New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

OPEN ACCESS

RECEIVED
17 January 2018

REVISED
15 March 2018

ACCEPTED FOR PUBLICATION
3 April 2018

PUBLISHED
16 April 2018

PAPER

Forecasting the magnitude and onset of El Niño based on climate network

Jun Meng[1,2], **Jingfang Fan**[1,2], **Yosef Ashkenazy**[1], **Armin Bunde**[3] **and Shlomo Havlin**[2,4]

[1] Department of Solar Energy & Environmental Physics, Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Midreshet Ben-Gurion 84990, Israel
[2] Department of Physics, Bar-Ilan University, Ramat-Gan 52900, Israel
[3] Institut für Theoretische Physik, Justus-Liebig-Universität Giessen, D-35392 Giessen, Germany
[4] Institute of Innovative Research, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8502, Japan

E-mail: j.fang.fan@gmail.com

Keywords: ENSO, climate networks, complex systems, dynamic networks

Supplementary material for this article is available online

Abstract

El Niño is probably the most influential climate phenomenon on inter-annual time scales. It affects the global climate system and is associated with natural disasters; it has serious consequences in many aspects of human life. However, the forecasting of the onset and in particular the magnitude of El Niño are still not accurate enough, at least more than half a year ahead. Here, we introduce a new forecasting index based on climate network links representing the similarity of low frequency temporal temperature anomaly variations between different sites in the Niño 3.4 region. We find that significant upward trends in our index forecast the onset of El Niño approximately 1 year ahead, and the highest peak since the end of last El Niño in our index forecasts the magnitude of the following event. We study the forecasting capability of the proposed index on several datasets, including, ERA-Interim, NCEP Reanalysis I, PCMDI-AMIP 1.1.3 and ERSST.v5.

1. Introduction

El Niño Southern Oscillation (ENSO) is an inter-annual coupled ocean-atmosphere climate phenomenon [1–3]. El Niño is the warm phase of ENSO and is characterized by several degrees warming of the eastern equatorial Pacific ocean. It occurs every 3–5 years, and is regarded as the most significant climate phenomenon on decadal time scales. Among other factors, it affects the surface temperature, precipitation and mid-tropospheric atmospheric circulation over extended regions in America, Australia, Europe, India, and East Asia [4–8]. In particular, strong El Niño can trigger a cascade events that can affect many aspects of human life [9–11].

As a result of the environmental, economical, and social impacts of El Niño, intensive efforts have been undertaken to understand and eventually forecast El Niño [12–14]. Extensive atmospheric and oceanic observations have been used to track variations in ENSO cycle, and complex computer models have been developed to forecast El Niño[15–21]. Still, reliable forecasts techniques for the onset and in particular the magnitude of El Niño with relatively long lead time (of more than half a year) are not fully satisfactory. We have just undergone one of the strongest El Niño events since 1948, which started in the end of 2014 and ended in mid-2016 [22]. The onset of this event was predicted one year ahead using the network approach [23]. Here, we develop a climate network-based index that can forecast the onset of El Niño approximately 1 year ahead (similar to [23–25]). In particular our approach forecasts the magnitude of El Niño, once it begins.

2. Methodology

The Oceanic Niño Index (ONI) is a standard index that is used to identify El Niño [26]. It is the running 3 month mean sea surface temperature (SST) anomaly averaged over the Niño 3.4 region, based on 30 years periods,

Table 1. Summary of the information (rows 1–5) and the forecasting power (rows 6–10) for different datasets. The 'Resolution of the data' refers to the spatial (zonal and meridional) resolution of the data. In the row of 'D', the symbol '*' indicates that there are not enough events in the datasets to perform the Kolmogorov–Smirnov statistic. In the columns of lead time, we show the mean value ± 1 standard deviation in units of months. For more details on the above results see figure S2 in the SI.

DATA		ERA-Interim [27]	NCEP Reanalysis I [28]	PCMDI-AMIP 1.1.3 [29]	ERSST.v5 [30]
Type of data		Daily near surface (1000 hPa) air temperature		Monthly sea surface temperature	
The first year of data		1979	1948	1870	1854
The first year of the FI		1984	1953	1950	1950
Resolution of the data		$2.5^\circ \times 2.5^\circ$	$2.5^\circ \times 2.5^\circ$	$1^\circ \times 1^\circ$	$2^\circ \times 2^\circ$
El Niño magnitude	r-value	0.76	0.39	0.70	0.32
	D	*	0.50	0.81	0.37
El Niño onset	Hit rate	$(7+2)/9 = 1$	$(13+4)/21 \approx 0.81$	$(18+3)/22 \approx 0.95$	$(19+2)/22 \approx 0.95$
	False alarm rate	$2/23 \approx 0.09$	$8/43 \approx 0.19$	$9/42 \approx 0.21$	$10/41 \approx 0.24$
	Lead time (month)	12.2 ± 2.6	9.3 ± 4.9	9.1 ± 4.7	6.9 ± 4.1

updated every 5 years. When the ONI exceeds 0.5 °C for at least five consecutive months, the corresponding year is considered to be an El Niño year. We use the ONI (whose first value is at 1950) to estimate the accuracy of our predictions for El Niño events occurred after 1950.

We analyze the variability of the daily mean near surface (1000 hPa) air temperature fields of the ERA-Interim reanalysis [27], the NCEP Reanalysis I [28], the AMIP SST boundary condition data (current version: PCMDI-AMIP 1.1.3) [29], and the extended reconstructed SST v5 (ERSST.v5) [30] in the Niño 3.4 region (i.e., 5°S–5°N, 120°W–170°W) using a climate network approach [31–38]. See table 1 (rows 1–5) for detailed information on the datasets. We find that the temporal variations of temperature anomaly (defined below in (i)) in different sites of the Niño 3.4 region become less coherent (more disordered) well before the onset of El Niño. In particular, the magnitude of the event is approximately proportional to the maximal degree of disorder (defined below in (ii)) that the Niño 3.4 region can reach before the onset of El Niño. We suggest a single index, the degree of disorder of the El Niño 3.4 region, that can forecast both the onset and magnitude of El Niño.

In the following, we first demonstrate the steps of the forecasting method we propose on 33 years (1984 to present) of the reanalysis data of the European Centre for Medium-Range Weather Forecasts Interim Reanalysis (ERA-Interim) [27]. We then examine the robustness and accuracy of the prediction method on longer periods using several other datasets (NCEP Reanalysis I [28], PCMDI-AMIP 1.1.2 [29] and ERSST.v5 [30]).

The daily mean near surface (1000 hPa) air temperature fields of the ERA-Interim reanalysis data have a spatial (zonal and meridional) resolution of $2.5^\circ \times 2.5^\circ$, resulting in 105 grid points in the Niño 3.4 region. Different locations (grid points) in the Niño 3.4 region correspond to nodes in the local climate network, and the weight of links are determined by the similarities (defined below in (ii)) of the temporal temperature anomaly variations between pairs of nodes [31, 38]. The forecasting algorithm is as follows:

(i) At each node k of the network, we calculate the daily atmospheric temperature anomalies $T_k(t)$ (actual temperature value minus the climatological average which then is divided by the climatological standard deviation) for each calendar day. For the calculation of the climatological average and standard deviation, only past data up to the prediction date have been used. For simplicity leap days were excluded. We have used the first 5 years of data (1979–1983) to calculate the first average value and start the prediction from 1984.

(ii) For obtaining the time evolution of the weight of the links between nodes i and j in the Niño 3.4 region, we follow [24, 25, 31] and compute, for each month t (the first day where the month starts) in the considered time span between 1 January 1981 and 31 August 2017, the time-delayed cross-correlation function defined as

$$C_{i,j}^{(t)}(-\tau) = \frac{\langle T_i^{(t)}(t)\, T_j^{(t)}(t-\tau)\rangle - \langle T_i^{(t)}(t)\rangle \langle T_j^{(t)}(t-\tau)\rangle}{\sqrt{\langle (T_i^{(t)}(t) - \langle T_i^{(t)}(t)\rangle)^2\rangle} \cdot \sqrt{\langle (T_j^{(t)}(t-\tau) - \langle T_j^{(t)}(t-\tau)\rangle)^2\rangle}} \tag{1}$$

and

$$C_{i,j}^{(t)}(\tau) = \frac{\langle T_i^{(t)}(t-\tau)\, T_j^{(t)}(t)\rangle - \langle T_i^{(t)}(t-\tau)\rangle \langle T_j^{(t)}(t)\rangle}{\sqrt{\langle (T_i^{(t)}(t-\tau) - \langle T_i^{(t)}(t-\tau)\rangle)^2\rangle} \cdot \sqrt{\langle (T_j^{(t)}(t) - \langle T_j^{(t)}(t)\rangle)^2\rangle}}, \tag{2}$$

where the brackets denote an average over the past 365 days, according to

$$\langle f(t)\rangle = \frac{1}{365}\sum_{a=1}^{365} f(t-a). \tag{3}$$

We consider, for the daily datasets, time lags of $\tau \in [0, 200]$ days, where a reliable estimate of the background noise level can be guaranteed (the appropriate time lag is discussed in [39]). For monthly updating datasets (PCMDI-AMIP 1.1.3 and ERSST.v5), the brackets denote an average over the past 12 months, according to $\langle f(t)\rangle = \frac{1}{12}\sum_{a=1}^{12} f(t-a)$ and we consider time lags of $\tau \in [0, 6]$ months. The similarity between two nodes (the weight of the link) is determined by the value of the highest peak of the cross-correlation function, $C_{i,j}^{(t)}(\theta)$, where θ is the corresponding time lag at the peak. The degree of coherence/disorder of the Niño 3.4 region is quantified by the average value of all links at their peaks, i.e.

$$C(t) = \frac{2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} C_{i,j}^{(t)}(\theta), \tag{4}$$

where $N = 105$ is the number of nodes in the Niño 3.4 region. Thus, higher values of $C(t)$ indicate higher coherence in the Niño 3.4 region.

We like to note that the strength of the link between nodes i and j is represented by the strength of the cross-correlation between the temperature records at the nodes, which is defined by [24, 35]

$$W_{i,j}^{(t)} = \frac{C_{i,j}^{(t)}(\theta) - \mathbf{E}(C_{\mathrm{i,j}}^{(t)})}{\sqrt{\mathbf{E}(C_{i,j}^{(t)} - \mathbf{E}(C_{i,j}^{(t)}))^2}}, \tag{5}$$

where $\mathbf{E}(g)$ denotes the average over 401 shifting days, according to

$$\mathbf{E}(g) = \frac{1}{401}\left(\sum_{\tau=0}^{200} g(\tau) + \sum_{\tau=1}^{200} g(-\tau)\right). \tag{6}$$

Thus, $W_{i,j}^{(t)}$ is high when the peak at $\tau = \theta$ is sharp and prominent, and it is low when the cross-correlation function $C_{i,j}^{(t)}(\tau)$ varies slowly with τ. In [24], Ludescher *et al* introduced a 12-mo forecasting scheme based on the observation that the mean strength of links that connect the 'El Niño basin' (equatorial Pacific corridor) with the surrounding sites tends to increase about one year before the El Niño event.

(iii) The forecasting index (FI) we propose here, is based on the temporal evolution of $C(t)$ (defined in (ii) equation (4)), representing the interactions or similarity (coherence) between the different sites within the Niño 3.4 region. We define the FI as a function of months as follows,

$$\mathrm{FI}(t) = \frac{1}{m+1}\sum_{a=0}^{m} \ln C(t-a) - \ln C(t), \tag{7}$$

where $a = 0$ indicates that the average of $\ln(C)$ includes the current month, while m is the total number of months preceding t since January 1981. We use a minus sign in the right hand side of equation (7) so that peaks in the FI will correspond to peaks in the ONI, see figure 1. We also use the log in $C(t)$ instead of just $C(t)$ in order to make small variations of $C(t)$ to become more significant so that it will be seen more clearly in figure 1. We start to evaluate $\mathrm{FI}(t)$ from January 1984. (For NCEP Reanalysis I, m equals the number of months before t since January 1950, and the $\mathrm{FI}(t)$ starts from January 1953; for PCMDI-AMIP 1.1.2, m equals the number of months before t since January 1872, and $\mathrm{FI}(t)$ starts from January 1950; for ERSST.v5, m equals the number of months before t since January 1856, and $\mathrm{FI}(t)$ starts from January 1950) Thus, it follows that $\mathrm{FI}(t)$ increases ($C(t)$ decreases) when the Niño 3.4 region is less coherent or more disordered (due to the minus sign). $\mathrm{FI}(t)$ is calculated for each month (red dotted line in figure 1) and one can easily see that usually $\mathrm{FI}(t)$ increases well before the onset of El Niño, and decreases once El Niño begins. In other words, the temporal variations of temperature anomaly in different sites of the Niño 3.4 region become less coherent (more disorder) prior to El Niño, and start to synchronize once El Niño begins. In particular, we find that the more disordered the Niño 3.4 region is before El Niño, the higher is the magnitude of the approaching El Niño.

We provide a flow chart entitled 'Steps in calculating FI' that briefly describes the above algorithm in the supplementary material is available online at stacks.iop.org/NJP/20/043036/mmedia.

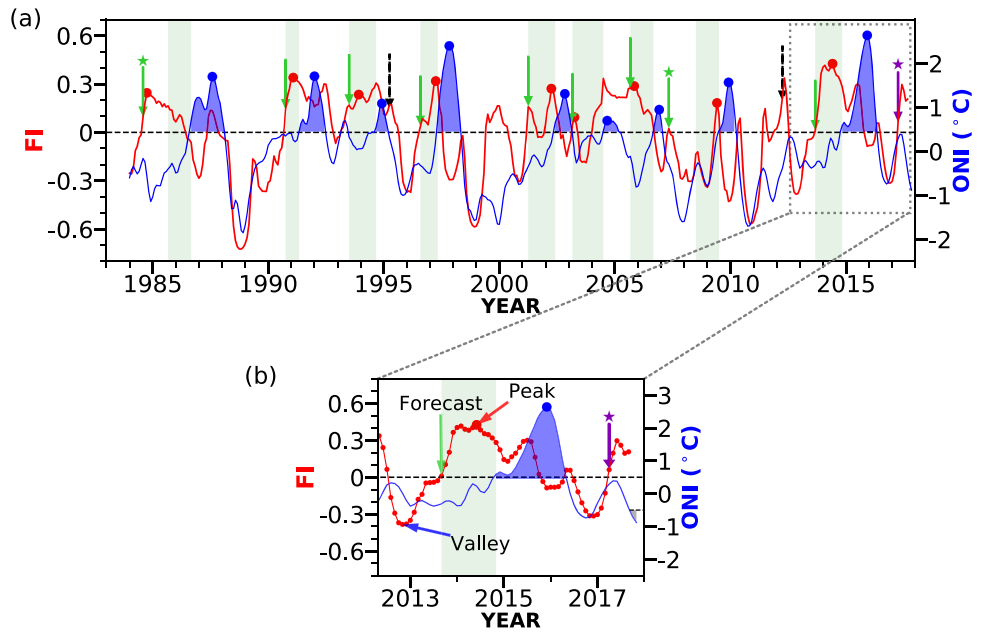

Figure 1. The forecasting algorithm. (a) The forecasting index (FI(t), red) and the ONI (blue). The blue shades under the ONI curve indicate the El Niño periods. True positive forecasts are marked by the green arrows, while false alarms are marked in black dashed arrows. The purple arrow is the newest potential alarm pending verification in the future. The green stars above some of the green arrows indicate the predictions after which La Niña occurred, leading to postponed El Niño prediction. The horizontal black dashed line indicates the FI $= 0$. The left edge of the light green rectangular shading indicates the time that FI(t) predicts El Niño and the right edge indicates the onset of El Niño, thus the width of the shading indicates the lead time before the onset of El Niño. The red dots indicate FI(t) values that are used to predict the magnitudes of El Niño and the blue dots indicate the actual magnitudes of El Niño. (b) A detailed view of FI(t) and the ONI since May, 2012. The green arrow (Forecast) indicates the true positive forecast of 2014–2016 El Niño event; the blue (Valley) and red (Peak) arrows indicate the nearest minimum before the prediction point and the peak in FI(t) is the value used to predict the magnitude of El Niño (blue dot). The purple star above the purple arrow indicates that we might be undergoing a new La Niña (the gray shades, the ONI has already been bellow -0.5 °C for the last two months), and therefore an El Niño is forecasted to come within 18 months after the end of the suspected ongoing La Niña.

3. The forecasting algorithm using index FI

Based on the above observation, we suggest the following algorithm to forecast simultaneously both the magnitude and onset of El Niño using FI(t). For demonstration see the example shown in figure 1(b).

(i) To forecast the magnitude, as soon as one month the ONI rises across 0.5 °C we regard the value of the highest peak of FI(t) ('Peak', as indicated by the red points in figure 1(a) and the red arrow in (b)) since the end of last El Niño as an estimate (forecasted magnitude) for El Niño strength (observed magnitude). However, if the peak value is negative or there is no peak during this period, we use zero as the forecasted magnitude and forecast a weak El Niño event (ONI < 1 °C) (we counted the results of all the datasets we used, and find that the ratio of such events is 13% on average of all the El Niño events, and most of them (84% on average of this kind of events) are indeed weak). In addition, we should clarify that if the ONI rises across but do not keep above 0.5 °C for at least five months, we do not have an El Niño event, thus the value of the highest peak is not a prediction of El Niño magnitude.

(ii) To forecast the onset, we track both FI(t) and the ONI, starting from the onset of the previous El Niño. If FI(t) increases from a local minimum ('Valley', as indicated by the blue arrow in figure 1(b)) continuously for at least two months (time segment that yielded the best forecast), the time at which FI(t) exceeds 0 (if it is not during ongoing El Niño/La Niña period, i.e. -0.5 °C $<$ ONI < 0.5 °C) is considered as a potential signal for the onset of either El Niño or La Niña event within approximately the next 18 months ('Forecast', as indicated by the green arrows in figure 1). Moreover, if La Niña is experienced within these 18 months, we forecast a new El Niño to occur within 18 months after the end of La Niña (the first month of ONI > -0.5 °C after La Niña). Given the above, a true positive prediction of El Niño is counted if within 18 months after the potential signal an El Niño occurs ('normal', as indicated by the green arrows in figure 1), or a La Niña that followed by an El Niño in the next 18 months occurs ('delayed', as indicated by the green arrows with stars on the top in figure 1); otherwise, a false alarm is counted.

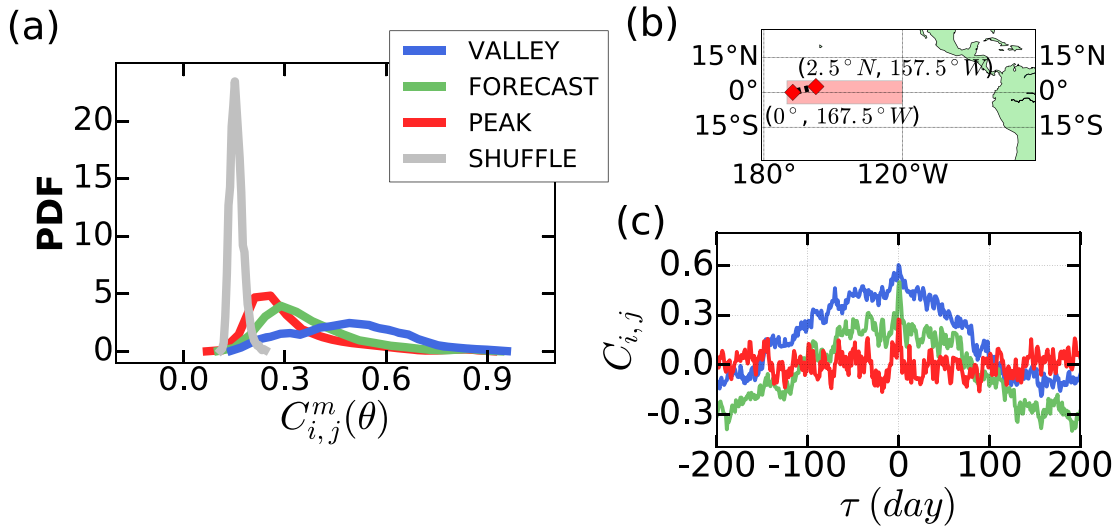

Figure 2. (a) PDF of $C^m_{i,j}(\theta)$ for all links within Niño 3.4 region, in networks at VALLEY (blue), FORECAST (green), PEAK (red) months (see figure 1), and for the random network (gray). (c) Typical cross-correlation functions of a link (shown in (b)) within Niño 3.4 region (indicated by the light red shaded area) for months of VALLEY (blue), FORECAST (green), and PEAK (red) before the 2014–2016 El Niño event.

Next, we elaborate on the reasoning behind our approach. In figure 2(a), we plot the probability density function (PDF) of $C^{(t)}_{i,j}(\theta)$ for all links in network windows at which 'Valley' ($m = Valley$, blue), 'Forecast' ($m = Forecast$, green) and 'Peak' ($m = Peak$, red) occur, respectively. We compare these PDFs with a PDF of random networks that are obtained by shuffling the order of the calendar days for each node within the Niño 3.4 region. We find the strongest correlations for the 'Valley' periods (as the PDF is stretched toward higher values), then weaker correlations for the 'Forecast' periods, and then the weakest correlations for the 'Peak' periods (closest to the shuffled correlations). Thus, the Niño 3.4 network (region) becomes less coherent when progressing from 'Valley' periods to the 'Peak' periods. The order is reestablished towards the actual peak of El Niño. The evolution of the cross-correlation of a typical link (shown in figure 2(b)), before the onset of 2014–2016 El Niño event, is shown in figure 2(c). The three cross-correlation functions (blue, green, and red) correspond to the 'Valley', 'Forecast' and 'Peak' points marked by blue, green and red arrows in figure 1(b). Consistently, we find that the maximal values of the cross-correlation function, $C^{(t)}_{i,j}(\theta)$, decreases from time of 'Valley' to 'Forecast' time, then to the 'Peak' time.

Moreover, while $C^{(t)}_{i,j}(\theta)$ is decreasing from Valley to Peak months, the strength of the link $W^{(t)}_{i,j}$ [24] is increasing. This difference is probably due to the autocorrelation of the temperature anomaly variations in the Niño 3.4 region [39]; see figure S1.

4. Results

4.1. Forecasting the magnitude of El Niño

We now examine the accuracy and robustness of our forecast for the magnitude of El Niño events between 1950 and present (since the ONI begin from 1950), using several datasets. For this purpose, we plot the predicted magnitude versus the observed magnitude of El Niño (scatter plot), and use the Pearson correlation coefficient, r, to quantify the correlation. We present such scatter plots in figure 3.

Next, we apply the Kolmogorov–Smirnov test to quantify the significance of the relationship between the predicted and observed magnitude of El Niño; figure 3 (insets). Each time we randomly choose ten events and calculate the correlation coefficient between their predicted and observed magnitudes; we repeated this procedure 1 million times, and obtained the PDF of r-values for each dataset (colored by green in figure 3). For a comparison, we also consider random cases as follows. Each time we choose randomly ten predicted values and randomly ten observed values and then perform a linear regression between them; also here we have performed 1 million selections, and obtain the PDF of r-values for each dataset (colored by gray in figure 3). Then we compare the PDFs of observed r-values to the random r-values using Kolmogorov–Smirnov statistic D [40]. For each dataset used here, D is relatively large ($D \geqslant 0.37$), indicating significant difference between the observed and predicted El Niño magnitude.

The results are summarized in table 1, in the rows heading 'El Niño magnitude'. We note however, that the prediction of the magnitude of El Niño is performed at the actual onset of El Niño, which on average occurs about half a year prior to the peak of El Niño.

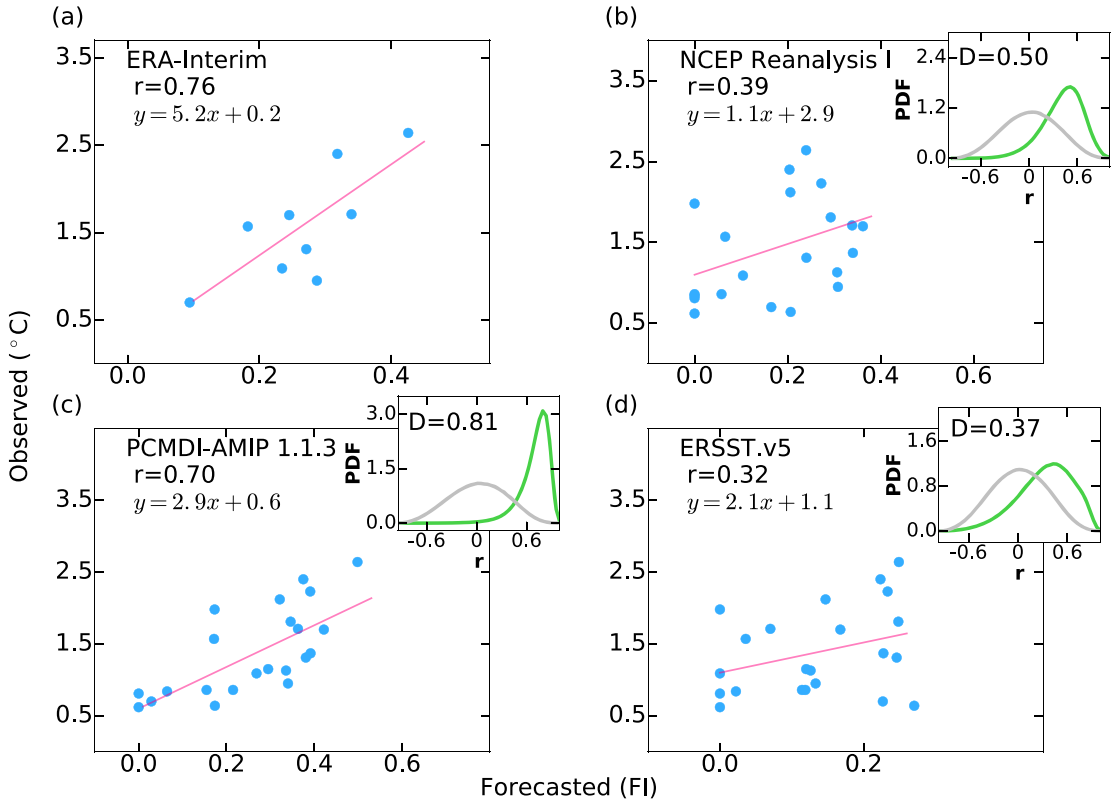

Figure 3. Scatter plots of the observed El Niño magnitude versus the forecast ones for different datasets (different panels) and the Kolmogorov–Smirnov statistic (inset of each panel). The different datasets are: (a) ERA-Interim [27], (b) NCEP Reanalysis I [28], (c) PCMDI-AMIP 1.1.1 [29] and (d) ERSST.v5 [30]. The red lines indicate the linear regression fits. We also show the corresponding correlation (r-values), Kolmogorov–Smirnov statistic (D), and the function of the fitting lines.

4.2. Forecasting the onset of El Niño

Next we examine the forecasting power of the onset of El Niño. The results are summarized in table 1, in the rows heading 'El Niño onset'. Here, the hit rate $= \frac{\text{hits}}{\text{hits} + \text{misses}}$, where in the numerator, the 'hits' is the number of true positive prediction of El Niño, which is composed of two parts, 'normal' + 'delayed', and the denominator is the number of all the El Niño events for each dataset. The False alarm rate $= \frac{\text{false alarms}}{\text{false alarms} + \text{correct rejections}}$, where in the denominator the number of 'correct rejections' equals the number of years where no El Niño started and no false alarm appeared in the past 18 months before the year. The lead time equals the time from the potential signal (or the end of La Niña, if La Niña is experienced after the prediction) to the actual onset of El Niño (shaded areas in figure 1).

Previous studies proposed various methods to forecast El Niño events. Some of these predict quite successfully the onset of El Niño, about one year in advance [24, 25]. We compare our prediction method to prediction of the 12-mo forecasting scheme based on climate network approach [24] and to the prediction of state-of-the-art models—the COLA anomaly coupled model [41] and the Chen–Cane model [13]; for this purpose we use the operating characteristics (ROC) [42], see figure S3. The resulting hit rate of our approach is $\geqslant$0.81, and the lowest (worse) hit rate of 0.81 is obtained for NCEP Reanalysis I. Meanwhile, the false alarm rate of our approach is $\leqslant$0.24, and the highest (worse) false alarm ratio of 0.24 is obtained for ERSST.v5. For prediction lead time of 12 months the hit rate is $<$ 0.4 for the COLA model [41] and $<$ 0.45 for the Chen–Cane model [13] with false alarm ratio of $\sim$0.2. The hit rate for the network approach in [24] is 0.667 and false alarm rate is 0.095.

The prediction scheme we proposed here improves the prediction of the onset of El Niño. An additional and also the most important advantage of the prediction scheme we propose is that it provide prediction both for the magnitude and the onset of El Niño based only on the temperature variability and their coherence in the Niño 3.4 region.

5. Summary

In summary, we introduce a new FI that is based on climate networks which accurately and simultaneously forecasts both the onset and magnitude of El Niño. The performance of the FI is examined successfully on

several datasets. Our forecasting algorithm is based on the finding that the similarity or the coherence of low frequency temporal variability of temperature anomaly between different sites (strength of links) in the Niño 3.4 region decreases well before El Niño and increases at the onset of El Niño. The magnitude of the predicted El Niño is positively related with the highest peak in the FI during the period between the end of last El Niño and the onset of the new one. The results presented here indicate an important characteristic of the phase of the ENSO cycle, i.e., significant increase of disorder occurs in the Niño 3.4 region well before the onset of El Niño. The relationship between El Niño and the variation of the degree of disorder in the Niño 3.4 region may be further explained by defining an entropy based on the coherence of temperature variations in different sites of the Niño 3.4 region, which oscillates periodically with the ENSO cycle. There is surely a room of further improvement of the forecasting algorithm proposed here, probably with combination with other forecasting techniques and models.

Acknowledgments

We thank Kai Xu for helpful discussions and suggestions. We acknowledge the Israel-Italian collaborative project NECST, the Israel Science Foundation, ONR, Japan Science Foundation, BSF-NSF, and DTRA (Grant No. HDTRA-1-10-1-0014) for financial support. JF thanks the fellowship program funded by the Planning and Budgeting Committee of the Council for Higher Education of Israel.

References

[1] Dijkstra H A 2005 *Nonlinear Physical Oceanography: A Dynamical Systems Approach to the Large Scale Ocean Circulation and El Niño* (New York: Springer)
[2] Clarke A J 2008 *An Introduction to the Dynamics of El Niño and the Southern Oscillation* (London: Academic)
[3] Cane M A and Sarachik E S 2010 *The El Niño-Southern Oscillation Phenomenon* (Cambridge: Cambridge University Press)
[4] Halpert M S and Ropelewski C F 1992 Surface temperature patterns associated with the Southern Oscillation *J. Clim.* **5** 577
[5] Diaz H F, Hoerling M P and Eischeid J K 2001 ENSO variability, teleconnections and climate change *Int. J. Climatol.* **21** 1845
[6] Ropelewski C F and Halpert M S 1986 North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO) *Mon. Weather Rev.* **114** 2352
[7] Kumar K K *et al* 2006 Unraveling the mystery of Indian monsoon failure during El Niño *Science* **314** 115
[8] Ihara C, Kushnir Y and Cane M A 2007 Indian summer monsoon rainfall and its link with ENSO and Indian Ocean climate indices *Int. J. Climatol.* **27** 179
[9] Hsiang S M, Meng K C and Cane M A 2011 Civil conflicts are associated with the global climate *Nature* **476** 438
[10] Burke M, Gong E and Jones K 2015 Income shocks and HIV in Africa *Econ. J.* **125** 1157
[11] Schleussner C F, Donges J F, Donner R V and Schellnhuber H J 2016 Armed-conflict risks enhanced by climate-related disasters in ethnically fractionalized countries *Proc. Natl Acad. Sci. USA* **113** 9216
[12] McPhaden M J *et al* 1998 The Tropical Ocean-Global Atmosphere observing system: a decade of progress *J. Geophys. Res.* **103** 14169
[13] Chen D and Cane M A 2008 El Niño prediction and predictability *J. Comput. Phys.* **227** 3625
[14] Cane M A, Zebiak S E and Dolan S C 1986 Experimental forecasts of El Niño *Nature* **321** 827
[15] Barnett T P *et al* 1988 On the prediction of the El Niño of 1986–1987 *Science* **241** 192
[16] Tang F T, Hsieh W W and Tang B 1997 Forecasting the equatorial Pacific sea surface temperatures by neural network models *Clim. Dyn.* **13** 135
[17] Kirtman B P *et al* 1997 Multiseasonal predictions with a coupled tropical ocean global atmosphere system *Mon. Weather Rev.* **125** 789
[18] Chen D, Cane M A, Kaplan A, Zebiak S E and Huang D 2004 Predictability of El Niño over the past 148 years *Nature* **428** 733
[19] Luo J, Masson S, Behera S K and Yamagata T 2008 Extended ENSO predictions using a fully coupled ocean-atmosphere model *J. Clim.* **21** 84
[20] Yeh S W *et al* 2009 El Niño in a changing climate *Nature* **461** 511
[21] Galanti E, Tziperman E, Harrison M, Rosati A and Sirkes Z 2003 A study of ENSO prediction using a hybrid coupled model and the adjoint method for data assimilation *Mon. Weather Rev.* **131** 2748
[22] ENSO Cycle: Recent Evolution, Current Status and Predictions, Climate Prediction Center/NCEP http://cpc.ncep.noaa.gov/products /analysis_monitoring/lanina/enso_evolution-status-fcsts-web.pdf (Accessed: August 2017)
[23] Ludescher J, Gozolchiani A, Bogachev M, Bunde A, Havlin S and Schellnhuber H J 2014 Very early warning of next El Niño *Proc. Natl Acad. Sci. USA* **111** 2064
[24] Ludescher J, Gozolchiani A, Bogachev M, Bunde A, Havlin S and Schellnhuber H J 2013 Improved El Niño forecasting by cooperativity detection *Proc. Natl Acad. Sci. USA* **110** 11742
[25] Meng J, Fan J, Ashkenazy Y and Havlin S 2017 Percolation framework to describe El Niño conditions *Chaos* **27** 035807
[26] ŁOceanic Niño Index (ONI) https://esrl.noaa.gov/psd/data/correlation/oni.data (Accessed: August 2017)
[27] Dee D P *et al* 2011 The ERA-Interim reanalysis: configuration and performance of the data assimilation system *Q. J. R. Meteorol. Soc.* **137** 656
[28] Kalnay E *et al* 1996 The NCEP/NCAR 40 year reanalysis project *Bull. Am. Meteorol. Soc.* **77** 437
[29] Taylor K E, Williamson D and Zwiers F 2000 *The Sea Surface Temperature and Sea Ice Concentration Boundary Conditions for AMIP II Simulations* PCMDI Report 60 Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory 25p available as a PDF (For further descriptive details, see http://pcmdi.llnl.gov/projects/amip/AMIP2EXPDSN/BCS/index.php)
[30] Huang B *et al* 2017 Extended reconstructed sea surface temperature, version 5 (ERSSTv5): upgrades, validations, and intercomparisons *J. Clim.* **30** 8179
[31] Yamasaki K, Gozolchiani A and Havlin S 2008 Climate networks around the globe are significantly affected by El Niño *Phys. Rev. Lett.* **100** 228501
[32] Tsonis A A and Swanson K L 2008 Topology and predictability of El Niño and La Niña networks *Phys. Rev. Lett.* **100** 228502
[33] Donges J F, Zou Y, Marvan N and Kurths J 2009 Complex networks in climate dynamics *Eur. Phys. J. Spec. Top.* **174** 157
[34] Donges J F, Zou Y, Marvan N and Kurths J 2009 The backbone of the climate network *Europhys. Lett.* **87** 48007

[35] Gozolchiani A, Havlin S and Yamasaki K 2011 Emergence of El Niño as an autonomous component in the climate network *Phys. Rev. Lett.* **107** 148501
[36] Wang Y, Gozolchiani A, Ashkenazy Y, Berezin Y, Guez O and Havlin S 2013 Dominant imprint of rossby waves in the climate network *Phys. Rev. Lett.* **111** 138501
[37] Zhou D, Gozolchiani A, Ashkenazy Y and Havlin S 2015 Teleconnection paths via climate network direct link detection *Phys. Rev. Lett.* **115** 268501
[38] Fan J, Meng J, Ashkenazy Y, Havlin S and Schellnhuber H S 2017 Network analysis reveals strongly localized impacts of El Niño *Proc. Natl Acad. Sci. USA* **114** 7543
[39] Guez O, Gozolchiani A and Havlin S 2014 Influence of autocorrelation on the topology of the climate network *Phys. Rev.* E **90** 062814
[40] Sheldon M R 1987 *Introduction to Probability and Statistics for Engineers and Scientists* (New York: Wiley)
[41] Kirtman B P 2003 The COLA anomaly coupled model: ensemble ENSO prediction *Mon. Weather Rev.* **131** 2324
[42] Mason S J and Graham N E 1999 Conditional probabilities, relative operating characteristics, and relative operating levels *Wea. Forecast.* **14** 713