Constructing the linear response function of SST using fluctuation-dissipation theorem and Green's function approach

Fukai Liu¹, Jian Lu², Oluwayemi Garuba, L. Ruby Leung², Yiyong Luo¹, and Xiuquan Wan¹

¹Physical Oceanography Laboratory/CIMST, Ocean University of China and Qingdao National Laboratory for Marine Science and Technology, Qingdao, China
²Atmospheric Sciences & Global Change Division, Pacific Northwest National Laboratory, WA, USA

Corresponding author: Jian Lu, 902 Battelle Boulevard, P.O. Box 999, MSIN K9-30, Richland, WA 99354. E-mail: Jian.Lu@pnnl.gov
Abstract

This paper explores the use of linear response function (LRF) to relate the mean sea surface temperature (SST) response to prescribed ocean heat convergence (q-flux) forcings. Two methods for constructing the LRF based on the fluctuation-dissipation theorem (FDT) and Green’s function (GRF) are examined. A 900-year preindustrial simulation from the Community Earth System Model with a slab ocean (CESM-SOM) is used to estimate the LRF using FDT. For GRF, 106 pairs of CESM-SOM simulations with warm and cold q-flux patches are performed. FDT is found to have skill in estimating the SST response to a q-flux forcing when the local SST response is strong, but it fails in inverse estimation of the q-flux forcing for a given SST pattern. In contrast, GRF is shown to be reasonably accurate in estimating both SST response and q-flux forcing. Possible degradation in FDT may be attributed to insufficient data sampling, significant departures of the SST data from Gaussian, and the non-normality of the constructed operator.

The accurately estimated GRF-based LRF is used to (i) generate a global surface temperature sensitivity map that shows the q-flux forcing in higher latitudes to be three to four times more effective than in low latitudes in producing global surface warming; (ii) identify the most excitable SST mode (neutral vector) resembling Interdecadal Pacific Oscillation; and (iii) estimate a time-invariant q-flux forcing needed for maintaining the GHG-induced SST warming pattern. The GRF experiments will be used to construct LRF for other variables to further explore climate sensitivities and feedbacks.
1. Introduction

Despite the growing interests in understanding the global distribution of the SST response to the forcing of increasing greenhouse gases (Xie et al. 2010; DiNezio et al. 2009, Lu and Zhao. 2012, Luo et al. 2015; Liu et al. 2017), progress has been hampered by the so-called surface energy budget fallacy. As noted by Pierrehumbert (2010), ‘if one is in a regime where the surface fluxes tightly couple the surface temperature to the overlying air temperature as is in the low-frequency SST evolution, there is no need to explicitly consider the surface balance in determining how much the surface warms or cools’. The surface energy budget terms and the oceanic convergence/divergence always conspire to ensure that the SST is approximately in equilibrium. As SST tendency is hardly detectable, it is even more difficulty to argue causation based on which energy budget term contributes the most to the tendency. A more informative approach to understanding the SST response to forcing is through linear response function (LRF). Considering one key feature of the coupled ocean-atmosphere system that the low-frequency (decadal and longer) variations of SST and the intrinsic atmospheric time scales of days to weeks are well separated, one can have a significant simplification in the treatment of the slow SST response or variations (see Palmer 1999, Winton et al. 2010, Liu et al. 2012): the SST \( \delta T \) can be thought of as in equilibrium resulting from the constant balance between the oceanic convergence/divergence of heat \( \delta f \); hereafter referred to q-flux) and the surface feedbacks:

\[
L\delta T = -\delta f, \quad (1)
\]

where the surface energy flux feedbacks are expressed to be linearly proportional to the SST perturbation, which is assumed small. Note that matrix \( L \) is the LRF that represents all the relevant physical processes such as evaporative damping effect (Knutson and Manabe 1995), wind-evaporation-SST feedback mechanism (e.g., Xie and Philander 1994), Planck feedback, and cloud-radiation feedbacks (e.g., Ramanathan and Collins 1991), etc. Since \( L \) is a function of
the mean climate under small forcing, one can obtain the response to a given q-flux without resorting to energy budget analysis. Note also that the atmospheric dynamics and feedbacks can spread the heat from a local ocean q-flux via teleconnection mechanisms, so a local \( \delta T \) response is the product of the q-flux forcing everywhere, and the mutual relationship between \( \delta T \) and the ocean q-flux at low enough frequency can be all encapsulated in the matrix \( \mathbf{L} \). Constructing the LRF of the SST to ocean q-flux is the central goal of the current study. Through this LRF, we will be able to construct the global SST response to a given arbitrary oceanic heat convergence/divergence without running expensive model experiments; meanwhile we can pinpoint the q-flux that is responsible for a given SST response pattern of interest (such as the El Nino-like warming in tropical Pacific in response to a globally uniform increase of CO\(_2\)).

What equally motivates this study is the urge to construct the sensitivity map of any given climate index/phenomenon of interest to the q-flux in the same spirit as Barsugli and Sardeshmukh (2006, BS06 hereafter). Through the Green’s function (GRF) approach, BS06 identified the areas in the tropical Indo-Pacific where the SST anomalies can most effectively drive the global mean surface warming and global increase of precipitation. Here we follow suit from BS06 except that GRF is applied to q-flux in place of SST. This is also partly motivated by the latest proposition that the ocean heat uptake (represented by negative q-flux) in the Southern Ocean and North Atlantic is key in delaying global surface temperature warming under a sudden increase of CO\(_2\), owing to the greater global cooling efficacy of the Southern Ocean heat uptake than other oceans (e.g., Winton et al. 2010; Armour et al. 2013; Rose et al. 2016). It is another goal of the current study to investigate the global distribution of the ocean heat uptake efficacy---a goal achievable via the LRF of surface temperature to q-flux.

There are at least two ways to construct LRF: Fluctuation-Dissipation Theorem (FDT) approach and GRF approach, with their corresponding LRFs denoted as \( \mathbf{L}_{\text{FDT}} \) and \( \mathbf{L}_{\text{GRF}} \).
respectively. FDT states that by observing the statistical characteristics of the unperturbed system a LRF can be estimated that gives the response of the quasi-Gaussian system to weak external forcing. It has a relatively long history of use as a powerful tool in the climate science research, beginning with the seminal work of Leith (1975), followed among others by Bell (1980), North et al. (1993), Gritsun and Branstator (2007), Majda et al. (2010), Liu et al. (2012), Lutsko et al. (2015), Fuchs et al. (2015) and Hassanzadeh and Kuang (2016b). Some recent studies have further proposed a time periodic (seasonal) operator with more accuracy based on larger data sample (e.g., Gritsun 2010, Fuchs and Sherwood 2016). Inspired by the above studies focusing mostly on mean atmospheric responses to momentum or thermal forcings in atmospheric general circulation models, here we apply FDT to develop the LRF relating SST to ocean q-flux.

A more straightforward approach to construct the LRF is the GRF approach (e.g., Branstator 1985, Barsuguli and Sardeshmukh 2002, BS06, Kuang 2010, Hassanzadeh and Kuang 2016a). In this paper, we perform a large set of simulations with a state-of-the-art atmospheric general circulation model coupled to a slab ocean model (AGCM-SOM) using an array of localized q-flux anomaly patches that cover most of the global ocean surface (Fig. 1). Thus, the SST response to an individual q-flux patch can be considered as akin to a Green’s function for the q-flux forcing in that location, and the SST responses to all imposed q-flux patches can then be used to construct $L_{GRF}$.

With sufficient data samples from long simulations, the two methods should lead to the same result in a purely Gaussian system. However, they could depart from each other significantly due to i) departure from Gaussianity of the SST statistics (Loikith and Neelin 2015, Berner and Branstator 2007); ii) sampling errors due to limited data length. Thus, this paper also attempts to systematically evaluate the LRFs calculated from FDT and GRF, and explore the
possible sources of errors and the repercussions in their applications.

The rest of the paper is structured as follows. Section 2 describes a long control slab model simulation for constructing $\mathbf{L}_{\text{FDT}}$ and the design of the q-flux patches for the estimate of $\mathbf{L}_{\text{GRF}}$, as well as a test case to assess the accuracy of the LRFs. Section 3 and 4 present the detailed procedure to construct and to assess the two LRFs. Section 5 discusses the sources of the inaccuracy in the LRFs, especially in $\mathbf{L}_{\text{FDT}}$. In section 6 we demonstrate a few applications of the estimated LRF, including i) creating sensitivity map of global SST response and the related feedbacks, ii) calculating the neutral vector of SST variations, and iii) calculating the q-flux forcing underpinning the greenhouse gas (GHG)-induced warming pattern. Finally, a summary of our findings is given in Section 7.

2. Model and Experiment

The model we use is Community Earth System Model version 1.1 with a slab ocean model (CESM-SOM) in which the active Community Atmospheric Model version 5 (CAM5), the Community Land Model version 4 (CLM4), and the Community Ice CodE (CICE) are coupled to a SOM. The horizontal resolution of CAM5 and CLM4 is 2.5° longitude x 1.9° latitude, with the atmospheric component having 30 vertical levels. The horizontal resolution of the CICE and SOM is at a nominal 1°, telescoped meridionally to ~0.3° at the equator. In this model, the ocean and atmosphere are only thermodynamically coupled and SST is computed from surface heat flux and q-flux that accounts for the missing ocean dynamics.

The construction of $\mathbf{L}_{\text{FDT}}$ requires a long unperturbed simulation, so we make use of a long, 900-year CESM1.1-SOM control simulation (CTRL) from NCAR, which is forced with preindustrial carbon dioxide levels and solar insolation. Both the mixed layer depth and the q-flux used in CTRL are derived from the climatology of a fully coupled CESM1.1 control
The creation of $L_{GRF}$ needs a large ensemble of simulations. Therefore, starting from January 1st of an arbitrary year from the equilibrated CTRL run, we perform 106 pairs of “warm patch” and “cold patch” simulations, in which the $q$-flux patch is added to or subtracted from the slab ocean. The locations of the 106 $q$-flux patches are illustrated in Fig. 1. Following Barsugli and Sardesmukh (2002), the $q$-flux patch is designed as a localized cosine hump

$$ Q \cos^2 \left( \frac{\pi}{2} \frac{\phi - \phi_k}{\phi_w} \right) \cos^2 \left( \frac{\pi}{2} \frac{\lambda - \lambda_k}{\lambda_w} \right), \quad (2) $$

where $Q$ is the maximum $q$-flux anomaly of $\pm 12 \text{ Wm}^{-2}$ at the center of the rectangular patch $(\phi_k \pm \phi_w, \lambda_k \pm \lambda_w)$, $\phi_w = 30^\circ$ and $\lambda_w \approx 12^\circ$ are half-widths of the rectangle in zonal and meridional directions respectively. Note what are contoured in Fig. 1 are only half of the $q$-flux anomaly maximum ($Q/2$), and adjacent patches actually overlap. The sum of the $q$-flux from all the patches amounts to a uniform heating (cooling) of $12 \text{ Wm}^{-2}$ over the global ocean except near the edges of the oceans. Given that approximately 20 years are needed for the CESM1.1-SOM to approach equilibrium, each of the patch experiments is integrated for 40 years, and only the output of the last 20 years is averaged and compared to that from the CTRL to give the anomalous SST response. The linear and nonlinear parts of the SST anomalies can be estimated as $T_l = (T_+ - T_-)/2$ and $T_n = (T_+ + T_-)/2$, where the subscript $+(-)$ denotes the response in $+12 (-12)$ $\text{ Wm}^{-2}$ simulation.

It should be noted that the choice of the forcing amplitude of $12 \text{ Wm}^{-2}$ and the simulation length of 40 years is a tradeoff between the requirement for generating a robust response signal (which entails long simulations and large forcing perturbation) and the need to stay within the linear regime (which entails long simulations and small forcing perturbation) within the computational affordability. As it turns out, the averaged magnitude of local SST
response at each patch is in the order of 0.5°C (Fig. 5c), which is comparable to BS06, who used SST patches with a mean of 0.667°C.

To test the accuracy of the constructed $L_{\text{FDT}}$ and $L_{\text{GRF}}$, we further perform a pair of AGCM-SOM experiments (denoted as CPX) forced by positive/negative q-flux perturbation with more complexity (Fig. 7a). The corresponding $T_1$ averaged over year 21 to 40 is shown in Fig. 7b. In addition, the patch experiments themselves can also serve as test cases for the LRFs.

3. Construction of the LRFs

3.1. FDT approach

In this paper, we follow previous studies (Gritsun and Branstator 2007, Lutsko et al. 2015, Fuchs et al. 2015, Hassanzadeh and Kuang 2016b) and use the most common quasi-Gaussian formulation of FDT to estimate $L_{\text{FDT}}$ as follows:

$$L_{\text{FDT}} = -\left[ \int_0^\infty C(\tau)C(0)^{-1}d\tau \right]^{-1}, \quad (3)$$

where $C$ represents covariance matrix and $\tau$ the time lag, $C(0)$ the autocovariance matrix. Note although we aim to estimate the equilibrium SST response, it is impractical to set the upper bound of the integral infinite, so we use a finite but sufficiently large upper boundary $N_{\text{Inf}}=120$ month to capture the time scale of the response. Further tests show that increasing $N_{\text{Inf}}$ would not significantly improve the performance of the FDT operator (not shown). Note also that the Gaussian approximation for SST may not be accurate in both observations (e.g., Loikith and Neelin 2015) and model simulations (e.g., Berner and Branstator 2007), and the departure from Gaussianity of the system may degrade the accuracy of the $L_{\text{FDT}}$.

For constructing $L_{\text{FDT}}$ for atmospheric variability, daily data might be needed as the dominant atmosphere modes manifest themselves mostly at intraseasonal time scales (e.g.,
Feldstein, 2000; Hoerling and Kumar 2003). On the other hand, the SST modes such as El Nino and Southern Oscillation (ENSO, Philander 1990), Interdecadal Pacific Oscillation (IPO; Zhang et al. 1997) and Atlantic Multidecadal Oscillation (AMO; Kerr 2000), are all characterized at interannual or interdecadal time scales, so monthly data are sufficient for the construction of SST $L_{FDT}$. To be sure, we repeat the construction using daily SST and the result show little difference from that using the monthly data, whereas the computing cost of the former approach is an order of magnitude greater. Thus, all the results reported here are based on monthly data.

3.2. Dimension Reduction

Although our model resolution is already relatively coarse, it still has over 8000 ocean grid points globally, so it is impractical to calculate the covariance matrix on the entire space of the system. As conventionally practiced (Gritsun and Branstator 2007, Lutsko et al. 2015, Fuchs et al. 2015, Hassanzadeh and Kuang 2016b), we project the raw SST data onto the Empirical Orthogonal Functions (EOFs) of the SST in the long CNTRL simulation, truncated based on the North criterion (North, 1982) for well-resolved EOFs. The resultant operator and LRF are all expressed based on the EOFs with reduced dimensions. In addition, dimension reduction can largely reduce the effective number of degrees of freedom $N_{eff}$ and reduce the likelihood that the auto-covariance matrix $C(0)$ becomes ill-conditioned (Martynov and Nechepurenko, 2004).

Some studies (e.g., Gritsun and Branstator 2007, Fuchs et al., 2015, 2016) simply used the first EOFs that explain a certain proportion (varies between 90% and 97.5%) of the total variance without a rigorous assessment of the number of well-resolved EOFs. Here, we adopt North criterion that demands the ratio of sampling error of a particular eigenvalue $\lambda_1$ to the spacing between neighboring eigenvalues to be less than 1:
\[ Cr = \frac{e^\lambda_i}{(\lambda_{i-1} - \lambda_{i+1})/2} < 1, \quad (4) \]

where \( e = \sqrt{1/N_{\text{eff}}} \) is the sampling error ratio. The effective degree of freedom is estimated following Trenberth (1984):

\[ N_{\text{eff}} = N \left[ \sum_{\tau=-N+1}^{N-1} (1 - \frac{1}{N}) \rho(\tau) \right]^{-1}, \quad (5) \]

where \( \rho(\tau) \) is the autocorrelation of each principal component (PC). According to Fig. 2, the first few EOFs easily satisfy the \( Cr < 1 \) condition with well-separated eigenvalues despite relatively small \( N_{\text{eff}} \). However, \( Cr \) increases gradually with EOF modes as the separation between the eigenvalues decreases much faster than the increase of \( N_{\text{eff}} \). A power law function fit (orange line in Fig. 2) to the raw \( Cr \) curve (black line in Fig. 2) helps estimate its intercept with the line \( Cr = 1 \) to identify the reasonably resolved EOFs. In this way, the first 62 EOFs, which explain \(~80\%\) of the total variance (Fig. 3d), are retained to construct the \( L_{FDT} \). Sensitivity tests using either more or fewer than 62 EOFs for the base of \( L_{FDT} \) (not shown) all result in deteriorate results (measured against the actual response to the q-flux patch experiments). All in all, both North’s criterion and empirical tests point to 62 EOFs as the optimal number for the dimension reduction.

The first three EOFs of the model’s monthly mean SST fields are shown in Fig. 3a-c. It should be noted that these EOFs are different from those derived in observational data (e.g., Messié and Chavez, 2011; Hartmann 2015) due to the lack of ocean dynamics in our AGCM-SOM. In the Pacific, all three leading modes appear to be dominated by amplitude away from the equator. The lack of strong SST variability in the equatorial Pacific (El Niño–Southern Oscillation) as observed is a common phenomenon in similar model configuration (e.g., Dommenget and Latif 2002, Clement et al. 2011, Dommenget 2010, Okumura 2013), suggesting that the SST variability in the equatorial Pacific are closely related to the interactive ocean...
dynamics, such as the Bjerknes feedback. Nevertheless, some interesting resemblance to the observed EOFs (though not in the same order) does emerge. The structure of the first mode is primarily the North Pacific Mode (Deser and Blackmon, 1995), and the third mode appears to be the South Pacific Meridional Mode (Zhang et al. 2014a) with more significant connections to the tropics. The second mode is more like a transition between the first and third modes. In the North Atlantic, the horseshoe patterns in all three leading modes bear some resemblance to the pattern of AMO.

3.3. Greens’ function approach

The mean SST response to an individual patch shown in Fig. 1 can be considered as akin to a GRF for forcing in that location. That is, the equilibrium SST response to any large-scale q-flux forcing pattern \( \mathbf{f} \) could be approximated as a linear combination of the responses to the local q-flux patches and can be expressed as

\[
\mathbf{T} = \mathbf{Gf}, \quad (6)
\]

where \( \mathbf{G} \) is a rectangular Green’s function matrix. Apparently, \( -\mathbf{G}^{-1} \) (referred to as \( \mathbf{G}_{\text{GRF}} \) hereafter) is the LRF derived from GRF approach. Following Barsugli and Sardesmukh (2002), each column vector \( \mathbf{G}_k \) can be calculated as

\[
\mathbf{G}_k = \frac{(\mathbf{T}_k)_{ij}}{\sum_j f_k(x_j)}, \quad (7)
\]

where \( k \), ranging from 1 to 106, is the index of the q-flux patches, \( \mathbf{T}_k \) is the linear parts of global SST response to the kth forcing, \( x_j \) represents the grids within the kth patch (see Barsugli and Sardesmukh 2002, Barsugli et al. 2006 for details). Since \( \mathbf{G}_{\text{GRF}} \) is not a square matrix and can only be calculated from pseudo-inversion, we project the SST response and q-flux forcing onto the selected EOFs to obtain a square matrix \( \mathbf{L}_{\text{GRF}} \). Employing this dimension reduction will
result in two major benefits. First, the comparison of performances of $L_{\text{GRF}}$ and $G_{\text{GRF}}$ will show if dimension reduction is a source of error in constructing LRF. Second, dimension reduction enable $L_{\text{GRF}}$ to have the same matrix dimension (62×62) as $L_{\text{FDT}}$, and their comparison can help reveal their different eigen properties.

3.4. Comparison of eigenvalues

Figure 4 shows the eigenvalues ($\lambda$) of $L_{\text{FDT}}$ and $L_{\text{GRF}}$, the decay rate (real part of $\lambda$) of each eigenmode is on the ordinate, and the frequency (imaginary part of $\lambda$) is on the abscissa. Note that the growing eigenmodes ($\lambda$ has positive real part) of LRF should be corrected as all eigenmodes are expected to decay in a steady system (Kuang 2010). For $L_{\text{FDT}}$, all the real parts of eigenvalues are negative and thus the eigenmodes are decaying. However, 8 of the 62 eigenmodes of $L_{\text{GRF}}$ are growing. We simply set the 8 eigenvalues with positive real part to zero and then reconstruct the matrix $L_{\text{GRF}}$. This results in only decaying modes in the two LRFs. Most of the eigenmodes in the two LRFs also have imaginary eigenvalues, indicating their oscillatory property. Furthermore, both the calculations of $L_{\text{FDT}}$ and $L_{\text{GRF}}$ involve the matrix inversion operation. According to Hassanzadeh and Kuang (2016a), the eigenvalues with the smallest (largest) magnitude are calculated with the highest (lowest) accuracy, as they are least (most) influenced by the errors in the matrix to be inverted. Figure 4 shows that the decay rates ($\lambda$) of eigenmodes of $L_{\text{FDT}}$ are much greater than those of $L_{\text{GRF}}$, and thus $L_{\text{FDT}}$ is more susceptible to the uncertainties in the estimate of eigenvalues and therefore less accurate. As shown later in the next section, $L_{\text{FDT}}$ indeed fails to estimate the q-flux forcing for a given SST response pattern via Eq. (1).

4. Validation of the two LRFs
In this section, we evaluate the performance of the two LRFs by testing (1) whether they can capture the true (modeled) time-mean SST response to q-flux forcing (referred to as forward problem), and (2) whether they can give accurate estimates for the q-flux forcing from a given SST response pattern of interest (referred to as inverse problem). Both the patch experiments and CPX experiment described in Section 2 will serve as the test cases.

4.1. Q-flux patches

We first evaluate the performance of the two LRFs in solving the forward problem to each localized q-flux patch. More specifically, the 106 q-flux patches shown in Fig. 1 are first projected onto the selected 62 EOFs, the SST responses are then directly calculated from $\mathbf{T} = -\mathbf{L}^{-1} \mathbf{f}$. The spatial pattern correlation between the true global SST response and the values linearly constructed through our LRFs is calculated for each q-flux patch. The pattern correlation maps shown in Fig. 5a and 5b serve as metrics of the skill for $\mathbf{L}_{FDT}$ and $\mathbf{L}_{GRF}$, respectively.

$\mathbf{L}_{FDT}$ has some skill in capturing the SST response to q-flux forcing over the equatorial Pacific, the Southern Ocean, the North Pacific and the North Atlantic, with the pattern correlations (with the modeled pattern) around 0.5. However, for q-flux forcings over the Atlantic Ocean, tropical Indian Ocean, as well as the banded regions in the Pacific around 20°S and 20°N, the predicted patterns exhibit weak and even negative correlations with the simulated counterparts. A further inspection of the correlation map and the local SST response (Fig. 5c) suggests that they are somehow related. In general, the weaker the local SST response, the lower the pattern correlation. Here we select two representative cases in the north tropical Indian Ocean (centered at 12°N, 60°E) and eastern equatorial Pacific (centered at 1°N, 90°W), regions with the lowest and highest pattern correlations respectively. In the former case, the constructed SST can hardly capture any remote true responses beyond the regions of forcing (comparing Fig. 6b with
There are myriad of reasons behind the case of failure. One may be related to the weak and locally confined SST response, as the q-flux fails to produce strong enough local convection anomalies and hence wave source for linear teleconnections, and thus the remote response is dominated by the nonlinear or chaotic processes (this can be seen from the very limited stippled regions in Fig. 6a and 6b). In contrast, the eastern Pacific case shows rather strong and broad warming near the q-flux perturbation, and most of the global ocean are featured with significant linear SST responses (denoted by stippling in Fig. 6c and 6d). As a result, $L_{FDT}$ shows sizable skill (pattern correlation is 0.69) in predicting the SST responses in the Pacific and Indian Ocean basins. Therefore, the accuracy of $L_{FDT}$ is severely degraded when the local SST response is weak and confined locally near the forcing and chaotic processes dominate remote true SST response. Similar spatial dependence has also been noticed in earlier studies using FDT operator to predict the true response (e.g. Gritsun and Branstator 2007, Fuchs et al. 2015, 2016).

Figure 5b shows that the skill of $L_{GRF}$ is much higher than $L_{FDT}$ in predicting the SST response to q-flux forcings. The pattern correlations are greater than 0.6 over most of the global ocean, and even in regions where $L_{FDT}$ fails to capture the sign of the true SST response. Part of the skill gain in $L_{GRF}$ may arise from retaining higher order ($3^{rd}$, $5^{th}$, …) nonlinear terms in its construction.

4.2. CPX test case

One might argue that the above comparison is not fair as $L_{GRF}$ is constructed with the patch experiments themselves and should work better if tested against the patch experiments. For a fairer test, we further make use of experiment CPX, which is forced by a more complex q-flux perturbation, to test the ability of the two LRFs in solving the response and inverse problems.

The target SST pattern for the forward problem is shown in Fig. 7b in response to the q-
flux shown in Fig. 7a, and the corresponding predictions constructed with \( \mathbf{L}_{\text{FDT}} \) and \( \mathbf{L}_{\text{GRF}} \) are presented in Fig. 7d and 7f, respectively. The \( \mathbf{L}_{\text{FDT}} \) prediction exhibits relatively low skill not only in predicting spatial pattern but also in magnitude of the target pattern. In particular, the magnitude of the IPO-like response in Pacific is severely overestimated. Worse yet, the AMO-like response and the Indian Ocean Dipole (IOD)-like response are both missing, and the pattern correlation is merely 0.42. On the other hand, the \( \mathbf{L}_{\text{GRF}} \) predictions are superior on all aspects in capturing both pattern and magnitude.

The application to inverse problem are also evaluated using the same test experiment. The constructed q-fluxes calculated via \( \mathbf{f} = -\mathbf{L}\mathbf{T} \) are compared with the actual q-flux shown in Fig. 7a. The \( \mathbf{L}_{\text{FDT}} \) completely fails (Fig. 7c), and this can be rationalized by the too large eigenvalues of \( \mathbf{L}_{\text{FDT}} \) compared to \( \mathbf{L}_{\text{GRF}} \), which tends to unrealistically amplify the uncertain eigenmodes in predicting the q-flux. By contrast, the prediction with \( \mathbf{L}_{\text{GRF}} \) can capture the major feature of the target q-flux, although with considerably weaker amplitude. Further analysis shows the underestimation of the amplitude to be due to the dimension reduction, as only 45% of the variance in the target q-flux forcing is retained after projecting onto the first 62 EOFs (Fig. 8).

Comparing Fig. 7e to Fig. 8, the \( \mathbf{L}_{\text{GRF}} \) prediction actually agrees reasonably well with the projected target q-flux. However, as discussed in section 3.2, we should not include as many EOFs as desired, because poorly resolved EOFs would severely impair the accuracy of the LRFs, as evidenced by the failed predictions using the \( \mathbf{L}_{\text{GRF}} \) constructed with 200 EOFs (supplementary Fig. S1).

5. Sources of the inaccuracy in the LRFs

We now attempt to examine more comprehensively the possible sources of the inaccuracy in the LRFs, especially in \( \mathbf{L}_{\text{FDT}} \). As we mentioned earlier, the desired forcing should
be large enough to generate a robust response signal yet small enough to stay within the linear regime. However, the forcing amplitude of 12 Wm$^{-2}$ used here can excite considerable nonlinear response, which cannot be captured by LRF (Fig. 6 and 7). The nonlinearity in the GRF experiments could be a source of error for both LRFs, but especially detrimental for $L_{FDT}$, as $L_{GRF}$ can at least capture the odd order nonlinear terms.

As discussed in Hassanzadeh and Kuang (2016b), uncertainties in choosing the upper bound of $N_{inf}$ in Eq. (3) and EOF basis for dimension reduction can both be a source of inaccuracy of $L_{FDT}$. The first factor is not likely an important one as we have tested different $N_{inf}$ and the result is insensitive as long as $N_{inf}$ is greater than 120 months (section 3.2). On the other hand, despite careful examinations of the EOF basis through the North criterion (section 3.1), the sampling size is still a limiting factor that determines the number of well-resolved EOFs, and this limits the spatial details (or the effective spatial degrees of freedom) the LRF can capture.

Another possible source of poor performance of $L_{FDT}$ is the non-Gaussian statistics in SST data, which may violate the assumption behind Eq. (3). Here we assess the Gaussianity of the state vectors by checking their skewness and kurtosis (White 1980; Sura and Sardeshmukh 2008). The skewness, which measures the asymmetry of a probability distribution function (PDF), is calculated as $\overline{T^3}/\sigma^3$, where overbar denotes the time mean, $T$ is the state vector that is the first 62 PCs of global SST. The kurtosis is defined as $\overline{T^4}/\sigma^4$, and it describes the peakiness of the PDF. To determine if the calculated skewness and kurtosis are significantly different from those obtained from a Gaussian distribution (0 for skewness and 3 for kurtosis), we follow Brooks and Carruthers (1953) and measure their standard errors as $e_S = \sqrt{6/N_{eff}}$ and $e_S = \sqrt{24/N_{eff}}$, respectively, where $N_{eff}$ can be estimated with Eq. (5). Note that ± two standard deviations are associated with the 95% confidence levels for a normal distribution. Shown in Fig.
9 is a scatterplot of kurtosis as a function of skewness for the first 62 leading PCs of global SST data. Overall, 12 of them (including the first two leading PCs) are significantly non-Gaussian, and the non-Gaussianity is more related to kurtosis than skewness: amongst all the 12 non-Gaussian PCs, only 2 PCs deviate from Gaussian due to large skewness. 

What further degrades the accuracy of $L_{gdt}$ is its non-normality. As Hassanzadeh and Kuang (2016b) pointed out, errors could be introduced by dimension reduction if the included and excluded EOFs are strongly coupled through a non-normal LRF. To illustrate the point, we project a linear system $\frac{dt}{dt} = L T + f$ onto $n$ EOF basis:

$$\frac{da}{dt} = [EOF^{-1}(EIG L EIG^{-1})EOF] a + f_v, \quad (8)$$

where $a = (a_1, a_2, ..., a_k)$ is a coefficient vector corresponding to the first $k$th EOFs, $A$ is a diagonal matrix containing the eigenvalues $\lambda$ of LRF, $EOF$ and $EIG$ are EOFs of SST and eigenvectors of LRF, respectively. $f_v$ is the projections of $f$ on the EOF basis. For a system governed by a normal LRF, $EOF$ and $EIG$ are identical, and the first term on the right hand side of Eq. (8) becomes $\Lambda a$ and the corresponding system becomes uncoupled, yielding $n$ uncoupled equations

$$\frac{da_n}{dt} = \lambda_n a_n + f_v, \quad (9)$$

If only the first $k$th EOFs are retained after dimension reduction, error can only be introduced by the omission of the projection of the forcing $f$ onto the excluded EOFs, but the first $k$th modes would not be affected due to the orthogonality.

On the other hand, if LRF is non-normal and hence $EOF$ is different from $EIG$, the matrix in the first term at the right hand side of Eq. (8) will not be diagonal, but have large off-diagonal elements $\beta_{ij}$. The corresponding equation for $a_n$ becomes
\[ \frac{d a_n}{dt} = \sum_{j=1}^{n} \beta_{nj} a_j + f_{en}, \quad (10) \]

As the consequence, all the EOF coefficients are coupled to one another, and when truncated to the first \( k \) th EOFs, the equation for the first \( k \) th coefficients become

\[ \frac{d \tilde{a}_n}{dt} = \sum_{j=1}^{k} \beta_{nj} \tilde{a}_j + f_{en} \quad (n \leq k). \quad (11) \]

Comparing (11) with (10), one can see that the accuracy of the first \( k \) th EOF coefficients is compromised by not accounting for the coupling with the higher ( \( j \geq k + 1 \) ) EOFs: \( \sum_{j=k+1}^{n} \beta_{nj} \tilde{a}_j \). Both \( \mathbf{L}_{\text{FDT}} \) and \( \mathbf{L}_{\text{GRE}} \) constructed in our paper show considerable non-normality as delineated by the off-diagonal elements in supplementary Fig. S2. To further verify this point, we follow Eq. (7) and employ the GRF method to construct \( \mathbf{G}_{\text{GRF}} \) on the model grid without dimension reduction, and the predicted results are shown in Figs. 7g and 7h. Significant improvements can be gained in both the forward and inverse problems compared to the \( \mathbf{L}_{\text{GRF}} \) prediction. For the forward problem, the predicted SST pattern shows improved pattern correlation in comparison to the result from \( \mathbf{L}_{\text{GRF}} \) (increasing from 0.69 to 0.89). For the inverse problem, the predicted q-flux captures better the magnitude of the target forcing.

6. Applications

In view of the moderate success of \( \mathbf{L}_{\text{GRF}} \) and \( \mathbf{G}_{\text{GRF}} \) in addressing both the forward and inverse problems in CESM1.1-SOM, we are ready to apply the LRF derived from the GRF approach to the following applications: (i) global surface temperature sensitivity; (ii) neutral vector for the most predictable SST mode; and (iii) retrieving the oceanic q-flux for the SST pattern under the forcing of quadrupling CO₂.

6.1 Global surface temperature sensitivity map
The sensitivity $S_k$ to q-flux of any scalar function of the model variables (e.g., temperature, precipitation, and top-of-the-atmosphere (TOA) radiative forcing) is a special case of Green’s function, and it can be used to evaluate the effectiveness of a q-flux forcing at different regions in exciting global changes. An immediate example is the global surface air temperature (TS) sensitivity map, which is calculated as the ratio of the global mean TS anomaly divided by the area-integrated q-flux anomaly over each patch

$$S_k = \frac{\langle T_k \rangle}{\sum f_k} = \langle G_k \rangle, \quad (12)$$

where $\langle \cdot \rangle$ represents global area-weighted mean. The equation for the sensitivity is very similar to Eq. (7) for the GRF matrix except that the numerator here represents the global mean TS anomalies in response to the kth q-flux forcing, and thus $S_k$ can be alternatively calculated as the global mean of $G_k$. Fig. 10a shows the sensitivity map $S_k$, with high sensitivity mainly concentrated in higher latitudes, especially in the southern hemisphere south of $30^\circ$ S. This is consistent with previous studies (Rose et al. 2014, Marshall et al. 2015) that highlight the role of heat uptake over Southern Ocean in affecting climate response to increased GHG concentrations. By contrast, the tropics overall features relatively low sensitivity values (roughly one-third of the sensitivity in higher latitudes), especially in the Indian Ocean and Atlantic Ocean. Barsugli et al. (2006) arrived at similar sensitivity pattern in the tropics by creating a global surface temperature sensitivity map to tropical SST forcing rather than q-flux. Our sensitivity map shows that the q-flux forcing in higher latitudes are three to four times more effective in causing global warming or cooling compared to the tropics. To understand this meridional structure in the sensitivity map, we decompose the net TOA radiative forcing into longwave (LW) and shortwave (SW) components from clear and cloudy sky (all defined as positive downward) and composite them for the q-flux at each latitudinal band. The resultant sensitivities of these fluxes per unit (PW) q-
flux from each latitude are shown in Figure 10b, from which one can see that the enhanced TS sensitivity (dashed black line) to high-latitude q-flux stems mostly from cloud shortwave feedback (magenta line) and clear sky shortwave feedback (red line). While the importance of cloud shortwave feedback in enhancing global TS sensitivity to high-latitude forcing is consistent with the earlier conclusion derived from idealized aquaplanet experiments (e.g., Rose et al. 2014; Rugenstein et al., 2016), the greater importance of the clear sky short wave feedback in the large TS sensitivity to northern hemisphere high-latitude q-flux seems to be unique for the sensitivity in more realistic climate. In-depth investigation into the feedback mechanisms behind the inter-hemispheric asymmetry is under way and the result will be reported elsewhere.

6.2 Neutral vector

Neutral vectors, obtained as the right singular vectors of the LRF with the smallest singular values, help reveal not only the leading patterns of the low-frequency internal variability, but also the least damped response to external forcing (Marshall and Molteni 1993, Goodman and Marshall 2002, 2003). Here we will focus the discussion on the first neutral vector of \( \mathbf{L}_{GRF} \) calculated using singular value decomposition (Fig. 11), as it has important bearing on the recent debate regarding the origin of the Pacific decadal variability (Clement et al. 2011).

The leading neutral vector of \( \mathbf{L}_{GRF} \) displays an IPO-like pattern in the Pacific basin, with horseshoe patterns in midlatitudes and ENSO-like pattern along the equator. Note that large SST anomalies in the tropical Pacific is not symmetric about the equator compared to the canonical IPO; this is common to the simulated SST variability in slab ocean AGCM (e.g., Clement et al. 2011, Okumara et al. 2013). In particular, the warm anomalies tend to be confined to the south of the intertropical convergence zone (ITCZ). Due to the northward placement of the ITCZ from the equator, only the internal atmospheric variability of southern hemisphere origin can reach the
equator and exert significant impact on equatorial eastern Pacific variability through surface turbulent heat flux (Liu and Xie 1994; Xie and Philander 1994; Okumara et al. 2013). As a result, the tropical Pacific tends to be more tightly coupled to the south Pacific than the north Pacific in slab ocean AGCM (e.g., Zhang et al. 2014a and 2014b).

In the North Atlantic basin, the neutral vector shows a horseshoe pattern resembling the AMO. However, since the ocean heat transport is prescribed here, this SST variability can only be atmospherically originated, lending some support to the notion of Clement et al. (2015) that AMO-like variability can be present in a slab ocean under stochastic atmospheric forcing.

Regressing the global SST against the leading neutral vector pattern yields a neutral vector time series, the wavelet analysis of which (Fig. 11b) exhibits significant low-frequency variability with periods around 60 years. While it might be coincidental that this matches the period of IPO and AMO found in observations (Steinman et al. 2015, Mann and Park, 1994), this provides an evidence that slab AGCM can produce the IPO-type multi-decadal variability in the absence of the ocean dynamical feedbacks. In addition, the dominant patterns of the wind stress and sea level pressure (SLP), obtained by regressing onto the neutral vector time series, also correspond well with their observed counterparts (Clement et al. 2011, Okumura 2013). Taken together, we argue that IPO is a dynamical mode intrinsic to the linear operator $L_{GRF}$, rather than just a product of statistical analysis, at least in the context of the modeling framework used here.

More interestingly, a striking similarity is found between our leading neutral vector and the most predictable mode of decadal internal variability in Srivastava and DelSole (2017). Additionally, the 2nd and 4th neutral vector (supplementary Fig. S3), characterized by interhemispheric asymmetric pattern and ENSO pattern, closely resemble their 2nd and 3rd most predictable modes.
6.3 q-flux forcing for a specified SST response

A third application is a hypothetical one. Figure 12a shows the change in SST at the end of the 100 years in the coupled model experiment forced by an abrupt quadrupling of CO$_2$ with CESM1.1. As the model is still in adjustment, the surface energy flux is mostly used to warm the ocean interior and surface energy budget will not be informative with regard to the formation mechanisms of the SST pattern. If the spatial pattern of the SST response were to remain unchanged even as the climate approaches equilibrium, what would be the oceanic energy convergence/divergence maintaining this pattern? This is the hypothetical question we try to address with $\mathbf{G}_{GRF}$.

The desired q-flux can be computed through equation (1) with $\mathbf{G}_{GRF}$ substituting $\mathbf{L}$. The result is shown in Figure 12b, implicating what ocean dynamics would be doing in maintaining the SST pattern seen in Figure 12a. An interesting point to note is that, although q-flux forcing is responsible for the local SST response in some regions, e.g., the minimum warming in the Southern Ocean attributable to the local ocean heat uptake, the calculated q-flux overall shows rather little resemblance to the SST response, cautioning against the exercise of understanding the SST pattern via surface energy budget analysis. For example, the minimum SST warming in the tropical southeast Pacific and the subtropical north Atlantic corresponds to local maximum q-flux heating. As argued by Xie et al. (2010) and Leloup and Clement (2009), these local SST warming minima there are controlled by the evaporative cooling (or Newtonian cooling in the language of Xie et al. 2010) of the SST, having little to do with the ocean dynamical feedbacks. In the tropical Pacific, the heating confined to the east end of the basin is actually responsible for the El Nino-like SST warming to the west along the equator (also see Fig. 6c); this is consistent with Luo et al. (2017) who found that the ocean dynamical heating there is responsible for the El Nino-like pattern, owing to the gradual warming of the equatorial thermocline and the reduction
of the subtropical cells strength. In addition, the largest q-flux heating in the north Indian Ocean does not give rise to large local SST warming, consistent with the patch experiment showing that q-flux forcing here is ineffective in driving local SST change (Fig. 5c).

7. Summary

Motivated by the need to understand the role of the ocean dynamics in the climate response to GHG-induced warming, in this study, we constructed the linear response function of SST to the q-flux of an AGCM coupled to a slab ocean using both FDT and GRF approaches. The former is predicated on the hypothesis that the SST can be approximated as a Gaussian variable and that its response to external forcing behaves the same way as the variability arising from internal noise, while the latter is achieved through a large set of numerical simulations with the patches of q-flux specified over the open ocean surface everywhere. The LRFs from the two approaches both show some skill in capturing the SST response to prescribed q-fluxes, especially those from the equatorial Pacific and Southern Ocean. The LRF from FDT is in general inferior to that from GRF, manifesting the challenges in the quantitative understanding of the SST variability and response. In addition to the limited data length used for the construction, $L_{FDT}$ suffers also from the non-Gaussianity in the internal variability of the SST and the non-normality of the LRF itself. Despite sharing these two difficulties, the GRF method exhibits considerable skill in both predicting the SST response to a given q-flux (forward problem) and retrieving the q-flux for a target SST pattern (inverse problem).

The immediate application of the GRF is the sensitivity of the global mean surface temperature to the q-flux from different geographic locations. Consistent with an earlier study using idealized aquaplanet model and idealized q-flux (e.g., Rose et al. 2014), we find that high latitudes are three to four times more effective in driving global surface temperature change.
Somewhat differing from the previous result, we also find an interesting interhemispheric asymmetry in the feedbacks to the amplified global warming sensitivity to q-flux from high latitudes: the TOA clear sky shortwave feedback plays a more important role than the shortwave cloud feedback in response to the NH high latitude forcing, while the positive cloud feedback is the leading positive feedback for the enhanced global warming response to SH high latitude forcing. Further detailed analysis using a radiative kernel to dissect the specific feedbacks is underway.

Another application of the LRF is its neutral vector analysis. Neutral vectors reveal the least damped, and thus most excitable modes of variability in SST. The leading neutral vector turns out to bear close resemblance to the IPO in spatial structure and have a spectral peak around 60 years. It is also almost identical to the most predictable SST mode found by Srivastava and DelSole (2017) via average predictability time analysis, indicating the higher predictability of low-frequency internal variability. The result also lends some support to the notion that AGCM coupled to a slab ocean can develop SST variability with decadal predictability in the absence of ocean dynamical variability (Clement et al. 2011, 2015). However, we caution against equating this result with support for an atmospheric origin of the Pacific decadal variability in the observations.

There is also a modest success in the inverse application of the LRF, that is, inferring the ocean heat convergence for the SST response pattern to quadrupling CO$_2$. The q-flux pattern inferred here remains to be verified with future more careful analysis and improved resolution in the q-flux array. One important lesson learned from this exercise, though, is that a positive SST anomaly does not necessarily imply a heating to the overlying atmosphere. While this is only the first attempt in this study, with high enough resolution, this technique can be a powerful tool for linking the SST anomalies to the ocean processes underneath and for designing pace-maker
experiments with slab models (e.g. Cash et al. 2008).

Owing to the limited scope of this paper, only a few issues related to the SST LRF are examined here. As the GRF experiments simulate all the atmospheric, land and ice variables, one, in principle, can build LRF for any variable of interest. Research is underway to explore the sensitivities of ITCZ, jet location and strength, and Hadley cell width and intensity to the q-flux, so more results will be forthcoming to address a broader set of questions on climate sensitivities and feedbacks.

Acknowledgments

The CESM1.1-SOM warm and cold q-flux patch simulation output are archived at the National Energy Research Scientific Computing Center (NERSC) and can be accessed by contacting Jian Lu (Jian.Lu@pnnl.gov). This study is supported by the U.S. Department of Energy Office of Science Biological and Environmental Research (BER) as part of the Regional and Global Climate Modeling program. Computational resources for the warm and cold q-flux patch simulations were provided by NERSC. PNNL is operated for DOE by Battelle Memorial Institute under contract DE-AC05-75RL01830.


Gritsun, A., 2010: Construction of the response operators onto small external forcings for the
general circulation atmospheric models with time-periodic right hand sides. Izv. Atmos.

Hassanzadeh, P., and Z. Kuang, 2016a: The Linear Response Function of an Idealized
73, 3423–3439.

Hassanzadeh, P., Z. Kuang, P. Hassanzadeh, and Z. Kuang, 2016b: The Linear Response
Function of an Idealized Atmosphere. Part I: Construction Using Green’s Functions and


doi:10.1126/science.1079053.

1986.

Increased C0 2 in a Coupled Ocean-Atmosphere Model. J. Clim., 8, 2181–2199.

Kuang, Z., 2010: Linear Response Functions of a Cumulus Ensemble to Temperature and
Moisture Perturbations and Implications for the Dynamics of Convectively Coupled Waves.

2026.

Leloup, J., and A. Clement, 2009: Why is there a minimum in projected warming in the tropical


of transient climate sensitivity and radiative feedbacks on the spatial pattern of ocean heat

Rose, B. E. J., and L. Rayborn, 2016: The Effects of Ocean Heat Uptake on Transient Climate

Rugenstein, M. A. A., K. Caldeira, and R. Knutti, 2016: Dependence of global radiative

Ring, M. J., and R. A. Plumb, 2008: The Response of a Simplified GCM to Axisymmetric
3898.

Steinman, B. a, M. E. Mann, and S. K. Miller, 2015: Atlantic and Pacific multidecadal


Trenberth, K. E., 1984: Some Effects of Finite Sample Size and Persistence on Meteorological

White, G. H., 1980: Skewness, Kurtosis and Extreme Values of Northern Hemisphere


**Figure Captions**

**Fig. 1** Configuration of q-flux perturbation patches illustrated by the 6 Wm$^{-2}$ contours. Note that the size of the patch is actually larger than the contoured area. For each patch, q-flux anomalies are added to or subtracted from the climatology over the ocean grid points.

**Fig. 2** North’s criterion $C_r$ (black line) as a function of the mode number of EOFs, overlaid with an of power law function fit (orange line).

**Fig. 3** First three empirical orthogonal functions computed from monthly mean SST (a-c) and fractional variance of monthly mean SST explained by the EOFs (d).

**Fig. 4** The eigenvalues of $L_{FDT}$ (red) and $L_{GRF}$ (blue). The abscissa (ordinate) represents the real (imaginary) part of the eigenvalues.

**Fig. 5** Pattern correlations between the modeled and constructed SST responses with (a) $L_{FDT}$ and (b) $L_{GRF}$ for each q-flux patch. (c) shows the local SST response (unit: K) averaged over each patch.

**Fig. 6** (left) Modeled and (right) constructed SST response with $L_{FDT}$ to q-flux perturbation centered at (top) ($12^\circ$N, $60^\circ$E) and (bottom) ($0^\circ$, $150^\circ$E). In this and the following figures, the crosses denote the regions of linearity (measured by $|T_1|/|T_n| > 1$).

**Fig. 7** Top: q-flux perturbation used to force CPX (a) and its corresponding SST response (b); Second row: constructed q-flux perturbation and SST response with $L_{FDT}$ (c & d); Third row:
constructed q-flux perturbation and SST response with $L_{GRF}$ (e & f); Bottom: constructed q-flux perturbation and SST response with $G_{GRF}$ (g & h).

**Fig. 8** The result of transforming the q-flux perturbation in Fig 7a onto the first 62 EOFs of SST.

**Fig. 9** Scatterplot of kurtosis vs skewness for the first 62 leading PCs of global SST. The stars represent the PCs statistically different from a Gaussian distribution at 95% confidence level.

**Fig. 10** (a) Map of global surface temperature (TS) sensitivity (unit: K PW$^{-1}$). (b) The zonal mean of the global TS sensitivity (dashed black line, unit: K PW$^{-1}$) and the corresponding radiative feedbacks: net TOA radiative flux (solid black line) and SW clear-sky radiative flux (red line), SW cloud radiative flux (magenta line), LW clear sky flux (blue line), and LW cloud radiative flux (green line). The unit for the flux feedbacks is W m$^{-2}$ PW$^{-1}$. A spatial smoother derived from the ‘inpaintn’ algorithm (Garcia 2010) is employed here and in Fig. 12b to ensure that only the statistically robust features were retained.

**Fig. 11** (a) The 1st neutral vector (shading) computed as the right singular vector of $L_{GRF}$ with the smallest singular number, and rescaled to have an amplitude of 1K. The neutral vector time series is defined as the regression of SST against the neutral vector pattern. Overlaid also are the regression of wind stress (vectors; unit: 10$^{-2}$ N m$^{-2}$) and SLP (black contours at intervals of 20 Pa; zero contours are thickened) upon the neutral vector time series. (b) Wavelet power spectrum...
of the neutral vector time series \((\log_2 \text{scaled})\), with the black line indicating the cone of influence by the edge effects.

**Fig. 12** (a) Annual-mean SST changes in the 4xCO2 simulation with CESM1.1; (b) the qflux needed to maintain the SST response in (a), calculated via \(G_{\text{GRF}} \delta T\).

**Fig. S1** Predicted (left) q-flux forcing and (right) SST response with \(L_{\text{GRF}}\) constructed with 200 EOFs.

**Fig. S2** (Left) \(L_{\text{FDT}}\) and (right) \(L_{\text{GRF}}\) constructed on the same 62 EOFs.

**Fig. S3** The 2\(^{\text{nd}}\) (left) and 4\(^{\text{th}}\) (left) neutral vectors of \(L_{\text{GRF}}\), rescaled to be of amplitude of 1K.
Fig. 1 Configuration of q-flux perturbation patches illustrated by the 6 Wm⁻² contours. Note that the size of the patch is actually larger than the contoured area. For each patch, q-flux anomalies are added to or subtracted from the climatology over the ocean grid points.
**Fig. 2** North’s criterion $C_r$ (black line) as a function of the mode number of EOFs, overlaid with an of power law function fit (orange line).
Fig. 3  First three empirical orthogonal functions computed from monthly mean SST (a-c) and fractional variance of monthly mean SST explained by the EOFs (d).
Fig. 4 The eigenvalues of $L_{FDT}$ (red) and $L_{GRF}$ (blue). The abscissa (ordinate) represents the real (imaginary) part of the eigenvalues.
**Fig. 5** Pattern correlations between the modeled and constructed SST responses with (a) $\mathbf{L}_\text{FDT}$ and (b) $\mathbf{L}_\text{GRF}$ for each q-flux patch. (c) shows the local SST response (unit: K) averaged over each patch.
Fig. 6 (left) Modeled and (right) constructed SST response with $L_{FDT}$ to q-flux perturbation centered at (top) (12°N, 60°E) and (bottom) (0°, 150°E). In this and the following figures, the crosses denote the regions of linearity (measured by $|\mathbf{T}_1|/|\mathbf{T}_n| > 1$).
Fig. 7 Top: q-flux perturbation used to force CPX (a) and its corresponding SST response (b); Second row: constructed q-flux perturbation and SST response with $L_{\text{FDT}}$ (c & d); Third row: constructed q-flux perturbation and SST response with $L_{\text{GRF}}$ (e & f); Bottom: constructed q-flux perturbation and SST response with $G_{\text{GRF}}$ (g & h).
Fig. 8 The result of transforming the q-flux perturbation in Fig 7a onto the first 62 EOFs of SST.
Fig. 9 Scatterplot of kurtosis vs skewness for the first 62 leading PCs of global SST. The stars represent the PCs statistically different from a Gaussian distribution at 95% confidence level.
Fig. 10 (a) Map of global surface temperature (TS) sensitivity (unit: K PW$^{-1}$). (b) The zonal mean of the global TS sensitivity (dashed black line, unit: K PW$^{-1}$) and the corresponding radiative feedbacks: net TOA radiative flux (solid black line) and SW clear-sky radiative flux (red line), SW cloud radiative flux (magenta line), LW clear sky flux (blue line), and LW cloud radiative flux (green line). The unit for the flux feedbacks is W m$^{-2}$ PW$^{-1}$. A spatial smoother derived from the ‘inpaintn’ algorithm (Garcia 2010) is employed here and in Fig. 12b to ensure that only the statistically robust features were retained.
Fig. 11 (a) The 1st neutral vector (shading) computed as the right singular vector of $L_{GRF}$ with the smallest singular number, and rescaled to have an amplitude of 1K. The neutral vector time series is defined as the regression of SST against the neutral vector pattern. Overlaid also are the regression of wind stress (vectors; unit: $10^{-2} Nm^{-2}$) and SLP (black contours at intervals of 20 Pa; zero contours are thickened) upon the neutral vector time series. (b) Wavelet power spectrum of the neutral vector time series ($log_2$ scaled), with the black line indicating the cone of influence by the edge effects.
Fig. 12 (a) Annual-mean SST changes in the 4xCO2 simulation with CESM1.1; (b) the qflux needed to maintain the SST response in (a), calculated via $G_{RFF}\delta T$. 
Fig. S1 Predicted (left) q-flux forcing and (right) SST response with $L_{GRF}$ constructed with 200 EOFs.
Fig. S2 (Left) $L_{FDT}$ and (right) $L_{GRF}$ constructed on the same 62 EOFs.
Fig. S3 The 2\textsuperscript{nd} (left) and 4\textsuperscript{th} (left) neutral vectors of $L_{\text{GRF}}$, rescaled to be of amplitude of 1K.