

## The Niño-3.4 Prediction Skill of Empirically Adjusted Wind Power

KERI KODAMA,<sup>a</sup> NATALIE J. BURLS,<sup>a</sup> AND LAURIE TRENARY<sup>b</sup>

<sup>a</sup> *Department of Atmospheric, Oceanic, and Earth Sciences, George Mason University, Fairfax, Virginia*

<sup>b</sup> *Center of Ocean–Land–Atmosphere Studies, George Mason University, Fairfax, Virginia*

(Manuscript received 22 January 2020, in final form 21 October 2020)

**ABSTRACT:** Wind power, defined as the energy received by the ocean from wind, has been identified as a potentially viable precursor of ENSO. The correlation between tropical Pacific wind power anomalies and eastern equatorial Pacific sea surface temperature anomalies can be enhanced over a range of lead times by applying an empirical adjusted framework that accounts for both the underlying climatological state upon which a wind power perturbation acts and the directionality of wind anomalies. Linear regression is used to assess the seasonal prediction skill of adjusted wind power in comparison to unadjusted, as well as the conventional ENSO predictors wind stress and warm water volume. The forecast skill of each regression model is evaluated in a 1800-yr preindustrial climate simulation (CESM-LENS), as well as 23 years of observations. The simulation results show that each predictor's effectiveness varies considerably with the sample record, providing a measure of the uncertainty involved in evaluating prediction skill based on the short observational record. The influence of climatological biases is however a demonstrable concern for results from the simulated climate system. Despite the short record, the observational analysis indicates that adjusted wind power skill is comparable to the conventional dynamical predictors and notably is significantly more predictable than unadjusted wind power when initialized in the summer. Moreover, the adjusted framework results in a reduction of error when evaluating wind power associated with wind bursts, reinforcing previous findings that the adjusted framework is particularly useful for capturing the ENSO response to westerly wind bursts.

**KEYWORDS:** El Niño; ENSO; Regression analysis


### 1. Introduction

El Niño–Southern Oscillation (ENSO) is a mode of atmosphere–ocean interaction whereby zonal wind anomalies give rise to oceanic pressure and sea surface temperature (SST) gradient anomalies along the equatorial Pacific, which in turn sustain the anomalous atmospheric circulation until delayed negative feedbacks act to reverse these perturbations (Bjerknes 1969; Neelin and Dijkstra 1995). El Niño is an extreme phase of ENSO, characterized by weaker trade winds, a relaxation of the SST and pressure gradients, and warmer eastern Pacific SST anomalies and vice versa for its respective counterpart La Niña. In the ENSO energetics perspective presented by Goddard and Philander (2000), a relationship is seen between equatorial Pacific oceanic available potential energy (APE) and ENSO state. APE is defined as the energy that would result from a transition from the current state to a minimum energy base state in the absence of the horizontal pressure gradients and generally quantifies the slope of the zonal mass field (Lorenz 1955; Oort et al. 1989). As such, high APE is associated with a steeper thermocline and La Niña and low APE with a flatter thermocline and El Niño.

Wind power, defined as the work imparted to the ocean by the winds, provides the dominant source of APE variability in the tropical Pacific through the conversion of kinetic energy

into APE. Changes in wind power force changes in APE and, by proxy, Niño SST on seasonal time scales, such that positive wind power anomalies correspond to cooling SSTs and negative wind power anomalies correspond to warming. Given this association, wind power has thus been identified as another potential precursor of ENSO variability by Goddard and Philander (2000) and Philander and Fedorov (2003). While the role of wind power in an ENSO energetics framework has been explored by numerous studies focusing primarily on climate model data (e.g., Brown and Fedorov 2008; Brown and Fedorov 2010; Hu et al. 2014), there is limited assessment of its actual viability as an ENSO predictor.

Kodama and Burls (2019, hereafter KB19) perform the first analysis of a tropical Pacific wind power index based on satellite observations. KB19 show that the correspondence between wind power anomalies and the associated ENSO response strongly depends on the expectation that the mean state of the tropical Pacific is characterized by westward currents, easterly winds, and a west–east thermocline slope at all times of the year and in all parts of the basin. KB19 find that this is not true everywhere, and in the regions where the configuration of the mean state does not align with the conventional expectation, inconsistencies in the sign association between wind power and SST can arise. For example, during the spring, the equatorial undercurrent surfaces, resulting in a shift in the climatological currents from westward to slightly eastward. This results in regionally dependent ambiguities between the sign of the wind power anomaly and the corresponding ENSO response. KB19 also find that a second type of inconsistency occurs when westerly wind events (WWE)—which are understood dynamically to induce warming in the eastern Pacific

 Denotes content that is immediately available upon publication as open access.

Corresponding author: Keri Kodama, [kkodama@gmu.edu](mailto:kkodama@gmu.edu)

DOI: 10.1175/JCLI-D-20-0045.1

© 2021 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

through the propagation of downwelling Kelvin waves (Lengaigne et al. 2004; Seiki and Takayabu 2007a,b)—are strong enough to outright reverse the underlying surface currents. The resulting wind power anomalies are positive, in contradiction to the aforementioned ENSO energetics convention that positive wind power corresponds to SST cooling. In such cases, sufficiently strong WWE can result in the same wind power anomalies as easterly wind events (EWE), creating further ambiguities in the overall wind power correspondence to SST anomalies.

KB19 propose an empirically modified ENSO energetics framework in which sign adjustments are made to the wind power computation to create an “adjusted wind power” index that accounts for the sign inconsistencies. The sign adjustments result in higher correlations with Niño-3.4 SST than wind power without adjustments (“unadjusted wind power”) up to 9 months in advance for a December target. The primary advantages of the adjusted wind power are that the inconsistencies resulting from the mean state dependence and WWE are removed, allowing the adjusted wind power to better capture the growth of major El Niño events. The analysis in KB19 motivates a more rigorous assessment of the comparative predictive skill of adjusted wind power versus unadjusted, as well as its skill in relation to the prevailing dynamical precursors, namely equatorial wind stress and warm water volume (WWV) (Meinen and McPhaden 2000; Kessler 2002). A major obstacle in ENSO forecasting is the well-known spring predictability barrier, a period in March–May in which forecasts initialized during this time have low skill, even when predicting only a month or two in advance (Torrence and Webster 1998; McPhaden 2003; Barnston et al. 2012). How the wind power fares against the predictability barrier compared to the other predictors is one of several points of interest in our investigation. Thus, the primary goals of this work are as follows: 1) to compare the predictive skill of the adjusted wind power index with that of unadjusted wind power using linear regression and 2) to determine the relative skill of the adjusted wind power with respect to other conventional predictors and dynamical models.

Section 2 introduces the methodology used to train and create the linear regression models as well as the metrics that will be used to evaluate the forecasts. Additionally, a summary of the adjusted framework conceptualization is included. Section 3 covers the different predictors that will be used in the analysis, which are also detailed in Table 1. An analysis of the much longer preindustrial control run from the Community Earth System Model Large Ensemble (CESM LENS) version 1 configuration is presented in section 4. This analysis is then used to inform the conclusions drawn from the observational analysis in section 5. Finally, a discussion of the results is given in section 6.

## 2. The predictive model

To assess predictive skill of adjusted wind power, we employ a simple linear regression model

$$\text{SST}(t) = \mathbf{X}(t - T)\boldsymbol{\beta} + \epsilon(t), \quad (1)$$

TABLE 1. Predictor variable notation and list of regression model configurations.

Predictor notation and definitions		
SST	Sea surface temperature anomalies in the Niño-3.4 region (170°–120°W, 5°S–5°N)	
WS	Wind stress anomalies averaged over the region 165°E–165°W, 5°S–5°N	
WPa	Basin integrated adjusted wind power anomalies in the region 140°E–70°W, 15°S–15°N (KB19)	
WPu	Basin integrated unadjusted wind power anomalies in the region 140°E–70°W, 15°S–15°N	
WWV	Warm water volume, volume of water above 20°C isotherm in the region 120°E–80°W, 5°S–5°N	
WB5/WB15	Basin integrated cumulative wind burst index in the region 140°E–70°W, 5°S–5°N/15°S–15°N	
WBPa5/WBPa15	Cumulative adjusted wind power due to wind bursts in the region 140°E–70°W, 5°S–5°N/15°S–15°N	
WBPu5/WBPu15	Cumulative unadjusted wind power due to wind bursts in the region 140°E–70°W, 5°S–5°N/15°S–15°N	
Regression model predictor configurations		
Main predictors (sections 4 and 5)	Wind burst predictors (section 5b)	
1. SST only	1. WB5	2. WB15
2. WS only	3. WBPa5	4. WBPa15
3. WPa only	5. WBPu5	6. WBPu15
4. WPu only		
5. WWV only		

where  $\text{SST}$  is the  $N \times 1$  vector of the November–January (NDJ) 3-month averaged Niño-3.4 SST anomalies being predicted (i.e., the predictand) as a function of monthly time  $t$ ;  $N$  is the number of data points in the sample. The matrix  $\mathbf{X}$  is the  $N \times M$  predictor matrix as a function of time at a seasonal lead time  $T$ , where  $M$  is the number of predictors used,  $\boldsymbol{\beta}$  is the  $M \times 1$  vector of regression coefficients that minimize the least squares error, and  $\epsilon$  is the residual. The models are trained to predict the NDJ Niño-3.4 SST anomalies using adjusted wind power, unadjusted wind power, wind stress, WWV, and SST each as single predictors in the  $\mathbf{X}$  matrix. One element of the findings from KB19 is that the adjusted wind power framework is important for correctly capturing the sign correspondence between WWEs and Niño-3.4 SST in the development of ENSO events. Thus, we also test single-predictor models this time using wind power derived specifically from wind bursts to determine if they contribute substantially to the overall predictability. Further detail on how the wind burst power indices are constructed is given in section 3. Acronyms used to

reference the different predictors are defined in Table 1. The observational datasets used to derive these predictors only overlap for years 1993–2015 and as such these are the years over which the observationally based skill assessment is conducted. All anomalies are estimated with respect to monthly climatologies over the training record, except in the case of wind bursts, which require daily data.

A primary focus of this study is to evaluate the dependence of prediction skill on seasonal leads for forecasts of NDJ Niño-3.4 SST. The predictors are converted to running 3-month seasonal averages [i.e., January–March (JFM), February–April (FMA), and so on through October–December (OND)], and the models are trained for each initialization season starting with JFM predicting NDJ Niño-3.4 SST. The lead time  $T$  in Eq. (1) is determined by the difference between the central month of the initialization season and December (e.g., models trained using JFM predictors have  $T = 10$ ). A total of 11 linear regression models will be evaluated with the main focus directed toward evaluating the predictive skill of the two wind powers, wind stress, SST, and WWV (Table 1). The skill of the observational regression models is assessed based on leave-one-out cross-validation (L1OCV), in which one data point is left out as validation while a model is trained on the remaining data to predict said data point. While computationally more expensive, L1OCV allows us to maximize the amount of data that can be used to train the models. Since data availability is not an issue with the LENS control run, there is more flexibility in the choice of training and verification datasets. These choices and the methodology used to evaluate the LENS-based regression models are the focus of section 4.

A linear regression framework is used to evaluate predictive skill because it is computationally inexpensive and straightforward to implement (e.g., Pegion and Selman 2017; Dominiak and Terray 2005; DelSole and Tippett 2016). Although linear regression models do not accommodate any dependence on the spatial structure of predictors and are simplistic forecast models, they are useful tools for establishing a baseline to evaluate predictors and more complex models. While their skill inherently depends on the linearity of ENSO, the analysis of an ENSO linear inverse model (LIM) by Newman and Sardeshmukh (2017) suggests that when the spatial variations of ENSO are accounted for, the phenomenon is indeed primarily linear.

The primary skill metrics used for this analysis are cross-validated root-mean-square error (RMSE), correlation skill of the model SST forecasts from the cross-validation, and an additional test known as the random walk skill score (DelSole and Tippett 2016). While RMSE and correlation skill are widely used skill assessments, they do not provide a robust basis by which we can determine if a given model can be considered significantly more skillful than another model on a case-by-case basis (DelSole and Tippett 2016). In the case of RMSE, determining the underlying probability density function of the error is difficult, making them difficult to compare. On the other hand, while correlation coefficients do have known distributions, their comparison assumes independent data samples, which is untrue for forecasts over the same time range.

The random walk skill score (DelSole and Tippett 2016) treats the pointwise comparison of two models as a binary

operation—either one model prediction or the other is closer to observation. The comparative prediction skill for each forecast value can then be treated like a coin toss, as each of the two models being compared has an equally likely chance of being closer to observation. For each forecast, one reference model is compared to any number of competing models. If the reference model's forecast is closer to observations than its competitor, it receives a score of +1 for that particular forecast. If it is worse than its competitor, it instead receives a  $-1$ . If the number of times the reference model outperforms its competitors exceeds the number of times it loses by a threshold of  $2\sqrt{N}$ , it can be said to be the more skillful model at 95% confidence. The threshold follows from the binomial distribution, where  $N$  is the number of forecasts compared. It quantifies the uncertainty in the skill assessment based on the sample size.

A common definition of skill is such that a forecast can be considered “skillful” when it has more skill than a climatological forecast (i.e., a zero anomaly). Hence, a model can be considered skillful if it significantly outperforms the zero-anomaly forecast in a random walk test. If the random walk score does not exceed the  $2\sqrt{N}$  threshold, then the differences in model skill are statistically indistinguishable for the set of forecasts being compared.

RMSE and correlation provide a measure of average error over all forecasts, while the random walk depicts the relative error of each predictor for each forecast point. As such, we evaluate all three metrics to get the broadest sense of each model's predictive skill.

### 3. Overview of predictors and data

#### a. Sea surface temperature and warm water volume

The Niño-3.4 index (averaged SST anomalies in the region  $170^{\circ}$ – $120^{\circ}$ W,  $5^{\circ}$ S– $5^{\circ}$ N) is computed using HadISST from the Met Office Hadley Centre (Met Office 2003; Rayner et al. 2003). HadISST is available from 1870 to the present as  $1^{\circ} \times 1^{\circ}$  gridded monthly data. The WWV index is defined as the volume of water above the  $20^{\circ}$ C isotherm in the region  $120^{\circ}$ E– $80^{\circ}$ W,  $5^{\circ}$ S– $5^{\circ}$ N and is computed from “Estimating the Circulation and Climate of the Ocean” (ECCO) version 4 re-release 3-monthly potential temperature, which is available on a global Lat-Lon-Cap90 (LLC90) grid from 1992–2015 and provided by the National Aeronautics and Space Administration Jet Propulsion Laboratory (NASA JPL; Fenty et al. 2015; Fukumori et al. 2017; Forget et al. 2015).

#### b. Wind stress

The wind stress data in this study use the zonal component of the TropFlux wind stress product (Praveen Kumar 2013). The TropFlux data are a precomputed surface wind stress product available from 1979 to 31 May 2016 at the time of our analysis. It is provided by the Indian National Center for Ocean Information Services (INCOIS). Further details on the processing and verification methods are provided in Praveen Kumar et al. (2012) and compared with other wind stress products. The TropFlux wind stress is used to compute the wind stress index and is defined as the spatially averaged wind

stress anomalies in the region 165°E–165°W, 5°S–5°N. The boundaries of this index region are chosen based on where the wind stress is most strongly correlated with Niño-3.4 SST at zero lag.

### c. Wind power

The wind power  $\int_S u\tau dS$  is area integrated over the region 140°E–70°W, 15°S–15°N, where  $u$  is the zonal surface currents,  $\tau$  is the wind stress, and  $S$  represents the integration area. The wind stress used to compute wind power is taken from the same TropFlux dataset used for the wind stress index. The zonal surface currents are taken from the Earth and Space Research (ESR) Ocean Surface Current Analyses Real-time (OSCAR) project (ESR 2007; Lagerloef et al. 1999). The surface current data are available as  $1^\circ \times 1^\circ$  and  $0.33^\circ \times 0.33^\circ$  grids, available in pentad format from October 1992 through the present. Following from KB19, we specifically focus our analysis on wind power anomalies. To represent wind power anomalies, we first decompose the wind stress and surface currents as  $\tau = \bar{\tau} + \tau'$  and  $u = \bar{u} + u'$ , where  $\bar{\tau}$  and  $\bar{u}$  represent the Reynolds averages and  $\tau'$  and  $u'$  are the deviations about the average. Then the product  $u\tau$  can be expanded into

$$u\tau = \bar{u}\bar{\tau} + \bar{u}\tau' + u'\bar{\tau} + u'\tau'. \quad (2)$$

The anomalous wind power about the climatological wind power is defined as

$$(u\tau)' = u\tau - \bar{u}\bar{\tau} = \bar{u}\tau' + u'\bar{\tau} + u'\tau' - \overline{u'\tau'}, \quad (3)$$

where  $\bar{u}\bar{\tau}$  is the climatological wind power. Note that  $\bar{u}\tau' + u'\bar{\tau}$  is referred to as “mixed perturbation wind power” and  $u'\tau'$  as the “perturbation wind power”;  $\overline{u'\tau'}$  is the Reynolds averaged perturbation wind power.

As discussed in section 1, there are two types of sign inconsistencies that need to be corrected in the conventional ENSO energetics framework. Specifically, the conventional framework associates positive wind power anomalies with negative SST anomalies and vice versa for negative wind power. This convention depends on both the assumption that the equatorial climatological state is negative wind stress and currents with a thermocline sloping up from west to east and the assumption that the perturbation wind power ( $u'\tau'$ ) is small relative to the mixed perturbation components ( $\bar{u}\tau' + u'\bar{\tau}$ ). The first type of adjustment addresses inconsistencies in the assumptions of the climatology. We invert the sign of the relevant mixed perturbation terms at each grid point in which the climatology does not align with convention—for example,  $\bar{u}\tau'$  is inverted where climatological surface currents are positive.

The second type of adjustment addresses instances where the perturbation wind power is not small. This is specifically relevant to WWEs, which locally reverse the surface currents and result in positive wind power anomalies, despite overall contributing to a reduction of APE on basinwide scale. Thus, for the perturbation adjustments, we invert the  $u'\tau'$  term wherever  $\tau'$  is positive, so that easterly anomalies correspond to La Niña and westerly anomalies to El Niño. Whether the respective components of the wind power are adjusted is

contingent on whether the current and wind stress climatological and perturbation fields meet the adjustment conditions detailed in KB19. Each time step and grid point is checked to determine whether an adjustment condition is met; if so, the appropriate adjustment is applied. KB19 notes that unlike the first, climatologically based adjustment, the second adjustment type no longer upholds energy conservation—hence the “empirical” element of the adjusted framework. Adjusted wind power still however contains added information related to ocean adjustment that is contained with the surface current field because it “weights” the atmospheric wind stress perturbations by the underlying surface currents—the key strength of the wind power metric emphasized in the conventional wind power framework. As discussed in section 1 and KB19, the conventional ENSO energetics framework depends on certain expectations of the mean state and perturbation magnitude for the sign association between wind power and SST to hold. The adjusted framework focuses on maintaining consistency of that sign association, even when those assumptions about the background state and perturbation strength no longer hold.

### d. Wind bursts

The wind burst analysis tests the prediction skill of a cumulative integrated wind burst power index. The index is created by first filtering out wind stress variability outside of the intraseasonal (5–90 day) range at each grid point based on the parameters in Puy et al. (2016). As in Puy et al. (2016), a magnitude threshold of  $\pm 0.04 \text{ N m}^{-2}$  is also imposed on the filtered daily wind stress anomalies such that all regions below the threshold are removed. The wind power adjustments and surface spatial integration are applied following the methodology of KB19 and the adjusted version of Eq. (3) is integrated over the wind power index region described above (140°E–70°W, 15°S–15°N). The anomalous wind stress terms in Eq. (3) are replaced with the intraseasonally filtered threshold-restricted wind stress with positive values corresponding to WWE and negative to EWE. Thus, only grid points in a region where a wind burst is occurring are included in the spatial integration. As such, terms in Eq. (3) that do not explicitly depend on anomalous wind stress, namely  $u'\bar{\tau}$ , are still incorporated into the wind burst power so that the surface current response in wind burst regions factors into the wind burst power index. Finally, an additional time integration is performed from the start of each year, and the monthly averaged seasonal cycle of the integrated index is removed in order to create a yearly cumulative anomaly index. The purpose of the time integration is both to temporally smooth and to capture a continuous accumulated effect of wind bursts over extended durations. This integrated adjusted wind burst power index (WBPa) is created for both the 5°S–5°N and 15°S–15°N latitude ranges to compare the importance of equatorial versus off-equatorial wind events. Integrated indices are similarly made for unadjusted wind burst power (WBPU)—that is, wind power due to wind bursts without applying the adjusted framework—as well as basin integrated wind bursts (WB), namely applying the same wind burst filtering and basin integration as in wind power, except with the  $u$  field set to 1. The integrated



wind burst and wind burst power indices are then used to train regression models. For a full description of the adjusted framework, see [KB19](#).

#### e. NMME

As a skill benchmark for the observations, we include the data from the North American Multimodel Ensemble project (NMME; [Kirtman et al. 2014](#)). The NMME consolidates ensemble forecasts from a variety of operational forecast models, producing both real-time and hindcast data based on standardized initialization procedures on time scales ranging from intraseasonal through interannual. The hindcast data are initialized every month with lead times ranging from 2 weeks to 12 months, depending on the model. The NMME multimodel mean forecast has been shown to have greater skill than even the most skillful individual model ([Kirtman et al. 2014](#); [Becker et al. 2014](#)). We will use “NMME” to refer to the forecast and skill metrics associated with the multimodel ensemble mean. Unlike the linear regression model in which 3-month seasonal means (e.g., JFM or FMA) are used to predict NDJ Niño-3.4 SST anomalies at a range of lead times, for the NMME skill metrics shown for each lead time correspond to simulation initialized at the beginning of the center month (e.g., in [Fig. 1](#) the FMA lead corresponds to NMME simulations initialized in March). Certain models do not have forecasts for all lead times used in this study. The models are included for any seasons for which they have data availability. The models used to create the multimodel mean are taken from the IRI database and listed with their associated centers in [Table 2](#). Bias correction is performed by computing and removing the monthly climatological SST from each model and computing the seasonal averages of the monthly anomalies.

#### 4. Linear regression forecasts in the CESM Large Ensemble

The predominant interest of this study is to gain insight into the performance of the adjusted wind power as a predictor in the observed world. However, the observational record is only 23 years, subsequently limiting the sample size for a seasonally based forecast assessment each year. To examine the sensitivity of predictor skill to sample size we analyze the 1800 years of coupled model data from the CESM LENS preindustrial (PI) control run ([Kay et al. 2015](#)). This analysis in turn will be used to better contextualize the observational analysis in the following section. The use of the PI control is itself not without caveats, the first being that the control run is not an operational forecast model, and therefore cannot be verified against observations on an event-by-event basis. A second caveat is that CESM, as with other coupled models, is subject to the well-known cold tongue and double ITCZ biases, which impact the estimation of wind power through the propagation of these biases to the tropical wind stress and zonal surface currents used to compute wind power ([Li and Xie 2014](#); [Burls et al. 2016](#); [KB19](#)). Specifically, [KB19](#) find that these climatological biases impact the efficacy of the wind power adjustments in LENS compared to observations (an important aspect to keep in mind when evaluating the skill of adjusted versus unadjusted

wind power within the LENS dataset). Despite these limitations, CESM is generally regarded as a good simulator of ENSO variability ([Gent et al. 2011](#); [Hu and Fedorov 2017](#)) and provides a long independent data sample to test the general behavior of the predictive models.

The PI control is analyzed in two different ways. The first analysis is used to understand the degree of uncertainty involved in using only 23 years of data to train and verify the models. The second analysis examines how the predictors fare against each other in a statistically robust 500-yr verification sample.

In the first, a series of 23-yr samples is drawn from the dataset to be consistent with observations (years 407–429, 490–512, 1166–1188, 1285–1307, 1503–1525, and 1995–2017). It has been shown in [KB19](#) that there is some association between ENSO variability and the effectiveness of the adjusted wind power relative to unadjusted. Therefore, the choice of sample years is based on the periods during which the LENS Niño-3.4 index has either the lowest, highest, or median standard deviations. These samples are then used to create seasonally stratified forecasts of the NDJ Niño-3.4 index using the L1OCV technique mentioned in [section 2](#) for apples-to-apples comparison with observations. The associated RMSE is shown in the left-hand columns of [Figs. 1](#) and [2](#).

In the second analysis method ([Figs. 1](#) and [2](#), right columns), the same aforementioned 23-yr samples are used in their entirety as training datasets. That is, instead of subdividing the 23-yr sample into a training section and forecast section, the full 23 years of a given sample is used to train a corresponding single regression model for each predictor season. The trained model is then used to forecast the following 500 years of NDJ Niño-3.4 SST. In the case of training periods for which the 500-yr forecast period will extend beyond the available years of the PI control, the forecast period loops around and samples from the very beginning of the control run instead.

We assess the skill of the single predictor models listed in [Table 1](#) (lower left column) for the PI control analysis. All forecasts use data normalized by their respective anomaly standard deviations to account for differences in the order of magnitude of variability of the predictor variables. The comparisons of the predictors can vary considerably when different 23-yr samples of the LENS data are used, and the envelope of possible RMSE values is already large for six training sets ([Figs. 1](#) and [2](#), left column). On the other hand, when the forecast length is increased to 500 years, the errors appear to converge ([Figs. 1](#) and [2](#), right column, blue). This indicates that given a sufficiently long record, the structure of the RMSE appears to be largely independent of training period and regression coefficients ([Figs. 1](#) and [2](#)). This convergence occurs despite the fact that variability of the Niño-3.4 changes considerably with training period. Some of this might potentially result from overlap in the different forecast samples, although the convergence occurs regardless of whether the forecast periods intersect. Another possibility is simply that the error sample distribution approaches the population distribution as sample size increases.

In general, the unadjusted wind power (WPU) has the lowest RMSE for forecasts trained on JFM and FMA data over all

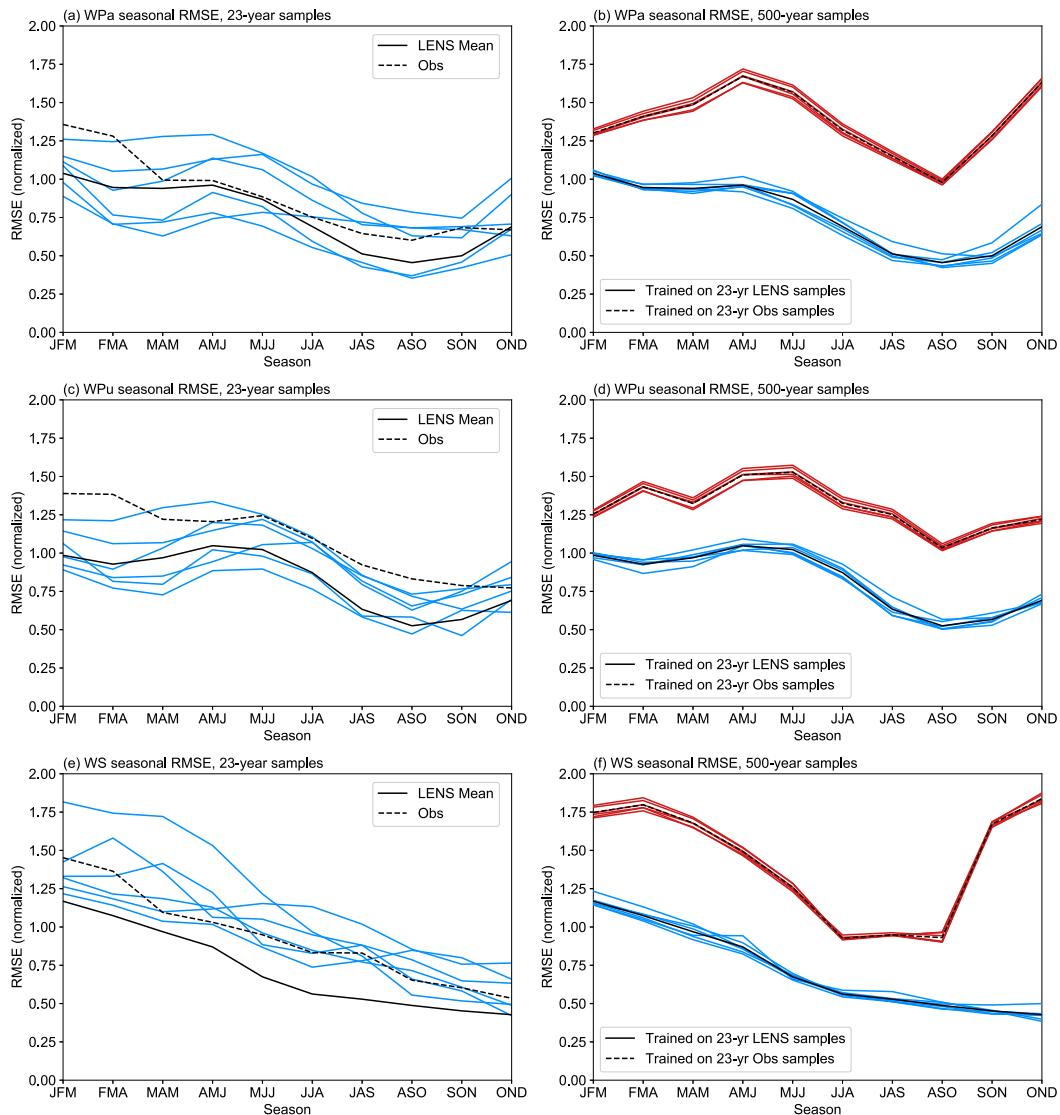


FIG. 1. (a),(c),(e) RMSE of 23-yr cross-validated samples of LENS for WPa, WPU, and wind stress, respectively. The observational RMSE is superimposed in the dotted black and the model 500-yr mean RMSE of the six sample sets tested in the solid black. (b),(d),(f) RMSE of 500-yr validation sample of LENS for WPa, WPU, and wind stress, respectively, for regression models trained using LENS data (blue; mean solid black) and models trained using the 23-yr observed dataset forecasting the same respective 500-yr periods (red; mean dotted black).

other predictors. From AMJ through JJA, integrated wind stress takes over, after which SST becomes the best predictor. MAM is the most sensitive time of the year, when WPa, WPU, and wind stress are close enough in value that even a small change can impact the comparative error.

Next, we evaluate the general skill of the five predictors using the random walk skill score. For the sampled periods, all models except SST and occasionally WWV show significant skill as early as JFM (Fig. 3), and this skill persists as lead time decreases. By MAM (not shown), SST becomes consistently skillful alongside the two wind power products, while WWV remains the most dependent on forecast sample throughout all lead seasons. It frequently takes more than 50 forecast

comparisons before this skill can be established, and in many cases, more than 100. For seasonal forecasts, this corresponds to 50–100 years of data.

Figures 4 and 5 show the random walk test with WPa as the reference model for AMJ and JFM in the different sample records, respectively. By AMJ (Fig. 4) and onward (not shown), WPa significantly outperforms WPU in most samples. However, at the longer lead times JFM (Fig. 5) and FMA (not shown) WPU significantly outperforms WPa in four of the six samples and two of the six samples, respectively. Moreover, while WPa continues to significantly outperform WPU during the rest of the year, at the same time, the wind stress index takes over and surpasses WPa in turn.

TABLE 2. NMME System Phase II models used to compute the multimodel mean, based on Kirtman et al. (2014).

Center	Models
National Centers for Environmental Prediction (NCEP)	CFSv1 CFSv2
Rosenstiel School of Marine and Atmospheric Science (RSMAS)	Community Climate System Model version 3 (CCSM3) Version 4 (CCSM4)
National Center for Atmospheric Research (NCAR)	Community Earth System Model version 1 (CESM1)
Geophysical Fluid Dynamics Laboratory (GFDL)	Climate Model 2.1 (CM2p1) Climate Model 2.5 Forecast-oriented Low Ocean Resolution (CM2p5-FLOR)
National Aeronautics and Space Administration (NASA)	Global Earth Observing System version 5 (GEOS5)
Canadian Centre for Climate Modeling and Analysis	Coupled Global Climate Model 3rd generation (CanCM3) 4th generation (CanCM4)

The random walk tests for each forecast sample show that there is some correspondence between the relative performance of the predictors in RMSE and their random walk scores, though it is not a perfect association. For example, MAM WPa has a lower RMSE than wind stress in five of the six samples, but never significantly outperforms wind stress from a random walk perspective (not shown). Even with minor variations in relative RMSE, the results of the corresponding random walk test can vary greatly. The discrepancy between these metrics results from the fact that they are measuring different aspects of model skill. The random walk test does not account for the magnitude of difference between the forecasts, only whether one model is closer to the verification than

another. However, the random walk does depend considerably on which events are included in the verification set. The RMSE, on the other hand, penalizes a model if it has one or two years that are dramatically inaccurate.

This prompts the question of whether there are specific validation periods during which any given predictor is more impactful, or whether the effectiveness of the predictors shows dependence on ENSO. KB19 find that the adjusted framework leads to increased correlations of WPa with Niño-3.4 SST during periods of high ENSO variability in LENS, reinforcing the notion that WPa is a particularly useful predictor of El Niño events. From a brief examination of the RMSE with respect to the ENSO statistics in the six 23-yr samples, the results

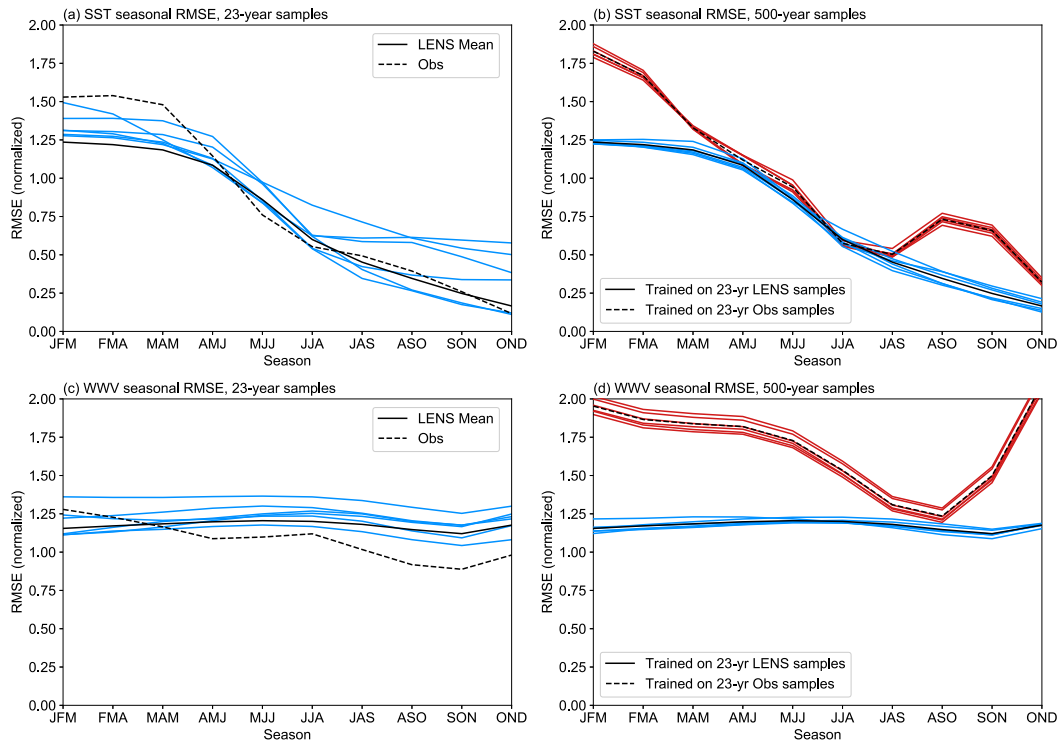


FIG. 2. (a),(c) As in Figs. 1a,c,e, but for SST and WWV, respectively. (b),(d) As in Figs. 1b,d,f, but for SST and WWV, respectively.

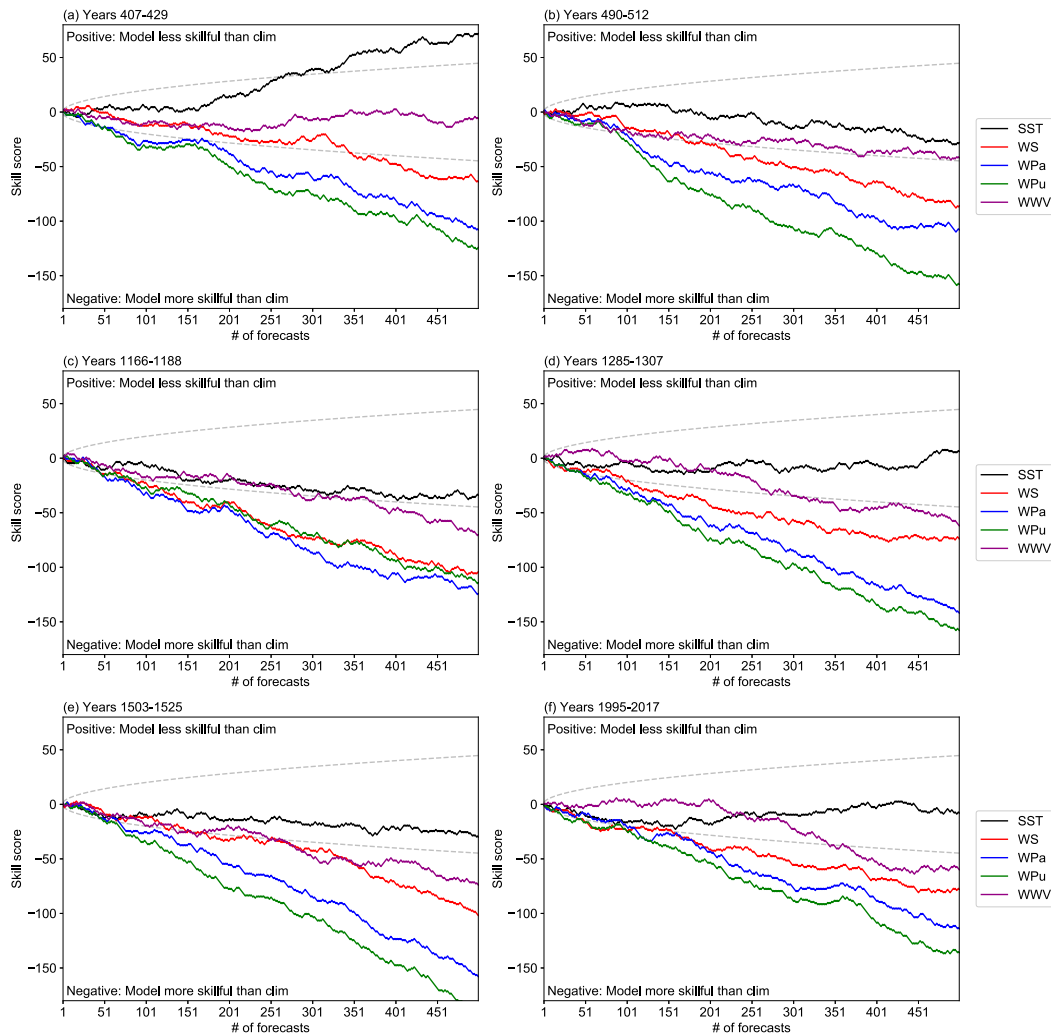


FIG. 3. Random walk skill test using climatological forecast as reference for 500-yr validation samples from LENS. Showing scores for JFM forecasts of NDJ SST trained over sample years (a) 407–429, (b) 490–512, (c) 1166–1188, (d) 1285–1307, (e) 1503–1525, and (f) 1995–2017. Gray curves represent the  $\pm 2\sqrt{N}$  significance threshold, where  $N$  is the number of forecasts. Positive values indicate the models do not have skill compared to a climatological forecast. Negative values indicate the model is more skillful than climatology.

appear to be consistent. For example, the samples 490–512 and 1995–2017 with standard deviations of  $1.268^{\circ}$  and  $1.145^{\circ}\text{C}$ , respectively, showed WPa as the lowest error predictor for the time periods JFM–MJJ and FMA–MJJ, respectively. Meanwhile, in years 1285–1307 and 1503–1525 with standard deviations  $0.642^{\circ}$  and  $0.644^{\circ}\text{C}$ , WPa only shows the lowest error for MAM in 1285–1307 and is nearly indistinguishable from WPu in 1503–1525. The higher standard deviations also correspond to a larger percentage of El Niño events, which are the event type that the adjusted wind power framework is intended to address.

The large uncertainty involved with 23-yr samples means that it is difficult to confidently determine if the differences in RMSE may simply be an artifact of that sample-dependent uncertainty or how the random walk scores would change when evaluating a different set of forecasts. This is still useful context when considering the observed results in the following section,

and the LENS analysis indicates some correspondence between the variability (number of ENSO events) and how the predictors perform in a shorter record that can help inform the conclusions drawn from observations. With this in mind, the observational results are presented in the following section and compared with the LENS analysis.

## 5. Observational analysis

### a. Comparison to LENS

The observational RMSE and anomaly correlation coefficients of the single predictors are shown in Fig. 6, plotted as functions of predictor season along with the NMME RMSE as a skill benchmark. Differences in the structure of lead-dependent errors can be seen when comparing the observed RMSE in Figs. 1 and 2 (black solid line, left column) to the



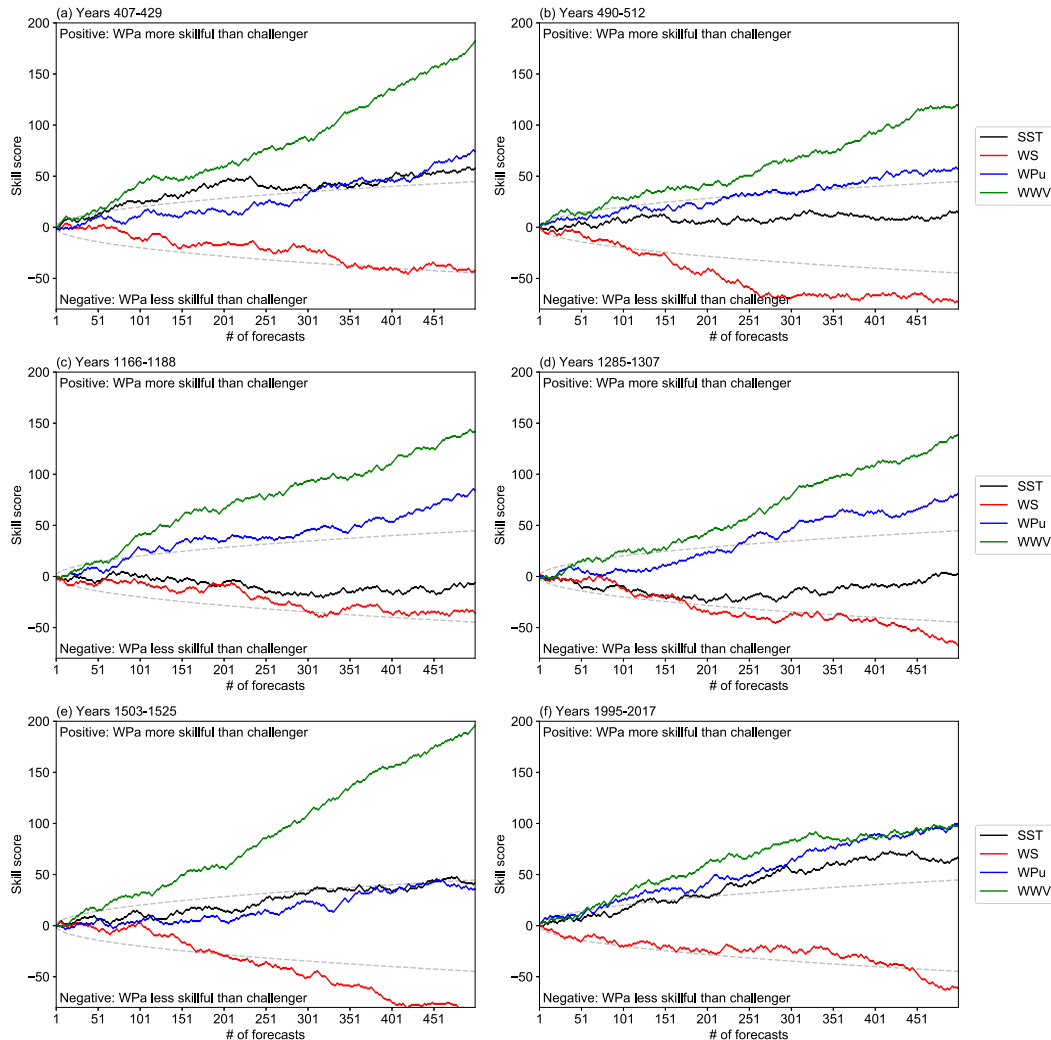


FIG. 4. Random walk test with WPa as the reference model for 500-yr validation samples in LENS. Scores of AMJ forecasts of NDJ SST are shown for same years as in Fig. 3.

LENS samples (blue) and comparing Figs. 6a and 6c. For observational forecasts initialized in JFM and FMA, the errors of all of the predictors are much higher in general than those of the LENS forecasts for those leads. The dropoff in error as lead decreases is subsequently much steeper than in LENS, especially for SST. That SST dominates from summer onward is somewhat expected as the ENSO anomaly growth associated with the Bjerknes positive feedbacks is strongest during the summer and fall (Eisenman et al. 2005). The differences between WPa and WPU are also far more distinct than in LENS, with observed WPa having consistently lower RMSE than WPU at every lead.

As seen in section 4, uncertainty is indeed high for this sample length. While it is possible that, given a much longer record, the advantages of WPa might fade out as they do in the LENS 500-yr samples, the opposite is equally possible, and the RMSE may converge in favor of WPa as the more useful framework for the observed climate system. There are a number

of key differences between the modeled and observed climate system that point to the latter. KB19 show that the lag correlations between WPa and WPU in the PI control do not differ to the same extent that they do in the observational analysis (also see Figs. 6b,d). They suggest that the adjusted framework has less impact in the PI control compared to observations due to biases in the climatological surface currents and wind stress, which are both excessively strong. Excessive tropical Pacific winds and their role in setting SST bias is a common feature of coupled models (Schneider 2002; Luo et al. 2005). Moreover, the LENS wind stress index and WWV correlation coefficients are considerably at odds with their observed counterparts (Fig. 6).

Another test is conducted to further exemplify the differences in how LENS represents the climate system in comparison to observations. In this test, the linear regression models are trained using the 23-yr observed dataset, but are verified using LENS data. In this analysis, the LENS data are treated as independent observations. RMSE is then computed for these

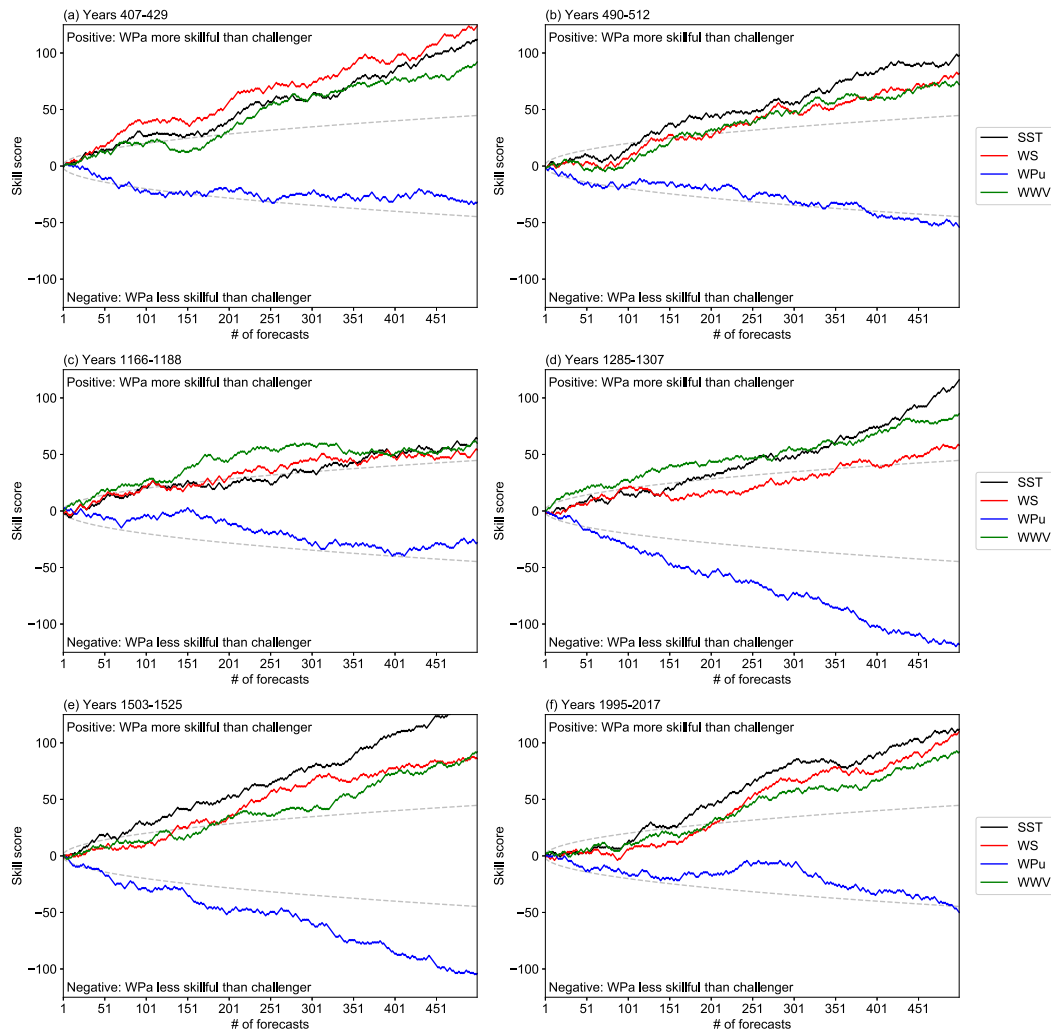


FIG. 5. As in Fig. 4, but for JFM forecasts of NDJ SST.

forecasts (red lines, Figs. 1 and 2, right column). The resulting error is much larger than that of the models trained using LENS data to forecast LENS data. The fact that the observationally trained models do so poorly when forecasting LENS data shows that the two climate systems do not represent each other well. In other words, the linear relationships associated with ENSO variability in LENS appear to have important differences from reality even when accounting for the influence of training period. These differences may indicate that in observations WPa has a more important impact on improving predictions of ENSO events than captured by LENS.

For comparison, the random walk skill test from section 4 is applied to the observations, using a climatological forecast as a reference (Fig. 7) and with the consideration that such a metric is more effective for much larger samples. Certain predictors quickly show indications of skill, for example WPu in FMA (Fig. 7b) and wind stress in MAM (Fig. 7c), though again, the sample size hinders the conclusiveness of the results. Interestingly, while the NMME outperforms both adjusted and unadjusted

wind power in the AMJ and JJA seasons in terms of RMSE and correlation (Fig. 6), it does not meet the random walk climatological forecast skill criterion at those leads. A closer analysis of the actual forecast values shows that NMME does not have large outlier forecasts but is also rarely the most accurate forecast among the competing models. In contrast, the regression models are generally more accurate for each forecast comparison but have a few outlier years in which their errors are large enough to inflate their RMSE. This illustrates how the random walk and RMSE provide different insights into the forecast behavior.

While it is difficult to infer how the predictors will compete in the long term in observations, we note some points of consistency that persist regardless of sample size in LENS to aid our assessment. KB19 observe that the predictor variables have higher correlations with SST in LENS than observations at lead times greater than 8 months. The comparison between observation and LENS in Fig. 6 complement this finding, with higher forecast correlations and lower RMSE for LENS-based regression

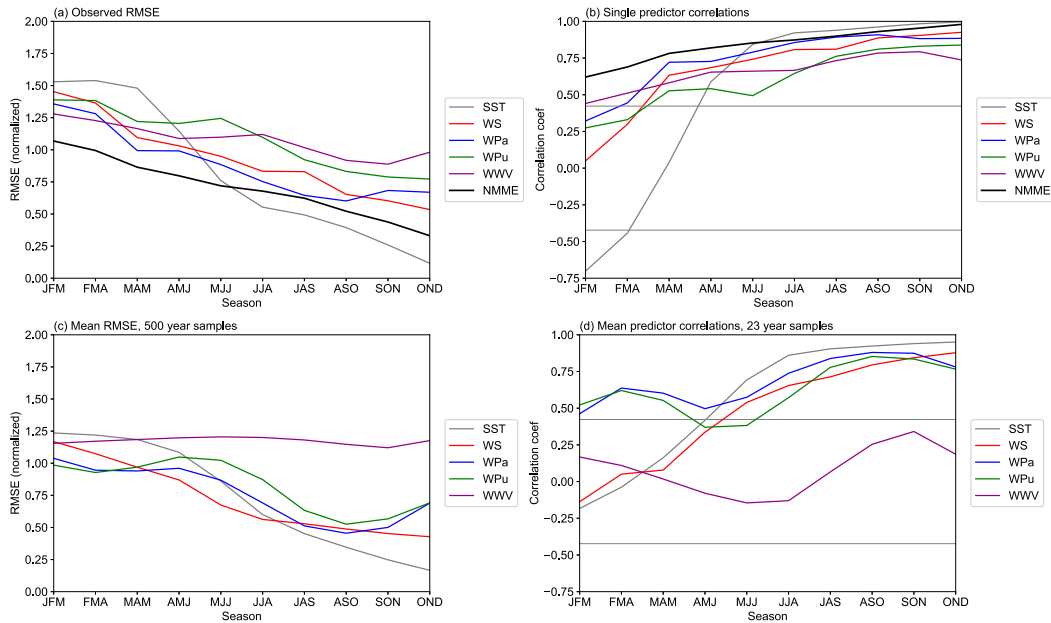


FIG. 6. (a) RMSE and (b) anomaly correlations of NDJ Niño-3.4 anomaly prediction initialized in each 3-month seasonal average denoted on the  $x$  axis for the single predictor regression models (see Table 1) and NMME, (c) the mean RMSE of the six sets of 500-yr forecast samples from the LENS analysis, and (d) mean correlations of the six sets of 23-yr LENS forecast samples.

models trained in JFM and FMA compared to observation-based models at the same leads. The random walk scores generally reflect the correlation and RMSE results in LENS, even when considering different sample sizes. In other words, taking each

of the metrics into consideration provides more robust insight into how each predictor will behave in the long run.

The random walk is also performed with observed WPa as the reference model (Fig. 8) for comparison with Figs. 4 and 5.

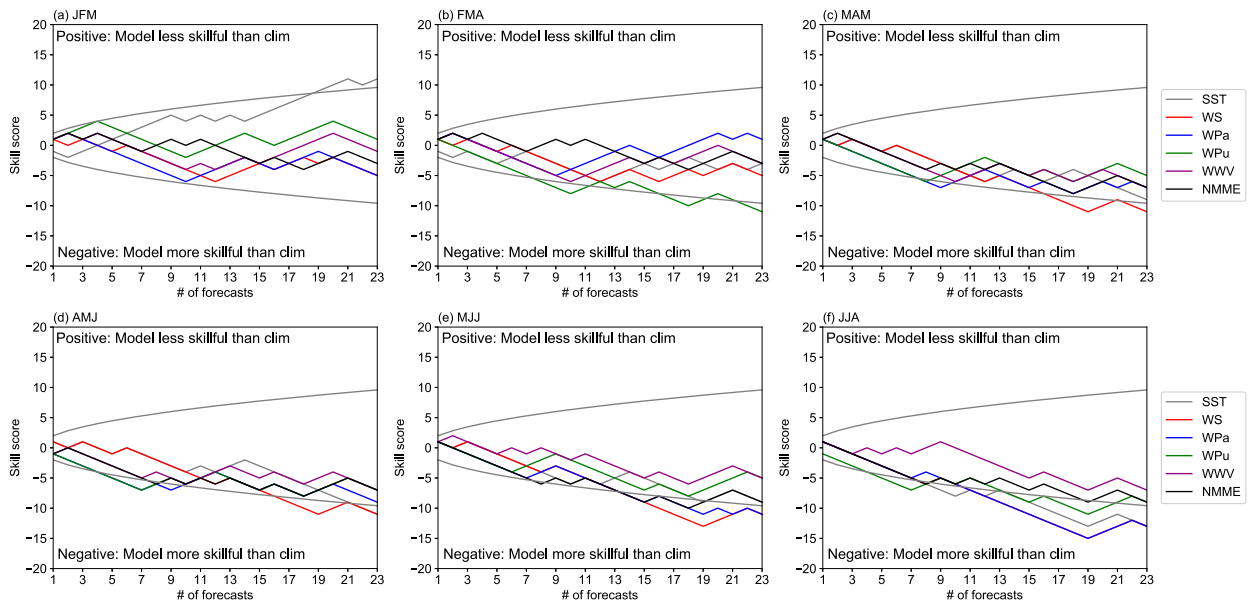


FIG. 7. Random walk test comparing climatology (zero anomaly forecast) skill to the regression models and NMME for initialization seasons (a) JFM, (b) FMA, (c) MAM, (d) AMJ, (e) MJJ, and (f) JJA. Each line represents the cumulative score of the reference model compared to its respective competitor (see legend). Gray curves represent the  $\pm 2\sqrt{N}$  significance threshold, where  $N$  is the number of forecasts. Positive values indicate the models do not have skill compared to a climatological forecast. Negative values indicate the model is more skillful than climatology.

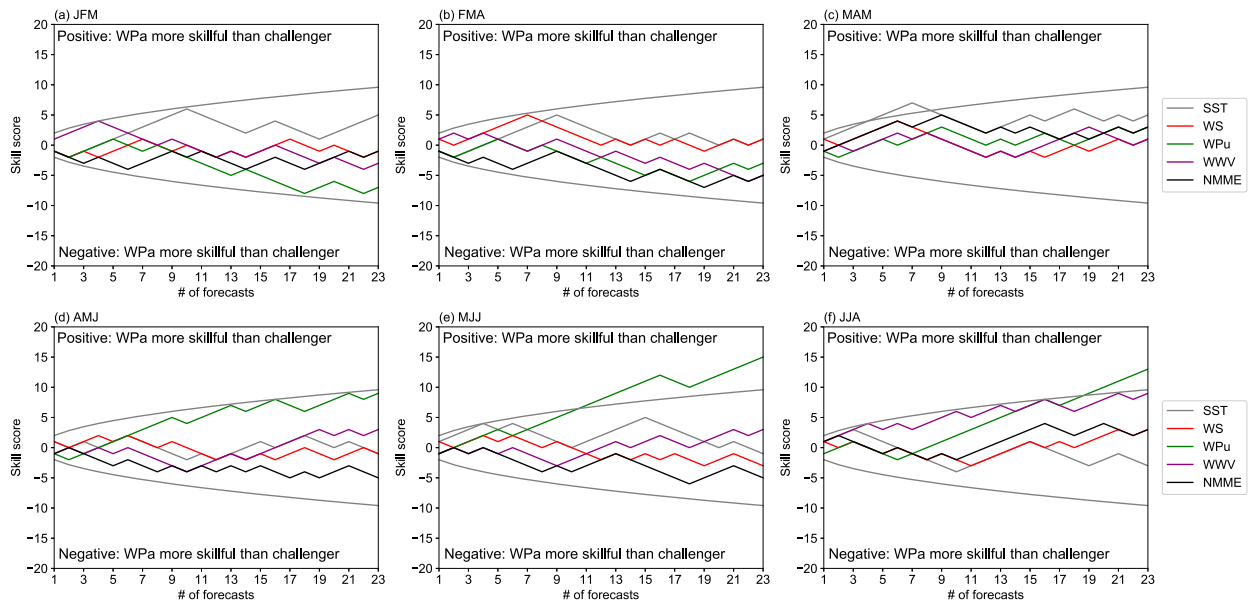


FIG. 8. As in Fig. 7, but comparing the skill of WPu as the reference model to the other models. Positive values indicate WPu has more skill, and negative values indicate WPu has less skill.

WPu is shown to be less skillful than WPa during MJJ through JJA in Fig. 8. The LENS analysis has shown that certain features of the predictors remain salient between 23- and 500-yr samples—for example, the reduced effectiveness of the adjusted framework in LENS. Given the higher correlations of WPu over WPa in the observations it is possible that the performance of WPu over WPa during MJJ–JJA in the observations may significantly persist with a longer record.

### b. Wind burst analysis

One of the highlighted advantages of the adjusted wind power framework is its ability to associate the direction of wind bursts with the correct dynamical SST response in the Niño-3.4 index (KB19). This is because WPu removes ambiguity in the wind power signal resulting from an EWE versus a WWE (KB19). This is particularly important during El Niño events where strong WWEs play a major role in the event development. The ability of adjusted wind power to more effectively capture the contributions of wind bursts motivates a brief analysis of adjusted wind power that is specifically associated with wind bursts.

The RMSE is computed and the random walk skill test is applied to the six wind burst–related indices (Fig. 9): the dynamical wind burst index from 5°S–5°N (WB5) and 15°S–15°N (WB15), the wind power due to wind bursts in the adjusted wind power framework from 5°S–5°N (WBPa5) and 15°S–15°N (WBPa15), and the wind power due to wind bursts without wind power adjustments (WBpu5 and WBpu15; see Table 1 for full summary of notation definitions). Figure 9a shows that the RMSE for the adjusted wind burst power in both the 5°S–5°N and 15°S–15°N regions is lower than the other regression models, although overall the errors for all indices are greater than 1. Interestingly, during JFM and FMA, the adjusted wind

burst power RMSE is also lower than that of the full wind power anomaly (WPu; model 3 in Table 1).

As in the previous skill assessments, the random walk score is computed for the wind burst forecasts using the climatological forecast as the reference model. The results of the skill test are only shown for the first three seasons (Figs. 9b–d) as early year wind bursts are often crucial ingredients in priming the system for El Niño events (Lengaigne et al. 2003; Fedorov 2002; Fedorov et al. 2014), and the adjusted wind power framework is shown to be most useful for improving representation of El Niño variability due to its accommodation of strong westerly wind events (KB19).

During FMA and MAM, the WBPa5 appears to have significant skill, but as in the previous section, it is well within the possibility that this is spurious in a robust sample. When taking both the random walks and the RMSE into account, it appears that even with temporal smoothing the wind burst indices remain too noisy to be systematically useful as predictors in a regression model at longer leads. However, the reduced RMSE in both WBPa indices suggests that the empirical adjusted framework does indeed serve its intended purpose to improve the representation of ENSO SST response to wind bursts.

## 6. Discussion and conclusions

The predictive skill of various ENSO precursors based on linear regression has been assessed with a variety of data sources and metrics. An initial analysis using the long LENS control illustrates the obstacles to assessing prediction skill for interannual events. After a sufficient number of forecasts, the RMSE begins to converge in the LENS analysis so that the error becomes independent of the training period used to train the regression models. While the random walk results show a greater dependence on

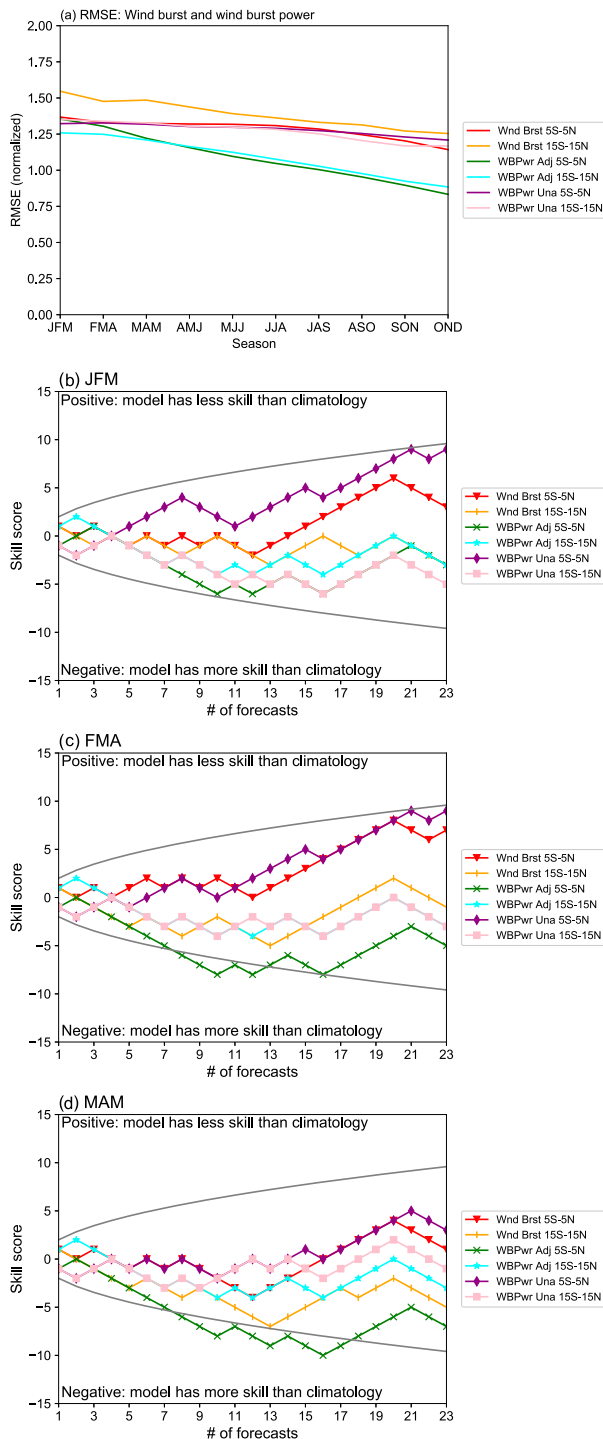


FIG. 9. (a) RMSE for the six integrated wind burst regression models, and random walk skill assessment with climatological zero anomaly reference model competing against the wind burst regression models (Table 1) for seasons (b) JFM, (c) FMA, and (d) MAM. In (b)–(d), the gray curves represent the  $\pm 2\sqrt{N}$  significance threshold, where  $N$  is the number of forecasts made. Positive values indicate the models do not have skill compared to a climatological forecast. Negative values indicate the model is more skillful than climatology.

verification sample, it takes a similarly long number of forecasts before the random walk scores establish persistent significance. For predictions on seasonal time scales, this involves around 100 years of data. If fewer years are included, the random walk scores have larger uncertainty, and significance of model skill is more sensitive to the years included in the sample.

While the LENS analysis points to several difficulties with an observational predictability assessment due to the shortness of the observational record, it also provides useful context for inferences drawn from observed results. While uncertainty is high, the differences from the LENS climate versus reality still manifest in the structure of the RMSE, correlations, and random walk scores. Even in the observationally consistent LENS samples, the wind power correlations are larger at the longest lead times compared to observations. When comparing the RMSE for the observations versus the short LENS samples, the observed errors for JFM and FMA are much higher (Fig. 1, left). The random walk test likewise shows that both wind power indices have skill in JFM and FMA in LENS but not observations.

In observations, WPA has lower RMSE than WPU for all lead seasons and among the lowest RMSE of all the regression models prior to MJJ. The advantages WPU appears to have over WPA at longer leads in LENS quickly disappear by the time forecasts are trained on MAM data, and from that lead onward, WPA significantly outperforms WPU in all of the sampled forecast records. The LENS analysis indicates that the effectiveness of WPA has some dependence on the number and strength of ENSO events, particularly El Niño. While there are not enough data to determine how the forecast errors will converge for the observed climate, in both LENS and observations, the adjusted framework appears to be most effective during periods of significant El Niño activity and larger ENSO variability.

The observed spring predictability barrier continues to be an obstacle, making ENSO in reality much harder to skillfully predict. The random walk skill test and the RMSE and correlation results at these leads (Figs. 6b,d) do not show the observational models as having skill from JFM to MAM. This is in contrast to LENS, where RMSE averaged lower than observed at the same leads and multiple predictors showed skill even during JFM. Even the NMME does not meet the zero-anomaly skill criterion prior to the spring predictability barrier. Whether this will change with a sufficiently long record is uncertain, but a more in-depth examination of the NMME forecasts themselves suggests that it is no exception to the influence of barrier. These results are similar to the comparison between the skill of the dynamic and statistical models of the IRI ENSO plume done by Tippett et al. (2012) and the regression models of DelSole and Tippett (2016), which found that the dynamical models, including the NMME, did not outperform the statistical models on a statistically significant level.

Simulating the nonlinear processes involved in ENSO is an inherently difficult task, illustrated by fact that the ensemble of dynamical models in NMME struggled to produce forecasts that are substantially more skillful than linear regression both on the subseasonal scales (DelSole and Tippett 2016) and on the seasonal scales examined here. Due to the difficulties with



representing the inherent uncertainty in the system that comes with the territory of nonlinear dynamics, there is increasing evidence that a move toward probabilistic forecasts over deterministic may be more practical (Fedorov et al. 2003; Kirtman 2003; Levine et al. 2016).

In other words, the challenges in ENSO forecasting are still substantial. However, even a baseline linear regression framework shows evidence that WPa is comparable in skill to the conventional predictors and even over the short observational record, shows more skill than Wpu for MJJ–JAS when considering all the skill metrics holistically. In addition, while the wind burst power regression models do not show skill based on the random walk criterion, this may also be an artifact of the short record. It is worth pointing out that the application of the adjusted framework nevertheless results in a reduction in RMSE. This finding further confirms the value in using the adjusted wind power framework to correct for ambiguity in the wind power–ENSO relation in the presence of strong wind bursts. As previous work has shown, state-dependent noise forcing is vital to the development of El Niño, especially major El Niño (Eisenman et al. 2005; Levine et al. 2016). Since the question of how noise forcing due to wind bursts is responsible for event development is still being discussed (Puy et al. 2017; Hu and Fedorov 2017), these results corroborate with KB19 that the adjusted wind power framework provides a viable additional perspective with which to evaluate the role of wind bursts.

Overall, the analysis here highlights the many difficulties of ENSO forecasting and, unsurprisingly, the empirical adjusted framework does not appear to be the magic bullet. However, even considering the differences between the LENS data and reality, the adjusted framework shows its advantages relative to the unadjusted wind power in both observations and state-of-the-art climate models. The adjusted wind power holds up comparably to the other predictors and demonstrates that correctly accounting for the wind bursts reduces error when representing ENSO.

*Acknowledgments.* This research is supported by grants from the National Science Foundation (NSF) (AGS-1613318 and AGS-1338427), National Oceanic and Atmospheric Administration (NOAA) (NA14OAR4310160), and National Aeronautics Space Administration (NASA) (NNX14AM19G). N.J.B. is supported by the Alfred P. Sloan Foundation as a Research Fellow. The authors thank Tim DelSole from George Mason University Department of Atmospheric, Oceanic, and Earth Sciences for his insight and assistance with applying the random walk skill assessment.

*Data availability statement.* The OSCAR data were developed by ESR and can be obtained from JPL Physical Oceanography DAAC from the dataset “OSCAR 1° ocean surface currents.” ECCO version 4 4-D potential temperature data are available through the NASA JPL ECCO Drive server, which requires user registration for an EarthData account to access and download the data. HadISST1 sea surface temperature data were provided by the U.K. Met Office Hadley Centre at <https://www.metoffice.gov.uk/hadobs/hadisst/as> either ASCII or NetCDF format. TropFlux surface zonal wind stress is accessible through the ESSO-INCOIS TropFlux product website <https://incois.gov.in/tropflux/overview.jsp> and requires registration with ESSO for access.

## REFERENCES

- Barnston, A. G., M. K. Tippett, M. L. L’Heureux, S. Li, and D. G. DeWitt, 2012: Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? *Bull. Amer. Meteor. Soc.*, **93**, 631–651, <https://doi.org/10.1175/BAMS-D-11-00111.1>.
- Becker, E., H. van den Dool, and Q. Zhang, 2014: Predictability and forecast skill in NMME. *J. Climate*, **27**, 5891–5906, <https://doi.org/10.1175/JCLI-D-13-00597.1>.
- Bjerknes, J., 1969: Atmospheric teleconnections from the equatorial Pacific. *Mon. Wea. Rev.*, **97**, 163–172, [https://doi.org/10.1175/1520-0493\(1969\)097<0163:ATFTEP>2.3.CO;2](https://doi.org/10.1175/1520-0493(1969)097<0163:ATFTEP>2.3.CO;2).
- Brown, J. N., and A. V. Fedorov, 2008: Mean energy balance in the tropical Pacific Ocean. *J. Mar. Res.*, **66** (1), 1–23, <https://doi.org/10.1357/002224008784815757>.
- , and —, 2010: How much energy is transferred from the winds to the thermocline on ENSO time scales? *J. Climate*, **23**, 1563–1580, <https://doi.org/10.1175/2009JCLI2914.1>.
- Burls, N. J., L. Muir, E. M. Vincent, and A. V. Fedorov, 2016: Extra-tropical origin of equatorial Pacific cold bias in climate models with links to cloud albedo. *Climate Dyn.*, **49**, 2093–2113, <https://doi.org/10.1007/s00382-016-3435-6>.
- DelSole, T., and M. K. Tippett, 2016: Forecast comparison based on random walks. *Mon. Wea. Rev.*, **144**, 615–626, <https://doi.org/10.1175/MWR-D-15-0218.1>.
- Dominiak, S., and P. Terray, 2005: Improvement of ENSO prediction using a linear regression model with a southern Indian Ocean sea surface temperature predictor. *Geophys. Res. Lett.*, **32**, L18702–, <https://doi.org/10.1029/2005GL023153>.
- Eisenman, I., L. Yu, and E. Tziperman, 2005: Westerly wind bursts: ENSO’s tail rather than the dog? *J. Climate*, **18**, 5224–5238, <https://doi.org/10.1175/JCLI3588.1>.
- ESR, 2007: OSCAR 1 Degree Ocean Surface Currents, version 1. Physical Oceanography Data Active Archive Center, accessed 21 February 2016, [https://podaac.jpl.nasa.gov/dataset/OSCAR\\_I4\\_OC\\_1deg](https://podaac.jpl.nasa.gov/dataset/OSCAR_I4_OC_1deg).
- Fedorov, A. V., 2002: The response of the coupled tropical ocean–atmosphere to westerly wind bursts. *Quart. J. Roy. Meteor. Soc.*, **128** (579), 1–23, <https://doi.org/10.1002/qj.200212857901>.
- , S. L. Harper, S. G. Philander, B. Winter, and A. T. Wittenberg, 2003: How predictable is El Niño? *Bull. Amer. Meteor. Soc.*, **84**, 911–920, <https://doi.org/10.1175/BAMS-84-7-911>.
- , S. Hu, M. Lengaigne, and E. Guilyardi, 2014: The impact of westerly wind bursts and ocean initial state on the development, and diversity of El Niño events. *Climate Dyn.*, **44**, 1381–1401, <https://doi.org/10.1007/s00382-014-2126-4>.
- Fenty, I., I. Fukumori, and O. Wang, 2015: ECCO-V4r3, version 4, release 3. ECCO drive, accessed 29 March 2019, <https://ecco.jpl.nasa.gov/drive/files>.
- Forget, G., J.-M. Campin, P. Heimbach, C. N. Hill, R. M. Ponte, and C. Wunsch, 2015: ECCO version 4: An integrated framework for non-linear inverse modeling and global ocean state estimation. *Geosci. Model Dev.*, **8**, 3071–3104, <https://doi.org/10.5194/gmd-8-3071-2015>.
- Fukumori, I., O. Wang, I. Fenty, G. Forget, P. Heimbach, and R. M. Ponte, 2017: ECCO version 4 release 3, <http://hdl.handle.net/1721.1/110380>, doi:1721.1/110380.
- Gent, P. R., and Coauthors, 2011: The Community Climate System Model version 4. *J. Climate*, **24**, 4973–4991, <https://doi.org/10.1175/2011JCLI4083.1>.
- Goddard, L., and S. G. Philander, 2000: The energetics of El Niño and La Niña. *J. Climate*, **13**, 1496–1516, [https://doi.org/10.1175/1520-0442\(2000\)013<1496:TEOENO>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<1496:TEOENO>2.0.CO;2).

- Hu, S., and A. V. Fedorov, 2017: The extreme El Niño of 2015–2016: The role of westerly and easterly wind bursts, and preconditioning by the failed 2014 event. *Climate Dyn.*, **52**, 7339–7357, <https://doi.org/10.1007/s00382-017-3531-2>.
- , —, M. Lengaigne, and E. Guilyardi, 2014: The impact of westerly wind bursts on the diversity and predictability of El Niño events: An ocean energetics perspective. *Geophys. Res. Lett.*, **41**, 4654–4663, <https://doi.org/10.1002/2014GL059573>.
- Kay, J. E., and Coauthors, 2015: The Community Earth System Model (CESM) Large Ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Meteor. Soc.*, **96**, 1333–1349, <https://doi.org/10.1175/BAMS-D-13-00255.1>.
- Kessler, W. S., 2002: Is ENSO a cycle or a series of events? *Geophys. Res. Lett.*, **29**, 2125, <https://doi.org/10.1029/2002GL015924>.
- Kirtman, B. P., 2003: The COLA anomaly coupled model: Ensemble ENSO prediction. *Mon. Wea. Rev.*, **131**, 2324–2341, [https://doi.org/10.1175/1520-0493\(2003\)131<2324:TCACME>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<2324:TCACME>2.0.CO;2).
- , and Coauthors, 2014: The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; Phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, <https://doi.org/10.1175/BAMS-D-12-00050.1>.
- Kodama, K., and N. J. Burls, 2019: An empirical adjusted ENSO ocean energetics framework based on observational wind power in the tropical Pacific. *Climate Dyn.*, **53**, 3271–3288, <https://doi.org/10.1007/s00382-019-04701-8>.
- Lagerloef, G. S. E., G. T. Mitchum, R. B. Lukas, and P. P. Niiler, 1999: Tropical Pacific near-surface currents estimated from altimeter, wind, and drifter data. *J. Geophys. Res.*, **104**, 23 313–23 326, <https://doi.org/10.1029/1999JC900197>.
- Lengaigne, M., J.-P. Boulanger, C. Menkes, G. Madec, P. Delecluse, E. Guilyardi, and J. Slingo, 2003: The March 1997 westerly wind event and the onset of the 1997/98 El Niño: Understanding the role of the atmospheric response. *J. Climate*, **16**, 3330–3343, [https://doi.org/10.1175/1520-0442\(2003\)016<3330:TMWWEA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<3330:TMWWEA>2.0.CO;2).
- , —, —, P. Delecluse, and J. Slingo, 2004: Westerly wind events in the tropical Pacific and their influence on the coupled ocean–atmosphere system: A review. *Earth's Climate*, C. Wang, S.-P. Xie, and J. A. Carton, Eds., Vol. 22, *The Ocean–Atmosphere Interaction*, Amer. Geophys. Union, 49–69.
- Levine, A., F.-F. Jin, and M. J. McPhaden, 2016: Extreme noise–extreme El Niño: How state-dependent noise forcing creates El Niño–La Niña asymmetry. *J. Climate*, **29**, 5483–5499, <https://doi.org/10.1175/JCLI-D-16-0091.1>.
- Li, G., and S.-P. Xie, 2014: Tropical biases in CMIP5 multimodel ensemble: The excessive equatorial Pacific cold tongue and double ITCZ problems. *J. Climate*, **27**, 1765–1780, <https://doi.org/10.1175/JCLI-D-13-00337.1>.
- Lorenz, E. N., 1955: Available potential energy and the maintenance of the general circulation. *Tellus*, **7**, 157–167, <https://doi.org/10.3402/tellusa.v7i2.8796>.
- Luo, J.-J., S. Masson, E. Roeckner, G. Madec, and T. Yamagata, 2005: Reducing climatology bias in an ocean–atmosphere CGCM with improved coupling physics. *J. Climate*, **18**, 2344–2360, <https://doi.org/10.1175/JCLI3404.1>.
- McPhaden, M. J., 2003: Tropical Pacific Ocean heat content variations and ENSO persistence barriers. *Geophys. Res. Lett.*, **30**, 1480, <https://doi.org/10.1029/2003GL016872>.
- Meinen, C. S., and M. J. McPhaden, 2000: Observations of warm water volume changes in the equatorial Pacific and their relationship to El Niño and La Niña. *J. Climate*, **13**, 3551–3559, [https://doi.org/10.1175/1520-0442\(2000\)013<3551:OOWWVC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<3551:OOWWVC>2.0.CO;2).
- Met Office Hadley Centre for Climate Change, 2003: HadISST, version 1. Met Office Hadley Centre for Climate Change, accessed 20 March 2016, <https://www.metoffice.gov.uk/hadobs/hadisst/>.
- Neelin, J. D., and H. A. Dijkstra, 1995: Ocean–atmosphere interaction and the tropical climatology. Part I: The dangers of flux correction. *J. Climate*, **8**, 1325–1342, [https://doi.org/10.1175/1520-0442\(1995\)008<1325:OAIATT>2.0.CO;2](https://doi.org/10.1175/1520-0442(1995)008<1325:OAIATT>2.0.CO;2).
- Newman, M., and P. D. Sardeshmukh, 2017: Are we near the predictability limit of tropical Indo-Pacific sea surface temperatures? *Geophys. Res. Lett.*, **44**, 8520–8529, <https://doi.org/10.1002/2017GL074088>.
- Oort, A. H., S. C. Ascher, S. Levitus, and J. P. Peixoto, 1989: New estimates of the available potential energy in the world ocean. *J. Geophys. Res.*, **94**, 3187–3200, <https://doi.org/10.1029/JC094iC03p03187>.
- Pegion, K. V., and C. Selman, 2017: Extratropical precursors of the El Niño–Southern Oscillation. *Climate Extremes Patterns and Mechanisms*, S. Y. S. Wang et al., Eds., John Wiley & Sons, 300–314.
- Philander, S. G., and A. V. Fedorov, 2003: Is El Niño sporadic or cyclic? *Annu. Rev. Earth Planet. Sci.*, **31**, 579–594, <https://doi.org/10.1146/annurev.earth.31.100901.141255>.
- Praveen Kumar, B., 2013: TropFlux Zonal Wind Stress, version 1. ESSO-Indian National Centre for Ocean Information Services, accessed 6 March 2016, <https://incois.gov.in/tropflux/overview.jsp>.
- , J. Vialard, M. Lengaigne, V. S. N. Murty, M. J. McPhaden, M. F. Cronin, F. Pinsard, and K. Gopala Reddy, 2012: TropFlux wind stresses over the tropical oceans: Evaluation and comparison with other products. *Climate Dyn.*, **40**, 2049–2071, <https://doi.org/10.1007/s00382-012-1455-4>.
- Puy, M., J. Vialard, M. Lengaigne, and E. Guilyardi, 2016: Modulation of equatorial Pacific westerly/easterly wind events by the Madden–Julian oscillation and convectively-coupled Rossby waves. *Climate Dyn.*, **46**, 2155–2178, <https://doi.org/10.1007/s00382-015-2695-x>.
- , and Coauthors, 2017: Influence of westerly wind events stochasticity on El Niño amplitude: The case of 2014 vs. 2015. *Climate Dyn.*, **52**, 7435–7454, <https://doi.org/10.1007/s00382-017-3938-9>.
- 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. Geophys. Res.*, **108**, 4407–, <https://doi.org/10.1029/2002JD002670>.
- Schneider, E. K., 2002: Understanding differences between the equatorial Pacific as simulated by two coupled GCMs. *J. Climate*, **15**, 449–469, [https://doi.org/10.1175/1520-0442\(2002\)015<0449:UDBTEP>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<0449:UDBTEP>2.0.CO;2).
- Seiki, A., and Y. N. Takayabu, 2007a: Westerly wind bursts and their relationship with intraseasonal variations and ENSO. Part I: Statistics. *Mon. Wea. Rev.*, **135**, 3325–3345, <https://doi.org/10.1175/MWR3477.1>.
- , and —, 2007b: Westerly wind bursts and their relationship with intraseasonal variations and ENSO. Part II: Energetics over the western and central Pacific. *Mon. Wea. Rev.*, **135**, 3346–3361, <https://doi.org/10.1175/MWR3503.1>.
- Tippett, M. K., A. G. Barnston, and S. Li, 2012: Performance of recent multimodel ENSO forecasts. *J. Appl. Meteor. Climatol.*, **51**, 637–654, <https://doi.org/10.1175/JAMC-D-11-093.1>.
- Torrence, C., and P. J. Webster, 1998: The annual cycle of persistence in the El Niño/Southern Oscillation. *Quart. J. Roy. Meteor. Soc.*, **124**, 1985–2004, <https://doi.org/10.1002/qj.49712455010>.