We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Methods of Estimating Uncertainty of Climate Prediction and Climate Change Projection

Youmin Tang, Dake Chen,
Dejian Yang and Tao Lian

Additional information is available at the end of the chapter

1. Introduction

A critical issue in climate prediction and climate change projection is to estimate their uncertainty. The estimation of uncertainty has been an intensive research field in recent years, which is also called the potential predictability study. The terminology of the uncertainty of prediction and the potential predictability are often alternatively used in literature due to their inherent linkage, although they have some difference in a rigorous framework of predictability theory. For example, when a system has a high potential predictability, we may think the uncertainty of its predictions to be small, and vice versa. In this chapter, unless otherwise indicated, the uncertainty of prediction and potential predictability have the similar meaning in describing and measuring the prediction utility, and are thus used alternatively. For simplicity, we also often use the term of predictability to denote the potential predictability.

The uncertainty of prediction or predictability study is usually conducted using the strategy of ensemble prediction, from which there are a couple of metrics to quantify the potential predictability. Among them are variance-based measure and information-based measure, both quantifying the predictability or prediction uncertainty from different perspectives. In this chapter, we will introduce the two kinds of metrics. Emphasis will be placed on the similarity and disparity of these measures, and the realistic applications of the measures in studying the uncertainty of climate prediction and climate change projection. It should be noted that these potential predictability metrics do not make use of observation, which is essentially different from the actual prediction skills measured against observations like correlation skill or root mean square of errors (RMSE).

INTECH
open science | open minds

2. Two methods of measuring potential predictability

2.1. Signal-to-Noise Ratio (SNR) and potential predictability

The SNR has been a widely used measure of potential predictability [1, 2]. At seasonal time scale, the signal is usually regarded as the atmospheric responses to the slowly varying external forcing such as sea surface temperature (SST), sea ice, snow cover, etc., whereas the noise is induced by the relatively high frequency atmospheric variability such as weather processes. In an ensemble seasonal climate prediction, the amplitude of signal and noise can be approximately quantified by the variance of ensemble mean and the averaged ensemble spread over all initial conditions [3-5], namely,

$$Var(S) = \frac{1}{M} \sum_{i=1}^{M} \left(\overline{X_i} - \overline{\overline{X}} \right)^2 \tag{1}$$

$$Var(N) = \frac{1}{MK} \sum_{i=1}^{M} \sum_{j=1}^{K} \left(X_{i,j} - \overline{X_i} \right)^2 \tag{2}$$

where $X_{i,j}$ is the j-th member of the ensemble prediction starting from the i-th initial condition. The X itself can be a scalar such as an index or a two dimensional field. K is the ensemble size and M is the total number of initial conditions (predictions); and $\overline{X_i} = \frac{1}{K} \sum_{j=1}^{K} X_{i,j}$, $\overline{\overline{X}} = \frac{1}{M} \sum_{i=1}^{M} \overline{X_i}$.

Considering the sampling errors in estimating signal variance, the more accurate estimation of signal variance S is modified as below:

$$Var(S) = \frac{1}{M} \sum_{i=1}^{M} \left(\overline{X_i} - \overline{\overline{X}} \right)^2 - \frac{1}{K} Var(N)$$

Two common measures of potentially predictability are the signal-to-noise ratio (SNR) and the signal-to-total variance ratio (STR), i.e.,

$$SNR = \frac{Var(S)}{Var(N)}, \quad STR = \frac{Var(S)}{Var(S) + Var(N)} \tag{3}$$

It can be derived that the square root of *STR* is equivalent to the correlation of the signal component (S) to the prediction target itself. Thus, the $\sqrt{STR}$ is often defined as potential correlation (PCORR).

It is easy to derive that the $\sqrt{STR}$ is actually a perfect correlation skill, which assumes that the observation is an arbitrary ensemble member. The perfect correlation skill ignores the imperfectness of model itself. To see this equality, we denote the ensemble mean μ as the prediction, thus the 'observation' can be written by $\mu + \varepsilon$, where the ε is a normally distributed white noise with the mean of zero and variance of σ_e^2.

The correlation of prediction against the 'observation' can be written as follows:

$$Corr_{pef} = \frac{E[\mu(\mu+\varepsilon)]}{\{E(\mu^2)E[(\mu+\varepsilon)^2]\}^{1/2}} = \frac{E(\mu^2)}{\sqrt{E(\mu^2)E[\mu^2+2\mu\varepsilon+\varepsilon^2]}} = \frac{\sqrt{E(\mu^2)}}{\sqrt{E(\mu^2)+E(\varepsilon^2)}} \tag{4}$$

Comparison between (3) and (4) reveals the equality of $Corr_{pef}$ and $\sqrt{STR}$.

2.2. Information-based potential predictability

2.2.1. Relative Entropy and predictive information

Entropy is a measure of dispersion level (e.g. uncertainty). The entropy of a continuous distribution $p(x)$ is defined as

$$H(x) = -\int p(x)\ln p(x)dx,$$

where the integral is understood to be a multiple integral over the domain of x. Larger entropy is associated with smaller probability and larger uncertainty.

The information-based potential predictability measures include relative entropy (RE), predictive information (PI) and predictive power (PP). The central idea of these information-based measures is that the difference between two probability distributions: the forecast distribution and climatology distribution, quantifies the extra information brought from the prediction.

Suppose that the future state of a climate variable is predicted/modeled as a random variable denoted by v with a climatological distribution $p(v)$. One ensemble prediction produces a forecast distribution which is the conditional distribution $p(v|i)$ given the initial condition i. The climatological distribution is also the unconditional distribution and we have

$$p(v) = \int p(v|i)p(i)di, \tag{5}$$

where the $p(i)$ is the probability distribution of the initial condition i. Usually various statistical tests are used to examine the difference between two distributions [6-7]. Relative entropy RE, or Kullback-Leibler distance, is a quantitative measure of the difference between two distributions from information theory [8]. In the context of predictability, it is defined as

$$RE = \int p(v|i)\ln \frac{p(v|i)}{p(v)}dv . \tag{6}$$

In terms of information theory, the quantity RE measures the informational inefficiency of using the climatological distribution $p(v)$ rather than the forecast distribution $p(v|i)$ and $RE \geq 0$ with

the equality if and only if $p(v \mid i) = p(v)$ [8]. In Bayesian terminology, the climatological distribution is *a prior* distribution which can be usually derived from the long term historical observations. An ensemble prediction augments this prior information, and the additional information measured by RE is a natural measure of the utility or usefulness of this prediction and thus implies the potential predictability. In practice, $p(v \mid i)$ and $p(v)$ can be estimated directly from samples or approximated alternatively using kernel density estimation.

Another natural measure of predictability is the predictive information (PI), defined as the difference between the entropy of the climatological and forecast distributions:

$$PI = H(v) - H(v \mid i) \tag{7}$$

Considering (7), then

$$PI = -\int p(v) \ln[p(v)] dv + \int p(v \mid i) \ln[p(v \mid i)] dv \tag{8}$$

The first term on the right hand side of Eq. (*8*) denotes the entropy of the prior distribution $p(v)$ (climatological distribution), measuring the uncertainty of a prior time when no extra information is provided from the observed initial condition and forecast model; whereas the second term represents the entropy of the posterior distribution $p(v \mid i)$ (forecast distribution), measuring the uncertainty after the observed initial condition and subsequent prediction becomes available (An elaborated illustration can be found in [9]). Thus a large PI indicates that the posterior uncertainty will decrease because of useful information being provided by a prediction (e.g., the larger $p(v \mid i)$ the smaller uncertainty) that is, the prediction is to be more reliable in a "perfect model" context.

The predictive power (PP) was defined by [10]

$$PP = 1 - \exp(-PI) \tag{9}$$

In the case where the PDFs are Gaussian distributions, which is a good approximation in many practical cases (including ENSO prediction, e.g., [11]). The predictive and climatological variances, and the difference between their means. The resulting analytical expression for the relative entropy *RE, PI and PI* are given as follows [1]:

$$RE = \frac{1}{2} \left\{ \ln \left[\frac{\det(\Sigma_q^2)}{\det(\Sigma_p^2)} \right] + trace \left[\Sigma_p^2 (\Sigma_q^2)^{-1} \right] + (\mu_p - \mu_q)^T (\Sigma_q^2)^{-1} \left(\mu_p - \mu_q \right) - n \right\} \tag{10}$$

$$PI = \frac{1}{2} \ln \left(\frac{\det(\Sigma_q^2)}{\det(\Sigma_p^2)} \right) \tag{11}$$

$$PP = 1 - \left(\frac{\det(\Sigma_q^2)}{\det(\Sigma_p^2)} \right)^{-\frac{1}{2}} \tag{12}$$

where, q and P are the climatological and predictive covariance matrices respectively; det is the determinant operator and tr is the trace operator; μ_q and μ_p are the climatological and predictive mean state vectors of the system, and n is the number of degree of freedom;. RE is composed of two components: (i) a reduction in climatological uncertainty by the prediction [the first two terms plus the last term on the right-hand side of (10)] and (ii) a difference in the predictive and climatological means [the third term on the rhs of (10)]. These components can be interpreted respectively as the dispersion and signal components of the utility of a prediction[12]. A large value of RE indicates that more information that is different from the climatological distribution is being supplied by the prediction, which could be interpreted as making it more reliable [1]. A key difference between relative entropy (RE) and predictive information (PI) is that RE vanishes if and only if the forecast and climatological distributions are identical (i.e., same mean and spread), while PI is zero as long as the two distributions have the same spread [9]. Remarkably, predictive information and relative entropy are invariant with respect to linear invertible transformations of the state [9-10].

For a scalar variable (e.g., an index), RE, PI, and PP can be simplified as

$$PI = \frac{1}{2}\ln\left(\frac{\sigma_q^2}{\sigma_p^2}\right) \tag{13}$$

$$RE = \frac{1}{2}\left[\underbrace{ln\left(\frac{\sigma_q^2}{\sigma_p^2}\right) + \frac{\sigma_p^2}{\sigma_q^2} - 1}_{Dispersion} + \underbrace{\frac{\mu_p^2}{\sigma_q^2}}_{Signal}\right] = PI + \frac{1}{2}\left[\frac{\sigma_p^2}{\sigma_q^2} - 1 + \frac{\mu_p^2}{\sigma_q^2}\right] \tag{14}$$

$$PP = 1 - \left(\frac{\sigma_p^2}{\sigma_q^2}\right)^{1/2} \tag{15}$$

2.2.2. Mutual information

RE or PI is a predictability measure for individual predictions. The average of REs or PIs over all initial conditions reflects the average predictability and was proved to be equal to mutual information (MI), another quantity from information theory [9]. In the context of predictability, MI is defined as [9]

$$MI = \iint p(v,i)\ln[\frac{p(v,i)}{p(v)p(i)}]dvdi \tag{16}$$

where $p(v,i)$ is the joint probability distribution between v and i. MI measures the statistical dependence between v and i, and vanishes when v and i are independent ($p(v,i) = p(v)p(i)$).The equality of MI and average RE indicates that predictability can be

measured in two equivalent ways: by the difference between forecast and climatological distributions or by the degree of statistical dependence between the initial condition i and the future state v [13]. If the future state v is on average unpredictable, individual forecasts should have the probability distribution identical to the climatological distribution, i.e., $p(v \mid i) = p(v)$ and RE =0 for all predictions. This is equivalent to independence between i and v. Therefore, independence indicates unpredictability and dependence implies predictability. MI is invariant with respect to nonlinear, invertible (nonsingular) transformations of state[9]. Thus, the MI between v and i equals to the MI between v and ensemble mean $\mu_{v|i}$. The latter is probably more straightforward in understanding MI-based predictability since the dependence between v and ensemble mean $\mu_{v|i}$ can be interpreted as the dependence between observation (v) and prediction ($\mu_{v|i}$) under the assumption of a perfect model.

When forecast and climatological distributions are Gaussian, MI can be expressed, using (13), by[13]

$$MI = \frac{1}{2}\left(\ln \sigma_v^2 - \left\langle \ln \sigma_{v|i}^2 \right\rangle\right) \tag{17}$$

Eq. (17) is the formula often used to calculate MI. *Joe* [14] and *DelSole* [15] showed that the transformations $\sqrt{1 - \exp(-2MI)}$ and $1 - \exp(-2MI)$ produce "potential" skill scores which exhibit proper limiting behavior: they have values between 0 and 1, and the minimum (maximum) value 0 (1) occurs when MI vanishes (approaches infinite). Here "potential" indicates that they are perfect model measures. In this study, we will use the two "potential" skills to represent MI. Furthermore, if the forecast and climatological distributions are Gaussian, and forecast variance is constant, the above two "potential" skills respectively reduce to another two conventional "potential" skills: "potential" anomaly correlation (AC_p) and "potential" mean square skill score ($MSSS_p$)[9,13].

$$AC_p = \sqrt{1 - \exp(-2MI)}, \tag{18}$$

$$MSSS_p = AC_p{}^2 = 1 - \exp(-2MI) \tag{19}$$

2.3. Relationship between SNR-based metrics and MI-based metrics

The averaged RE and PI ($\overline{RE}$ and $\overline{PI}$) over all predictions (initial conditions) are identical to MI, as mentioned before. For seasonal climate prediction, the total variance (i.e., climate variance) can be decomposed into signal (S) variance and noise (N) variance, if the signal and noise are assumed to be independent of each other ([16-17]), namely,

$$Var(T) = Var(S) + Var(N) \tag{20}$$

where

$$Var(T) = \frac{1}{MK}\sum_{i=1}^{M}\sum_{j=1}^{K}\left(X_{i,j} - \bar{\bar{X}}\right)^2,$$

$$var(S) = \frac{1}{M}\sum_{i=1}^{M}(\bar{X_i} - \bar{\bar{X}})^2$$

$$Var(N) = \frac{1}{MK}\sum_{i=1}^{M}\sum_{j=1}^{K}\left(X_{i,j} - \bar{X_i}\right)^2.$$

$X_{i,j}$ is the j-th member of the ensemble prediction starting from the i-th initial condition. The K is the ensemble size and M is the total number of initial conditions (predictions); and $\bar{X_i} = \frac{1}{K}\sum_{j=1}^{K}X_{i,j}$, $\bar{\bar{X}} = \frac{1}{M}\sum_{i=1}^{M}\bar{X_i}$.

Without the loss of generality, the climatological mean is assumed to be zero, thus (20) can be expressed by

$$\overline{\sigma_q^2} = \overline{\mu_p^2} + \overline{\sigma_p^2} \tag{21}$$

where the overbar denotes the expectation over all predictions (initial conditions).Eq (14) and Eq. (21) can easily verify the property of MI, for example,

$$MI = \overline{RE} = \overline{PI} + \frac{1}{2}\left[\frac{\overline{\mu_p^2} + \overline{\sigma_p^2} - \overline{\sigma_q^2}}{\overline{\sigma_q^2}}\right] = \overline{PI} \tag{22}$$

Using (21), the information-based potential predictability measures MI, ($\overline{RE}$ or $\overline{PI}$) and $\overline{PP}$ can be rewritten as the function of the mean signal and noise, or their ratio SNR. The σ_q and σ_p in (21) are actually σ_v and $\sigma_{v|i}$ in (17), thus we have [18]

$$MI = \frac{1}{2}\left(\ln\sigma_v^2 - \langle\ln\sigma_{v|i}^2\rangle\right) \geq \frac{1}{2}\left(\ln\sigma_v^2 - \ln\langle\sigma_{v|i}^2\rangle\right)$$
$$= -\frac{1}{2}\ln\left(\frac{\langle\sigma_{v|i}^2\rangle}{\sigma_v^2}\right) = -\frac{1}{2}\ln(\frac{\overline{\sigma_p^2}}{\overline{\sigma_q^2}}) = -\frac{1}{2}\ln(1 - STR) \tag{23}$$

The inequality in (23) is due to the fact that arithmetic mean is larger than or equal to geometric mean, or more strictly is a result of Jensen's inequality from information theory. Therefore, we have

$$AC_p \leq \sqrt{1 - \exp(-2MI)}, \tag{24}$$

$$MSSS_p \leq 1 - \exp(-2MI). \tag{25}$$

The equalities in (24), (25) and (23) hold if and only if $\sigma^2_{v|i}$ is constant, as addressed in (18) and (19). The conditions that the forecast and climatological distribution are both Gaussian and the forecast variance $\sigma^2_{v|i}$ is constant are equivalent to the condition that i and v are joint normally distributed [13,18]. If i and v are joint normally distributed, the probability distributions $p(i)$, $p(v)$ and $p(v|i)$ are all Gaussian distributions and there are [13,18-19]

$$\mu_{v|i} = \rho_0 \frac{\sigma_v}{\sigma_i} \left(i - \langle i \rangle \right), \tag{26}$$

$$\sigma^2_{v|i} = (1 - \rho_0^2)\sigma_v^2 = cons\tan t, \tag{27}$$

$$MI = -\frac{1}{2}\ln(1 - \rho_0^2), \tag{28}$$

where(26) is obtained using a linear regression with ρ_0 being the linear correlation between the initial state i and the future state v. As mentioned earlier, MI measures the statistical dependence between i and v. As can be seen from (26), the statistical dependence reduces to a linear correlation ρ_0 if the two variables are joint normally distributed. Because conditional mean $\mu_{v|i}$ is a linear function of initial state i (see (26)), ρ_0 is also the linear correlation between $\mu_{v|i}$ and v, which is the potential anomaly correlation skill AC_p. Note that if i and v are not joint normally distributed, ρ_0 is usually different from AC_p.

One interesting question arises here, namely that, how we understand the MI-SNR discrepancy when there is significant variability of prediction variance, as expressed in (24) and (25)? As discussed earlier, the MI-based potential predictability measures the statistical dependence, liner or nonlinear, between the ensemble mean prediction $\mu_{v|i}$ and the hypothetical observation v (an arbitrary ensemble member), whereas the SNR-based potential skill only measures their linear correlation. When $\mu_{v|i}$ and v are joint normally distributed, their statistical dependence reduces to linear correlation. When $\mu_{v|i}$ and v are not joint normally distributed, MI naturally disagrees with SNR. The joint normally distributed variables have constant conditional variance. Note that prediction variance is also the conditional variance of v given the ensemble mean $\mu_{v|i}$. Thus, if the prediction variance is varied, $\mu_{v|i}$ and v are definitely not joint normally distributed, making the SNR-based potential skill, a linear correlation between $\mu_{v|i}$ and v, underestimate the nonlinear statistical dependence between $\mu_{v|i}$ (or i) and v, which is a strict statistical definition of potential predictability.

It should be noted that the above conclusion should not be challenged by a possible fact that SNR-based skill might have a better relationship to actual skill than MI-based skill, simply

because the actual skill is often measured by the linear correlation (or related quantity), which is inherent to the SNR-based skill. Thus, a more challenging issue is how to design new metrics to measure actual forecast skill which could appreciate the MI-based extra predictive information beyond SNR. In principle, the MI between ensemble mean prediction $\mu_{v|i}$ and actual observation O could have the potential capability to quantify the MI-based potential predictability [15]. However, how to effectively estimate MI in this context is not an easy issue.

In summary, there are connections between information-based potential predictability and SNR-based potential predictability, as built by the above equations. In other words, all the averaged information-based potential predictability measures are better than SNR-based predictability in characterizing 'true' potential predictability. When the climatology and prediction distribution are both Gaussian and the prediction variances are constant, the information-based measure is equivalent to the SNR-based potential measure.

3. Maximum SNR and PrCA

The signal and noise are theoretically statistically irrelevant when the ensemble size is infinite. However, the ensemble size is always finite in reality, thus the estimation of the signal is often contaminated by the noise. An optimal estimate for the largest potential predictability should be to maximize the SNR, from which the resultant signal component is the most predictable.

We denote by $\mathbf{S}$ and $\mathbf{N}$ signal and noise of variable $\mathbf{X}$, where $\mathbf{S}$ and $\mathbf{N}$ are matrixes of a two-dimension field describing temporal and spatial variation of the signal and noise of one variable of interest, namely, this section is at the framework of the multivariate statistics.

where, $\mathbf{S} = \bar{X}_i - \bar{\bar{X}}$; $\mathbf{N} = X_{i,j} - \bar{X}_i$

$X_{i,j}$ is the j-th member of the ensemble prediction starting from the i-th initial condition. The K is the ensemble size and M is the total number of initial conditions (predictions); and $\bar{X}_i = \frac{1}{K}\sum_{j=1}^{K} X_{i,j}$, $\bar{\bar{X}} = \frac{1}{M}\sum_{i=1}^{M}\bar{X}_i$.

Our goal is to look for a vector q, which can maximize the ratio of the variance of signal and noise that are projected onto the vector, namely,

$$r_S = q^T * \mathbf{S}; \ r_N = q^T * \mathbf{N}$$

$$\text{SNR} = \frac{\sigma^2_{r_S}}{\sigma^2_{r_N}} \quad ===> \quad \max \tag{29}$$

where

$$\sigma_{r_S}^2 = \frac{1}{M} r_S * r_S{}^T = \frac{1}{M} q^T S * S^T q = q^T \Sigma_S q$$

$$\sigma_{r_N}^2 = \frac{1}{KM} r_N * r_N{}^T = \frac{1}{KM} q^T N * N^T q = q^T \Sigma_N q$$

Mathematically such an optimization by (29) leads to a generalized eigenvalue-eigenvector problem based on the Rayleigh Quotient theorem[10,20]

$$q^T \Sigma_S = \lambda q^T \Sigma_N \tag{30}$$

Where, Σ_S is the covariance matrix of signal, Σ_N is the covariance matrix of noise. The solution of (30) can be obtained by solving the below eigenvalue equation

$$q^T \Sigma_S \Sigma_N^{-1} = \lambda q^T \tag{31}$$

Thus, the analysis of the largest potential predictability is also called the maximum signal-to-noise EOF (MSN EOF) analysis, first introduced by Allen and Smith [21] to estimate the signal optimally by suppressing the influences of noise, and widely used already in climate predictability study [20,22-23].

Practically, the number of grid points is always much larger than the number of total samples in climate studies, thus usually Σ_N doesn't have full-rank, leading to a solution of ill-conditioned inversions. There are two common methods to solve this issue, as introduced below.

3.1. The SNR is optimized in a truncated EOF space

Denote by e_i^T (i=1,2…,k) the EOF modes[1] , and the signal and noise components projected on them are

$$T_S = e_i^T * S$$

$$T_N = e_i^T * N \tag{32}$$

T_S and T_N are PC components with dimension of k*n where the n is the time samples. Thus, the signal and noise variance used in (31) should be calculated by $\Sigma_S = T_S * T_S^T$, $\Sigma_N = T_N * T_N^T$, respectively. If k truncated modes remain, where k is usually much smaller than the number of spatial grids, the signal and noise covariance matrix is a full-rank of k matrix. Thus, eq. (31) can be easily solved and the vector q (denoted as q_{eof}) is called filter pattern, *which is* a k-element vector, the filter pattern on the truncated EOF space. The leading predictable component is

---

[1] e_i is a matrix of m*k, where m is the number of spatial grids and k is the number of the truncated modes. EOF could be employed using signal, or noise matrix or corresponding data matrix.

$$r_S = q_{eof}{}^T * T_S \tag{33}$$

If projecting q_{eof} back to data space, we have

$$\Pr C\_data = q^T * S = q_{eof}^T * e^T * S = q_{eof}{}^T * T_S \tag{34}$$

Eq. (33) and (34) are identical each other, i.e., PrC is invariant with respect to a linear transformation.

The q is the filter pattern, rather than the most predictable pattern. The most predictable pattern v can be obtained using the regression method, i.e.,

$$N = r_N * V \tag{35}$$

$$V = N * r^T{}_N / (r_N * r_N{}^T) = \frac{1}{MK} N * r^T{}_N$$

$$= \frac{1}{MK} N * T_N{}^T * q$$

$$= \frac{1}{MK} N * N^T * e * q$$

$$= \Sigma_N * e * q$$

Also it can be written by

$$S = r_S * V \tag{36}$$

$$V = S * r_S^T / r_S * r_S^T =$$

$$\frac{1}{M\lambda} S * r_S^T =$$

$$\frac{1}{M\lambda} S * S^T * e * q = \frac{1}{\lambda} \Sigma_S * e * q$$

$$= \Sigma_N * e * q$$

3.2. Solving eq (29) using whitening approach

The approach is to whiten the noise variance (i.e., the denominator), making Σ_N an identity matrix and whitening the covariance matrix of signal(Σ_S) simultaneously. Thus, eq. (29) becomes

$$SNR = \frac{q^T \Sigma_S q}{q^T \Sigma_N q} = \frac{q'^T \Sigma_{WS} q'}{q'^T q'} \rightarrow \max imum \tag{37}$$

Based on the matrix theory, the SNR in eq. (37) reaches maximum when q' is the eigenvector of Σ_{WS}, the whitened signal covariance associated with the whitening noise Σ_N. The q' is a modified q by a whitening factor. The algorithm is briefly summarized as follows:

i.   Make the covariance matrix of noise (Σ_N) identity, namely,

$$D^{-1/2}E^T\Sigma_N ED^{-1/2} = I \tag{38}$$

D and E are the eigenvalue and eigenvector matrices of Σ_N. $ED^{-1/2}$ is the transformation matrix that makes the covariance matrix of noise (Σ_N) identity.

ii. Whiten the signal covariance matrix by the transformation matrix $ED^{-1/2}$, using the k leading modes

$$\Sigma_{WS} = D^{-1/2}E^T\Sigma_S ED^{-1/2} \tag{39}$$

iii. The SNR of (*37*) reaches maximum when q' is the leading eigenvector of the whitened signal covariance matrix Σ_{WS} (in descent order). It is easy to see the relationship between q and q'

$$q'^T \Sigma_{WS}q' = q'^T D^{-1/2}E^T\Sigma_S ED^{-1/2}q' = q^T\Sigma_S q \tag{40}$$

Thus,

$$q = ED^{-1/2}q'$$

iv. After the filter pattern q is known, the most predictable component is easy to derive as shown in method 1, namely projecting the signal (ensemble mean) on the filter patterns

$$\begin{aligned} \mathrm{Pr}Cs\_s &= q^T S \\ \mathrm{Pr}Cs\_n &= q^T N \\ \mathrm{Pr}Cs\_t &= q^T T \end{aligned} \tag{41}$$

The most predictable component is the one corresponding to the largest signal-to-noise ratio. All PrCs are temporally orthogonal (uncorrelated) with each other. It is noted that (*41*) is a little different from (*33*) or (*34*) where the PCs of truncated EOF spaces are used. It is due to a different truncation procedure in the two methods. In this first method, the truncation is applied before optimization whereas in the second method, the truncation is implicitly integrated into the whitening process. However both should be equivalent, which can be seen by another expression of (*41*)

$$\mathrm{Pr}Cs\_s = q^T S = q'^T D^{-1/2}E^T S = q'^T D^{-1/2}T_S$$

v.   Obtain the corresponding predictable patterns V by,

$$V = N * \mathrm{Pr}Cs\_n^T / \mathrm{Pr}Cs\_n * \mathrm{Pr}Cs\_n^T = \frac{1}{MK} N * \mathrm{Pr}Cs\_n = \frac{1}{MK} N * N_N^T * q = \Sigma_N q$$

A reconstructed forecast based on PrCS leading modes can be obtained by

$$
\begin{aligned}
\hat{N} &= V * \mathrm{Pr}Cs\_n \\
\hat{S} &= V * \mathrm{Pr}Cs\_s \\
\hat{X} &= V * \mathrm{Pr}Cs\_T
\end{aligned}
\tag{42}
$$

$\hat{X}$ only remains the leading PrCS modes and removes noise components, thus it can be expected to have a better skill than simple ensemble mean.

The variance explained by a PrCA mode can be obtained using (42). If all modes are remained in (42), the reconstructed filed should explain 100% of original field. We rewrite (42), applied into signal and noise, thus,

$$
\Sigma_S = \frac{1}{M}S^T {*} S; \quad \hat{\Sigma}_S = \frac{1}{M}V{*}\mathrm{r}_S{}^T * \mathrm{r}_S * V^T = V\Lambda V^T = \sum_{j=1} v_j * \lambda_j {*} v_j^T
$$

$$
\Sigma_N = \frac{1}{MK}N^T * N; \quad \hat{\Sigma}_N = \frac{1}{MK}V * r_N^T * r_N * V^T = V * V^T = \sum_{j} v_j * v_j^T
$$

Where the $\hat{\Sigma}$ is the estimated variance using PrCA modes. Thus, the variance explained by a specific mode measured in the original space, and the truncated space, is respectively as below:

$$
\begin{aligned}
&\text{relative to signal}: \quad \frac{\lambda_j v_j * v_j{}^T}{tr(\Sigma_S)} \quad ; \quad \frac{\lambda_j v_j * v_j^T}{tr(\hat{\Sigma}_S)} \\[2mm]
&\text{relative to noise}: \quad \frac{v_j * v_j^T}{tr(\Sigma_N)} \quad ; \quad \frac{v_j * v_j^T}{tr(\hat{\Sigma}_N)} \\[2mm]
&\text{relative to total variance}: \quad \frac{\lambda_j v_j * v_j^T + v_j * v_j^T}{tr(\Sigma_S) + tr(\Sigma_N)} \quad ; \quad \frac{\lambda_j v_j * v_j^T + v_j * v_j^T}{tr(\hat{\Sigma}_S) + tr(\hat{\Sigma}_N)}
\end{aligned}
$$

4. Maximizing PI and PrCA

Another interpretation to MSN EOF is its connection with information-based measure PI or PP defined in (11) - (15). For example, as argued in [10], the predictive power PP is a positively oriented predictive index, defined by the difference between posterior (prediction) entropy and prior (climatology) entropy, thus measuring the decrease of uncertainties due to prediction.

The PrCA analysis is an approach to maximize PI, or maximum PP, equivalent to minimizing σ_p^2 / σ_q^2 if the prediction variance is little changed, to derive the most predictable component. The σ_q^2 is climatology variance, often referred as to the total variance Var (T), which is composed of the signal variance and noise variance. Under the 'perfect model' assumption, the noise variance equals to the forecast error variance [24], namely,

$$\sigma_q^2 = Var(S) + Var(N) = Var(S) + \sigma_p^2 \tag{43}$$

Thus, the minimization of σ_p^2 / σ_q^2 is equal to the maximization of $1 - \sigma_p^2 / \sigma_q^2$, i.e., STR, which is equivalent to the maximization of SNR, i.e., MSN EOF. In some literature, the term of MSN EOF and PrCA are alternatively used due to their complete equivalence. Actually, both the MSN EOF and PrCA methods belong to the discriminant analyses because the two methods, though from different perspectives, can be understood to seek a best linear combination of variables that separates the signal and the noise as much as possible [13]. The both methods identify the "filter pattern", or weight matrix, providing an optimized filter to discriminate the signal and noise, where the time series reflects the temporal evolution of the dominant mode of the signal, and the spatial pattern characterizes the spatial distribution of the dominant mode of signal, which are respectively referred to as spatial pattern, or the most predictable pattern.

It should be noted that the equivalence of SNR-based and information-based PrCA approach is based on the condition that the climatology and forecast distribution are both Gaussian. It is apparent since the PI and PP cannot be only expressed by the form of prediction and climatology variance as (11) – (15) under non-Gaussian assumption. It is difficult to derive the optimization solution for PI or PP from their general definitions of (8) and (9).

A remark to the algorithm of Schneider and Griffies [10] is a technical issue. In Schneider and Griffies [10], the PrCA is proposed to derive by minimizing PP, i.e., minimizing σ_p^2 / σ_q^2, leading to the below eigenvalue equation:

$$f^T \Sigma_N \Sigma_T^{-1} = \xi f^T \tag{44}$$

where Σ_T is the total variance. The optimal filter resulting in the most predictable is the eigenvector f with the smallest eigenvalue ξ. Comparing (44) with (31) reveals that eigenvalue ξ and q are reciprocal, indicating the equivalence of PrCA using maximization of (31) and minimization of (44). Usually, the eigenvector with the smallest eigenvalue often lacks of a stable, large scale-like pattern, making the approach of (44) impractical in real application. The truncated EOF space, which is used in solving (31) and (44), can greatly reduce this concern but still the most predictable pattern contains some noise. Thus, the MSN EOF approach, introduced above, is a better option.

5. A practical application – Potential predictability of climate change projection in AR5

In this section, we will explore the uncertainty of climate change projection using the above theoretical framework. The estimation of uncertainty is based on the Coupled Model Intercomparison Project Phase 5 (CMIP5), a new set of climate model experiments involved in the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5). The CMIP5 is promoted to address some crucial issues on climate modeling and future

climate state. More than 20 climate models were employed in this project with main focus on: 1) evaluate model predictability of future climate on different time scales (near term (out to about 2035) and long term (out to 2100 and beyond)), 2) understanding key mechanisms responsible for differences in model projections, and 3) quantify some important feedbacks of climate system like clouds and carbon cycle.

One of experiments used in CMIP5 is the Representative Concentration Pathways (RCPs) scenario. All model experiments involved in this scenario are forced by four kinds of mixing greenhouse gases (GHGs) boundary conditions which will finally lead increasing of radiation by 2.6, 4.5, 6.0 and 8.0 watt per square at the end of 21 century.

In this chapter, we will use the sea surface temperature (SST) projection of scenario R60 (the increasing of 6.0 watt per square experiment) to evaluate the potential predictability of climate projection of the scenario R60. At present, only nine models collected in R60 are available to download (From ESG-PCMDI Gataway), as summarized in table 1.

Model	Country	Ocean Model Resolutions	Projection
CCSM4-version16	USA (NCAR)	60 levels; 1.0 lon. x 0.5 lat.	2051-2100
CSIRO-MK3.6.0	Australia (BMRC)	30 levels; 1.875 lon x 0.9375lat.	2051-2100
GISS-E2-R	USA (NASA)	32 levels;1 lon x 1.25 lat.	2051-2100
GFDL-ESM2M	USA (GFDL)	50 levels; 1- 1/3 lon. x 1 lat.	2051-2100
HadGEM2-ES	UK (Hadley Center)	40 levels;1-1/3 lon. x 1 lat.	2051-2100
CM5A-LR	France (IPSL)	ORACA2 resolution in OPA	2051-2100
MIROC5-Coco 4.5	Japan	Varied resolution	2051-2100
MRI-CGCM3	Japan	Varied resolution	2051-2100
NorESM1-M	Norway	Varied resolution	2051-2100

Table 1. Models used for evaluation

The SST outputs from these models are all monthly averaged data. For the purpose of the study of the climate change, we use annual mean in the following discussions. Because the lack of uniformity of ensemble member, only one member is used for each model here. In this study, we confine the domain to the Pacific across 60S to 60N.

Shown in Fig. 1 and Fig 2a are the spatial pattern and time series of the first EOF (Empirical Orthogonal Function) for the Pacific Ocean from 2051-2100. As can be seen in Fig.2a, the Pacific SST has a striking increase, with the strongest response to the forcing of GHGs in the tropical Pacific along the equator as shown in Fig.1. In the extra-tropical beyond the 30S and 30N, the increase in SST is relatively weaker. On average, the mean temperature of the Pacific ocean of 60S to 60N increases around 0.5 to 1C from 2051-2010 in these models, as shown in Fig. 2b, the evolution of the mean temperature over the Pacific ocean. The mean of multiple models has the increase rate of around 0.75C as shown by the red line in Fig. 2b.
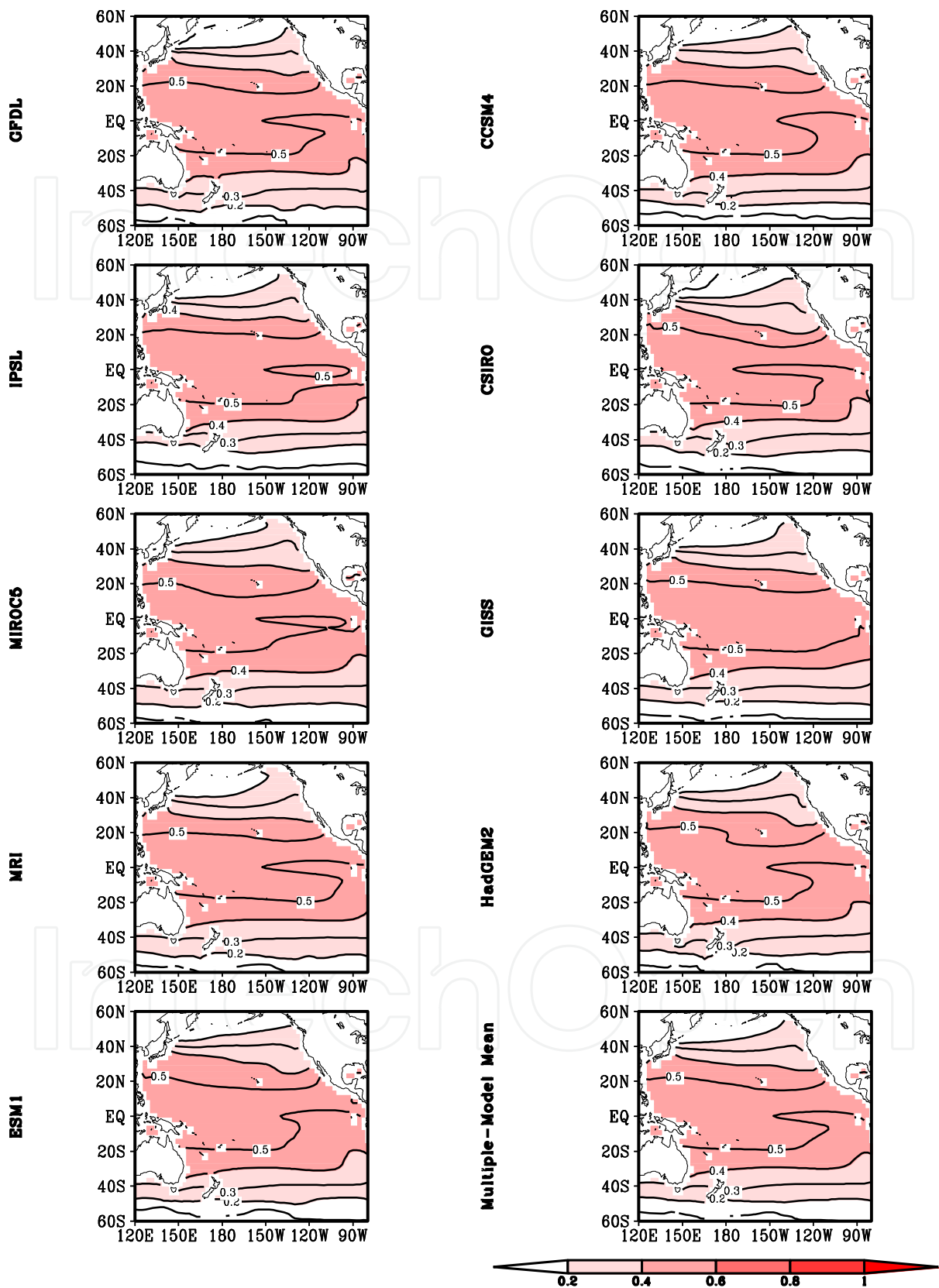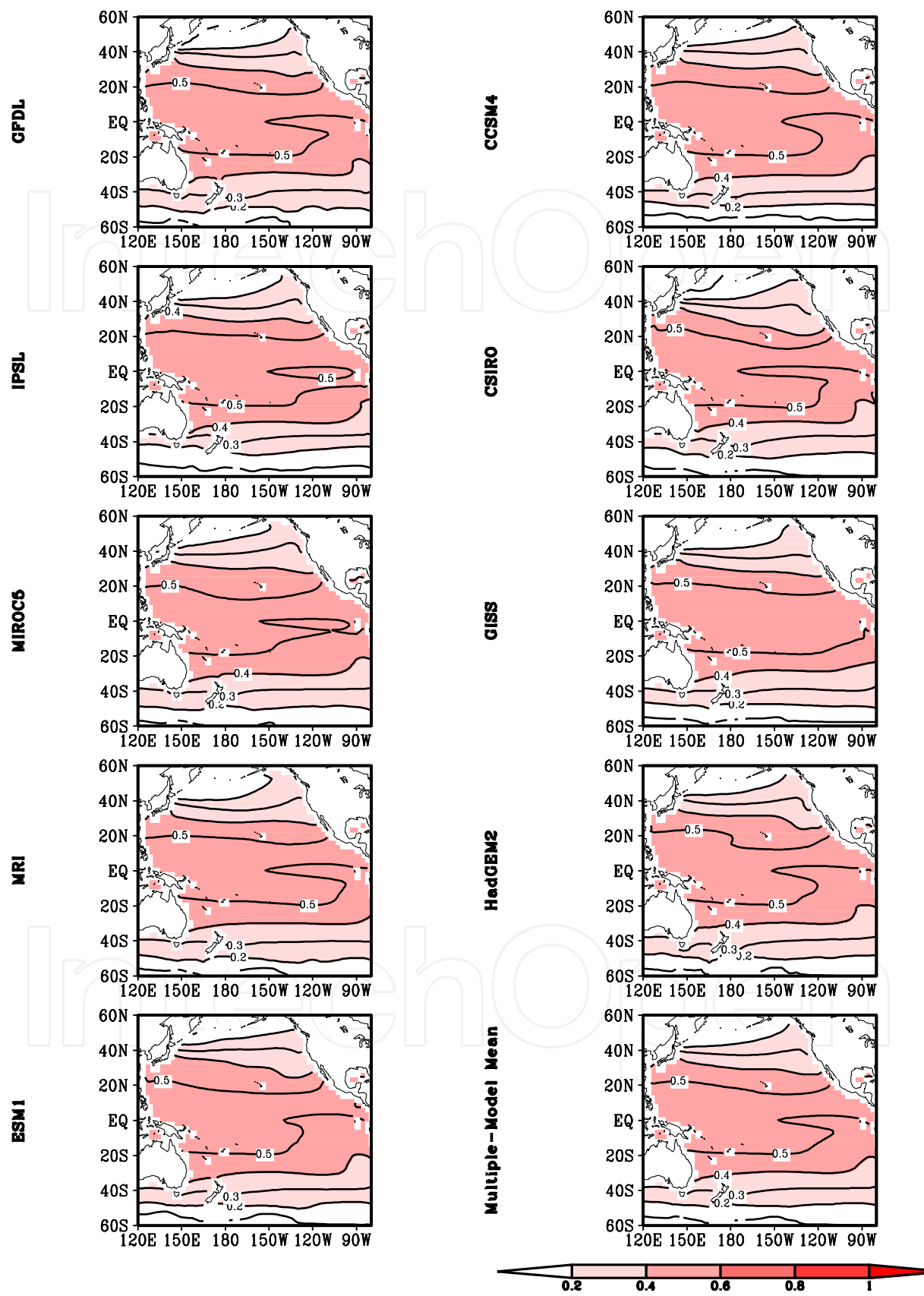
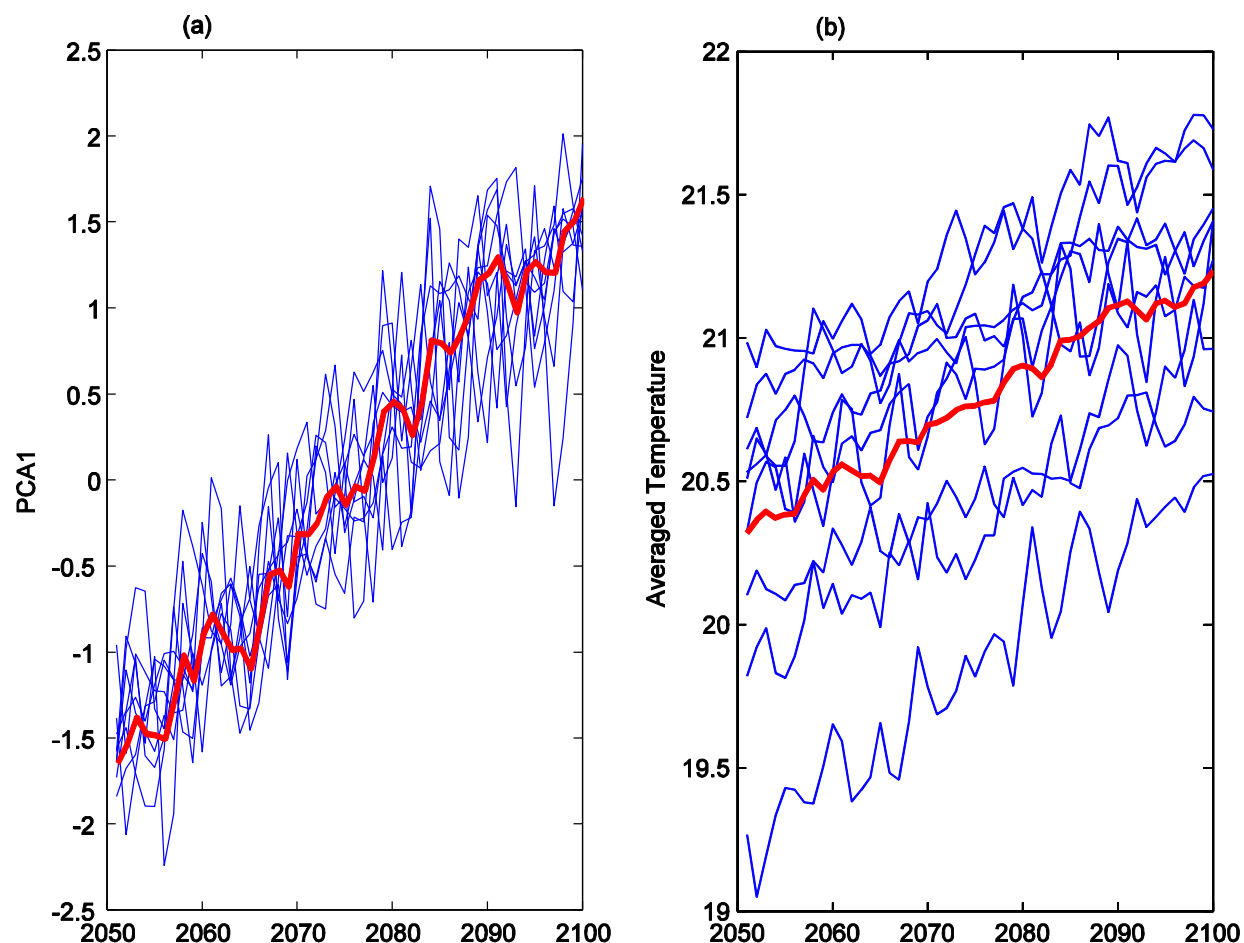**Figure 1.** The first EOF mode of the project from each model.

Figure 2. (a) The temporal variation of the time series of first EOF (99% variance) and (b) the averaged temperature over the north America. The blue line is for each model and red line is the mean of all models. The PCA1 of each model in Fig. 2a is normalized prior to plotting.

Fig. 2 shows a visible divergence of projections among models, suggesting uncertainties existed in the responses of these models to the GHGs forcing. It should be noted that little divergence in Fig 2a is due to the normalization, a post-processing just for a good-looking of this figure.

It is of great interest to explore the uncertainty of the above projections. As introduced aforementioned, one can use the above information-based framework to measure the uncertainty of climate prediction, given the multiple ensembles available. Apparently, there are several challenges here: 1) there is only one-member projection for each model, lacking sufficient ensembles; 2) the projection is not dependent on initial condition, thus any measures based on multiple initial conditions are invalid here; 3) the climatological distribution used in estimating the uncertainty may be uncertain under the background of global warming. For the first issue, we propose to solve it using multiple model strategy, i.e. pool all model projections to construct a 9-member ensemble. Under the framework of potential predictability, the model is assumed to be perfect. Thus the disparities among

these model projections can be viewed as the ensembles of a perfect model, perturbed by initial conditions or other parameters. For the second issue, we assume that the projection is a long-term prediction at a given initial condition. The distribution for the average of multiple model projections is used as climatological distribution here.

Displayed in Fig. 3 are the variations of projection utility RE during the projection time from 2051 to 2010.The climatological mean and variance are estimated from all 'ensemble' members and years (sample size is 50*9) as in [25]. The projection mean and variance are estimated each year from the 9-member ensemble. As can be seen, it is apparent that the utility R continues to decrease until around 2070 and then bounce after 2080. For the projection during 2070 - 2080, RE is small. When projection (prediction) and climatology distributions are identical, the relative entropy R is zero from (14). In theory, a nonzero value of R indicates predictability. However, in practice, a finite sample size introduces sampling errors that lead to a nonzero R even though there is no extra information supplied by the prediction. Therefore the statistical significance level should exceed the extent of uncertainty due to the finite sample size. We quantify the extent of uncertainty using a Monte Carlo method as in [26]. A sample with 9 members is randomly drawn from the climatology distribution and its relative entropy R is computed with respect to the climatology distribution. This process is repeated 10 000 times, and the value above 95% of 10 000 RE is considered to be the significant level as shown in Fig. 3 (dashed line). As can be seen, the projections between around 2070 and 2080 have statistically 'zero' relative entropy, and the other projections beyond this period have significant relative entropy.

A striking feature of RE in Fig. 3 is its U-shape variation with the projection time (i.e., the time step of integration), which is quite different from actual ensemble forecast at time scales from days to seasons. Typically, the RE monotonously decreases with the lead time of prediction at the time scales from days to seasons (e.g., [1-2, 11]), i.e., the predictability decreases with lead time. The monotonous variation of RE with lead time of predictions well characterizes the nature and attributes of realistic atmospheric and oceanic system, which is chaotic and stochastic, leading to the information at initial conditions gradually dissipated with lead time. Apparently it is not this case here, since the projection is not an initial value problem, and mainly is a response to external forcing (e.g., CO_2).

One possible explanation for this U-shape is related to the climatological distribution used here. We used the average of multiple model projections that have an apparent trend as the climatology distribution. If the RE is dominated by the ensemble mean (ensemble mean square) and the contribution of ensemble spread is relatively much smaller, the RE can show such a U-shape structure. Another plausible explanation is based on a hypothesis, namely, the climatology from multiple models is close to the true value. Under this assumption, the projection with small RE in figure 3 has high fidelity and vice versa. Here, we use the RE to measure the difference between the distribution of projection and the designed distribution, which has been also used in previous studies [17]. However such a hypothesis may cause concerns. One may argue to use present climatological distribution as a reference

distribution in the above discussions. However, it can be expected that the climatology of the scenario of R60 should be quite different from the present one. Thus, a further study on the reference distribution is highly demanded in estimating uncertainty of climate change projection.
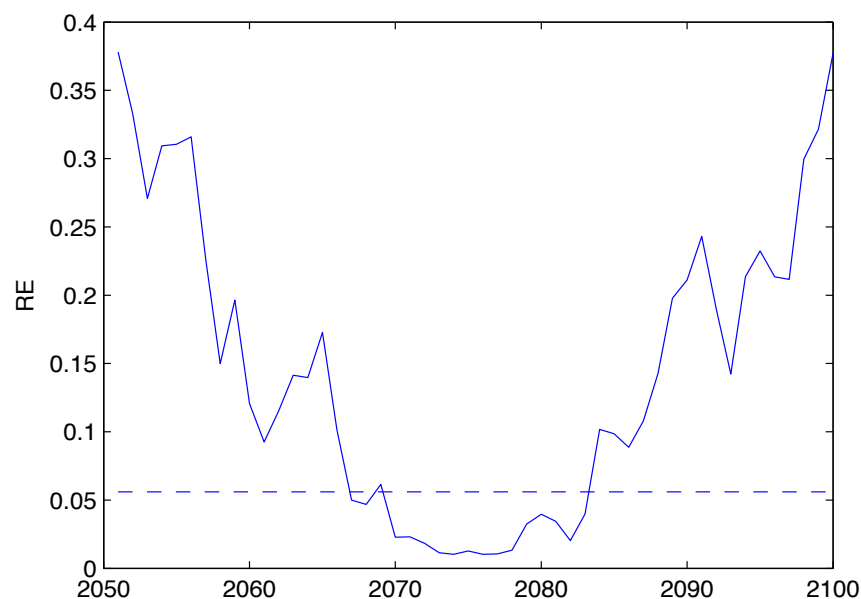

Figure 3. RE as a function of projection time

6. Conclusion

In this chapter, the SNR-based and information-based measures of potential predictability were introduced. They include the signal to noise ratio (SNR) and two measures of information-based predictability. One is relative entropy (RE) that measures individual potential predictability whereas the other is mutual information (MI), the average of RE over all initial conditions, which measures the average potential predictability. From statistical derivation and theoretical analysis, we have below conclusions:

i.   The SNR usually measures the average predictability with the assumption that signal inherent to slowly varying external forcing is predictable and the noise is unpredictable;

ii. A new measure of prediction utility that is derived from information theory is introduced. It measures the additional information provided by a prediction (p) over that already available from the climatological or reference distribution (q). One natural measure is their relative entropy *RE* defined as the relative difference of entropy between *p* and *q*. For the case of Gaussian distributed *p* and *q*, the RE can be expressed in terms of the prediction and reference means and covariance.

iii. Averaged RE over all initial conditions, called the mutual information (MI), a measure of the statistical dependence of the forecast state and the initial (boundary) conditions, measure the averaged predictability. The MI-based metrics can measure more potential prediction utility than the SNR-based counterpart. The MI-based predictability measures the statistical dependence, linear or nonlinear, between ensemble mean (prediction) and an ensemble member (hypothetical observation), whereas the SNR-based predictability only measures a linear relationship between prediction and hypothetical observation.

iv. When the prediction and climatological distribution are Gaussian and the ensemble spread is constant with predictions, both measures are identical to each other. When the ensemble spread is not constant, the SNR-based predictability often underestimates the potential predictability.

v. The predictable component analysis (PrCA), a method that assesses the most predictable patterns, is introduced. The PrCA decomposes the predictability into patterns accounting for different fractions of the total predictability. Distinguishing spatial structures that are unpredictable from those that are predictable is important for practical prediction problems, particularly when the predictable patterns are few in number.

As an example, the uncertainty of the climate change projection from scenario R60 of AR5 was evaluated, with the Pacific SST as the target. Nine models from different countries were participated in this evaluation. It was found that the most striking warming occurs at the tropical Pacific along the equator. In the extra-tropics beyond 30S to 30N, the increase in SST is relatively weaker. On average, the mean temperature of the Pacific ocean of 60S to 60N increases around 0.5 to 1C from 2051-2010 in these models. The relative entropy RE, measuring the utility of climate projection, continues to decrease until around 2070 and then bounce after 2080. For the projection during 2070 - 2080, RE is small. Under the assumption that the climatology from multiple models is close to the true value, the projection during the period with small RE suggests high fidelity and vice versa.

Author details

Youmin Tang
Environmental Science and Engineering, University of Northern British, Columbia, Prince George, Canada
State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, State Oceanic Administration, Hangzhou, P. R. China

Dake Chen
State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography,
State Oceanic Administration, Hangzhou, P. R. China
Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA

Dejian Yang
School of Atmospheric Sciences, Nanjing University

Tao Lian
State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography,
State Oceanic Administration, Hangzhou, P. R. China

Acknowledgement

7. References

[1] Tang, Y., Kleeman, R., and Moore, A. M. Comparison of information-based measures of forecast uncertainty in ensemble ENSO prediction. J. Clim. 2008a; 21, 230–247.

[2] Tang, Y., Lin, H. and Moore, A. M. Measuring the potential predictability of ensemble climate predictions. J. Geophys. Res. 2008b; 113, D04108, doi:10.1029/2007JD008804.

[3] Shukla, J. Predictability in the midst of chaos: A scientific basis for climate forecasting. Science, 1998; 282, 728–731.

[4] Kumar, A., Barnston, A. G., Peng, P., Hoerling, M. P., and Goddard, L.Changes in the spread of the variability of the seasonal mean atmospheric states associated with ENSO. J. Clim. 2000;13, 3139-3151.

[5] Peng, P., Kumar, A., Wang, W. An analysis of seasonal predictability in coupled model forecasts. Clim. Dyn.2009; 36, 637-648.

[6] Anderson, J. L., and W. F. Stern, 1996: Evaluating the potential predictive utility of ensemble forecasts. J. Climate, 9, 260–269.

[7] Sardeshmukh, P. D., Compo, G. P., and Penland C. Changes of probability associated with El Ni~no. J. Climate, 2000; 13, 4268–4286.

[8] Cover, T. M., and Thomas, J.A. Elements of Information Theory, pp. 576, Wiley;1991.

[9] DelSole, T. Predictability and information theory. Part I: Measures of predictability. J. Atmos. Sci. 2004; 61, 2425–2440.

[10] Schneider, T., and Griffies, S. A conceptual framework for predictability studies. J. Clim.1999; 12, 3133–3155.

[11] Tang, Y., Kleeman, R. and Moore, A. M. On the reliability of ENSO dynamical predictions. J. Atmos. Sci.2005; 62, 1770–1791.

[12] Kleeman, R. Measuring dynamical prediction utility using relative entropy. J. Atmos. Sci.2002; 59, 2057–2072.

[13] DelSole, T., and Tippett M. K. Predictability: Recent insights from information theory. Rev. Geophys.2007; 45, RG4002, doi:10.1029/2006RG000202.

[14] Joe, H. Relative entropy measures of multivariate dependence, *J. Amer. Stat. Assoc., 1989: 84,* 157–164.

[15] DelSole, T. Predictability and information theory. Part II: Imperfect forecasts. J. Atmos. Sci.2005; 62, 3368–3381.

[16] Kumar, A. and Hoerling, M.P., 1998. 'Annual cycle of Pacific/North American seasonal predictability associated with different phases of ENSO', J. Climate, 11, 3295-3308.

[17] Shukla, J., DelSole,T. Fennessy, M. Kinter, J. and Paolino, D. Climate Model Fidelity and Projections of Climate Change. Geophys. Res. Lett.2006; 33, L07702, doi:10.1029/2005GL025579.

[18] Yang, D. Tang, Y., Zhang, Y. and Yang, X. Information-Based Potential Predictability of the Asian Summer Monsoon in a Coupled Model. J. Geophys. Res.2012; doi:10.1029/2011JD016775.

[19] Tippett, M. K., Barnston,A.G. and DelSole, T. Comments on "finite samples and uncertainty estimates for skill measures for seasonal prediction". Mon. Weather Rev. 2010; 138, 1487-1493.

[20] Venzke S, Allen MR, Sutton RT, Rowell DP. The atmospheric response over the North Atlantic to decadal changes in sea surface temperatures. J Clim 1999; 12: 2562‒2584

[21] Allen and Smith 1997 Optimal filtering in singular spectrumanalysis. *Phys. Lett.* 234:419–428

[22] Fukunag. Introduction to statistical pattern recognition, Second Ed., Academic press, Boston. 1990.

[23] Sutton RT, Jewson SP, Rowell DP. 2000. The elements of climate variability in the tropical Atlantic region. *J. Climate* 13:3261–3284

[24] Delsole, T. and Shukla, J. Linear prediction of the Indian monsoon rainfall. Centre for Ocean–Land–Atmosphere Studies (COLA) Technical Report, CTR 114, 2002, p. 58.

[25] Tang, Y., Lin, H. Derome, J. and Tippett, M. K. A predictability measure applied to seasonal predictions of the Arctic Oscillation. J. Clim.2007; 20, 4733–4750.

[26] Tippett, M. K., Kleeman, R. and Tang Y. Measuring the potential utility of seasonal climate predictions. Geophys. Res. Lett. 2004; 31, L22201, doi:10.1029/2004GL021575.