

ENSO Model Validation Using Wavelet Probability Analysis

SAMANTHA STEVENSON AND BAYLOR FOX-KEMPER

*Department of Atmospheric and Oceanic Sciences and Cooperative Institute for Research in the Environmental Sciences,
University of Colorado, Boulder, Colorado*

MARKUS JOCHUM

National Center for Atmospheric Research, Boulder, Colorado

BALAJI RAJAGOPALAN

Department of Civil, Environmental, and Architectural Engineering, University of Colorado, Boulder, Colorado

STEPHEN G. YEAGER

National Center for Atmospheric Research, Boulder, Colorado

(Manuscript received 5 January 2010, in final form 1 June 2010)

ABSTRACT

A new method to quantify changes in El Niño–Southern Oscillation (ENSO) variability is presented, using the overlap between probability distributions of the wavelet spectrum as measured by the wavelet probability index (WPI). Examples are provided using long integrations of three coupled climate models. When subsets of Niño-3.4 time series are compared, the width of the confidence interval on WPI has an exponential dependence on the length of the subset used, with a statistically identical slope for all three models. This exponential relationship describes the rate at which the system converges toward equilibrium and may be used to determine the necessary simulation length for robust statistics. For the three models tested, a minimum of 250 model years is required to obtain 90% convergence for Niño-3.4, longer than typical Intergovernmental Panel on Climate Change (IPCC) simulations. Applying the same decay relationship to observational data indicates that measuring ENSO variability with 90% confidence requires approximately 240 years of observations, which is substantially longer than the modern SST record. Applying hypothesis testing techniques to the WPI distributions from model subsets and from comparisons of model subsets to the historical Niño-3.4 index then allows statistically robust comparisons of relative model agreement with appropriate confidence levels given the length of the data record and model simulation.

1. Introduction

Predicting changes to the El Niño–Southern Oscillation (ENSO) has important societal implications, including drought management in Australia and the American Southwest (Seager 2007; Trenberth et al. 1998; Ropelewski and Halpert 1996; Nicholls et al. 1996). However, accurate ENSO simulation is limited by the short extent of observations in the tropical Pacific (Guilyardi et al. 2009), as well as model errors (Guilyardi 2006; AchutaRao and Sperber 2006; van Oldenborgh et al. 2005). Here, the

goal is to provide a tool that can be used to test models' performance relative to observations and to other models in a statistically reliable manner, which will both guide improvements to model performance and shed light on the impacts of external forcing on ENSO.

Both modeling (Wittenberg 2009) and observational (Zhang and McPhaden 2006) studies agree that modulations in ENSO dynamics occur on long time scales, meaning that longer records are necessary to capture the full behavior of the system. Paleoclimatic proxy reconstructions from sources such as fossil coral records are often used to extend the temporal baseline, but their use may be complicated by the need to account for the physical location of the proxy record, sampling variability within the record, or small-scale effects leading to

Corresponding author address: Samantha Stevenson, CIRES, 216 UCB, Boulder, CO 80303.
E-mail: samantha.stevenson@colorado.edu

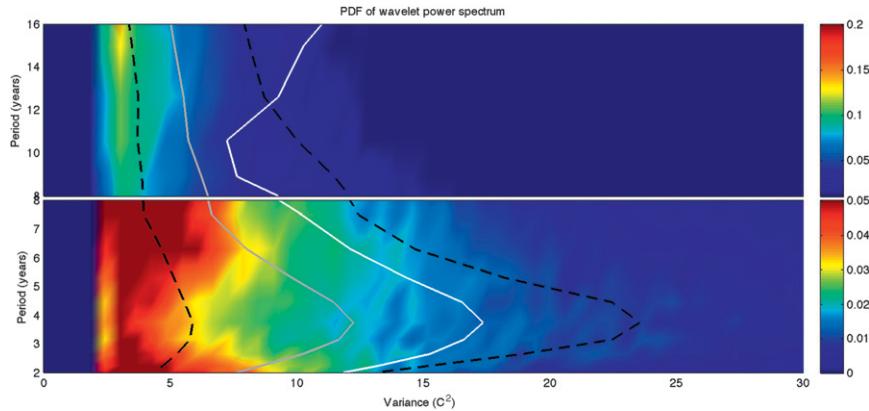


FIG. 1. Probability distribution functions for CCSMcontrol Niño-3.4 wavelet power. The gray line represents the median value for the model simulation, while the white line is the mean value generated using the CORE hindcast. Dashed black lines correspond to the 25th and 75th percentile values for the model simulation (interquartile range).

artificial fluctuations in the signal (McGregor and Gagan 2004; Brown et al. 2008). This implies that long coupled climate model integrations are the best option for generating long-term, spatially complete ENSO records (Wittenberg 2009).

Coupled models suffer from some biases (Capotondi et al. 2006; AchutaRao and Sperber 2006; van Oldenborgh et al. 2005), but the present generation of coupled models shows increased accuracy (Guilyardi et al. 2009). In particular, the updated version of National Center for Atmospheric Research (NCAR)'s Community Climate System Model (CCSM3.5; Neale et al. 2008) is much improved relative to the IPCC Fourth Assessment Report (AR4)-class climate models at both fine and coarse resolutions (Jochum et al. 2009). Although biases remain, this version of CCSM performs as well as the current generation of coupled models at a relatively low computational cost, and it has therefore been used as the primary coupled model for this study.

This paper uses long integrations of the coarse-resolution CCSM3.5 to illustrate a new, wavelet-based probabilistic model validation method, capable of dealing with skewed and temporally variable distributions and useful both for ENSO and for other climate indices. Traditional tests (χ^2 or Kolmogorov-Smirnov) are not suitable for non-Gaussian distributions; however, wavelet probability analysis can provide quantitative statistical measures even for highly nonnormal distributions of spectral power. This method is extremely versatile: it may be used to predict the necessary length for a model simulation (section 2a), to quantify agreement between a model and observations (section 2b), or to examine the relative performance of multiple models compared to observations (section 2c).

2. Wavelet probability analysis

This method relies on the probability distribution function (PDF) of wavelet power. Here, Niño-3.4 SST from a 1200-yr integration of the CCSM3.5 (hereafter CCSMcontrol) forms the primary dataset. CCSMcontrol is configured as in Jochum et al. (2010) and validated against SST from the Common Ocean-Ice Reference Experiment (CORE) of Large and Yeager (2008; hereafter the CORE hindcast), covering the period from 1949 to 2003 and chosen for convenience. The Large and Yeager (2008) Niño-3.4 is highly correlated (≥ 0.95) with other data products over this time window. However, we note that other data products can easily be used as well.

Figure 1 shows the PDF of wavelet power, generated using the wavelet toolkit of Torrence and Compo (1998). This is equivalent to plotting a histogram of wavelet power at each period; the variance is shown on the horizontal axis, and the wavelet period on the vertical, with probability densities shown in color. The median and 25th/75th percentile values (interquartile range) of wavelet power for each period are overplotted as the gray line and pair of black dashed lines, respectively, and the median wavelet PDF for the CORE hindcast is shown for comparison as the white solid line. Looking at the positions of the two median curves, one sees that the CORE hindcast falls well within the model's interquartile range for all wavelet periods below 10 yr: in other words, the range of short-period variability shown by CCSM is consistent with the behavior of the recent observational record. Some offsets do remain at long periods, most likely due to errors in CCSM3.5's representation of ENSO or other decadal variability (e.g., the Pacific Decadal Oscillation) but with some potential

contribution from undersampling the true range of ENSO dynamics. Understanding how the contribution from natural variability compares to real errors in model physics is the purpose of wavelet probability analysis.

Let $f_1(\sigma, \nu)$ and $f_2(\sigma, \nu)$ be two PDFs of wavelet power σ at frequency ν . Then the joint PDF $F(\sigma, \nu)$ is the probability that a given level of wavelet power is observed in both datasets at frequency ν , and the integral of $F(\sigma, \nu)$ is the overlap between the two. We refer to the latter quantity as the wavelet probability index, or WPI:¹

$$\text{WPI}(\nu) = \int_0^\infty F(\sigma, \nu) d\sigma = \int_0^\infty f_1(\sigma, \nu) f_2(\sigma, \nu) d\sigma. \quad (1)$$

The two wavelet PDFs f_1 and f_2 have been assumed to be independent, which allows us to interpret F as the product of the two distributions: $F(\sigma, \nu) = f_1(\sigma, \nu) f_2(\sigma, \nu)$. By definition, WPI lies between 0 and 1 and measures statistical agreement between time series. WPI can be used to measure internal variability (self-overlap; section 2a), or to quantify agreement between records: for example, model simulations, or a model versus data (sections 2b and 2c).

The choice of wavelet basis has a minor effect on the results; here, we use the Morlet wavelet, of degree 6 (Daubechies 1990; Torrence and Compo 1998):

$$\Psi(\eta) = \pi^{-1/4} e^{-i\omega_0\eta} e^{-\eta^2/2}, \quad (2)$$

where η is the nondimensionalized time parameter and ω_0 the nondimensionalized frequency. We note that the known bias in the wavelet spectrum (Liu et al. 2007) affects the numerical outcome of some of the analyses below, but does not affect the results of hypothesis tests or the major conclusions of this work. The same is true for the choice of wavelet basis; the Morlet wavelet is used throughout this analysis, but choosing a real-valued rather than a complex wavelet does affect the outcome to a minor extent. A full analysis of the wavelet basis effect is beyond the scope of this introductory paper; further analysis will be forthcoming.

The relevant steps for this analysis are as follows:

- (i) Choose the two time series to compare (e.g., subsets of a model versus entire simulation, subsets of a model versus data).
- (ii) Create a time series for the region of interest.
- (iii) Perform a wavelet analysis on the two time series.
- (iv) Compute the probability distribution function of the wavelet power, for all time series of interest.

- (v) Calculate the WPI according to Eq. (1).
- (vi) Subsample the data to find the WPI distribution due to internal variability.² Confidence intervals at the $1 - \alpha$ significance level may then be obtained using the $\alpha/2$ and $1 - \alpha/2$ percentiles of the WPI distribution.

Steps (i)–(vi) yield a quantitative measure of spectral agreement between time series, accompanied by well-defined significance levels. In this sense, the wavelet probability method is a natural extension of qualitative estimates of model–observed ENSO agreement, such as Neale et al. (2008) or van Oldenborgh et al. (2005). Three examples of using wavelet probability analysis are presented here using the Niño-3.4 wavelet PDF: a self-overlap calculation (section 2a), a data–model comparison (section 2b), and a demonstration of the use of hypothesis testing to accept or reject a climate model based on ENSO variability (section 2c) are shown. A suite of Matlab codes developed for this purpose has been used in all three calculations, and is available for download as both command-line codes and a graphical user interface. (More information is available online at <http://atoc.colorado.edu/~slsteven/wpi/>.)

a. Self-overlap

The first goal of this analysis is to quantify the internal variability within an ENSO record, which is expected to be quite large. Measuring the WPI distribution generated from subsamples of an (ideally equilibrated, constantly forced) control time series yields the expected degree of self-agreement as a function of time series length. The 90% confidence interval is then the distance between the 5th and 95th percentiles of the resulting WPI distribution. As the simulation length and subsample length increase, the 90% WPI confidence interval becomes narrower, which allows a prediction of the length needed for a given level of spectral convergence. The self-overlap dependence on subinterval length is shown in Figs. 2a,b, for the CCSM3.5 and the CM2.1, respectively. For each of those models, example subintervals of length 50, 100, and 200 yr have been used to generate 90% WPI confidence intervals, plotted as the black, blue, and red lines in each panel.

Subintervals of a time series are by definition drawn from the same distribution. Therefore, the upper limit of the WPI distribution should approach 1 for long subintervals, a behavior that is observed in Fig. 2. Ultimately, one expects a self-overlap WPI value of 1 for very long comparison windows, but it is the width of the

¹ One can also integrate WPI over frequency to obtain a single value, but this destroys useful information.

² This analysis uses only nonoverlapping subsamples, to eliminate any possibility of introducing effects due to autocorrelation.

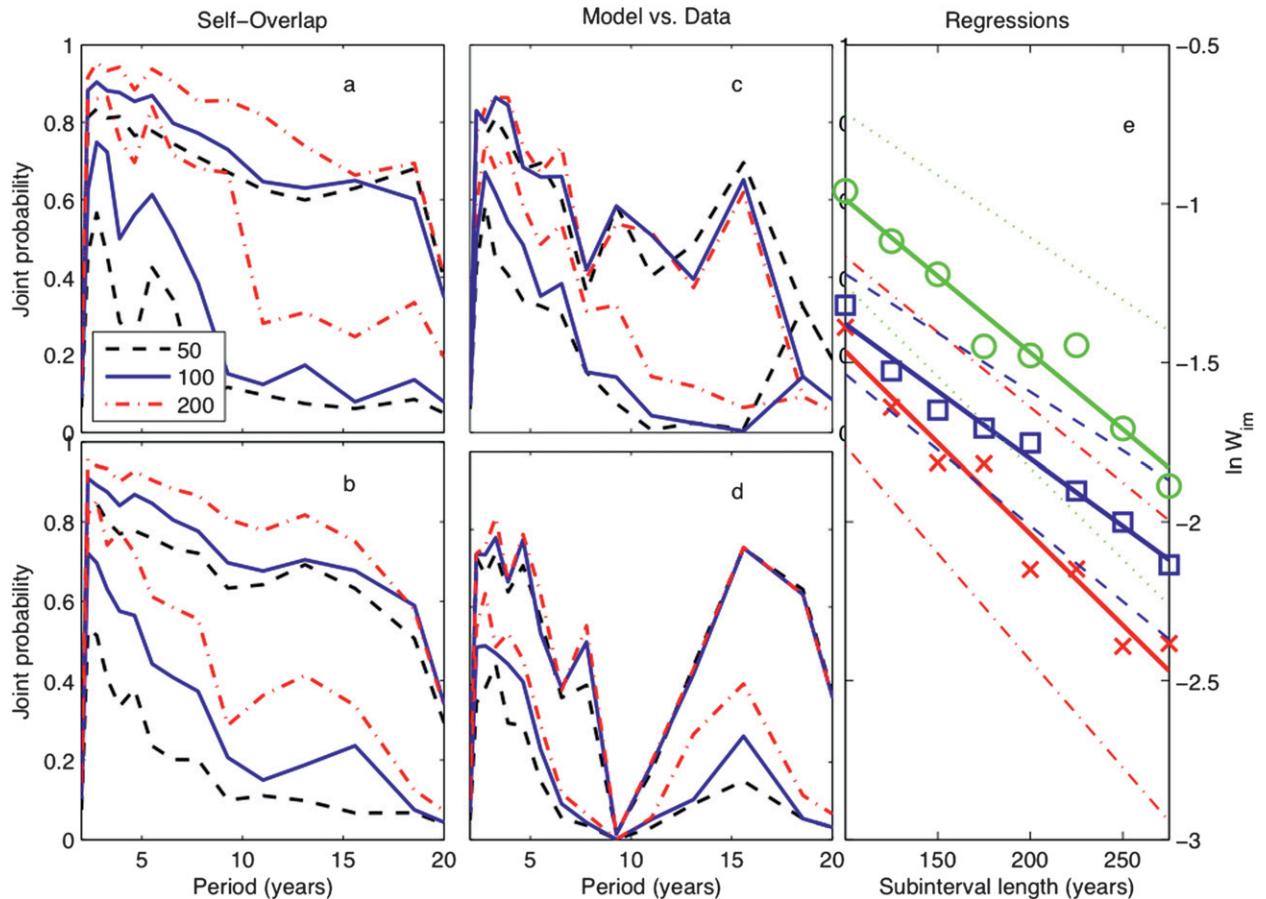


FIG. 2. (a),(b) The 90% confidence interval on WPI distributions for self-overlap calculations (CCSMcontrol and CM2.1, respectively). (c),(d) As in (a),(b), but for model–data WPI distributions. (a)–(d) Higher values of WPI indicate better agreement, ranging from 0 to 1. (e) The regression of 90% confidence interval widths against subinterval length, for self-overlap calculations. CCSMcontrol (NCAR CCSM3.5) data appear as red \times 's, GFDL CM2.1 as blue squares, and IPSL CM4 as green circles.

confidence interval that quantifies the confidence one has in the wavelet spectrum estimate. Thus, it is proposed here that a good measure of confidence in the convergence of the WPI is the width of the 90% confidence interval W_{im} .

For ENSO, we have assumed that the most relevant wavelet periods are those between 2 and 6 years and have performed a regression on the average WPI confidence interval width W_{im} against the model subinterval length L , averaged over those wavelet periods. An exponential relationship provides a good fit:

$$\ln W_{im} = \beta_0 + \beta_1 L. \quad (3)$$

Values for the slope and intercept coefficients β_1 and β_0 are given in Table 1. The regression lines are shown graphically in Fig. 2e; CCSM3.5 is shown in red, CM2.1 in blue, and CM4 in green, with the corresponding 90% confidence interval for each fit given as dot–dashed lines of the corresponding color.

The slope of Eq. (3) is demonstrated to be quite robust. Tests of the relationship have been performed with a 2000-yr integration of the Geophysical Fluid Dynamics Laboratory (GFDL) CM2.1 (Wittenberg et al. 2006; Wittenberg 2009), and an 1155-yr preindustrial control simulation of the L'Institut Pierre-Simon Laplace Coupled Model, version 4 (IPSL CM4). For all three of these models, the slope is statistically indistinguishable. However, the same cannot be said of the regression intercept; this is specific to the individual model. In some sense, the regression intercept may be said to inform the degree to which self-overlap is built in to the coupled model in question, since the total length of the simulation does not greatly affect the value of the intercept.

This exponential convergence allows the determination of the necessary simulation length for any given coupled model to settle down in terms of its ENSO statistics. If Niño-3.4 is the index of choice, then to sample 90% of the true ENSO variability in CCSM3.5, find the

TABLE 1. Dependence of the 90% WPI confidence interval width on model subinterval length, from confidence intervals averaged over the 2–6-yr band. The $\Delta\beta_0$ and $\Delta\beta_1$ refer to the bounds of the 90% confidence intervals on those coefficients; L_{\min} is the minimum length required to achieve 90% convergence in Niño-3.4 statistics for each model and ΔL_{\min} is the range between the upper and lower limits on L_{\min} .

Simulation	β_0	β_1	$\Delta\beta_0$	$\Delta\beta_1$	L_{\min}	ΔL_{\min}
CCSMcontrol	-0.891	-0.0057	-1.09 to -0.694	-0.0067 to -0.0047	247	180–342
GFDL CM2.1	-0.956	-0.0042	-1.06 to -0.852	-0.048 to -0.0037	320	258–391
IPSL CM4	-0.504	-0.0048	-0.683 to -0.324	-0.0057 to -0.0039	374	283–507

value of L in Eq. (3) where $W_{\text{im}} = 0.1$ ($\ln W_{\text{im}} = -2.3$). This simulation length is between 247 and 374 yr for all models (see Table 1). Thus, 250 yr is a reasonable initial simulation length for quantification of ENSO statistics. However, it may take longer than 250 yr for the model to equilibrate; note that each of the three models tested have a different value of β_0 in Eq. (3). It is recommended that a run of 250 yr (or shorter) be used to determine β_0 , which will allow for a more accurate estimate of the simulation length needed to reach $W_{\text{im}} = 0.1$. The online tools at <http://atoc.colorado.edu/~slsteven/wpi/> allow automated calculation of the necessary simulation length.

Before comparing models to data in the next section, it is interesting to use the convergence rate of W_{im} in models to estimate the observational record length with $W_{\text{im}} = 0.1$, just as if the observations were a preliminary (short) model simulation. Doing so with the Hadley Centre Global Sea Ice and Sea Surface Temperature (HadISST) dataset, for instance, yields $\beta_0 = -0.927$. Then using the β_1 value from CCSM3.5 indicates that 90% confidence in convergence of the Niño-3.4 data will require an observational record 240 yr long. In other words, both twentieth-century climate simulations and the observational record are too short to sample the full range of ENSO.

Ideally, the modern record would be used to generate its own WPI regression, from which upper and lower bounds on the needed observation length could be calculated. This is precluded by its short length, but a rough estimate may be generated using the 90% confidence interval width on β_0 from the CCSMcontrol regression. When this is done, we find that the minimum observation length needed for 90% convergence is 175 yr, and the maximum is 334. Thus, as many as 334 yr of observations may be required to uncover the true picture of ENSO.

b. Validation against data

Even if the model and data records are too short to reach $W_{\text{im}} = 0.1$ as above, one can use them to formulate quantitative estimates of model/data agreement and agreement uncertainty. Estimating the expected agreement between distinct time series (e.g., a model and observations) as a function of their lengths is another use of

wavelet probability analysis, which helps deal with validation against an observational record too short to sample the full range of variability. The method follows section 2a, except that now the WPI values are derived from comparing the entirety of the CORE hindcast to subintervals of various lengths taken from the model integrations.

Figures 2c and 2d show model–data agreement for CCSMcontrol and CM2.1, using the same plotting conventions as their self-overlap counterparts in Figs. 2a,b. Below 5 yr, the model–data WPI ranges from 40% to 80%, and it is much lower from 8 to 12 years. CM2.1’s lower agreement with CORE relative to CCSM is consistent with CM2.1’s known overestimate of ENSO amplitude (Wittenberg et al. 2006). However, the upper bound of WPI never approaches 1 for either model–data comparison; this implies that there is a real offset, or equivalently that both models differ from CORE in a statistically significant way. It should also be noted that the dependence of the model–data WPI confidence interval on simulation length is exponential (not shown), as was the model self-overlap of section 2a.

A visual examination of Fig. 2a versus Fig. 2c and Fig. 2b versus Fig. 2d allows for a qualitative assessment of how CORE/model and model–model agreement compare. For 50-yr model subintervals (black dashed lines in Figs. 2a–d), the data/model and model self-overlap confidence interval pairs overlap for both the CCSM3.5 and CM2.1; the internal variability present in both models spans the range of WPI values observed for model–data comparison. In other words, the comparison of Fig. 2a to Fig. 2c and Fig. 2b to Fig. 2d implies that 50-yr subintervals of both CCSM3.5 and CM2.1 are indistinguishable from the data. In contrast, for intervals longer than 200 years, self-overlap and model–data WPI confidence intervals do not overlap; simulations (or data records) longer than 200–300 yr are needed to identify real offsets, a result which will be made more precise in the next section.³

³ This work focuses on stable control simulations. In transient forcing scenarios, longer simulations may not be applicable, thus simulation ensembles can be used to reduce the uncertainty to an acceptable level. Extending the WPI calculations to account for mutual information between ensemble members is presently underway.

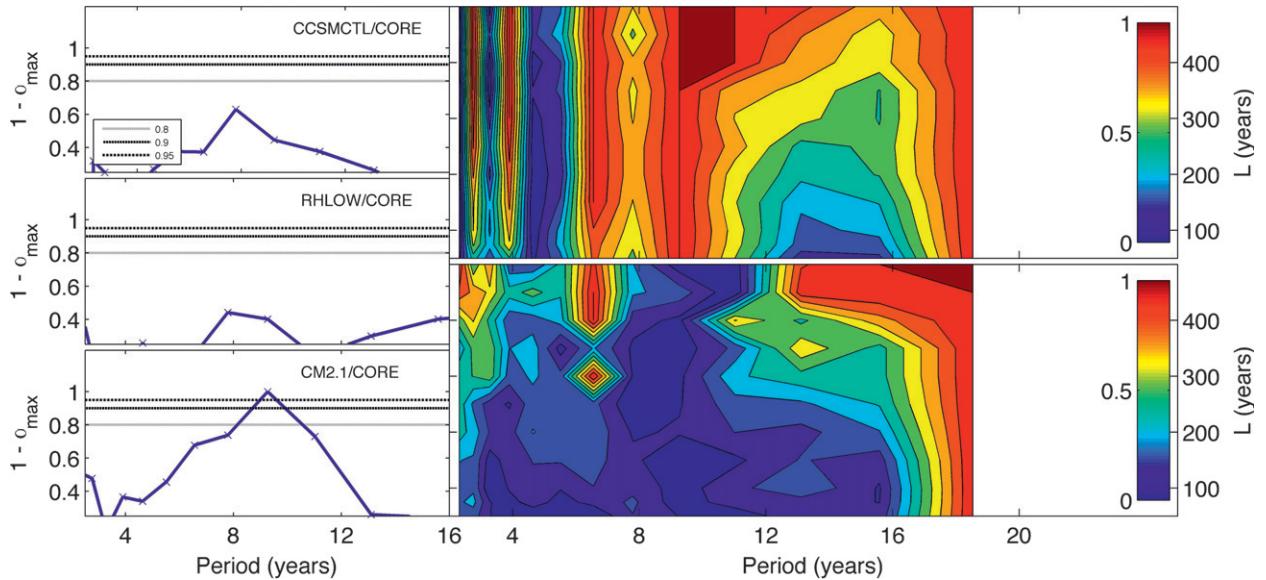


FIG. 3. Results of hypothesis testing procedure. (left) Validation of (top) CCSMcontrol, (middle) RHLOW, and (bottom) CM2.1 against the CORE hindcast. (right) Comparison of (top) CCSMcontrol vs CM2.1 and (bottom) CCSMcontrol vs RHLOW. In all panels, confidence levels plotted range from 0 (agreement) to 1 (disagreement). Note: see text for description of model simulations.

In general, rather than aiming to get a climate model to agree as closely as possible with observations, the most appropriate goal might be for the model to lie inside the range of acceptable agreement.

c. Empirical hypothesis testing

The power of wavelet probability analysis lies in the specification of the significance level at which two time series disagree, which is done through hypothesis testing on WPI distributions (i.e., Fig. 2). Empirical methods are used, since preliminary analysis showed that the WPI distributions of sections 2a and 2b can be highly nonnormal (see Fig. 1). This makes the use of traditional tests, such as the Student's t test or the F test, impossible (Montgomery and Runger 2007). In this case, even the application of the nonparametric Kolmogorov–Smirnov (KS) test is questionable because samples drawn from different distributions cannot be dismissed without a priori knowledge of the correct distribution.

The procedure for conducting a hypothesis test on wavelet probability data includes the following steps:

- 1) Determine the type of test to perform: model–model or model–data.
- 2) Create the appropriate WPI distributions from subsets of the input time series. For a model–data comparison, model self-overlap (section 2a) will be tested against the model–data WPI distribution (section 2b). For a model–model comparison, the two model–data distributions will be compared.

- 3) To determine whether two distributions differ at significance level α , compute the $\alpha/2$ to $1 - \alpha/2$ confidence intervals on the two WPI distributions. If these intervals overlap, the distributions are equivalent; otherwise, they differ.
- 4) To determine the level of confidence one may have in differences between distributions, repeat step c at many values of α . The largest α for which the confidence intervals overlap is then equivalent to the smallest significance level at which the distributions differ. Where $\alpha_{\max} \leq 0.1$ ($1 - \alpha_{\max} \geq 0.9$), for example, the null would be rejected at the 90% level. In the limit of identical distributions, α_{\max} (minimum significance) approaches 1(0); when there is no overlap, α_{\max} (minimum significance) approaches 0 (1).

The end result of steps 1–3 is a map of locations in parameter space where the two time series differ at confidence level α . This can be depicted as a graph of the test result [0 (agree) or 1 (disagree)] versus wavelet period. Step 4 provides the significance level at which the time series differ at every wavelet period; now everywhere the time series differ at or above 90%, for example, the plotted values will meet or exceed 0.9. For consistency with statistical conventions, the plotting convention for the WPA hypothesis tests in Fig. 3 differs from that in Figs. 2a–d, in which lower values indicated less agreement; in Fig. 3, lower values indicate greater agreement. The left-hand panels of Fig. 3 show examples of hypothesis test results for the case where a single

model subinterval (55 yr) is used; the right-hand panels show similar results, plotted as a function of both wavelet period and model subinterval length (see below).

The analyses conducted in this study were meant to validate different climate models against observations, then to compare different models to one another. Examples of each are shown in Fig. 3c (see below). If the CORE hindcast is tested against a version of itself contaminated with an AR(1) red noise spectrum of varying amplitude (not pictured), reliable results are found; CORE does not differ from itself by this metric for modest noise levels, and disagreement grows with additional noise.

Validation against observations is performed on three model simulations: CCSMcontrol, the CM2.1 simulation discussed earlier, and an additional 600-yr CCSM simulation using a lower value of the threshold relative humidity for cloud formation, hereafter RHLOW. Frequency bleeding, or shifting of wavelet power to different periods depending on the length of the time series, is prevented by using model subintervals of the same length as CORE (in this case, 55 yr). Results are found in Fig. 3 (left) where horizontal lines indicate differences at the 80%, 90%, and 95% levels. CCSMcontrol agrees relatively well with CORE at all wavelet periods; the same is true for RHLOW. CM2.1 is similar to CORE at 90% confidence for all periods except the 8–12-year band, which is the same period range where model–data agreement drops off sharply in Fig. 2d below.

Model/model comparison is then performed for CCSMcontrol/CM2.1 and CCSMcontrol/RHLOW (Fig. 3, right): CCSMcontrol and CM2.1 differ throughout the 8–20-year band, but only at long (≥ 300 yr) subinterval lengths. In contrast, for the CCSMcontrol/RHLOW comparison, the areas of disagreement are smaller than for CCSMcontrol/CM2.1. Some areas of disagreement between CCSMcontrol and RHLOW are seen near 2–3, 6–8, and 12–16 years, but only at subinterval lengths longer than 300 yr.

The above test cases constitute sanity checks: CCSM simulations that differ only by a parameter adjustment are closer to one another than to CM2.1. We expect this method to usefully quantify true physical differences between models, and prevent overtuning based on statistically insignificant model–data and model–model differences.

3. Conclusions

Wavelet probability analysis is a robust method of measuring agreement between one or more datasets. Using the PDF of the Niño-3.4 wavelet power, CCSM3.5 is seen to agree well with the ocean hindcast product of

Large and Yeager (2008). Although NIÑO-3.4 is not necessarily the best measure of ENSO variability (Guilyardi et al. 2009; van Oldenborgh et al. 2005), the wavelet PDF at least indicates that probabilistic measures of Niño-3.4 statistics yield results consistent with previous work on validating CCSM3.5 (Neale et al. 2008). Also, wavelet probability analysis can easily be applied to other time series aside from Niño-3.4, and this is anticipated to provide additional information on model performance.

Self-agreement depends on the record length; the 90% confidence interval on the self-overlap WPI distribution narrows exponentially with record length, and in general halves for every increase of 80 yr in simulation length. Using a 1200-yr CCSM3.5 simulation, a 2000-yr GFDL CM2.1 simulation, and an 1155-yr IPSL CM4 simulation, statistically identical regression slopes are found; this property may be exploited to provide the expected level of agreement for a model simulation of arbitrary length. It is likely that 250 yr is sufficient to illustrate 90% of the range of ENSO behavior, and should be viewed as a minimum length for future long baseline simulations. Additionally, when the regression is calibrated against HadSST, we find that approximately 240 yr of observations are required to achieve convergence, indicating that the observational record is itself temporally limited. We note that the length estimates from both models and observations are longer than typical IPCC scenario simulations, which are roughly 110 yr long. Further analysis may be conducted on any arbitrary dataset using the online wavelet probability analysis toolkit, which among other things, provides for the rapid assessment of the needed simulation length for any model.

A procedure for validating shorter model simulations against observations is demonstrated using an empirical hypothesis testing procedure on CCSM and CM2.1, with the ocean hindcast of Large and Yeager (2008) as a reference. Differences between CCSM3.5 and CM2.1 at some frequencies are detectable only for model subintervals longer than 200 yr, suggesting this value as a reasonable minimum length for CCSM model intercomparison studies. More dramatic changes to model physics lead to more dramatic intermodel differences, providing evidence that the method is sensitive to the degree of physical changes.

Wavelet probability analysis is a simple but powerful tool that provides robust statistical limits on the expected level of agreement between time series of any length, from any source; this technique should prove useful for the development of future climate models.

Acknowledgments. This work was supported by NASA Headquarters under the NASA Earth and Space

Science Fellowship Program Grant NNX09A020H to SS. A. Wittenberg is gratefully acknowledged for providing the Niño-3.4 time series from GFDL CM2.1, and E. Guilyardi and M. Khodri for providing the Niño-3.4 time series from IPSL CM4.

REFERENCES

- AchutaRao, K., and K. R. Sperber, 2006: ENSO simulation in coupled ocean–atmosphere models: Are the current models better? *Climate Dyn.*, **27**, 1–15.
- Brown, J., A. W. Tudhope, M. Collins, and H. V. McGregor, 2008: Mid-Holocene ENSO: Issues in quantitative model-proxy data comparisons. *Paleoceanography*, **23**, PA3202, doi:10.1029/2007PA001512.
- Capotondi, A., A. Wittenberg, and S. Masina, 2006: Spatial and temporal structure of tropical Pacific interannual variability in 20th century coupled simulations. *Ocean Modell.*, **15**, 274–298.
- Daubechies, I., 1990: The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans. Inf. Theory*, **36**, 961–1004.
- Guilyardi, E., 2006: El Niño-mean state-seasonal cycle interactions in a multi-model ensemble. *Climate Dyn.*, **26**, 329–348.
- , A. Wittenberg, A. Fedorov, M. Collins, C. Wang, A. Capotondi, G. J. van Oldenborgh, and T. Stockdale, 2009: Understanding El Niño in ocean–atmosphere general circulation models: Progress and challenges. *Bull. Amer. Meteor. Soc.*, **90**, 325–340.
- Jochum, M., B. Fox-Kemper, P. Molnar, and C. Shields, 2009: Differences in the Indonesian seaway in a coupled climate model and their relevance to Pliocene climate and El Niño. *Paleoceanography*, **24**, PA1212, doi:10.1029/2008PA001678.
- , S. Yeager, K. Lindsay, K. Moore, and R. Murtugudde, 2010: Quantification of the feedback between phytoplankton and ENSO in the community climate system model. *J. Climate*, **23**, 2916–2925.
- Large, W. G., and S. G. Yeager, 2008: The global climatology of an interannually varying air–sea flux dataset. *Climate Dyn.*, **33**, 341–364, doi:10.1007/s00382-008-0441-3.
- Liu, Y., X. S. Lian, and R. H. Weisberg, 2007: Rectification of the bias in the wavelet power spectrum. *J. Atmos. Oceanic Technol.*, **24**, 2093–2102.
- McGregor, H. V., and M. K. Gagan, 2004: Western Pacific coral $\delta^{18}\text{O}$ records of anomalous Holocene variability in the El Niño–Southern Oscillation. *Geophys. Res. Lett.*, **31**, L11204, doi:10.1029/2004GL019972.
- Montgomery, D., and G. Runger, 2007: *Applied Statistics and Probability for Engineers*. John Wiley & Sons, 768 pp.
- Neale, R. B., J. H. Richter, and M. Jochum, 2008: The impact of convection on ENSO: From a delayed oscillator to a series of events. *J. Climate*, **21**, 5904–5924.
- Nicholls, N., B. Lavery, C. Frederiksen, and W. Drosowsky, 1996: Recent apparent changes in relationships between the El Niño–Southern Oscillation and Australian rainfall and temperature. *Geophys. Res. Lett.*, **23**, 3357–3360.
- Ropelewski, C. F., and M. S. Halpert, 1996: Quantifying Southern Oscillation–precipitation relationships. *J. Climate*, **9**, 1043–1059.
- Seager, R., 2007: The turn of the century North American drought: Global context, dynamics, and past analogs. *J. Climate*, **20**, 5527–5552.
- Torrence, C., and G. Compo, 1998: A practical guide to wavelet analysis. *Bull. Amer. Meteor. Soc.*, **79**, 61–78.
- Trenberth, K. E., G. W. Branstator, D. Karoly, A. Kumar, N.-C. Lau, and C. Ropelewski, 1998: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *J. Geophys. Res.*, **103C**, 14 291–14 324.
- van Oldenborgh, G., S. Philip, and M. Collins, 2005: El Niño in a changing climate. *Ocean Sci.*, **1**, 81–95.
- Wittenberg, A. T., 2009: Are historical records sufficient to constrain ENSO simulations? *Geophys. Res. Lett.*, **36**, L12702, doi:10.1029/2009GL038710.
- , A. Rosati, N.-C. Lau, and J. J. Ploshay, 2006: GFDLs CM2 global coupled climate models. Part III: Tropical Pacific climate and ENSO. *J. Climate*, **19**, 698–722.
- Zhang, X., and M. J. McPhaden, 2006: Wind stress variations and interannual sea surface temperature anomalies in the eastern equatorial Pacific. *J. Climate*, **19**, 226–241.