Responses to reviewer#2:

All the authors are extremely grateful to you for providing your excellent comments and valuable advices for this paper. Your major four suggestions that Construction of the first two predictors ieT1 and T2; Selection of the other predictors; Structure of the model and Model validation are very helpful for us. Based on your suggestions, we have made major revisions to on our paper. We have added the discussion of the selection of the predictors, the structure of the model and the model validation based on your specific comments.

Thank you again for your valuable comments to improve our submission. If there are still any problems on the method, diction, phrasing, grammar, and spelling, please do not hesitate to tell us and we'll try our best to improve them.

In the following, kind comments you suggested before are in black text with corresponding actions taken by us following in blue.

1 . Section 2.2 EOF deconstruction. This section requires some more detail. While the given reference describes the EOF method, we need to know how it is applied here. Is the correlation or covariance matrix used? How are the anomalies constructed – simple removal of the monthly means? How are the anomalies smoothed - how strong is the smoothing and is it applied spatially or over time? More importantly, why are only the first 2 EOFs considered? A similar analysis has recently been reported by L'Heureux et al (Clim Dyn 2013, DOI 10.1007/s00382-012-1331-2). Their first two EOFs are similar to those described here (but with no smoothing and hence lower explained variance). Using different data sets and time periods, they

show that the 2nd EOF is not stable, being entirely due to the strong trend (also evident in Figure 1d). The pattern does not appear if the data is detrended, and also becomes less important if different time periods and/or domains are used. Most importantly, they do not interpret it as indicating "the ENSO signal beginning to decay".

Responses：Good suggestions. We have used covariance matrix，because the covariance matrix was selected to diagnose the primary patterns of co-variability in the basin-wide SSTs, rather than the patterns of normalized covariance (or correlation matrix). We have used the smooths function with MATLAB, which is five points two times moving, mainly filtering out some noise points and outliers.

Because the variance contribution of the first EOF mode is 61.33% and the variance contribution of the second EOF mode is 14.52%, so the first two EOF modes account for 75.85% of the total variance contribution, which has occupied most of the variance contribution and also contains most of the information of the field decomposition. So the first 2 EOFs are considered.

Based on the reference of L'Heureux et al. (Clim Dyn 2013, DOI 10.1007/s00382-012-1331-2), we need to do more experiments to prove that we choose the second mode of EOF to be appropriate, and whether different time periods will make us forecast unstable or not. Our original data is the monthly average SST data from January 1951 to Dec. 2010, which are 60 years. We will increase the length of the data for 20 years (Jan.1931 –Dec.2010), for 10 years (Jan.1941- Dec.2010) and decrease the length of the data for 10 years (Jan.1961- Dec.2010), for 20 years

(Jan.1971- Dec.2010). And then we use the same method to reconstruct a model and forecast the ENSO index as section5.4. The prediction results are shown in the following table:

Table5. The forecast results of the different data periods

Forecast events	The data periods (Jan. 1951-Dec.2010) Lead time of all seasons combined		The data periods (Jan. 1931-Dec.2010) Lead time of all seasons combined		The data periods (Jan. 1941-Dec.2010) Lead time of all seasons combined		The data periods (Jan. 1961-Dec.2010) Lead time of all seasons combined		The data periods(Jan. 1971- Dec.2010) Lead time of all seasons combined	
	TC	MAPE	TC	MAPE	TC	MAPE	TC	MAPE	TC	MAPE
The average of 18 El Niño examples	0.604	9.70%	0.683	9.02%	0.642	9.35%	0.572	10.15%	0.551	10.44%
The average of 22 La Niña examples	0.625	8.97%	0.701	8.33%	0.675	8.55%	0.589	9.42%	0.567	9.82%
The average of 20 Neutral examples	0.798	5.96%	0.845	5.12%	0.821	5.56%	0.746	6.21%	0.721	6.58%
The average of total 60 examples	0.712	7.62%	0.771	7.14%	0.740	7.38%	0.680	7.96%	0.652	8.15%

From the table, we can see that in the 60 experiments, the prediction results of the data period increased by 20 years are the best, and the prediction results of the data period decreased by 20 years is the worst. This is because the more data we use, the more information it contain. But from the table we can also see the difference among forecast results of both TC and MAPE of five different sample data are less, and no abnormal change suddenly worse or better appear. All these indicate that using different data sets and time periods, even though may have a certain impact on the pattern of the 2nd EOF, but the impact on our forecast is not great and it will not make our forecast unstable.

The "indicating the ENSO signal beginning to decay" in our previous paper is a mistake of writing, which is not seen from the space mode of Figure 1 (c), but from the time mode of Figure 1 (d). From Figure1 (d) we can see the time coefficient has a significant upward trend over time, indicating "the ENSO signal beginning to enhanced".

We have added the discussion about the stability of our forecast in page6-7 and page28-29 and revised as "the ENSO signal beginning to enhanced " in page7.

We sincerely hope for your satisfaction with our revision. Thank you again for your kind suggestion.

2. Section 2.3 Predictor selection The selection of other potential predictors is confusing. Apart from T1 and T2, the other potential predictors come from a fairly limited set, and are not well supported by the referenced works. In lines 157-160, zonal winds in the western and eastern equatorial Pacific are mentioned, and it is well known that westerly wind anomalies in the western equatorial Pacific can (and do) trigger equatorially trapped oceanic Kelvin waves. There is an extensive amount of literature on the relationship between western equatorial Pacific zonal wind and ENSO, but here no references are given and only the eastern equatorial winds is

considered. Trenberth et al. discuss a link between ENSO and the PNA pattern (amongst other modes of extratropical variability), but this is the context of ENSO forcing of the PNA, ie ENSO leads to PNA teleconnections, but PNA does not predict ENSO. Yang et al introduce the EAWM index, but they note that "the relationship between ENSO and the east Asian winter monsoon is relatively weak". Nowhere do they suggest that the EAWMI is closely related to any ENSO indices. It is not surprising that the east Pacific wind and PNA do not feature in the final model.

Responses：Good suggestions. Your opinion is very good. In pervious paper the factors that we may consider are relatively few. But we are a complex coupled model of four factor differential equations and are not the similar with a simple statistical model (such as stepwise regression). So in our pervious paper using the stepwise regression method to select factors also has a problem. According to your opinion, we have read more literatures. We have expanded the scope of factor selection and revised the criterion of selecting factors, and the paragraph has revised as follows:

Considering the complexity of computation, the amount of variables in the equations of our model can’t be too large, usually 3 or 4 for the best. This has been explained in our previous studies (Zhang et al., 2006; Zhang et al., 2008). If there are more than 4 variables in the modeling equation, it will cause the amount of parameters such as $a_1, a_2, ...a_n, b_1, b_2, ...b_n, ...$ too large. The huge computation makes it difficult to be precisely modeled. Thus, the total number of parameters in the model of five variables was 102, which may cause an overfitting problem. Hence, when we selected the model of five or six variables which entailed large amounts of computation that made precision difficult, and too many parameters might cause an overfitting phenomenon. If we choose only two or even fewer variables, the forecast performance is poor too. Too few variables cause too small reconstructed parameters, resulting in amounts of important information missing out in the model. Thus, four variables are best for dynamically and accurately modeling. Because we have chosen two time series in section2.2 as the modeling objects, now we should select the other two ENSO intensity impact factors.

The ENSO intensity impact factor is an important issue in ENSO prediction. Previous studies have been completed in this area, which found that teleconnection patterns, temperature, precipitation, wind and SSH may affect ENSO strength. For example, Trenberth et al. (1998) noted that PNA, SOI and OLR in the Pacific Intertropical Convergence Zone (ITCZ) are all closely related to ENSO. Webster(1999) pointed out after the 1970, Indian Ocean dipole (IOD) is not only affected by ENSO, but also affected the strength of ENSO (Ashok et al., 2001). Yoon and Yeh (2010) reported that the Pacific Decadal Oscillation (PDO) disrupts the linkage between El Ni˜no and the following Northeast Asian summer monsoon (NEASM) through inducing the Eurasian pattern in the mid-high latitudes. The vast majority of studies (Tomita and Yasunari, 1996; Zhou and Wu, 2010; Kim et al., 2017)have concentrated on the impacts of ENSO on the East Asian winter monsoon(EAWM). During the EAWM season, ENSO generally reaches its mature phase and has the most prominent impact on the climate. Wang et al. (1999a) and

Wang et al. (1999b) suggested that the zonal wind factors in the eastern and western equatorial Pacific play a critical role in the phase of transition of the ENSO cycle, which could excite eastward propagating Kelvin waves and affect the SSTA in the equatorial Pacific. Zhao et al. (2012) analyzed the characteristics of the tropical Pacific SSH field and its impact on ENSO events.

Based on the above analysis, we have selected nine factors, which may be closely related with the ENSO index (Niño3.4).

(1)The zonal wind in the eastern equatorial Pacific factor (u1) was calculated as the grid-point average of zonal wind in the area [5 °S ~ 5 °N, 150 °W ~ 90 °W].

(2) The zonal wind in the western equatorial Pacific factor (u2) was calculated as the grid-point average of zonal wind in the area [0 °~ 10 °N; 135 °E ~ 180 °E].

(3) The PNA teleconnection factor was obtained from the CPC.

(4) the dipole mode index factor (DMI) was obtained from SSTA for June-July-August (JJA) based on Saji(1999) method.

(5) The SOI factor was obtained from the CPC.

(6) The PDOI factor was obtained from department of Atmospheric Sciences in the university of Washington. The web is http://tao.atmos.washinton.edu/pdo/RDO.latest.

(7) The EAWM index (EAWMI) factor was proposed by Yang et al. (2002), which is defined by the meridional 850-hPa winds averaged over the region (20 ° ~40 °N, 100 °~140 °E).

(8) The OLR in the ITCZ factor was calculated as the grid-point average of

OLR in the area [10° N～20° N，120° E～150° E].

(9) The SSH factor was calculated as the grid-point average of the SSH data in the area [10 °S ~ 10 °N; 120 °E ~ 60 °W].

A correlation analysis of the above factors was carried out and the results are shown in Table 2.

Table 2 shows that SOI and EAWMI have the stronger correlation with the front two time series T_1, T_2 than the other 7 factors. The results are also consistent with previous research (Clarke and Van Gorder, 2003; Drosdowsky, 2006; Zhang et al., 1996; Wang et al., 2008; Yang and Lu, 2014). Therefore, the first time series T_1, the second time series T_2, SOI and EAWMI will be selected as prediction model factors.

Table 2. The correlation analysis between the front two time series T_1, T_2 and nine impact factors

factors	u_1	u_2	PNA	DMI	SOI	PDOI	EAWMI	OLR	SSH
T_1	0.3161	0.5684	0.4386	-0.3457	0.7734	0.4081	0.6284	0.3287	0.3363
T_2	0.2118	0.4181	0.2560	-0.2345	0.5232	0.3065	0.4825	0.1816	0.2169

Actually, how many variables and which variables are used in our model become a key issue to be resolved. We are a complex four factor differential equations coupling model. We are a complex coupled model of four factor differential equations, so we are more concerned with the correlation between each other. The correlation must be considered as an important criterion to select the factors, but in order to further verify the correctness of the selection criterion, we have carried out the prediction experiments (the 60 cross-validated retroactive hindcasts experiments of the ENSO index for all seasons combined at lead times of 8 months) of different variables. The forecast results of the models of different variables are as following:

Table3. The forecast results (The temporal correlation (TC) and the root mean square error (RMSE))of the models of different variables

The forecast results	Three variables of the model					
	T_1,T_2,u_1	T_1,T_2,u_2	$T_1,T_2,$PNA	$T_1,T_2,$DMI	$T_1,T_2,$SOI	$T_1,T_2,$PDOI
TC	0.4423	0.5628	0.3852	0.3226	**0.6027**	0.3809
RMSE	0.9025	0.7855	0.9244	1.0041	**0.7275**	1.0642
	$T_1,T_2,$EAWMI	$T_1,T_2,$OLR	$T_1,T_2,$SSH			
TC	0.5829	0.3205	0.4288			
RMSE	0.7516	0.9814	0.9090			
	four variables of the model					
	T_1,T_2,u_1,u_2	$T_1,T_2,u_1,$PNA	$T_1,T_2,u_1,$DMI	$T_1,T_2,u_1,$SOI	$T_1,T_2,u_1,$ *PDOI*	$T_1,T_2,u_1,$ EAWMI
TC	0.4672	0.3628	0.5617	0.5201	0.5028	0.5822
RMSE	0.8824	0.9902	0.7617	0.8233	0.8092	0.7132
	$T_1,T_2,u_1,$OLR	$T_1,T_2,u_1,$SSH	$T_1,T_2,u_2,$PNA	$T_1,T_2,u_2,$DMI	$T_1,T_2,u_2,$SOI	$T_1,T_2,u_2,$ *PDOI*
TC	0.3815	0.4128	0.3107	0.4125	0.5910	0.5504
RMSE	0.9702	0.9017	1.0255	0.9392	0.7128	0.7503
	$T_1,T_2,u_2,$ EAWMI	$T_1,T_2,u_2,$OLR	$T_1,T_2,u_2,$SSH	$T_1,T_2,$PNA, DMI	$T_1,T_2,$PNA,SOI	$T_1,T_2,$PNA, *PDOI*
TC	0.6048	0.4528	0.5308	0.3022	0.3875	0.2876
RMSE	0.6910	0.9028	0.8344	1.0578	0.9706	1.1305

	T_1,T_2,PNA, EAWMI	T_1,T_2,PNA, OLR	T_1,T_2,PNA, SSH	T_1,T_2,DMI, SOI	T_1,T_2,DMI, *PDOI*	T_1,T_2,DMI, EAWMI
TC	0.3527	0.2556	0.2175	0.5688	0.2028	0.5807
RMSE	0.9518	1.2024	1.3244	0.7425	1.2905	0.7015
	T_1,T_2,DMI, OLR	T_1,T_2,DMI, SSH	T_1,T_2,SOI, *PDOI*	T_1,T_2,SOI, EAWMI	T_1,T_2,SOI,OLR	T_1,T_2,SOI, SSH
TC	0.3504	0.4833	0.6022	**0.6344**	0.5876	0.5476
RMSE	1.1624	0.8530	0.7054	**0.6728**	0.7408	0.7895
	T_1,T_2,PDOI, EAWMI	T_1,T_2,PDOI, OLR	T_1,T_2,PDOI, SSH	T_1,T_2,EAWMI, OLR	T_1,T_2,EAWMI, SSH	T_1,T_2,OLR, SSH
TC	0.4217	0.2017	0.2044	0.5872	0.4607	0.2028
RMSE	0.9147	1.2085	1.2542	0.7233	0.8925	1.3524

From the table, we can see that for all the forecast results of the models of different variables, the prediction results of T_1,T_2,SOI is the best among those of the three factors and the prediction result of T_1,T_2,SOI,EAWMI is the best among those of the four factors. But the prediction result of T_1,T_2,SOI,EAWMI is best among all, which proves that our selection factors are correct. In our previous study (Hong et al., 2015), the model of the Western Pacific subtropical high was established by using the correlations as a criterion to select factors and their forecast results are also good. Now we use the correlations as a criterion to select factors is also in line with our previous research.

With the deepening of the research, there are still a lot of new literatures that reveal the relationship between ENSO and the East Asian winter monsoon. For example:

[1] Kim Ji-Won ,Soon-Il An,Sang-Yoon Jun,Hey-Jin Park,Sang-Wook Yeh. 2017.ENSO and East

Asian winter monsoon relationship modulation associated with the anomalous northwest Pacific anticyclone, Climate Dynamics, Volume 49, Issue 4, pp 1157–1179.

[2] Yang Se-Hwan and Lu Riyu . 2014. Predictability of the East Asian winter monsoon indices by the coupled models of ENSEMBLES, Advances in Atmospheric Sciences, Volume 31, Issue 6, pp 1279–1292.

[3] Wang L., Chen W. and Huang R. H., 2008. Interdecadal modulation of PDO on the impact of ENSO on the east Asian winter monsoon, Geophysical Research Letter, DOI: 10.1029/2008GL035287.

So there is a good correlation between ENSO and the East Asian winter monsoon.

The specific revision can be seen in section2.3 in page7-10 and line616 to632 in page29-30.We sincerely hope for your satisfaction with our revision. Thank you again for your kind suggestion.

References:

[1] Mei Hong, Ren Zhang, et al., Reconstruction and forecast experiments of a statistical-dynamical model of the Western Pacific subtropical high and Eastern Asian summer monsoon factors，Weather and Forecasting, 2015,30:206-216.

[2]Zhang R. and Hong M., et al.: Non-linear Dynamic Model Retrieval of Subtropical High Based on Empirical Orthogonal Function and Genetic Algorithm,Applied Mathematics and Mechanics,27(12),1645-1654, 2006.

[3]Zhang R. and Hong M.,et al.: Retrieval of the non-linear dynamic forecast model of El Nino/La Nina index based on the genetic algorithm optimization. Chinese Journal of

Geophysics,51(5),1354-1362, 2008.

[4]Trenberth, E. K., et al.: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures,J. Geophys. Res., 107, C7, 14291-14324,1998.

[5]Webster P. J., Moore A. M., Loschnigg J. P., et al.: Coupled ocean-atmosphere dynamics in the Indian Ocean during 1997- 98, Nature, 401(6751),356-360, 1999.

[6]Ashok K, Guan Z, Yam agata T : Impact of the Indian Ocean Dipole on the decadal relationship between the Indian mon soon rainfall and ENSO, Geophys Res Let,28(23), 4499-4502, 2001.

[7]Yoon, J., and S. W. Yeh: Influence of the Pacific Decadal Oscillation on the relationship between El Ni˜no and the northeast Asian summer monsoon, J. Climate,23, 4525–4537, 2010.

[8]Tomita, T., and T. Yasunari: Role of the northeast winter monsoon on the biennial oscillation of the ENSO/monsoon system,J. Meteor. Soc. Japan, 74,399–413 , 1996.

[9]Zhou, L.-T., and R. G. Wu: Respective impacts of the East Asian winter monsoon and ENSO on winter rainfall in China,J. Geophys. Res.,115, doi: 10.1029/2009JD012502, 2010.

[10]Wang B., Wu R., Lukas R.: Roles of western North Pacific wind variation in thermocline adjustment and ENSO phase transition,J Meteor Soc Japan,77,1-16,1999a.

[11]Wang C., Weisberg R. H. and Virmani J. I.: Western Pacific interannual variability associated with the El Niño-Southern Oscillation,J Geophy Res.,104,5131-5149, 1999b.

[12]Zhao J. H., Liu X. Y. and Jiang H. Y., et al.: Characteristics of Sea Surface Height in Tropical Pacific and its relationship with ENSO events,Meteorological and Environmental Sciences, 35(2),33-39, 2012.

[13]Yang, S., K. M. Lau, and K. M. Kim: Variations of the East Asian jet stream and

Asian-Pacific-American winter climate anomalies, J. Climate, 15,306–325 , 2002.

[14]Saji N. H., Goswami B. N., V. inayachandran P. N., et al.: A dipole mode in the tropical Indian Ocean,Nature, 401(6751),360-363, 1999.

[15]Clarke A. J. and S. Van Gorder: Improving El Niño prediction using a space-time integration of Indo-Pacific winds and equatorial Pacific upper ocean heat content, Geophys. Res. Lett.,30,1399. doi:10.1029/2002GL016673, 2003.

[16]Drosdowsky W.: Statistical prediction of ENSO (Niño 3) using sub-surface temperature data,Geophys. Res.Lett., 33 , L03710. doi:10.1029/2005GL024866, 2006.

[17]Zhang, R. H., A. Sumi, and M. Kimoto: Impact of El Ni˜noon the East Asian monsoon: A diagnostic study of the ’86/87 and ’91/92 events,J. Meteor. Soc. Japan, 74, 49–62, 1996.

[18]Wang, L., W. Chen, and R. H. Huang: Interdecadal modulation of PDO on the impact of ENSO on the east Asian winter monsoon, Geophys. Res. Lett., 35, L20702, doi:10.1029/2008GL035287, 2008.

3-1. The remainder of section 2.3, concerned with determining the number of predictors is difficult to follow. It is not until section 3 (page11) that it is revealed that the model is a dynamical system of four second order coupled equations, involving the products of the various predictors as well as the predictors themselves. Nowhere is the inclusion of these terms discussed or justified. What physical processes do these terms represent? What do the predictors squared represent?, and the cross products ie what do T1 * SOI or T2 * EAWMI mean? Since the model is not a linear regression model, is stepwise regression a valid procedure for determining the significance of the predictors?

Responses：Good suggestions. Your opinion is very good. Based on your suggestion of question2, we have revised the discussion of how to determine the number of predictors. Our model is not a linear regression model, the stepwise regression may be a valid procedure for determining the significance of the predictors, so we also have revised the method for determining the significance of the predictors, the specific revision can be seen our answer of the question2.

The inclusion of these terms and the physical processes do these terms represent are important, especially for the discussion of dynamical characteristics of the dynamical model. But now we are difficult to give a clear meaning. Now the main work of our paper is the prediction experiments of the model. For the reason of time and length, this paper mainly discusses the prediction results of the model. The physical processes do these terms represent and the discussion of the dynamical characteristics of the model will be the focus of our next work. Before this, we have also used the Takens' delay embedding theorem to reconstruct the dynamical model of the Western Pacific subtropical high(WPSH). And Based on the reconstructed dynamical model, dynamical characteristics of WPSH are analyzed and an aberrance mechanism is developed, in which the external forcings resulting in the WPSH anomalies are explored, which have been published(Hong et al., 2016). We also study the bifurcation and catastrophe of the West Pacific subtropical high ridge index of a nonlinear model (Hong et al., 2017). Based on our previous method and work, our next work is to analyze the physical processes and the dynamical characteristics of the SST field.

The specific revision can be seen from line689 to704 in page33.We sincerely hope for your satisfaction with our revision. Thank you again for your kind suggestion.

References:

[1] Mei Hong*, Ren Zhang, et al.,. Catastrophe and Mechanism Analyses of Multiple Equilibria in the Western Pacific Subtropical High System Based on Objective Fitting of Spatial Basis Functions. Monthly Weather Review, 2016,144:997-1015.

[2] Mei Hong*, Ren Zhang, et al., Bifurcations and catastrophes in a nonlinear dynamical model of the western Pacific subtropical high ridge line index and its evolution mechanism, Theor. Appl. Climatol., 129, 363-384, 2017.

3-2. line 195. The idea that a model with the number of predictors less than 10% of the sample size can avoid overfitting is new to me. The reference given (Tetko et al) is about neural networks. Is this applicable to the system of coupled equations used here? (I could only see the first page) Also I am not sure if the discussion in 198-203 is incorrect. Even if only 34 parametres are accepted, the full set of 56 parameters must be estimated to know which to accept or reject. This may be more a problem of introducing artificial skill, which has long been recognised as a problem in statistical forecasting. It generally arises when you try enough predictors, and retain those that "work" and discard the others.

This question of the number of parametrs / predictors is exacerabated in Section 4 and 5 where the number of predictors is increased again by including lagged values. On first inspection Equations 3 and 7 involve 112 parameters. There are 28 alphas, 28 thetas, as given in lines 395 and 396. (In line 202, it is stated that there are 28 self memorization parameters beta; but there are no betas in Eqs 3 and 5, but there are in Appendix B) In addition each of the four F "dynamical cores" involve 14 parameters as shown in Equation 1, assuming that the same F is used at each lagged time. Given that the input data (the xi) are different at each lag, is the same F a valid assumption? Even with the authors 34 accepted values in the Fs, there is still a total of 90 parameters. This is well over 10%, and on the authors own criterion, this would suggest that the system is perhaps overfit. Additionally, all the 720 observations are not statistically independent. Both T1 and the SOI (and probably T2 with its strong trend) are strongly auto-correlated, and the effective sample size is probably significantly less than 720. All in all, this discussion is very confusing!

Responses：Good suggestions. Our final number of 90 parameters is still a little large for a sample size of 720. In the previous paper, this discussion of overfitting is a little confusing. So it is still necessary to further discuss whether our model has the overfitting problem or not. Thank reviewers to remind us this problem.

The definition of overfitting: The learned hypothesis may fit the training set very well, but fail to predict to new examples (fail to fit additional data or predict future observations reliably).

The potential for overfitting depends not only on the number of parameters and data but also the conformability of the model structure with the data shape, and the magnitude of model error compared to the expected level of noise or error in the data(Burnham and Anderson, 2002). So there are many reasons causing the overfitting phenomenon. But this does not mean having many parameters relative to the number of observations inevitably causes the overfitting problem (Golbraikh et al., 2003). There is no evidence that more parameters will be certain to result in overfitting. Based on the definition of overfitting and the previous studies(Golbraikh et al., 2003; Everitt and Skrondal,2010), we can judge whether a model is overfitting or not by the accuracy of prediction results of independent samples (Golbraikh and Tropsha, 2002; Qi and Li, 2006).

In the sample training, our model does not purposely pursue the high degree of the training samples fitting and improve the effectiveness of the independent generalization. In fact in our paper the forecast results of the Cross-validated retroactive hindcasts (section 5.2) and the independent samples validation (table3 and table4) are both good. Especially, the independent samples validation of the ENSO index as the table4, we have carried out the 240 independent sample validation prediction of four seasons of different ENSO events and the coverage of independent samples test is very wide. Moreover, compared with 6 mature prediction models, the forecast results of our model are also good, which prove the overfitting problem does not exist in our model. According to the previous literature (Islam and Sivakumar, 2002; Sivakumar et al.,2001), we can see that prediction principle and structure of the phase space reconstruction (PSR) of dynamical system is not the same with the traditional neural network and in the small sample situation the forecasting results of PSR model are better than those of the traditional neural network (Sivakumar et al. ,2002), which can be verified in the independent sample test (table3 and table4). So according to the definition of overfitting, we can say the over fitting phenomenon does not exist in our model.

Now we have added the new discussion of the overfitting problem from line633 to663in page30-31.

We sincerely hope for your satisfaction with our revision. Thank you again for your kind suggestion.

References：

[1]Burnham, K. P.; Anderson, D. R. (2002), Model Selection and Multimodel Inference (2nd ed.), Springer-Verlag.

[2] Everitt B.S., Skrondal A. (2010), Cambridge Dictionary of Statistics, Cambridge University Press.

[3] Golbraikh A.,Shen M., Xiao Z. Y., Xiao Y. D., Lee Kuo-Hsiung, Tropsha A., 2003: Rational selection of training and test sets for the development of validated QSAR models. Journal of Computer-Aided Molecular Design, 17(2), 241-253.

[4] Golbraikh A. and Tropsha A., 2002: Beware of q 2 ! Journal of Molecular Graphics and Modelling, 20 , 269–276.

[5] Qin G. H. and Li Z. H., 2006: Over-fitting of BP NN research and its application, Engineering Journal of Wuhan University,39(6),1671-1679

[6] Leinweber, D. J., 2007: Stupid Data Miner Tricks. The Journal of Investing, 16, 15–22.

[7] Islam M.N. Sivakumar B., 2002. Characterization and prediction of runoff dynamics:a

nonlinear dynamical view. Advances in Water Resources, 25, 179-190.

[8]Sivakumar B, Berndtsson R, Persson M. 2001. Monthly Runoff Prediction Using Phase -space Reconstruction. Hydrological Sciences Journal, 46(3), 377 -388.

[9]Sivakumar B., Jayawardena A.W., Fernando T.M.K.G., 2002. River flow forecasting: use of phase-space reconstruction and artificial neural networks approaches. Journal of Hydrology, 265, 225-245.

4. Model Validation

4-1.line 281-288. This paragraph took me a long time to understand, especially how one could obtain correlations and MAPE values based on a single forecast. As I understand it, "at this time" refers to the forecast at five months, and the correlation and MAPE are calculated over the first five months forecasts, and in general the values at the Nth month are based on the first N months forecast. (I assume that this is the "n" in the equation for MAPE on line 283)

Responses：Good suggestions. Your understanding is right. "at this time" refers to the forecast at five months, and the correlation and MAPE are calculated over the first five months forecasts, and in general the values at the Nth month are based on the first N months forecast. Now we revise the sentence “Using T_1 as an example, at this time, the temporal correlation between model predictions and corresponding observations was 0.8966 and the mean absolute percentage error (MAPE) (Hu et al., 2001), $\mathrm{MAPE}=\frac{1}{n}\sum_{i=1}^{n}\left|\frac{D_e(i)-D_0(i)}{D_0(i)}\right|\times 100$, was 8.32%.” as “Using as an example, the CC between model predictions and corresponding observations over the first five months forecasts was 0.8966 and MAPE was 8.32%. ” for readers’ better understanding.

The specific revision can be seen from line275 to276 in page13. We sincerely hope for your satisfaction with our revision. Thank you again for your kind suggestion.

4-2. This method would suggest that the correlation at one month is undefined, and 1.0 (perfectly accurate) at two months? This same type of calculation appears to be used in Tables 3 and 4.

Responses：Good suggestions. In previous paper, we have not explained the concept of correlation. There two different correlations in our paper. The first correlation in our paper is the pearson correlation coefficient (CC), which also can be called the linear correlation coefficient. It measures the strength and direction of a linear relationship between two variables (for example model output and observed values). The mathematical formula for computing r is:

$$r=\frac{\sum_{i=1}^{n}(D_e(i)-\bar{D}_e)\cdot(D_0(i)-\bar{D}_0)}{\sqrt{\sum_{i=1}^{n}(D_e(i)-\bar{D}_e)^2\cdot\sum_{i=1}^{n}(D_0(i)-\bar{D}_0)^2}}$$

Where n is the number of pairs of data, D_e, D_0 is a series of n observations and n forecast values.

The CC (Wang et al. 2009) and the mean absolute percentage error (MAPE)(Hu et al. 2001) are employed as objective functions to calibrate the model. The CC evaluates the linear relationship between the observed and predicting values and MAPE measures the difference between the observed and predicting values. The forecast results of T_1, T_2 in Section3, table2 and table3 have used the above two evaluation criteria (r and MAPE).

While the evaluation criteria of the ENSO index in table4 is the temporal correlation (TC), its definition and specific calculation steps can be seen in these literatures (Kathrin et al.,2016; Nicosia et al. 2013); The TC is often used to measure the prediction effect of the ENSO index. For example, in 1995,Chen et al. used TC as the evaluation criteria to test the improved Predictability of El Nino Forecasting of their model and Barnston et al.in 2012 also used the TC to compare the forecast skill of 21 real-time seasonal ENSO models.

In the previous paper, we didn't explain two different correlations clearly, which will be easy for readers to misunderstand. Now we have explained two different correlations and the specific revision can be seen in all my paper.

We sincerely hope for your satisfaction with our revision. Thank you again for your kind suggestion.

References：

[1] Wang, W. C., K. W. Chau, C. T. Cheng, and L. Qiu, 2009: A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. J. Hydrol., 374, 294–306, doi:10.1016/j.jhydrol.2009.06.019.

[2] Hu, T. S., K. C. Lam, and S. T. Ng, 2001: River flow time series prediction with a range-dependent neural network. Hydrol. Sci. J., 46, 729–745, doi:10.1080/02626660109492867.

[3] Kathrin Büttner, Jennifer Salau, and Joachim Krieter,2016: Temporal correlation coefficient for directed networks. Springerplus, 5(1): 1198-1203.

[4] Nicosia V, Tang J, Mascolo C, Musolesi M, Russo G, Latora V. Graph metrics for temporal networks. In: Holme P, Saramäki J, editors. Temporal networks. Berlin: Springer; 2013. pp. 15–40.

[5] Chen D., S. E. Zebiak, A. J. Busalacchi and M. A. Cane: An Improved Procedure for El Nirio Forecasting: Implications for Predictability,Science, 269, 1699-1702, 1995

[6] Barnston A. G., et al.: Skill of real-time seasonal ENSO model predictions during 2002-2011,Bull. Amer. Meteor. Soc.,93, 631-651, 2012.

4-3.line 289-298. Another confusing paragraph. January 1951 to January 1952 inclusive? is 13, not 12 months. How was the omitted section forecast, ie was it simply a 12 (or 13) month forecast starting at the last point before the omitted data?

Responses： Good suggestions. This is a mistake in writing and thanks the reviewers' comments. The omitted forecast section is 12 months, Jan. 1951 to Dec.1951, and the training sample time is Jan.1952 to December 2010. Then in the next prediction experiment, the omitted segment is Jan.1952 to Dec. 1952 and the training samples are Jan. 1951 to Dec.1951 and Jan.1953 to Dec.2010. So the forecast time series is Jan.1952 to Dec. 1952. We then repeated this procedure the by moving the omitted segment along the entirety of the available time series. The similar process of the cross-validated retroactive hindcasts has also been used in the previous literatures (Hu et al., 2017).

The specific revision can be seen from line 284 to 293 in page14.

We sincerely hope for your satisfaction with our revision. Thank you again for your kind suggestion.

References:

[1] Hu Y. J., Zhong Z., Zhu Y. M. et al., A statistical forecast model using the time-scale decomposition technique to predict rainfall during flood period over the middle and lower reaches of the Yangtze River Valley, Theoretical and Applied Climatology, doi: 10.1007/s00704-017-2094-9.

4-4.it is difficult to judge how "good" the forecast was based on Figure 3.

Responses：Good suggestions. From Fig3, the prediction values (blue line) and the actual values (red line) are relatively close in some places, but in many places, especially in the peaks, the error is large, which in accordance with the analysis of Figure 2. The forecast results within 5 months of the simple dynamical reconstruction model in section3 are good, but the long term prediction results after 5 months become bad and the error increases quickly. So this is why we have to introduce the self -memorization principle to improve the long term prediction results.

We sincerely hope for your satisfaction with our revision. Thank you again for your kind suggestion.

4-5.Again it is not clear how the correlation and MAPE statistics were calculated - only one value is given, so presumeably it is taken over all (720 months) forecast?

Responses：Good suggestions. In pervious paper we haven’t explained clearly how the correlation and MAPE statistics in Fig.3 were calculated. It isn’t taken over all (720 months) forecast when only one value is given (The forecast for such a long time is not possible). The figure 3 merges the 60 experiments (each experiment is the prediction of the 12 month similar as Fig.2) on one picture. The Fig.3 is equivalent to 60 experiments instead of the results of only one experiment, because the results of one experiment are not entirely representative. And through multiple cross experiments can more objectively reflect the forecast capability of our model. So the forecast results of 60 cross experiment (each experiment is the prediction of the 12 month as Fig.2) according to the time sequence can merger into a new time series (from Jan.1951-Dec.2010), and then the pearson correlation coefficient (CC) and the mean absolute percentage error (MAPE) can be calculated by the new prediction time series and the time series of the actual value based on the formula in the above answer of 4-2 problems. Actually, the CC and MAPE are the average of the prediction values of the 60 cross experiments. That's how the correlation and MAPE statistics were calculated in Fig. 3.

Now we have added the above explanation from line 294 to 300 in page14 for readers’ better understanding.

We sincerely hope for your satisfaction with our revision. Thank you again for your kind suggestion.

4-6. However the discussion in lines 310-312 suggest that individual 12 month forecasts were also evaluated. Overall the discussion of the forecast process and its validation in not clear.

Responses：Good suggestions. The CC and MAPE in Fig.3 are the average of the prediction values of the 60 cross experiments. But each MAPE value of the above 60 experiments is not the same and the difference between the maximum and the minimum MAPE value is quite large, which means that the prediction results of the simple dynamical reconstruction model in section3 is not stable. So that is another reason why we need to introduce self -memorization principle to improve our model.

We sincerely hope for your satisfaction with our revision. Thank you again for your kind suggestion.

Some minor points

1. In line 170, all 4 data sets range from Jan 1951 to Jan 2010, yet in at least 4 places,

Responses：Good suggestions. Now we have deleted the other 3 places about the description of the length of the data. And in pervious paper, " all 4 data sets from Jan. 1951 to Jan. 2010" is mistake in writing. Now we revised as "The time series of all data were from Jan. 1951 to Dec. 2010, 720 months in total" from line129 to line130 in page6.

We sincerely hope for your satisfaction with our revision. Thank you again for your kind suggestion.

2. lines 292, 373, 402 and 416 forecasts are evaluated up to December 2010?

Responses：Good suggestions. In previous paper, " all 4 data sets from Jan. 1951 to Jan. 2010" is mistake in writing. Now we revised as "The time series of all data were from Jan. 1951 to Dec. 2010, 720 months in total." So the lines 292, 373, 402 and 416 forecasts are surely evaluated up to December 2010.

We sincerely hope for your satisfaction with our revision. Thank you again for your kind suggestion.

3. lines 249-253. Why does normalising the raw values avoid the overfitting problem?

Responses：Good suggestions. Now we have revised the sentences" To avoid the overfitting problem, we used $x_{nor} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$ to normalize the raw value of each of the four predictors, then we used the normalized value to model and forecast." as "In order to eliminate the dimensionless relationship between variables, data standardization is to transform data from different orders of magnitude to the same order of magnitude, thus making the data comparable. So we used $x_{nor} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$ to normalize the raw value of each of the four predictors, then we used the normalized value to model and forecast." from line243 to line248 in page12.

We sincerely hope for your satisfaction with our revision. Thank you again for your kind suggestion.

4. line 254. What criterion is used to determine what are "weak items" with "small dimension coefficient".

Responses：Good suggestions. In the previous paper, we have neglected to explain the criterion is used to determine what are "weak items" with "small dimension coefficient".

In order to quantitatively compare the relative contribution of each item of our model to the evolution of the system, we calculated the relative variance contribution. The formula is as follows: $R_i = \frac{1}{n}\sum_{j=1}^{n}[\frac{T_i^2}{\sum_{i=1}^{14} T_i^2}], i = 1, 2, ..., 14$, Where n is the length of the data, $T_i = a_1x_1, a_2x_2, ..., a_{14}x_3x_4$ is the item in the equation. According to our previous research (Hong et al., 2007), the variance contribution of the real item reflecting the performance of the model has a large proportion, while the variance contribution of the false term is almost zero, so we delete the weak items of $R_i < 0.01$.

Now we have added the above explanation about the criterion is used to determine what are "weak items" from line250-257 in page12.

We sincerely hope for your satisfaction with our revision. Thank you again for your kind suggestion.

References:

[1] Hong Mei, Zhang Ren, Wu Guoxiong, et al., 2007. A Nonlinear Dynamic System Reconstruction of the Subtropical High Characteristic Index based on Genetic Algorithm. Chinese Journal of Atmospheric Sciences,31(2):346-352.

5. line 280 "forecast performance ... was better" than what??

Responses：Good suggestions. Now we have revised the sentence “From Fig. 2, forecast performance of T_1 and T_2 within 5 months was better.” as “From Fig. 2, forecast performance of T_1 and T_2 within 5 months was good.”

We sincerely hope for your satisfaction with our revision. Thank you again for your kind suggestion.

6.Section 6.2 - Table 5 The values reported here do not make sense. By construction, EOFs (the spatial patterns) are orthogonal, and the PCs (the time series) are uncor-related. L'Heureax et al report that the correlation between PC1 and PC2 (using the same HADISST data set) is 0.4 when the time series are detrended. This is the same value quoted in Table 5. Has T2 been detrended here also?

Responses：Good suggestions. In table 5, the values reported here do not make sense. Now we have deleted the Table5. In previous paper, we don't have detrended T_2. We have just smoothed the SSTA field before EOF. But due to a careless mistake, we use the data of a prediction experiment of 12 months to calculate the correlation coefficient in table5 and this is a mistake. We should use the all data from Jan.1951 to Dec.2010, a total of 720 months to calculate the correlation coefficient, so the correlation coefficients in the table5 are not correct in our pervious paper. Now we have recalculated with the right data. And after the time series are detrended, we have recalculated that the correlation between PC1 and PC2 is 0.4024, which is the similar as L'Heureax et al.

We sincerely hope for your satisfaction with our revision. Thank you again for your kind suggestion.

7.EOF1 is the cannonical ENSO pattern, and its time series is stronly correlated with the standard Nino indices (l'Heureaux et al give a value of 0.94 between their first EOF and the Nino3.4 index). In turn the Nino3.4 index is strongly correlated to the SOI, so that is difficult to see the correlation beteen T1 and the SOI being as small as the 0.4 given in Table 5.(This correlation is where the term ENSO ie El Nino - Southern Oscillation arises)

Responses：Good suggestions. In the answer of the pervious question, we mentioned that because of a careless mistake, correlation coefficient in the table5 formula is not correct. Now we have recalculated with the right data. In the answer to question 2, the correlation coefficient of T_1 and SOI in table2 is 0.773, which is consistent with the fact that the Nino3.4 index is strongly correlated to the SOI.

We sincerely hope for your satisfaction with our revision. Thank you again for your kind suggestion.

8. Acronyms need to be defined the first time they are used, eg EOF on lines 128-130

Responses：Good suggestions. Now we have defined Acronyms in the first time they are used.

We sincerely hope for your satisfaction with our revision. Thank you again for your kind suggestion.

9. Figure caption (line 912) for figure 1 in List of figures is incorrect, and different to that given with the figure itself (line 959).

Responses：Good suggestions. Now we have revised the figure caption (line

1027) for figure 1 in List of figures.

We sincerely hope for your satisfaction with our revision. Thank you again for your kind suggestion.

10.References are incomplete; there are at least 15 references that are not cited in the text, and a number that are cited but referenced.

Responses：Good suggestions. Now we have revised the list of references carefully and make all the references complete.

We sincerely hope for your satisfaction with our revision. Thank you again for your kind suggestion.