Multi-year statistical prediction of ENSO enhanced by the Tropical Pacific Observing System

Desislava Petrova*

Climate and Health Programme, Barcelona Institute for Global Health, Barcelona, Catalonia, Spain

Joan Ballester

Climate and Health Programme, Barcelona Institute for Global Health, Barcelona, Catalonia, Spain

Siem Jan Koopman

Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Xavier Rodó

Climate and Health Programme, Barcelona Institute for Global Health, Barcelona, Catalonia, Spain

Institució Catalana de Recerca i Estudis Avancats (ICREA), Barcelona, Catalonia, Spain

**Corresponding author address:* Desislava Petrova, Barcelona Institute for Global Health, Carrer del Doctor Aiguader 88, Barcelona, Catalonia, Spain 08003.

E-mail: desislava.petrova@isglobal.org

ABSTRACT

The theoretical predictability limit of El Niño-Southern Oscillation has been shown to be on the order of years, but long-lead predictions of El Niño (EN) and La Niña (LN) are still lacking. State-of-the-art forecasting schemes traditionally do not predict beyond the spring barrier. Recent efforts have been dedicated to the improvement of dynamical models, while statistical schemes still need to take full advantage of the availability of ocean subsurface variables, provided regularly for the last few decades as a result of the Tropical Ocean Global Atmosphere Program (TOGA). Here we use a number of predictor variables, including temperature at different depths and regions of the equatorial ocean, in a flexible statistical dynamic components model to make skilful long-lead retrospective predictions (hindcasts) of the Niño3.4 Index in the period 1970-2016. The model hindcasts the major EN episodes up to two-and-a-half years in advance, including the recent extreme 2015/16 EN. The analysis demonstrates that events are predicted more accurately after the completion of the observational array in the tropical Pacific in 1994, as a result of the improved data quality and coverage achieved by TOGA. Therefore, there is potential to issue long-lead predictions of this climatic phenomenon at a low computational cost.

1. Introduction

Skilful long-range forecasts of El Niño-Southern Oscillation (ENSO) are still in high demand. After decades of extensive efforts, dynamical models nowadays represent the best available tools to issue ENSO forecasts at lead times of up to two seasons, although they are still largely constrained by the lack of complete understanding of the physics of the phenomenon, by problems arising from the initialization of the components of the climate system or by the need for accurate parametrization of important physical processes (Barnston et al. 2012). Statistical models, on the other hand, largely depend on the availability of ocean and atmosphere historical data, so that the longer the length of the data, the more robust is the predictor-predictand relationship identified by the model (Barnston et al. 2012). In addition to these factors, the low signal-to-noise ratio in boreal spring (Sarachik and Cane 2010), the influence of high-frequency atmospheric winds (Fedorov et al. 2003, 2015), as well as the natural irregularity of the climate system (Wittenberg 2009) all limit the long-term dynamical and statistical forecasting of the phenomenon. Some of the classical ENSO theories view the oscillation as self-sustained (Cane et al. 1990; Jin et al. 1994; Jin 1997), and support the claim that it is potentially predictable several years in advance (Cane et al. 1986; Goswami and Shukla 1991; Latif et al. 1999; Chen and Cane 2008; Wittenberg et al. 2014; Gonzalez and Goddard 2016; Luo et al. 2016; DiNezio et al. 2017; Astudillo et al. 2017), but only a handful of studies document such long-lead retrospective forecasts of past events (Latif et al. 1999; Chen et al. 2004; Luo et al. 2008; Izumo et al. 2010; Ludescher et al. 2013, 2014; Petrova et al. 2017; Gonzalez and Goddard 2016; Ramesh et al. 2016; Luo et al. 2017), and most of them use dynamical models. Statistical models are assumed to be less skilful at long lead times, and comparable in performance to dynamical schemes at shorter lead times of about half a year (Barnston 1994; Chen and Cane 2008). To some extent this is explained by the fact that a new generation of statistical models has not been added to the ENSO forecasting plume, while the majority of the old models have not been substantially revised in the recent years, and some since they were created in the 1980s and early 1990s (Barnston et al. 2012).

One of the strongest events on record - the 1982/83 EN - surprised the scientific community (Cane et al. 1986; McPhaden and Yu 1999) as it was neither predicted, nor identified until very late in its development. This triggered a decade-long effort to put in place a monitoring system in the tropical Pacific with the aim of studying ENSO better and improving the predictive capacity of models (McPhaden and Yu 1999), which led to the inauguration of the TOGA research program in 1985 (McPhaden and Yu 1999). It deployed a three-dimensional array in the tropical Pacific that since then regularly samples the subsurface temperature down to 500 metres depth. The number of monthly temperature profiles increased dramatically (Figure S1). The system was completed in 1994, just in time to track the stronger-than-normal trade winds in 1995/96, which generated a buildup of warm waters in the western tropical Pacific more than one year before the peak of the record-breaking 1997/98 EN (McPhaden and Yu 1999). This was the first time when the scientific community and the public could see the benefits of TOGA. A number of studies now fully recognize the fundamental role that the intensification of the trade winds and the subsurface heat buildup in the western equatorial Pacific play in the onset of EN events (Wyrtki 1985; Cane et al. 1986; Jin 1997; Clarke and Van Gorder 2003; McPhaden 2003, 2004; McPhaden et al. 2006; Ramesh and Murtugudde 2013; Ballester et al. 2015; Petrova et al. 2017), and statistical models can benefit from available data to represent in more detail these processes that occur early on in the generation of the events.

In the present study we use an improved version of the flexible statistical dynamic components ENSO model described in Petrova et al. (2017). At long lead times it incorporates predictor variables designed to capture the three-dimensional shape of the warm pool subsurface heat buildup at different depths, as well as zonal wind stress anomalies in the central and western equatorial Pacific (see Methods). The aim is to capture the low-frequency deterministic and state-dependent portions of the variability and coupling between the ocean and atmosphere (Eisenman et al. 2005; Gebbie and Tziperman 2009; Hu and Fedorov 2016; Levine and Jin 2017), from which predictability can be derived (Latif et al. 1999; Chen et al. 2015). The model consists of several stochastic cycle components with frequencies corresponding to the main peaks in the spectrum of the Niño3.4 Index (see Petrova et al. (2017)), as well as predictor regression variables such as sea surface and subsurface temperature and zonal wind stress. These variables enter the model equations in the form of lagged time series with respect to the monthly value of the Niño3.4 index, and are selected to be consistent with the EN dynamical evolution. In this way, different covariates are used for predictions at different lead times, depending on the average temporal progression of EN events. In the present study we show hindcasts from the model, and as is normally the case, a somewhat poorer performance is expected in operational mode. In fact, the model has been operational since 2015, and while it correctly detected the 2015 EN and the mild La Niña (LN) in 2016, it failed to foresee the recent LN in 2017 and predicted neutral conditions instead (not shown). This paper is organized as follows: in Section 2 we describe the data and methods used in the analysis; in Section 3 we present the results and discuss them in Section 4, where we also provide concluding remarks.

2. Data and Methods

The model used in this study is an advanced version of the statistical dynamic components model proposed by Petrova et al. (2017) and developed specifically for prediction of the average sea surface temperature in the Niño3.4 region defined as the box [5°N–5°S, 170°W–120°W]. It is a statistical model that belongs to the class of dynamic components time series models. The distinctive feature of this type of models is that they decompose the time series of interest into dynamic components that represent linear stochastic processes with separate evolutions (Durbin and Koopman 2012). The addition of predictor variables, in this case derived from lead-lag climate composites, is done using regression. We refer to the Appendix for complete and mathematically precise details.

The model first presented in Petrova et al. (2017) is built in terms of two main subgroups of elements. The first subgroup contains the so-called dynamic components, which include a trend (level), a seasonal, and three time-varying cyclical (quasi-periodic) components. The second subgroup contains a number of individually selected predictor variables, which enter the model equation in the form of regressed and lagged time series and will be described later. All of these separate components are added together in a linear fashion to form the final ENSO model given by:

$$\underbrace{y_t}_{\textbf{Niño3.4 index}} = \underbrace{\mu_t}_{\textbf{trend}} + \underbrace{\gamma_t}_{\textbf{seasonality}} + \underbrace{\psi_{1t} + \psi_{2t} + \psi_{3t}}_{\textbf{quasiperiodic cycles}} + \underbrace{X_t\beta}_{\textbf{predictors}} + \underbrace{\varepsilon_t}_{\textbf{noise}}$$

where y_t represents the average monthly temperature in the Niño3.4 region at time t, μ_t is the trend component, γ_t is the seasonal component with 12 seasonal effects (one fixed value for every month of the year), and ψ_{1t}, ψ_{2t} and ψ_{3t} are the stochastic cycle components. $X_t\beta$ is a vector that contains the predictor variables, while ε_t is the noise term in the model.

Here we improve this first version of the model, by replacing the previously fixed seasonal component with two slowly-varying annual and semi-annual periodic components, and also by including one additional time-varying cycle component, so that the new model equation becomes:

$$\underbrace{y_t}_{\textbf{Niño3.4 index}} = \underbrace{\mu_t}_{\textbf{trend}} + \underbrace{\psi_{1t} + \psi_{2t}}_{\textbf{seasonality}} + \underbrace{\psi_{3t} + \psi_{4t} + \psi_{5t} + \psi_{6t}}_{\textbf{quasiperiodic cycles}} + \underbrace{X_t\beta}_{\textbf{predictors}} + \underbrace{\varepsilon_t}_{\textbf{noise}}$$

Thorough information about the different components and how they are modelled and estimated is provided in the Appendix.

The previous model presented in Petrova et al. (2017) used three quasi-periodic cycle components that generally correspond to the near-annual (NA), quasi-biannual (QB) and quasi-quadrennial (QQ) modes of ENSO variability, while here we added one more stochastic cycle component associated with ENSO variability on decadal (D) time scales. In Petrova et al. (2017) we established that this low-frequency variability is important for the simulation of some EN events, and this feature was not explicitly resolved in the previous model version. We have also replaced the fixed seasonal component in the previous version of the model with two new cyclical components bound to annual ($\sim$ 12 months) and semi-annual ($\sim$ 6 months) periodicities. They are allowed to vary slowly over time in order to address the finding in our previous study that the annual frequency of the seasonal component was not sufficiently well-simulated, because the annual periodicity of the Niño3.4 temperature is not strictly fixed at 12 months (Chen et al. 2016), and especially because during EN events the amplitude of the annual cycle is suppressed (Guilyardi 2006). As a result, we have a total of 6 stochastic cycle components in the new model version.

There are also different regression predictors included in the model at different lead times, all selected based on the general evolution of an average EN event. LN is assumed to be symmetrical, although we are aware that important asymmetries exist between the two and this problem will be addressed in future work. In the ocean we used both surface and subsurface temperatures at different depths (between 0 and 500 metres) and regions for the extraction of the predictors. Regions are selected in the western and central equatorial Pacific where the ocean is typically warmed abnormally prior to EN and a heat buildup occurs during the growing and recharge phase (Ballester et al. 2016a; Petrova et al. 2017). Figures S2 and S3, as well as Table 1 show the selected regions and depths considered at different lead times. The selection is based on climate composites of EN events from the period 1978-2012 (also see Petrova et al. (2017)). The sea surface temperature data sets used for the predictors and for the Niño3.4 temperature time series are the NOAA-ERSST-V3 before 1982 and the NOAA-OISST-V2 thereafter (www.esrl.noaa.gov/psd/). The subsurface temperature data set used for the subsurface ocean predictors is the Subsurface Temperature and Salinity Analyses by Ishii et al. (2005) archived at (https://rda.ucar.edu/datasets/) before 2012 and the Hadley Centre EN4.0.2 analyses data thereafter (Good et al. 2013). In the atmosphere three different regions are used to extract zonal wind stress predictors for the model. The three regions are again located in the western and central equatorial Pacific (see Figure S4 and Table 1) and the data set is the NCEP/NCAR Reanalysis (Kalnay et al. 1996).

During forecasting, the dynamic components (especially the stationary cycles) have larger weights for the mid- and short-term forecasts, while the impact of the predictors remains the same for short- and long-term forecasts. Hence, the predictors become relatively more important for long-term forecasting (also see the Appendix for more information). Importantly, the predictor variables also affect the estimation of the cycle components parameters for each forecast. Parameter estimation relies on the Kalman filter methods (Kalman 1960; Harvey 1989) and on state space methods (Durbin and Koopman 2012).

Results in Figure 3 are obtained as follows: the Niño3.4 predictions in the period 1972–1993 are based on parameter estimates (calibration process) from data in the period 1952–1970, while the predictions in the period 1994–2015 are based on parameter estimates from data in the period 1974–1992. In this way, to avoid the heavy computations, we have produced the predictions in Figure 3 using a pre-fixed period for calibration purposes. In comparison, the predictions (including parameter estimation) presented in Figures 1, 2 and 4 are based on the observations available before each starting prediction point. Still, data for the prediction estimations was progressively excluded, in order to include only the more recent samples and discard earlier data of assumingly lesser quality. Thus, predictions up to 1990 were made using the data from 1952 onwards, predictions between 1991-1996 were made using the data from 1970 onwards, while predictions thereafter were made using the data from 1982 onwards.

A limitation of the study is that the performed predictions are not operational, as they are based on retrospective hindcasting experiments. Our system also strongly relies on the model variability skeleton, contributed, among others, by different cyclical components. However, all ENSO forecasting systems, including operational dynamical models, implicitly or explicitly rely on intrinsic ENSO variability generated at the cyclical low-frequency modes for prediction (Kirtman and Schopf 1998). As an example, we include the spectrum of Niño3.4 from a long (500 years) spin-up simulation with the GFDL CM2.1, which is one of the operational models for ENSO prediction, in order to compare it with the spectrum of both the Niño3.4 observations and their predictions with the model proposed here (Figure S5). What can be clearly noticed is that the power density is distributed similarly in all cases, with main peaks corresponding to the NA, QB, QQ and D modes of variability, respectively, also used as cyclical time-varying components in our model.

3. Results

The observed and hindcast monthly Niño3.4 anomalies at 6 and 24 months lead time are presented in Figure 1. The 6-month lead hindcast predicts the timing and magnitude of all EN and LN events, and no false alarms are generated ($RMSE = 0.54$; Figure 1a). Since an ENSO event is typically already under-way half a year before its peak in December-January-February (DJF), the majority of the operational forecasting schemes are able to produce accurate predictions at this lead time (Barnston et al. 2012). The 24-month lead hindcast, and in general any lead time hindcast beyond the spring barrier (i.e. from 8 months onward; not shown), generally reproduces the crests and troughs in the time series ($RMSE = 0.62$; Figure 1b). However, for the period before the prominent 1997/98 EN, we find that the predicted amplitudes of the larger events are notably smaller than the observed and sometimes an event is hardly or not detected. We highlight that this cannot be explained by a change in the interannual ENSO activity in the different time periods, as three sizeable EN (1972/73, 1982/83, 1986/87 and 1997/98, 2009/10, 2015/16) and LN (1973/74, 1975/76, 1987/88 and 1998/00, 2007/08, 2010/11) episodes have occurred before and after 1994 (CPC 2016). In addition, it cannot be simply attributed to the design of the model and the predictor variables used, because EN events from both periods were considered for the composites on which the selection of predictor variables was based (see Petrova et al. (2017) for details).

To characterize better the difference between periods, Figure 2 displays the regressions between the observations and hindcasts for two consecutive 22-year sub-periods (1972-1993 in blue and 1994-2015 in red) at 6- and 24-month lead. No substantial difference is observed between the slopes of the regression lines for the two periods at the shorter lead time ($regr_{1972-1993} = 0.65$, $t = 23.88$, $regr_{1994-2015} = 0.74$, $t = 27.34$, $p < 0.001$; Figure 2a), indicating that the model performance is comparable. Conversely, the regression coefficients significantly increase for the long-range hindcasts made after 1994 ($regr_{1972-1993} = 0.35$, $t = 17.12$, $regr_{1994-2015} = 0.65$, $t = 30.93$, $p < 0.001$; Figure 2b), which represents a major improvement in the capacity of the model. The change in the overall similarity between the observations and the hindcasts at 24-month lead time is also assessed by the sixteen-year moving root mean square error (RMSE) shown in Figure 1c. The RMSE decreases monotonically with time until the early 1990s and then stays relatively constant afterwards. At the same time, data availability was constantly improving during TOGA, until the tropical Pacific network array of moorings was fully into place at the end of the program in 1994 (McPhaden et al. 1998).

To further explore the difference in the model performance over the two periods, Figure 3 shows correlations and root mean square errors for the whole range of lead times up to 24 months. For lead times of about 2 seasons both the correlations and RMSE are similar among periods, while for lead times beyond 6 months they start to diverge. We also observe that correlations and RMSE stay relatively constant beyond this lead time. One possible reason for this stable behaviour is that the stochastic quasi-periodic cycles are the main contributors to the skill of the predictions and their unknown parameters are estimated similarly by the Kalman filter at different lead times beyond 2 seasons. Generally, the skill is derived from information about subsurface heat anomalies of approximately the same intensity, and the different cycles capture similar oscillation phases. Previous studies (Chen et al. 1995, 2004; Chen and Cane 2008; Stockdale et al. 2011; Duan et al. 2016; Lee et al. 2018) have already concluded that the spring predictability barrier may not be an intrinsic barrier to the system itself, but it could rather depend on model skill, observational data availability, especially in the subsurface western tropical Pacific (Lee et al. 2018), and precursors used. Warm water volume (WWV) in the tropical Pacific as a predictor (i.e. subsurface information) is not associated with a spring persistence barrier, and its correlation with the Niño3.4 is above 0.7 for the February-April season when SST anomalies have the lowest correlation (McPhaden 2003). Here we add evidence to such claims, as we also find that the drop in forecast skill is slow and gradual for longer-lead predictions than a couple of seasons (Figure 3).

The statistical model we use is linear, and while its stochastic cyclical components are mainly responsible for capturing the correct phase of the oscillation, the lagged predictor variables are expected to contribute to the correct forecasts of the amplitudes of the events, especially at longer lead times (see Methods, the Appendix and Petrova et al. (2017) for details). Below we analyse if the predictor variables add significantly to the EN hindcasts of the earlier period, which also coincides with a time when no regular subsurface temperature and wind stress data were being provided yet (McPhaden et al. 1998).

The hindcasts at several lead times of the strongest EN events in the study period (see CPC (2016)) are displayed in Figure 4. In all cases the model is capable of detecting a warming 29 months in advance (magenta curve), although there are evident errors in the amplitude and timing in some cases. A much better representation of the amplitudes in the long-lead hindcasts of the events in the second period (1997/98, 2009/10 and 2015/16), as compared to those occurring in the first period (1972/73, 1982/83 and 1986/87), is also clearly visible in the figure. The estimated coefficients and the corresponding t- and p-values for the predictor variables used in the 24-month lead predictions of all the warm events in the study period are listed in Table 2. Remarkably, none of the three predictor variables is found to be significant at the 90% level for the hindcasts of any of the events before 1994, while there is at least one significant variable for each hindcast of the episodes that occurred afterwards. Similar results hold for the other long-lead predictions shown in Figure 4 (Tables S1 and S2).

4. Discussion and Conclusions

We demonstrated that the Tropical Pacific Observing System, and especially the provision of subsurface temperature data on a regular basis, has a vital contributing role (Newman et al. 2011) for the long-lead predictive capabilities of the model proposed here. With the end of TOGA in 1994 nearly the whole equatorial band between 10°N-10°S was covered with moorings (McPhaden et al. 1998), and this is also the start of altimetry data (Stockdale et al. 2011). As can be seen from the Tropical Atmosphere Ocean-Triangle Trans-Ocean Buoy Network (TAO-TRITON) array development (NOAA 2018a), some subsurface data from the central Pacific was already streamed at the end of 1987, while at the end of 1991 data was also coming in from the western Pacific, which represents a key region for the forecast of the phenomenon at lead times beyond the spring barrier. Thus, almost three decades have passed since the three-dimensional observations began in the tropical Pacific, and the limited span of the data is now less of a problem for the robust definition of statistical predictive schemes (Barnston et al. 2012).

As seen in the previous section, there is a well-defined shift between the lack of significance of the predictor variables for the hindcasts of the warm events before the end of TOGA and their significance thereafter. Our results strongly support the view that the improved hindcasts are due to the availability of regular and higher resolution subsurface data ensured by the implementation of the observational network array (NOAA 2018b). This is also confirmed by Figure S6, which shows the same hindcasts as in Figure 4, but made without the inclusion of the predictor variables in the model framework. The lack of predictors in the model results into a clear deterioration of the hindcasts of the EN events from the period after 1994, but in no substantial difference in the hindcasts of the events from the earlier period (also see Table S3).

The correct and relevant subsurface information also has implications for the forecasting of the magnitudes of the warm events (Ballester et al. 2016b, 2017). In the linear framework of the model that we use, at the longer lead times the predictor variables have more forecast weight than they do at the shorter lead times (see Methods and the Appendix). The predicted amplitudes of the three earlier events shown in Figure 4a-c do not exceed 1.5°C at the long lead times of 21 and 29 months (green and magenta curves). At the same time, the predicted amplitudes of the three events that took place in the later period, when the predictor variables are shown to have an impact (Table 2), are consistent with the occurrence of a strong EN event (green and magenta curves in Figure 4d-f). Some of the underestimation of the amplitudes of the events predicted at long lead times is also due to the stochastic noise component of the zonal wind (Penland 1996;

Hu and Fedorov 2016; Levine and McPhaden 2016) as more extreme EN events have been found to result from more intense and frequent westerly wind bursts (Chen et al. 2015). Additionally, in the case of the 1982/83 EN anomalous solar radiation and suppressed convection may have played a more decisive role, setting this event apart from the others (Kim and An 2018) and making the predictors used here at long lead times less relevant (Kirtman and Zebiak 1997).

In essence, the same conclusion as the one reached here has been made by Stockdale et al. (2011), where a large reduction of the errors in Niño3.4 SST forecasts made after 1994 is detected with the European Centre for Medium-Range Weather Forecasts (ECMWF) Seasonal Forecast System 3. The results are also in agreement with an earlier study with the same system (Balmaseda and Anderson 2009), in which the effect of ARGO floats are removed from the observations, and it is established that improvements in the forecast are clearly explained by the improved observing system. Further, it was found that the information from the mooring array is the main contributor for the increased skill of the prediction system in the equatorial Pacific region. In addition, McPhaden et al. (2006) using an empirical ENSO model with two predictors - WWV in the equatorial Pacific and an index of the Madden-Jullian Oscillation - documents much better estimations of the Niño3.4 after 1995, with lower than observed amplitudes before this year, just as it is in the results presented here. Similarly, the authors attribute the improvement to the better observations after the placement of the TAO array.

Conversely, a more recent study (Kumar et al. 2015) concluded that the increase of the number of observations after 1994 did not result in a clear improvement of the prediction skill of the National Center for Environmental Prediction (NCEP) System 2. We note, however, that our results are not directly comparable, because the forecasts discussed therein are performed at up to 6 months lead time, when essentially an SST anomaly signature of a developing EN or LN event is already present in the eastern equatorial Pacific, and subsurface information is generally not as crucial as it is at the longer lead times discussed here. The authors themselves admit that the evolution of the ocean-atmosphere system at this short lead is affected much more by the surface wind and ocean circulation feedbacks. SST is in fact found to be a more useful predictor for forecasts started 6 months before the event than WWV (McPhaden 2003).

Some of the existing statistical systems already include measures of integrated equatorial heat content (Barnston et al. 2012). However, our model uses temperature data from a selection of dynamically relevant regions and depths to maximize its predictive power. These values may not always be well-represented by spatially-integrated measures of heat content, and our analysis suggests that the integration sometimes masks the intensity of the heat buildup in specific regions in the subsurface at long lead times, and more importantly, does not allow the systems to properly track the eastward propagation of heat along the equatorial thermocline (Ballester et al. 2015; Petrova et al. 2017). WWV anomalies along the whole equatorial Pacific present in late boreal winter and spring (February-May) are persistent until next boreal winter, but those in early boreal winter are not. Hence, as a predictor it could extend the lead time to about a year in advance, but not more (Izumo et al. 2018). Alternatively, WWV calculated only in the western equatorial Pacific is significantly correlated with Niño3 SST anomalies for much longer lead times of more than 20 months (McPhaden 2003), and is a significantly better predictor beyond the spring barrier (Izumo et al. 2018). Sea surface height (SSH), on the other hand, is also not always representative of the heat accumulation in the warm pool, because sometimes positive and negative heat anomalies exist at different depths of the water column near the thermocline, and the net result is a lack of a prominent SSH anomaly (see Figures S7 and S8). The combination of the memory effect represented by subsurface information, weakly-varying seasonality and nonlinearity, on the other hand, could be sufficient for reproducing the overall ENSO variability (Chen et al. 2016), and our model design attempts to incorporate these particular effects.

Although there is a marked difference in the predictive capacity of the model during the earlier and later sub-periods, it still exhibits high skill (i.e. correlations and RMSE in Figure 3) in both periods. We conclude that statistical models should be improved in the direction of using the available subsurface information that is fundamental for ENSO in a more discrete and targeted way, so that they can provide early and useful information about EN and LN events to decision makers around the world, which could prevent threats to human lives and reduce economic costs.

Acknowledgments. J.B. gratefully acknowledges funding from the European Commission through a Marie Curie International Outgoing Fellowship (project MEMENTO from the FP7-PEOPLE-2011-IOF call), and from the European Commission and the Catalan Government through a Marie Curie - Beatriu de Pinós Fellowship (project 00068 from the BP-DGR-2014-B call).

X.R. gratefully acknowledges funding from the Daniel Bravo Foundation (project WINDBIOME), as well as from the PERIS PICAT project (SLT002/16/466) and the NEW INDIGO project (PCIN-2013-038).

APPENDIX

Model Description

The most basic version of the class of dynamic components time series models is the local level model for a univariate time series y_t and is given by $y_t = \mu_t + \varepsilon_t$ where μ_t is a linear stochastic process, dynamically evolving over time, and ε_t is a noise term. We can consider μ_t to follow a random walk process that captures the long-term trend features in the time series and ε_t to be an independent and identically distributed (IID) variable that represents the short-term deflections from the trend or the noise in the time series. The trend signal μ_t is the key feature of interest in the local level model, also for generating long- or medium-term forecasts of y_t. Its random walk process is given by $\mu_{t+1} = \mu_t + \eta_t$ where the noise term η_t is IID. Under the assumption that both noise terms ε_t and η_t are normally distributed with mean zero and variances σ_ε^2 and σ_η^2, respectively, the celebrated Kalman filter equations (Kalman 1960) compute the minimum mean squared error (MMSE) estimates of μ_t given realizations for $y_1, y_2, \ldots, y_t$, in a recursive real-time fashion, for $t = 1, 2, \ldots, T$ where T is the length of the time series under investigation. The estimate of μ_t can be expressed as the weighted average $\sum_{j=0}^{T} w_j y_{t-j}$ where the weights are normalised (they sum up to unity, $w_0 + w_1 + w_2 + \ldots = 1$), are exponentially decaying and are a function of the signal-to-noise ratio $q = \sigma_\eta^2 / \sigma_\varepsilon^2$. When q is relatively large (σ_ε^2 is small relative to σ_η^2, implying that y_t behaves close to a random walk process as $y_t \approx \mu_t$), the weights are decaying fast to zero and we obtain a ”noisy” estimate of μ_t. This estimate may be representative as it is close to the local level (small estimation bias), but given that only a few observations are used for the estimation, the precision is typically small (i.e. large estimation variance). When q is relatively small (σ_η^2 is small relative to σ_ε^2, implying that μ_t is evolving slowly over time as $\mu_{t+1} \approx \mu_t$), the weights are decaying slowly to zero and we obtain a smooth estimate of μ_t. In the latter case, the estimation bias may be larger (less local targeting), but the estimation variance is smaller since more observations are used for estimation. The appropriate value for the signal-to-noise ratio q for a particular time series depends on the dynamic features of the time series. We estimate q by the method of maximum likelihood, which entails the numerical maximization of the likelihood function that is computed using the Kalman filter (for a specific value of q). The h-step ahead forecasting (that is the estimation of y_{T+h}, given realizations for $y_1, y_2, \ldots, y_T$) is also computed by the forward-moving Kalman filter (from 1 to T). The estimation methodology provides the MMSE optimal weights for forecasting: the forecasting weights gradually decline when observations are increasingly remote from the forecast point as these become increasingly less relevant. The estimation of all μ_t's given the realizations $y_1, y_2, \ldots, y_T$ (all data) is referred to as signal extraction and relies on Kalman smoothing which is a backward-moving filter (from T to 1); see Durbin and Koopman (2012, Chapter 2) with all methods for filtering, forecasting, signal extraction and parameter estimation, and with related details for the local level model.

The local level model is a special case of the dynamic components model adopted in Petrova et al. (2017) where the observation equation $y_t = \mu_t + \varepsilon_t$ is extended with regression effects (predictors) and more linear stochastic processes that represent key dynamic features of the Niño3.4 temperature time series including seasonal and cyclical effects. The model then becomes $y_t = \mu_t + X_t \beta + \sum_{i=1}^{M} \psi_{it} + \varepsilon_t$ where X_t is the exogenous $1 \times K$ vector of covariates (or predictor variables) measured at time t, β is the $K \times 1$ vector of predictor coefficients, ψ_{it} is the ith dynamic cycle component which is modelled as a stationary process for $i = 1, \ldots, M$, where M is the number of cycles in the model (in our case $M = 6$). The model specification for the cycle component is given by $\psi_{i,t+1} = \rho_i \cos(\lambda_i) \psi_{it} + \rho_i \sin(\lambda_i) \psi_{it} + \omega_{it}$ with the auxiliary dynamic process given by $\psi^{\dagger}_{i,t+1} = \rho_i \cos(\lambda_i) \psi^{\dagger}_{it} - \rho_i \sin(\lambda_i) \psi^{\dagger}_{it} + \omega^{\dagger}_{it}$ where ρ_i is the autoregressive coefficient (determines the persistence of the cycle process), λ_i is the frequency of the cycle measured in radians, and ω_{it} and $\omega_{it}^{\dagger}$ are two IID noise terms which are independent of each other, and all other noise terms, but they have the same (common) variance $\sigma^2_{\omega,i}$, for $i = 1,\ldots,M$. It can be shown that we can formulate ψ_{it} as a stationary autoregressive moving average (ARMA) process. As long as the coefficient pairs (ρ_i, λ_i) are sufficiently different for different $i = 1,\ldots,M$, the M cycle components $\psi_{1,t},\ldots,\psi_{M,t}$ can be uniquely extracted from the observed time series y_t. The parameter constraints for each cycle process are $0 < \rho_i < 1$ (stationarity), $0 < \lambda_i < 2\pi$ (circularity) and $\sigma^2_{\omega,i} > 0$, for $i = 1,\ldots,M$. The signal-to-noise coefficient for the ith cycle process is given by $q_{\psi,i} = \sigma^2_{\omega,i} / \sigma^2_{\varepsilon}$, for $i = 1,\ldots,M$. The complete model for y_t (in our case for the Niño3.4 temperature time series) can be represented as a linear Gaussian state space model such that the Kalman filter methods can be used in a similar way as for the local level model. The dynamic level and cycle (including the auxiliary cycle variables) components are placed in the state vector, denoted by α_t, which is subject to a multivariate dynamic stochastic process. The predictor coefficients in vector β are treated as time-invariant, fixed parameters. Both α_t and β are simultaneously estimated as part of the Kalman filter (see Harvey (1989) and Durbin and Koopman (2012, Part I) for its general treatment). Also in this more general context of the state space model, the Kalman filter methods remain to provide the MMSE optimal weights to the observations for signal extraction and forecasting.

The statistical dynamic components model can be viewed as a linear time series model with time-varying parameters. The introduction of time-varying parameters as done with stochastically evolving level and cycle components can address and approximate non-linear features in the time series via piece-wise linearization. The number of nodes for the linearization (or the smoothness of the piece-wise approximation) is implicitly determined via the signal-to-noise parameters of the time-varying components. We therefore may claim that the introduction of the dynamic components also make the analysis more robust to non-linear features in the time series.

References

Astudillo, H., R. Abarca-del Río, and F. Borotto, 2017: Long-term potential nonlinear predictability of El Niño–La Niña events. *Climate Dynamics*, **49**, 131–141.

Ballester, J., S. Bordoni, D. Petrova, and X. Rodó, 2015: On the dynamical mechanism explaining the western Pacific subsurface temperature buildup leading to ENSO events. *Geophysical Research Letter*, **42**, 2961–2967.

Ballester, J., S. Bordoni, D. Petrova, and X. Rodó, 2016a: Heat advection processes leading to El Niño events as depicted by an ensemble of ocean assmiliation products. *Journal of Geophysical Research: Oceans*.

Ballester, J., D. Petrova, S. Bordoni, B. Cash, M. García-Díez, and X. Rodó, 2016b: Sensitivity of El Niño intensity and timing to preceding subsurface heat magnitude. *Scientific Reports*.

Ballester, J., D. Petrova, S. Bordoni, B. Cash, and X. Rodó, 2017: Timing of subsurface heat magnitude for the growth of El Niño events. *Geophysical Research Letters*.

Balmaseda, M., and D. Anderson, 2009: Impact of initialization strategies and observations on seasonal forecast skill. *Geophysical Research Letters*, **36**.

Balmaseda, M., M. Davey, and D. Anderson, 1995: Decadal and seasonal dependence of ENSO prediction skill. *Journal of Climate*, **8**, 2705–2715.

Barnston, A., 1994: Long-lead seasonal forecasts: where do we stand? *Bulletin of the American Meteorological Society*, **75**, 2097–2114.

Barnston, A., M. Tippett, M. L′Heureux, S. Li, and D. DeWitt, 2012: Skill of real-time seasonal ENSO model predictions during 2002-11. Is our capability increasing? *Bulletin of the American Meteorological Society*, **93**, 631–651.

Cane, M., M. Munnich, and S. Zebiak, 1990: A study of self-excited oscillations of the tropical ocean-atmosphere system. Part 1: linear analysis. *Journal of Atmospheric Sciences*, **47**, 1562–1577.

Cane, M., S. Zebiak, and S. Dolan, 1986: Experimental forecasts of El Niño. *Nature*, **321**, 827–832.

Chen, C., M. Cane, N. Henderson, E. Dong, D. Chapman, D. Kondrashov, and M. Chekroun, 2016: Diversity, nonlinearity, seasonality, and memory effect in ENSO simulation and prediction using empirical model reduction. *Journal of Climate*, **29**, 1809–1830.

Chen, D., M. Cane, A. Kaplan, S. Zebiak, and D. Huang, 2004: Predictability of El Niño over the past 148 years. *Nature*, **428**, 15.

Chen, D., and M. A. Cane, 2008: El Niño prediction and predictability. *Journal of Computational Physics*, **227**, 3625–3640.

Chen, D., S. Zebiak, A. Busalacchi, and M. Cane, 1995: An improved procedure for El Niño forecasting: implications for predictability. *Science*, **269**, 1699–1702.

Chen, D., and Coauthors, 2015: Strong influence of westerly wind bursts on El Niño diversity. *Nature Geoscience*, **8**, 339–345.

Clarke, A., and S. Van Gorder, 2003: Improving El Niño prediction using space-time integration of Indo-Pacific winds and equatorial Pacific upper ocean heat content. *Geophysical Research Letters*, **30**, 1944–8007.

CPC, 2016: Cold and warm episodes by season. *http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml*.

DiNezio, P. N., C. Deser, Y. Okumura, and A. Karspeck, 2017: Predictability of 2-year La Niña events in a coupled general circulation model. *Climate Dynamics*, **49**, 4237–4261.

Duan, W., P. Zhao, J. Hu, and H. Xu, 2016: The role of nonlinear forcing singular vector tendency error in causing the spring predictability barrier for ENSO. *Journal of Meteorological Research*, **30**, 853–866.

Durbin, J., and S. J. Koopman, 2012: *Time Series Analysis by State Space Methods*. 2nd ed., Oxford University Press.

Eisenman, I., L. Yu, and E. Tziperman, 2005: Westerly wind bursts: ENSO's tail rather than the dog? *Journal of Climate*, **18**, 5224–5238.

Fedorov, A., S. Harper, S. G. Philander, B. Winter, and A. Wittenberg, 2003: How predictable is El Niño? *Bulletin of the American Meteorological Society*, **84**, 911–919.

Fedorov, A., S. Hu, M. Lengaigne, and E. Guilyardi, 2015: The impact of westerly wind bursts and ocean initial state on the development, and diversity of El Niño events. *Climate Dynamics*, **44**, 1381–1401.

Gebbie, G., and E. Tziperman, 2009: Predictability of SST-modulated westerly wind bursts. *Journal of Climate*, **22**, 3894–3909.

Gonzalez, P., and L. Goddard, 2016: Long-lead ENSO predictability from CMIP5 decadal hindcasts. *Climate Dynamics*, **46**, 3127–3147.

Good, S. A., M. J. Martin, and N. A. Rayner, 2013: EN4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, **118**, 6704–6716.

Goswami, B. N., and J. Shukla, 1991: Predictability of a coupled ocean-atmosphere model. *Journal of Climate*, **4**, 3–22.

Guilyardi, E., 2006: El Niño-mean state-seasonal cycle interactions in multi-model ensemble. *Climate Dynamics*, **26**, 329–348.

Harvey, A., 1989: *Forecasting, structural time series models and the Kalman Filter*. Cambridge University Press.

Hu, S., and A. Fedorov, 2016: Exceptionally strong easterly wind burst stalling El Niño of 2014. *Proceedings of the National Academy of Sciences*, **113**, 2005–2010.

Ishii, M., A. Shouji, S. Sugimoto, and T. Matsumoto, 2005: Objective analyses of SST and marine meteorological variables for the 20th century using COADS and the Kobe Collection. *International Journal of Climatology*, **25**, 865–879.

Izumo, T., M. Lengaigne, J. Vialard, I. Suresh, and Y. Planton, 2018: On the physical interpretation of the lead relation between Warm Water Volume and the El Niño Southern Oscillation. *Climate Dynamics*, **https://doi.org/10.1007/s0038**.

Izumo, T., and Coauthors, 2010: Influence of the state of the Indian Ocean Dipole on the following years El Niño. *Nature Geoscience*, **3**, 168.

Jin, F. F., 1997: An equatorial ocean recharge paradigm for ENSO. Part I: conceptual model. *Journal of Atmospheric Sciences*, **54**, 811–829.

Jin, F. F., J. D. Neelin, and M. Ghil, 1994: El Niño on the devil's staircase - annual subharmonic steps to chaos. *Science*, **264**, 70–72.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *Journal of Basic Engineering , Transactions, ASMA, Series D*, **82**, 35–45.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77**, 437–471.

Kim, J. W., and S. An, 2018: Origin of early-spring central Pacific warming as the 1982-1983 El Niño precursor. *International Journal of Climatology*.

Kirtman, B. P., and P. S. Schopf, 1998: Decadal variability in ENSO predictability and prediction. *Journal of Climate*, **11**, 2804–2822.

Kirtman, B. P., and S. Zebiak, 1997: ENSO simulation and prediction with a hybrid coupled model. *Monthly Weather Review*, **125**, 2620–2641.

Kumar, A., M. Chen, and Y. Xue, 2015: An analysis of temporal evolution of ENSO prediction skill in the context of the equatorial Pacific Ocean observing system. *Monthly Weather Review*, **143**, 3204–3213.

Latif, M., and Coauthors, 1999: A review of the predictability and prediction of ENSO. *Journal of Geophysical Research*, **103**, 14 375–14 393.

Lee, H., A. Kumar, and W. Wang, 2018: Effects of ocean initial perturbation on developing phase of ENSO in a coupled seasonal prediction model. *Climate Dynamics*, **50**, 1747–1767.

Levine, A., and F. Jin, 2017: A simple approach to quantifying the noise-ENSO interaction. Part I: deducing the state-dependency of the windstress forcing using monthly mean data. *Climate Dynamics*, **48**, 1–18.

Levine, A., and M. McPhaden, 2016: How the July 2014 easterly wind burst gave the 2015-2016 El Niño a head start. *Geophysical Research Letters*, **43**, 6503–6510.

Ludescher, J., A. Gozolchiani, M. I. Bogachev, A. Bunde, S. Havlin, and H. J. Schellnhuber, 2013: Improved El Niño forecasting by cooperativity detection. *Proceedings of the National Academy of Sciences*, **110**, 11 742–11 745.

Ludescher, J., A. Gozolchiani, M. I. Bogachev, A. Bunde, S. Havlin, and H. J. Schellnhuber, 2014: Very early warning of next El Niño. *Proceedings of the National Academy of Sciences*, **111**, 2064–2066.

Luo, J., G. Liu, H. Hendon, O. Alves, and T. Yamagata, 2017: Inter-basin sources for two-year predictability of the multi-year La Niña event in 2010–2012. *Scientific Reports*, **7**, 2276.

Luo, J., S. Masson, S. K. Behera, and T. Yamagata, 2008: Extended ENSO predictions using a fully coupled ocean–atmosphere model. *Journal of Climate*, **21**, 84–93.

Luo, J., C. Yuan, W. Sasaki, S. K. Behera, Y. Masumoto, T. Yamagata, J. Lee, and S. Masson, 2016: Current status of intraseasonal–seasonal-to-interannual prediction of the Indo-Pacific climate (in Indo-Pacific climate variability and predictability). 63–107.

McPhaden, M., 2003: Tropical Pacific Ocean heat content variations and ENSO persistence barriers. *Geophysical Research Letters*, **30**, 1480.

McPhaden, M., 2004: Evolution of the 2002/2003 El Niño. *American Meteorological Society*, **85**, 677–695.

McPhaden, M., and X. Yu, 1999: Equatorial waves and the 1997/98 El Niño. *Geophysical Research Letters*, **26**, 2961–2964.

McPhaden, M., X. Zhang, H. Hendon, and M. Wheeler, 2006: Large scale dynamics and MJO forcing of ENSO variability. *Geophysical Research Letters*, **33**, 16.

McPhaden, M., and Coauthors, 1998: The Tropical Ocean-Global Atmosphere observing system: A decade of progress. *Journal of Geophysical Research: Oceans*, **103**, 14 169–14 240.

Newman, M., M. A. Alexander, and J. D. Scott, 2011: An empirical model of tropical ocean dynamics. *Climate Dynamics*, **37**, 1823.

NOAA, 2018a: TAO Array Development. *http://www.pmel.noaa.gov/gtmba/wmo-numbers-0*.

NOAA, 2018b: TAO-TRITON Array Figure. *http://www.pmel.noaa.gov/gtmba/taotriton-map*.

Penland, C., 1996: A stochastic model of Indo-Pacific sea surface temperature anomalies. *Physica D*, **98**, 534–558.

Petrova, D., S. Koopman, J. Ballester, and X. Rodó, 2017: Improving the long-lead predictability of El Niño using a novel forecasting scheme based on a dynamic components model. *Climate Dynamics*, **48**, 1249–1276.

Ramesh, N., M. Cane, R. Seager, and D. Lee, 2016: Predictability and prediction of persistent cool states of the Tropical Pacific Ocean. *Climate Dynamics*, **49**, 2291–2307.

Ramesh, N., and R. Murtugudde, 2013: All flavours of El Niño have similar early subsurface origins. *Nature Climate Change*, **3**, 42–46.

Sarachik, E., and M. Cane, 2010: *The El Niño Southern Oscillation Phenomenon*. Cambridge University Press.

Stockdale, T., and Coauthors, 2011: ECMWF Seasonal Forecast System 3 and its prediction of sea surface temperature. *Climate Dynamics*, **37**, 455–471.

Wittenberg, A., 2009: Are historical records sufficient to constrain ENSO simulations. *Geophysical Research Letters*, **36**, L12 702.

Wittenberg, A. T., A. Rosati, T. L. Delworth, G. A. Vecchi, and F. Zeng, 2014: ENSO modulation: Is it decadally predictable? *Journal of Climate*, **27**, 2667–2681.

Wyrtki, K., 1985: Water displacements in the Pacific and the genesis of El Niño cycles. *Journal of Geophysical Research*, **90**, 7129–7132.

Zebiak, S. E., and M. A. Cane, 1987: A model El Niño-Southern Oscillation. *Monthly Weather Review*, **115**, 2262–2278.

Predictor variable	**Region I**	**Region II**	**Region III**
Zonal Wind Stress	$(180e-220e)$x$(4s-4n)$	$(180e-210e)$x$(10s-0)$	$(160e-200e)$x$(0-10n)$
Sea Surface Temperature	$(140e-160e)$x$(5s-5n)$	$(140e-180e)$x$(10s-5n)$	$(120e-170e)$x$(10s-5n)$
Subsurface Temperature	$(120e-140e)$x$(10s-7n)$	$(150e-200e)$x$(10s-7n)$	$(140e-210e)$x$(5n-10n)$

TABLE 1: Regions over which wind stress and temperature variables are averaged to calculate predictors used in the ENSO model.

El Niño event	**250m. RI**	**300m. RI**	**400m. RI**
1972/73			
Coefficient	0.12	−0.17	−0.29
t	0.78	−0.82	−0.86
p	0.43	0.41	0.39
1982/83			
Coefficient	0.09	0.01	0.20
t	0.78	0.03	0.90
p	0.43	0.97	0.36
1986/87			
Coefficient	−0.03	−0.12	−0.02
t	−0.30	−0.88	−0.09
p	0.76	0.37	0.92
1991/92			
Coefficient	0.07	−0.09	0.09
t	0.64	−0.56	0.38
p	0.52	0.57	0.70
1997/98			
Coefficient	**0.24**	0.35	0.46
t	**1.61**	1.52	1.46
p	**0.10**	0.12	0.14
2002/03			
Coefficient	**0.21**	0.31	0.38
t	**1.67**	1.57	1.44
p	**0.09**	0.11	0.15
2006/07			
Coefficient	**0.23**	**0.32**	**0.43**
t	**2.07**	**1.80**	**1.75**
p	**0.04**	**0.07**	**0.08**
2009/10			
Coefficient	**0.17**	0.24	**0.46**
t	**1.68**	1.46	**1.95**
p	**0.09**	0.14	**0.05**
2014/15			
Coefficient	**0.15**	**0.25**	0.34
t	**1.61**	**1.63**	1.59
p	**0.10**	**0.10**	0.11
2015/16			
Coefficient	0.14	**0.28**	0.32
t	1.55	**1.85**	1.56
p	0.12	**0.06**	0.12

TABLE 2: Coefficients, t-values and p-values for **subsurface temperature** predictor regression variables at **24-month** lead. Values significant at the 90% level are bold.

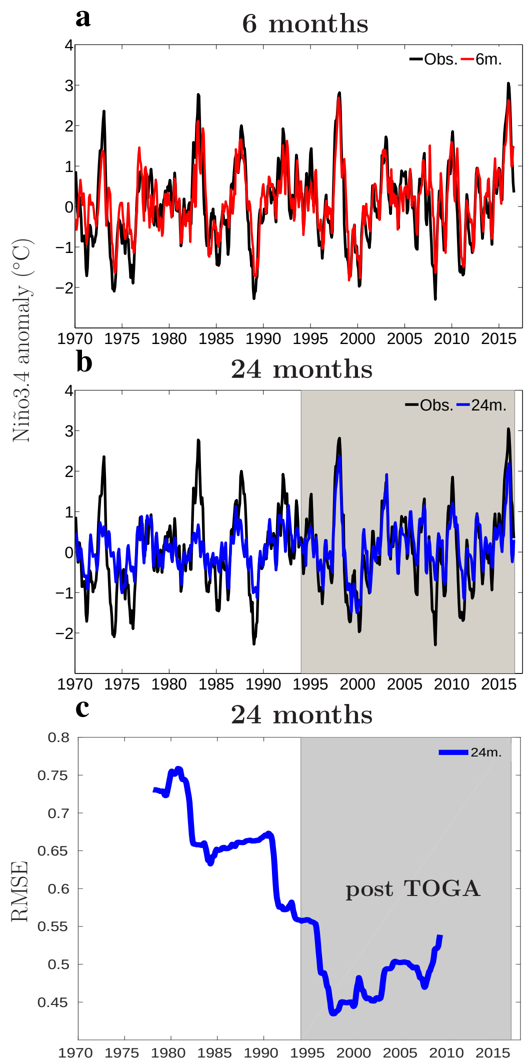
[FIG. 1 about here.]

[FIG. 2 about here.]

[FIG. 3 about here.]

[FIG. 4 about here.]

LIST OF FIGURES



608 FIG. 1. **Retrospective prediction of the Niño3.4 Index.** Monthly observations (black curve) and model
609 prediction at **a**, 6-month lead (red curve) and **b**, 24-month lead (blue curve). **c**, 16-year moving root mean
610 square error (RMSE) of the prediction in (**b**) (blue curve) before and after (shading) the completion of the
611 Observing System in 1994.

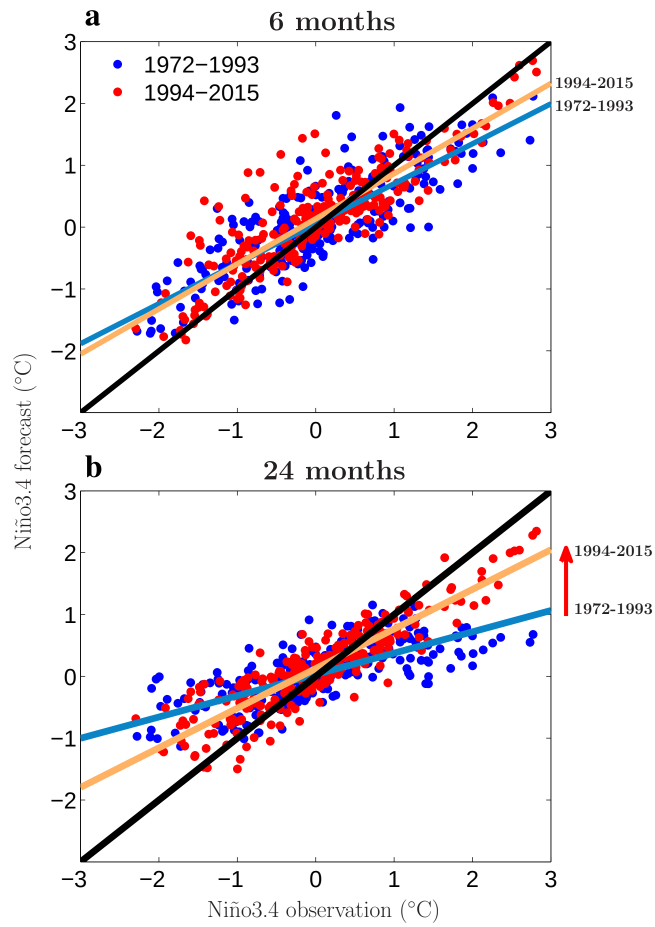


FIG. 2. **Relationship between observations and model predictions.** Scatter plots of the Niño3.4 Index observations and the model predictions at **a**, 6-month lead and **b**, 24-month lead. The blue dots correspond to the period 1972-1993 with a linear regression line in light blue, and the red dots correspond to the period 1994-2015 with a linear regression line in beige. The red arrow indicates the improvement in the slope of the regression line for the period 1994-2015 with respect to the slope of the regression line for the period 1972-1993.

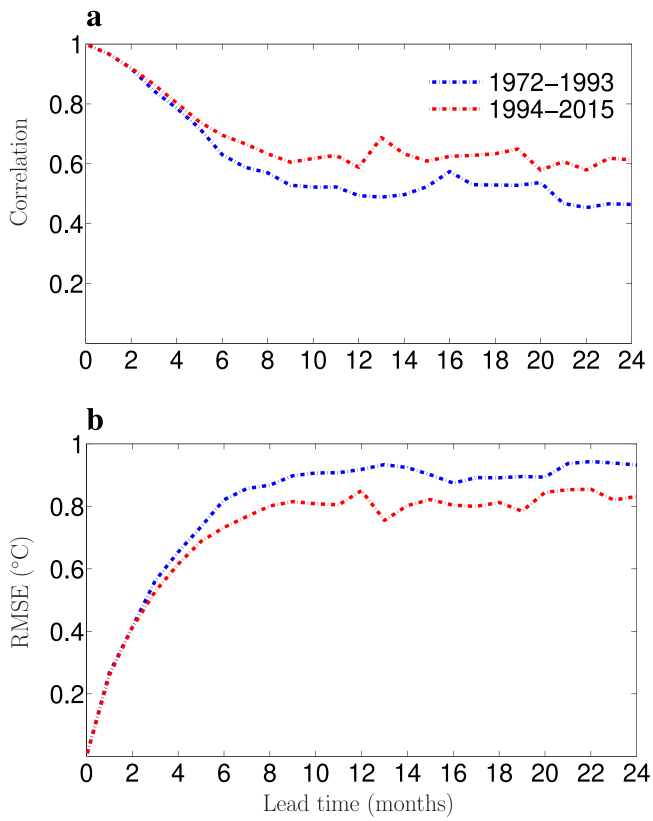


FIG. 3. **General forecast skill of the model.** **a**, Correlations between the Niño3.4 Index observations and model predictions and **b**, root mean square errors (RMSE) as functions of lead time for two consecutive 22-year periods, 1972-1993 (blue) and 1994-2015 (red).

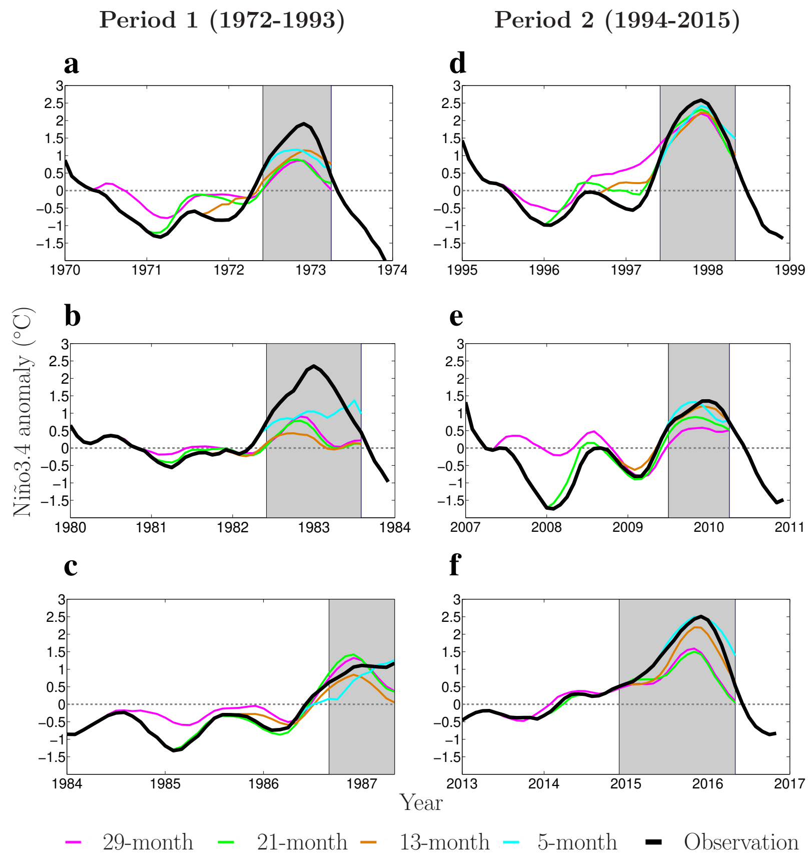


FIG. 4. **Forecasts of the major El Niño events since 1970. a-c**, El Niño events in the period 1972-1993 and **d-f**, 1994-2015. The thick black curves are the observed Niño3.4 Index anomalies, and the thin magenta, green, beige and cyan curves are predictions started 29, 21, 13 and 5 months in advance.

Supplementary material for "Multi-year statistical prediction of ENSO enhanced by the Tropical Pacific Observing System"

Desislava Petrova[1], Joan Ballester[1], Siem Jan Koopman[2], and Xavier Rodó[1,3]

[1]Climate and Health Programme, Barcelona Institute for Global Health, Barcelona, Catalonia, Spain
[2]Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[3]Institució Catalana de Recerca i Estudis Avancats (ICREA), Barcelona, Catalonia, Spain

May 20, 2019

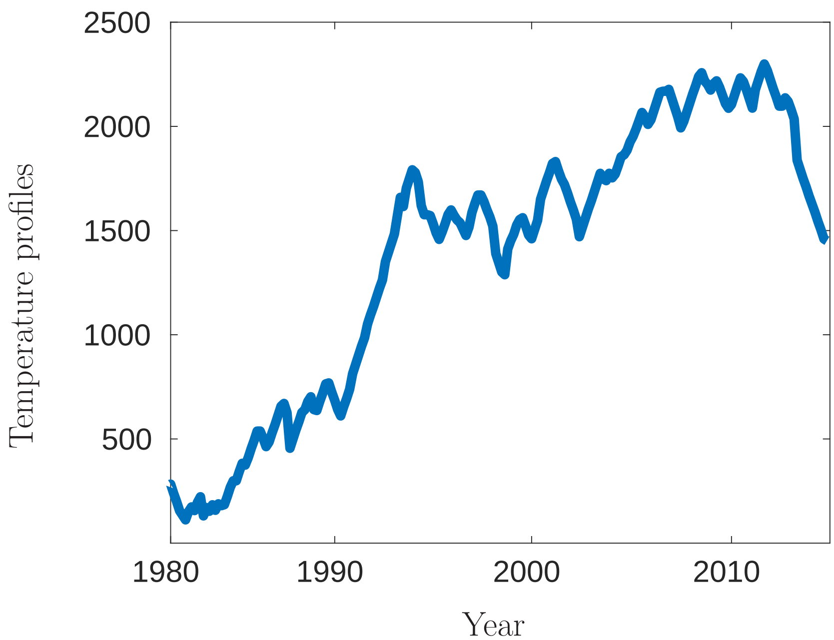


Figure S1: Number of temperature profiles in the equatorial Pacific Ocean (including XBT, TAO and Argo profiles). Adapted from *Kumar et al. (2015).*

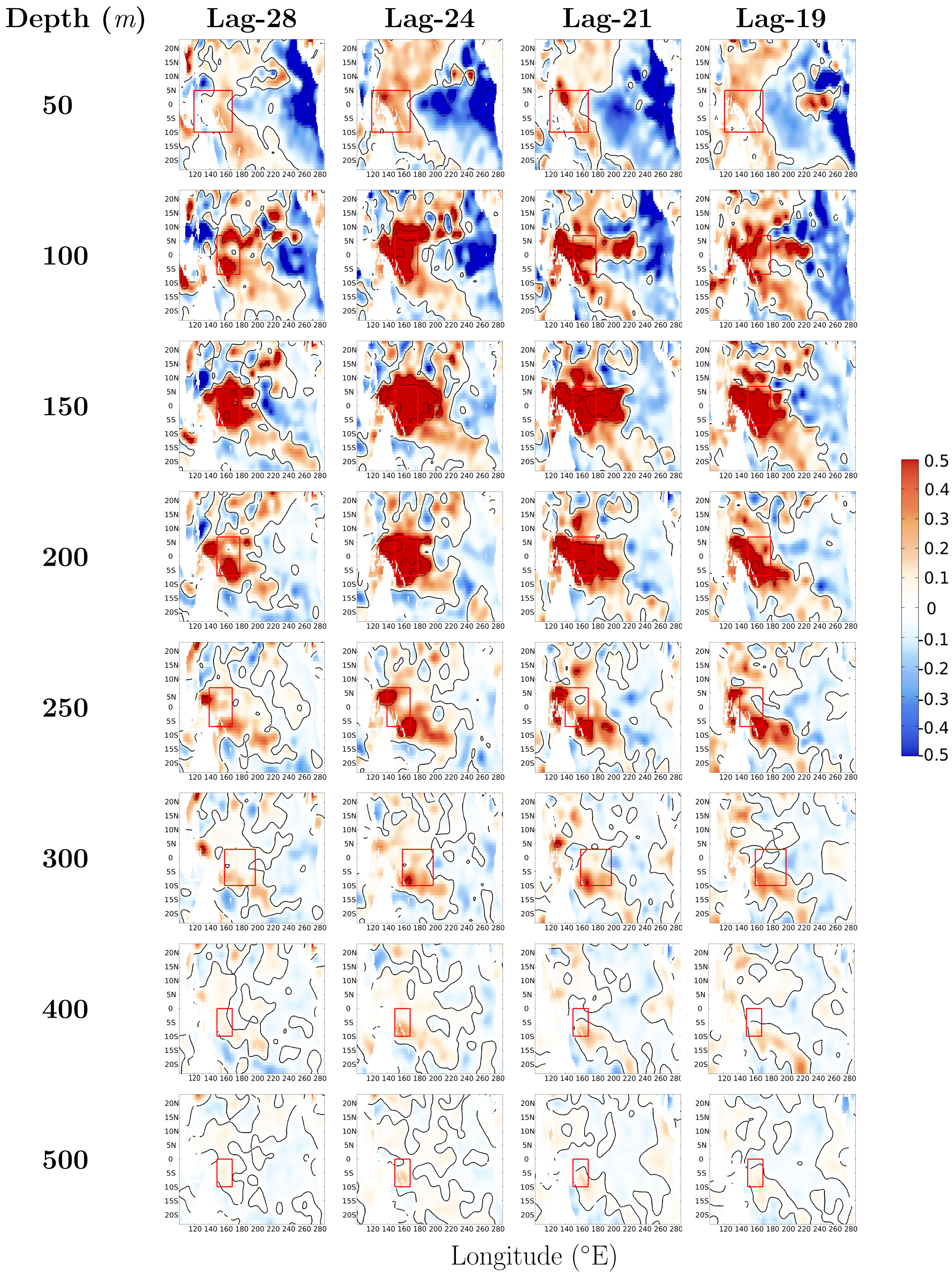


Figure S2: Composites of interannual monthly subsurface temperature anomalies (in [$^\circ C$], shading) between 50-500 metres depth from the Subsurface Temperature and Salinity Analyses by *Ishii et al. (2005)* at lead times of 19, 21, 24 and 28 months ahead of the EN peak. Red boxes indicate regions for derivation of predictors. Composites are with respect to all El Niño events in the period 1978-2012.

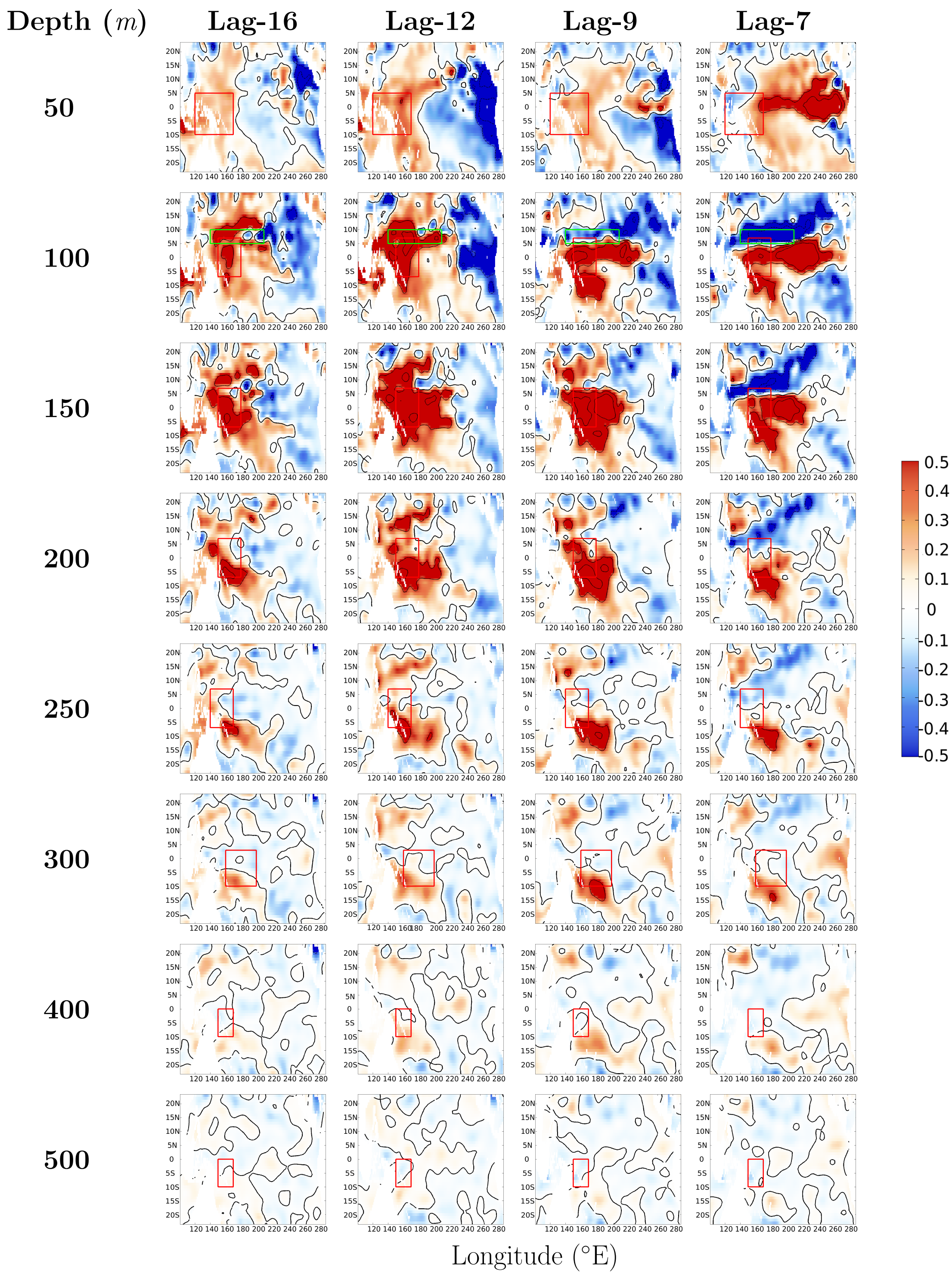


Figure S3: As in Figure S2, but at lead times of 7, 9, 12 and 16 months ahead of the El Niño peak. Red and green boxes indicate regions for derivation of predictors.

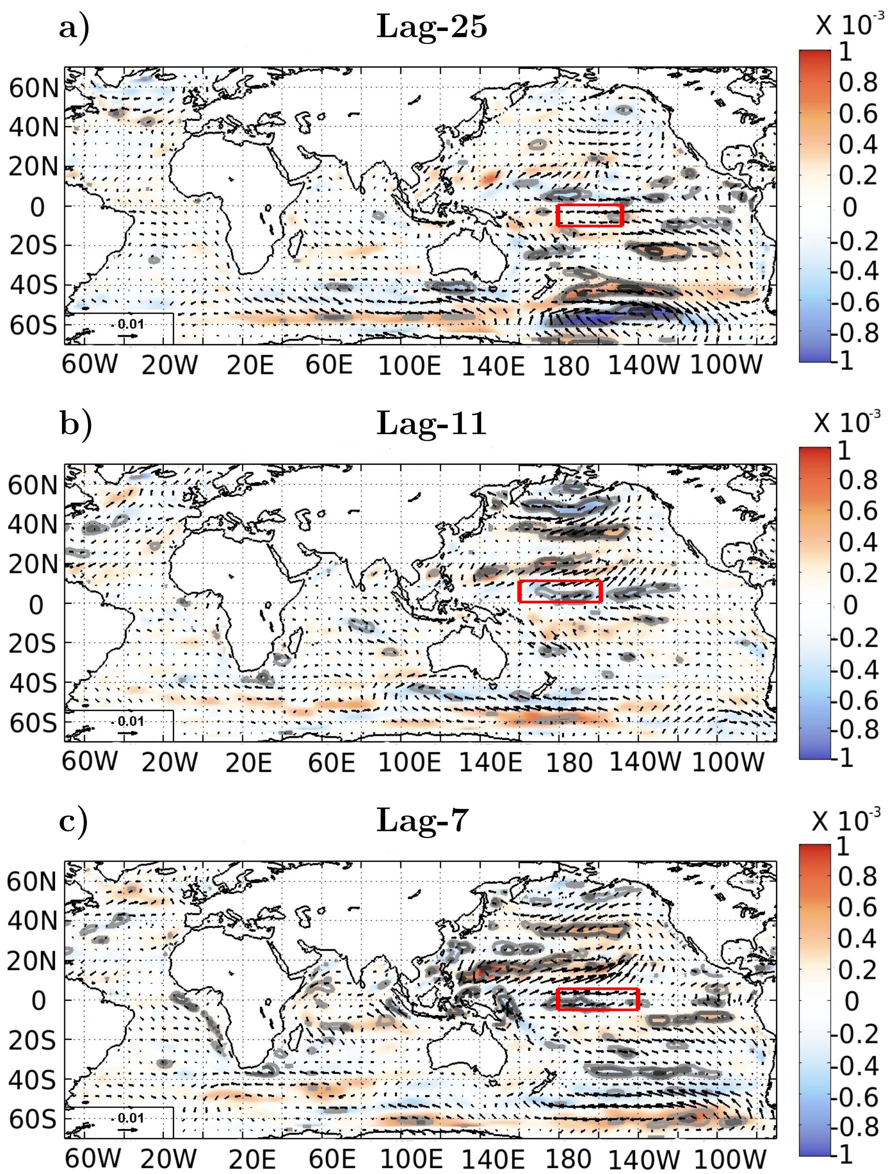


Figure S4: Composites of interannual monthly surface zonal and meridional wind stress anomalies (in $[Nm^{-}2]$, arrows) and wind stress curl (in $[Nm^{-}3]$, shading) from the NCEP/NCAR reanalysis (*Kalnay et al. (1996)*) for **a)** 25, **b)** 11, and **c)** 7 months before the winter peak of El Niño. Red boxes indicate regions for derivation of predictors. Composites are with respect to all El Niño events in the period 1978-2012.

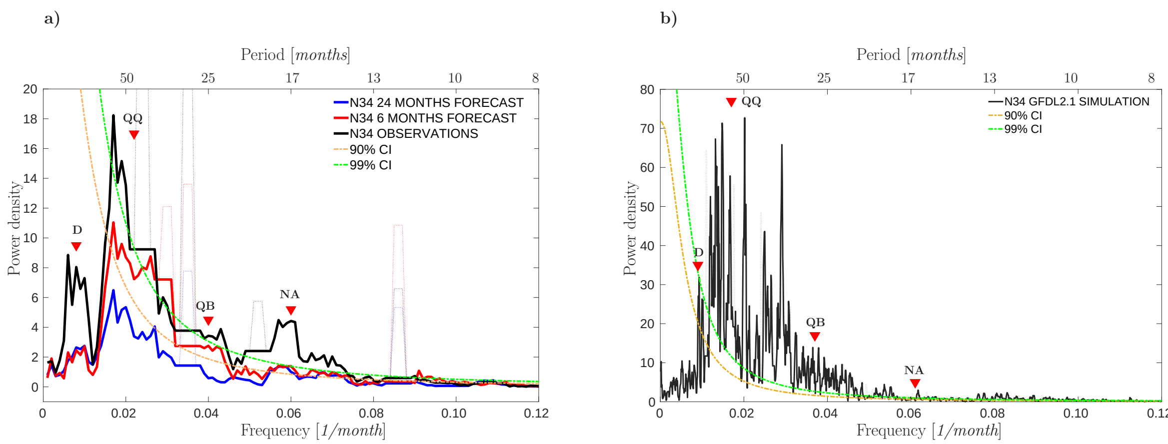


Figure S5: Multi Taper Method (MTM) power spectra for a) the observed Niño3.4 time series (black), and for predictions with the dynamic components model at 6 months lead time (red) and 24 months lead time (blue); b) the simulated Niño3.4 time series (black) with GFDL2.1 ENSO dynamical model (500-year spin-up simulation). The solid lines indicate the power density, dotted lines harmonic peaks and dashed lines confidence levels based on a red noise null hypothesis. The red markers indicate the regions of the spectrum associated with the near-annual (NA), quasi-biannual (QB), quasi-quadrennial (QQ) and decadal (D) modes of ENSO variability.

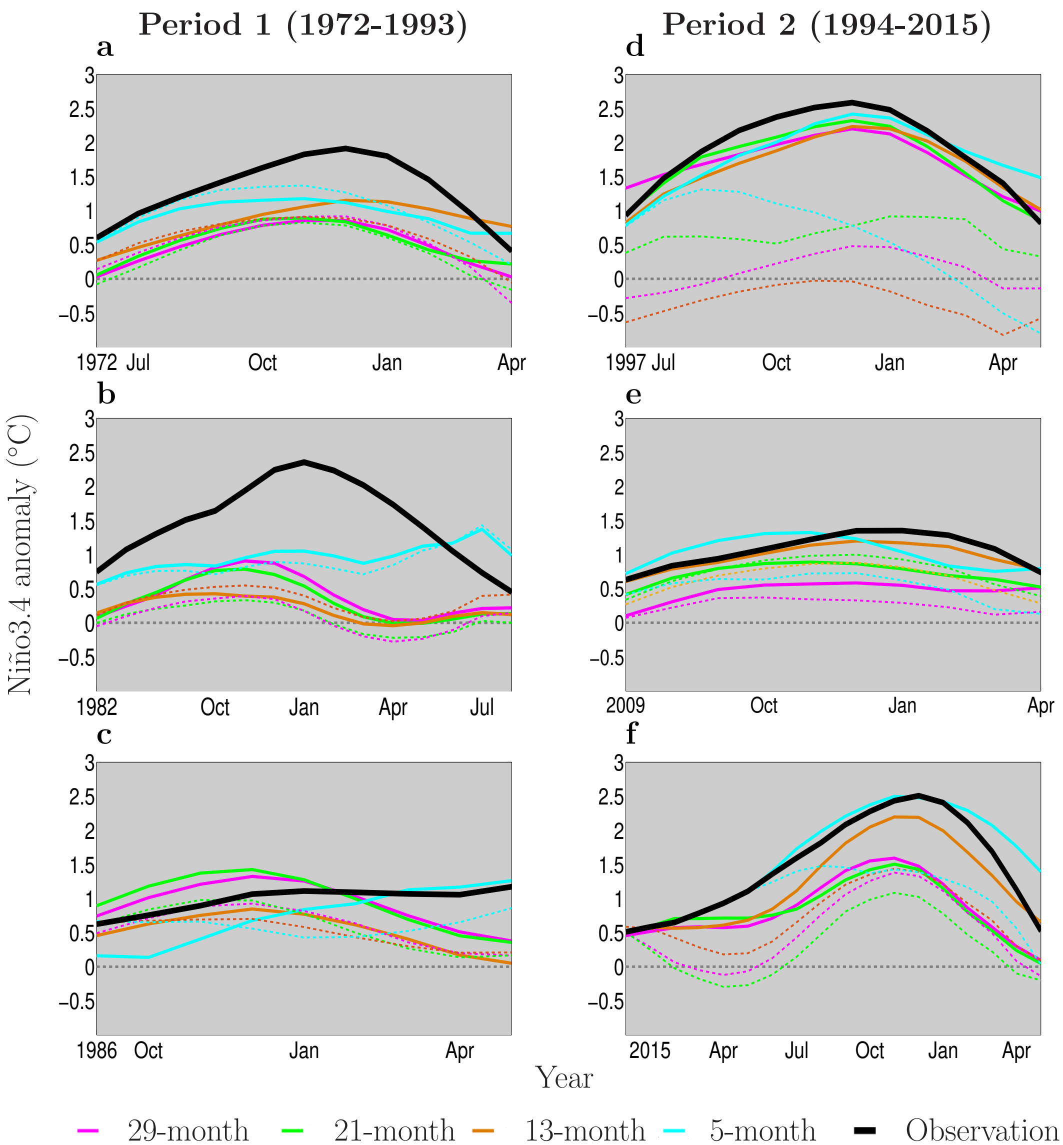
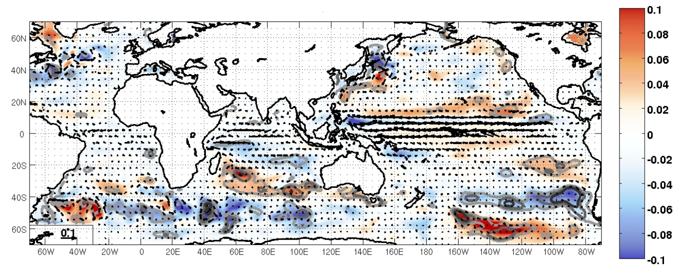


Figure S6: Forecasts of the major El Niño events since 1970. **a-c**, El Niño events in the period 1972-1993 and **d-f**, 1994-2015. The solid black curves are the observed Niño3.4 Index anomalies, the solid (dashed) magenta, green, beige and cyan curves are predictions started 29, 21, 13 and 5 months in advance running the model with the predictor variables (without the predictor variables).

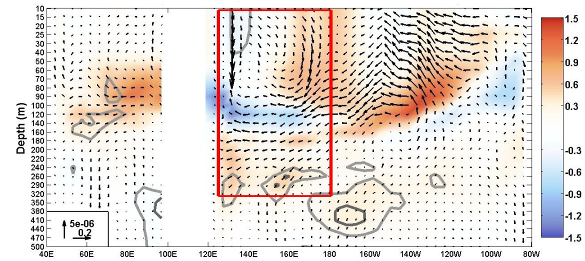
SSH

SUBT

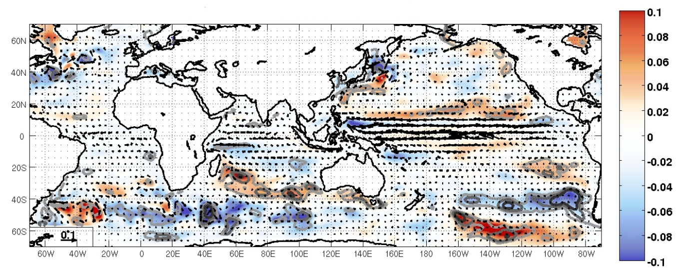
Anomaly for 10/1980



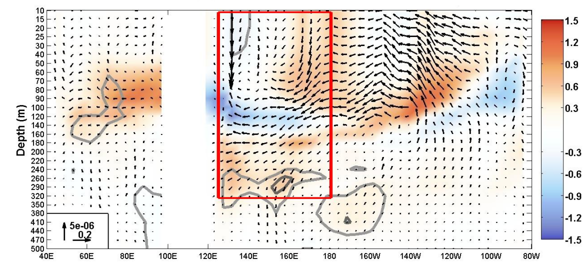
Anomaly for 10/1980



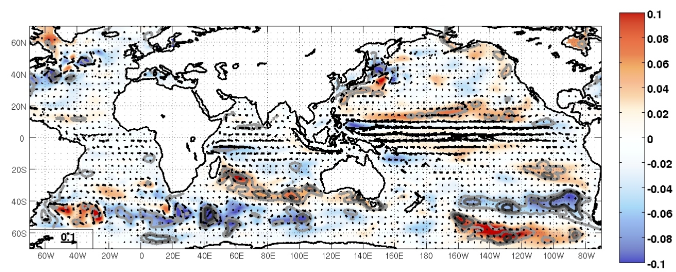
Anomaly for 11/1980



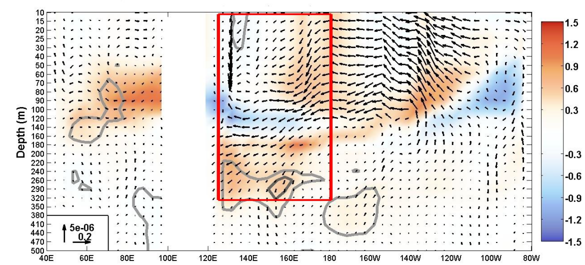
Anomaly for 11/1980



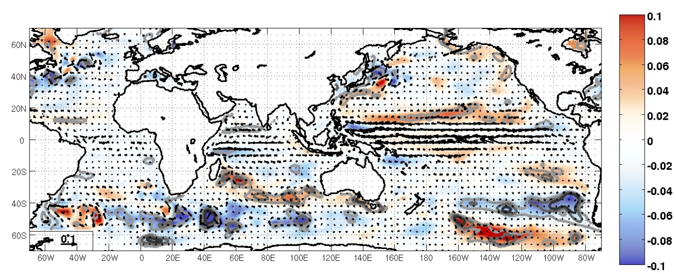
Anomaly for 12/1980



Anomaly for 12/1980



Anomaly for 1/1981



Anomaly for 1/1981

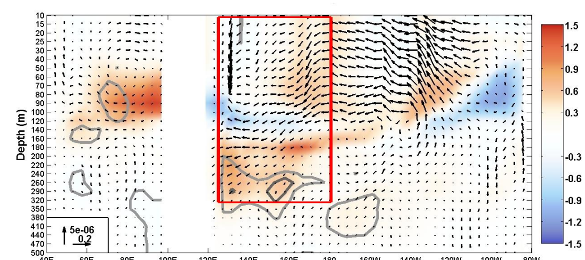


Figure S7: Composite anomalies of (left) sea surface height (in $[cm]$, shading) and zonal wind stress (in $[Nm^{-2}]$, arrows); (right) subsurface potential temperature (in $[°C]$, shading) and zonal and vertical currents (in $[m/s]$, arrows) between latitudes 2°S-2°N of the ORAS Version 3 reanalysis product at lags 23-26 with respect to the 1982/83 El Niño event. Red boxes contain opposite-sign temperature anomalies in the western equatorial Pacific.

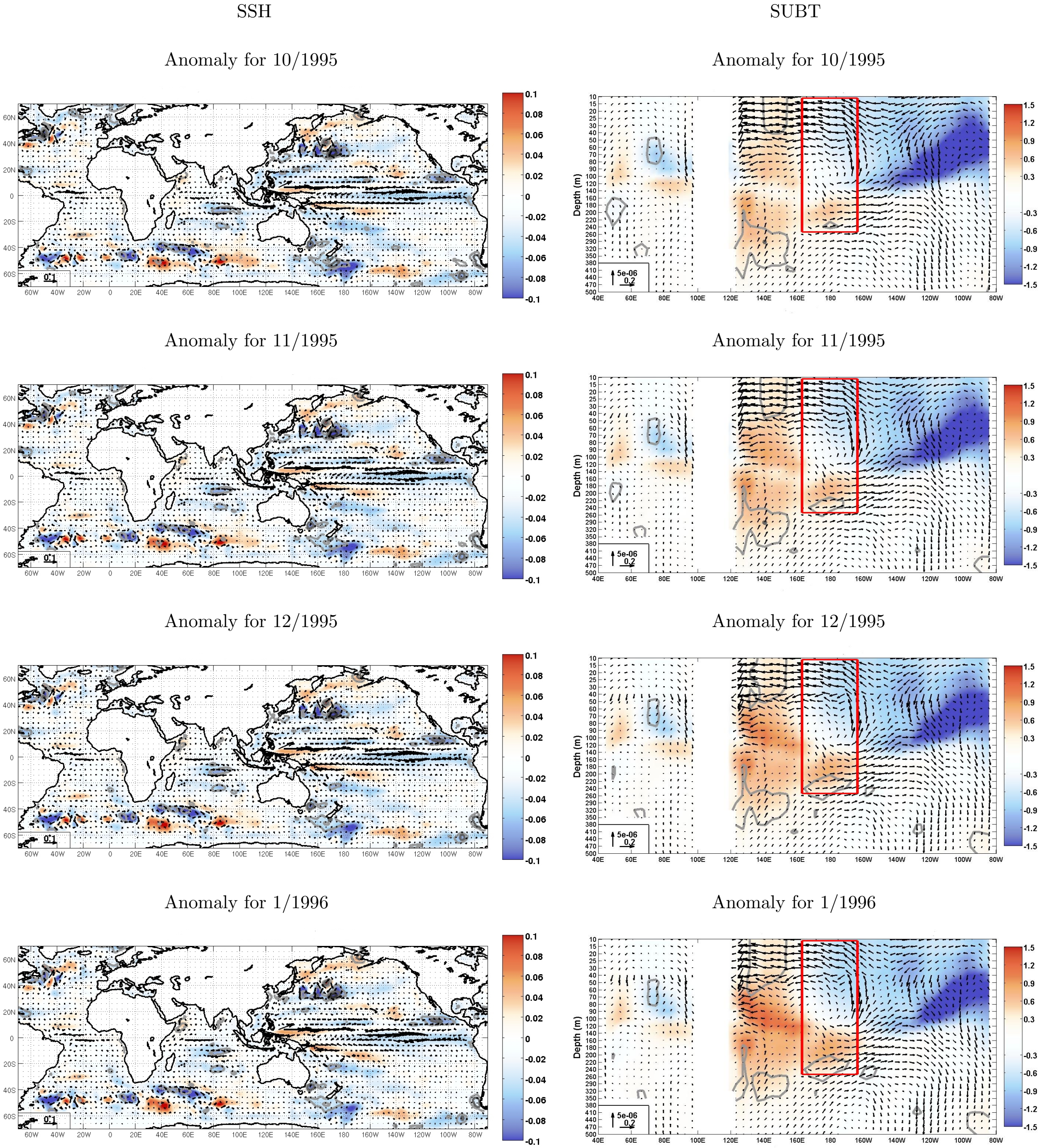


Figure S8: Composite anomalies of (left) sea surface height (in $[cm]$, shading) and zonal wind stress (in $[Nm^{-2}]$, arrows); (right) subsurface potential temperature (in $[^{\circ}C]$, shading) and zonal and vertical currents (in $[m/s]$, arrows) between latitudes 2°S-2°N of the ORAS Version 3 reanalysis product at lags 23-26 with respect to the 1997/98 El Niño event. Red boxes contain opposite-sign temperature anomalies in the western equatorial Pacific.

Table S1: Coefficients, t-values and p-values for **subsurface temperature** predictor regression variables at **21-month** lead. Values significant at the 90% level are bold.

El Niño event	**250m. RII**	**wnd RI**
1972/73		
Coefficient	-0.08	-0.70
t	-0.62	-0.26
p	0.53	0.79
1982/83		
Coefficient	-0.07	-0.88
t	-0.70	-0.40
p	0.48	0.68
1986/87		
Coefficient	-0.05	-0.32
t	-0.50	-0.16
p	0.61	0.87
1991/92		
Coefficient	-0.14	0.48
t	-1.27	0.23
p	0.20	0.81
1997/98		
Coefficient	**-0.33**	3.96
t	**-1.96**	1.34
p	**0.05**	0.18
2002/03		
Coefficient	**-0.34**	**4.67**
t	**-2.08**	**1.83**
p	**0.03**	**0.06**
2006/07		
Coefficient	-0.24	**4.41**
t	-1.60	**1.92**
p	0.11	**0.05**
2009/10		
Coefficient	**-0.30**	**4.07**
t	**-2.10**	**2.02**
p	**0.03**	**0.04**
2014/15		
Coefficient	**-0.20**	**4.05**
t	**-1.67**	**2.53**
p	**0.09**	**0.01**
2015/16		
Coefficient	-0.14	**3.59**
t	-1.19	**2.25**
p	0.23	**0.02**

Table S2: Coefficients, t-values and p-values for **subsurface temperature** predictor regression variables at **29-month** lead. Coefficients significant at the 90% level are bold.

El Niño event	**250m. RI**	**300m. RI**	**400m. RI**
1972/73			
Coefficient	0.06	0.04	−0.51
t	0.38	0.18	−1.25
p	0.70	0.85	0.21
1982/83			
Coefficient	−0.07	−0.01	0.07
t	−0.55	−0.07	0.30
p	0.58	0.94	0.76
1986/87			
Coefficient	0.04	0.07	0.18
t	0.37	0.50	0.81
p	0.71	0.61	0.41
1991/92			
Coefficient	−0.04	0.13	0.22
t	−0.36	0.72	0.91
p	0.71	0.47	0.36
1997/98			
Coefficient	**0.25**	**0.37**	**0.57**
t	**1.73**	**1.68**	**1.88**
p	**0.08**	**0.09**	**0.06**
2002/03			
Coefficient	0.15	0.28	**0.50**
t	1.15	1.35	**1.78**
p	0.25	0.17	**0.07**
2006/07			
Coefficient	0.18	**0.38**	**0.40**
t	1.60	**2.09**	**1.64**
p	0.11	**0.03**	**0.10**
2009/10			
Coefficient	**0.18**	**0.29**	0.32
t	**1.68**	**1.71**	1.38
p	**0.09**	**0.08**	0.16
2014/15			
Coefficient	**0.16**	**0.29**	**0.41**
t	**1.64**	**1.87**	**1.89**
p	**0.10**	**0.06**	**0.06**
2015/16			
Coefficient	0.11	0.25	**0.38**
t	1.23	1.60	**1.79**
p	0.22	0.11	**0.07**

Table S3: Predicted values for the peak of El Niño events shown in Figure 4 from the model with and without predictor variables. Indicated are also the observed values for the peaks.

El Niño event/Lag	**5 months**	**13 months**	**21 months**	**29 months**
1972/73				
no predictors	1.37	0.91	0.82	0.92
predictors	1.18	1.15	0.89	0.85
observation	1.91	1.91	1.91	1.91
1982/83				
no predictors	0.89	0.55	0.33	0.39
predictors	1.05	0.43	0.79	0.91
observation	2.36	2.36	2.36	2.36
1986/87				
no predictors	0.66	0.70	0.98	0.93
predictors	0.93	0.85	1.42	1.32
observation	1.11	1.11	1.11	1.11
1997/98				
no predictors	1.31	-0.03	0.92	0.48
predictors	2.42	2.24	2.32	2.20
observation	2.59	2.59	2.59	2.59
2009/10				
no predictors	0.73	0.88	0.99	0.37
predictors	1.32	1.20	0.89	0.58
observation	1.35	1.35	1.35	1.35
2015/16				
no predictors	1.48	1.44	1.08	1.38
predictors	2.50	2.20	1.59	1.50
observation	2.51	2.51	2.51	2.51