand side terms. In particular, from (A.11) and (A.10), we have the ratio of the two terms in (A.2) as

$$\frac{\left\langle \dfrac{x(t+L) - x(t-L)}{2L}, \dfrac{x(t+L) - x(t-L)}{2L} \right\rangle}{a^2 \langle \overline{x}(t), \overline{x}(t) \rangle} \to \frac{\lambda_x}{2La^2} \sim \frac{\lambda_x}{2L\lambda_x^2} \sim \frac{1}{2L\lambda_x} \ll 1. \qquad \text{(A.12)}$$

Here, the first "$\sim$" sign has used the fact that the SST variability is much slower than the atmospheric variability so that the persistence time of atmosphere $x(t)$ is approximately $1/\lambda_x \sim 1/a$ in Eq. (A.1).

APPENDIX B

Tau Test in the Multipoint Model and the Observations

Our study shows that the estimation of the feedback matrix **B** is better than that of the system matrix **A** in LIM and FDT. This has implication to the tau test for LIM and FDT, which are designed to validate the linear stochastic model (1.1a) and (1.1b) in terms of the sensitivity of the estimated **A** to lags (Penland 1989; Penland and Sardeshmukh 1995a). The linear stochastic model is more likely to be valid if the matrix **A** has a smaller lag sensitivity or a more robust tau test. This tau test also applies to the feedback matrix **B** here. The smaller lag sensitivity of **B** than **A** implies that the tau test is more robust for **B** than **A**. This has been shown in the one-point model (Figs. 1, 2) and can be

shown to be valid in the multipoint model and the observations.

Figure B1 shows the amplitude ratio between the estimation and the truth similar to the **B** estimation in the six-point model in Figs. 3d,f, but now for the submatrices $\mathbf{A_{xx}}$ and $\mathbf{A_{xy}}$. For LIM, within the first 60 lag days, the amplitude ratio decreases dramatically from 1.2 to 0.2 for $\mathbf{A_{xx}}$ (Fig. B1a) and from 1.7 to 0.5 for $\mathbf{A_{xy}}$ (Fig. B1b) but increases only modestly from 1.2 to 1.6 for **B** (Fig. 3d). Similarly, in FDT, the amplitude ratio stays above 1.6 for $\mathbf{A_{xx}}$ (Fig. B1c) and are close to $\sim$2 for $\mathbf{A_{xy}}$ (for large lags in Fig. B1d) but stays below $\sim$1.4 for **B** (Fig. 3f). The implication is that some common sampling error factor in the magnitude is canceled between $\mathbf{A_{xx}}$ and $\mathbf{A_{xy}}$ such that the estimation of **B** is improved.

Figure B2a shows the tau test, or lag sensitivity, for the daily LIM estimation in a randomly selected realization. For a given point $\mathbf{x}_i$, the atmospheric response to the 6 SST forcing (points) is represented by a vector, which is the ith row of the submatrix $\mathbf{A_{xy}}$ (denoted as $\mathbf{A_{ixy}}$)[7] or the

---

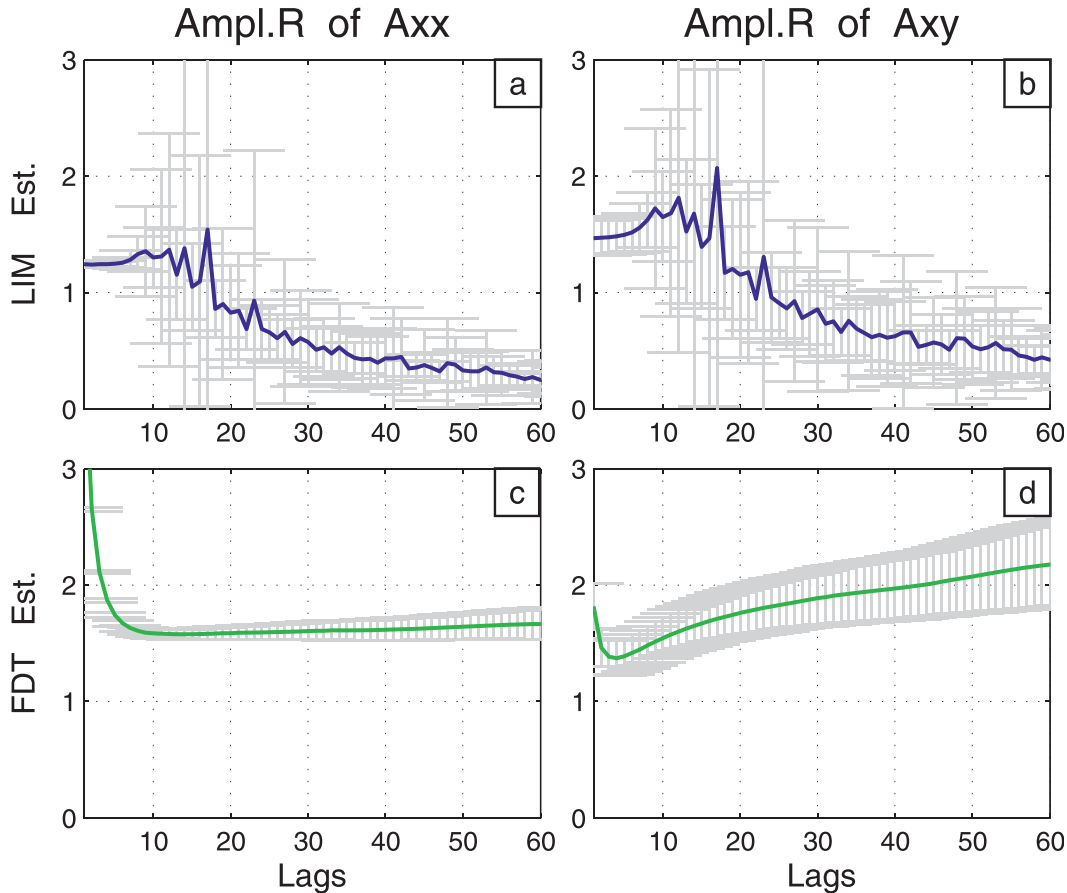[7] This form of the tau test is suggested by Dr. Sang-Ik Shin.

FIG. B1. The amplitude ratio between the estimation and truth at different lags for the submatrices $\mathbf{A}_{xx}$ and $\mathbf{A}_{xy}$ in the six-point model. Daily LIM estimation of (a) $\mathbf{A}_{xx}$ and (b) $\mathbf{A}_{xy}$. A 100-member ensemble is used with the ensemble mean as the solid line and the ensemble spread as the gray error bar. (c),(d) As in (a),(b), but for daily FDT estimation.

ith row of the feedback matrix $\mathbf{B}$ (denoted as $\mathbf{B}_i$). Figure B2a plots the norm of the forcing vector as a function of the lag for the $\mathbf{A}_{ixy}$ (dashed) and for $\mathbf{B}_i$ (solid) for all the six atmospheric points (six solid curves and six dashed curves). It is seen that the norms of the forcing vectors decrease dramatically with lag in the system matrix $\mathbf{A}_{ixy}$ but change only slightly in the feedback matrix $\mathbf{B}_i$. This smaller lag sensitivity in the feedback matrix than in the system matrix is consistent with the discussion in Fig. B1. Figures B2b–d further illustrates the lag sensitivity of the forcing vectors in $\mathbf{B}$ over $\mathbf{A}$ for the midpoint of the atmosphere ($i = 3$) in 10 randomly selected realizations. Consistent with the discussion on the realization in Fig. B2a, the sensitivity to lag is much smaller in the feedback matrix (Fig. B2b) than in the system matrices (Figs. B2c,d).

Now, we turn to the tau test for the pentad LIM estimation of the observed Z200 response to the set of 10 SST forcings as discussed in Fig. 7. First, we select 60 grid points in the atmosphere uniformly (area weighted) across the globe to illustrate the lag sensitivity. As in Fig. B2, for a given point $\mathbf{x}_i$, Figs. B3a,b show the norm of each forcing vector as a function of lag for $\mathbf{A}_{ixy}$ [denoted as $m_{iA}(\tau)$] and $\mathbf{B}_i$ [denoted as $m_{iB}(\tau)$], respectively. It is seen that the norms decrease with lag substantially for $\mathbf{A}_{ixy}$ but increase modestly for $\mathbf{B}_i$, reminiscent of the tau test in the simple model in Fig. B2. The lag sensitivity can be further quantified using relative lag sensitivity, which is defined as the standard deviation of the lag sensitivity relative to its mean, $\sigma[m_{iA}(\tau)]/\overline{m}_{iA}(\tau)$ and $\sigma[m_{iB}(\tau)]/\overline{m}_{iB}(\tau)$. The relative lag sensitivity is plotted for every point over the globe for $\mathbf{A}_{xy}$ (Fig. B4a) and $\mathbf{B}$ (Fig. B4b). It is seen that the relative sensitivity is usually below 10%–20% for $\mathbf{B}$ (global mean $\sim$ 0.1) but below 40%–50% for $\mathbf{A}_{xy}$ (global mean $\sim$ 0.30). Therefore, the relative sensitivity of $\mathbf{B}$ is below 50% of $\mathbf{A}_{xy}$ over most of the globe, as seen in their ratio (Fig. B4c) (global mean of $\sim$0.5).
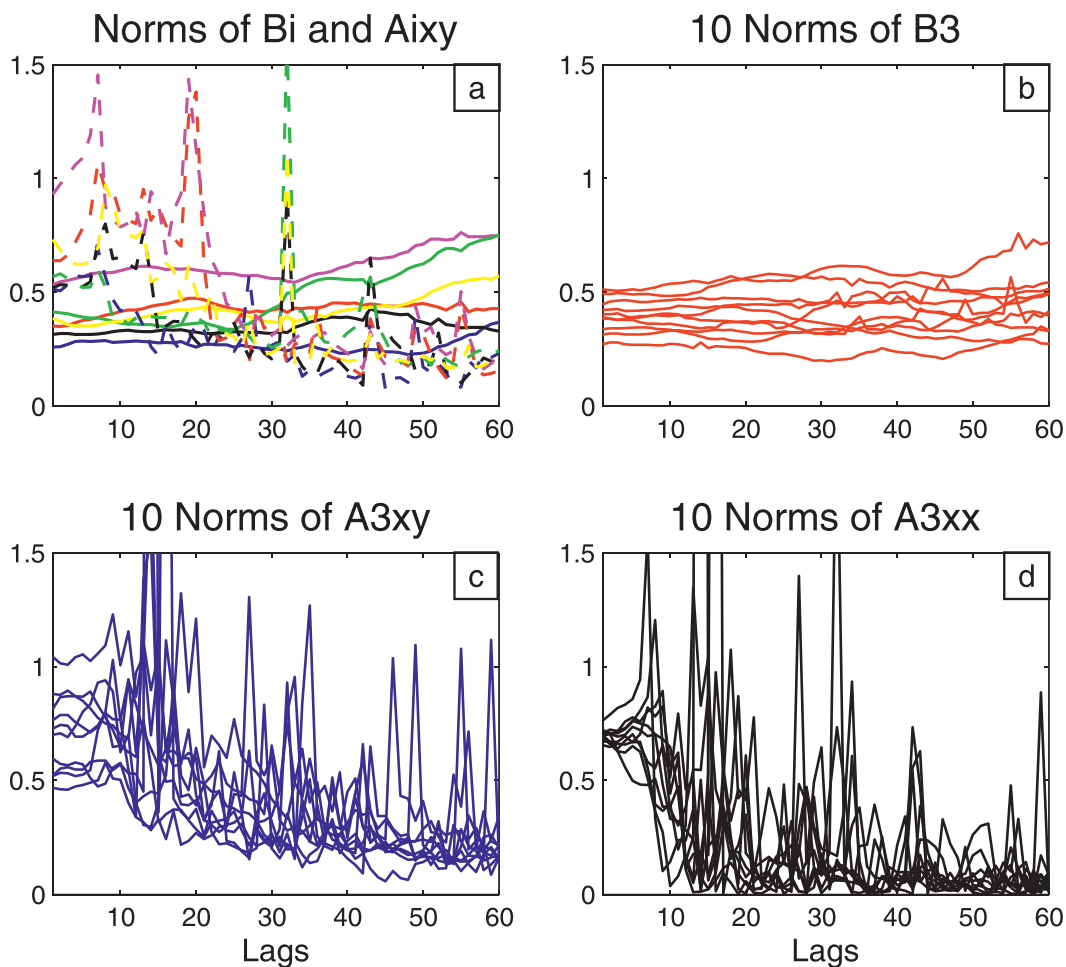
FIG. B2. Tau test for daily LIM estimation in the six-point model for the norm of forcing vectors as a function of lags. (a) The solid lines show the norms of the SST forcing vectors to the six points of the atmosphere in the feedback matrix ($\mathbf{B}_i$, $i = 1, \ldots, 6$, with the response at each point indicated by a different colors) in a randomly selected realization. The dashed lines show the norms of the corresponding system matrix $\mathbf{A}_{ixy}$ ($i = 1, \ldots, 6$, with the response at each point indicated by the same color as in the feedback matrix). (b) Tau test of the feedback matrix for atmospheric point 3 ($\mathbf{B}_3$) in 10 random selected realizations. (c),(d), As in (b), but for the corresponding system matrices (c) $\mathbf{A}_{3xy}$ and (d) $\mathbf{A}_{3xx}$.
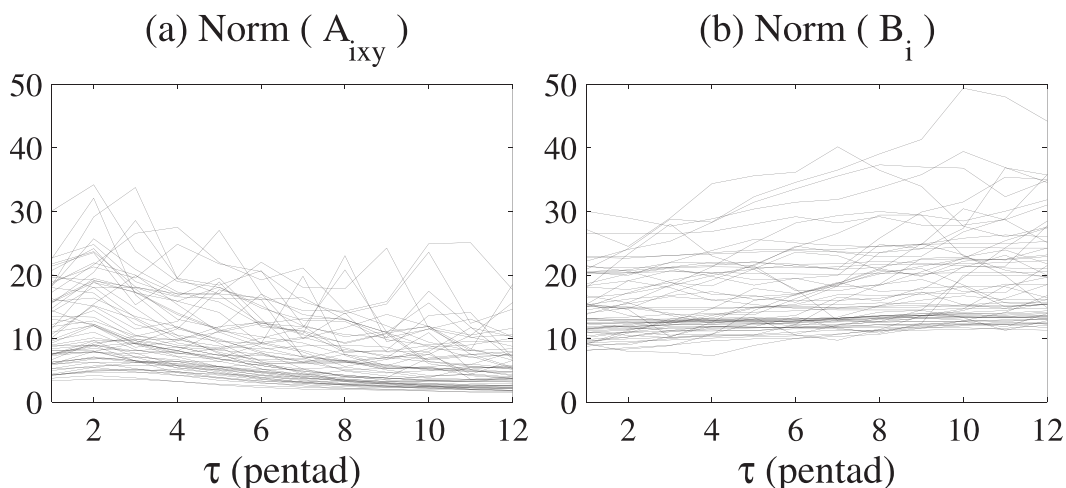

FIG. B3. Tau test in pentad LIM for the norm of the forcing vectors to the Z200 in the observations. A total of 60 lines are plotted in each panel, representing 60 atmospheric grids selected uniformly (area weighted) across the globe.

(a) Std [norm (A$_{ixy}$)] / mean [norm (A$_{ixy}$)]

(b) Std [norm (B$_i$)] / mean [norm (B$_i$)]
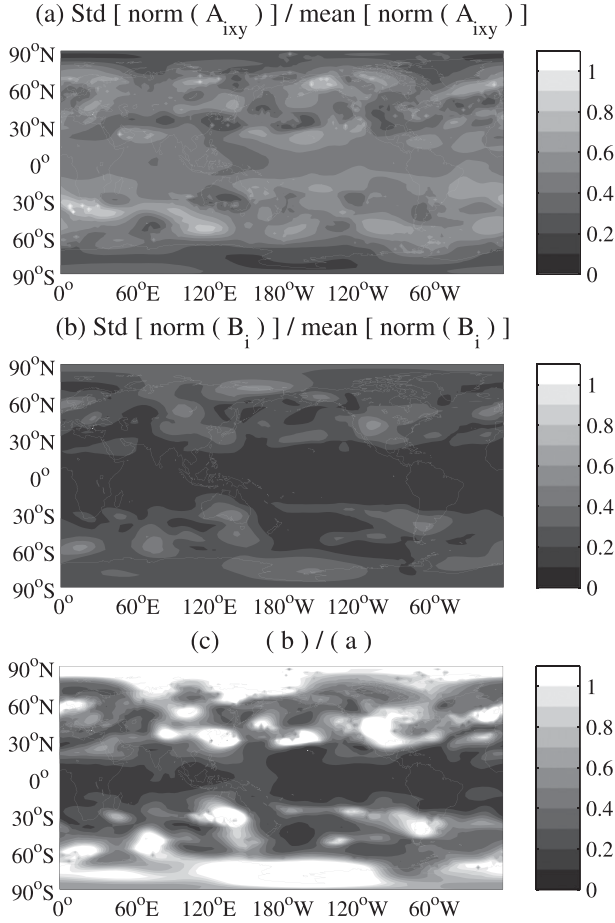

(c)   (b) / (a)

FIG. B4. Tau test in pentad LIM for the norm of the forcing vector to Z200 for each grid point over the globe in the observations. The relative standard deviation (a) $\sigma[m_{iA}(\tau)]/\overline{m}_{iA}(\tau)$ and (b) $\sigma[m_{iB}(\tau)]/\overline{m}_{iB}(\tau)$ and (c) the ratio of (b) to (a).

APPENDIX C

GEFA for High-Frequency Data

GEFA can also be applied to high-frequency data, such as daily and weekly data, as long as the dynamics can be approximately as a linear stochastic system. This can be understood from the single point coupled atmosphere (x) and ocean (y) model. The atmospheric Eq. (1.1a) or (2.1a), for a specific t, can be rewritten by substituting t with $t + \tau$ as

$$\frac{dx(t + \tau)}{d\tau} = -ax(t + \tau) + by(t + \tau) + n(t + \tau).$$

Multiply the equation by $y(t)$ and perform ensemble covariance with respect to t, and we have

$$\frac{d\langle x(t + \tau), y(t)\rangle}{d\tau} = -a\langle x(t + \tau), y(t)\rangle + b\langle y(t + \tau), y(t)\rangle$$
$$+ \langle n(t + \tau), y(t)\rangle.$$

Notice, for example,

$$\langle x(t + \tau), y(t)\rangle = \langle x(t), y(t - \tau)\rangle \equiv C_{xy}(\tau),$$

and we have the covariance equation

$$\frac{dC_{xy}(\tau)}{d\tau} = -aC_{xy}(\tau) + bC_{yy}(\tau) + C_{ny}(\tau). \quad \text{(C.1)}$$

Because the oceanic response time is very slow, the atmospheric response time can be approximated as $\tau_{atm} \sim 1/a$. For lags much longer than the atmospheric response time $\tau^* \gg \tau_{atm}$, the ratio between the first two terms is

$$\frac{\dfrac{dC_{xy}(\tau)}{d\tau}}{aC_{xy}(\tau)} \sim \frac{\tau_{atm}}{\tau^*} \ll 1. \quad \text{(C.2)}$$

Local variability is negligible and the covariance equation degenerates to similar to the quasi-equilibrium dynamics (1.2),

$$0 \approx -aC_{xy}(\tau) + bC_{yy}(\tau) + C_{ny}(\tau). \quad \text{(C.3)}$$

Therefore, the feedback parameter can be estimated the same as using monthly data from the quasi-equilibrium response.

Relative to GEFA on high-frequency data, the monthly GEFA is more economical because it only requires monthly data. However, monthly GEFA tends to exhibits a greater sensitivity with lag with the amplitude of the feedback parameter increasing rapidly with lags, making it difficult to perform tau test on the feedback matrix **B**. This larger lag sensitivity for monthly **B**, however, is not caused by the failure of GEFA itself; instead, it is caused by the decorrelation of SST forcing after 2–3 months, which leads to an increased sampling error (Liu et al. 2006). With daily or pentad GEFA, however, a range of lags can be tested within 2–3 months before a significant increase of sampling error. For example, for our observational assessment in Fig. 7, pentad GEFA (not shown) gives almost the same response as monthly GEFA (pattern correlation of ~0.99 and amplitude ratio of ~1.0) but now performs well in tau test. Indeed, the lag sensitivity of the feedback matrix **B** from pentad GEFA is almost the same as that for pentad LIM (Figs. B3b, B4b) (not shown), giving us more confidence on the GEFA estimation.

REFERENCES

Alexander, M. A., L. Matrosova, C. Penland, J. D. Scott, and P. Chang, 2008: Forecasting Pacific SSTs: Linear inverse model predictions of the PDO. *J. Climate,* **21,** 385–402.

Barsugli, J., and D. Battisti, 1998: The basic effects of atmosphere–ocean thermal coupling on midlatitude variability. *J. Atmos. Sci.,* **55,** 477–493.

Bell, T. L., 1980: Climate sensitivity from fluctuation dissipation: Some simple model tests. *J. Atmos. Sci.,* **37,** 1700–1707.

Bretherton, C. S., C. S. Smith, and J. M. Wallace, 1992: An intercomparison of methods for finding coupled patterns in climate data. *J. Climate,* **5,** 541–560.

Cionni, I., G. Visconti, and F. Sassi, 2004: Fluctuation dissipation theorem in a general circulation model. *Geophys. Res. Lett.,* **31,** L09206, doi:10.1029/2004GL019739.

Czaja, A., and C. Frankignoul, 2002: Observed impact of North Atlantic SST anomalies on the North Atlantic Oscillation. *J. Climate,* **15,** 606–623.

Fan, L., Z. Liu, and Q. Liu, 2011: Robust GEFA assessment of climate feedback to SST EOF modes. *Adv. Atmos. Sci.,* **28,** 907–912, doi:10.1007/s00376-010-0081-5.

Ferreira, D., and C. Frankignoul, 2008: Transient atmospheric response to interactive SST anomalies. *J. Climate,* **21,** 576–583.

Frankignoul, C., 1999: A cautionary note on the use of statistical atmospheric models in the middle latitudes: Comments on "Decadal variability in the North Pacific as simulated by a hybrid coupled model." *J. Climate,* **12,** 1871–1872.

——, A. Czaja, and B. L'Heveder, 1998: Air–sea feedback in the North Atlantic and surface boundary conditions for ocean models. *J. Climate,* **11,** 2310–2324.

——, N. Chouaib, and Z. Liu, 2011: Estimating the observed atmospheric response to SST anomalies: Maximum covariance analysis, generalized equilibrium feedback assessment, and maximum response estimation. *J. Climate,* **24,** 2523–2539.

Gardiner, C. W., 1997: *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences.* Springer, 442 pp.

Gill, A. E., 1980: Some simple solutions for heat-induced tropical circulation. *Quart. J. Roy. Meteor. Soc.,* **106,** 447–462.

Gritsun, A., and G. Branstator, 2007: Climate response using a three-dimensional operator based on fluctuation–dissipation theorem. *J. Atmos. Sci.,* **64,** 2558–2572.

Honda, M., H. Nakamura, J. Ukita, I. Kousaka and K. Takeuchi, 2001: Interannual seesaw between the Aleutian and Icelandic lows. Part I: Seasonal dependence and life cycle. *J. Climate,* **14,** 1029–1042.

——, S. Yamane, and H. Nakamura, 2005: Impacts of the Aleutian–Icelandic low seesaw on surface climate during the twentieth century. *J. Climate,* **18,** 2793–2802.

Hoskins, B. J., and D. J. Karoly, 1981: The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.,* **38,** 1179–1196.

Kirk-Davidoff, D. B., 2009: On the diagnosis of climate sensitivity using observation of fluctuations. *Atmos. Chem. Phys.,* **9,** 813–822.

Kumar, A., and M. Hoerling, 1998: Annual cycle of Pacific–North American seasonal predictability associated with different phases of ENSO. *J. Climate,* **11,** 3295–3308.

Kushnir, Y., W. A. Robinson, I. Blade, N. M. J. Hall, S. Peng, and R. Sutton, 2002: Atmospheric GCM response to extratropical SST anomalies: Synthesis and evaluation. *J. Climate,* **15,** 2233–2256.

Leith, C. E., 1975: Climate response and fluctuation dissipation. *J. Atmos. Sci.,* **32,** 2022–2026.

Liu, Z., and L. Wu, 2004: Atmospheric response to North Pacific SST: The role of ocean–atmosphere coupling. *J. Climate,* **17,** 1859–1882.

——, and M. Alexander, 2007: Atmospheric bridge, oceanic tunnel, and global climatic teleconnections. *Rev. Geophys.,* **45,** RG2005, doi:10.1029/2005RG000172.

——, and N. Wen, 2008: On the assessment of nonlocal climate feedback. Part II: EFA-SVD and optimal feedback modes. *J. Climate,* **21,** 5402–5416.

——, M. Notaro, J. Kutzbach, and N. Liu, 2006: Assessing global vegetation–climate feedbacks from observations. *J. Climate,* **19,** 787–814.

——, N. Wen, and Y. Liu, 2008: On the assessment of nonlocal climate feedback. Part I: The generalized equilibrium feedback analysis. *J. Climate,* **21,** 134–148.

Newman, M., P. D. Sardeshmukh, and C. Penland, 2009: How important is air–sea coupling in ENSO and MJO evolution? *J. Climate,* **22,** 2958–2977.

Notaro, M., Z. Liu, and J. Williams, 2006: Assessing climate–vegetation feedback in the United States. *J. Climate,* **19,** 763–786.

Peng, S., and J. S. Whitaker, 1999: Mechanisms determining the atmospheric response to midlatitude SST anomalies. *J. Climate,* **12,** 1393–1408.

——, W. A. Robinson, and M. P. Hoerling, 1997: The modeled atmospheric response to midlatitude SST anomalies and its dependence on background circulation states. *J. Climate,* **10,** 971–987.

Penland, C., 1989: Random forcing and forecasting using principal oscillation pattern analysis. *Mon. Wea. Rev.,* **117,** 2165–2185.

——, and T. Magorian, 1993: Prediction of Niño-3 sea surface temperatures using linear inverse modeling. *J. Climate,* **6,** 1067–1076.

——, and P. D. Sardeshmukh, 1995a: Error and sensitivity analysis of geophysical eigensystems. *J. Climate,* **8,** 1988–1998.

——, and ——, 1995b: The optimal growth of tropical sea surface temperature anomalies. *J. Climate,* **8,** 1999–2024.

Wen, N., Z. Liu, Q. Y. Liu, and C. Frankignoul, 2010: Observed atmospheric responses to global SST variability modes: A unified assessment using GEFA. *J. Climate,* **23,** 1739–1759.