Predictability of SST-Modulated Westerly Wind Bursts

GEOFFREY GEBBIE

Harvard University, and Massachusetts Institute of Technology, Cambridge, Massachusetts

ELI TZIPERMAN

Department of Earth and Planetary Sciences, School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts

(Manuscript received 6 March 2008, in final form 21 January 2009)

ABSTRACT

Westerly wind bursts (WWBs), a significant player in ENSO dynamics, are modeled using an observationally motivated statistical approach that relates the characteristics of WWBs to the large-scale sea surface temperature. Although the WWB wind stress at a given location may be a nonlinear function of SST, the characteristics of WWBs are well described as a linear function of SST. Over 50% of the interannual variance in the WWB likelihood, zonal location, duration, and fetch is explained by changes in SST. The model captures what is seen in a 17-yr record of satellite-derived winds: the eastward migration and increased occurrence of wind bursts as the western Pacific warm pool extends. The WWB model shows significant skill in predicting the interannual variability of the characteristics of WWBs, while the prediction skill of the WWB seasonal cycle is limited by the record length of available data. The novel formulation of the WWB model can be implemented in a stochastic or deterministic mode, where the deterministic mode predicts the ensemble-mean WWB characteristics. Therefore, the WWB model is especially appropriate for ensemble prediction experiments with existing ENSO models that are not capable of simulating realistic WWBs on their own. Should only the slowly varying component of WWBs be important for ENSO prediction, this WWB model allows a shortcut to directly compute the slowly varying ensemble-mean wind field without performing many realizations.

1. Introduction

Westerly wind bursts (WWBs) have occurred in the tropical Pacific during the onset and development of every major El Niño event of the last 25 years (Kerr 1999; McPhaden 2004). Recent observational evidence shows that the timing and characteristics of these seemingly random wind bursts are, in fact, modulated by the large-scale SST and, in particular, by the phase of ENSO (Yu et al. 2003; Batstone and Hendon 2005; Tziperman and Yu 2007, hereafter TY). Recent modeling studies have suggested that the proper simulation of the feedback between sea surface temperature and WWBs significantly affects the characteristics and dynamical regime of the ENSO system (Eisenman et al. 2005; Perez et al. 2005; Gebbie et al. 2007; Jin et al.

DOI: 10.1175/2009JCLI2516.1

2007). For example, the amplitude of the simulated ENSO in intermediate complexity models (Eisenman et al. 2005) and hybrid coupled general circulation models (Gebbie et al. 2007) is twice as large with SST-modulated WWBs than with WWBs that are a white-noise process.

Given the critical role of modulated WWBs in ENSO dynamics, their proper representation in models is important for ENSO prediction. At present, some ENSO prediction models represent WWBs as a white-noise external forcing, excluding the modulating influence of the ocean (see the reviews by Mason and Mimmack 2002; Fedorov et al. 2003). Some atmospheric general circulation models forced with the observed SST are capable of producing wind bursts (e.g., Vecchi et al. 2006), but the situation is worse for coupled GCMs. Fully coupled models still contain major biases in their mean state (Wittenberg et al. 2006), and thus the characteristics of wind bursts are likely to be similarly biased. ENSO prediction models that use a linear statistical

Corresponding author address: Geoffrey Gebbie, Harvard University, 24 Oxford St., Cambridge, MA 02138. E-mail: gebbie@eps.harvard.edu

atmosphere also misrepresent WWBs because less than half of the WWB wind stress can be explained by a linear relationship with SST (Gebbie et al. 2007). The other half of the wind residuals was shown by Wittenberg (2002) to fail a white noise test. Furthermore, the results of Batstone and Hendon (2005) suggest that these residuals are also correlated with SST.

To what degree are wind bursts a deterministic function of the sea surface temperature field? Because WWBs are associated with many different processes, including extratropical cold surges, tropical cyclones, and the Madden-Julian oscillation, it is sometimes assumed that they are essentially part of the "weather noise," with the implied rapid loss of predictability. Previous works (e.g., Yu et al. 2003; Batstone and Hendon 2005) have shown, however, that, while the timing of individual bursts is certainly difficult to predict, the statistics of wind bursts may be more predictable. Specifically, the modulation of WWBs by the SST seems to induce the bursts to occur when the warm pool is extended, while their stochasticity may result in a relatively small range of uncertainty of their precise time of occurrence (possibly a few weeks). One may hope, therefore, that the longer-term ocean memory should make it possible to predict an ensemble of WWB realizations based on their modulation by the SST and that this may result in improved ENSO prediction skill.

The relationship between wind bursts and SST has been previously analyzed by using ocean composites over multiple WWBs (Harrison and Vecchi 1997) and by correlation studies (Batstone and Hendon 2005; TY). In particular, TY found that, if instead of correlating the WWB wind stress field with SST one correlates the WWB characteristics including their amplitude, width, duration, location, and probability of occurrence, then more of the WWB variability can be linked to SST. A correlation, of course, does not imply causality between two events, and the correlations of TY cannot serve as a prognostic model for WWBs given SST. The goal of this paper is to build on these observed correlations and to present a quantitative, observationally based approach for predicting WWBs and representing them in ENSO models in a way that accounts for the WWB modulation by the SST. This goal differs from TY in that causality must be taken into account, as we wish to find the WWB response to a given SST field. For this reason, this study uses statistical tools to determine the robustness and significance of the postulated cause-effect relationship, something that has not been attempted before.

The WWB model developed here is used to demonstrate that there is a strong deterministic component in WWBs and a strong connection between the western Pacific warm pool extent and WWB characteristics. This approach may be considered an extension of statistical atmospheric models to include WWBs. The relevance of the resulting WWB model to the modeling and predictability of the tropical Pacific interannual variability is discussed. Many studies (e.g., Chang et al. 1995) model the atmospheric response as a combination of deterministic and white noise processes. However, these approaches are not meant to explicitly represent WWBs as a state-dependent noise process as attempted here. In addition to the relevance to ENSO prediction, our WWB model may be viewed as an efficient way to represent the observational evidence of SST modulation on WWBs. It may then be used to evaluate the ability of complicated coupled models to reproduce the same behavior. We note that the implications of the WWB prediction model to ENSO prediction were examined by Gebbie and Tziperman (2008), using a hybrid coupled model, with encouraging, even if tentative, results.

To achieve our stated goal, we first introduce the WWB dataset and analyze the statistics of occurrence times of WWBs to determine whether one can reject the hypothesis that WWBs are a random system where the probability of WWB occurrence is constant in time (section 2). The observationally based empirical WWB model is presented in section 3. The WWB model predicts the WWB characteristics from the SST, and leads to new insights regarding the interannual variability of WWBs (section 4) and the seasonal cycle (section 5). We conclude in section 6.

2. Statistics of observed WWB timing

To identify WWBs, TY used a 1° gridded surface wind analysis with 6-h time resolution based on an analysis combining the European Centre for Medium-Range Weather Forecasts 10-m surface wind analyses, Special Sensor Microwave Imager (SSM/I) wind speeds, and ship and buoy winds. A westerly wind burst was defined as any zonal wind anomaly greater than 5 m s⁻¹ with a duration between 2 and 40 days and with a longitudinal extent greater than 500 km. The anomaly is defined relative to the seasonal climatology of the trade winds, which are as strong as 4 m s^{-1} , and thus a WWB signifies a reversal of the total wind direction. This definition is necessarily arbitrary, as are all previous WWB definitions. It is important, however, that this definition does not preclude that WWBs have some low-frequency interannual power, in addition to the high frequency (2–40 days) power inherent in the length of an individual event. The existence of interannual power is thought to be critical to the ENSO response to WWBs (Roulston and Neelin 2000; Eisenman et al. 2005; Gebbie et al. 2007).

Previous studies (Luther et al. 1983; Fasullo and Webster 2000; Yu et al. 2003; Eisenman et al. 2005) have suggested that the zonal wind speed of a westerly wind burst can be approximately represented by a Gaussian in space and time. We use this idealization, yet note that this assumption is far from being perfect and has not been justified by a sufficiently rigorous analysis of observed WWBs. In particular, we model a WWB as

$$u_{\text{wwb}}(x, y, t) = A \exp\left(-\frac{(t - T_o)^2}{T^2} - \frac{(x - x_o)^2}{L_x^2} - \frac{(y - y_o)^2}{L_y^2}\right),$$
(1)

where x_o and y_o are the central longitude and latitude of the wind event, T_o is the time of peak wind, A is the peak wind speed, T is a measure of event duration, and L_x and L_y are the spatial scales. Using the TY catalog of WWB events from the SSM/I satellite-derived winds for 1988–2004, the number of events identified is $N_{WWB} =$ 127, at an average of 7.5 WWBs per year.

Of particular importance is whether the temporal distribution of WWBs is consistent with a Bernoulli process, a discrete-time stochastic process that has two possible outcomes: failure or success. Success, as we define it, indicates that a WWB is triggered. We therefore first consider the null hypothesis that a WWB event is equally likely at every time, $P(t) = P_o$, where P_o is the probability of a WWB occurring on any given day (day^{-1}) . If the occurrence times of individual WWB events are independent of each other, the expected number of WWBs over an interval of N days follows a binomial distribution with a mean $\langle N_{\text{wwb}} \rangle = NP_o$ and a variance of $\langle N_{\text{wwb}}^2 \rangle = NP_o(1 - P_o)$. The observation of 127 WWBs over 17 yr (i.e., $N_{\text{wwb}} = 127$) allows us to determine a plausible value of P_o , namely, 1/(58.8 days) < $P_o < 1/(40.8 \text{ days})$ with 95% confidence. The most likely value of P_o is simply the number of WWBs divided by the number of days, $P_o = N_{\text{wwb}}/N = 1/(48.9 \text{ days})$, shown as a straight line in Fig. 1. There are prolonged periods, such as 1988-90, with relatively few WWBs, and other periods, such as 2002-05, with more frequent WWBs. The long periods without WWBs call into question whether a constant P_o explains the data.

To judge more quantitatively whether the observed distribution could possibly be from the expected binomial distribution, we analyze the separation times between WWBs. Here, $T_r(n)$ is defined as the elapsed time between WWB number "*n*" and "*n* + 1" (days). The expected distribution of T_r between any two WWBs is denoted $\Phi_{Tr}(N)$ and is given by

$$\Phi_{Tr}(N) = (1 - P_0)^{N-1} P_0, \qquad (2)$$



FIG. 1. The timing of observed WWBs, 1988–2004. The cumulative sum of the number of WWBs (solid) is compared to the mean of a binomial distribution with $P = P_0$ and N, the number of trials, equal to the number of days since 1 Jan 1988. Long runs with relatively few WWBs (1998–2001) and many WWBs (1990–95) are both apparent.

where N is an integer number of days. The observed distribution of T_r is taken by differencing the T_o values in the WWB catalog. A higher than expected number of WWBs occur within 15 to 50 days of each other (Fig. 2). Furthermore, there are gaps of 400 and 600 days without any WWBs at all, such as the period from January 1998 to May 2001. Based on the Kolmogorov–Smirnov test, a nonparametric statistical tool that does not rely on any assumption of Gaussian statistics, the null hypothesis is rejected at the 5% insignificance level with a P value of 0.02. The observed WWB distribution is therefore unlikely to be from a stochastic process in which the probability of occurrence is constant in time.

If P is defined instead as a continuous function of time, rather than a daily value, a Poisson distribution would describe the timing of WWBs. Given the large number of days in the 17 years of data, there is no appreciable difference between the binomial distribution and Poisson distribution. To confirm this, we repeated the null hypothesis test with an assumed Poisson distribution with $P = P_o$ and found the same results: that the null hypothesis can be rejected.

The observed histogram can possibly be explained by the likelihood of a WWB changing in time; that is, P = P(t). When P(t) is large, the time between WWBs is shorter. In the case smaller values of P(t) persist over multiple years, long breaks without WWBs are possible. We hypothesize that low frequency variations in SST modulate P(t) and explain the distribution of observed WWBs through time (in the nomenclature of statistics, a time-inhomogeneous Bernoulli process). The WWB



FIG. 2. The recurrence time T_r between WWB events. The observed distribution of T_r (gray histogram) is compared to T_r for the null hypothesis H_0 (black line), a binomial distribution with $P = P_0$, and one time step per day.

model of the next section attempts to make this connection explicit.

3. The empirical WWB model

For a prognostic WWB model, the necessary parameters are slightly different from those in the WWB dataset. Instead of the observed start times of individual WWBs, T_o , we need an estimate of P, the likelihood of the triggering of a WWB. TY defined an observational estimate of the likelihood of a WWB as the number of WWBs in a period of 3 months centered on the time of interest, divided by 3 months.

The WWB model state can therefore be represented by the following vector of characteristics:

$$\mathbf{r} = [A; x_o; y_o; L_x; L_y; T; P],$$
(3)

corresponding to the parameters in Eq. (1). Note that the WWB characteristics, apart from the probability of occurrence, are only defined when a WWB occurs. A regression model for the WWB parameters may be developed by creating a list of WWB events with their observed characteristics and the average SST field during the event, as was done in the correlation study of TY. When this is attempted, the model strongly underestimates the variance of WWB parameters. This result is not surprising since it should be necessary to train the model with both the SST fields that lead to WWBs and those that do not, in order to reproduce the full range of behavior in the model. Instead, we proceed with a second method of pairing the WWB and SST data that produces better results. This second method uses an estimate of WWB characteristics and SST for every month during the data record. We use a linear interpolation from the nearest WWBs in time to evaluate the WWB parameters at non-WWB times. The only exception is that the probability parameter is set to zero if no WWBs are observed, rather than being interpolated. When more than one WWB occurred in a month, we take the average WWB characteristics for that month.

The observed WWB dataset for all months may now be written as a $N_{\text{mon}} \times 7$ matrix,

$$\mathbf{R} = [\mathbf{r}_1; \mathbf{r}_2; \dots; \mathbf{r}_{N \text{mon}}], \qquad (4)$$

where $N_{\text{mon}} = 204$ is the total number of months in 17 years. We wish to predict the WWB parameter anomalies, so the *R* dataset is referenced to an appropriate time average. We explore two approaches to defining the anomalies. In section 4, we reference the anomalies to the seasonally varying climatology and model the interannual WWB variability. In section 5, we reference the anomalies to a mean that is not a function of month and attempt to model both interannual and seasonal variability.

The monthly SST field of the tropical Pacific (25°N– 25°S, 120°E–80°W) is arranged in a similar way such that **S** is the SST data matrix that contains N_{xy} rows and N_{mon} columns. The SST dataset has a 2° resolution in longitude and latitude and a total of N_{xy} = 1890 ocean grid points. Again, the appropriate time-averaged monthly climatology is subtracted from the data. A complete WWB model is described by a set of simultaneous equations of the form

$$\mathbf{R} = \mathbf{SW} + \mathbf{E},\tag{5}$$

where **W** is the desired WWB model matrix relating the WWB characteristics to the SST, and **E** is the misfit. We solve for the matrix **W**, which minimizes the sum of squared elements (i.e., the Frobenius norm) of \mathbf{ED}_{R}^{-1} , where \mathbf{D}_{R}^{-1} is a square diagonal matrix with the inverse of the standard deviation of each WWB parameter along the diagonal. Column weighting with \mathbf{D}_{R} accounts for the differing units and magnitudes of the WWB parameters. In other words, the model is designed with equal weight given to the variability in each parameter. Such a weighting scheme does not necessarily result in a WWB model that leads to optimal ENSO predictions, and future work would need to test the sensitivity of ENSO prediction using this WWB model to the weighting strategy.

The **W** matrix contains $N_{xy} \times 7$ elements, to be calculated from the $N_{wwb} \times 7$ available data points (one sevenelement vector for each of the 127 observed WWBs).



FIG. 3. The first two SST (top) singular vectors and (middle) EOFs, with the projection of the 1988–2005 observations onto those patterns, that is, (bottom) the principal components. Positive values are shaded. Both the projection onto the singular vectors and EOFs are plotted in the bottom panels, but they are nearly identical and cannot be distinguished.

Given that $N_{xy} \gg N_{wwb}$, the 17-yr observational record of WWBs is not long enough to robustly estimate **W** by the above regression of each WWB parameter onto the SST at all geographic locations. Instead, we express the SST as a sum over large-scale modes, to be defined shortly, and regress the WWB parameters only with these large-scale patterns. (Base function might be a more appropriate description than "mode.")

We explore two alternatives for the SST modes. In the first, the modes are the left singular vectors of the covariance matrix between the SST and the westerly wind burst parameters, $\mathbf{C} = \mathbf{S}^{\mathrm{T}} \mathbf{R}$ (the matrix multiplication is effectively an average over all months in the dataset). These left singular vectors, also called singular value decomposition (SVD) modes, are the modes of SST variability that explain the most covariance with the wind. Another possibility for the reduced SST space is to use the leading EOFs of the SST field. We find that the first two leading EOFs are remarkably similar to the first two singular vectors (Fig. 3). In the calculations below we use the first seven left singular vectors as the reduced space representation for the SST. Given that a few such vectors explain most of the covariance between the SST and WWB parameters, the leading results of the WWB model are not significantly altered using the EOFs instead. Note that the first SST singular vector resembles the El Niño pattern, demonstrating

that the characteristics of WWBs covary with the phase of ENSO.

With a set of N_Q large-scale SST modes, the SST field is represented by N_Q nondimensional, time-varying coefficients. These coefficients are stored in a N_{mon} by N_Q matrix given by $\tilde{\mathbf{S}}_Q^* = \mathbf{SQD}_S^{-2}$, where the rows of matrix **Q** are the dimensional SST modes, **Q** itself is a mapping from the full SST field to the reduced space of SST mode amplitudes, and \mathbf{D}_S^{-2} is a square, diagonal matrix with the inverse of the variance of each expansion coefficient time series along the diagonal. The final equation for the WWB model is now rewritten as

$$\mathbf{R} = \tilde{\mathbf{S}}_O^* \mathbf{W}_O + \mathbf{E},\tag{6}$$

where the desired WWB model matrix \mathbf{W} is now replaced by the much smaller matrix \mathbf{W}_Q of size $N_Q \times 7$. Now there are $N_{\text{wwb}} \times 7$ data points to constrain $N_Q \times 7$ values, and $N_Q < N_{\text{wwb}}$.

This combined regression-singular vectors approach is similar to that used by Harrison et al. (2002) and Wittenberg (2002) to develop their linear statistical atmospheric models. An alternative approach is a purely SVD-based approach (e.g., Syu et al. 1995). Such an SVD-based approach, however, effectively assumes a perfect time correlation between SST and WWB SVD modes, either at zero lag or some other fixed lag. Future





studies may need to look into the possibility of allowing for a nonzero lag between the two. In general, such a scheme does not account for the possible statistical insignificance of a particular SVD mode, and it is unlikely to be the best fit to the observations in a least squares sense (cf. Bretherton et al. 1992).

4. Interannual WWB model

We now form a model for the interannual WWB variability, while prescribing the climatological seasonal cycle of WWBs. This "interannual-only" model is formed by solving (6) after subtracting the seasonally varying climatology from the datasets **R** and **S**.

a. Modes of variability

The interannual-only WWB model uses the SST singular vectors of the covariance matrix as the basis for all calculations (Fig. 4) by first calculating the projection of the observed SST onto the SST singular vectors (Fig. 5), and then regressing these projections onto the WWB parameters. The first SST mode explains 97% of the covariance between SST and the WWB parameters, and its spatial pattern is directly related to El Niño. The time series of the projection onto the second SST mode leads the first mode with a maximum lagged cross-correlation of r = 0.6 at 11 months. SST mode 2 captures the eastward propagation of SST anomalies during recent El Niño events. The third mode calculated here is different from that of TY. This is partially because, as explained above, we use a correlation between WWB parameters and SST that is calculated at all times, not only those times when WWB events occurred. The sensitivity of the structure of the third mode indicates, of course, that this structure is not a robust result of our study even if the third mode is useful for the WWB prediction. We will attempt to interpret the higher SST modes with respect to particular WWB parameters later in this work.

In a regression model, the addition of another SST mode will always lead to a better fit to the observations, but we wish to eliminate modes that are simply fitting noise. To determine the number of significant modes, N_Q , we use an iterative statistical procedure. At each step, a single mode is added to the WWB model and the residual between the observed and modeled wind burst





FIG. 5. The nondimensional SST expansion coefficients for the first seven SST modes corresponding to Fig. 4.

parameters is calculated. The reduction in the sum of the squared errors is considered significant when a two-tailed F test finds the variance of the residual before and after adding the new mode to be different at the 5% insignificance level (see Wittenberg 2002). The WWB model can be based on a maximum of seven modes, corresponding to the seven westerly wind burst parameters.

Figure 6 displays the variance explained for each WWB parameter as additional modes are added. We find that all modes significantly reduce the error (increasing the explained variance) in at least two wind burst parameters, and we therefore keep all seven modes in our model. The ENSO SST pattern (the first SVD mode) explains over 30% of the variance in four wind burst parameters. The zonal location and likelihood of occurrence are explained best (63% and 58% of their variance, respectively, is explained by seven SVD modes). Not all WWB parameters are explained equally well by SST, however. Meridional extent (33%) and amplitude (7%) are least well explained. These results reflect the considerable high-frequency variance that cannot be explained by the slowly varying SST. When considering only the low frequency atmospheric components (using a 6-month running mean), the WWB model explains more than 65% of four WWB parameters (probability, zonal location, zonal extent, and duration) and 25% of the variance in amplitude (lower panel, Fig. 6).

Despite our attempts to determine N_Q by objective statistical means, some subjectivity cannot be eliminated in the choice of N_Q . For example, we determined



FIG. 6. The fraction of explained variance (top) for each WWB parameter and a WWB model with a given number of modes 1–7 and (bottom) at periods greater than six months.

-							
	1	2	3	4	5	6	7
$A (m s^{-1})$	(0.00)	-0.94	(0.19)	(0.33)	-0.60	0.40	-0.76
x_o (° lon)	12.29	(-2.56)	5.16	(-0.31)	(-1.32)	(0.42)	-2.63
L_x (° lon)	4.87	-2.00	(0.55)	2.53	(-0.71)	(0.10)	(-0.62)
L_{v} (° lat)	1.62	(0.34)	(0.21)	1.60	0.99	1.30	(0.21)
$T(\text{days}^{-1})$	2.80	-1.41	-0.52	(0.52)	-0.22	0.84	(0.02)
$P(yr^{-1})$	3.93	2.22	-1.46	0.64	(-0.31)	(-0.98)	(-0.32)
y_o (° lat)	-0.41	(0.22)	-1.16	(0.09)	(0.79)	-0.30	(-0.01)

TABLE 1. The transpose of the interannual WWB model matrix \mathbf{W}_Q . The first seven rows give the seven regression vectors for the WWB parameters. Entries in parentheses do not significantly reduce the model–observations misfit.

the statistical significance of the improvement in each WWB parameter separately, while the significance of the covariance between WWB parameters may be more important to other investigators. Furthermore, the explanation of 7% variance in WWB amplitude is barely significant (P = 0.05) with the *F* test and the available degrees of freedom. Figure 4 underscores the fact that modes 4–7 are probably less significant than the first three modes, as might be deduced by the existence of smaller-scale spatial patterns. We note that the WWB model may be easily rederived with a different number of modes. For the results emphasized later in this work, the exclusion of the higher modes does not appreciably change our conclusions.

b. Key relationships

The coefficients of the resulting regression model (elements of the matrix \mathbf{W}_{O}) are shown in Table 1. Entries in parentheses do not significantly decrease the residuals between the observed WWB characteristics and those predicted by the WWB model. The values in the table can be used to form a linear relationship between SST and WWB characteristics. For the first SST mode amplitude (s_1^*) and the zonal location of WWBs (x'_o) , for example, we have $x'_o = 12.29^{\circ}s_1^*$. This relationship indicates that the WWBs move by 12.29° longitude in response to a one standard deviation (1σ) change in the amplitude of the first SST singular vector (see the best-fit line in Fig. 7). For the 1997 El Niño (a 3σ event), the regression coefficient of first mode of the WWB model predicts that wind bursts move more than 30° to the east. As shown in the figure, the linear function is a good approximation to the observed relationship between SST and the zonal location of WWBs. The misfit between the linear model and observations (bottom-left panel of Fig. 7) cannot be distinguished from a Gaussian distribution (p = 0.82), a posterior check that the linear regression form of the model (based upon minimizing a sum of squares) is reasonable.

The eastward migration of WWBs with the edge of the warm pool during ENSO has been shown to be an important feedback in the evolution of ENSO events (e.g., Chang et al. 1995; Picaut et al. 1997). This feedback was not built into the WWB model by the assumptions used here, and we can therefore check to see if such an effect is captured by our model. Defining the warm pool edge, γ , to be the longitude of the 28.0° SST isotherm, the temporal correlation between movements of the warm pool and the first SST mode is r = 0.85, with a best-fit linear relationship $\gamma = 10.03^{\circ}s_1^* + 159^{\circ}E$. This linear relationship is nearly identical to the aforementioned anomaly equation given by the regression model; thus, our model captures the observed movement of WWBs with the extension of the warm pool. The regression model is linearized around the seasonally varying WWB location, between 155° and 165°E, so there appears to be no significant offset between the



FIG. 7. Relationship of the first SST mode to (left) the zonal location of WWBs and (right) the probability of WWB occurrence. (top) Observations of x_o and P plotted vs the observed projection onto SST mode 1 (dots) and including the model prediction from one SST mode (solid line); (middle) the single-mode model prediction (solid line) reproduced and compared to the model prediction with seven modes (dots); and (bottom) histograms of the misfit between the seven-mode model output and observations.



FIG. 8. Time series of WWB parameters as observed (diamonds), as predicted by the cross-validated WWB model (thick line), and as predicted by the model trained with all 17 years (thin line).

warm pool edge and the center longitude of the WWB, x_o , in the data.

Yu et al. (2003) suggested that WWBs are more likely with an extended warm pool. The WWB model predicts the following relationship for the first mode P' = $3.93[\text{yr}^{-1}] \times s_1^*$, variations of almost four WWBs per year in response to a one standard deviation increase in the amplitude of the first SST singular vector (see upper-right panel of Fig. 7). The average number of WWBs per year is 7.5; thus the variability is large relative to the mean and potentially important to ENSO dynamics. Again, the linear relationship seems to be a good first-order explanation for the majority of variance in P. It is not obvious what physical mechanism leads to an increase of WWBs with an extended warm pool, but the results here suggest that it has to do with the changing large-scale SST and extension of the warm pool, as opposed to an atmospheric-only explanation based, say, purely on MJO dynamics. Note that this discussion refers to the propagation of sequential events, not to a propagation of the wind within a given event.

The second SST mode of the WWB model contains important information, as it precedes mode 1 by 11 months and may therefore be a precursor to El Niño events. The coefficient relating this mode to x_o corresponds to a -2.56° longitude change in the location of the WWBs in response to a one standard deviation change in the amplitude of the second SST singular vector. Therefore, WWBs occur farther to the west than average when the second SST mode increases in amplitude. This seems sensible because a positive projection onto mode 2 reflects the warm pool edge being farther to the west. Interestingly, the coefficient relating the second SST mode to the WWB probability P is positive, 2.22 yr⁻¹, indicating that WWBs are more likely when mode 2 is excited. If WWBs occur when the SST has the structure of mode 2, the ocean response can lead to a warming of the central and eastern Pacific, thus initiating the development of an El Niño event some months later. Mode 2 explains much less variance of the WWB parameters than mode 1, but the aspects of the WWB evolution explained by this mode may still be crucial for ENSO.

c. Retrospective WWB predictions

Given the observed SST during 1988–2005, a retrospective prediction of the WWB parameters is produced by the regression model (Fig. 8). The robustness of the WWB model is checked by predicting the WWB parameters for a given year with a cross-validated model that does not explicitly use data from that year. For example, the prediction for 1997 is made with a model fitted only to data from years 1988–96 and 1998–2005. In prediction mode, the WWB model is

$$\mathbf{r}(t) = \tilde{\mathbf{s}}_{O}^{*}(t) \mathbf{W}_{O},\tag{7}$$

where $\mathbf{r}(t)$, a vector containing the estimated WWB parameters at a time *t*, is predicted from $\tilde{\mathbf{s}}_Q^*(t)$, the vector of SST mode coefficients for that time, normalized by the standard deviation of the expansion coefficients over the training period. The cross-validated estimate is not significantly degraded from an estimate made with the model fitted to data from all years (cf. thick and thin lines in Fig. 8), an indication that the statistical model is robust. One exception may be the discrepancy in model prediction for 1997, perhaps because an event of that amplitude only occurred once in the record.

The amplitude A, duration T, and probability P of wind bursts are all nonnegative by definition. Formally, this situation could be handled by a linear programming (e.g., Luenberger 1984) or a nonnegative least squares method (e.g., Tziperman and Hecht 1988). In this work, we have found only rare instances of negative values predicted by the WWB model for these parameters, in which case they are simply set to zero.

The model prediction of the WWB characteristics is deterministic, but we now introduce a scheme for creating multiple stochastic realizations of WWBs using the probability parameter P(t) in (3). A random number η is sampled at every model time step from a flat distribution between 0 and 1, and an event is triggered if $\eta < P(t)$. Given the time series of P, therefore, we can produce many realizations of the time series of WWB occurrence times.

Our model calculates the WWB wind stress field from the probability *P*, which in turn is calculated from the SST, and is therefore a stochastic model. It is instructive to compare the form of our model to models using explicit multiplicative noise forcing in the SST equation (Perez et al. 2005; Jin et al. 2007; Sura and Sardeshmukh 2008). Both types of models rely upon a random number generator and the inclusion of SST information into the stochastic process. Our model is different in that the SST determines the probability for a WWB, and a random number is then used to obtain specific realizations of the WWB wind field. Both formulations can be considered state-dependent noise processes, although our model is not derived from mathematical principles but, instead, from physical reasoning.

To address the relative importance of deterministic versus stochastic processes affecting the wind bursts, consider the number of WWBs recorded in each year from 1988 to 2004 (Fig. 9). Without any interannual variability, 7.5 WWBs per year ($P_0 \times 365 = 7.5$) are expected, with a standard deviation of 2.7 WWBs per year given the expression for the variance of the binomial distribution in section 2. The WWB model predicts a distribution of WWBs for each year, where the mean represents the deterministic modulation by SST. One



FIG. 9. Number of WWBs per year, 1988–2005: observed (gray histogram), predicted by the cross-validated WWB model (bold line with bold bars), and predicted by the WWB model with all 17 years of data (dashed line with thin bars). The 10th and 90th percentile limits (vertical bars) of the stochastic model prediction, the mean of the cross-validated distribution (squares), and the mean of the full WWB model (circles). The model captures the interannual variability of WWBs without any tuning parameters. The total number of WWBs is also consistent with observations.

thousand realizations of the WWB model were run, and the 10th and 90th percentile of the number of WWBs per year are plotted as the vertical bar in Fig. 9, representing the impact of the stochastic part of the WWB model. The observed number of WWBs lies within the modeled error bars for 14 of 17 yr, or 82% of the time, consistent with the expected value of 80% (because we use the range 10%–90%). Overall, it appears that the interannual variability in the number of WWBs can be predicted as a deterministic function of SST and the stochastic component of WWBs seems less dominant, raising hopes for the potential to improve ENSO prediction by including the SST modulation effect on WWBs in ENSO prediction models.

During some years, such as 1997, the observed number of WWBs is outside the one standard deviation level of the WWB model. Figure 10 shows the detailed model distribution computed from 1000 realizations of the year 1997. The observed number of WWBs $[N_{wwb}(yr = 1997) = 17]$ lies at the 95th percentile, indicating that if the model distribution is correct, such a high number of WWBs would only happen in 1 year out of 20. Year 1997 may have been a year when a special dynamical regime was reached, and therefore, the model distribution may be an underestimate owing to a model bias. At this point, however, our record is only 17 years long and such a result cannot be ruled out as being due to statistical chance.



FIG. 10. Model distribution of the number of WWBs in calendar year 1997 over 1000 realizations. The 5th and 95th percentiles of the distribution (dashed vertical lines), the mean of the distribution is 10 WWBs (solid, vertical line); 17 WWBs were observed in 1997.

d. The ensemble-mean WWB wind field

To further examine the importance of deterministic versus stochastic processes affecting the WWBs, we plot the observed WWBs on a longitude-time diagram, with three realizations of the WWB model with the same observed SST (Fig. 11). During El Niño years, WWBs migrate eastward into the central Pacific and they occur in groups. All three model realizations show this general behavior even though the exact timing of individual WWBs is different. Roulston and Neelin (2000) suggested that the high frequency component of WWBs is significantly less important to the excitation of ENSO events than the slow interannual component (see also Eisenman et al. 2005). The modeled WWB parameters vary with the same time scales as the SST and are, therefore, dominated by the slow interannual component.

It is possible to calculate an approximation of the slowly evolving ensemble-mean WWB wind field without simulating an ensemble of realizations. The probability of triggering a WWB at a time *t* is P(t) and the expected ensemble-mean amplitude is P(t)A(t). We must also account for the temporal variations in wind speed in the WWB life cycle, assumed in our model to vary with time as $u(t) \propto \exp(-t^2/T^2)$. At any given time and location, there is a contribution to the ensemble-mean wind by WWBs that are exactly at their peak and those that peaked earlier or later by an interval *t'*. The events that affect a given location may be centered at neighboring locations. Integrating over the ensemble of WWBs with different peak times, the ensemble-mean zonal wind speed at *t* is



FIG. 11. Westerly wind burst zonal wind speed as a function of time and longitude: (from top to bottom) a reconstruction from the observed WWB parameter vector and three different realizations calculated by the WWB model from the observed SST field. Values greater than 5 m s⁻¹ are shown, with a contour interval of 5 m s⁻¹.



FIG. 12. Comparison of three time–longitude representations of WWBs: (top) the low-frequency component (3-month running mean) of the observed WWBs, (middle) the ensemble average WWB wind field calculated from 1000 realizations, and (bottom) the direct calculation of the ensemble-mean WWB wind field from Eq. (9). Values greater than 1.8 m s⁻¹ (5/e = 1.8) are shown with contours at 5, 10, 15, and 20 m s⁻¹.

$$\langle u(x,y,t)\rangle = \int_{-\infty}^{\infty} P(t') A(t') \exp\left\{-\frac{(t-t')^2}{T(t')^2} - \frac{[x-X_o(t')]^2}{L_x(t')^2} - \frac{[y-Y_o(t')]^2}{L_y(t')^2}\right\} dt',$$
(8)

where P(t') denotes the likelihood of having a WWB peaking at t', whose amplitude, duration, location, and spatial extent are given by A(t'), T(t'), $X_o(t')$, $Y_o(t')$, $L_x(t')$, and $L_y(t')$. Because the WWB characteristics vary with the same time scale as SST, which is much

longer than the duration of an individual WWB event, we may assume $P(t') \approx P(t)$ to a good approximation. Similarly, the other WWB characteristics at a time t'may also be replaced by those at t. With this approximation, the integral is solved to give

$$\langle u(x,y,t) \rangle \approx \sqrt{\pi} P(t) A(t) T(t) \exp\left\{-\frac{[x - X_o(t)]^2}{L_x(t)^2} - \frac{[y - Y_o(t)]^2}{L_y(t)^2}\right\}.$$
 (9)

This formula allows a shortcut to directly calculate the slowly varying component of WWBs, which seems to be the most important part for ENSO dynamics and predictability (Roulston and Neelin 2000).

The low frequency component of the observations is shown in the top panel of Fig. 12. The ensemble average WWB wind field was calculated from 1000 realizations of the WWB model (middle panel) and from Eq. (9) (bottom panel). The middle and lower panels are virtually indistinguishable, suggesting that the approximations of the previous paragraph are valid. We also note that the expected ensemble-mean model WWB wind is similar to the low frequency component of the observed wind.



5. Model for both seasonal and interannual WWB variability

In the interannual-only WWB model discussed above, the seasonal cycle in wind burst parameters was prescribed from climatology. We next attempt to develop a model that will predict both the seasonal and the interannual WWB variability from the SST.

We first remove the 17-yr time-mean fields (constant in time) from the wind and SST observations. The SST modes of variability are now diagnosed as the left singular vectors of the new covariance matrix between SST and the WWB parameters (Fig. 13). The first SST mode still resembles the typical El Niño pattern, but the second mode, with opposite-sign structure across the equator, is related to the seasonal cycle. The singular vector expansion coefficients (Fig. 14) reveal that all seven SST modes contain variability at the annual period, demonstrating that the seasonal cycle is not completely represented by the second mode, as all seven modes are strongly influenced by seasonality. Philander (1990), for example, showed that the seasonal cycle in the tropical Pacific evolves via a complicated spatially propagating signal. The difference in the structure of the singular vectors 2–7 between Fig. 13 and Fig. 4 also serves as a warning that it is not straightforward to attach a dynamical meaning to any particular pattern.

Now that the WWB parameter anomalies contain seasonal variability, there is much more variance for the WWB model to explain. After performing the same procedure described in section 3, we find that WWB model still explains more than 50% of the variance in the zonal location and probability of WWBs. At interannual periods, the seasonal and interannual model explains similar amounts of variance as the interannualonly model—over 50% of the total variance in five of seven parameters. Figure 15 shows that all seven modes significantly reduce the model error in at least two WWB parameters according to the two-tailed F test. The 7 × 7 model matrix is included in Table 2.

The mean of 1000 realizations of the WWB model predicts more WWBs during the spring and fall, consistent with the observations (Fig. 16). Overall, the boreal summer is predicted to have a smaller number of WWBs than other

2005

2005

2005



seasons, as shown by Harrison and Vecchi (1997). The model predicts that July is the month with the least number of WWBs, as observed. The observed number of July WWBs over the last 17 years is outside the 10th percentile of the model distribution, but this should occur by chance one-tenth of the time. As this only occurs for the month of July, the observations are consistent with the modeled seasonal cycle. The size of the error bars are relatively large compared to the seasonal variations, and therefore, a longer observational record is necessary before we can have greater statistical confidence in these results.

Further explanation of the seasonal cycle of WWBs could shed light on some observed ENSO characteristics. The number of WWBs per month shows a doublepeak pattern, similar to the evolution of the sea surface during many El Niño events when the Niño-3.4 index peaks in both May and December. Furthermore, the seasonal cycle of WWBs can potentially explain the phase locking of ENSO to the seasonal cycle.

6. Conclusions

This paper developed and tested an observationally motivated approach for modeling and predicting westerly wind bursts from SST, which may be used to represent WWBs in ENSO simulation and prediction models. This approach is motivated by the analysis of TY who found that WWB characteristics are related to the largescale spatial patterns of SST, even if the linear correlation of the WWB wind stress itself with SST does not explain much of the WWB wind stress variance (Gebbie et al. 2007). The WWB model of this study goes beyond previous works by taking cause and effect into account in determining the statistically significant WWB response to SST. As a prognostic rather than diagnostic tool, it is successful in capturing the observed number of WWBs per year, the interannual variability of WWBs, and their eastward propagation and increased likelihood during a warm event. Apart from the hope that such an empirical WWB model may serve to improve ENSO prediction, a largely successful prediction of major WWB characteristics as is achieved in this paper also provides an important lesson about the dynamics of the tropical Pacific.

For our WWB model to be successful, we needed to determine in which ways SST can be used to determine a WWB response, and to understand the robustness (statistical significance) of the model. In addition, we tested the model performance by explicitly examining the model data misfit of WWBs over the last 20 years.



FIG. 15. The fraction of explained variance (top) for each WWB parameter and a WWB model with a given number of modes, 1–7, and (bottom) at periods greater than one year. (Anomaly model based on the 17-yr mean.)

We note that the correlation study of TY did none of these things, did not attempt to make a convincing case for the causality in the relationship between WWBs and SST, and could not be sensibly used for modeling studies.

Our WWB model can be integrated in a stochastic way, providing an ensemble of WWB realizations for a given SST history. Although the stochastic spread in the number of model WWBs can be up to five events in a given year, there is still a strong resemblance between the model-predicted and observed interannual variability of these events. This similarity, and the ability to directly calculate the ensemble-mean WWB wind stress, indicates that the deterministic modulation of WWBs by SST may dominate the stochastic signal in these events. This, in turn, may imply greater ENSO predictability than if the WWBs were an independent, white noise process. In a recent companion work (Gebbie and Tziperman 2009), it was found that the incorporation of the WWB model developed here to a hybrid coupled

TABLE 2. The transpose of the interannual and seasonal WWB model matrix \mathbf{W}_{Q} . The first seven rows give the seven regression vectors for the WWB parameters.

	1	2	3	4	5	6	7
$A (m s^{-1})$	0.01	1.02	0.40	0.61	-0.21	0.10	0.00
x_o (° lon)	18.76	-10.90	6.91	5.15	5.62	3.83	-0.86
L_x (° lon)	7.56	-2.87	1.14	3.64	1.75	-2.20	0.92
L_{v} (° lat)	2.62	-2.19	-0.60	0.30	1.77	-1.31	-0.21
T (day)	4.65	-0.94	0.44	3.81	-0.84	0.26	0.49
$P(yr^{-1})$	6.03	-5.37	-4.42	0.40	-0.05	-2.20	0.18
y_o (° lat)	-0.60	0.28	-1.19	1.05	-1.05	-0.98	-0.69



FIG. 16. Comparison of number of WWBs per calendar month as observed (gray, histogram) and predicted by the seasonal and interannual WWB model (lines and error bars) for the years 1988–2004: The cross-validated model (thick lines) and the model trained with the full 17-yr time series (dashed lines with thin error bars); the 10th and 90th percentiles of the modeled distributions (error bars), the mean of the cross-validated distribution (squares), and the mean of the full WWB model (circles).

model led to some improvement in the prediction of large El Niño events. These results are not conclusive because of the preliminary nature of the ENSO model used, and work is now in progress to test this WWB model within a state-of-the-art coupled general circulation ENSO model.

We assume that the WWB statistics are a function of SST and are therefore state dependent. Our model, when combined with an ENSO prediction model, is a form of a multiplicative noise process, although this terminology emphasizes the stochastic element of the WWBs, while an important message of this paper is that WWB statistics can largely be predicted from the SST. This implies that the WWBs are essentially predictable even if their precise occurrence time is random to within a few weeks. Only the interannual frequency component of the WWB variability affects ENSO (Roulston and Neelin 2000), and this slow part may be expected to be similar to the ensemble average WWB amplitude, which was found here to be quite predictable. Therefore, the essential contribution of WWBs to ENSO dynamics does not seem to involve the stochastic component of the WWBs. As noted, for example, by Eisenman et al. (2005), the SST-modulated WWBs act as an enhanced ocean-atmosphere coupling during warm events only, an effect which cannot be captured by linear statistical atmospheric models yet is important for our understanding of ENSO dynamics. Note that some other related works (Perez et al. 2005; Jin et al. 2007; Sura and

Sardeshmukh 2008) use multiplicative noise forcing in the more straightforward sense, following a detailed analysis of the heat equation, as was done by Penland and Matrosova (1994). We take a different, more empirical approach to deriving our model, and the relationship between the two approaches requires further study.

Most coupled GCMs do not simulate realistic WWB activity. Given the importance of WWBs during the onset of El Niño events, the skill of such models in predicting ENSO is clearly degraded. Our empirical model could be used within ENSO prediction models as part of an ensemble approach, provided that any WWBlike internal variability in the coupled model is taken into account first. The deterministic part of WWBs has not been quantitatively captured before, and thus implementation of the WWB model could lead to improved ENSO prediction skill. The stochastic part of the model allows the isolation and quantification of the spread in forecasts, and hence loss of predictability due to WWBs. It would be interesting to use this tool to examine the relative importance of the deterministic and stochastic parts of WWBs on ENSO forecasts, and future work can now directly address this problem.

Acknowledgments. This work was funded by the NSF Climate Dynamics program, Grant ATM-0351123, NASA (ECCO2 project), and the McDonnell Foundation. Martin Tingley provided assistance with the statistics. ET thanks the Weizmann Institute where some of this work was done during a sabbatical.

REFERENCES

- Batstone, C., and H. H. Hendon, 2005: Characteristics of stochastic variability associated with ENSO and the role of the MJO. *J. Climate*, **18**, 1773–1789.
- Bretherton, C. S., C. Smith, and J. M. Wallace, 1992: An intercomparison of methods for finding coupled patterns in climate data. J. Climate, 5, 541–560.
- Chang, P., L. Ji, B. Wang, and T. Li, 1995: On the interactions between the seasonal cycle and El Niño–Southern Oscillation in an intermediate coupled ocean–atmosphere model. *J. Atmos. Sci.*, 52, 2353–2372.
- Eisenman, I., L. S. Yu, and E. Tziperman, 2005: Westerly wind bursts: ENSO's tail rather than the dog? J. Climate, 18, 5224–5238.
- Fasullo, J., and P. Webster, 2000: Atmospheric and surface variations during westerly wind bursts in the tropical western Pacific. *Quart. J. Roy. Meteor. Soc.*, **126**, 899–924.
- Fedorov, A. V., S. L. Harper, S. G. Philander, B. Winter, and A. Wittenberg, 2003: How predictable is El Niño? *Bull. Amer. Meteor. Soc.*, 84, 911–919.
- Gebbie, G., and E. Tziperman, 2009: Incorporating a semistochastic model of ocean-modulated westerly wind bursts into an ENSO prediction model. *Theor. Appl. Climatol.*, doi:10.1007/s00704-008-0069-6, in press.

- —, I. Eisenman, A. Wittenberg, and E. Tziperman, 2007: Modulation of westerly wind bursts by sea surface temperature: A semistochastic feedback for ENSO. J. Atmos. Sci., 64, 3281–3295.
- Harrison, D. E., and G. A. Vecchi, 1997: Westerly wind events in the tropical Pacific, 1986–95. J. Climate, 10, 3131–3156.
- Harrison, M. J., A. Rosati, B. J. Soden, E. Galanti, and E. Tziperman, 2002: An evaluation of air–sea flux products for ENSO simulation and prediction. *Mon. Wea. Rev.*, **130**, 723–732.
- Jin, F.-F., L. Lin, A. Timmermann, and J. Zhao, 2007: Ensemblemean dynamics of the ENSO recharge oscillator under statedependent stochastic forcing. *Geophys. Res. Lett.*, 34, L03807, doi:10.1029/2006GL027372.
- Kerr, R. A., 1999: Atmospheric science: Does a globe-girdling disturbance jigger El Niño? *Science*, 285, 322–323.
- Luenberger, D. G., 1984: Linear and Nonlinear Programming. Addison-Wesley, 491 pp.
- Luther, D. S., D. E. Harrison, and R. A. Knox, 1983: Zonal winds in the central equatorial Pacific and El Niño. Science, 222, 327–330.
- Mason, S. J., and G. M. Mimmack, 2002: Comparison of some statistical methods of probabilistic forecasting of ENSO. *J. Climate*, **15**, 8–29.
- McPhaden, M. J., 2004: Evolution of the 2002/03 El Niño. Bull. Amer. Meteor. Soc., 85, 677–695.
- Penland, C., and L. Matrosova, 1994: A balance condition for stochastic numerical-models with application to the El Niño– Southern Oscillation. J. Climate, 7, 1352–1372.
- Perez, C. L., A. M. Moore, J. Zavala-Garay, and R. Kleeman, 2005: A comparison of the influence of additive and multiplicative stochastic forcing on a coupled model of ENSO. *J. Climate*, **18**, 5066–5085.
- Philander, S. G. H., 1990: El Niño, La Niña, and the Southern Oscillation. Academic Press, 293 pp.
- Picaut, J., F. Masia, and Y. DuPenhoat, 1997: An advectivereflective conceptual model for the oscillatory nature of the ENSO. *Science*, **277**, 663–666.
- Roulston, M. S., and J. D. Neelin, 2000: The response of an ENSO model to climate noise, weather noise and intraseasonal forcing. *Geophys. Res. Lett.*, **27**, 3723–3726.
- Sura, P., and P. D. Sardeshmukh, 2008: A global view of non-Gaussian SST variability. J. Phys. Oceanogr., 38, 639–647.
- Syu, H.-H., J. Neelin, and D. Gutzler, 1995: Seasonal and interannual variability in a hybrid coupled GCM. J. Climate, 8, 2121–2143.
- Tziperman, E., and A. Hecht, 1988: Circulation in the Eastern Levantine Basin determined by inverse methods. J. Phys. Oceanogr., 18, 506–518.
- —, and L. Yu, 2007: Quantifying the dependence of westerly wind bursts on the large-scale tropical Pacific SST. J. Climate, 20, 2760–2768.
- Vecchi, G. A., A. T. Wittenberg, and A. Rosati, 2006: Reassessing the role of stochastic forcing in the 1997–1998 El Niño. *Geophys. Res. Lett.*, **33**, L01706, doi:10.1029/2005GL024738.
- Wittenberg, A. T., 2002: ENSO response to altered climates. Ph.D. thesis, Princeton University, 475 pp.
- —, A. Rosati, N.-C. Lau, and J. J. Ploshay, 2006: GFDL's CM2 global coupled climate models. Part III: Tropical Pacific climate and ENSO. J. Climate, 19, 698–722.
- Yu, L., R. A. Weller, and T. W. Liu, 2003: Case analysis of a role of ENSO in regulating the generation of westerly wind bursts in the Western Equatorial Pacific. J. Geophys. Res., 108, 3128, doi:10.1029/2002JC001498.