Reconstructing the past wind stresses over the tropical Pacific Ocean from 1875 to 1947

Ziwang Deng and Youmin Tang*

Environmental Science and Engineering,

University of Northern British Columbia,

3333 University Way, Prince George, BC, Canada, V2N 4Z9

Revised version for Journal of Applied Meteorology and Climatology

Dec. 17, 2008

Corresponding author: Youmin Tang, E-mail: ytang@unbc.ca

Abstract

An important step in understanding the climate system is simulating and studying the past climate variability, using either oceanic or atmospheric models, or both. Towards this goal, long term wind stress data, as the forcing of oceanic or climate models, is often required. In this study, we explored the possibility of reconstructing the past winds of the tropical Pacific using the historic sea surface temperature (SST) and sea level pressure (SLP) datasets. Four statistical models, based on principle component (PC) regression and singular vector decomposition (SVD), were developed for reconstructing monthly pseudo wind stress over the tropical Pacific for the period 1875-1947. The high frequency noise was removed from the raw data prior to the reconstruction. These models are SST-based PC regression (model 1), SLP-based PC regression (model 2), SST-based SVD (model 3) and SLP-based SVD (model 4).

The results show that reconstructed wind stresses from all models can account for more than half of the total variances. In general, the SLP is better than SST as predictor and the SVD method is superior to the PC-regression. Forced by these reconstructed wind stresses, an oceanic general circulation model (OGCM) can simulate realistic interannual variability of the tropical Pacific SST. However, the wind stress reconstructed by SST-based models leads to better simulation skill compared with that from SLP-based models. Finally a long-term wind stress dataset was constructed for the period from 1875 to 1947 by the SST-based SVD model, which provides a useful tool for studying the past climate variability over the tropical Pacific, especially for ENSO (El Niño and Southern Oscillation).

1. Introduction

Surface winds play an important role in the ocean-atmosphere interaction over the tropical Pacific Ocean. It determines the oceanic surface currents, the distribution of SST and thermocline depth, as well as the equatorial upwelling process. Surface winds have been a critical component in climate modeling and prediction. For example, a crucial task in ENSO prediction is to forecast variations of the surface winds as accurately as possible.

Currently, there are several reanalysis datasets of surface winds, which were generated by observations and the state-of-the-art models. They include the European Centre for Medium-Range Weather Forecasts (ECMWF) 40-year Re-Analysis 10m wind speed at a resolution of 2.5° x 2.5° from 1957 to 2006 (Simmons and Gibson 2000, hereafter referred to as ECMWF wind), the NCEP/NCAR Reanalysis 10m wind speed on a monthly T62 Gaussian grid for 1948-2006 (Kalnay et al.1996; hereafter referred to as NCEP wind) and the FSU/COAPS Tropical Pacific Ocean Pseudo-stress Analyses for 1961-1999 (Stricherz et al. 1997; hereafter referred to as FSU wind). There are also other wind datasets including the Southampton Oceanography Center (SOC) global wind stress (Josey et al. 2001), the University of Wisconsin at Milwaukee's Comprehensive Ocean–Atmosphere Dataset (Da Silva et al. 1994) etc. These reanalysis wind products have been playing a significant role in studying and predicting atmospheric and oceanic variations at time scales from days to decades (e.g., An and Wang 2000; Wang 1995). However these datasets only cover approximately 30-50 years because wind observations before the 1950s are unavailable. Vecchi et al.(2006) used the large-scale Indo-Pacific SLP gradient to study the long-term trend of the tropical Pacific atmospheric circulation but the reconstructed data is only a mean index. Long-term gridded surface wind data is often required for many purposes. For

example, a long-term oceanic simulation over 100 years, which would be more informative and insightful for studying the decadal/interdecadal variability in the Pacific Ocean, typically uses surface winds as forcing data. Other examples are ENSO predictability studies. A widely used strategy for the study of ENSO predictability involves performing retrospective predictions using dynamical models, by which the variations of predictability and the causes responsible for the variations are examined. So far, because of the limited length of forcing data, hindcasts performed for the study of ENSO predictability only cover a period of 20-40 years (e.g., Moore et al. 1998; 2006; Tang et al 2005). This test period covers few ENSO cycles, roughly 10-15, basically precluding statistically robust conclusions. To derive robust and statistically significant conclusions of ENSO predictability, a long period retrospective prediction should be generated, which requires a long period surface wind forcing data to initialize forecasts. Therefore, it is useful and has practical importance to explore the possibility of using long-term datasets available to prolong wind stress data.

The ocean has a long-term memory in the coupled atmosphere-ocean system. The adjustment of the atmosphere to ocean variables such as SST and upper ocean heat content is fast, making the atmosphere like a "slave" to the ocean at monthly or longer time scales. At these time scales, the contemporaneous correlation between surface wind stress and oceanic variables is associated with the atmosphere's rapid non-local adjustment to the oceanic anomaly patterns throughout the basin (Syu et al.1995; Tang et al. 2001).

The close relationship between SST and surface wind stress over the tropical Pacific has been documented in many studies (Kirtman et al. 1997,1998; Barnett et al. 1993; Eckert and Latif 1997; Tang et al, 2002; Wittenberg 2004; Galanti and Tziperman. et al., 2003, Tang et al. 2008). The hybrid coupled models of ENSO, composed of dynamical ocean and statistical

atmosphere, have similar or even better prediction skills compared with fully dynamical coupled models (Tang et al 2003; Tang et al. 2008). These facts strongly suggest the possibility of using historic SST datasets to reconstruct signals of the surface wind at monthly or longer time scales by statistical techniques. The direct dynamic link of large-scale structures between the surface wind and sea level pressure also allows to use SLP to reconstruct the surface wind (Zebiak 1990; Kushnir and Kaplan 1994).

This study attempts to reconstruct the surface wind for the period from 1875-1947 using statistical techniques and historic SST and SLP datasets. Emphasis is placed on the reconstruction of low-frequency components that the SST and SLP datasets can resolve well. The low-frequency components of the surface wind are important for modeling and predicting climate variability. The original observations that produced SST or SLP historic datasets are gappy before 1950s and rather sparse for most of the 19th century, thereby reliable signals that could be resolved and characterized in these historic datasets are probably low-frequency components with large spatial scales. Indeed, it has been found that ENSO signals can be fairly well resolved and represented in these datasets (e.g., Chen et al. 2004; Kaplan et al. 1998; Xue et al. 2003). On the other hand, the statistical reconstruction of the surface winds using SST is based on the assumption that the contemporaneous correlation between wind stress and oceanic variables is associated with the atmosphere's rapid non-local adjustment to oceanic anomaly patterns throughout the basin. This assumption only holds for low-frequency components on monthly or longer time scales. The high frequency components of the winds are primarily determined by the atmospheric internal dynamical process. Thus the surface winds (or wind stress) reconstructed in this study would mainly be used for studying large-scale low frequency variabilities such as ENSO or decadal variability.

The remainder of this paper is organized as follows. In section 2, a brief description of the used data and methods is presented. Section 3 presents the results of the cross validation of the models. Section 4 describes the ENSO simulations forced by the reconstructed wind stresses. The possible uncertainties in the reconstructed wind stress are discussed in section 5. Finally, a brief summary of this paper is included in section 6.

2. Data and methods

2.1 Data

In this study, we chose the NCEP pseudo wind stress as the predictand (Kalnay et al.1996), and the Extended Reconstructed Sea Surface Temperature (Smith et al. 1996; 1998; 2002; 2003; 2004, hereafter referred to as ERSST) and the Global Mean Sea Level Pressure Data Set GMSLP2 (Basnett and Parker 1997) as predictors. These datasets were chosen because they span a long period and have fine resolution. They also have been well verified and widely used in climate analysis and prediction (e.g., Trigo et al. 2004; Klotzbach 2007;Ummenhofer et al. 2008; Vecchi et al. 2008).

To examine the possible impact of the selection of datasets on the reconstructed winds, we also explored other datasets (see section 5) including the ECMWF wind, FSU wind, the Global Ocean Surface Temperature on a monthly 1° area grid for the period 1870-2007 (Rayner et al. 2003, referred to as HadISST) and the Kaplan Extended SST V2 on a 5.0 degree latitude by 5.0 degree longitude global grid for the period 1856-2006 (Kaplan et al. 1998, referred to as KESST). The domain of all the data is confined within the tropical Pacific (30°S-30°N, 120°E-70°W).

The FSU wind is actually a pseudo wind stress. The NCEP and ECMWF wind are converted into the pseudo wind stress by:

$$\tau = \| \mathbf{U_{10}} \| \mathbf{U_{10}} \qquad (1)$$

where $\mathbf{U_{10}}$ is vector wind at 10m above the sea surface.

2.2 Data preprocessing

The wind usually contains high frequency noise that is not predictable, thus filtering techniques are applied to the datasets prior to building reconstruction models. The filtering techniques also allow us to address the statistical relationship of two variables in the frequency bands of interest.

The frequency decomposition method used is the Fourier transform (William et al 2001). Given a time series $y(t_n)$ of N samples ($t_n = 1, 2, ..., N$),

$$y(t_n) = \frac{1}{2} a_0 + \sum_{k=1}^{N/2} [a_k \cos(\omega_k t_n) + b_k \sin(\omega_k t_n)] \qquad (2)$$

where the angular frequency is $\omega_k = 2\pi k / (N\Delta t)$. When all frequency components with $\omega_k > 2\pi M / N$ are removed, the truncated time series at M<N/2 is obtained by

$$\tilde{y}(t_n) = \frac{1}{2} a_0 + \sum_{k=1}^{M} [a_k \cos(\omega_k t_n) + b_k \sin(\omega_k t_n)] \qquad (3)$$

Using equations (2) and (3), anomalies of the wind, SST, and SLP are decomposed into three frequency bands: the low frequency band (LFB), the middle frequency band (MFB) and the high frequency band (HFB). The anomalies are computed relative to climatology of the

period from 1961-1990. In this study, the LFB is defined by the period greater than 30 years , the HFB by the period less than 6 months, and the remaining, i.e., subtracting the sum of the LFB and HFB from the raw data, is defined as the MFB. In this study, we only reconstruct the LFB and MFB components of wind, and ignore the HFB that is almost unpredictable by statistical models and only plays little roles in simulating low-frequency climate variability (Wittenberg, 2004).

The physical basis for the wind reconstruction using SST is the no-local large-scale response of the atmosphere to the ocean. In order to extract the large-scale structure of atmospheric and oceanic variables, the empirical orthogonal function (EOF) analysis was employed for the LFB and the MFB of the winds, SST and SLP respectively. The leading EOF modes accounting for statistically significant variances will be used for model construction, which is determined by the statistical test method of North (1982). It should be noted that the EOF analysis of the predictor field was performed for the whole period from 1875-2005 to keep consistence of the spatial pattern between training and reconstruction period. Because the wind stress is a vector, its EOF analysis is thus performed for a joint matrix composed of both the zonal and meridional wind components.

2.3 Wind reconstruction models

For each statistical model, the reconstructed wind is denoted by:

$$\hat{\tau} = \overline{\tau} + \hat{\tau}'_{LFB} + \hat{\tau}'_{MFB} \qquad (4)$$

where $\overline{\tau}, \hat{\tau}'_{LFB}$ and $\hat{\tau}'_{MFB}$ are the annual cycle, and the predicted LFB and MFB respectively. The annual cycle is the climatology for the period 1961-1990, which is assumed to be

unchanged for the entire period from 1875-2005. The term $\hat{\tau}'_{LFB}$ or $\hat{\tau}'_{MFB}$ can be written as follows:

$$\hat{\tau}'(x,y,t) = \sum_{k=1}^{K} \hat{a}(t,k) e(x,y,k) \qquad (5)$$

where $\hat{\tau}'(x,y,t)$ is $\hat{\tau}'_{LFB}$ or $\hat{\tau}'_{MFB}$, $\hat{a}(t,k)$ is the principal component (PC) or SVD expansion coefficient at time t, corresponding to the *k*th wind EOF or SVD mode $e(x,y,k)$, and K is the number of truncated wind EOF (or SVD) modes used in statistical models.

Four statistical models are developed for the wind reconstruction. They are the SST-based PC-regression model (model 1), SLP-based PC-regression model (model 2), SST-based singular vector decomposition (SVD) model (model 3) and the SLP-based SVD model (model 4).

For the PC-regression models (model 1 and model 2), the wind PCs are predicted by:

$$\hat{a}(t,k) = \sum_{j=1}^{J} C_{\tau}(j,k) b(t,j) \qquad (6)$$

where $C_{\tau}(j,k)$ is the regression coefficient between the *j*th PC of predictors and the *k*th PC of wind; $b(t,j)$ is the value of the *j*th PC of predictors at the time *t* during the reconstructed period from 1875 to 1947, and *J* is the total number of PCs used in the reconstruction models (shown in Table 1). The coefficient $C_{\tau}(j,k)$ is obtained based on the training period from 1948-2005 with the cross-validation scheme.

For the SVD models (model 3 and model 4), the scheme employed in this study is similar to that by Syu and Neelin (1995), namely that the wind SVD expansion coefficients for the reconstructed period are obtained by:

$$\hat{a}(t,k) = C(k)\hat{h}(t,k) \qquad (7)$$

where *C(k)* is the regression coefficient between the *k*th pair SVD expansion coefficients between predictors and wind, obtained from the training period from 1948-2005 with the cross-validation scheme , $\hat{h}(t,k)$ is the corresponding SVD expansion coefficient of predictors at the time *t* during the reconstructed period from 1875-1947.

The details of individual statistical models are summarized Table 1. In constructing these statistical models, we only consider the statistical relationship between predictands and predictors in the same frequency bands. In other words, the LFB (MFB) wind is only related to the LFB (MFB) of SST or SLP. Since the LFB (MFB) spans a broad range of frequency and the statistical methods are linear, this strategy should be reasonable and realistic.

A K-fold cross validation scheme is used for the validation of the statistical models (Kohavi 1995). The cross-validation procedure involves dividing the data record into several segments, selecting one segment as test data and the rest as training data. The statistical models are built using the training data only, and the model predictions are tested on the test data. Next, another segment is selected as the test data, and a new version model is built. This is repeated until the entire data record has been used for testing. The cross-validation scheme ensures that no training data are used for testing the prediction skills, hence artificial skill associated with the overfitting problem (von Storch and Zwiers, 1999) can be effectively eliminated.

In this study, the record of 1948-2002 was divided into 11 segments of 5 years each. Thus there are 11 models, each of which produces a reconstructed wind data for the period from 1875-1947. The final reconstructed wind stress is the ensemble mean averaged over the 11 reconstructed winds. To validate the prediction models, several measures are used including

anomaly correlation, root mean square error (RMSE) and the explained variance. The explained variance is defined as below

$$V = 1 - \frac{\frac{1}{NM-1}\sum_{t=1}^{N}\sum_{i}^{M}(\tau o_{ti} - \tau p_{ti})^2}{\frac{1}{NM-1}\sum_{t=1}^{N}\sum_{i}^{M}(\tau o_{ti})^2} \qquad (8)$$

Where τo_{ti} and τp_{ti} are observation and prediction at time *t* and on grid *i*, *N* is the length of samples and *M* is the number of total grids.

2.4 Ocean model

The wind reconstruction models could be evaluated by cross validation skill for the period when observed wind is available. An ocean general circulation model (OGCM) is used to test the reliability of the reconstructed wind for the past period when the observed wind is not available. The ocean model used is the latest version of Ocean Parallelise Model (OPA 9.1). This is a primitive equation global oceanic model widely used in the ocean community (Tang et al. 2004; Moore et al. 2006). Detailed formulation of the ocean model is described in Madec et al. (2008). The ocean model is configured for the tropical Pacific Ocean between 30°N and 30°S and between 122°E and 70°W. The horizontal resolution in the zonal direction is 2°, while the resolution in the meridional direction is 0.5° within 5° of the equator, smoothly changing up to 2.0° at 30°N and 30°S. There are 31 vertical levels with 17 concentrated in the top 250m of the ocean. The time step for integration was 1.5 hours. The boundaries were closed.

The ocean model was spun up by the climatologic wind for 500 years, and then forced by the following wind stress

$$\tau_s = C_D \rho_a \tau \qquad (9)$$

where $C_D = 0.00153$ is the drag coefficient (Huang et al. 1986; Balmaseda et al. 1994), $\rho_a = 1.2\text{kgm}^{-3}$ is the air density, τ is the pseudo wind stress.

3 Statistical models and their validations

3.1 Signal decomposition

As mentioned above, the NCEP wind (1948-2005), the ERSST (1875-2005) and the GMSLP2 (1875-2005) are chosen for reconstructing the historic wind. The training period is 1948-2005 during which all data are available. We focus on the low-frequency component wind here, i.e., the LFB and the MFB. The LFB and MFB components of wind, SLP and SST account for around 10% and 45%, 8% and 47%, and 19% and 72% of total variance[1], respectively.

We will first examine the predictands and the predictors before building the statistical models. Figure 1 shows the first EOF mode of the LFB and MFB components of SLP, SST and the wind, derived from the period 1948-2005 for the wind and the period 1875-2005 for both SLP and SST. As can be seen, the LFB signals appear to be spatially uniform for SLP (Fig.1a) and SST (Fig.1c) but have relatively large amplitude in the equatorial central Pacific for the wind (Fig.1e), whereas the strongest MFB signals occur in the central and the eastern equatorial Pacific for SST (Fig.1d) and in the extropical Pacific for both SLP (Fig. 1b) and

[1] The total variance is defined as: $S = \frac{1}{MN-1}\sum_{t=1}^{N}\sum_{i=1}^{M} Z(t,i)^2$, where N and M are the time length and the total number of spatial grids for variable Z.

wind (Fig.1f). Their spatial distributions of standard deviation also display similar features (not shown). The MFB signals of these variables characterize the primary features of the mature phase of El Niño, namely that when the prevailing easterly trade wind weakens over the central and the eastern tropical Pacific (Fig.1f), the upwelling weakens and the SST increases over the equatorial central and eastern Pacific (Fig. 1d), resulting in negative SLP anomaly in the east and positive SLP anomaly in the west (Fig. 1b). Figure 2 shows the corresponding time series (PC1) of the leading EOF modes. As seen in Fig. 2a, the three PC1s of LFB show a similar trend during their common period although there are some shifts in their phase. There is a phase shift between SST and SLP from 1875-1960, suggesting that the predicted LFB winds have obvious differences when different variables are used as predictors. This uncertainty in the prediction of LFB winds will be further addressed in section 5. Figures 2b and 2c indicate that all the PC1s of the MFB agree well with each other and display strong positive correlation with Niño3 SSTA index (not shown). Figure 2 discloses the possibility of reconstructing the winds for the period from 1875-1947 using historic SST/SLP datasets, especially for the MFB components.

In this study, we only reconstruct the low-frequency winds (the LFB and the MFB) using several leading EOF (SVD) modes. The omission of higher-order EOF modes and high frequency wind might influence ENSO simulation. To investigate this issue, we run the OGCM twice for the period from 1948 to 2002, one forced by the raw wind stress and the other by the filtered wind stress constructed by the leading EOF modes of LFB and MFB components. The leading EOF modes are determined by the North rule (North et al. 1982) as shown in Table 1. Comparing SSTA simulation skills shown in figure 3c with figure 3a

reveals that the filtered wind stress can produce simulation skills comparable to those by the original wind stress, even better for some regions. The Niño3 SSTA index produced with the filtered wind stresses is in good agreement with that produced with the raw wind stresses and their correlation coefficient is up to 0.9. Thus, high frequency winds are not important in ENSO simulation. In fact, there are many hybrid coupled models that have good ENSO prediction skills comparable with those of fully coupled models. In these hybrid coupled models, the atmosphere components are statistical and only have low frequencies derived from the relationship between winds and oceanic variables.

3.2 Cross-validation skills

As discussed above, four statistical models were developed for the LFB and the MFB winds respectively, using either PC regression or SVD techniques and cross-validation scheme. These models are summarized in Table 1. The total component of the reconstructed wind is composed of the predicted LFB and MFB winds plus climatologic seasonal cycle. In this section, we will validate these models.

Figure 4 shows cross validation skills of the predicted MFB zonal wind against the observed counterpart for the period from 1948-2002. The four models present good prediction skills over the Niño4 area, with correlation coefficients greater than 0.6. Given the same statistical method, the SLP-based models seem better than SST-based models for the extratropical regions off the equator (model 2 vs model 1; model 4 vs model 3). This is not surprising since the wind-SLP quasi-geostrophical balance holds better in the regions farther from the equator. For the same predictors, the SVD technique is superior to the PC-regression (model 3 vs model 1; model 4 vs model 2). In general, model 4 seems the best, followed by model 3, 2 and 1.

Figure 5 shows cross validation skills of the predicted low-frequency zonal wind (LFB+MFB) by the four models for the period 1948-2002. The statistical test indicates that the cross-validation correlation skills of the four models are all significant over the tropical Pacific at a confidence level of 99%. There are large correlations in the equatorial central Pacific in each of the four models. The similarity of spatial patterns between Figures 4 and 5 suggest the dominance of MFB in the reconstructed winds, due to the fact that the MFB accounts for much more variance than LFB as discussed in section 3.1. In addition, Fig.5 shows higher skills than Fig. 4 in the whole domain, indicating that the inclusion of LFB components increases the overall skill of reconstructed winds. This is because the long-term trend of the LFB during the period 1948-2002, as shown in Fig. 2, can be captured effectively by these models. A further comparison between Figures 5 and 4 reveals that the improvement of LFB predictions to the overall skill of low-frequency winds is more significant in models 1 and 2 than in models 3 and 4, especially in the RMSE skill. This is because that PC-regression uses all leading PCs of predictor to predict each individual wind PC. Thus, the phase shift between the predictor (SST or SLP) PC1 and wind PC1, as shown in Fig. 2a, can be considered by other PCs of the predictor. This is different in SVD models where the regression is only built for a pair of time series, i.e., the expansion coefficient of wind SVD mode and corresponding SVD mode of predictor, as indicated in Eq. (7).

It should be noticed that the capability of these models to capture the linear trend of the training period might be biased in the reconstructed period due to trend uncertainties as reflected in SST and SLP data (Fig 2a). We will further discuss these uncertainties in following sections.

Compared with model 1, the other three models also have higher correlations in the off-equatorial regions, especially for model 2 and 4. The large RMSE appears in the region where the standard deviation of the MFB is large (not shown), indicating that all models have difficulty in predicting the amplitude of strong interannual variability. This is a common problem of all statistical models. In terms of the cross-validated correlation and RMSE skills, model 4 is the best for the zonal wind with high correlations not only in the equatorial belt but also in a large off-equatorial domain. However in the following discussions, we find that a good prediction in the off-equatorial region has little contribution to SST simulation. The cross-validation skills of the meridonal winds show that the large correlations occur in the off-equatorial regions where there are relatively large variances as shown in the first EOF mode (not shown). Correspondingly, the RMSE has large values in these regions due to the difficulty in predicting strong variability (not shown). These skill differences among the four models are also reflected in the total variance accounted for as calculated by (8) and shown in Table 2. Several cases are chosen in Table 2, including the variance of both zonal and meridional winds in the whole domain of data (V0); the variance of both zonal and meridional winds across the basin between 15°S-15°N (V1); the variance of zonal winds in the whole domain of data (V2); the variance of the meridional winds in the whole region of data (V3) and the variance of the zonal winds over Niño4 region (V4). As shown in Table 2, all predicted winds explain more than 50% of total variance, especially for model 4. In general, SLP is better than SST as the predictors whereas the SVD method is superior to the PC-regression, which is consistent with the correlation and RMSE skills. The better skills derived using the SLP as predictors are probably due to a direct dynamic link

between winds and SLP. On the other hand, the SVD method directly extracts the correlated modes between the predictors and predictands, reducing the uncorrelated noise.

An interesting feature in Table 2 is that while the four models have skill differences in many regions, they all produce a consistent good skill for the zonal wind prediction in the Niño4 region. This may explain why the predicted winds by the four models lead to similar SSTA simulation skills when they are used to force the OGCM. It is well-known that the zonal wind over Niño4 plays a very important role in ENSO development and formation, and usually is a good precursor of ENSO.

4. The reconstructed winds and their driven ENSO simulation.

After these statistical models are built, we can reconstruct the winds for the period 1875-1947 period using SST or SLP historic datasets. The reconstructed winds are the ensemble mean of individual models built using cross-validation scheme as discussed above. Figure 6 shows the PC1s of the LFB and MFB wind stress reconstructed by the four models for the period 1875-1947. Figure 6b and 6c show that all models produce consistent reconstruction of MFB components, suggesting the validity of the reconstructed MFB winds. There are differences in the LFB predicted by different models as shown in Fig. 6a, indicating uncertainties in the reconstructed LFB due to differing predictors and a relatively short training period, as discussed above.

The validation of the reconstructed wind and its capability in modeling climate variability are further explored by the OGCM. Figure 7 shows the simulated Niño3 SSTA indices by the four reconstructed winds from 1875-1947. As has been shown, Niño3 SSTA amplitude is

always weak in all simulations compared with the observation. This should be due to the problems inherent to the OGCM itself, most probably related to parameterization of vertical advections. In fact, the weak amplitude of simulated Niño3 SSTA also exists even using the observed winds to force the OGCM (Deng et al. 2008). The simulated Niño3 indices by different winds have some differences. The correlation coefficients between the simulated and observed Niño3 SSTA indices are 0.88, 0.60, 0.87 and 0.71, respectively, for the winds of model 1 2, 3 and 4. Obviously, the reconstructed winds by model 1 and 3 are better than model 2 and 4 in terms of Niño3 SSTA simulation. This is interesting because the SLP-based models (model 2 and 4) have better skills for the winds themselves, but their driven SSTA skills are worse than those by SST-based models. It might be due to the fact that the observed SSTA information has been utilized in constructing statistical models. In addition, the SLP-based models may include some components that are not related to the SSTA variability, probably spoiling the SSTA simulation.

Shown in figure 8 are the correlation and RMSE skills of the SSTA simulation produced by the reconstructed winds against the observed SSTA. From figure 8, one can find that the winds produced by the four models lead to useful correlation skills in the equatorial central and eastern Pacific. A comparison of these skills shows that their differences occur in the central and the eastern equatorial Pacific where the models 1 and 3 outperform the models 2 and 4. The winds reconstructed by model 3 produce the best SSTA simulation in the tropical equatorial Pacific. In contrast, the winds by model 2 appear to have the worst ENSO simulation skill, which might be due to a spurious trend in the reconstructed LFB wind as shown in Fig. 6a. Figure 8 is consistent with figure 7, namely that the winds from SST-based

models (model 1 and model 3) are superior to those from SLP-based models (model 2 and mode 4) in terms of their capability in simulating ENSO variability. Comparing figure 8 with figure 3 reveals that the simulated SSTA correlation skills produced by the reconstructed winds of model 1 and model 3 are very similar to those produced by the filtered winds and observed winds, in both the magnitude and pattern structure, justifying the use of the reconstructed winds in ENSO simulation and prediction.

A further analysis was performed for the zonal wind reconstructed by model 3 that produces the best ENSO simulation skill. Shown in figure 9 is the wavelet power spectrum(Torrence and Compo 1998) of Niño4 wind index and observed Niño3 SSTA index. Since the reconstructed winds have only LFB and MFB components, the Niño3 SSTA index is also decomposed into corresponding LFB and MFB components prior to the analysis. The Niño4 wind index shows a striking interannual variability, similar to the Niño3 SSTA index. Their correlation is 0.87, 0.83, and 0.86 during the period from 1875-2005, 1875-1947 and 1948-2005. A consistent correlation during the training and test periods suggests that the reconstructed wind probably contains as much useful information as the observed wind. The wavelet power spectrum of the Niño4 wind index indicates that there are significant interannual variability with the period of 4-8 years from 1900-1920, approximately 4 years from 1930-1960 and 2-4 years from 1970-2005 (Figure 9a), basically consistent with that in the Niño3 SSTA (Figure 9b), except the period 1890-1920 when the variability of the 2-4 yr period is stronger in SST than in winds. Fig 9b shows stronger ENSO variability than Fig.9a before 1947, suggesting that the reconstructed wind reduces signal amplitude, which is a common shortcoming in statistical models. Also there are some visible differences in LFB components between Fig. 9a and Fig9b, suggesting the uncertainty of reconstructed LFB.

Comparing figure 9a and figure 9b reveals that the interannual signals are better resolved in the wind than in SSTA. This seems reasonable because a statistical model acts like a filter, emphasizing the interannual variability and alleviating other frequencies.

5. Possible impacts of dataset selection on the reconstructed winds

Generally, any statistical model is data dependent. As mentioned in section 2.1, there are several datasets of SST, SLP and wind available, thus, it is necessary to explore the possible impacts of data selection on the reconstructed wind stress. Since the SST-based model (model 3) is chosen to reconstruct the past winds in this study, we confine our following analysis to different datasets of SST and wind. Also, the wind reconstruction in all models is implemented by predicting the time series of SVD/EOF leading modes. For simplicity, we only explore and compare the time series of the first mode of the datasets. This first mode explains a major part of the total variance of SST and winds.

Figure 10 shows the variation of normalized PC1s of the low frequency zonal winds (LFB+MFB) of the three datasets from 1961-1999: NCEP wind, ECMWF wind and FSU wind, accounting for 16.0%, 17.9%, 13.2% of total variance, respectively. As can be seen, the ECMWF and NCEP PC1 agree well with each other, with a correlation coefficient of up to 0.89, and both have a visible difference from FSU wind. In terms of MFB components (shown in Fig.11a), all three PC1s have good relationships with the lowest correlation of 0.62, suggesting that the selection of zonal wind is not very critical in terms of ENSO study. However, the FSU wind has a different trend from the NCEP and ECMWF winds, suggesting some uncertainties in the LFB components of winds.

Figure 10b shows the PC1s of three meridional winds. As can be seen, the ECMWF and FSU winds agree well each other with a good correlation coefficient up to 0.95, and both are quite different from the NCEP meridional wind. This significant difference is mainly due to the LFB components. Josey et al.(2001) and Kistler (2001) also reported that the climatological mean of NCEP zonal wind is weaker than that of ECMWF and FSU zonal wind over the equatorial region. Thus, the choice of wind datasets could have significant impacts on the reconstructed LFB wind. However, the relationships of MFB winds are very good with a minimum correlation of 0.88, as shown in Figure 11b, indicating that the choice of wind datasets probably has little impact on MFB winds.

Figure 12 shows the normalized PC1s of the LFB and MFB components of three SST datasets, ERSST, HadISST and Kaplan SST, for the period from 1875 to 2005. Generally, the three SST datasets agree with one another well in both LFB and MFB components, especially for the MFB components. These results suggest that the reconstructed winds are not sensitive to different SST datasets, especially for the MFB components.

We also compared the other leading modes of the three SST datasets, such as PC2s and PC3s, and found that these leading PCs of different datasets also have a very consistent variation, especially for the MFB components (not shown).

6. Discussion and summary

An important step in understanding the climate system is to simulate and study the past climate variability using oceanic or atmospheric models, or both. Towards this goal, long term wind stress data, as the forcing of ocean or climate models, is often required. In this study, we explored the possibility of reconstructing the past winds over the tropical Pacific using long-term SST/SLP datasets. Four statistical models were developed with a cross-

validation scheme for the low-frequency component wind reconstruction, including SST-based and SLP-based PC-regression and SVD models. The results show all four models have statistically significant cross-validation correlation skills over the tropical Pacific, especially over the equatorial central and eastern Pacific where the surface winds play a critical role in forming and maintaining ENSO cycle. The models are capable of capturing interannual variability in the winds, and the reconstructed winds can generate realistic ENSO variability when the OGCM is forced.

It is impractical to retrieve high frequency signals of winds using the SST/SLP historic dataset and statistical models. Thus, the temporal filtering and EOF truncation are performed prior to building statistical models. It was found that the treated wind stress could generate the tropical Pacific SST simulation as well as the raw wind stress, suggesting that the high frequency winds are not important in modeling large-scale climate variability at the interannual or longer time scales.

It was found that the choice of wind datasets has significant impact on the reconstructed LFB components, but probably little impact on MFB winds. On the other hand, the reconstructed winds might not be sensitive to different SST datasets, especially for the MFB components.

We also performed several extra analyses to seek better reconstruction models of winds. First, the nonlinear regression based on neural network (NN) was explored to build nonlinear statistical prediction model of winds using SST or SLP. This is similar to PC-SST and PC-SLP regression models (model 1 and 2) presented in the proceeding sections, but using NN instead of linear regression. With a large amount training of NN, we found that the NN is not superior to linear regression in terms of cross-validation skills of winds. This is most

probably because the relationship of winds to SST or SLP is approximately linear at the interannual and longer time scale. In addition, the data used here are monthly averages which can drastically relieve the nonlinearity (Yuval and Hsieh 2002). Second, the wind has direct dynamical association with the pressure gradient rather than pressure itself, described by the momentum equation. Thus, we also constructed statistical models using pressure gradients as the predictors. The cross-validation skills from this approach are very poor, much worse than the skills presented in this paper. Third, both SST and SLP were used as the predictors to construct statistical models. The results show that such joint predictors cannot improve cross-validation skills. Finally, the raw winds instead of low-frequency signals were directly reconstructed by statistical models. The results show that these statistical models neither have good cross-validation skills nor explain a large portion of the total variance. This is not surprising because the high frequency signals (noise) often spoil the statistical training, leading to poor cross-validation skills.

Several cautions should be borne in mind when reconstructing the past wind. First, the predictors, SLP and SST, are coarse and gappy before the 1950s, thus, the datasets might only contain useful information on large-scale interannual or decadal/interdecadal climate variability. Such information might be sufficient to describe and characterize some large-scale climate modes such as ENSO, but not enough for relatively short and small scale variability such as Madden Julian Oscillation. Thus, the reconstructed winds should be mainly used for modeling large-scale climate signals. Second, the reconstructed winds do not contain high frequency signals, thereby precluding the use for studying the stochastic role in ENSO. Third, the reconstructed winds, subject to a common statistical problem, that of generally underestimating the amplitude of wind anomaly. Finally, we used a running

window of 5 years to implement cross-validation scheme for developing statistical models. The window length of 5 years was arbitrary. We also performed several sensitivity experiments, with changing the window length, and got similar results. As well, we defined the HFB, the LFB and the MFB using the oscillation period under 6 months, over 30 years, and 6 months to 30 years. One may argue these definitions might not be an objective choice. Nevertheless, this study is, to date, the first work to reconstruct the past winds for the tropical Pacific. The reconstructed wind dataset has provided a useful source for studying ENSO (Deng et al. 2008; Deng and Tang 2008;Tang et al 2008). It has been posted on the internet and is freely loadable (http://web.unbc.ca/~ytang/wind.html). Therefore, while this work has limited theoretical contribution, it has practical significance in studying climate variability of the tropical Pacific, especially for ENSO.

Acknowledgment This work was supported by Canadian Foundation for Climate and Atmospheric Sciences Grant GR-523 to Y. Tang. We thank reviewers for helpful comments.

Reference

An, S.I. and B. Wang, 2000: Interdecadal Change of the Structure of the ENSO Mode and Its Impact on the ENSO Frequency. *J. Climate*,**13**,2004-2015.

Balmaseda, M.A., D.L.T.,Anderson and M.K., Davey. 1994: ENSO prediction using a dynamical ocean model coupled to statistical atmospheres: *Tellus*, **46(A),** 497-511.

Barnett, T.P., M Latif, Graham, N. E., Flügel, M., Pazan, S., and W. B., White, 1993: 'ENSO and ENSO related predictability, I, Prediction of equatorial Pacific sea surface temperature with a hybrid coupled ocean-atmosphere model'. *J. Climate*, **6**, 1545- 1566.

Basnett, T. and Parker, D. E., 1997: Development of the Global Mean Sea Level Pressure Data Set GMSLP2. CRTN 79, Hadley Centre, Met Office, Exeter, UK, 16pp.

Chen D., M. A. Cane, A. Kaplan, S. E. Zebiak, and D. Huang, 2004: Predictability of El Niño over the past 148 years. *Nature*, **428**,733–736.

Da Silva, A. M., C. C. Young, and S. Levitus, 1994: Algorithms and Procedures. *Atlas of Surface Marine Data. NOAA Atlas NESDIS,* **6**, 74 pp.

Deng, Z., and Y. Tang, 2008: The retrospective prediction of ENSO from 1881-2000 by a hybrid coupled\ model - (II) Interdecadal and decadal variations in predictability . *Climate Dynamics* (in press).

Deng, Z., Y. Tang and X. Zhou, 2008: The retrospective prediction of ENSO from 1881-2000 by a hybrid coupled model - (I): SST Assimilation with Ensemble Kalman Filter. *Climate Dynamics* (in press).

Eckert, C., and M., Latif, 1997: Predictability of a Stochastically Forced Hybrid Coupled Model of El Niño. *J. Climate*, **10**, 1488-1504.

Galanti, E. and E. Tziperman et al., 2003: A study of ENSO prediction using a hybrid coupled model and the adjoint method for data assimilation. *Monthly Weather Review*,131,2748-2764.

Huang, N. E., L. F. Bliven, S. R. Long, and P. S. Deleonibus, 1986: A Study of the Relationship Among Wind Speed, Sea State, and the Drag Coefficient for a Developing Wave Field, *J. Geophys. Res.*, ***91***, 7733–7742.

Josey, S. A., 2001: A Comparison of ECMWF, NCEP/NCAR and SOC surface heat fluxes with moored buoy measurements in the subduction region of the north-east Atlantic. *J. Climate*, **14**, 1780–1789.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project, *Bull. Amer. Meteor. Soc.*, **77**, 437-470.

Kaplan, A., M. Cane, Y. Kushnir, A. Clement, M. Blumenthal, and B. Rajagopalan, 1998: Analyses of global sea surface temperature 1856-1991, *Journal of Geophysical Research*, **103**, 18567-18589.

Kirtman, B. P., 1997: Oceanic Rossby wave dynamics and the ENSO period in a coupled model. *J. Climate*, **10**, 1690–1704.

__, and P. S. Schopf, 1998: Decadal variability in ENSO predictability and prediction. *J. Climate*, **11**, 2804-2822.

Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.*, **82**, 247–267.

Klotzbach, P.J., 2007: Revised Prediction of Seasonal Atlantic Basin Tropical Cyclone Activity from 1 August**.** *Wea. Forecasting*, **22**, 937–949.

Kohavi, R., 1995: A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence,* **2**, 1137–1143.

Kushnir, Y. and A. Kaplan, 1994: Dynamical constraints for the analysis of sea level pressure and surface wind over the world ocean. *Proceedings, International Winds Workshop*, 91-101.

Madec, G., 2008: NEMO ocean engine. Note du Pôle de modélisation, Institut Pierre-Simon Laplace (IPSL), France, NXX, 193pp.

Moore, A. M., and R. Kleeman, 1998: Skill assessment for ENSO using ensemble prediction. *Quart. J. Roy. Meteor. Soc.*, **124**,557–584.

—, J. Zavala-Garay, Y. Tang, R. Kleeman, J. Vialard, A. Weaver, K. Sahami, D. L. T. Anderson and M. Fisher 2006: Optimal Forcing Patterns for Coupled Models of ENSO *J. Climate*, **19**, 4683-4699.

North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions, *Mon. Weather Rev.*, **110**, 699– 706.

Rayner, N.A., D.E. Parker, E.B. Horton, C.K. Folland, L.V. Alexander, D.P. Rowell, E.C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geoghys. Res.*, **108(D14)**, 4407, doi:10.1029/2002JD002670.

Simmons, A. J., and J. K. Gibson, 2000: The ERA-40 project plan. *ECMWF ERA-40 Project Report Series 1*, 63 pp.

Smith, T.M., R.W. Reynolds, R.E. Livezey, and D.C. Stokes, 1996: Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *J. Climate*, **9**, 1403-1420.

Smith, T.M., and R.W. Reynolds, 1998: A high-resolution global sea surface temperature climatology for the 1961-90 base period. *J. Climate*, **11**, 3320-3323.

Smith, T.M., and R.W. Reynolds, 2002: Bias corrections for historic sea surface temperatures based on marine air temperatures. *J. Climate*, **15**, 73-87.

Smith, T.M., and R.W. Reynolds, 2003: Extended reconstruction of global sea surface temperatures based on COADS data (1854-1997*). J. Climate*, **16**, 1495-1510.

Smith, T.M., and R.W. Reynolds, 2004: Improved extended reconstruction of SST (1854-1997). *J. Climate*, **17**, 2466-2477.

Stricherz, J. N., D. M. Legler, and J. J. O Brien,1997: TOGA pseudo-Stress atlas 1985-1994:II Tropical Pacific Ocean. COAPS Rep. 97-2,177 p. [Available from COAPS, The Florida State University, Tallahassee FL 32306].

Syu H.-H., J. D. Neelin, and D. Gutzler, 1995: Seasonal and interannual variability in a hybrid coupled GCM. J. Clim., 8, 2121-2143.

Tang, Y. and W. W, Hsieh, 2001: Coupling neural networks to incomplete dynamical systems via variational data assimilation, *Mon. Wea. Rev*., **129**, 818-834.

Tang, Y., and W. W, Hsieh, 2002: Hybrid coupled models of the tropical Pacific -- ENSO prediction . *Climate Dynamics*, **19**, 343-353.

— and —, 2003: ENSO simulation and predictions using a hybrid coupled model with data assimilation. *J. of Japan Met. Soc*., **81**, 1-19.

Tang Y, Z.,Deng, X.,Zhou, Y., Cheng and D., Chen, 2008: Interdecadal variation of ENSO predictability in multiple models. *J Clim* (in press).

Tang Y, R.,Kleeman and A., Moore, 2004: SST assimilation experiments in a tropical Pacific Ocean model, *J Phys. Oceangr*. **34**, 623-642

—,— and A., Moore, 2005: On the reliability of ENSO dynamical predictions. *J. Atmos. Bci*., **62**, 1770-1791.

Torrence, C. and G. P. Compo, 1998: A Practical Guide to Wavelet Analysis. *Bull. Amer. Meteor. Soc.*, **79**, 61-78.

Trigo R.M, I. F. Trigo, C. C. DaCamara and T. J. Osborn, 2004: Climate impact of the European winter blocking episodes from the NCEP/NCAR Reanalyses. *Climate Dynamics*, 23: 17–28.

Ummenhofer, C.C., A. Sen Gupta, M.J. Pook, and M.H. England, 2008: Anomalous Rainfall over Southwest Western Australia Forced by Indian Ocean Sea Surface Temperatures. *J. Climate,* 21, 5113–5134.

Vecchi, G. A., B. J., Soden et al., 2006: Weakening of tropical Pacific atmospheric circulation due to anthropogenic forcing. *Nature*, **44**,73-76.

Vecchi, G. A., A. Clement and B. J. Soden, 2008: Examining the tropical Pacific's response to global warming. *EOS,Trans. Amer. Geophys. Union*, **89(9)**, pp. 81, 83.

von Storch, H., and F. W. Zwiers, 1999: Statistical Analysis in Climate research, Cambridge University Press, Cambridge.,494 pp.

Wang, B., 1995: Interdecadal changes in El Niño onset in the lastfour decades. *J. Climate*, **8**, 267–285.

William J. Emery and Richard E. Thomson, 2001: Data Analysis Methods in Physical Oceanography. Elsevier Science B.V. 650 pp.

Wittenberg, A. T., 2004: Extended wind stress analyses for ENSO. *J. Climate*, **17**, 2526-2540.

Xue, M., D.-H. Wang, J.-D. Gao, K. Brewster, and K. K. Droegemeier, 2003: The Advanced Regional Prediction System (ARPS), storm scale numerical weather prediction and data assimilation. *Meteor. Atmos. Physics*, **82**, 139-170.

Yuval, and W. W. Hsieh 2002: The impact of time-averaging on the detectability of nonlinear empirical relations. *Quart. J. Roy. Met. Soc*. **128**, 1609-1622.

Zebiak S.E.,1990: Diagnostic studies of Pacific surface winds. *J. Climate*, **3**, 1016–1031.

Figure Caption

TABLE 1. Model description. The digit in the table indicates the number of leading modes used for model reconstruction. The percentage in the bracket is the variance explained by the leading modes related to the total variance.

TABLE 2. Portion of variance explained by reconstructed wind against the filtered wind for several regions (the number in bracket is the portion of variance explained by the fitted wind against the filtered wind. Also see context for further information).

FIG. 1. The first EOF pattern of (a) LFB of SLPA, (b) MFB of SLPA, (c) LFB of SSTA, (d) MFB of SSTA, (e) LFB of vector wind and (f) MFB of vector wind. The areas with amplitude >0.02 are shaded in (e) and (f). The variance explained by this mode is shown in bracket.

FIG. 2. The PC1 of (a) LFB component and (b-c) MFB component, for SST (thick solid line), SLP (thin solid line) for the 1875-2005 period and for the wind stress (thin dashed line) for the 1948-2005 period. The PC1s are normalized by their standard deviations of the period 1961-1990.

FIG. 3. The correlation (left) and RMSE (right) of simulated SSTA against the observed SSTA. The simulated SSTA is obtained by the OGCM forced with the raw wind stress (a and b); temporal-spatial filtered wind stress (c and d). The correlation coefficients and RMSE are

calculated based on the period from 1948 to 2002. The areas with the correlation over 0.5 and with RMSE over 0.5℃ are shaded.

FIG.4. The anomaly correlation (left) and RMSE (right) of the reconstructed MFB zonal wind against the observed counterpart. The reconstructed wind was produced by model 1 (a and b); model 2 (c and d); model 3 (e and f) and model 4 (g and h). The correlation and RMSE are evaluated using cross-validation scheme from 1948-2002. The contour interval is 0.2 for correlation and 2 m^2 / s^2 for RMSE. The areas with correlation over 0.8 and RMSE over 5 m^2 / s^2 are shaded.

FIG. 5. The anomaly correlation (left) and RMSE (right) of the reconstructed zonal wind (LFB+MFB) against the observed counterpart. The reconstructed wind was produced by model 1 (a and b); model 2 (c and d); model 3 (e and f) and model 4 (g and h). The correlation and RMSE are evaluated using cross-validation scheme from 1948-2002. The contour interval is 0.2 for correlation and $2m^2/s^2$ for RMSE. The areas with correlation over 0.8 and RMSE over $5m^2/s^2$ in are shaded.

FIG. 6. PC1s of the reconstructed wind stress by the four models for (a) LFB for the 1875-1947; (b) MFB for 1875-1910 and (c) MFB for 1911-1947.

FIG. 7. The observed and simulated Niño3 SSTA indices by the OGCM forced with the past wind reconstructed by (a) model 1 and model 2 for 1881-1914, (b) model 1 and model 2 for

1915-1947, (c) model 3 and model 4 for 1881-1914 and (d) model 3 and model 4 for 1915-1947.

FIG. 8. The correlation (left) and RMSE (right) of simulated SSTA against observed SSTA when the OGCM is forced with the wind stress reconstructed by different models: (a, b) model 1; (c, d) model 2; (e, f) model 3 and (g, h) model 4. The correlation coefficients and RMSE are calculated based on the period 1881-1947. The correlation contour interval is 0.2 and the RMSE contour interval is 0.2ºC. The areas with correlation over 0.28 (0.01 significance level) are shaded with grey color and the areas with correlation over 0.6 are shaded with black color. The areas with RMSE over 0.5℃ are shaded with grey color.

FIG. 9. Wavelet power spectrum for (a) Niño4 index of the zonal wind and (b) filtered Niño3 SSTA index. The contour levels are chosen so that 75%, 50%, 25%, and 5% of the wavelet power is above each level, respectively. The cross-hatched region is the cone of influence, where zero padding has reduced the variance. Black contour is the 10% significance level, using a red-noise (autoregressive lag1) background spectrum. The right panels are the global wavelet power spectrum (black line and the significance for the global wavelet spectrum, assuming the same significance level and background spectrum (dashed line).

FIG. 10. The normalized PC1 for the period 1961-1999 of (a) zonal and (b) meridional low frequency (LFB+MFB) wind stress, for NCEP (thick solid line), ECMWF (thin solid line) and FSU wind stresses (thin dashed line).

FIG. 11. As in FIG. 10 but for MFB winds.

FIG. 12. Comparison of normalized PC1s of SSTA for the period 1870-2005 for (a) LFB and (b) MFB component

TABLE 1. Model description. The digit in the table indicates the number of leading modes used for model reconstruction. The percentage in the bracket is the variance explained by the leading modes related to the total variance.

model	Model type	LFB		MFB	
		predictands	predictors	Predictands	predictors
Model 1	PC regression	Wind 2(99%)	SST 4(93%)	Wind 5(45%)	SST 5(73%)
Model 2	PC regression	Wind 2(99%)	SLP 4(95%)	Wind 5(45%)	SLP 5(67%)
Model 3	SVD	Wind	SST	Wind	SST
		2(99%)		5(99%)	
Model 4	SVD	Wind	SLP	Wind	SLP
		2(99%)		5(99%)	

TABLE 2. Portion of variance explained by reconstructed wind against the filtered wind for several regions (the number in bracket is the portion of variance explained by the fitted wind against the filtered wind. Also see context for further information).

Model	V0	V1	V2	V3	V4
Model 1	0.51 (0.55)	0.61 (0.66)	0.49 (0.54)	0.57 (0.61)	0.77 (0.79)
Model 2	0.65 (0.70)	0.64 (0.71)	0.65 (0.70)	0.66 (0.71)	0.73 (0.76)
Model 3	0.59 (0.62)	0.67 (0.70)	0.58 (0.60)	0.63 (0.65)	0.79 (0.80)
Model 4	0.76 (0.78)	0.77 (0.79)	0.76 (0.78)	0.76 (0.78)	0.82 (0.83)

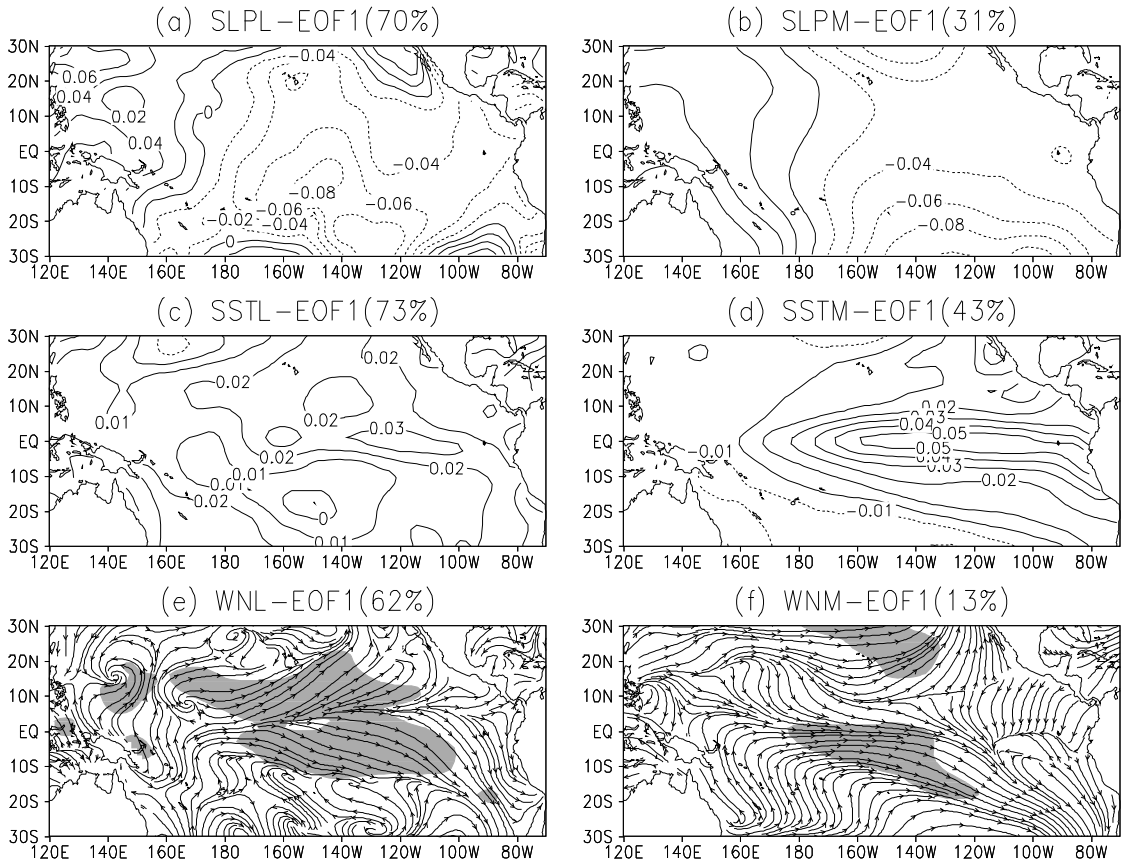


FIG. 1. The first EOF pattern of (a) LFB of SLPA, (b) MFB of SLPA, (c) LFB of SSTA, (d) MFB of SSTA, (e) LFB of vector wind and (f) MFB of vector wind. The areas with amplitude >0.02 are shaded in (e) and (f). The variance explained by this mode is shown in bracket.

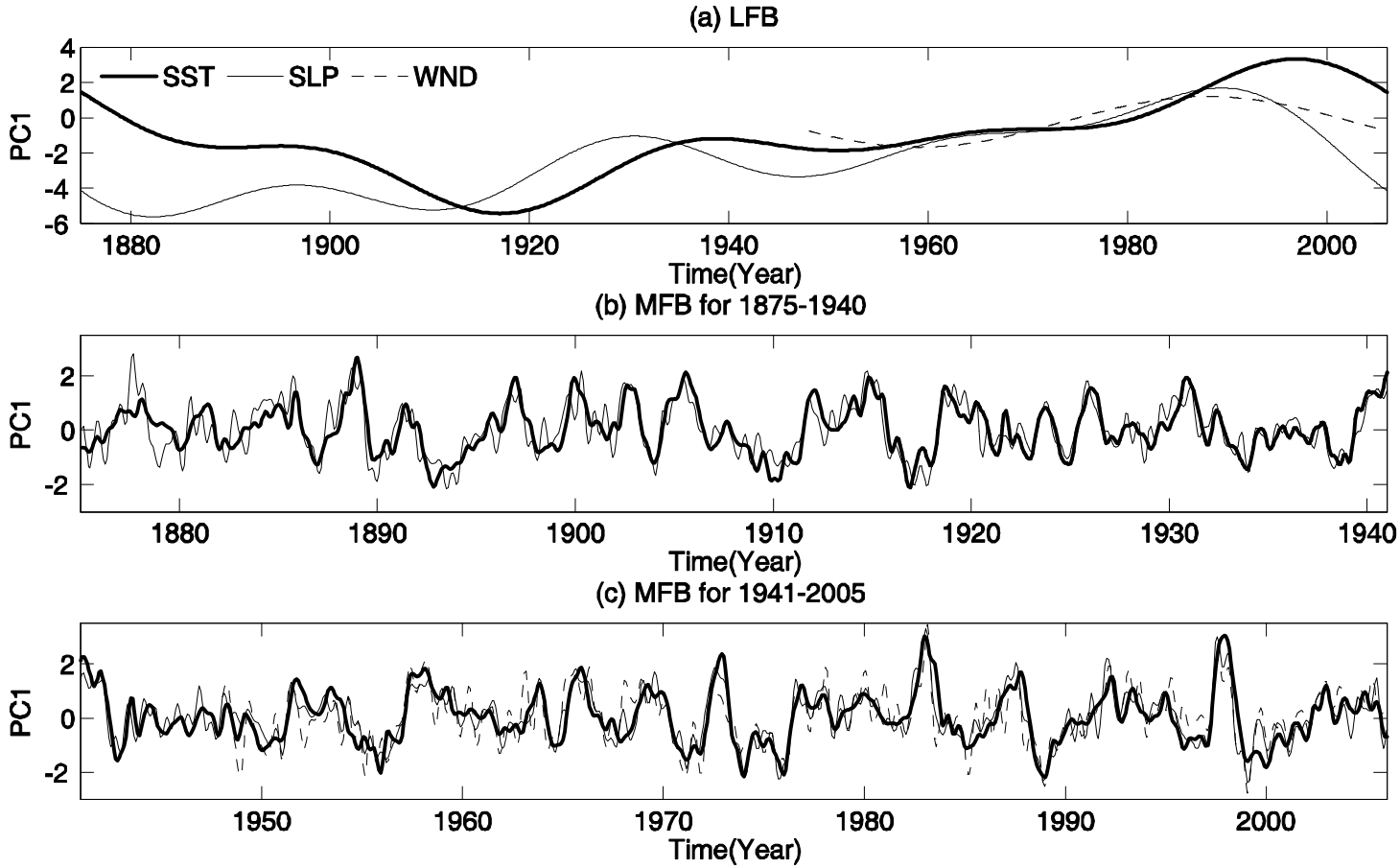


FIG. 2. The PC1 of (a) LFB component and (b-c) MFB component, for SST (thick solid line), SLP (thin solid line) for the 1875-2005 period and for the wind stress (thin dashed line) for the 1948-2005 period. The PC1s are normalized by their standard deviations of the period 1961-1990

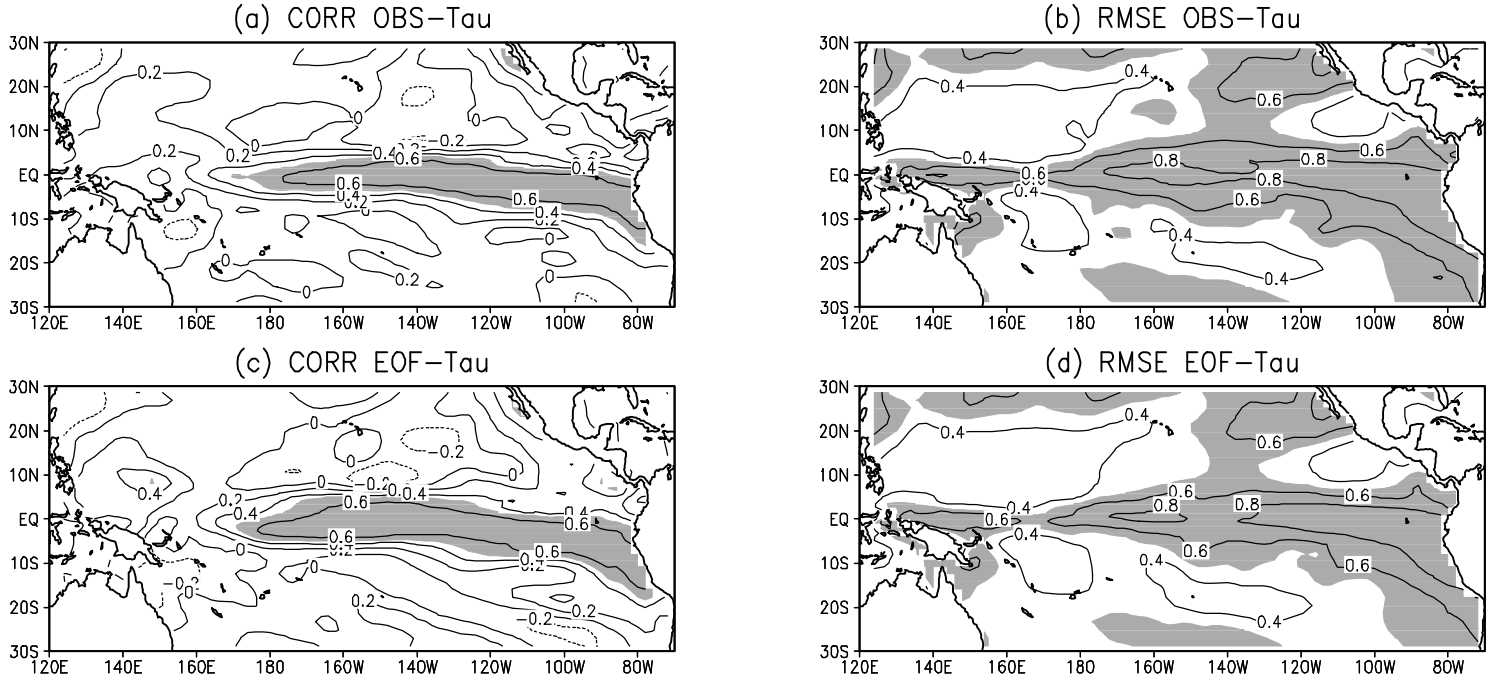


FIG. 3. The correlation (left) and RMSE (right) of simulated SSTA against the observed SSTA. The simulated SSTA is obtained by the OGCM forced with the raw wind stress (a and b); temporal-spatial filtered wind stress (c and d). The correlation coefficients and RMSE are calculated based on the period from 1948 to 2002. The areas with the correlation over 0.5 and with RMSE over 0.5℃ are shaded.

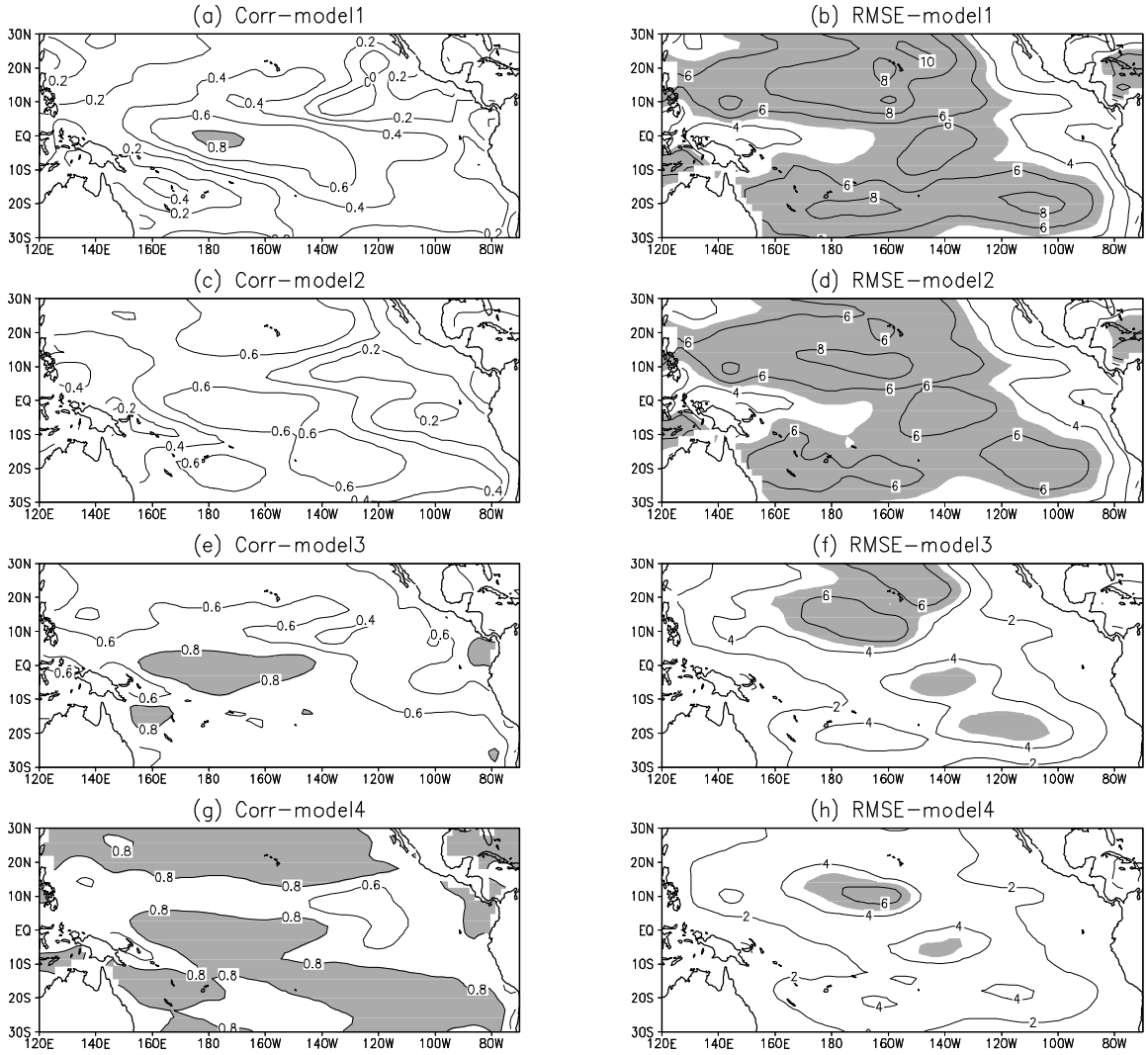


FIG.4. The anomaly correlation (left) and RMSE (right) of the reconstructed MFB zonal wind against the observed counterpart. The reconstructed wind was produced by model 1 (a and b); model 2 (c and d); model 3 (e and f) and model 4 (g and h). The correlation and RMSE are evaluated using cross-validation scheme from 1948-2002. The contour interval is 0.2 for correlation and $2m^2/s^2$ for RMSE. The areas with correlation over 0.8 and RMSE over $5m^2/s^2$ are shaded.

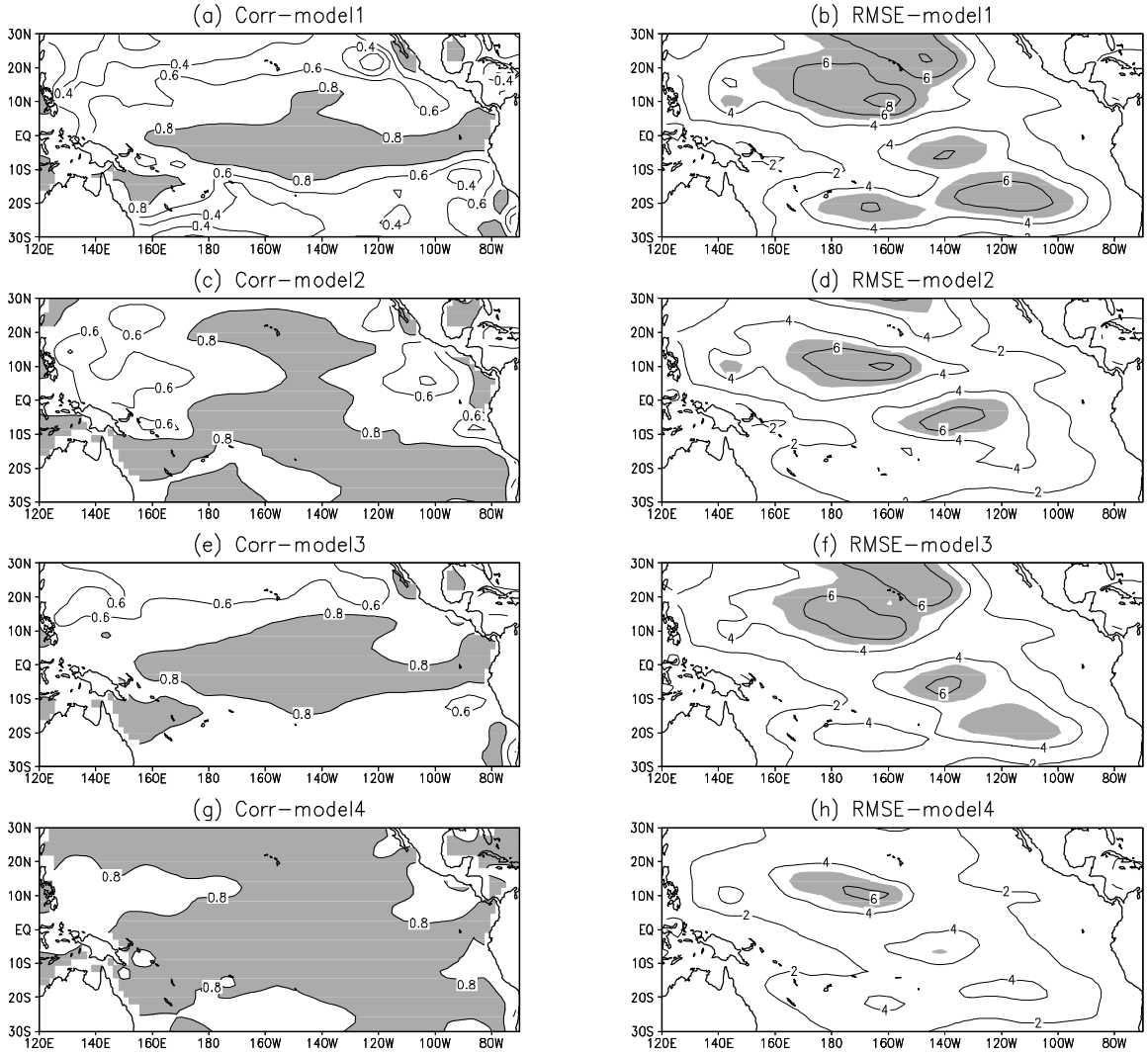


FIG. 5. The anomaly correlation (left) and RMSE (right) of the reconstructed zonal wind (LFB+MFB) against the observed counterpart. The reconstructed wind was produced by model 1 (a and b); model 2 (c and d); model 3 (e and f) and model 4 (g and h). The correlation and RMSE are evaluated using cross-validation scheme from 1948-2002. The contour interval is 0.2 for correlation and $2m^2/s^2$ for RMSE. The areas with correlation over 0.8 and RMSE over $5m^2/s^2$ in are shaded.

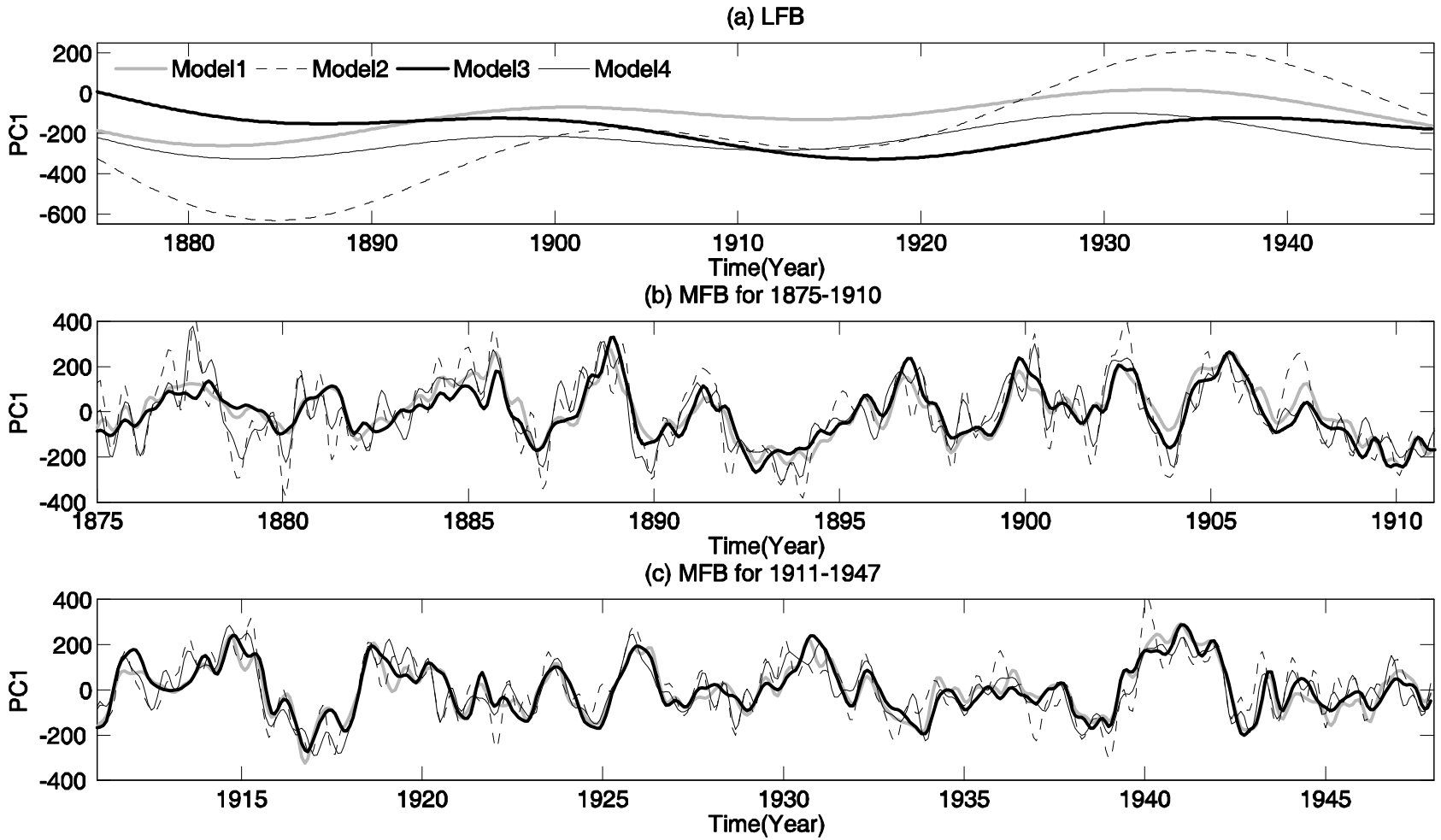


FIG. 6. PC1s of the reconstructed wind stress by the four models for (a) LFB for the 1875-1947; (b) MFB for 1875-1910 and (c) MFB for 1911-1947.

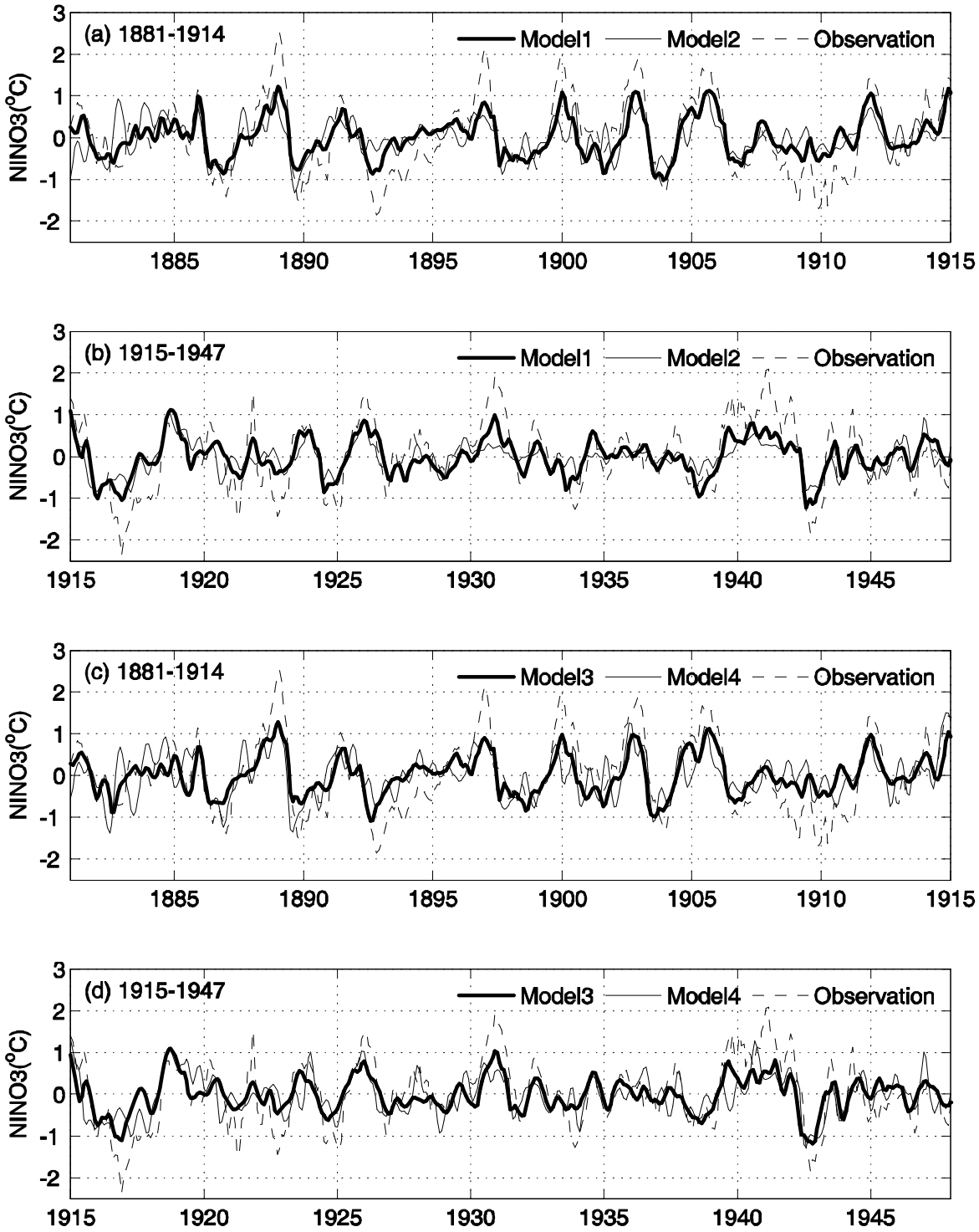


FIG. 7. The observed and simulated Niño3 SSTA indices by the OGCM forced with the past wind reconstructed by (a) model 1 and model 2 for 1881-1914, (b) model 1 and model 2 for 1915-1947, (c) model 3 and model 4 for 1881-1914 and (d) model 3 and model 4 for 1915-1947.

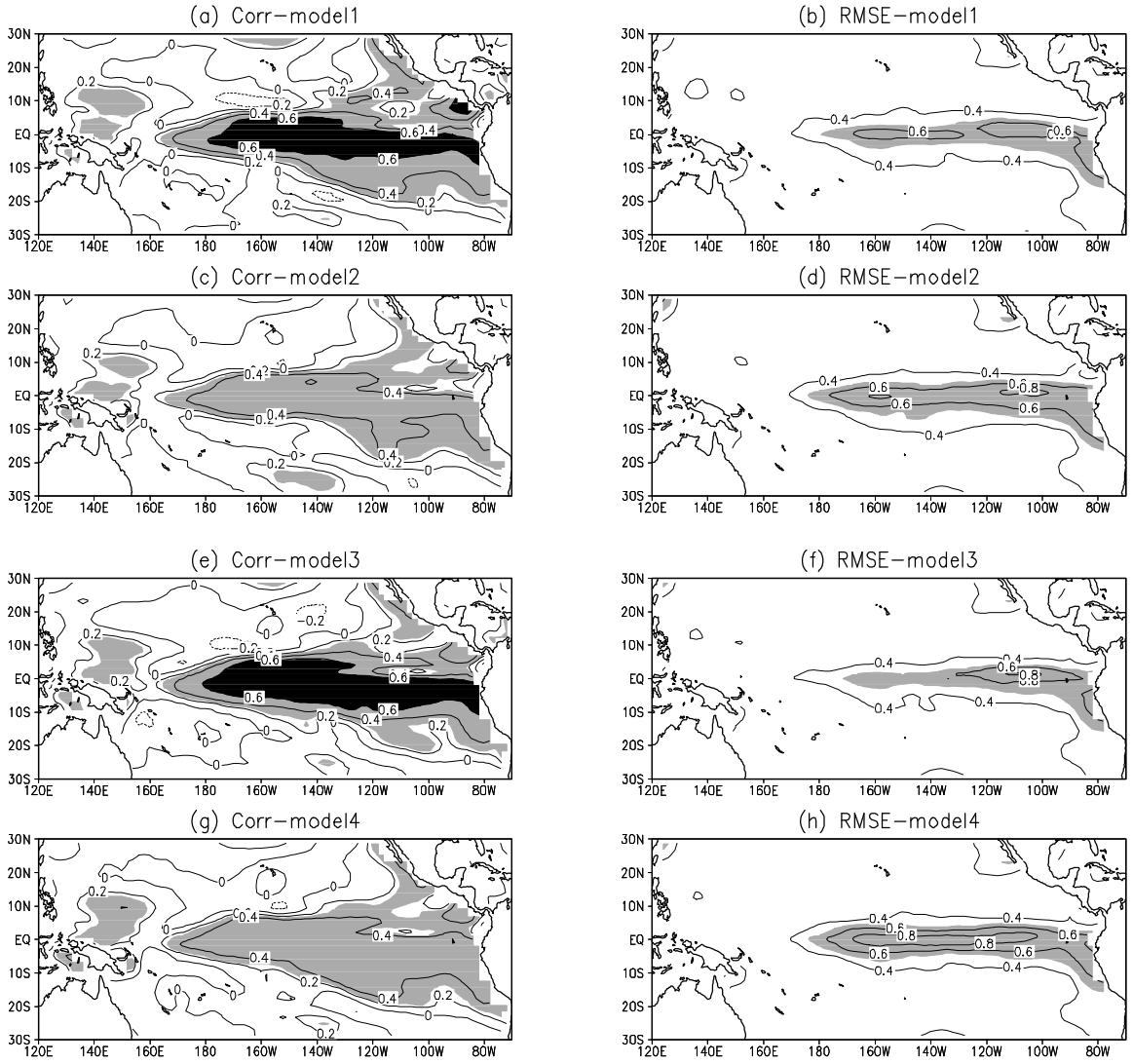
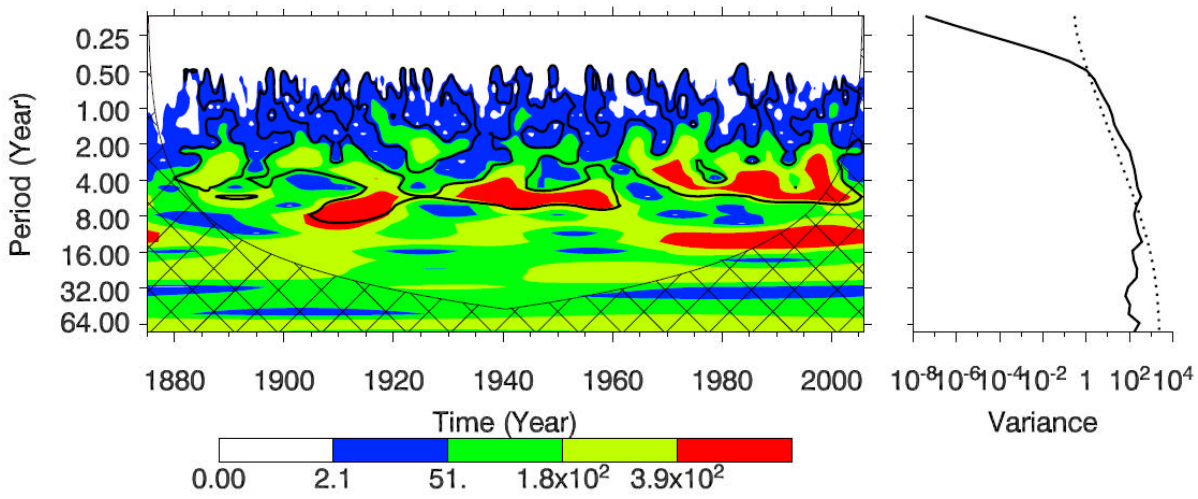


FIG. 8. The correlation (left) and RMSE (right) of simulated SSTA against observed SSTA when the OGCM is forced with the wind stress reconstructed by different models: (a, b) model 1; (c, d) model 2; (e, f) model 3 and (g, h) model 4. The correlation coefficients and RMSE are calculated based on the period 1881-1947. The correlation contour interval is 0.2 and the RMSE contour interval is 0.2ºC. The areas with correlation over 0.28 (0.01 significance level) are shaded with grey color and the areas with correlation over 0.6 are shaded with black color. The areas with RMSE over 0.5℃ are shaded with grey color.

(a) Niño4 wind anomaly index



(b) Niño3 SSTA index

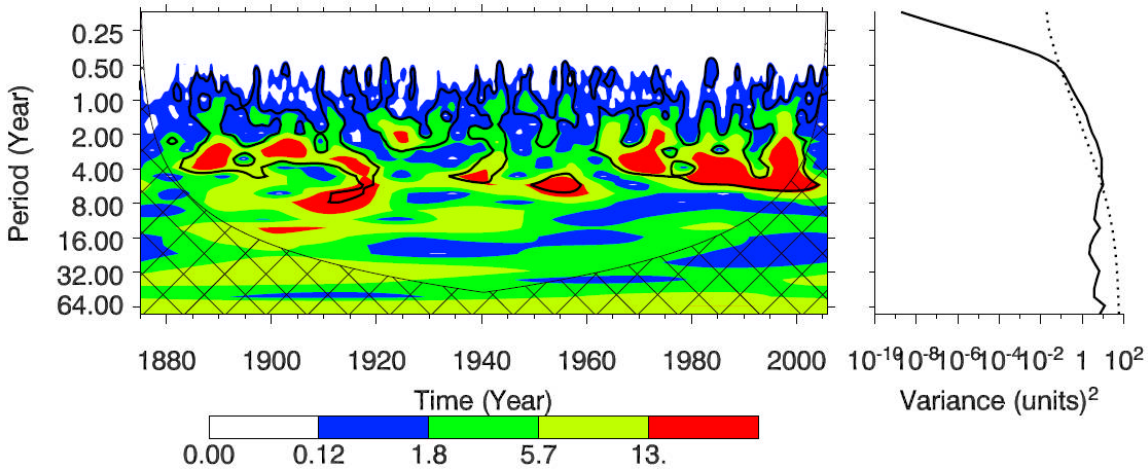


FIG. 9. Wavelet power spectrum for (a) Niño4 index of the zonal wind and (b) filtered Niño3 SSTA index. The contour levels are chosen so that 75%, 50%, 25%, and 5% of the wavelet power is above each level, respectively. The cross-hatched region is the cone of influence, where zero padding has reduced the variance. Black contour is the 10% significance level, using a red-noise (autoregressive lag1) background spectrum. The right panels are the global wavelet power spectrum (black line and the significance for the global wavelet spectrum, assuming the same significance level and background spectrum (dashed line).

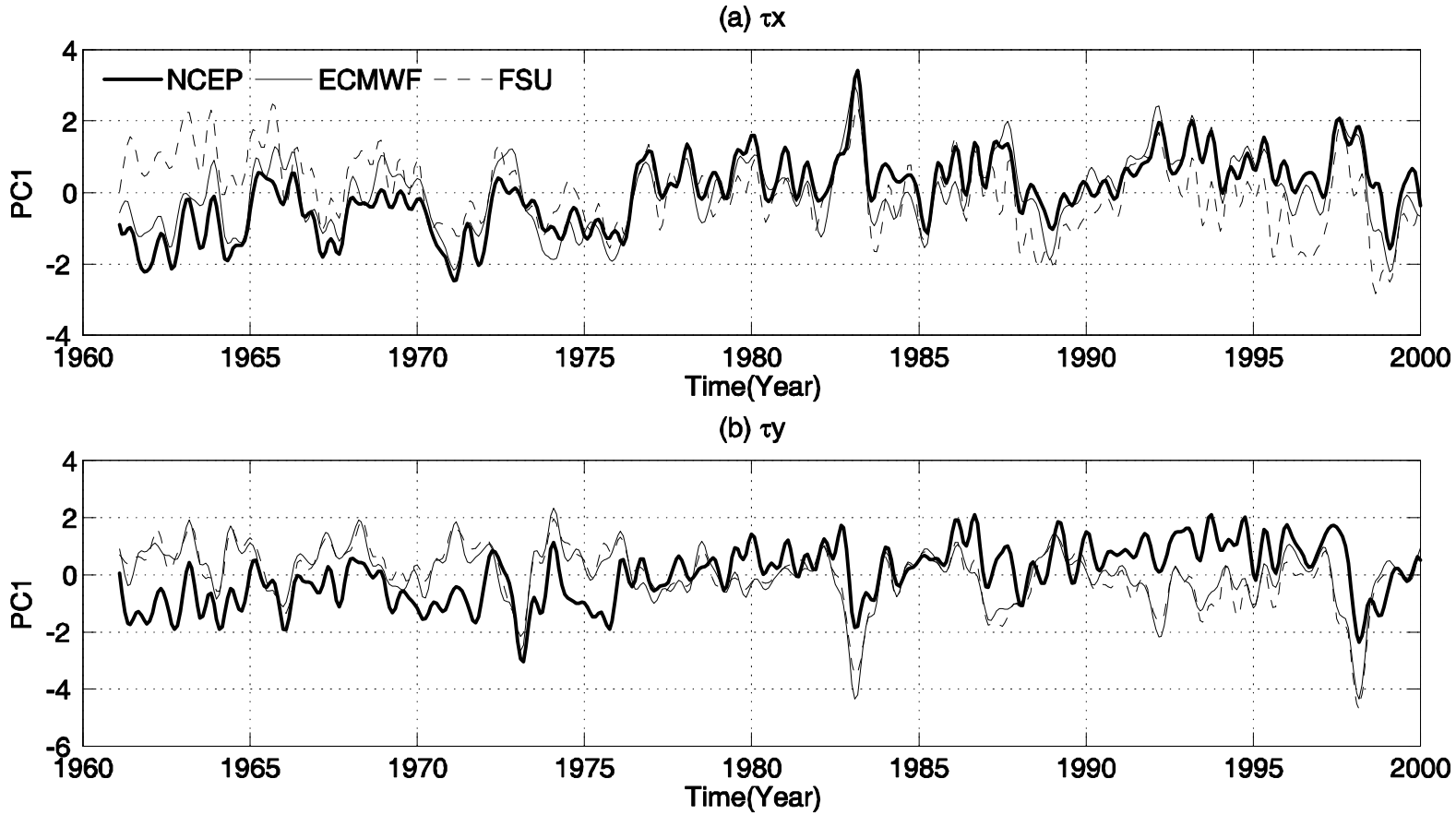


FIG. 10. The normalized PC1 for the period 1961-1999 of (a) zonal and (b) meridional low frequency (LFB+MFB) wind stress, for NCEP (thick solid line), ECMWF (thin solid line) and FSU wind stresses (thin dashed line).

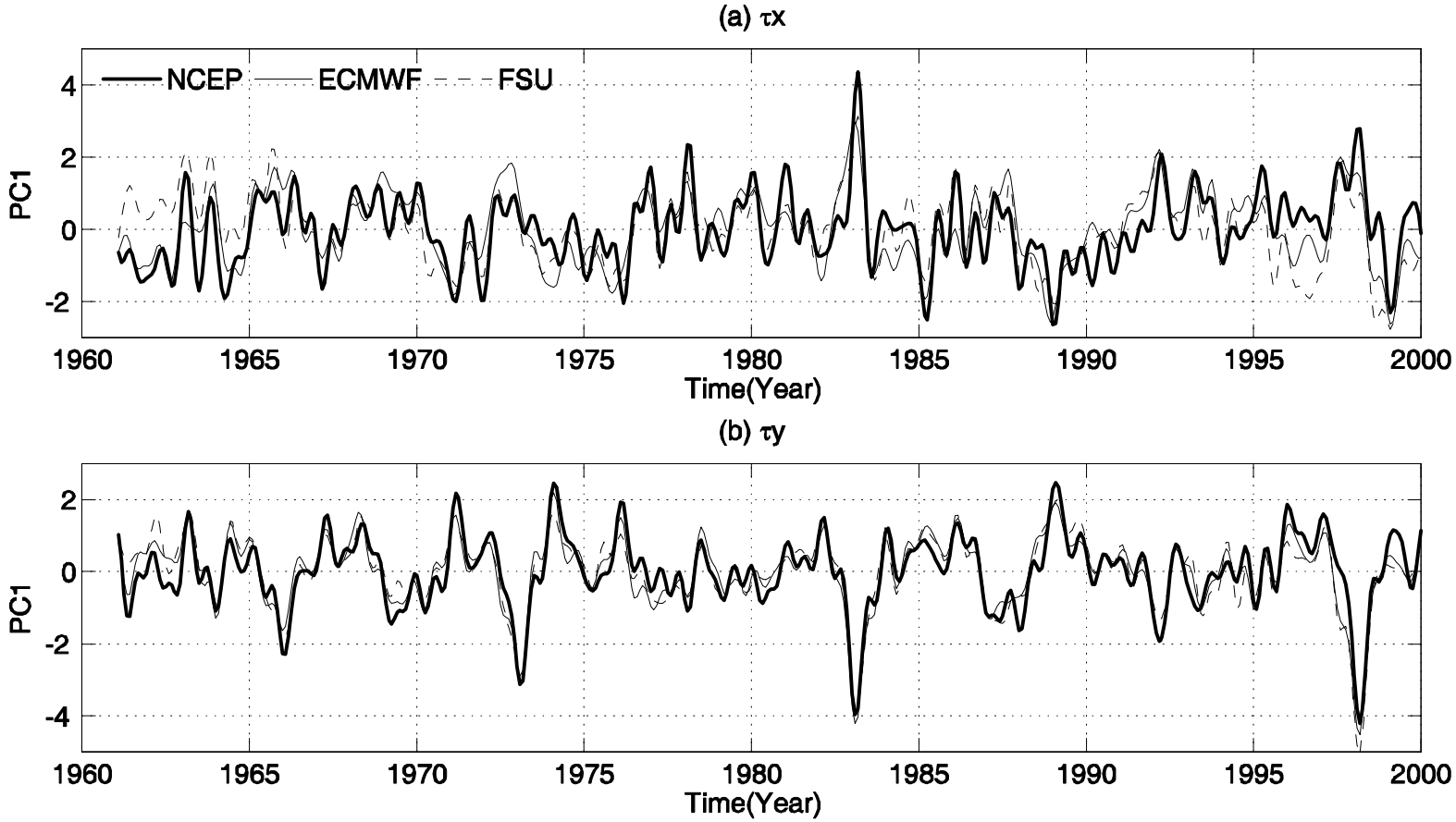
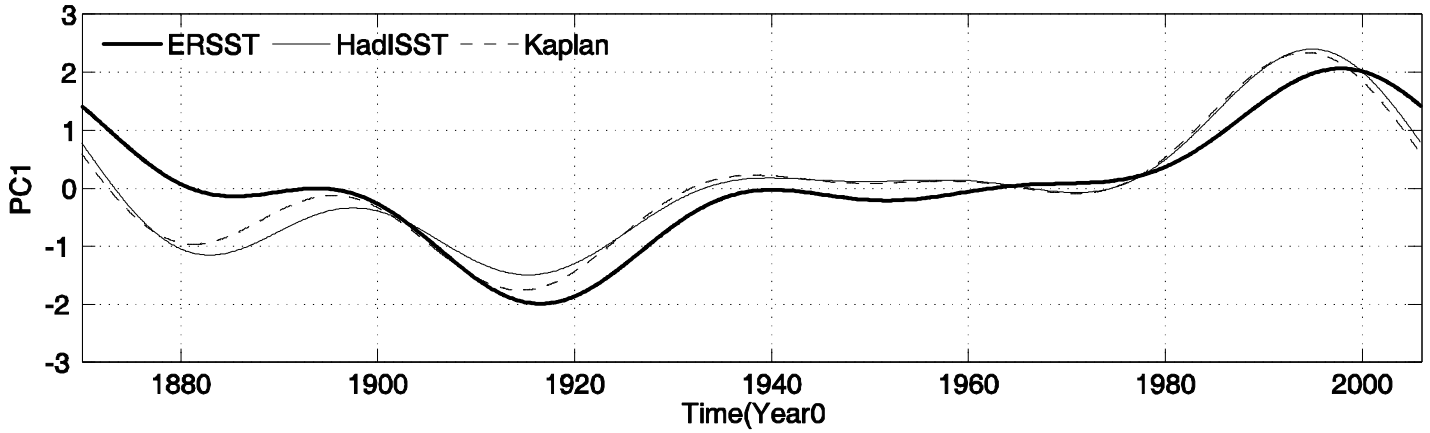


FIG. 11. As in FIG. 10 but for MFB winds.

(a) LFB



(b) MFB

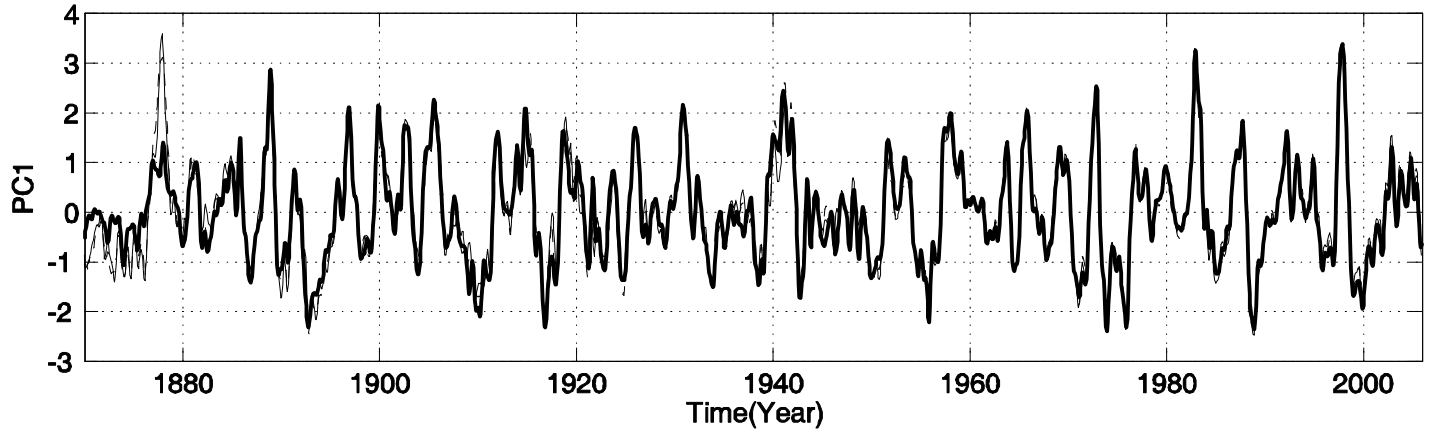


FIG. 12. Comparison of normalized PC1s of SSTA for the period 1870-2005 for (a) LFB and (b) MFB component