Earth-Science Reviews 150 (2015) 201–218

Contents lists available at ScienceDirect

Earth-Science Reviews

journal homepage: www.elsevier.com/locate/earscirev

Theoretical foundation of cyclostationary EOF analysis for geophysical and climatic variables: Concepts and examples

Kwang-Yul Kim [a,*], Benjamin Hamlington [b], Hanna Na [c]

[a] School of Earth and Environmental Sciences, Seoul National University, Seoul, Republic of Korea
[b] Department of Ocean, Earth and Atmospheric Research, Old Dominion University, Norfolk, USA
[c] Faculty of Science, Hokkaido University, Japan

ARTICLE INFO

Article history:
Received 14 January 2015
Received in revised form 11 May 2015
Accepted 9 June 2015
Available online 29 July 2015

Keywords:
Space-time eigen analysis
Data analysis
Physical and dynamical modes

ABSTRACT

Natural variability is an essential component of observations of all geophysical and climate variables. In principal component analysis (PCA), also called empirical orthogonal function (EOF) analysis, a set of orthogonal eigenfunctions is found from a spatial covariance function. These empirical basis functions often lend useful insights into physical processes in the data and serve as a useful tool for developing statistical methods. The underlying assumption in PCA is the stationarity of the data analyzed; that is, the covariance function does not depend on the origin of time. The stationarity assumption is often not justifiable for geophysical and climate variables even after removing such cyclic components as the diurnal cycle or the annual cycle. As a result, physical and statistical inferences based on EOFs can be misleading.

Some geophysical and climatic variables exhibit periodically time-dependent covariance statistics. Such a dataset is said to be periodically correlated or cyclostationary. A proper recognition of the time-dependent response characteristics is vital in accurately extracting physically meaningful modes and their space–time evolutions from data. This also has important implications in finding physically consistent evolutions and teleconnection patterns and in spectral analysis of variability—important goals in many climate and geophysical studies. In this study, the conceptual foundation of cyclostationary EOF (CSEOF) analysis is examined as an alternative to regular EOF analysis or other eigenanalysis techniques based on the stationarity assumption. Comparative examples and illustrations are given to elucidate the conceptual difference between the CSEOF technique and other techniques and the entailing ramification in physical and statistical inferences based on computational eigenfunctions.

Contents

1. Introduction

Measurements of physical and climate variables show the presence of seemingly random fluctuations in addition to such deterministic components as the diurnal cycle and the annual cycle. This random

* Corresponding author at: School of Earth and Environmental Sciences, Seoul National University, Seoul 151-742, Republic of Korea.
E-mail address: kwang56@snu.ac.kr (K.-Y. Kim).

http://dx.doi.org/10.1016/j.earscirev.2015.06.003

component of variability is often referred to as natural variability as opposed to "forced variation" in the presence of external forcing to a physical system. Natural variability has been the focus of numerous studies not only because it is an essential concept in addressing climate and environmental changes due to such external forcing agents as greenhouse gasses (Jones et al., 1994; Andreae et al., 2005) or volcanic aerosols (Rampino and Self, 1992; Zielinski et al., 1996; Jones et al., 2005) but also because natural variability is important on its own accord (Luterbacher et al., 2004; Bengtsson et al., 2006; Swanson et al., 2009). Much national and international effort focuses on the study of natural variability in understanding the current and future climate changes (NAC, 1995; IPCC, 2007, 2013).

The stochastic component of variability is often treated as a random variable. This does not mean that the source of random fluctuation is truly and solely the random nature of our physical and climate systems; this rather reflects our incomplete comprehension of how the system works in its full detail. Such a notion is often reflected in our attempts to assign physical meaning to seemingly stochastic undulations of some physical and climate variables. It may be fair to say that the study of natural variability in many applications is best approached by treating it as a random variable. Thus, data analysis often deals with multivariate random variables in space and time.

A random variable cannot be described in a deterministic manner. It can only be described in terms of probability, $p(X)$, of a random variable X having values in an event space A, i.e., $X \in A$. The event space is a subset of sample space, Ω, which represents a collection of all possible values of X. A probability density function (pdf) is often awkward to use and is also difficult to estimate accurately from limited observational datasets. Thus, the moment statistics of random variables are computed instead in a simplified approach. The nth moment is defined as

$$E(X^n) = \int_\Omega x^n p(x)dx = \langle X^n \rangle, \tag{1}$$

where x is a particular realization of the random variable X, $E(\bullet)$ is the expectation, and $\langle \bullet \rangle$ represents ensemble averaging. Having all of the moment statistics of a random variable is equivalent to knowing its pdf. The attractiveness of the moment statistics over the pdf lies in the alternative way of computing the former as suggested in Eq. (1)—namely, ensemble averaging without invoking a true pdf.

Typically, it is assumed that a random variable, $T(x)$, is reasonably described in terms of its first two moment statistics, i.e., the mean function and the covariance function:

$$\mu(x) = \int_\Omega T(x)p(T(x))dT = \langle T(x) \rangle, \tag{2}$$

$$C(x,x') = \int_\Omega T'(x)p\left(T'(x)\right)T'(x')p\left(T'(x')\right)dT' = \left\langle T'(x)T'(x') \right\rangle, \tag{3}$$

where $T'(x) = T(x) - \mu(x)$ and independent variable x represents time, space, or both. According to the independent variable, the covariance function, $C(x, x')$, may be called a temporal, spatial or spatio-temporal covariance function. Thus, the computation and analysis of the first two moment statistics of a given dataset constitute an essential step of analyzing random variables.

Computation of the two moment statistics in Eqs. (2) and (3) may look simple, but it actually is difficult because we do not know the pdf nor do we usually have enough realizations for ensemble averaging. In geophysical studies, we typically have only one realization (one observational record) for a physical variable of interest because we can carry out only one experiment (one Earth). Modeling studies, of course, can provide us with as many realizations as we want. The veracity of model statistics, on the other hand, should be tested against the statistics of observational data. Thus, it is necessary to introduce a simplifying assumption to find the first two moment statistics of observed data.

An assumption we often introduce in analyzing a random variable is stationarity. The essence of this assumption is that the moment statistics of a random variable are independent of time. When this assumption applies to the first two moment statistics, such a random variable is called "weakly" stationary. What is implicit in Eqs. (2) and (3) is that $T(x)$ and $T(x')$ are different random variables since their statistics, in general, are different. Under the stationarity assumption, however, the two random variables at two different times x and x' have the same statistics and, henceforth, are regarded the same. Thus, $T(x)$ and $T(x')$ may be viewed as two different realizations of the random variable T. As a result, the ensemble average in Eqs. (2) and (3) can be replaced by averaging in the time direction:

$$\mu = \langle T(t) \rangle = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} T(t), \tag{4}$$

$$C(t,t') = \left\langle T'(t)T'(t') \right\rangle = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} T'(t)T'(t+\tau) = R(\tau), \tag{5}$$

where $\tau = t - t'$ corresponds to lag and N is the number of samples. Here, the independent variable, time t, has been introduced explicitly. Note that $C(t, t')$ is no longer a function of time because of the stationarity assumption but depends only on lag τ. The function $R(\tau)$ is called the autocovariance function, meaning the covariance function of the same variable (self). In a similar manner, a spatial covariance function can be constructed as

$$C(x,x') = \langle T(x,t)T(x',t) \rangle = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} T(x,t)T(x',t), \tag{6}$$

where $T(x, t)$ hereafter denotes a zero-mean random variable and x is a spatial independent variable. One can, of course, extend the definition further to define a space–time covariance function:

$$C(x,t;x',t') = \langle T(x,t)T(x',t') \rangle = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} T(x,t)T(x',t+\tau). \tag{7}$$

Mainly because of the difficulty involved in the computation of Eq. (7), rarely seen in geophysical research is the analysis of a spatio-temporal covariance function.

While the stationarity assumption allows us to carry out the seemingly impossible calculations in Eqs. (2) and (3) from a limited dataset, statistics of many geophysical variables unfortunately do not exhibit the stationary behavior. Instead, many geophysical variables have statistics that are time dependent and periodic (see Section 6 in Gardner et al. (2006) and references therein). For example, the monthly variance of the NINO3 time series is strongly time dependent as shown in Fig. 1; NINO3 is a region in the equatorial Pacific defined by [150°–90°W, 5°S–5°N]. While there is no significant time-dependent component in the insolation statistics after removing the annual cycle, Fig. 1 shows that the variability of the equatorial climate system is stronger in winter

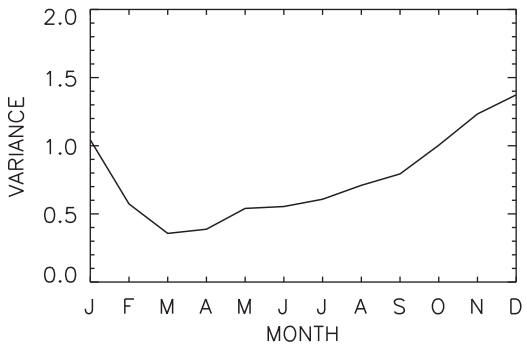

Fig. 1. Monthly variance of the NINO3 time series after removing the monthly mean values.

than in summer. This time dependence of statistics is a prevalent property of many geophysical variables. This raises two important concerns: (1) many statistical tools based on the stationarity assumption will not be accurate thereby leading to inaccurate statistical inferences; and (2) statistical tools should be developed that explicitly account for this time dependence. The second concern is difficult to address in that the development of general data analysis techniques for nonstationary space–time data would be nearly impossible.

In addition to the time dependence, statistics of many geophysical variables exhibit periodicity in time. Such a process is said to be cyclostationary or periodically correlated. The periodicity of statistics may primarily come from the periodic nature of external forcing and boundary conditions associated with our climate and geophysical systems although they may not necessarily be an exclusive source of periodicity in the statistics. With this added assumption of periodicity, the second concern raised above can be dealt with efficiently as will be shown later. It is this special class of data for which a more accurate concept of data analysis will be discussed.

It should be stressed that the periodic nature of our physical system manifests itself not only in the mean field, such as the annual cycle and the diurnal cycle, but also in higher-moment statistics. The periodic mean component of data does not pose any problem since it can be identified and removed prior to analysis. The periodic components in higher-moment statistics, however, are difficult to deal with. For simplicity, let us confine our discussion to the second-moment statistics.

The time dependence of variance statistics such as shown in Fig. 1 can be eliminated by using the so-called variance normalization or stabilization procedure. This procedure is summarized as

$$T_{norm}(t) = \frac{T(t)}{\sigma(t)}, \tag{8}$$

where raw data, $T(t)$, are divided by the respective standard deviation, $\sigma(t)$, at each time step t. Then, the variance of the resulting time series will be uniform in time. This procedure, however, does not alter the correlation structure between two different times. As shown in Fig. 2, lagged correlation function of the NINO3 time series remains to be time dependent even after the variance normalization procedure is applied. There is also an interesting asymmetry between positive and negative lags in the correlation structure. This asymmetry also varies in time. Thus, time dependence of covariance statistics should be dealt with explicitly; there is no circumventing this fundamental structure of the second-moment statistics.

The second-moment statistics are often recast in different forms including the spectral density function and EOFs. An EOF analysis represents decomposition of the second-moment statistics in space (spatial covariance function) into computational orthogonal functions, which will be discussed below in detail. The spectral density function is a Fourier transformation of the autocovariance function, and is conceptually equivalent to an EOF analysis in the time domain. This point will also be addressed in more detail in Section 6. The lack of stationarity in the temporal statistics of many geophysical variables means that statistical inferences made through the two very popular techniques—spectral analysis and EOF analysis—are subject to potentially serious faults. The discussion below elaborates on this point and, as an alternate method, CSEOF analysis is introduced. The conceptual foundation of the technique is explained with illustrations and examples in comparison with the EOF technique and also with the extended EOF and complex EOF techniques as needed. Computational algorithms for the CSEOF technique have been published elsewhere (Kim et al., 1996; Kim and North, 1997b) and will not be discussed in this paper.

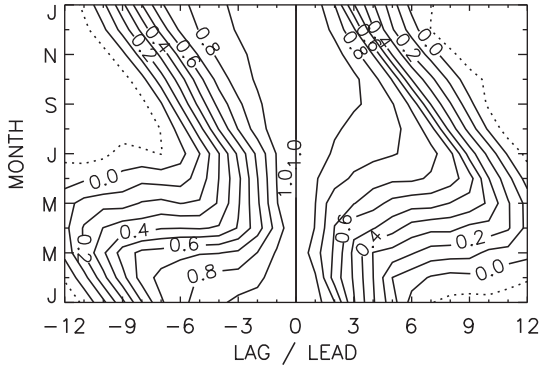

Fig. 2. Lagged correlation, $C(t,\tau) = \langle T(t)T(t+\tau)\rangle$, of the NINO3 time series, where the ordinate denotes t and the abscissa τ (month). A positive lag refers to a future time with respect to the reference time t. The NINO3 time series has been variance normalized by dividing each monthly value by the respective month's standard deviation after removing the monthly mean values.

2. Cyclostationary EOF analysis

The modal decomposition in terms of a set of basis functions is often useful in understanding the complicated response in a physical system. The complicated response is decomposed into less complicated basic patterns, which, in general, may be easier to understand and shed more insight into the nature of variability in a given physical system. Many theoretical basis functions have been studied extensively in the context of the Sturm–Liouville problem (Richtmyer, 1978; Arfken and Weber, 1995). Unfortunately, exact theoretical basis functions are very difficult to find, since our physical and climate systems are extremely complicated. Therefore, empirical basis functions (EOFs) are sought instead (Jolliffe, 2002; Hannachi et al., 2007). Finding computational basis functions are further complicated if the stationarity assumption is abandoned.

Let us consider a simple temporally discrete physical system defined by

$$T(x,t) = B(x,t)S(t), \tag{9}$$

where $B(x,t)$ is a physical process modulated by a stochastic time series $S(t)$. Then, it can be shown that

$$\begin{aligned}\mu(x,t) &= \langle T(x,t)\rangle \\ &= B(x,t)\langle S(t)\rangle = B(x,t)\mu_S,\end{aligned} \tag{10}$$

$$\begin{aligned}C(x,t;x',t') &= \langle T(x,t)T(x',t')\rangle \\ &= B(x,t)B(x',t')\langle S(t)S(t')\rangle = B(x,t)B(x',t')R_S(\tau),\end{aligned} \tag{11}$$

where μ_S and $R_S(\tau)$ are the mean and the autocovariance function of the purely stochastic component, $S(t)$, respectively. Thus, the first two moment statistics are shown to be time dependent in the presence of a time-dependent physical process, $B(x,t)$. Further, the time dependence of the statistics comes, in theory, only from the physical component not from the stochastic component of data in Eq. (9).

There are ample examples from geophysical observations suggesting that physical processes, and, henceforth, corresponding statistics, are time dependent. This can be seen, for example, in the covariance function, $C(t,t')$, of the NINO3 time series (Fig. 3). As shown, there is a prominent time dependence in the covariance statistics. Such a time-dependent structure cannot come from stationary random fluctuations. This should be contrasted with the stationarity assumption, under which

$$C(t,t') = R(t-t') = R(\tau). \tag{12}$$

In other words, covariance statistics of a stationary time series do not depend on time; they depend only on lag, τ. Note specifically in EOF analysis that physical response characteristics of a physical process are assumed to be stationary and not dependent on time in EOF analysis.

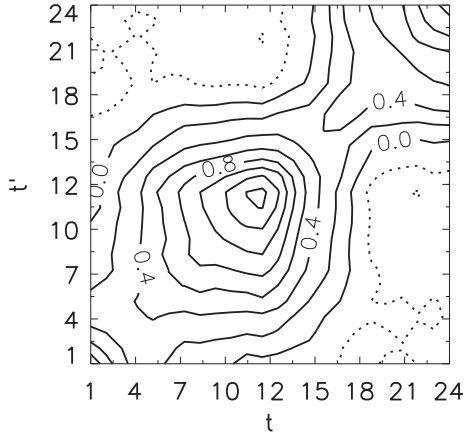

Fig. 3. Covariance function, $C(t, t')$, of the NINO3 sea surface temperature anomaly time series with the nested period of 24 months.

In fact, many geophysical observations indicate that the stationarity assumption is not really valid. A question arises, then, how these time-dependent response characteristics can properly be accounted for in dealing with the variability of a physical variable.

If the covariance function is time dependent, computational eigenfunctions, $B_n(x, t)$, can be defined as solutions of the Karhunen–Loève equation (Loève, 1978):

$$\int_D \int_T C(x,t;x',t')B_n(x',t')dt'dx' = \lambda_n B_n(x,t), \tag{13}$$

where D and T are space and time domains, respectively. As can easily be imagined, the solution of Eq. (13) is a computationally intensive procedure. For example, if the number of spatial points is 1000 and the length of record is 1000 at each station, then Eq. (13) is a matrix problem with the rank of one million—hopelessly too big to solve in practice. To make this problem tractable, a simplification should be introduced.

Let us make a further assumption in Eq. (9) that the response characteristics of the physical process is periodic in time, i.e.,

$$B(x,t) = B(x,t+d), \tag{14}$$

where d is a period. Note that this is an assumption although many observed physical processes often suggest such a periodicity. Then, the two moment statistics, with the aid of Eqs. (10) and (11), can be shown to be periodic:

$$\begin{aligned}\mu(x,t) &= \langle T(x,t)\rangle, \\ &= \langle T(x,t+d)\rangle = \mu(x,t+d),\end{aligned} \tag{15}$$

$$\begin{aligned}C(x,t;x',t') &= \langle T(x,t)T(x',t')\rangle \\ &= \langle T(x,t+d)T(x',t'+d)\rangle = C(x,t+d;x',t'+d).\end{aligned} \tag{16}$$

It should be noted that the stochastic component, $S(t)$ in Eq. (9), is assumed stationary. A process for which the two moment statistics satisfy Eqs. (15) and (16) is said to be (weakly) cyclostationary. Stationarity is a very special case of cyclostationarity, in which the physical period is $d = 1$.

With this added assumption, finding eigenfunctions as solutions of Eq. (13) can be made computationally tractable. Note that the eigenfunctions should also be periodic in time with the same period of the corresponding statistics leading to the terminology "cyclostationary" EOFs (Kim et al., 1996; Kim and North, 1997b). In a CSEOF analysis, space–time data are written as:

$$T(x,t) = \sum\nolimits_n B_n(x,t)T_n(t), \tag{17}$$

where $B_n(x, t) = B_n(x, t + d)$ are CSEOF loading vectors (CSLV) and $T_n(t)$ are corresponding principal component (PC) time series. Thus, each eigenfunction represents not one spatial pattern but multiple (d) spatial patterns, which repeat in time. Obviously, this is a more reasonable representation of a physical process, whose response characteristics vary in time periodically. It is crucial to realize that the temporal variation of data, as in Eq. (9), has two distinct sources: time-dependent physics, $B_n(x, t)$, and stochastic modulation of physical processes, $T_n(t)$. It is important to distinguish these two disparate sources of variability in order to make sound physical and statistical inferences. This issue will be addressed in detail in the following sections.

3. Nested periodicity

While the assumption of periodic statistics may be reasonable for many geophysical variables, it is difficult to prove the periodicity of statistics and identify the period. This so-called "nested period" is often determined based on a priori physical understanding of the physical process to be investigated. In many cases, however, there exists an obvious choice for the nested period because of the natural period in the Earth's physical and climate systems (e.g., Kim and Chung, 2001). Sometimes, however, the period of a physical process is not obvious mainly because of the lack of understanding of the underlying physical process. For example, Rasmusson et al. (1990) suggested that a significant power in the two-year period exists in the tropical Pacific sea surface temperature anomaly (SSTA) field. This biennial physical process is not obvious intuitively (Kim, 2002; Yeo and Kim, 2014). Yet, in another case, physical processes may not have any definite periods, to which CSEOF analysis may not be applicable.

We have already shown that in this model the nested period is identical with the period of a physical process in Eqs. (14–16). How is the nested period determined when multiple physical processes have different periods? Let us consider a dataset consisting of several physical processes

$$T(x,t) = \sum\nolimits_n B_n(x,t)S_n(t), \quad B_n(x,t) = B_n(x,t+d_n), \tag{18}$$

where d_n is the period of a physical process $B_n(x, t)$. If we assume that $S_n(t)$ are stationary time series, then

$$\begin{aligned}\mu(x,t) &= \langle T(x,t)\rangle = \sum\nolimits_n B_n(x,t)\langle S_n(t)\rangle \\ &= \sum\nolimits_n B_n(x,t+d)\langle S_n(t+d)\rangle = \langle T(x,t+d)\rangle = \mu(x,t+d),\end{aligned} \tag{19}$$

$$\begin{aligned}C(x,t;x',t') &= \langle T(x,t)T(x',t')\rangle = \sum\nolimits_n\sum\nolimits_m B_n(x,t)B_m(x',t')\langle S_n(t)S_m(t')\rangle \\ &= \sum\nolimits_n\sum\nolimits_m B_n(x,t+d)B_m(x',t'+d)\langle S_n(t+d)S_m(t'+d)\rangle \\ &= \langle T(x,t+d)T(x',t'+d)\rangle = C(x,t+d;x',t'+d),\end{aligned} \tag{20}$$

if d is given as the least common multiple (LCM) of $\{d_n\}$, i.e., $d = LCM(d_n)$. Note that

$$\langle S_n(t)S_m(t')\rangle = C_{nm}(t-t') = C_{nm}(\tau) \tag{21}$$

defines the cross-covariance function of two stationary time series, which is a function only of lag τ. Thus, the period of the first two moment statistics of a given dataset is the least common multiple of all physical periods in the dataset. If we further assume that $S_n(t)$ are independent of each other, i.e., uncorrelated at all lags, then

$$\begin{aligned}C(x,t;x',t') &= \langle T(x,t)T(x',t')\rangle = \sum\nolimits_n\sum\nolimits_m B_n(x,t)B_m(x',t')\langle S_n(t)S_m(t')\rangle \\ &= \sum\nolimits_n\sum\nolimits_m B_n(x,t)B_m(x',t')R_n(\tau)\delta_{nm} = \sum\nolimits_n B_n(x,t)B_n(x',t')R_n(\tau),\end{aligned} \tag{22}$$

a familiar expression for the covariance function. This expression clearly shows that the time dependence in the covariance function

comes only from the physical evolutions constituting the dataset and signifies the motivation of the CSEOF technique—physical decomposition of variability. Note that $R_n(\tau)$ in Eq. (22) does not contribute to the temporal covariance structure but only affects the magnitude of the covariance function.

The consequence of d being the LCM of all physical periods is that physical processes with a period less than d are shown to repeat in cyclostationary loading vectors, $B_n(r, t)$. For example, a physical process with a period of 6 months will repeat twice in the corresponding CSEOF loading vectors if d is set to 12 months. Although the nested period can be set to be an integral multiple of the LCM of all physical periods, the minimum period should be used; this is to minimize the contamination of covariance statistics by sampling errors. Typically, one has to work with a relatively short record so that the stationarity of the stochastic components in Eqs. (19) and (20) is difficult to establish. This contaminates the estimation of covariance statistics, which, in turn, distorts physical structures in CSEOF loading vectors.

4. A simple example of CSEOF analysis

Fig. 4 shows the first CSEOF with the nested period of 12 months of the tropical Pacific sea surface temperatures (SST). The PC time series indicates that the first CSEOF presents the annual cycle; the amplitude is always positive indicating that the loading patterns appear every year without any sign change. Unlike a conventional analysis such as composite analysis or harmonic analysis, the amplitude of the annual cycle is not a constant; it fluctuates interannually. The spatial loading patterns depict the evolution of the tropical Pacific sea surface temperature in the annual cycle. As addressed above, the loading patterns describe the evolution of a physical process called the annual cycle whereas the PC time series describes the "stochastic" amplitude variation of the annual cycle.

Fig. 5 shows the first two EOFs of the tropical Pacific SST. The first EOF resembles the March and September patterns and the second EOF the June and December patterns of the first CSEOF. Undoubtedly, the first two EOFs capture the characteristic features of the annual cycle. The corresponding time series are highly correlated (correlation = 0.75) at 3-month lag with the second PC time series leading the first PC time series. This is consistent with the interpretation of the spatial patterns in the context of the annual cycle.

Fig. 6 shows the loading vector of the second CSEOF mode of the tropical Pacific SST and represents El Niño; the loading patterns look similar to the second EOF pattern. Indeed, the two PC time series are highly correlated (correlation = 0.60), and the 1982/1983 and 1997/1998 El Niño events are captured in both the time series. This is clearly a serious problem, since the second EOF mode reflects both the seasonal cycle and the El Niño. An identical evolution pattern in two or more distinct physical processes is captured as a single EOF mode, and this leads to an ambiguous physical and statistical inference.

Fig. 7 shows the evolution of equatorial SST in the first CSEOF and that in the first two extended EOFs (EEOFs; see Appendix 1 for the definition of EEOFs) in comparison with the monthly composite (see

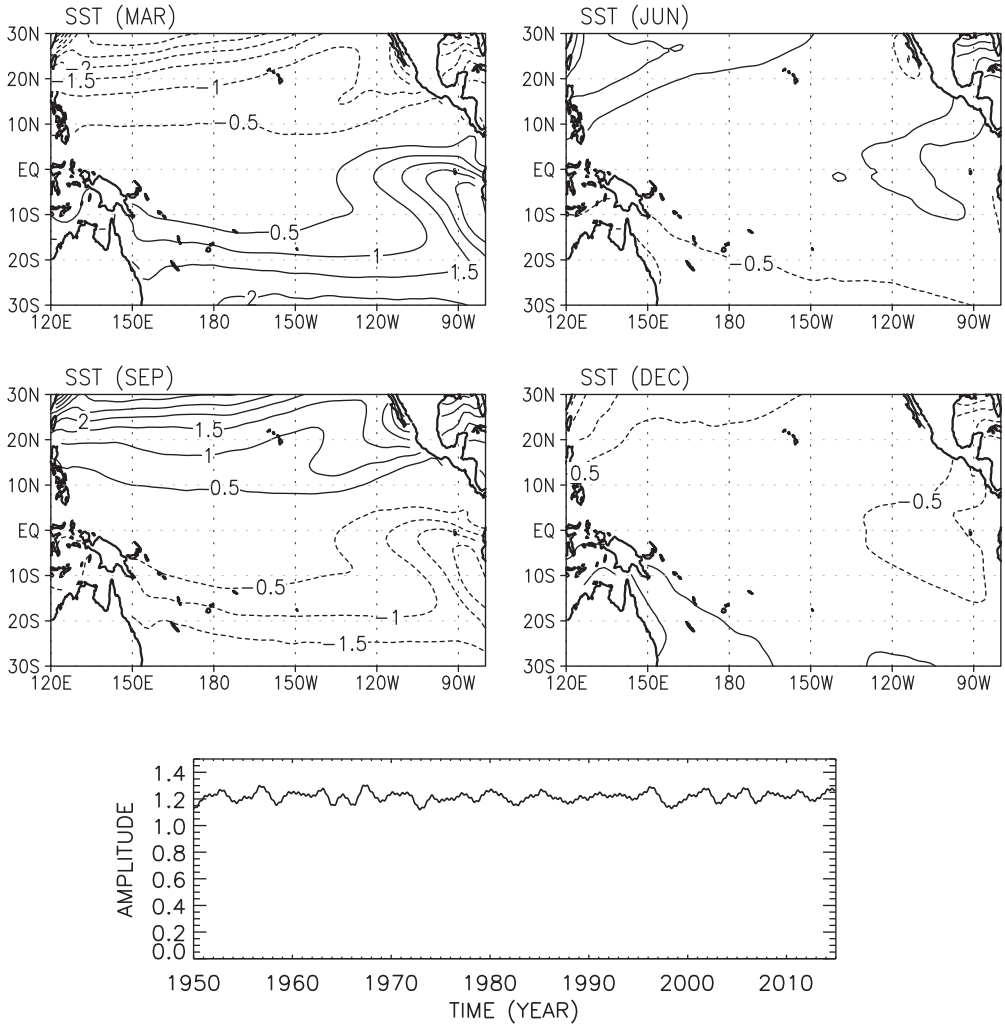

Fig. 4. The first CSEOF of the monthly sea surface temperatures in the tropical Pacific (upper panel) and the corresponding PC time series (lower panel).

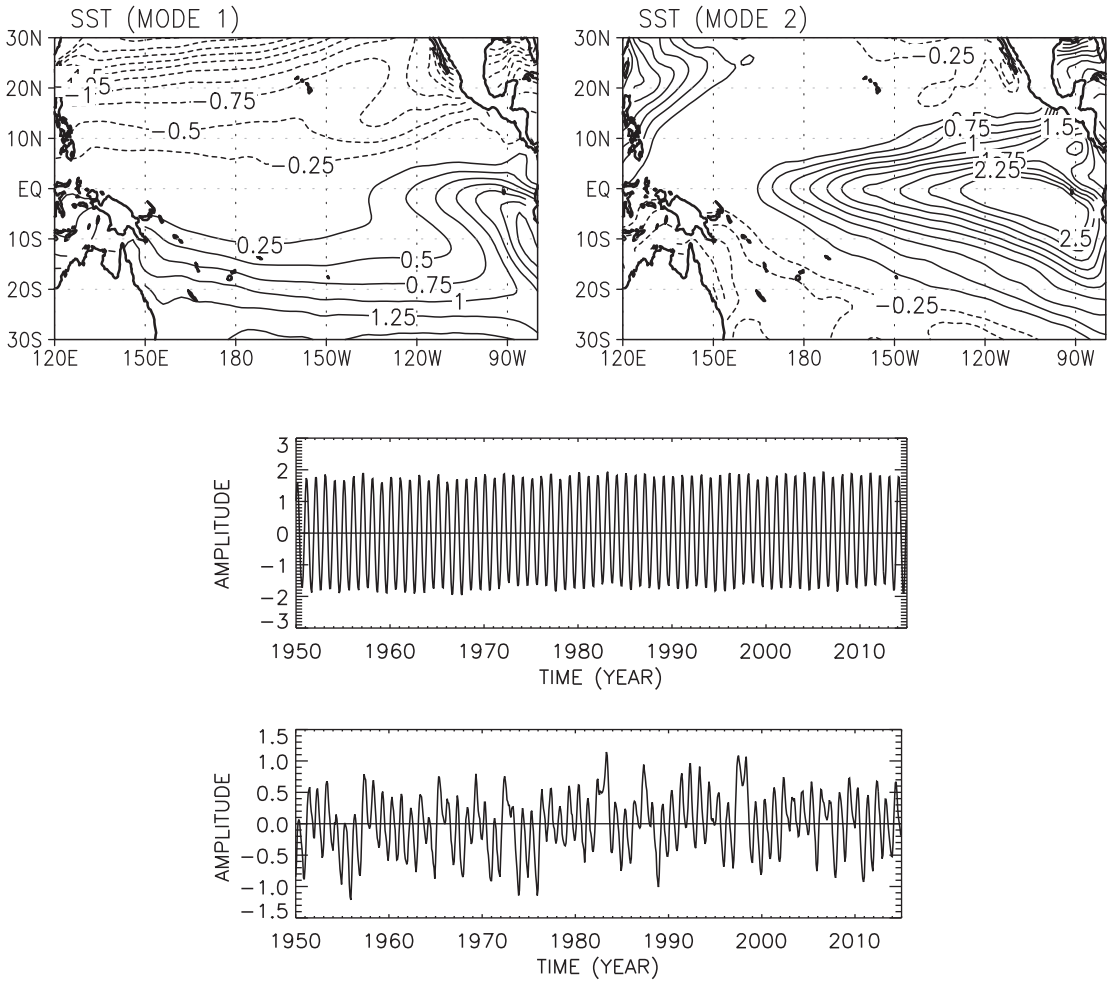

Fig. 5. The first two EOF modes of the monthly tropical Pacific sea surface temperatures (upper panels) and the corresponding PC time series (middle and lower panels).

Appendix 1). The first CSEOF and the composite essentially describe the same evolution of equatorial SST whereas the first two extended EOFs describe different evolutions. This discrepancy in the extended EOF representation is due to the use of temporal lag instead of specific time reference in extended EOF analysis (see Appendix 1). The second extended EOF is 90° out of phase with the first one. In the absence of specific time reference, the second extended EOF describes exactly the same evolution of the equatorial SST of the first extended EOF. This redundancy is clearly visible in the corresponding lagged correlogram in Fig. 8. More examples of comparisons between CSEOF analysis and extended EOF analysis are presented by Kim and Wu (1999) and Seo and Kim (2003).

Fig. 9 shows the evolution of equatorial Pacific SST based on the first two EOFs and the first two extended EOFs assuming that the evolution of the seasonal cycle is captured by the first two modes. The evolution of the seasonal cycle is reasonably described in terms of the first two modes. While EOF and extended EOF analyses appropriately capture the spatial patterns associated with the annual cycle in this simple example, it is obvious that two spatial patterns are rarely sufficient to depict complicated physical evolution and possibly mislead physical interpretation of the spatially propagating signals (e.g., Kim et al., 2010).

5. Implications on physical inferences

Once physically interesting modes are identified, one may be interested in identifying spatial patterns of other physical variables, which evolve in a consistent way with the identified modes. This is an important question for many reasons. For one thing, one may want to know how two or more physical variables interact in association with a certain physical process. For example, how is the surface wind change related to the sea surface temperature change in the tropical Pacific during El Niños (e.g., Kim, 2002)? How does the sea level height change (e.g., Kim and Kim, 2002)? What happens to the thermocline depth when sea level height changes (Kim and Kim, 2004)? For another, we may want to know how a certain physical process affects other physical processes at remote areas. For example, how is the mid-latitude jet affected by El Niño-Southern Oscillation (ENSO) events or vice versa (Kim et al., 2003)? Answers to such questions may be explored by finding physically consistent evolution patterns among different physical or derived variables.

Patterns of two physical variables may be called “consistent” when they have a “common” evolution history (see Fig. 10). When a physical system (or a process) undergoes a stochastic variation for some reason, two physical variables describing the system may evolve in the same fashion. This may not be quite true if the relationship between two physical variables is a nonlinear one. Also, if the response of a physical system as manifested in the variation of physical variables to imposed external forcing is nonlinear, this argument may not truly hold. On the other hand, it can be argued that the response of a physical system is essentially linear when a forcing has a much longer time scale than that of the physical system. If a physical system is strongly nonlinear, it does not make much sense to decompose data into physical modes in the first place. Therefore, let us take a very simple-minded view and assert that our initial assumption is reasonable; that is, two spatial patterns having the same evolution history are consistent patterns.

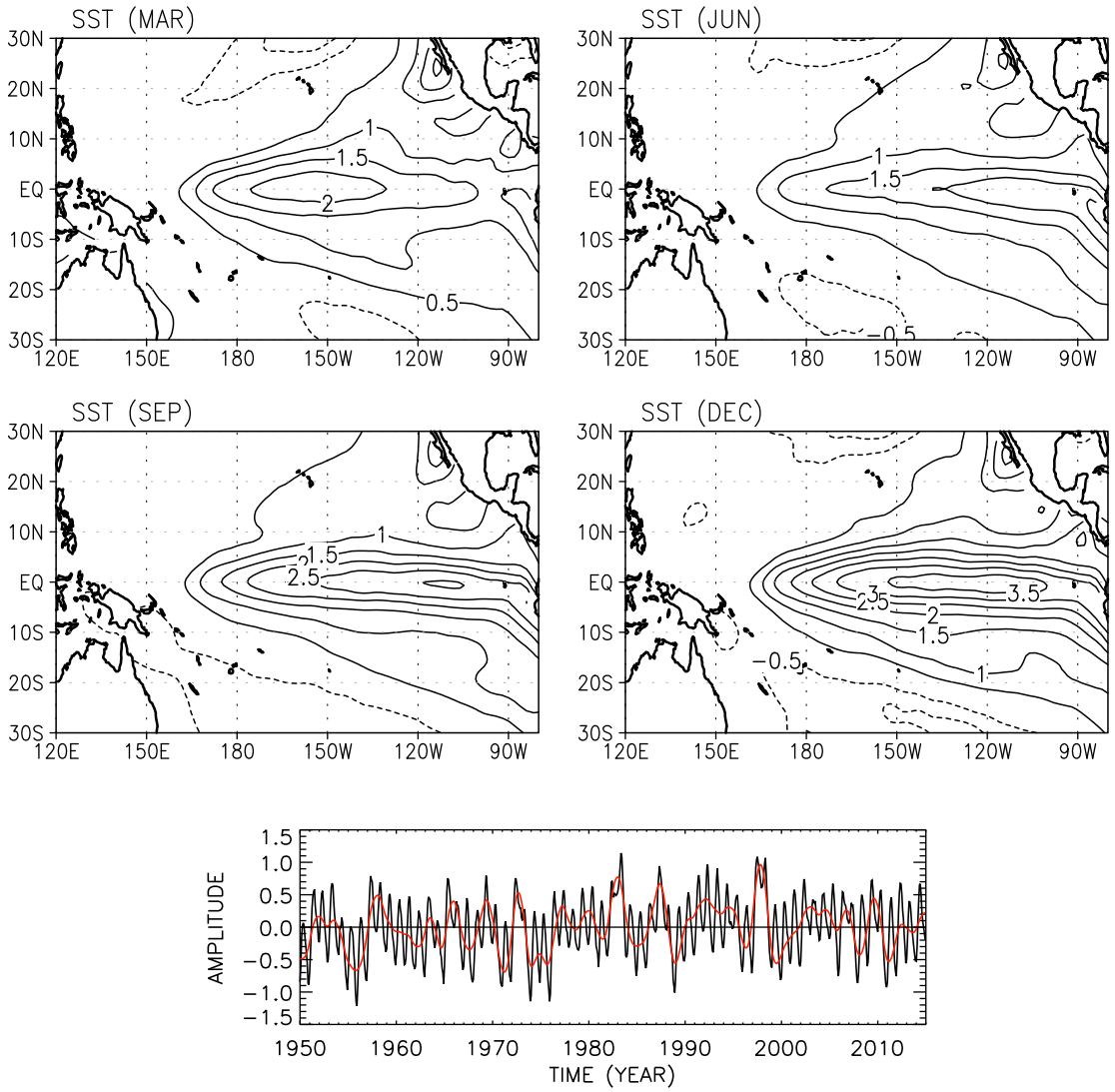

Fig. 6. The second CSEOF mode of the monthly tropical Pacific SST (upper panel) and the corresponding PC time series (lower panel).

One way to find a consistent pattern is to simply project the data, $P(x, t)$, on the target time series, say, $T(t)$:

$$P_{\text{proj}}(x) = \sum_t P(x,t) \cdot T(t), \tag{23}$$

where $T(t)$ is the amplitude time series of target spatial pattern, $\phi(x)$. This procedure is essentially the inner product in time and results in the component of $P(x, t)$, which is parallel to (consistent with) the time series $T(t)$. Thus, the target pattern, $\phi(x)$, and the consistent pattern, $P_{\text{proj}}(x)$, have the same evolution history:

$$\phi(x) \Rightarrow T(t) \Leftarrow P_{\text{proj}}(x). \tag{24}$$

This procedure can also be written as a regression problem in EOF space. Let

$$P(x,t) = \sum_n P_n(t)\varphi_n(x), \tag{25}$$

where $\varphi_n(x)$ and $P_n(t)$ are the EOFs and the corresponding PC time series of $P(x, t)$. Then, a regression relationship is sought between the target time series and a number of "predictor" PC time series:

$$T(t) = \sum_{n=1}^{N} \beta_n P_n(t) + \varepsilon(t). \tag{26}$$

That is, we want to express the target time series, $T(t)$, as a linear combination of predictor time series, $P_n(t)$. Regression coefficients, β_n, are determined such that the residual error variance, $E(\varepsilon^2)$, is minimized. Once regression coefficients are found, a consistent pattern is given by

$$P(x) = \sum_{n=1}^{N} \beta_n \varphi_n(x). \tag{27}$$

The number of predictor time series, N, used in the regression should be small in order to avoid over fitting. Typically a number between 10 and 20 is a good choice.

Note that the procedure described in Eqs. (26) and (27) is the same as the projection method (Eq. (23)), which can be proved as follows. Eq. (23) can be rewritten as

$$P(x) = \sum_t \sum_n P_n(t)\varphi_n(x) \cdot T(t) = \sum_n \alpha_n \varphi_n(x), \tag{28}$$

where α_n is the correlation between the two normalized time series $T(t)$ and $P_n(t)$. Note from Eq. (26) that

$$\alpha_n = \sum_t T(t) \cdot P_n(t) = \sum_m \beta_m \sum_t P_m(t) \cdot P_n(t) = \beta_n, \tag{29}$$

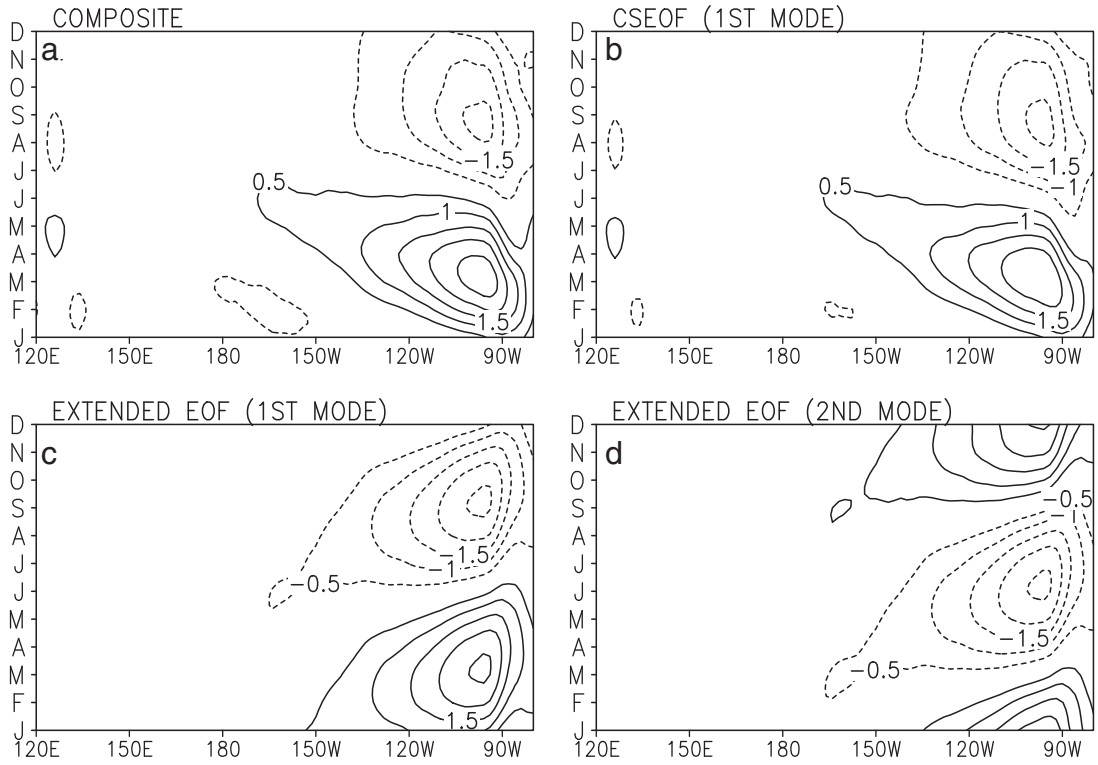

Fig. 7. Longitude–time plots of equatorial sea surface temperature evolutions (°C) in (a) the monthly composite, (b) the first CSEOF, (c) the first extended EOF, and (d) the second extended EOF of tropical Pacific sea surface temperatures.

from the uncorrelatedness of the PC time series, $P_n(t)$. Thus, regression coefficients are identified as correlation coefficients, and Eqs. (23) and (27) are essentially identical approaches. One advantage in using the regression method is that certain unwanted portions of data, say noise, can be excluded by choosing a proper EOF truncation level. This may result in a physically more meaningful spatial pattern without noise, which is easier to understand and interpret.

The two approaches of finding consistent physical patterns as discussed above will fail when a physical system (process) produces two different evolution histories for two physical variables. One important reason why this may happen is that "physical response characteristics" of two variables are generally different (see Fig. 11). For example, evolution of sea surface temperature may not be the same as that of surface wind at the same location even for a very simplified physical system with only one physical process. In fact, their evolution is determined by a physical relationship between the two variables aside from the stochastic undulation of the physical system. As depicted in Fig. 11, response characteristics of two physical variables may not necessarily coincide, and this results in different evolution for distinct variables even though they represent the same physical process. We can only stipulate that the stochastic component of variation is identical in the evolutions of two variables pertaining to the same physical process.

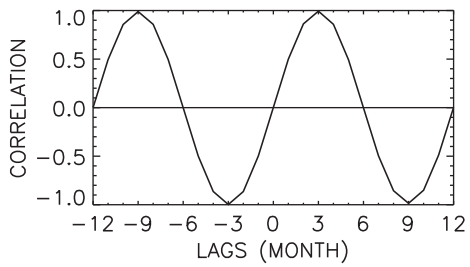

Fig. 8. Lagged correlogram of the first two PC time series of the extended EOFs of the tropical Pacific sea surface temperatures.

Therefore, a proper distinction from stochastic undulation and explanation of time-dependent physical response is vital in accurately determining physically consistent evolutions or teleconnection responses in different variables.

Fig. 12 depicts the difference in the representation of variability between CSEOF analysis and EOF analysis. In the former technique, temporal dependence due to physics is resolved in CSLVs and is separated from the stochastic fluctuation in the PC time series. By contrast, the latter technique assumes stationary (uniform) response characteristics for all physical variables. Thus, in the context of CSEOF analysis, the view that two consistent patterns have the same "stochastic" evolution history still holds:

$$B(x,t) \Rightarrow T(t) \Leftarrow C(x,t). \tag{30}$$

Therefore, the regression method, Eqs. (26) and (27) with EOFs being replaced by CSLVs, can be used in finding physically consistent patterns in CSEOF space. Let $T(x, t)$ be a "target" variable, say wind, which is decomposed into:

$$T(x,t) = \sum\nolimits_n B_n(x,t) T_n(t). \tag{31}$$

Let $P(x, t)$ be a "predictor" variable, say geopotential height, which is decomposed into:

$$P(x,t) = \sum\nolimits_n A_n(x,t) P_n(t). \tag{32}$$

Then, the evolution of the predictor variable, which is physically consistent with $B_n(x, t)$ is obtained as follows. Regression coefficients are obtained from

$$T_n(t) = \sum\nolimits_m \beta_m^{(n)} P_m(t) + \varepsilon_n(t), \tag{33}$$

where the superscript (n) denotes that the regression coefficients are for the nth mode of the target variable. Then the evolution of the

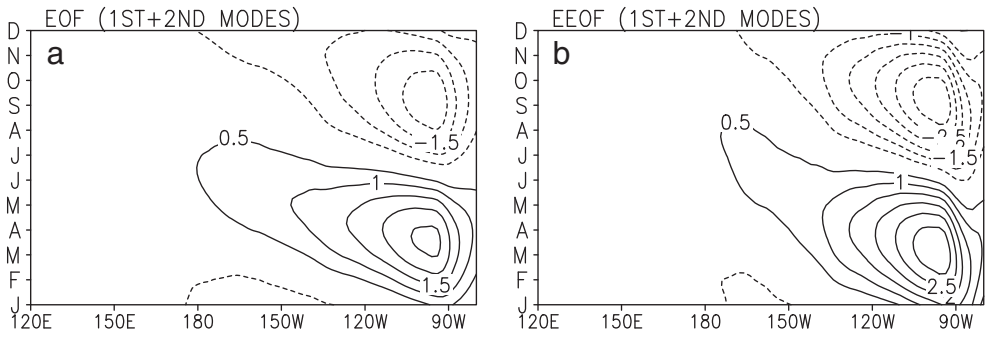

Fig. 9. Longitude–time plot of the evolution of sea surface temperatures at the equator (a) based on the first two EOFs and (b) based on the first two extended EOFs.

predictor variable, $C_n(x, t)$, which is physically consistent with the evolution, $B_n(x, t)$, is obtained via

$$C_n(x,t) = \sum_m \beta_m^{(n)} A_m(x,t), \tag{34}$$

where $A_m(x, t)$ are the CSLVs of the predictor variable. In this way, evolution of any variables can be derived to be physically consistent with that of the target variable.

It should be mentioned that regressed patterns, $C_n(x, t)$ in Eq. (34), are, in theory, orthogonal to each other. From Eq. (33), it can be shown, within the limit of small regression errors, that

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} T_n(t)T_m(t) &= \frac{1}{T}\sum_{t=1}^{T}\left(\sum_k \beta_k^{(n)} P_k(t) + \varepsilon_n(t)\right)\cdot\left(\sum_l \beta_l^{(m)} P_l(t) + \varepsilon_m(t)\right)\\ &= \sum_k\sum_l \beta_k^{(n)}\beta_l^{(m)}\frac{1}{T}\sum_{t=1}^{T} P_k(t)\cdot P_l(t)\\ &= \sum_k\sum_l \beta_k^{(n)}\beta_l^{(m)}\delta_{kl} = \sum_k \beta_k^{(n)}\beta_k^{(m)} = \delta_{nm}.\end{aligned} \tag{35}$$

Eq. (35) derives from the uncorrelatedness of PC time series and shows that the regression coefficients of the nth target time series and those of the mth target time series are orthogonal to each other. Thus, for $n \neq m$,

$$\begin{aligned}&\frac{1}{Nd}\sum_{t=1}^{d}\sum_x C_n(x,t)\cdot C_m(x,t)\\ &\quad= \frac{1}{Nd}\sum_{t=1}^{d}\sum_x\left(\sum_k \beta_k^{(n)} A_k(x,t)\right)\cdot\left(\sum_l \beta_l^{(m)} A_l(x,t)\right)\\ &\quad= \sum_k\sum_l \beta_k^{(n)}\beta_l^{(m)}\frac{1}{Nd}\sum_{t=1}^{d}\sum_x A_k(x,t)A_l(x,t)\\ &\quad= \sum_k\sum_l \beta_k^{(n)}\beta_l^{(m)}\delta_{kl} = \sum_k \beta_k^{(n)}\beta_k^{(m)} = \delta_{nm}.\end{aligned} \tag{36}$$

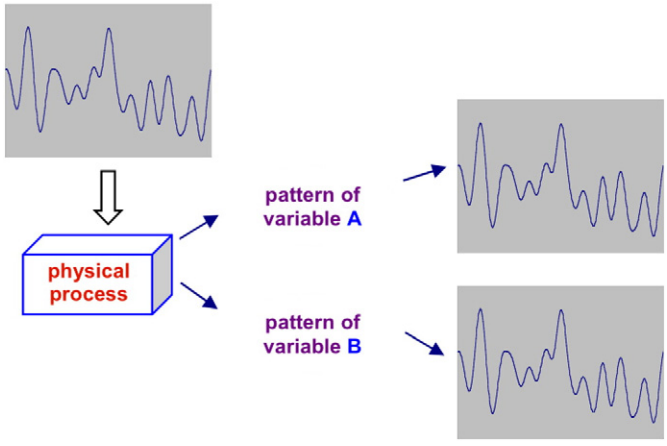

Fig. 10. Schematic diagram explaining the concept of physically consistent spatial patterns of two variables. The two consistent spatial patterns pertaining to the same physical process, which is driven by a stochastic external forcing, may have the same evolution history. Reproduced from Kim et al. (2003).

Thus, regressed patterns for two different modes should also be orthogonal to each other within the limit of small regression errors. It, then, can be shown that

$$P(x,t) \doteq \sum_n C_n(x,t)T_n(t). \tag{37}$$

In order to prove Eq. (37), let us rewrite Eq. (37) with the aid of Eqs. (33) and (34) as

$$\begin{aligned}P(x,t) &\doteq \sum_n\sum_k \beta_k^{(n)} A_k(x,t)\sum_l \beta_l^{(n)} P_l(t)\\ &= \sum_k\sum_l A_k(x,t)P_l(t)\sum_n \beta_k^{(n)}\beta_l^{(n)}\\ &= \sum_k\sum_l A_k(x,t)P_l(t)\delta_{kl} = \sum_k A_k(x,t)P_k(t) = P(x,t),\end{aligned} \tag{38}$$

where the small regression error has been ignored.

The consistent patterns of two physical variables, $B_n(x, t)$ and $C_n(x, t)$, respectively in Eqs. (31) and (37), may not generally have the same physical response characteristics. In fact, the physical relationship between physically consistent patterns of two or more physical variables (e.g., relationship between $B_n(x, t)$ and $C_n(x, t)$) should be dictated by a governing equation describing the particular physical process they represent. It is the "stochastic" component of undulation that should be identical in the evolution of two physical variables originating from the same physical process. On the other hand, space–time evolution (CSLVs) of two consistent patterns should be compatible with respect to their governing physics. In this sense, CSEOF analysis followed by regression analysis may be viewed as

$$Data(x,t) = \sum_n \{T_n(x,t), H_n(x,t), U_n(x,t), V_n(x,t), \ldots\}T_n(t), \tag{39}$$

where the entire data collection (physical variables) is decomposed into a series of physical modes. Specifically, $\{T_n(x, t), H_n(x, t), U_n(x, t), V_n(x, t), \ldots\}$ represents the evolution of the nth mode as reflected in various physical variables; evolution of each variable should be physically consistent with those of the other variables. For this reason, CSEOF decomposition may be regarded as "physical" decomposition with individual modes describing short time-scale physical processes and corresponding PC time series denoting long-term amplitude undulation of the physical processes.

As an example, Fig. 13 shows physically consistent evolutions of geopotential height and wind as discussed in conjunction with Eq. (30). As a comparison between different panels shows, the evolution of geopotential height anomalies (predictor) is not identical with that of the wind anomalies (target) whereas the physical relationship between the two appears to be reasonable in the context of the geostrophic balance. The evolution of the geopotential height anomalies is very consistent with that of the wind anomalies for the entire 90-day period of the CSLV (figure not shown). Further, a reasonable physical relationship is also seen at other pressure levels as well. In fact, a reasonable physical relationship is seen not only for the first mode but also for all the first ten modes as can be seen in terms of the R^2 values of regression in Tables 1 and 2.

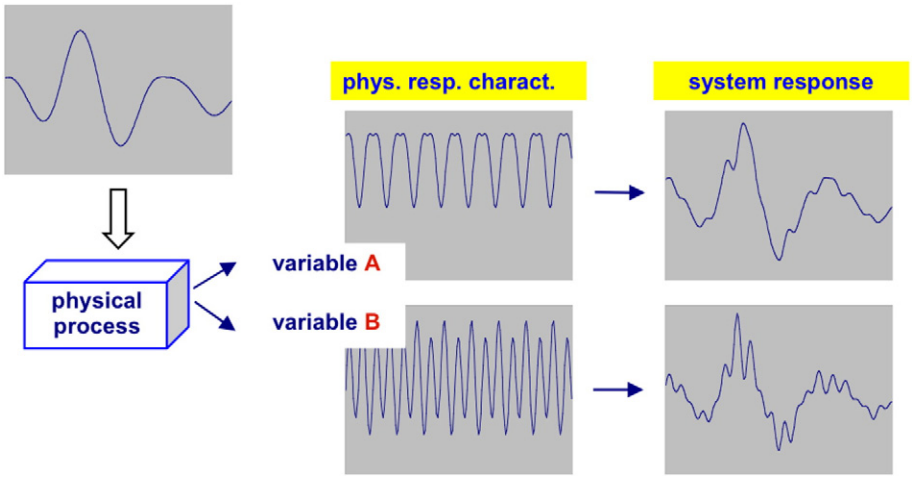

Fig. 11. Physical response characteristics of two variables pertaining to a physical process. When an independent physical system (or a process) fluctuates due to an external forcing, the response of two physical variables may be different from each other because of different physical response characteristics of the two variables. The resulting evolution histories, then, will be different between the two.
Reproduced from Kim et al. (2003).

As the second example, Fig. 14 shows the vertical section of the stratospheric seasonal cycle averaged in the 60°–80°N latitude band; the seasonal cycle was extracted from the 150-day (November 17–April 15) boreal winter data derived from the 1.5° × 1.5° ERA interim data for the period of 1989–2008 (Dee et al., 2011). The first CSEOF mode of 10-hPa air temperature, which is the target variable, is shown in Fig. 14a. The regressed patterns of geopotential height and potential vorticity are shown in Fig. 14b and e. As can be seen in the corresponding PC time series, the stratospheric seasonal cycle exhibits strong interannual variability; the PC time series derived from the ERA interim reanalysis data is very similar to that derived from the 1948–2008 NCEP/NCAR reanalysis data (Kalnay et al., 1996).

All physical variables exhibit seemingly downward propagations of anomalies; as can be seen, significant time lags exist between anomalies at 10- and 200-hPa levels. Not only so, this lag appears to vary from one variable to another. It is obvious that a conventional analysis based on co-evolving spatial patterns cannot capture such distinct temporal evolutions of physical variables as one physical mode. Not being able to render physical evolution in its entirety results in at most partial or incomplete pictures of the true nature of the physical mechanism. For example, Fig. 14f shows the zonally average tropopause pressure in comparison with the zonally-averaged 200-hPa potential vorticity; the evolutions of these two variables are highly correlated as can be seen. Also, the zonally averaged 10 hPa potential vorticity is negatively correlated with the tropopause pressure. It is not entirely obvious from Fig. 14f why there is a strong negative correlation in the evolution of the upper and lower stratospheric potential vorticity fields. It is through the detailed vertical evolution pattern in Fig. 14e that the negative correlation of the upper and lower stratospheric potential vorticity fields makes sense.

The accuracy of regression analysis in CSEOF space in deriving physically consistent evolutions from different variables makes this "statistical" technique uniquely adapted for detailed physical inferences. For example, Fig. 14c represents the evolution pattern of potential vorticity derived from the potential temperature based on the physical relationship between them (see figure caption). Likewise, Fig. 14d represents temperature derived from geopotential height based on their physical relationship (essentially hydrostatic

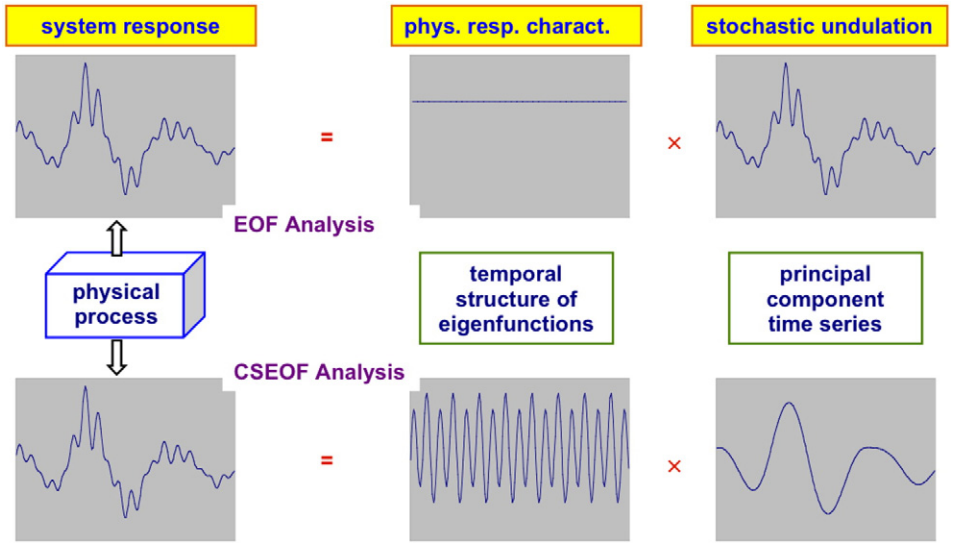

Fig. 12. Schematic diagram explaining the conceptual difference between stationary EOF analysis and cyclostationary EOF analysis. In EOF analysis, physical response is uniform (stationary) in time while in cyclostationary EOF analysis physical response characteristic is periodically time dependent with a given nested period.
Reproduced from Kim et al. (2003).

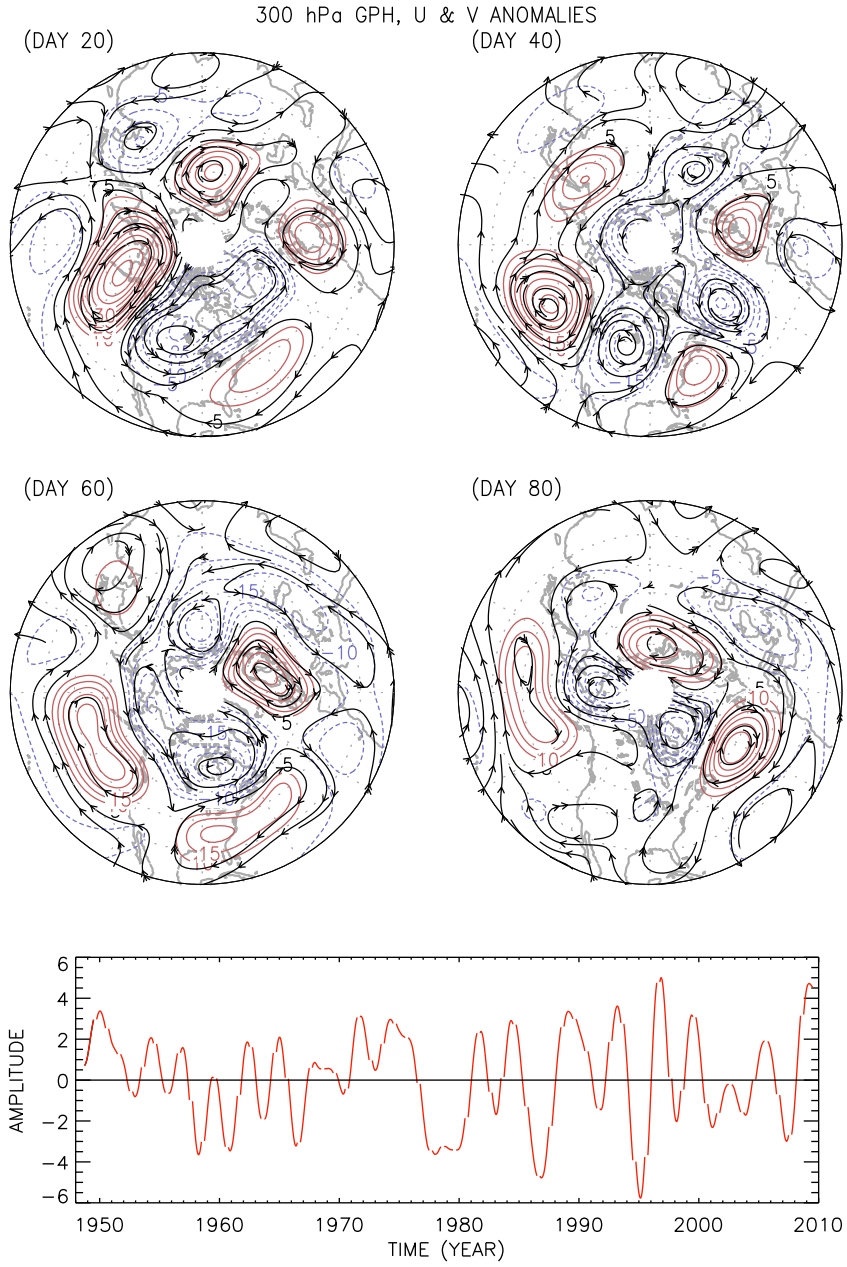

Fig. 13. Spatial patterns of 300 hPa geopotential height anomalies (predictor: colored contours) and the streamlines of winds (target: black contours) on four different days of the 90-day long first CSEOF mode (upper panel) and the corresponding PC time series (lower panel).

equation; see figure caption). As can be seen, the physical relationship holds accurately between the regressed fields. This example firmly demonstrates one of the key benefits of the CSEOF analysis technique: physically consistent evolutions can be extracted from different variables.

6. Implications on spectral inferences

Time series of physical variables are often analyzed spectrally to identify any dominant periodicities in the fluctuation. Spectral analysis is based on Fourier decomposition. Sinusoids are proper basis functions (EOFs) of stationary time series as can be proved as follows:

$$\int_T R(t-t')\exp(2\pi i\omega_n t')dt' = f(\omega_n)\exp(2\pi i\omega_n t), \tag{40}$$

where $f(\omega_n)$ is the spectral density function. Eq. (40) follows from the Fourier relationship between the autocovariance function and the spectral density function (Newton, 1988). Under the stationarity assumption, the temporal covariance function is identified as autocovariance function, $R(\tau)$. Then, it can be shown that Fourier functions are the eigenfunctions of the Karhunen–Loève equation and the eigenvalues are the spectral density function of the time series. Thus, a spectrum is nothing but a plot of eigenvalues.

When the stationarity assumption is not met as in the case of cyclostationary processes, spectral analysis should be carried out with a strong caution since Fourier functions are not a proper basis set. This means that Fourier expansion coefficients of time series are neither independent of each other nor orthogonal to each other and, henceforth, the variability of the time series is not properly partitioned among independent modes. This, of course, is not a serious concern since spectral estimation based on only one realization is already prone to some

Table 1
The R^2 values of regression for the first ten CSEOF modes of the 300-hPa wind anomalies at the standard pressure levels. Regression was conducted using the first 20 PC time series of wind anomalies at each standard pressure levels as predictors.

	1	2	3	4	5	6	7	8	9	10
200 hPa	0.993	0.994	0.992	0.989	0.987	0.996	0.976	0.992	0.984	0.965
250 hPa	0.999	0.998	0.998	0.998	0.998	0.998	0.995	0.998	0.995	0.993
300 hPa	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
400 hPa	0.997	0.997	0.998	0.997	0.997	0.997	0.997	0.997	0.994	0.992
500 hPa	0.993	0.992	0.994	0.992	0.995	0.992	0.994	0.994	0.982	0.981
600 hPa	0.990	0.988	0.989	0.988	0.992	0.986	0.989	0.991	0.968	0.971
700 hPa	0.986	0.982	0.982	0.983	0.989	0.976	0.983	0.985	0.940	0.957
850 hPa	0.981	0.971	0.960	0.972	0.978	0.972	0.963	0.972	0.911	0.928
925 hPa	0.977	0.963	0.950	0.955	0.965	0.966	0.952	0.957	0.890	0.912
1000 hPa	0.976	0.963	0.944	0.946	0.928	0.963	0.952	0.951	0.882	0.894

sampling error. A more serious concern is in the erroneous estimation of the positions and magnitudes of spectral peaks as hinted in Fig. 12. Whether the physical response characteristics of a variable are resolved accurately or not affects the outcome of spectral analysis significantly (Kim and Chung, 2001). This can be demonstrated as follows.

Let us suppose in Eq. (9) that

$$B(t) = A\cos\omega_1 t \ \text{ and } \ S(t) = C\cos\omega_2 t, \tag{41}$$

where the physical time scale, ω_1^{-1}, is much shorter than the stochastic time scale, ω_2^{-1}. Then, the raw time series can be written as

$$T(t) = B(t)S(t) = A\cos\omega_1 t \cdot C\cos\omega_2 t = \frac{AC}{2}\left(\cos\omega_1' t + \cos\omega_2' t\right), \tag{42}$$

where

$$\omega_1' = \omega_1 + \omega_2 \ \text{ and } \ \omega_2' = \omega_1 - \omega_2. \tag{43}$$

Thus, spectral analysis would indicate peaks at ω_1' and ω_2' instead of ω_2, which is the frequency of stochastic undulation. This is a critical issue since the periods corresponding to spectral peaks are among the important statistical inferences about a random variable. In CSEOF analysis, the physical response function is separated from stochastic modulation of a physical process and the PC time series (stochastic modulation) is stationary in theory (Kim and North, 1997a). Therefore, spectral analysis on a CSEOF PC time series does not pose an intrinsic problem.

7. Utility of CSEOF analysis

A primary advantage of CSEOF analysis lies in the physical consistency of loading vectors extracted from different physical variables, which we already discussed in conjunction with Figs. 13 and 14. The key idea of physical consistency is expressed in Eq. (39) through regression analysis in CSEOF space. The concept of physical consistency can be extended to finding teleconnection responses. If the target domain in Eq. (31) and the predictor domain in Eq. (37) are different, then regression in CSEOF space, in essence, is used to find a remote response to a physical process in the target domain. Figs. 15 and 16 illustrate the concept of teleconnection. There are three sets of CSEOF patterns—one for the North Pacific domain (20°–60°N; red contours), another for the entire domain (30°S–60°N; blue contours), and the other for the tropical Pacific domain (30°S–30°N; black contours). The first two sets of spatial patterns are from the regressed loading vectors onto the first and second CSEOF modes of the tropical Pacific SSTA aside from the seasonal cycle. As can be seen in the figures, the regressed patterns are fairly similar to the CSEOF patterns of the tropical Pacific SSTA and to each other although SST variability differs from one domain to another significantly. The similitude of the seasonal evolution patterns is also remarkable. Fig. 17 indicates that the temporal evolutions of SSTA exhibited in the regression patterns are similar to each other and to that in the target patterns. It should be emphasized that the evolution of SSTA in the target pattern is quite different from that in the regression pattern at a location apart from the target region as expected. The distinctiveness of the temporal evolution in the tropical region from that in the northern North Pacific makes it clear that it is important to link physical evolutions rather than two spatial patterns between two geographically unconnected regions (see Yeo et al., 2012 for more details). Examining the consistent physical evolutions has been applied to determine the atmospheric forcing responsible for the ocean response in a remote region (Na et al., 2010, 2012a) and to investigate their lead–lag relationship (Na et al., 2012b). Not only distinct physical evolution but also temporal lead and lag between two places make it difficult to apply a conventional analysis technique to a teleconnection problem. An accurate method of teleconnection should be able to account for distinct physical evolution and lead/lag relationship between two remote regions.

Another important application of the CSEOF technique is the construction of synthetic data in a physically consistent manner. The idea can be elucidated in terms of the following equation:

$$\hat{T}(r,t) = \sum\nolimits_n B_n(r,t)\hat{T}_n(t), \tag{44}$$

Table 2
The R^2 values of regression for the first ten CSEOF modes of the 300–hPa wind anomalies at the standard pressure levels. Regression was conducted using the first 20 PC time series of geopotential height anomalies at each standard pressure level as predictors.

	1	2	3	4	5	6	7	8	9	10
200 hPa	0.953	0.938	0.960	0.931	0.925	0.903	0.838	0.878	0.878	0.735
250 hPa	0.967	0.949	0.976	0.958	0.964	0.934	0.878	0.906	0.916	0.816
300 hPa	0.975	0.958	0.973	0.963	0.975	0.930	0.898	0.914	0.920	0.838
400 hPa	0.975	0.951	0.967	0.962	0.973	0.915	0.903	0.922	0.916	0.841
500 hPa	0.974	0.940	0.962	0.960	0.972	0.905	0.898	0.926	0.908	0.839
600 hPa	0.973	0.936	0.953	0.950	0.969	0.904	0.893	0.914	0.909	0.880
700 hPa	0.963	0.933	0.933	0.948	0.963	0.904	0.888	0.906	0.905	0.892
850 hPa	0.956	0.940	0.909	0.952	0.945	0.897	0.882	0.886	0.824	0.878
925 hPa	0.950	0.939	0.908	0.940	0.932	0.893	0.859	0.818	0.749	0.858
1000 hPa	0.946	0.937	0.900	0.923	0.918	0.884	0.868	0.777	0.754	0.845

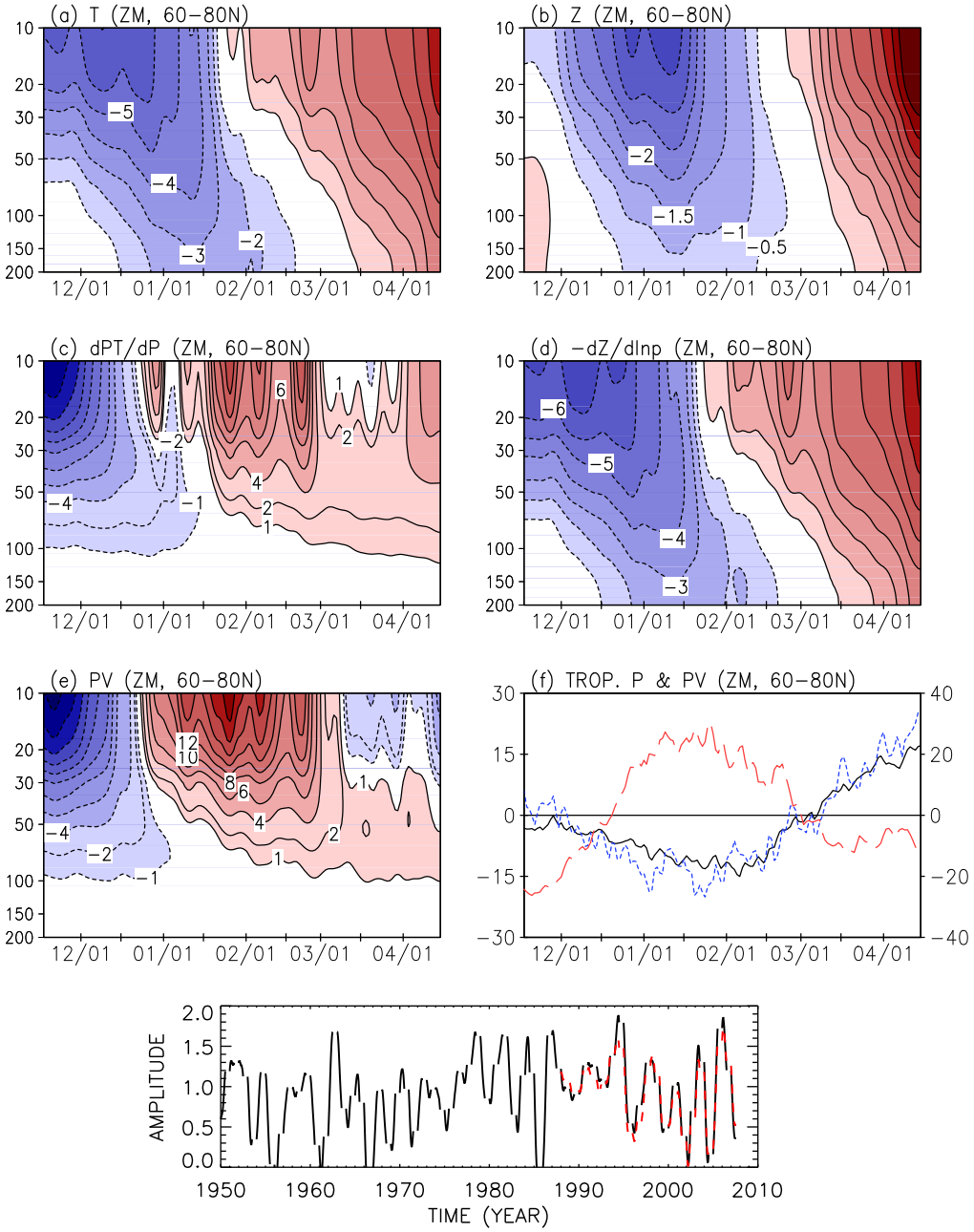

Fig. 14. The wintertime evolution of (a) temperature, (b) geopotential height, (c) $-g(f+\zeta)(d\theta/dp)$, (d) $-(g/R)(dZ/d\ln p)$, (e) potential vorticity, (f) tropopause pressure (black solid), 200-hPa potential vorticity (blue dotted), and 10-hPa potential vorticity (red dashed), (g) and the PC time series (black: NCEP, red: ERA interim) for the first CSEOF mode (stratospheric seasonal cycle). All the panels represent 60°–80°N zonal averages.

where the hat sign denotes "estimate". There are many different contexts and approaches in obtaining the estimates $\hat{T}_n(t)$. In developing a weather generator, each PC time series can be fit to an ARMA (autoregressive–moving average) model:

$$T_n(t)+\sum_{j=1}^{p}\alpha_j T_n(t-j)=\varepsilon(t)+\sum_{k=1}^{q}\beta_k\varepsilon(t-k) \sim ARMA\left(p,q,\tilde{\alpha},\tilde{\beta},\sigma^2\right), \quad (45)$$

where p and q are respectively the AR and MA orders, $\tilde{\alpha}=\{\alpha_i|i=1,\ldots,p\}$ and $\tilde{\beta}=\{\beta_j|j=1,\ldots,q\}$ are the AR and MA coefficients, and σ^2 is the variance of the white noise time series. Then, synthetic PC time series can be generated based on Eq. (45) by using different realizations of the white noise time series. In this way, as many synthetic amplitude time series as needed can be constructed in such a way that they have identical statistical properties with the original PC time series. Once synthetic time series are constructed, synthetic datasets can be constructed from Eq. (44).

As demonstrated in J.W. Kim et al. (2013) and K.Y. Kim et al. (2013), synthetic datasets exhibit statistical properties that are very close to those of the original data sets. One superior advantage of the CSEOF-based weather generator is that other variables can be generated in a physically consistent way as the target variable. This can be accomplished via regression analysis in CSEOF space. Note that entire data can be written as in Eq. (39) and distinct physical evolutions as manifested in different variables are all governed by an identical PC time

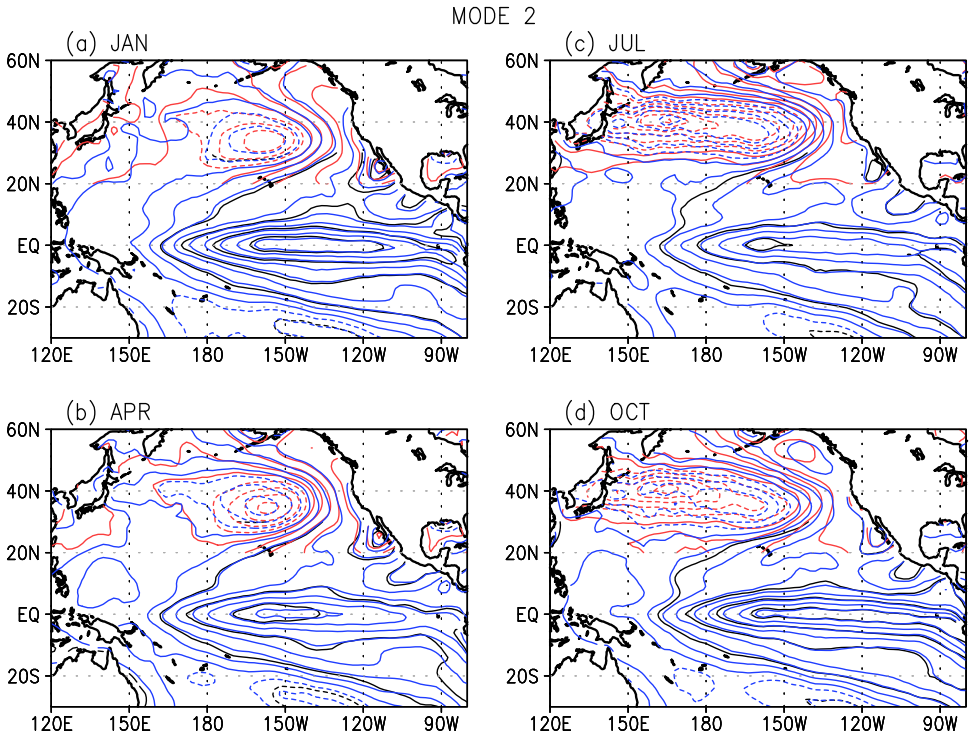

Fig. 15. The first CSEOF mode (aside from the seasonal cycle) of tropical (30°S–30°N) SSTA (black contours), the regressed SSTA patterns (red contours) over the North Pacific (30°–60°N), and the regressed SSTA patterns (blue contours) over the combined domain (30°S–60°N). The nested period is 24 months. Contour intervals are 0.1 °C.

series for each mode n. Thus, the synthetic time series generated by using Eq. (45) apply to the regressed loading vectors derived from different physical variables. As a result, physical consistency is preserved among synthetic datasets of different variables.

The concept delineated above can also be extended into a statistical prediction study, which is another interesting application of the CSEOF technique. Eq. (45) can be fitted to detrended PC time series. Then, Eq. (45) can be used to generate longer PC time series together with

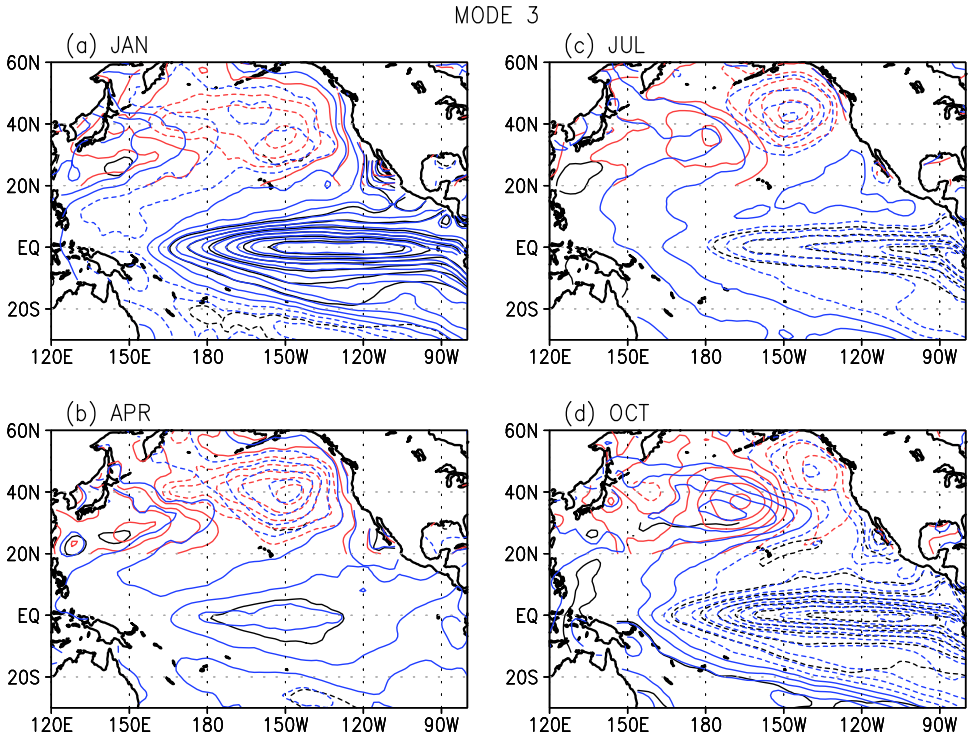

Fig. 16. The second CSEOF mode (aside from the seasonal cycle) of tropical (30°S–30°N) SSTA (black contours), the regressed SSTA patterns (red contours) over the North Pacific (20°–60°N), and the regressed SSTA patterns (blue contours) over the combined domain (30°S–60°N). The nested period is 24 months. Contour intervals are 0.1 °C.

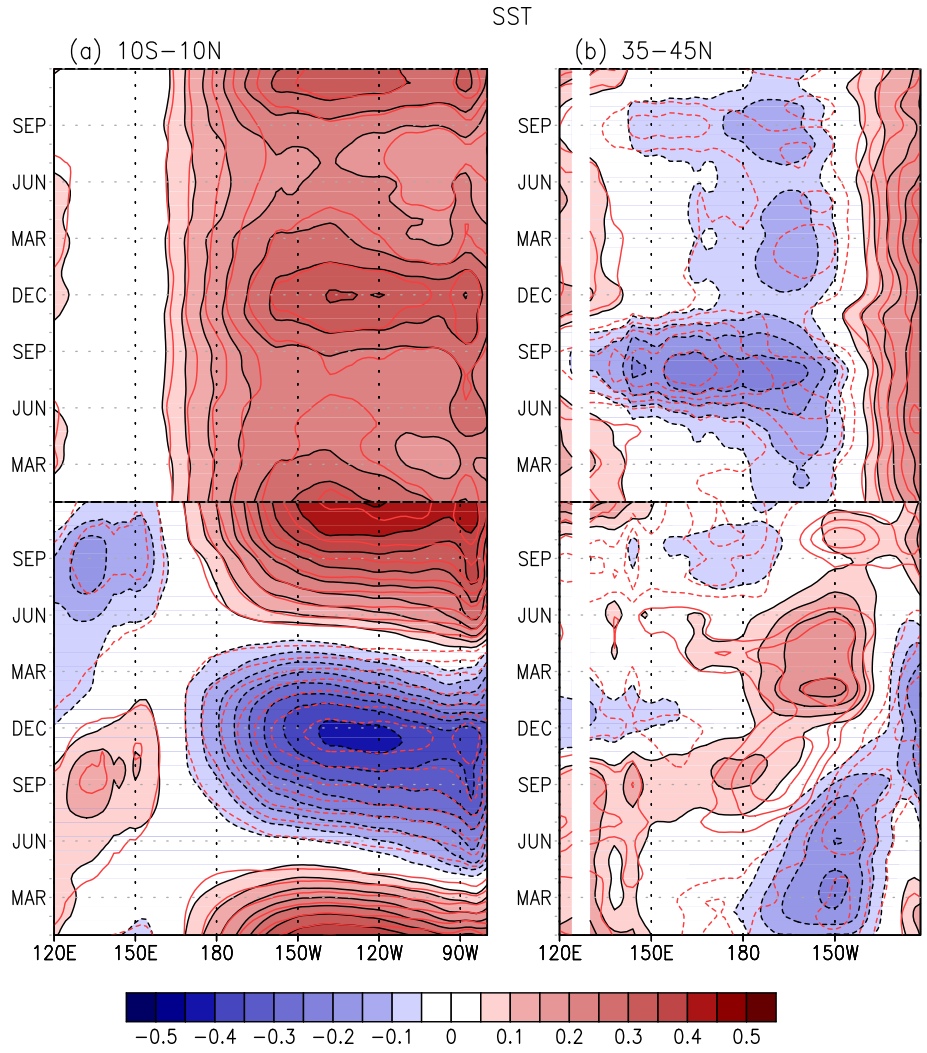

Fig. 17. Temporal evolution of SSTA for the second CSEOF mode (upper panels) and the third CSEOF mode (lower panels). The left panels denote SSTA variability averaged over the 10°S–10°N band from the tropical Pacific SSTA (shade with black contours) and that from the regressed patterns from the entire domain (30°S–60°N; red contours). The right panels denote SSTA variability averaged over the 35°–45°N band derived from the entire domain (30°S–60°N; shade with black contours) and that from the North Pacific domain (20°–60°N; red contours). Contour intervals are 0.1 °C.

"projected" trends for future periods. Eq. (44) can be used in conjunction with the extended PC time series to generate "future" synthetic datasets by utilizing $B_n(r,t) = B_n(r,t+d)$. In this sense, Eq. (44) serves essentially as a statistical prediction model. Of course, an important caveat in such an exercise is that loading vectors (physical evolutions) do not change in any significant manner under a climate change scenario.

In forecast studies, $\hat{T}_n(t)$ should be calculated at a future time. In a conventional prediction approach, $\hat{T}_n(t)$ is obtained from

$$\hat{T}_n(t+h) = \sum_{m=0}^{M} w_m T_n(t-m), \quad t \in D, \quad t+h \in R, \tag{46}$$

where h is called the horizon, D implies the data domain and R implies the prediction domain (Kim and North, 1998; Kim, 2000). The optimal weight $\{w_m\}$ can be determined from the prediction normal equation (Newton, 1988). Eq. (46) can be modified as

$$\hat{T}_n(t+h) = \sum_{m=1}^{M} w_m^{(n)} P_m^{(n)}(t-\tau_m^{(n)}), \quad t \in D, \quad t+h \in R, \tag{47}$$

where $\{P_m^{(n)}(t)\}$ are predictor time series and $\{\tau_m^{(n)}\}$ are corresponding lags. We determine appropriate predictor time series and corresponding lags to make the variance of error in Eq. (47), $Var(\hat{T}_n(t) - T_n(t))$, minimized over a training period. Once all the PC time series, $\hat{T}_n(t)$, are estimated, the forecast field, $\hat{T}(r,t)$, can be obtained from Eq. (44). Some examples are presented in Kim and North (1999) and Eq. (46) was also used to forecast summer monsoon precipitations over East Asia (Lim and Kim, 2006).

It should be pointed out that relatively short temporal scales associated with individual physical processes shorten autocorrelation time scales thereby seriously limiting the predictability of geophysical phenomena. By separating short "physical" time scales from relatively long "stochastic" time scales, predictability can be improved significantly as demonstrated in Lim and Kim (2006).

The CSEOF technique can also be used to develop a downscaling method. In downscaling studies, $\hat{T}(r,t)$ over a small and high-resolution domain can be estimated based on data over a wider and coarse-resolution domain. Let us consider an experiment, in which temperature over a small domain is estimated based on a coarse model temperature over a much wider domain. By first finding the regression relationship between the target PC time series and the predictor PC time series as in Eq. (33), one can estimate the target PC time series based on the predictor PC time series. That is,

$$\hat{T}_n(t) = \sum_{m=1}^{M} \beta_m^{(n)} P_m(t) + \varepsilon^{(n)}(t), \quad t \in R \tag{48}$$

where R is the prediction interval. Then, downscaling onto a high-resolution domain, $\hat{T}(x,t)$, can be achieved by writing

$$\hat{T}(x,t) = \sum\nolimits_n B_n(x,t)\hat{T}_n(t), \tag{49}$$

where $B_n(x,t)$ are loading vectors of high-resolution data.

This approach was used for regional-scale forecasts of temperatures and precipitations based on global model forecasts (e.g., Lim et al., 2007, 2010). Namely, the PC time series of high-resolution data can be estimated for future times by using the PC time series of global model forecasts as in Eq. (48). Such statistical downscaling methods can also be useful in developing modeling strategies, in which resolving small scale climate features is essential.

Application of CSEOF analysis as a reconstruction tool is considered of increasing importance these days. A reconstruction is a way to produce complete data fields from sparse historical data. The idea is to use spatial information from a data-rich period to interpolate between sparse measurements during data-poor periods. A full theory of reconstruction algorithm is beyond the scope of this study, but a brief sketch is given instead. Let data over the training period be given as

$$X(r,t) = \sum\nolimits_n B_n(r,t)T_n(t), \quad t \in D, \tag{50}$$

where D denotes the training period. In the reconstruction period, we have only a limited number observations (say, tide gauge observations), that is,

$$X(r_j,t), \quad j = 1,\ldots,N, \quad t \in R, \tag{51}$$

where N is the number of observations and R is the reconstruction period. In order to generate data over the whole domain in the prediction period, we need to estimate $T_n(t)$ based on N observations at each time step. That is,

$$\hat{T}_n(t) = \sum\nolimits_{k=-\delta}^{\delta}\sum\nolimits_{j=1}^{N} w_{jk}B_n(r_j,t+k)X(r_j,t+k), \tag{52}$$

where $\{w_{jk}\}$ are optimal weights. Optimal weights are determined such that the variance of $\hat{T}_n(t) - T_n(t)$ is minimized over the training period.

The advantage of a CSEOF-based estimation algorithm is that data at multiple $(2\delta+1)$ time steps can be used as shown in Eq. (52). It is possible to use multi-level data, since CSEOF basis functions reflect both spatial and temporal correlation structures. In other words, data at past time steps provide useful information for estimating data at the present time step. Likewise, data at future time steps provide useful information for estimating data at the present time step. An example of CSEOF-based reconstruction is found in Hamlington et al. (2011a, 2011b), in which global sea level height from 1950 to 2010 was estimated based on a limited number of tide-gauge observations (see also http://podaac.jpl.nasa.gov/dataset/RECON_SEA_LEVEL_OST_L4_V1). In an additional recent study, CSEOF-based sea level reconstructions were also found to be superior to more widely-used EOF-based sea level reconstructions with regards to sensitivity to sparse historical observations and the ability to capture natural climate variability. By every tested metric, the CSEOF-based reconstruction outperformed the EOF-based reconstruction (Strassburg et al., 2014).

An additional advantage of the CSEOF-based estimation technique lies in the fact that it allows related physical variables to be utilized in reconstructing a specific variable. This has important implications as reconstructions extend further into the past when historical measurements become sparser and have poorer coverage of the global ocean. For example, sea surface temperatures were used together with tide gauge data to estimate sea level height from 1900 to 2010 (Hamlington et al., 2012). The resulting dataset is the longest global reconstruction of sea level height and serves as an important tool for investigating long-term variability in sea level height. Such an estimation algorithm can be used to reconstruct global patterns of variables based on a few satellite observations; at any given time, there are only a few satellite observations, which can optimally be averaged to produce data with global coverage.

8. Concluding remarks

In this study, the conceptual foundation of CSEOF analysis has been addressed in comparison with EOF analysis, which is based on the stationarity assumption. The assumption of stationarity is often not justifiable and cyclostationarity may be a better assumption for a wide range of geophysical processes. CSEOF analysis finds computational modes of a cyclostationary process, for which the first two moment statistics are periodically time dependent. The computational modes, eigenfunctions of the covariance function, then, are also time dependent and periodic with the same period of the first two moment statistics. The time dependence of covariance statistics, and henceforth the time dependence of eigenfunctions (CSLVs), comes primarily from time-dependent physical evolutions and should be distinguished from stochastic components of variability, which are described in corresponding PC time series. The time dependence of physical processes aside from stochastic undulations is clearly realized in the time dependence of corresponding covariance statistics.

Accounting for the time-dependent characteristics of physical processes is important in wide areas of climate research. Specifically, CSLVs may provide clearer pictures of physical processes extracted from complex observational and model datasets. The importance of extracting accurate physical evolution cannot be emphasized too much. More importantly, it is extremely beneficial to extract physical evolutions from multiple variables in such a way that they are physically consistent with each other. Such physical consistency is ensured in a regression analysis in CSEOF space. Physically consistent evolutions in two or more variables can be derived, since physical evolutions are not assumed to be identical between the two or more variables. In fact, such an assumption is not physically valid, for the evolution of variables in a specific physical process is determined according to the equation governing the physical process.

Being able to extract independent and accurate physical evolutions from a given dataset is a pivotal characteristic of the CSEOF technique. Not only does this lead to more accurate physical and statistical inferences on physical processes but also allows the development of more accurate statistical algorithms. The distinction between physical and stochastic components of variability is extremely crucial in addressing the teleconnection between two geographically unconnected regions, since two regions should, in general, exhibit disparate physical evolutions. Unless such non-identical physical evolutions between two places can be dealt with appropriately, addressing a teleconnection between two places is bound to be difficult and erroneous.

If the physical processes captured by a variable can be assumed to be periodic, then the PC time series of CSEOF loading vectors are stationary; therefore, spectral analysis on them makes sense. As can easily be verified, the stationarity assumption is often a poor characterization of geophysical and climatic data. Thus, EOF PC time series, which are supposed to represent stochastic components of variability, are contaminated by physical evolutions. As a result, conventional spectral analysis may lead to incorrect statistical inferences. On the other hand, CSEOF PC time series are stationary and represent the stochastic amplitude fluctuations of time-dependent physical processes. Thus, spectral analysis on CSEOF PC time series does not pose any problem.

A more accurate representation of space–time covariance statistics provides additional benefits in statistical estimation studies. Separate accounts for temporal and spatial covariance statistics should not suffice; they should be dealt with in a physically consistent manner. Distinction of physical evolutions of a deterministic nature from stochastic components of variability in CSEOF analysis provides a superior advantage in developing more accurate statistical algorithms in estimation,

prediction, detection and reconstruction. Examples addressed in Section 7 are but a few possible applications of the CSEOF technique.

As with any other analysis technique, the CSEOF analysis has obvious limitations. The most important caveat is the periodicity of the statistics. Thus, CSEOF analysis may not be suitable for analyzing physical processes without well-defined periodicity. The technique, nevertheless, is a significant improvement over existing eigen-techniques based on the stationarity assumption. Specifically, the CSEOF technique is a three-dimensional extension of the EOF technique. By relating multiple variables in a physically consistent way, the CSEOF technique further expands the analysis dimensions into four—space, time, and variables.

Acknowledgments

This research was supported by the SNU-Yonsei Research Cooperation Program through Seoul National University (SNU) in 2014.

Appendix 1. A comparison between CSEOF analysis and extended EOF analysis

The motivation of CSEOF analysis may look similar to extended EOF analysis (Weare and Nasstrom, 1982; Plaut and Vautard, 1994), i.e., extract temporal evolution of physical processes from a given dataset. There are, however, some essential differences. In a general form of extended EOF analysis, eigenfunctions are found by diagonalizing the matrix

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}(0) & \mathbf{C}(1) & \cdots & \mathbf{C}(d) \\ \mathbf{C}(1) & \mathbf{C}(0) & \cdots & \mathbf{C}(d-1) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}(d) & \mathbf{C}(d-1) & \cdots & \mathbf{C}(0) \end{pmatrix}, \tag{A1}$$

where lag-d spatial covariance matrix is given by

$$\mathbf{C}(d) = \langle T(x,t)T(x',t+d)\rangle. \tag{A2}$$

One significant difference between the two techniques is that the covariance matrix of the extended EOF technique in Eq. (A1) does not have any specific time reference. Note that Eq. (A1) is very different from Eq. (16) although Eq. (16) may collapse to Eq. (A1) under the stationarity assumption. Note in CSEOF analysis that

$$C(x,t=1;x',t'=1) \neq C(x,t=2;x',t'=2), \tag{A3}$$

both of which represent zero-lag spatial covariance function, $\mathbf{C}(0)$, in Eq. (A1). Likewise,

$$C(x,t=1;x',t'=3) \neq C(x,t=6;x',t'=8), \tag{A4}$$

although, they may be termed $\mathbf{C}(2)$ in extended EOF analysis. Further,

$$C(x,t=3;x',t'=5) \neq C(x,t=3;x',t'=1) \tag{A5}$$

may also be termed $\mathbf{C}(2)$. Bear in mind that the covariance function should be explicitly time dependent in light of Fig. 2.

Note similarly in Eq. (16) that the covariance matrix is symmetric, i.e.,

$$C(x,t;x',t') = C(x',t';x,t). \tag{A6}$$

On the other hand,

$$C(x,t;x',t') = K(x,t;x',\tau) \neq K(x,t;x',\tau+d) = C(x,t;x',t'+d), \tag{A7}$$

where $\tau = t' - t$ and $K(x,t;x',\tau)$ is the covariance function as a function of time and lag. The first and the last terms in Eq. (A7) are generally different because of different lags. This asymmetry of covariance statistics implies that there is a preferred direction of physical processes in space and time.

The cyclostationary covariance function, on the other hand, is not limited in its representation and is well defined for all combination of points t and t'. This is possible because the cyclostationary covariance function is conceptually infinite in length. A key to the implementation of the rigorous concept of cyclostationarity as described in Eqs. (15) and (16) lies in the Fourier representation of the cyclic covariance function (Kim et al., 1996). Although the covariance function repeats itself indefinitely, it can be decomposed into a finite number of Fourier expansion coefficients, the number of which is determined by the periodicity d.

References

Andreae, M.O., Jones, C.D., Cox, P.M., 2005. Strong present-day aerosol cooling implies a hot future. Nature 435, 1187–1190.
Arfken, G.B., Weber, H.J., 1995. Mathematical Methods for Physicists. 4th ed. Academic Press (1029 pp.).
Bengtsson, L., Hodges, K.I., Roeckner, E., Brokopf, R., 2006. On the natural variability of the pre-industrial European climate. Clim. Dyn. 27, 734–760.
Dee, D.P., et al., 2011. The ERA-interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137, 553–597.
Gardner, W.A., Napolitano, A., Paura, L., 2006. Cyclostationarity: half a century of research. Signal Process. 86, 639–697.
Hamlington, B.D., Leben, R.R., Nerem, R.S., Kim, K.-Y., 2011a. The effect of signal-to-noise ratio on the study of sea level trends. J. Clim. 24, 1396–1408.
Hamlington, B.D., Leben, R.R., Nerem, R.S., Han, W., Kim, K.-Y., 2011b. Reconstruction sea level using cyclostationary empirical orthogonal functions. J. Geophys. Res. 116, C12015. http://dx.doi.org/10.1029/2011JC007529.
Hamlington, B.D., Leben, R.R., Kim, K.-Y., 2012. Improving sea level reconstructions using non-sea level measurements. J. Geophys. Res. Oceans 117, C10025. http://dx.doi.org/10.1029/2012JC008277.
Hannachi, A., Jolliffe, I.T., Stephenson, D.B., 2007. Empirical orthogonal functions and related techniques in atmospheric science: a review. Int. J. Climatol. 27, 1119–1152.
Intergovernmental Panel on Climate Change (IPCC), 2007. In: Solomon, S., et al. (Eds.), Climate Change 2007: The Physical Science Basis. Cambridge University Press (996 pp.).
Intergovernmental Panel on Climate Change (IPCC), 2013. In: Stocker, T.F., et al. (Eds.), Climate Change 2013: The Physical Science Basis. Cambridge University Press (1552 pp.).
Jolliffe, I.T., 2002. Principal Component Analysis. 2nd ed. Springer (488 pp.).
Jones, A., Roberts, D.L., Slingo, A., 1994. A climate model study of indirect radiative forcing by anthropogenic sulfate aerosols. Nature 370, 450–453.
Jones, G.S., Gregory, J.M., Stott, P.A., Tett, S.F.B., Thrope, R.B., 2005. An AOGCM simulation of the climate response to a volcanic super-eruption. Clim. Dyn. 25, 725–738.
Kalnay, E., et al., 1996. The NCEP/NCAR 40-year reanalysis project. Bull. Am. Meteorol. Soc. 77, 437–471.
Kim, K.-Y., 2000. Statistical prediction of cyclostationary processes. J. Clim. 13, 1098–1115.
Kim, K.-Y., 2002. Investigation of ENSO variability using cyclostationary EOFs of observational data. Meteorol. Atmos. Phys. 81, 149–168. http://dx.doi.org/10.1007/s00703-002-0549-7.
Kim, K.-Y., Chung, C., 2001. On the evolution of the annual cycle in the tropical Pacific. J. Clim. 14, 991–994.
Kim, K.-Y., Kim, Y.Y., 2002. Mechanism of Kelvin and Rossby waves during ENSO events. Meteorol. Atmos. Phys. 81, 169–189.
Kim, K.-Y., Kim, Y.Y., 2004. Investigation of tropical Pacific upper-ocean variability using cyclostationary EOFs of assimilated data. Ocean Dyn. 54, 489–505.
Kim, K.-Y., North, G.R., 1997a. Statistical interpolation using cyclostationary EOFs. J. Clim. 10, 2931–2942.
Kim, K.-Y., North, G.R., 1997b. EOFs of harmonizable cyclostationary processes. J. Atmos. Sci. 54, 2416–2427.
Kim, K.-Y., North, G.R., 1998. EOF-based linear prediction algorithm: theory. J. Clim. 11, 3046–3056.
Kim, K.-Y., North, G.R., 1999. EOF-based linear prediction algorithm: examples. J. Clim. 12, 2077–2092.
Kim, K.-Y., Wu, Q., 1999. A comparison of study of EOF techniques: analysis of nonstationary data with periodic statistics. J. Clim. 12, 185–199.
Kim, K.-Y., North, G.R., Huang, J., 1996. EOFs of one-dimensional cyclostationary time series: computations, examples and stochastic modeling. J. Atmos. Sci. 53, 1007–1017.
Kim, K.-Y., O'Brien, J.J., Barcilon, A.I., 2003. The principal modes of variability over the tropical Pacific. Earth Interact. 7, 1–31. http://dx.doi.org/10.1175/1087-3562(2003)007<0001:TPPMOV>2.0.CO;2.
Kim, K.-Y., Park, R.J., Kim, K.-R., Na, H., 2010. Weekend effect: anthropogenic or natural? Geophys. Res. Lett. 37, L09808. http://dx.doi.org/10.1029/2010GL043233.
Kim, J.W., Kim, K.-Y., Kim, M.-K., Cho, C.-H., Lee, Y., Lee, J., 2013a. Statistical multisite simulations of summertime precipitation over South Korea and its future change based on observational data. Asia-Pac. J. Atmos. Sci. 49 (5), 687–702.
Kim, K.-Y., Kim, J.W., Kim, M.-K., Cho, C.-H., 2013b. Future trend of extreme value distributions of wintertime surface air temperatures over Korea and the associated physical changes. Asia-Pac. J. Atmos. Sci. 49 (5), 675–685.
Lim, Y.-K., Kim, K.-Y., 2006. A new perspective on the climate prediction of Asian summer monsoon precipitation. J. Clim. 19, 4840–4853.
Lim, Y.-K., Shin, D.W., Cocke, S., LaRow, T.E., Schoof, J.T., O'Brien, J.J., Chassignet, E.P., 2007. Dynamically and statistically downscaled seasonal simulations of maximum surface

air temperature over the southeastern United States. J. Geophys. Res. 112, D24102. http://dx.doi.org/10.1029/2007JD008764.
Lim, Y.-K., Cocke, S., Shin, D.W., Schoof, J.T., LaRow, T.E., O'Brien, J.J., 2010. Downscaling large-scale NCEP CFS to resolve fine-scale seasonal precipitation and extremes for the crop growing seasons over the southeastern United States. Clim. Dyn. 35, 449–471.
Loève, M., 1978. Probability Theory. vol. II. Springer-Verlag (413 pp.).
Luterbacher, J., Dietrich, D., Xoplaki, E., Grojean, M., Wanner, H., 2004. European seasonal and annual temperature variability, trends and extremes since 1500. Science 303, 1499–1503.
Na, H., Kim, K.-Y., Chang, K.-I., Kim, K., Yun, J.-Y., Minobe, S., 2010. Interannual variability of the Korea Strait Bottom Cold Water and its relationship with the upper water temperatures and atmospheric forcing in the Sea of Japan (East Sea). J. Geophys. Res. 115, C09031. http://dx.doi.org/10.1029/2010JC006347.
Na, H., Kim, K.-Y., Chang, K.-I., Park, J.J., Kim, K., Minobe, S., 2012a. Decadal variability of the upper-ocean heat content in the East/Japan Sea and its relationship to Northwestern Pacific variability. J. Geophys. Res. 117, C02017. http://dx.doi.org/10.1029/2011JC007369.
Na, H., Park, J.-H., Watts, D.R., Donohue, K.A., Lee, H.J., 2012b. Near 13-day barotropic ocean response to the atmospheric forcing in the North Pacific. J. Geophys. Res. 117, C12. http://dx.doi.org/10.1029/2012JC008211.
NAC, 1995. In: Martinson, D.G., et al. (Eds.), Natural Climate Variability on Decade-to-Century Time Scales. National Academy Press, Washington D.C. (630 pp.).
Newton, H.J., 1988. TIMESLAB: A Time Series Analysis Laboratory. Wadsworth and Brooks/Cole (63 pp.).
Plaut, G., Vautard, R., 1994. Spells of low-frequency oscillations and weather regimes over the Northern Hemisphere. J. Atmos. Sci. 51, 210–236.
Rampino, M.R., Self, S., 1992. Volcanic winter and accelerated glaciation following the Toba super-eruption. Nature 359, 50–52.
Rasmusson, E.M., Wang, X., Roplelewski, C.F., 1990. The biennial component of ENSO variability. J. Mar. Syst. 1, 71–96.
Richtmyer, R.D., 1978. Principles of Advanced Mathematical Physics vol. 1. Springer-Verlag.
Seo, K.-H., Kim, K.-Y., 2003. Propagation and initiation mechanisms of the Madden–Julian oscillation. J. Geophys. Res. 108 (4384), D13. http://dx.doi.org/10.1029/2002JD002876.
Strassburg, M.W., Hamlington, B.D., Leben, R.R., Kim, K.-Y., 2014. A comparative study of sea level reconstruction techniques using 20 years of satellite altimetry data. J. Geophys. Res. http://dx.doi.org/10.1002/2014JC009893.
Swanson, K.L., Sugihara, G., Tsonis, A.A., 2009. Long-term natural variability and 20th century climate change. Proc. Natl. Acad. Sci. 106 (38), 16120–16123.
Weare, B.C., Nasstrom, J.N., 1982. Examples of extended empirical orthogonal function analyses. Mon. Weather Rev. 110, 481–485.
Yeo, S.-R., Kim, K.-Y., 2014. Global warming, low-frequency variability, and biennial oscillation: an attempt to understand physical mechanism of the major ENSO events. Clim. Dyn. 43 (3), 768–771. http://dx.doi.org/10.1007/s00382-013-1862-1.
Yeo, S.-R., Kim, K.-Y., Yeh, S.-W., Kim, W., 2012. Decadal changes in the relationship between the tropical Pacific and the North Pacific. J. Geophys. Res. 117, D15102. http://dx.doi.org/10.1029/2012JD017775.
Zielinski, G.A., Mayewski, P.A., Meeker, L.D., Whitlow, S., Twickler, M.S., Taylor, K., 1996. Potential atmospheric impact of the Toba Mega-Eruption ~71,000 years ago. Geophys. Res. Lett. 23, 837–840.