ARTICLE OPEN

Low-dimensional representations of Niño 3.4 evolution and the spring persistence barrier

Michael K. Tippett¹✉ and Michelle L. L'Heureux²

The El Niño-Southern Oscillation (ENSO) is the dominant mode of climate variability on interannual time scales, and its temporal evolution can be summarized using the Niño 3.4 index. Here we used EOF analysis to construct low-dimensional representations of the 12-month evolution of Niño 3.4 during different times of the year. The leading EOF explains more than 90% of the variance of the Niño 3.4 evolution from June to May, which means that the differences in evolution from one year to another are essentially differences in amplitude. Two EOFs explained 94% or more of the evolution variance for other 12-month periods of the year. The two-dimensional nature of the Niño 3.4 trajectories is a direct expression of the spring persistence barrier since the first EOF describes wintertime ENSO events, and the second EOF describes independent behavior during the antecedent spring. A periodic second-order autoregressive (AR2) model reproduced the observed properties, but a first-order model did not. Niño 3.4 EOFs in ocean-atmosphere coupled model forecasts matched observed EOFs with varying levels of fidelity depending on model and time of year. Forecast models with more accurate climatological covariance also have lower mean-squared error (MSE). Low-dimensional EOF-based statistical corrections reduced forecast MSE.

npj Climate and Atmospheric Science (2020)3:24; <https://doi.org/10.1038/s41612-020-0128-y>

INTRODUCTION

Sir Gilbert Walker identified three global-scale patterns of pressure variability in his statistical studies of “world weather”: the Southern Oscillation, the North Atlantic Oscillation, and the North Pacific Oscillation¹. The Southern Oscillation is a seesaw of atmospheric pressure between the eastern Pacific and Indonesia. Walker was especially interested in the Southern Oscillation because of the broad geographic extent of its influence and because he saw its persistence from one season to the next as an indication of its potential to be predicted in advance². Walker used sea-level pressure at Darwin to characterize the Southern Oscillation and developed autoregressive models to explain the quasi-periodic behavior of the Darwin pressure time series³. Eventually the Southern Oscillation was understood to be related, through the Walker circulation, to so-called El Niño events which were observed near the coast of Peru⁴. Characterizing and understanding the properties of the El Niño-Southern Oscillation (ENSO) remains an important scientific challenge⁵ because of ENSO's role as the dominant and most predictable mode of climate variability on interannual time scales. Following in the tradition of Walker, ENSO studies have often leaned heavily on statistical methods.

For instance, the leading Empirical Orthogonal Function (EOF) of monthly sea surface temperature (SST) anomalies is associated with ENSO⁶, and its time series is highly correlated with the Niño 3.4 index^{7,8}. EOF analysis, also known as principle component analysis, decomposes data into uncorrelated components that optimally approximate the data. The first EOF explains the most variance of the data. The second EOF explains the most variance of the data subject to the constraint of being uncorrelated with the first EOF, and so on. When a few EOFs of a dataset explain a large fraction of the total variance, they can be used as a low-dimensional approximation. A single number (the coefficient of the leading EOF, or equivalently the value of the Niño 3.4 index) is sufficient to characterize the overall state of ENSO at any given

time and is more representative than an observation at a single location (e.g., Darwin pressure). Two EOFs are sufficient to describe the diversity of ENSO expressions or flavors⁹. While the state of ENSO is well-described by one or two spatial EOFs, ENSO forecasts are more skillful when more EOFs are included in statistical ENSO predictions^{10–12}. Linear inverse models, for instance, require higher-order EOFs to capture the initial states that optimally evolve into mature ENSO events¹³.

The fact that the state of ENSO can be described using a small number of spatial patterns (EOFs) indicates that ENSO is spatially low-dimensional. However, such EOF analysis fails to provide direct information about the complexity or dimensionality of ENSO evolution in time. In fact, the same spatial patterns are obtained from EOF analysis if the time sequence of the data is arbitrarily reordered. There is evidence that ENSO evolution might be relatively simple. First, physical reasoning and observational data tell us that the magnitude of Niño 3.4 changes from one month to another is constrained and that Niño 3.4 values tend to persist from one month to another¹⁴. Second, the good performance of linear inverse models means that the complexity of the predictable part of ENSO evolution is limited, and this seems to be the case for dynamical models as well^{15,16}. However, to date, there has been no comprehensive assessment of the dimensionality of ENSO evolution or of the degree to which ENSO evolution varies from one year to another.

Information about ENSO evolution is of particular interest because of the impacts of ENSO on weather and climate variability^{17–22}. Operational ENSO forecasts are issued regularly by national and international agencies based on guidance from statistical and physics-based prediction systems²³. These forecasts provide information, often in probabilistic form, about the expected range of future Niño 3.4 values. A typical ENSO forecast for the next year provides information regarding the evolution of monthly Niño 3.4 values over next 12 months. These 12 Niño 3.4

¹Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY 10027, USA. ²NOAA/NWS/NCEP/Climate Prediction Center, College Park, MD 20740, USA. ✉email: mkt14@columbia.edu

values are not independent of each other, and it seems plausible that a sequence of Niño 3.4 values might be well-approximated by a low-dimensional representation. Here we addressed the question of the extent to which ENSO evolution admits low-dimensional representations by application of EOF analysis to time sequences of the Niño 3.4 index.

RESULTS

EOF analysis of Niño 3.4 observations

EOF analysis was applied to non-overlapping, consecutive 12-month segments of Niño 3.4 anomalies (see Methods). The EOF analysis was done 12 times, once for each of the 12 possible segment-start months. The EOF patterns (12-month segments) and fraction of variance that they explain depend on the choice of starting month of the segments, which may be surprising since the 12 EOF calculations use nearly the same data. The fundamental reason for this dependence is that the Niño 3.4 data are nonstationary in the sense that their statistics are not invariant to shifts in time. For instance, the statistics of January–December segments are different from those of February–January segments. In particular, there will be little correlation (-0.084) between January and December values in the January–December segments because they are separated by 12 months. On the other hand, January and December values will be highly correlated (0.99) in the February–January segments because they are adjacent. These differences are reflected in the covariance matrices used to compute the EOFs, and are the reason why the EOFs depend on the segment-start month.

For all segment-start months, the first EOF (EOF1) explains more than 50% of the total variance (Fig. 1a). The fraction of variance explained by EOF1 is smallest for segments that start in December (51%) and largest (90%) for segments that start in May and June. This result means that 12-month sequences of Niño 3.4 values running from June to May are essentially one-dimensional, which means that year-to-year differences are essentially differences in amplitude. The fraction of variance explained by the second EOF (EOF2) has the opposite dependence on segment-start month—it is smallest for segments starting in May (5%) and largest (44%) for segments starting in December (Fig. 1b). The variance explained by EOF1 is only modestly greater than that explained by EOF2 for segments starting in November, December, and January, and this similarity of explained variance indicates that EOF1 and EOF2 are not well-separated and that mixing of modes is possible (Supplementary Information Fig. 1 and see Methods). The fraction of variance explained by EOF3 is between 1 and 4%, depending on segment-start month, and peaks for segments starting in April and September (Fig. 1c). The residual variance after accounting for EOF1 and EOF2 varies between 4 and 6% (Fig. 1d), which means that a 2-dimensional representation captures over 94% of the variance of 12-month segments of Niño 3.4, regardless of segment-start month. This 2-dimensional representation is especially effective for segments that start in June, in which case, EOF1 and EOF2 combine to capture 97% of the evolution variance. The projection of the Niño 3.4 index onto the leading 2 EOFs is an accurate approximation of the Niño 3.4 index that uses only 2 numbers per 12-month (June–May) period (Fig. 1e). The difference between the Niño 3.4 index and the projection is seldom larger than 0.5°C , and the standard deviation of the difference is $<0.2^{\circ}\text{C}$.

We aligned the 12 sets of EOFs according to the months used in their construction so that they covered a 23-month period (Fig. 2). Specifically, the EOFs for segments that start in January cover the first 12 months, the EOFs for segments that start in February cover months 2–13, and so on, ending with the EOFs for segments that start in December which cover months 12–23. This arrangement provides a coherent picture of Niño 3.4 evolution in which EOF1 represents an El Niño event (with this choice of sign) that grows

during the first spring and summer, peaks in December, and decays during the second spring (Fig. 2a). EOF2 provides complementary structure in the beginning (first winter/spring) and end (second summer/fall) of the 23-month period, which are the months when the amplitude of EOF1 is small (Fig. 2b). There is some modest indication of mode mixing of EOF1 and EOF2 for the December segment-start month. By construction, EOF1 and EOF2 are orthogonal (zero correlation between the 12 pairs of values), and this orthogonality is achieved with EOF1 and EOF2 having approximately unimodal structures which are out of phase and are locked to the annual cycle. EOF3 is orthogonal to EOF1 and EOF2, but its structure is not locked to the annual cycle. Instead, EOF3 is roughly bimodal with values that are positive at the beginning and ending months of the segment and that are negative in the middle months of the segment, regardless of segment-start month.

The fraction of variance explained by an EOF in a particular month, as opposed to the fraction of variance explained over a 12-month segment, is the square of the EOF amplitude in that month divided by climatological variance of Niño 3.4 in that month. We examined the variance explained by EOF1 and EOF2 in individual months and found that EOF1 and EOF2 divide the 23-month period into distinct parts (Fig. 3). This separation is a direct expression of the spring persistence barrier, which is the idea that “spring is the transition time from one climate state to another”¹⁴. EOF1 explains more than 69% of the monthly variance during the months of June–April, depending on segment-start month, and sharply less in other months (Fig. 3a). EOF2 explains 76% or more (a higher per month fraction than EOF1) of the variance during the months of January–April of the first year and June–December of the second year (Fig. 3b). For fixed start-month segment, the variance explained by EOF1 is uncorrelated with the variance explained by EOF2. Consequently, nearly all of the Niño 3.4 variance in late winter and early spring (when EOF2 explains most of the variance) is independent of the Niño 3.4 variance during the following winter (when EOF1 explains most variance). In other words, the spring persistence barrier^{14,24} (related in the forecast context to the spring predictability barrier^{25–27}) emerges in the EOF analysis from the structure and independence of EOF1 and EOF2. EOF3 explains little variance except in May for segments that start in May, when it explains nearly 50% (Fig. 3c). The fact that EOF3 is uncorrelated with the other EOFs means that 50% of the May variance is uncorrelated with the variance that follows it, which is a direct expression of the spring persistence barrier. The variance unexplained by EOF1 and EOF2 is substantially lower in other months (Fig. 3d). The spring persistence barrier, which has been explained by a seasonally varying growth rate^{14,28}, is the dominant feature in the lag correlation of Niño 3.4 with the values that follow it (Fig. 4a). Projection of Niño 3.4 onto EOF1 and EOF2 gives almost the same lag correlation (Fig. 4b), except the May lag correlations are slightly too strong. Projecting onto the leading two EOFs generally increases lag correlations by filtering out the variance in the higher-order EOFs.

Although the time series of EOF1 (PC1) and the time series of EOF2 (PC2) are uncorrelated by construction, EOF analysis does not constrain the correlation of PC1 and PC2 for different 12-month periods. In fact, PC1 and PC2 display substantial cross-correlation depending on segment-start month (Fig. 5). This cross-correlation is explained by noting that the relative smoothness of the two-dimensional reconstruction (Fig. 1e) from year t to year $t + 1$ would be impossible if values of PC1 and PC2 for year t were completely independent of their values for year $t + 1$. Consider EOF1 and EOF2 of segments that start in January. The requirement of smoothness means that $\text{PC1}(t)$ must be highly correlated with $\text{PC2}(t + 1)$ because the December value of EOF1 and the January value of EOF2 are both strongly positive and explain most of the variance in those months. There is no constraint on $\text{PC1}(t)$ and $\text{PC1}(t + 1)$ because EOF1 explains little variance in January. A similar

Fig. 1 Variance explained by the EOFs of Niño 3.4 for the period 1982–2019. As a function of segment-start month for **a** EOF1, **b** EOF2, and **c** EOF3. Error bars in **a** and **b** indicate the 95% confidence intervals from 1000 simulations of an AR1 model (black), an AR2 model (orange), and an AR2 model with constant variance (light blue). **d** The variance unexplained by EOF1 and EOF2. **e** Niño 3.4 time series (black), its projection onto EOF1 and EOF2 (red), and the projection residual (blue) for segments that start in June.

argument noting that EOF2 explains much variance in January and little variance in December shows that $PC2(t)$ and $PC2(t+1)$ need not be correlated. Now apply the same reasoning to EOF1 and EOF2 of segments that start in December. The requirement of smoothness now means that $PC1(t)$ must be highly correlated with $PC2(t-1)$ because the December value of EOF1 and the November value of EOF2 are both strongly positive and explain most of the variance in those months. The smoothness argument does not apply to segments that start in May and June because EOF1 explains little variance in the beginning and ending months of those segments. Consequently, $PC1$ and $PC2$ show little correlation with their prior values for segments that start in May

and April, which again is an expression of the spring persistence barrier.

The main features of the Niño 3.4 EOFs: explained variance as a function of segment-start month, temporal structure, and explained variance as a function of target month were found to be robust when we repeated the analysis on SST data from an independent 37-year period, 1945–1981, that uses a different methodology (Supplementary Information Figs. 2–4). Some differences are lower overall amplitude, which may reflect differing SST estimation methodologies as well as period (Supplementary Information Fig. 2), slightly less variance explained by EOF1 and EOF2 (Supplementary Information Fig. 5), the 2-EOF reconstruction has noticeably smaller residuals

Fig. 2 EOFs of Niño 3.4 for the period 1982–2019. Colors indicate **a** EOF1, **b** EOF2, and **c** EOF3 values of as a function of segment-start month (vertical axis) and segment month (horizontal axis). Note the smaller colorbar range in **c**.

Fig. 3 Segment month variance explained by the EOFs of Niño 3.4 for the period 1982–2019. **a** EOF1, **b** EOF2, and **c** EOF3 of Niño 3.4 arranged by segment-start month. **d** The variance unexplained by EOF1 and EOF2. Note the smaller colorbar range in **c** and **d**.

Fig. 4 Lag correlation as function of base month and lag for the period 1982–2019. **a** Observed Niño 3.4, and **b** observed Niño 3.4 projected onto EOF1 and EOF2 (January segment starts).

Fig. 5 Lag correlations between PC1 and PC2 for 12-month segments starting in January, ..., December. Blue line: PC1 in year t and PC2 in the following year $t+1$. Red: PC1 in year t and PC2 in the preceding year $t-1$.

1948–1981 (standard deviation 0.2°C) compared with the period 1945–1947 (standard deviation 0.5°C), which is possibly related to SST estimates which have an uncertainty maximum around World War II²⁹ (Supplementary Information Fig. 3e), and the order of EOF1 and EOF2 is reversed for segments that start in December (Supplementary Information Fig. 4). The overall robustness of the EOF analysis to period and dataset suggests that the results are tied to the underlying physics and should be reproduced by physics-based models.

Autoregressive modeling

To explain the low-dimensional behavior of the Niño 3.4 index, we fitted first-order (AR1) and second-order (AR2) periodic autoregressive models to the Niño 3.4 index (see Methods). The coefficients and error variances of the models depend on target month. The error variance of the AR2 model was statistically significantly smaller than that of the AR1 model (indicating better fit) for the target months of May, June, and July (Fig. 6). For other target months, the reduced error variance of the AR2 model was statistically insignificant. The error variance of the AR2 model is nearly uniform except for June and October target months. The June error variance of the AR2 model is more than twice the average of all months.

Fig. 6 Unbiased estimates of the error variance by target month of AR1 and AR2 models fit to the Niño 3.4 index 1982–2019. The error variance of the AR2 model is statistically significantly lower than the error variance of the AR1 model in the target months of May, June, and July.

We applied EOF analysis to 1000 simulations from the AR1 and AR2 models. For both models, the variance explained by EOF1 and EOF2 is qualitatively similar to that of the observations (first two error bars in Fig. 1a, b). However, the observed explained variance of EOF1 and EOF2 falls outside the 95% confidence intervals of the AR1 model (Fig. 1a, b), and consequently we rejected (0.05 significance level) the hypothesis that the observations come from the AR1 model. The AR1 model is especially inadequate in matching the observed EOF variance for segments with winter start months (November, December, and January). The observed explained variance of EOF1 and EOF2 does fall within the 95% confidence intervals of the AR2 model (Fig. 1a, b), and we cannot reject the AR2 model on this basis. We also applied EOF analysis to 1000 simulations from the AR2 model with constant error variance, and on the basis of explained variance (third set of error bars in 1a, b), we did not reject it either. The explained variance by month of the AR2 EOFs also matched well the observations (compare Fig. 3 and Supplementary Information Fig. 6). Notably the high variance explained by EOF3 in May for segments starting in May is present both in observations and in the AR2 simulations (compare Fig. 3c and Supplementary Information Fig. 6c). We hypothesize that this feature is related

Fig. 7 Lag correlation as function of base month and lag. Averaged over **a** AR2 simulations with varying noise variance, and **b** AR2 simulations with constant noise variance.

to the month varying error variance of the AR2 model since it is substantially reduced when the monthly varying error variance is replaced by its annual average (not shown). The average lag correlations of the AR2 model (Fig. 7a) compare well to the observed ones (Fig. 4a). The spring persistence barrier is less pronounced in May for the AR2 model with constant error variance (Fig. 7b). This behavior makes sense because the large error variance for June targets serves to reduce the dependence of June values on those from May and April.

EOF analysis of Niño 3.4 forecasts

We compared the first two EOFs of North American Multimodel Ensemble (NMME) Niño 3.4 forecasts (see Methods) with the first two EOFs of observations by computing the earth moving distance (EMD; see Methods) between the corresponding observation and forecast covariance matrices for each forecast start month (Fig. 8a). The EMD measures the overall fidelity of the forecast covariance spanned by the leading two forecast EOFs and depends on differences in variances (diagonal elements of the covariance matrix) and differences in covariance (off-diagonals of the covariance matrix). The EOFs are the eigenvectors of the covariance matrix, which is why errors in EOFs are related to the EMD between the model and observed covariance. EMD is a measure of the climatological fidelity of the forecasts (not skill) and is small when the forecast and observation EOFs are similar. EMD is not a direct measure of skill. The EMD for NCEP-CFSv2 was smallest overall and fairly uniform with respect to start month. The EMD for COLA-RSMAS-CCSM4 was similar to that of NCEP-CFSv2 except in the last third of the year when it is higher. The EMD for COLA-RSMAS-CCSM4 and CMC2-CanCM4 was lower on average for segments starting in the first half of the year and higher for segments starting in the second half. The EMD for the other models showed a roughly U-shaped dependence on segment-start month with the lowest values near June or July and the highest values for winter months.

We compared observation and model EOFs for forecasts starting in February, June, and October. We chose June because the spread in EMD across models is relatively small, and because the decomposition of the observations into 2 EOFs is most accurate then. We chose February and October because the spread in EMD across models is relatively large, and because of contrasting skill levels at long leads—low for February starts and high for October starts. A comparison of the model and observation EOFs showed that forecasts with large EMD values have relatively poor representations of the Niño 3.4 EOFs. EOF1 is dominant for forecasts starting in June, and these forecasts EOFs generally compared well with observations. CMC2-CanCM4 is the model with highest June EMD, and the amplitude of its EOF1 is

too large (Fig. 8f). GFDL-CP2p1-aer04 has the next highest June EMD, and the peak of its EOF1 is in August rather than December (Fig. 8b). GFDL-CM2p5-FLOR and COLA-RSMAS-CCSM4 have similar June EMD, and their EOF1s have too small amplitude (Fig. 8e, d). NOAA-CFSv2 and COLA-RSMAS-CCSM4 have the smallest June EMD, and their EOFs match observations except for a September inflection point (Fig. 8c, g). Notably, this September feature is robust because it appears both in the earlier period (1945–1981) and in the AR2 model. For February starts, the EMD for GFDL-CM2p1-aer04 was largest, and the peak of its EOF1 is in August rather than December (Supplementary Information Fig. 7a). The EMD for the CMC1-Can3 and FLOR models were similar to each other, and their EOF1s grow more sharply than observed in spring and change slope in August (Supplementary Information Fig. 7c, d). The EMD for COLA-RSMAS-CCSM4 and NOAA-CFSv2 were smaller than for other models, and their EOFs match the observations well (Supplementary Information Fig. 7b, f). For October starts, the EMD for CMC2-CanCM4 was largest, and its EOF1 peaks in February instead of December (Supplementary Information Fig. 7k). The EMD for GFDL-CM2p1-aer04 was also large, and its EOF2 begins to grow too soon in the year (Supplementary Information Fig. 7g). Again, the EMD for NOAA-CFSv2 was smallest for October starts, and its EOFs match observations well (Supplementary Information Fig. 7i).

The ability of forecast models to match the observed climatological variability is different from the ability to make skillful forecasts. For instance, shuffling forecast years randomly would eliminate skill but not change the EMD. EMD is a lower bound on mean-squared error (MSE) and directly contributes to MSE in simple cases (see Methods). Consequently, incorrect variability in forecasts would generally be expected to increase MSE. We found that forecast skill is indeed related to correctly simulated variability. The correlation of squared EMD with MSE across models (squared EMD is averaged over starts; MSE is averaged over starts and leads) is 0.79 (Fig. 8h), and the correlation across models and start months (MSE is averaged over leads) is 0.74 (Fig. 8h). Models with lower EMD have lower MSE.

The presence of systematic deficiencies in the forecast variability is an indication that statistical corrections of forecast evolution might reduce forecast errors. Such statistical correlations are feasible when the correction is performed in the space of the leading 2 EOFs of ENSO evolution, which reduces the number of regression coefficients to be estimated. We corrected systematic forecast error by a multivariate linear regression between forecast (ensemble mean) and observation EOFs. The first lead was not corrected. Cross-validated MSE is generally reduced, mostly at longer leads and by greater amounts for the models whose EOFs differ most from observed (Fig. 9). MSE reductions are smallest for COLA-RSMAS-CCSM4 and largest for GFDL-CM2p1-aer04,

Fig. 8 Earth moving distance (EMD), EOFs, and mean-squared error (MSE) of NMME Niño 3.4 forecasts. **a** EMD between observed and NMME rank-2 Niño 3.4 segment covariances for segment-start months January–December. February, June, and October values are marked with circles. **b–g** Observed (gray) and NMME EOF1 (blue) and EOF2 (red) for June starts. Scatter plots along with best-fit lines for mean-squared EMD and MSE **h** across models and **i** across models and start months.

Fig. 9 Reduction in mean-squared error of NMME Niño 3.4 forecasts from EOF regression by target month and lead. Note the different colorbar ranges.

matching their EMD behavior. The statistical correction has higher correlations at many long leads with winter targets but slightly lower correlations at some shorter leads for summer targets when skill is low (Supplementary Information Fig. 8), which may reflect the tendency of correlation to be reduced under cross-validation, especially at lower skill levels³⁰.

DISCUSSION

Empirical orthogonal function (EOF) analysis revealed that the observed ENSO evolution is well-approximated by 2 EOFs which isolate the intrinsic seasonal cycle of ENSO evolution and its phase-locking to the annual cycle. The first EOF (EOF1) explains the most variance during the months of June–April and represents the tendency of ENSO to grow during summer, peak in winter (December), and then decay in spring. The second EOF (EOF2) explains the most variance in the months prior to (late winter–early spring) and following (early summer–fall) the ENSO event described by EOF1. Although the observed ENSO evolution contains features beyond these two patterns, these two patterns explain more than 94% of the overall variance of 12-month ENSO evolution sequences. This finding is both unexpected and expected. The fact that two EOFs explain so much of the evolution variance is unexpected because the common perception is that “each ENSO event is unique”³¹ and that ENSO evolution is complex⁵ and “varies greatly from event to event”³². On the other hand, the existence of a 2-dimensional representation of ENSO evolution is unsurprising, at least in hindsight, because the spring persistence barrier, which is the dominant coherent feature of ENSO evolution, splits the year into approximately independent components. Also, the low-dimensionality of ensemble-mean ENSO forecasts³³ and predictable components¹⁵ has been noted previously.

Our explanation of the low-dimensionality of ENSO evolution in terms of autoregressive (AR) models connects this work with some of the earliest analysis of ENSO-related variability by Walker who used AR models to represent quasi-periodic behavior in Darwin pressure records^{3,34}. In particular, Walker focused on the relation between the autocorrelation and periodicities, seeing periodicity as related to predictability. The same issue of periodicity was examined more than a half a century later in similar models for the

normalized pressure difference between Tahiti and Darwin, and periodicities between 3 and 4 years were found^{35,36}. However, those models contained no seasonal dependence and thus were unable to simulate any nonstationary aspects of ENSO, such as seasonally dependent variance, the spring predictability barrier, or the low-dimensional representation documented here. While the use of seasonally dependent AR1 models to model ENSO has a long history¹⁴, we are aware of only one other comparison of seasonally varying AR1 and AR2 models for ENSO³⁷. Our analysis identified the target months (May–July) when the AR2 model was a statistically significant improvement over the AR1 model. Also, while the leading EOFs of the AR1 simulations are similar in some qualitative aspects to the observed ones, they failed to match quantitatively the observed low-dimensional properties of ENSO evolution. Consequently, we would argue that the AR2 model is the more realistic model for ENSO evolution and that its properties are deserving of further study. In the AR2 model, the properties of the spring persistence barrier depend on the seasonally varying noise forcing, which is largest for June targets and which reflects the lack of linear relation between the June ENSO state and previous values.

We speculate that the superior performance of the AR2 model in predicting May–June Niño 3.4 values might come from implicitly including subsurface information. The AR2 model bases its predictions on the prior value and prior tendency, while the AR1 model uses only the prior value. For instance, an AR2 May prediction is based on the April value and the March–April tendency, while the AR1 May prediction is based only on the April value. Warm water volume (WWV) in the early part of the calendar year is a strong indicator of future ENSO evolution³⁸, and WWV is related to Niño 3.4 tendency since the correlation of March WWV (west; 5N–5S, 120 E–155 W) with the March–April Niño 3.4 tendency is 0.72. In this sense, the AR2 model for May might be implicitly benefiting from subsurface information.

The low-dimensional description presented here of observed ENSO evolution provides a new metric for the assessment of model fidelity. Metrics related to the spring persistence barrier intensity have been used to diagnose ENSO evolution in historical simulations²⁴, and the EOF analysis here provides an additional perspective for judging the realism of ENSO simulations. This metric also has application in the forecast context since we found

that forecast models with more realistic EOFs have smaller forecast errors. The low-dimensionality of ENSO evolution opens up new potential for statistical correction of ENSO forecasts in which the predictor is the 12-month ENSO forecast, and the predictand is the observed ENSO trajectory. Such a correction would be impossible without low-dimensional representations since a multivariate regression would require estimating 144 coefficients for each start month, while the low-dimensional correction requires only 4 coefficients.

An important caveat is that the ability of the Niño-3.4 index to capture the full range of ENSO diversity³⁹ is limited, especially with regard to spatial diversity. Other SST indices⁴⁰ and other variables^{21,41,42} provide additional perspectives on ENSO dynamics, evolution, and impacts. A question for future research is whether the evolution of other ENSO indices is also dominated by low-dimensional features.

METHODS

Niño 3.4 observations

The Niño 3.4 index is the average sea surface temperature (SST) over the Equatorial Pacific region 5°S–5°N and 170°–120°W⁸. For most of the calculations, the Niño 3.4 index was computed using the Reynolds-OISST dataset⁴³ for the period January 1982–June 2019. To assess the robustness of the results, we repeated some of the calculations using the Niño 3.4 index computed from version 5 of the Extended Reconstructed Sea Surface Temperature dataset^{44,45} for the period January 1945–December 1981.

Niño 3.4 forecasts

We analyzed Niño 3.4 forecasts from 6 forecast systems in the North American Multimodel Ensemble (NMME) database⁴⁶: GFDL-CM2p1-aer04, COLA-RSMAS-CCSM4, GFDL-CM2p5-FLOR, CMC1-CanCM3, CMC2-CanCM4, and NCEP-CFSv2. GFDL-CM2p5-FLOR consists of FLOR-A and FLOR-B pooled. The models are a subset of those in refs. ^{47,48}, where further details of their configurations can be found. Nominal monthly start dates are January 1, 1982–December 1, 2018. All forecasts extend 12 months past their start date except NCEP-CFSv2 forecasts which are 10 months. Anomalies were formed by removing the forecast climatology computed over the period. Similar to refs. ^{47,48}, two climatology periods were used for COLA-RSMAS-CCSM4 and NCEP-CFSv2.

EOF analysis

EOFs were computed for segments of Niño 3.4 anomalies. Each segment contains 12 consecutive months, and there are no gaps between segments. Twelve sets of EOFs were computed. The first set of EOFs are for segments that start in January, the second set of EOFs for segments that start in February, and so on. Using data from the period January 1982–June 2019, EOFs of segments that start in January were computed by first forming the 12×37 data matrix X_1 , in which each column is a 12-month segment. Row 1 contained the 37 values: January 1982, January 1983, ..., January 2018, and column 1 contained the 12 values: January 1982, February 1982, ..., December 1982. Anomalies with respect to the annual cycle were formed by removing the means of each of the 12 rows. The same procedure was used to construct the data matrices $X_2, \dots, X_{12}$ for segments that start with February, ..., December, respectively. The data matrices for August, ..., December segment-start months have 36 columns (years).

The 12 sets of EOFs are given by the eigenvectors of the 12 covariance matrices

$$C_i = \frac{1}{N_i - 1} X_i X_i^T, \quad i = 1, \dots, 12, \quad (1)$$

where N_i is the number of columns (years), which is 37 for segments with January–July starts, and 36 for segments with August–December starts. We normalized the EOFs so that the data are equal to the sum of the EOFs (units °C) multiplied by their time series (PCs with unit variance). That is,

$$X_i = E_i F_i^T, \quad (2)$$

where each column of E_i is an eigenvector of C_i (vector of length 12) and EOF, and each column of F is the corresponding time series (vector of length N_i) with unit variance. EOFs are ordered from most variance

explained (largest eigenvalue) to least variance explained (smallest eigenvalue). In terms of the singular value decomposition of the data matrix, $X_i = U_i S_i V_i^T$, the EOFs and normalized times series are $E_i = U_i S_i / \sqrt{N_i - 1}$ and $F_i = \sqrt{N_i - 1} V_i$; the columns of U_i and V_i are orthonormal vectors. The sign of the EOFs is arbitrary. To compare EOFs with different start months, we found the following choices to be effective: EOF1 is positive in December, EOF2 is positive in May, and the middle values of EOF3 are negative.

EOF were computed in the same way for Niño 3.4 values from the period 1945–1981. Twelve sets of EOFs were computed for each of the NMME models for the ensemble mean and for the ensemble members for forecasts starting in January, February, ..., December.

When two eigenvalues of C_i are close, their eigenvectors may be mixed and the eigenvalues are said to be degenerate⁴⁹. North's rule of thumb for deciding whether eigenvalues are close is based on the error variance σ_j^2 of the j -th eigenvalue λ_j being

$$\sigma_j^2 = \frac{2\lambda_j^2}{N_i}. \quad (3)$$

Here we judged whether eigenvalues are potentially degenerate by checking whether their 95% confidence intervals overlap (Supplementary Fig. 1). The interval widths are constant on a log scale. By this criterion, EOFs 1 and 2 are potentially degenerate for segments with January, November, and December start months. The greatest overlap is for December start segments. The eigenvalues corresponding to EOFs 2 and 3 are also well-separated for segments that start in February, March, and June–October (Supplementary Information Fig. 1).

Periodic autoregressive models

The periodic AR1 model for the random process y_τ (here Niño 3.4) is

$$y_\tau = A(i)y_{\tau-1} + B(i) + \epsilon_\tau(i), \quad (4)$$

where τ is the time index in months, i is the month of the year of τ , $i = 1, \dots, 12$, and $\epsilon_\tau(i)$ is white noise with mean zero and variance $\sigma_\epsilon^2(i)$. Similarly, the periodic AR2 model is

$$y_\tau = a(i)y_{\tau-1} + b(i)y_{\tau-2} + c(i) + \eta_\tau(i), \quad (5)$$

where $\eta_\tau(i)$ is white noise with mean zero and variance $\sigma_\eta^2(i)$. The model coefficients and noise variances depend on calendar month i and were estimated by ordinary least-squares from the Niño 3.4 values January 1982–June 2019.

To examine the question of whether the AR2 model better fits the data than the AR1 model, we considered the null hypothesis that $b(i) = 0$. Under this hypothesis, the quantity

$$F(i) = \frac{1 - a(i)}{a(i)} (N_i - 3), \quad (6)$$

has an F-distribution with 1 and $N_i - 3$ degrees of freedom where $a(i)$ is the ratio of the sum squared residuals of the AR2 model to the sum squared residuals of the AR1 model; in terms of the unbiased noise variance estimates $a(i) = (N_i - 3)\hat{\sigma}_\eta^2(i)/(N_i - 2)\hat{\sigma}_\epsilon^2(i)$. The sum squared residuals of the AR2 model are always smaller, and $a(i) < 1$. The null hypothesis is rejected for sufficiently large values of F . This is reasonable since if the AR2 model offers little improvement over the AR1 model, the sum squared residuals are similar, a is close to 1, and F is small. Conversely, when the AR2 model gives a large reduction in sum square residual, a is small, and F is large. Here, at the 5% significance level and with a Bonferroni correction, we rejected the null hypothesis for i corresponding to May, June, and July, and concluded that the AR2 model offers a substantial improvement over the AR1 model in these months.

We performed 1000 Monte Carlo simulations of the AR1 and AR2 model of length 450, which is the length of the observations January 1982–June 2019. In addition, we performed a set of AR2 simulations with constant noise forcing variance, which was equal to the average of the monthly varying variance. We computed the same 12 sets of EOF analysis for each simulation as we did for the observations. The 95% confidence intervals for the explained variance in Figs. 1a and b are the 2.5 and 97.5 percentiles of the Monte Carlos simulations. The explained variance in Supplementary Information Fig. 6 and lag correlations in Fig. 7 were computed by averaging the results from the 1000 simulations.

The earth mover's distance

The earth mover's distance (EMD; Wasserstein metric; Fréchet distance; Mallows distance) between two multivariate Gaussian distributions with equal means and covariances C_o (observations) and C_m (ensemble members) is defined⁵⁰:

$$\text{EMD}^2(C_o, C_m) = \text{trace}(C_o + C_m - 2(C_o C_m)^{1/2}). \quad (7)$$

For diagonal covariance matrices, the squared EMD is the sum of squared differences of the square-root of the diagonal elements. The squared EMD is a lower bound for mean-squared error (MSE) since the expected squared sum difference of random vectors $A \sim N(0, C_o)$ and $B \sim N(0, C_m)$ satisfies the bounds⁵⁰:

$$\text{EMD}^2 \leq E \|A - B\|^2 = E \left(\sum_i (A_i - B_i)^2 \right) \leq \text{trace}(C_o + C_m + 2(C_o C_m)^{1/2}), \quad (8)$$

where the expectation is over any joint distribution with the specified marginals for A and B . Rank-2 covariance matrices were formed for observations and ensemble members by projection onto the leading 2 eigenvectors. The rank-2 covariance matrix contains all variability that is spanned by the leading 2 EOFs. To make a fair comparison with NCEP-CFSv2 which only has leads 1–10, only leads 1–10 were used in the EMD calculations.

MSE of ensemble forecasts

To see a further relation between EOFs and forecast MSE, suppose that forecasts and observations are approximated using their leading EOF (or equivalently their leading singular vector) for a given segment-start month. Then, the data matrix of observations is a rank-1 matrix and can be written $s_o u_o v_o^T$ where u_o is a unit vector of length 12 that is proportional to the EOF (see above), v_o is a unit vector that is proportional to the EOF time series, and s_o is a positive number that sets the amplitude of the EOF. Likewise, the data matrix of forecasts can be written $s_f u_f v_f^T$. The squared error summed over all starts and leads is

$$\text{MSE} = \|s_f u_f v_f^T - s_o u_o v_o^T\|_F^2, \quad (9)$$

where $\|\cdot\|_F$ denotes Frobenius norm. The square of the Frobenius norm of a matrix M is

$$\|M\|_F^2 = \sum_{ij} M_{ij}^2 = \text{trace} M^T M. \quad (10)$$

A direct calculation gives

$$\text{MSE} = s_f^2 + s_o^2 - 2s_f s_o (v_f^T v_o) (u_f^T u_o). \quad (11)$$

The correlation ρ between the time series of the observation and forecast EOFs is given by $\rho = v_f^T v_o$. The observation and forecast variances are proportional to s_f^2 and s_o^2 , respectively. For an ensemble-mean forecast, we expect that the ensemble-mean forecast variance is less than the observed variance, and we take $s_f^2 = r^2 s_o^2$, with $0 < r < 1$. Then, the normalized MSE is

$$\frac{\text{MSE}}{s_o^2} = r^2 + 1 - 2r\rho(u_f^T u_o) = (1 - \rho^2) + (r - \rho)^2 + 2\rho(1 - (u_f^T u_o)). \quad (12)$$

The normalized MSE is the sum of three positive terms in Eq. (12). The first, $(1 - \rho^2)$, is the fraction of unexplained variance and depends on skill. The second, $(r - \rho)^2$, reflects the match between the ensemble-mean variance and correlation skill, and is zero for a well-calibrated ensemble⁵¹. The third term depends on $u_f^T u_o$, which is the pattern correlation between the forecast and observation EOFs, and is minimized when $u_f = u_o$. Differences between the forecast and observation EOFs increase MSE.

Forecast corrections

We applied corrections using multiple linear regression to forecast starts for which observations were available for the entire forecast trajectory. The multiple linear regression used the amplitudes of the first two ensemble-mean EOFs to predict the amplitudes of the first two observation EOFs. Leave-out-five-year cross-validation was used to compute the regression coefficients and the EOFs.

DATA AVAILABILITY

The Reynolds-OISST data are available at http://iridl.ldeo.columbia.edu/expert/SOURCES/NOAA/NCEP/EMC/CMB/GLOBAL/Reyn_SmithOlv2/monthly/sst/. NMME forecasts are available at <https://iridl.ldeo.columbia.edu/SOURCES/Models/NMME>. Monthly values of warm water volume (west) are available at https://www.pmel.noaa.gov/tao/www/data/www_west.dat.

Received: 8 October 2019; Accepted: 29 May 2020;

Published online: 24 June 2020

REFERENCES

- Walker, G. T. Correlation in seasonal variations of weather. VIII A preliminary study of world weather. *Mem. Indian Met. Dept.* **24**, 275332 (1923).
- Katz, R. W. Sir Gilbert Walker and a connection between El Niño and statistics. *Statistical Sci.* **17**, 97–112 (2002).
- Walker, G. T. On periodicity in series of related terms. *Proc. Roy. Soc. Lond. A* **131**, 518–532 (1931).
- Bjerknes, J. Atmospheric teleconnections from the equatorial Pacific. *Month. Weather Rev.* **97**, 163–172 (1969).
- Timmermann, A. et al. El Niño–Southern Oscillation complexity. *Nature* **559**, 535–545 (2018).
- Deser, C., Alexander, M. A., Xie, S.-P. & Phillips, A. S. Sea surface temperature variability: patterns and mechanisms. *Annu. Rev. Mar. Sci.* **2**, 115–143 (2010).
- L'Heureux, M. L., Collins, D. C. & Hu, Z.-Z. Linear trends in sea surface temperature of the tropical Pacific Ocean and implications for the El Niño–Southern Oscillation. *Clim. Dyn.* **40**, 1223–1236 (2013).
- Barnston, A. G., Chelliah, M. & Goldenberg, S. B. Documentation of a highly ENSO-related SST region in the equatorial Pacific. *Atmosphere–Ocean* **35**, 367–383 (1997).
- Takahashi, K., Montecinos, A., Goubanova, K. & Dewitte, B. ENSO regimes: reinterpreting the canonical and Modoki El Niño. *Geophys. Res. Lett.* **38**, L10704 (2011).
- Penland, C. & Magorian, T. Prediction of Niño 3 Sea Surface Temperatures Using Linear Inverse Modeling. *J. Climate* **6**, 1067–1076 (1993).
- Kondrashov, D., Kravtsov, S., Robertson, A. W. & Ghil, M. A hierarchy of data-based ENSO models. *J. Climate* **18**, 4425–4444 (2005).
- Chapman, D., Cane, M. A., Henderson, N., Lee, D. E. & Chen, C. A vector autoregressive ENSO prediction model. *J. Climate* **28**, 8511–8520 (2015).
- Penland, C. & Sardeshmukh, P. D. The optimal growth of tropical sea surface temperature anomalies. *J. Climate* **8**, 1999–2024 (1995).
- Torrence, C. & Webster, P. J. The annual cycle of persistence in the El Niño/Southern Oscillation. *Quart. J. Roy. Meteor. Soc.* **124**, 1985–2004 (1998).
- Zheng, Z., Hu, Z.-Z. & L'Heureux, M. Predictable components of ENSO evolution in real-time multi-model predictions. *Sci. Rep.* **6**, 35909 EP (2016).
- Newman, M. & Sardeshmukh, P. D. Are we near the predictability limit of tropical Indo-Pacific sea surface temperatures? *Geophys. Res. Lett.* **44**, 8520–8529 (2017).
- Gray, W. M. Atlantic Seasonal Hurricane Frequency. Part I: El Niño and 30 mb Quasi-Biennial Oscillation Influences. *Mon. Wea. Rev.* **112**, 1649–1668 (1984).
- Ropelewski, C. & Halpert, M. Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Mon. Wea. Rev.* **115**, 1606–1626 (1987).
- Ropelewski, C. F. & Halpert, M. S. Quantifying Southern Oscillation–precipitation relationships. *J. Climate* **9**, 1043–1059 (1996).
- Curtis, S. et al. Precipitation extremes estimated by GPCP and TRMM: ENSO relationships. *J. Hydrometeorol.* **8**, 678–689 (2007).
- L'Heureux, M. L., Tippett, M. K. & Barnston, A. G. Characterizing ENSO coupled variability and its impact on North American seasonal precipitation and temperature. *J. Climate* **28**, 4231–4245 (2015).
- Allen, J. T., Tippett, M. K. & Sobel, A. H. Influence of the El Niño/Southern Oscillation on tornado and hail frequency in the United States. *Nat. Geosci.* **8**, 278–283 (2015).
- L'Heureux, M. L. et al. Observing and predicting the 2015–16 El Niño. *Bull. Amer. Meteor. Soc.* **98**, 1363–1382 (2017).
- Tian, B., Ren, H.-L., Jin, F.-F. & Stuecker, M. F. Diagnosing the representation and causes of the ENSO persistence barrier in CMIP5 simulations. *Clim. Dyn.* **53**, 2147–2160 (2019).
- Jin, E. et al. Current status of ENSO prediction skill in coupled ocean–atmosphere models. *Clim. Dyn.* **31**, 647–664 (2008).
- Barnston, A. G., Tippett, M. K., L'Heureux, M. L., Li, S. & DeWitt, D. G. Skill of real-time seasonal ENSO model predictions during 2002–2011. Is our capability increasing? *Bull. Amer. Meteor. Soc.* **93**, 631–651 (2012).
- Larson, S. M. & Kirtman, B. P. Drivers of coupled model ENSO error dynamics and the spring predictability barrier. *Clim. Dyn.* **48**, 3631–3644 (2017).

28. Levine, A. F. Z. & McPhaden, M. J. The annual cycle in ENSO growth rate as a cause of the spring predictability barrier. *Geophys. Res. Lett.* **42**, 5034–5041 (2015).
29. Liu, W. et al. Extended reconstructed sea surface temperature version 4 (ERSST.v4): Part II. parametric and structural uncertainty estimations. *J. Climate* **28**, 931–951 (2015).
30. Barnston, A. G. & van den Dool, H. M. A degeneracy in cross-validated skill in regression-based forecasts. *J. Climate* **6**, 963–977 (1993).
31. Kirtman, B. Special issue: ENSO diversity. *Clim. Dyn.* **52**, 7133–7133 (2019).
32. Wu, X., Okumura, Y. M. & DiNezio, P. N. What controls the duration of El Niño and La Niña events? *J. Climate* **32**, 5941–5965 (2019).
33. Tippett, M. K., Barnston, A. G. & Li, S. Performance of recent multi-model ENSO forecasts. *J. Appl. Meteor. Climatol.* **51**, 637–654 (2012).
34. Walker, G. T. On periodicity. *Quart. J. Roy. Meteor. Soc.* **51**, 337–346 (1925).
35. Chu, P.-S. & Katz, R. W. Spectral estimation from time series models with relevance to the Southern Oscillation. *J. Climate* **2**, 86–90 (1989).
36. Trenberth, K. E. & Hoar, T. J. The 1990–1995 El Niño–Southern Oscillation event: longest on record. *Geophys. Res. Lett.* **23**, 57–60 (1996).
37. Liu, Z., Jin, Y. & Rong, X. A theory for the seasonal predictability barrier: threshold, timing, and intensity. *J. Climate* **32**, 423–443 (2018).
38. McPhaden, M. J. Tropical Pacific Ocean heat content variations and ENSO persistence barriers. *Geophys. Res. Lett.* **30**, 1480 (2003).
39. Capotondi, A. et al. Understanding ENSO diversity. *Bull. Amer. Meteor. Soc.* **96**, 921–938 (2014).
40. Ashok, K., Behera, S. K., Rao, S. A., Weng, H. & Yamagata, T. El Niño Modoki and its possible teleconnection. *J. Geophys. Res.: Oceans* **112**, C11007. <https://agupubs.onlinelibrary.wiley.com/action/showCitFormats?doi=10.1029%2F2006JC003798> (2007).
41. Chiodi, A. M. & Harrison, D. E. Characterizing warm-ENSO variability in the equatorial Pacific: an OLR perspective. *J. Climate* **23**, 2428–2439 (2010).
42. Yu, J.-Y., Kao, H.-Y., Lee, T. & Kim, S. T. Subsurface ocean temperature indices for Central-Pacific and Eastern-Pacific types of El Niño and La Niña events. *Theor. Appl. Climatol.* **103**, 337–344 (2011).
43. Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C. & Wang, W. An improved in situ and satellite SST analysis for climate. *J. Climate* **15**, 1609–1625 (2002).
44. Huang, B. et al. Extended reconstructed sea surface temperature, version 5 (ERSSTv5): upgrades, validations, and intercomparisons. *J. Climate* **30**, 8179–8205 (2017).
45. IRI Data Library. <http://iridl.ldeo.columbia.edu/SOURCES/NOAA/NCDC/ERSST/version5/sst> (2019).
46. Kirtman, B. et al. The North American Multi-Model Ensemble (NMME): Phase-1 seasonal to interannual prediction, Phase-2 toward developing intra-seasonal prediction. *Bull. Amer. Meteor. Soc.* **95**, 585–601 (2014).
47. Barnston, A. G., Tippett, M. K., Ranganathan, M. & L'Heureux, M. L. Deterministic skill of ENSO predictions from the North American Multimodel Ensemble. *Clim. Dyn.* **53**, 7215–7234 (2019).
48. Tippett, M. K., Ranganathan, M., L'Heureux, M. L., Barnston, A. G. & DelSole, T. Assessing probabilistic predictions of ENSO phase and intensity from the North American Multimodel Ensemble. *Clim. Dyn.* **53**, 7497–7518 (2019).
49. North, G. R., Bell, T. L., Cahalan, R. F. & Moeng, F. J. Sampling errors in the estimation of empirical orthogonal functions. *Month. Weather Rev.* **110**, 699–706 (1982).
50. Dowson, D. C. & Landau, B. V. The Fréchet distance between multivariate normal distributions. *J. Multivar. Anal.* **12**, 450–455 (1982).
51. Murphy, A. H. Skill scores based on the mean square error and their relationships to the correlation coefficient. *Mon. Wea. Rev.* **116**, 2417–2424 (1988).

ACKNOWLEDGEMENTS

M.K.T. acknowledges a Severo Ochoa Mobility grant that supported a visit to the Barcelona Supercomputing Center.

AUTHOR CONTRIBUTIONS

M.K.T. conceived and made the calculations. M.K.T. and M.L.H. wrote the paper.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at <https://doi.org/10.1038/s41612-020-0128-y>.

Correspondence and requests for materials should be addressed to M.K.T.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020