



# The Application of Machine Learning Techniques to Improve El Niño Prediction Skill

Henk A. Dijkstra<sup>1,2\*</sup>, Paul Petersik<sup>1</sup>, Emilio Hernández-García<sup>3</sup> and Cristóbal López<sup>3</sup>

<sup>1</sup> Department of Physics, Institute for Marine and Atmospheric Research Utrecht, Utrecht University, Utrecht, Netherlands,

<sup>2</sup> Department of Physics, Center for Complex Systems Studies, Utrecht University, Utrecht, Netherlands, <sup>3</sup> IFISC (Spanish National Research Council - University of the Balearic Islands), Instituto de Física Interdisciplinar y Sistemas Complejos, Palma de Mallorca, Spain

We review prediction efforts of El Niño events in the tropical Pacific with particular focus on using modern machine learning (ML) methods based on artificial neural networks. With current classical prediction methods using both statistical and dynamical models, the skill decreases substantially for lead times larger than about 6 months. Initial ML results have shown enhanced skill for lead times larger than 12 months. The search for optimal attributes in these methods is described, in particular those derived from complex network approaches, and a critical outlook on further developments is given.

## OPEN ACCESS

**Keywords:** El Niño, prediction, machine learning, neural networks, attributes, climate networks

### Edited by:

Raul Vicente,  
Max-Planck-Institut für Hirnforschung,  
Germany

### Reviewed by:

William W. Hsieh,  
University of British Columbia, Canada  
Diego R. Amancio,  
University of São Paulo, Brazil

### \*Correspondence:

Henk A. Dijkstra  
h.a.dijkstra@uu.nl

### Specialty section:

This article was submitted to  
Interdisciplinary Physics,  
a section of the journal  
Frontiers in Physics

**Received:** 15 June 2019

**Accepted:** 23 September 2019

**Published:** 10 October 2019

### Citation:

Dijkstra HA, Petersik P,  
Hernández-García E and López C  
(2019) The Application of Machine  
Learning Techniques to Improve  
El Niño Prediction Skill.  
*Front. Phys.* 7:153.  
doi: 10.3389/fphy.2019.00153

## 1. INTRODUCTION

Techniques of Artificial Intelligence (AI) and Machine Learning (ML) are very well developed [1], and massively applied in many scientific fields, like in medicine [2], finance [3], and geophysics [4]. Although the application to climate research has been around for a while [5–7], there is much renewed interest recently [8–10]. A main issue in which breakthroughs are expected is the representation of unresolved processes (e.g., clouds, ocean mixing) in numerical weather prediction models and in global climate models. For example, recently a ML-inspired (random-forest) parameterization of convection gave accurate simulations of climate and precipitation extremes in an atmospheric circulation model [11]. ML has also been used to train statistical models which mimic the behavior of climate models [12, 13]. Another area of potential breakthrough is the skill enhancement of forecasts for weather and particular climate phenomena, such as the El Niño-Southern Oscillation (ENSO) in the tropical Pacific.

During an El Niño, the positive phase of ENSO, sea surface temperatures in the eastern Pacific increase with respect to average values and upwelling of colder, deep waters diminishes. The oscillation phase opposite to El Niño is La Niña, with a colder eastern Pacific and increased upwelling. A measure of the state of ENSO is the NINO3.4 index (**Figure 1A**), which is the area-averaged Sea Surface Temperature (SST) anomaly (i.e., deviation with respect to the seasonal cycle) over the region 170–120°W × 5°S–5°N. Averaging over other areas defines other indices such as NINO3. For ENSO predictions, often the Oceanic Niño Index (ONI) is used which refers to the 3-months running mean of the NINO3.4 index.

El Niño events typically peak in boreal winter, with an irregular period between two and seven years, and strength varying irregularly on decadal time scales. The most recent strong El Niño had its maximum in December 2015 (**Figure 1A**). The spatial pattern of ENSO variability is often represented by methods from principal component analysis [14], detecting patterns of maximal variance. The first Empirical Orthogonal Function (EOF) of SST anomalies, obtained from the

Hadley Centre Sea Ice and Sea Surface Temperature (HadISST) dataset [15] over the period 1950–2010, shows a pattern strongly confined to the equatorial region with largest amplitudes in the eastern Pacific (**Figure 1B**).

El Niño events typically cause droughts on the western part of the Pacific and flooding events on the eastern part and hence affect climate worldwide. Estimated damages for the 1997–1998 event were in the order of billions of US\$ [16]. The development of skillful forecasts of these events, preferably with a one year lead time, is hence important. These forecasts will enable policy makers to mitigate the negative impacts of the associated weather anomalies. For example, farmers can be advised to use particular types of corn in El Niño years and others during La Niña years (see e.g., <http://globalagrisk.com>).

Although more detailed regional measures are sometimes desired in a forecast, most focus is on spatially averaged indices such as the NINO3.4 (cf. **Figure 1A**). Forecasting this time series is an initial value problem requiring the specification of initial conditions (of relevant observables) and a model, which can be either statistical or dynamical. With this model, one can predict future values of these observables or of other ones from which meaningful diagnostics, such as the NINO3.4 index, can be obtained. Due to many efforts in the past, detailed observations of relevant oceanic and atmospheric variables are available (since the mid-1980s) through the TAO-TRITON observation array in the tropical Pacific, and satellite data of sea surface height, surface wind stress and sea surface temperature [17]. In addition, reanalysis data (i.e., model simulations which assimilate existing observations) such as ERA-Interim [18] provide a rather detailed characterization of present and past state of the Pacific, essential for successful prediction of the future.

This paper provides an overview of efforts to use ML, mainly Artificial Neural Network (ANN) approaches, to predict El Niño events, and putting them in the context of classical prediction methodologies. In section 2, we describe the state-of-the-art in current prediction practices, the efforts to understand the results, and in particular what determines the skill of these forecasts. Then results of ML-based approaches are described in section 3 and challenges and outlook are described in section 4.

## 2. EL NIÑO PREDICTION: STATE OF THE ART

There have been many reviews on El Niño predictability (e.g., [19–22]) and a recent one [23], reviewing also most of the Chinese-community studies on this topic. Over the last decade, a multitude of models is used for El Niño prediction and results are available at several websites. Multi-model ensemble results are given at the International Research Institute for Climate and Society (IRI)<sup>1</sup> providing results from both dynamical models (i.e., models based on underlying physical conservation laws) and statistical models (those capturing behavior of past statistics). The NCEP Climate Forecast System CFSv2 [24]<sup>2</sup>, provides a

dynamical single-model ensemble forecast. The forecast systems developed in China, such as the SEMAP2 and the NMEFC/SOA are discussed in detail in Tang et al. [23] so they are further discussed here. It is illustrative to show the results of both the IRI and CFSv2 model systems for the last strong El Niño event, that of 2015–2016, which was discussed in detail by L'Heureux et al. [25]. Forecasts starting in June 2015 are shown in **Figure 2** indicating that these models are able to provide a skillful forecast of NINO3.4. Nevertheless, the dispersion in the predictions of the different models is huge, and even between ensemble members of the same model, highlighting the difficulty of reliable prediction.

The US National Oceanic and Atmospheric Administration (NOAA) will release an El Niño advisory when (i) the 1-month NINO-3.4 index value is at or in excess of 0.5°C, (ii) the atmospheric conditions are consistent with El Niño (i.e., weaker low-level trade winds, enhanced convection over the central or eastern Pacific Ocean), and (iii) at least five overlapping seasonal (3-months average) NINO3.4 SST index values are at or in excess of 0.5°C, supporting the expectation that El Niño will persist. The purpose of the forecasting efforts such as those in **Figure 2** is to predict in advance when those conditions will occur. Both the IRI and CFSv2 predicted already in June 2014 an El Niño event for next winter which turned out to be wrong as there was a dip in NINO3.4 at the end of 2014 (due to easterly winds). However, most models did very well in predicting (from June 2015) the winter 2015–2016 strong event (see **Figure 2**).

The skill of El Niño forecasts is usually measured by the anomaly correlation coefficient (AC) given by:

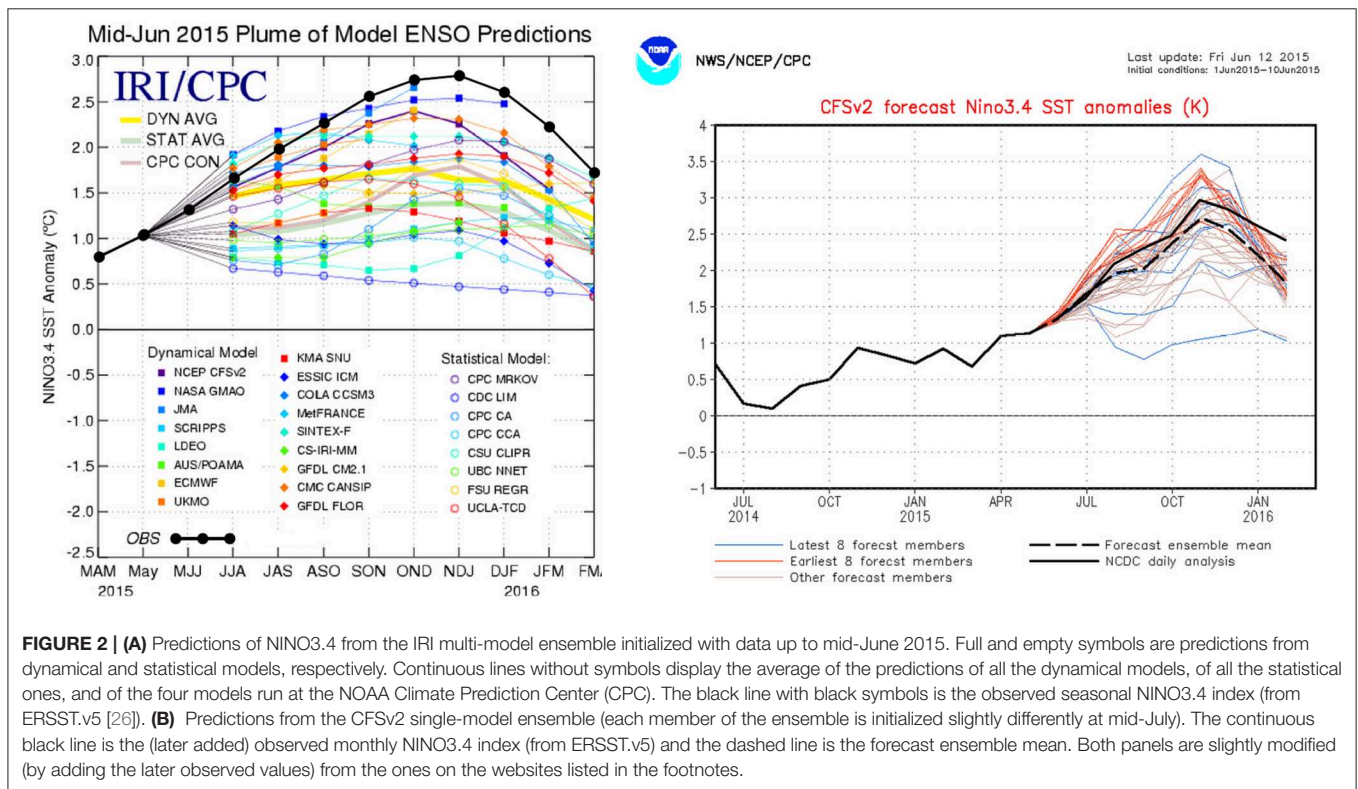
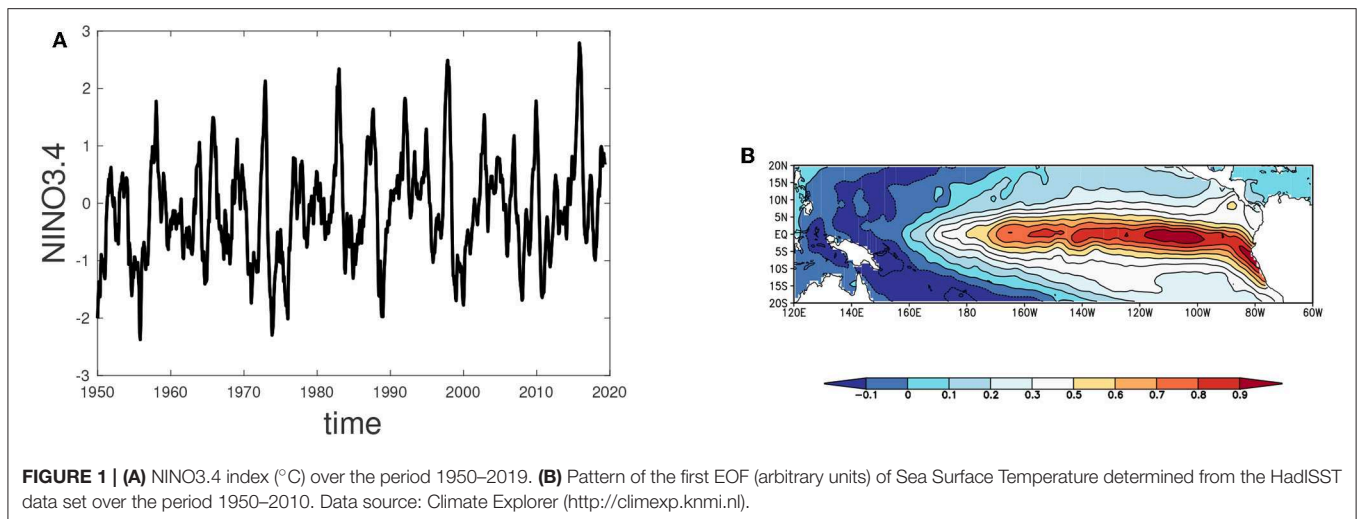
$$AC = \frac{\overline{m'o'}}{\sigma_m \sigma_o}, \quad (1)$$

where  $m'$  indicates NINO3.4 index of the model,  $o'$  that of observations and  $\sigma_x$  indicates the standard deviations of the time series  $x$ . The overbar indicates averaging of all time series elements. The AC is the Pearson correlation coefficient between prediction and observation. In Barnston et al. [21], the skill of the models over the period 2002–2011 was summarized with help of **Figure 3A** which indicates that skill beyond a 6-months lead time becomes overall lower than 0.5. Some general conclusions from these and many other prediction exercises are that (i) dynamical models do better than statistical models and (ii) models initialized before the Northern-hemispheric spring perform much worse than models initialized after spring. The latter notion is known as the “spring predictability barrier” problem. The concept of *persistence* is another way to look at the predictability barrier. It can be defined in terms of autocorrelation coefficients and in particular their decay with increasing lead times. SST anomalies originating from the spring seasons have the least persistence while those originating from summer seasons (**Figure 3B**) tend to have the greatest persistence [27].

El Niño events are difficult to predict as they have an irregular occurrence, and each time have a different development [17, 22]. The ENSO phenomenon is thought to be an internal mode of the coupled equatorial ocean-atmosphere system which can be self-sustained or excited by random noise [28]. The interactions

<sup>1</sup>[https://iri.columbia.edu/our-expertise/climate/forecasts/enso/current/?enso\\_tab=enso-sst\\_table](https://iri.columbia.edu/our-expertise/climate/forecasts/enso/current/?enso_tab=enso-sst_table)

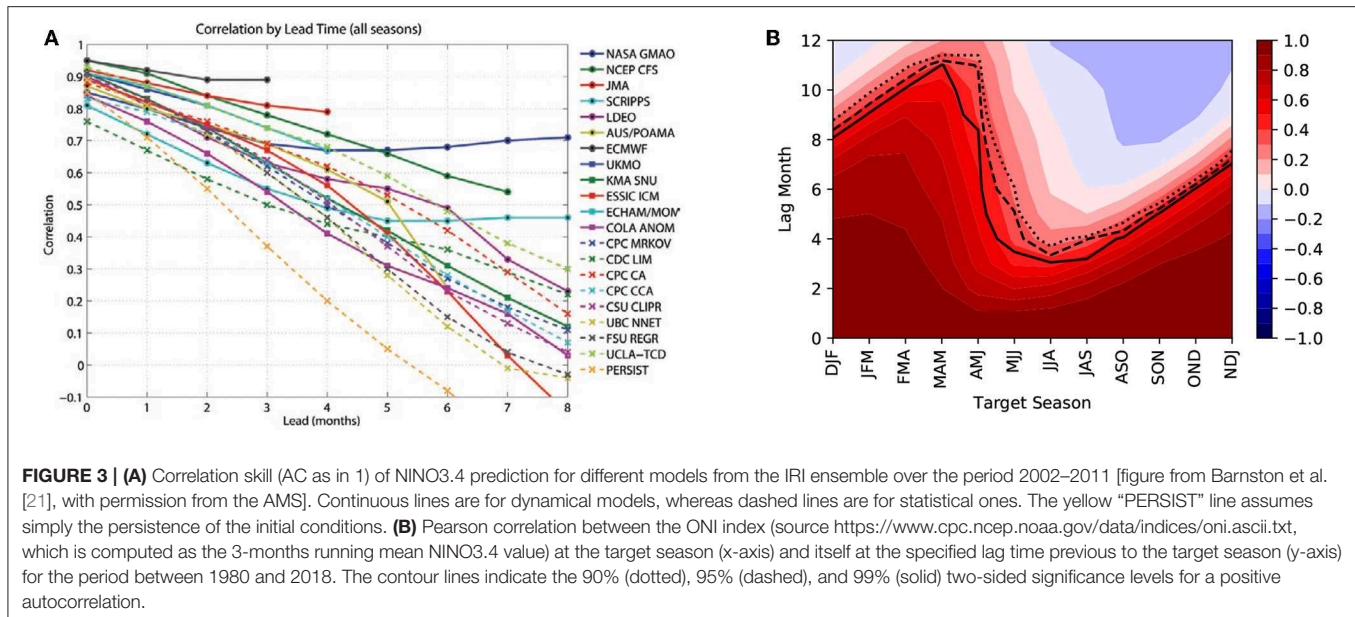
<sup>2</sup><https://www.cpc.ncep.noaa.gov/products/CFSv2/CFSv2seasonal.shtml>



of the internal mode and the external seasonal forcing can lead to chaotic behavior through nonlinear resonances [29, 30]. On the other hand, the dynamical behavior can be strongly influenced by noise, in particular westerly wind bursts [31] which can either be viewed as additive [32] or multiplicative noise [33]. Coupled processes between the atmosphere and ocean are seasonally dependent. During boreal spring the system is most susceptible to perturbations [34] leading to a spring predictability barrier [35]. The growth of perturbations from a certain initial state has been investigated in detail from one of the available intermediate-complexity models, the Zebiak-Cane model (ZC,

[36]), using the methodology of optimal modes [37–39]. It was indeed shown that spring is the most sensitive season as perturbations are amplified over a 6-months lead time.

In summary, the low skill after 6 months as seen in **Figure 3A** is believed to be due to both effects of smaller scale processes (noise) and nonlinear effects. Moreover, it is supposed that the period 2002–2011 was particularly difficult to predict because of the frequent occurrence of central Pacific (CP) El Niño types [40] for which the zonal advection feedback plays a key role for the development. In contrast, for an eastern Pacific (EP) El Niño, the thermocline feedback is most important [41]. As can be seen from



the NINO3.4 time series (Figure 1A) strong events appear about every 15 years (1982, 1996, 2015). There are likely other factors involved in the prediction skill of these strong events [42, 43], which we do not further discuss here.

### 3. MACHINE LEARNING APPROACHES

ML is being used in a variety of tasks that include regression and classification. ML algorithms can be divided into three main categories [44]: supervised, unsupervised, and reinforcement learning. In supervised learning, a model is build from labeled instances. In an unsupervised model, there are no labeled instances and the goal is to find hidden patterns (e.g., clustering) in the available data. In reinforcement learning, a particular target is pursued and feedbacks from the environment drive the learning process [1]. The usual procedure in supervised learning is as follows: the predictor model [e.g., an artificial neural network (ANN) or genetic programming (GP)] is trained with data from a training set in order to determine a set of optimal parameter values. Then the generalization capabilities of the model are tested on a validation data set. Once the predictor model is validated, a third so-called test data set that was hold out during training and validation can be used to evaluate the prediction skill.

Many types of ML methodologies have been developed. In ANNs, the basic element is the neuron, or perceptron (i.e., logistic or other function units which locally discriminate different inputs). An ANN has a multilayer structure—an input layer, an output layer and a few (or zero) hidden layers, in which each neuron is connected to all neurons in the previous and following layers. The system thus maps some input applied to the input layer to some output or prediction. The weights of the neuron connections are tuned to provide the optimal predictor model. Another ML technique, GP, is a symbolic regression

method used to find, by optimization procedures inspired by biological evolutionary processes, the functional form that fits the available data [45]. *Reservoir computing* [46] is another type of ML methodology in which input is injected into a high-dimensional dynamical system called “reservoir.” The response of the reservoir is recorded at particular output nodes with associated “output weights,” and linear regression is used to optimize these weights so that the recorded response performs the desired prediction.

Although there are a few ML attempts to forecast El Niño events by evolutionary or genetic algorithms [47, 48], and by other methods [49], we will focus here on ML prediction schemes based on the most popular approach, which is the use of feed-forward ANNs. Such ANNs with at least one hidden layer, also called multilayer perceptrons, have the powerful capability to approximate any nonlinear function to an arbitrary accuracy given enough input data and hidden-layer neurons (e.g., [1, 50]).

#### 3.1. Early ML Approaches

There is a large freedom on the implementation of ANN methods and choices have to be made regarding which variables to use as inputs (called in this context the *attributes*, *features*, or *predictors*), the architecture of the ANN and training method. ANNs, as any other supervised learning technique, require to split the available data in at least two parts: a training set, on which parameters of the ANN are optimized, and a test set, on which the skill of the optimized ANN is evaluated. Furthermore, it is good practice to use a third data set, often called validation data set, to tune hyperparameters and to check for overfitting. For ENSO prediction, it is of particular importance to split the data into connected time series. If instead the data set would be split by random sampling, training and test data points would be temporally close to each other. Due to the strong autocorrelations within the ENSO system, the test data set would be strongly



correlated with the training data set and hence could not serve as an independent data set. Because of the shortness of the available time series, the fact that El Niño repeats only about every four years on average, and that only a subset of them are strong, training and validation sets do not contain many significant events and statistical estimation of ANN skills is not very precise.

In using ANN's one can basically focus on two different supervised learning tasks: classification (will there be an El Niño event or not) and regression (predicting an index, e.g., the NINO3.4, with a certain lead time). Early ANN-based El Niño predictions [51] for the regression task used as predictors wind-stress fields and the NINO3.4 time series itself. More explicitly, the time series of the seven leading principal components of the wind-stress field (i.e., the amplitudes of their seven leading EOFs) in a large region of the tropical Pacific were averaged seasonally in each of the four seasons previous to the start of prediction. These numbers, together with the last value of the NINO3.4 time-series make a total of  $4 \times 7 + 1 = 29$  inputs to be fed into the ANN. Tangang et al. [52] noticed that using sea level pressure (SLP) fields gave better results at long lead times than using wind-stress fields. Also, averaging forecasts from an ensemble of ANNs with different random weights assigned to the neurons at the start of the learning phase improved results with respect to using the results of a single ANN. Maas et al. [53] further analyzed this fact and suggested using it to estimate prediction reliability. Tangang et al. [54] simplified the ANN architecture by using extended EOFs (EEOFs), which project the observed fields (wind stress or SLP) onto spatio-temporal patterns, instead of on spatial ones (using EOFs). In this way, input from the year previous to the forecast start was compressed to  $7 + 1 = 8$  variables, instead of the previous 29. In these earlier studies, all of which used a single hidden layer in the ANN, high forecast skills (values of the correlation  $AC$  above 0.6 even at lead times above one year) were reported. However, there were large differences in performance depending, for example, on the season of the year or the particular year or decade being predicted.

Later implementations of these early ANN methods indicated that the skill was relatively low. For example, the curve labeled UBC-NNET in **Figure 3A** (coming from the ANN model by the University of British Columbia group, based in Tangang et al. [54]) has the second lowest skill at 6 months lead time, improving only the forecast made by the simple “persistence” assumption. This can partially be attributed to the fact that the model architecture of the UBC-NNET changed in May 2004 from predicting the ONI to predicting the amplitudes of the leading EOFs. There are also differences in the climatology used by UBC-NNET and the one used for the tests in Barnston et al. [21]. Moreover, only during December 2004 and November 2005 the UBC-NNET included subsurface temperature data, while the thermal state of the subsurface can contain important information about the future state of the ENSO [55]. For the remaining period, the model lacked this subsurface information. Ideas to improve ANN performance included optimization regularization [56] and linear corrections that help to quantify prediction errors [57]. Another one to focus on forecasting individual principal components of the SST field and combining them to obtain climatic indices such as NINO3.4 [58], instead

of trying to predict directly the climatic index. In general, for a one year lead time, no  $AC$  values higher than 0.5–0.6 were obtained with these methods, although larger  $AC$  values have been reported for specific seasons or years.

An exception is the work of Baawain et al. [59] in which very high correlations (above 0.8 for lead times between 1 and 12 months) were reported for prediction of the NINO3 index using as inputs the two surface-wind components and the SAT at four selected locations in the Pacific (thus, 12 inputs). The high forecast skill may arise from the careful and systematic determination of the ANN architecture (again a single hidden layer but with up to 16 neurons, and different activation functions), or perhaps from the choices of training and validation data sets. Some of the practices in Baawain et al. [59], however, are rather questionable and can lead to substantial overfitting for the ENSO prediction. First, they perform the hyperparameter optimization on their test data set. A better practice is to tune hyperparameters on an additional validation data set and hold out the test data set completely during the training and hyperparameter optimization. Second, they do not precisely report how the data is split into the training and test data set. If they split the data by random splitting, the model is likely overfitted due to the problem mentioned earlier. The very small difference between the prediction skill on the training ( $r = 0.91$ ) and the test ( $r = 0.90$ ) data set indicates that they might split their data set by random splitting. A better practice would be to split the data into two connected time series. In addition to pure ANN prediction, also hybrid approaches that use a dynamical ocean model driven by wind stresses provided by an ANN fed by the ocean state have been applied [60, 61]. Skill in predicting El Niño is similar to purely dynamical models, but at a smaller computational cost.

The key for a successful application of ANNs to ENSO prediction is to determine the correct attributes to include in the training of the model. The attributes used in Tangang et al. [54], based on EEOFs of SLP and SST, may be not optimal considering where the memory of the coupled ocean-atmosphere system originates from, i.e., from the subsurface ocean.

### 3.2. Attributes: Role of Network Science

Although network science had been applied to many other branches of science, it was not applied to climate science until one realized that easy mappings between continuous observables (e.g., temperature) and graphs could be made [62–64]. One can consider these observables to be on a grid (observation locations or of model grid points), which then are the “nodes” of the graph. A measure of correlation between the time series of an observable at two locations, such as the Pearson correlation or mutual information, can then be used to define a “connection” or “link,” and eventually to assign a “strength” or “weight” to that connection [65].

Ludescher et al. [66] used the link strength concept for El Niño prediction. They determined the average link strength  $S$  of the climate network constructed from a Surface Air Temperature (SAT) data set. They suggested that when  $S$  crosses a threshold  $\Theta$  while monotonically increasing, an El Niño will develop about one year later. The rationale behind this is that, during El Niño,

correlations of climatic variables at many locations with variables in the tropical Pacific are very high, so that an increase of these correlations, conveniently revealed by the connectivity (or “cooperativity”) of the climate network, is an indicator of an approach to an El Niño state. A training set over the period 1950–1980 was used to determine the threshold  $\Theta$ . The result for the test period 1980–2011 showed a remarkable skill of this predictor [66]. By using this method, also a successful prediction of the onset of the weak 2014 El Niño was made [67].

The time-varying characteristics of climate networks (where correlations defining link strength between nodes are calculated on successive time windows) has also been used in a different way. The increase of connectivity of the climate network, occurring when approaching an El Niño event, may lead to a *percolation transition* [68] in which initially disconnected parts of the network become connected into a single component.

The study by Rodríguez-Méndez et al. [69] introduced percolation-based early warnings in climate networks for an upcoming El Niño/La Niña event. Here, the climate networks are generated with a relatively high threshold for the cross-correlation between two nodes to be considered as connected. Hence, one finds a lot of isolated nodes in these networks.

However, even long before an El Niño event is approached these isolated nodes become connected to other ones building clusters of size two since correlations between nodes increase. If the correlation building continues, more small clusters of size two emerge and the proportion of nodes in clusters of size two, indicated by  $c_2$ , increases. Approaching further the transition, small clusters can connect to more nodes and form even bigger clusters counter-balancing the increased probabilities for smaller clusters. Hence, in a typical percolation transition, the first sign of the transition is indicated by a peak of  $c_2$ . This peak is followed by peaks in the proportion of nodes in clusters of increasing size, the closer the system is to the percolation point. At the percolation point spatial correlations in the system become so strong that a giant component in the network emerges and incorporates nearly all nodes of the system [69, 70]. If the system moves again away from the transition point, peaks in the proportion of nodes in clusters of different sizes appear in reversed order, starting with peaks coming from larger clusters followed by peaks from smaller clusters.

### 3.3. Recent ML-Based Predictions

Networks are often associated with machine learning techniques, either to generate attributes or to create the learning method itself [71]. The advantage of using the network approach in climate research is that, during network construction, the temporal information is often included to determine the properties of climate networks. In this way, the machine learning techniques will, by default, take the temporal information into account in making predictions of the future states of the system.

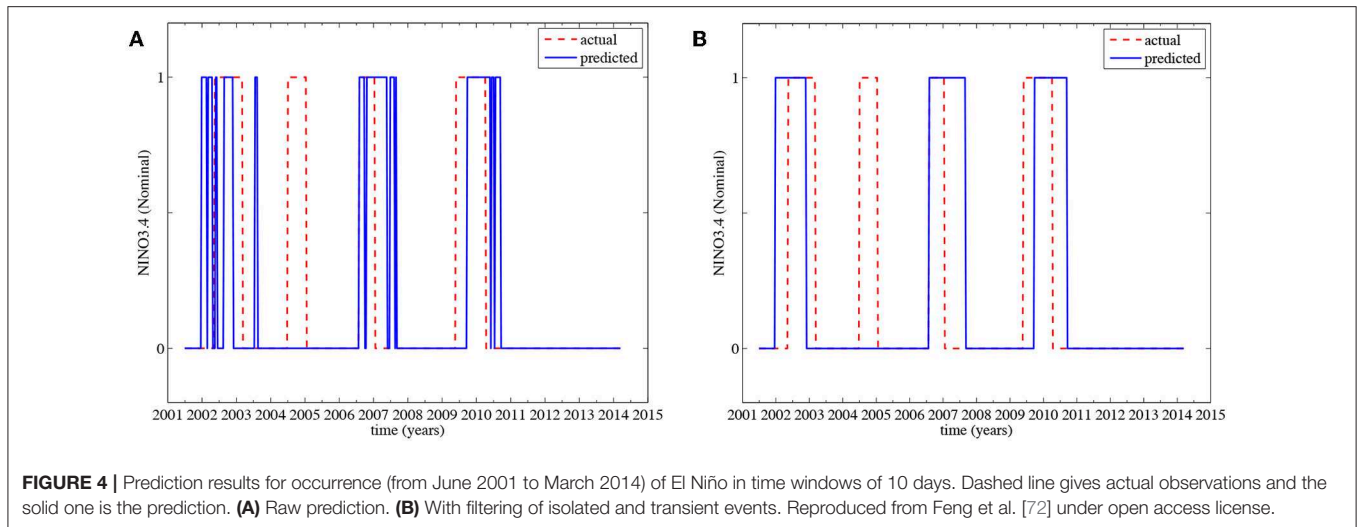
A first effort to combine complex network metrics with ANN’s for the prediction of the NINO3.4 index was made in Feng et al. [72]. They considered the classification problem (determining if El Niño will occur) with an ANN (two hidden layers with three neurons each) in which attributes were only the climate-network-based quantities from Gozolchiani et al. [64]. The period May

1949 to June 2001 was used as a training set, and the period June 2001 to March 2014 as the test set. The prediction lead time was set to 12 months. Classification results on the test set are shown in **Figure 4A**. Here a 1 indicates the occurrence of an El Niño event (in a 10-days window) and 0 indicates no event. When a filter is applied which eliminates the isolated and transient events and joins the adjacent events, the result is shown in **Figure 4B**. This forecasting scheme can hence give skillful predictions 12 months ahead for El Niño events.

The regression problem, i.e., forecasting the values of time series such as NINO3.4, was addressed by Nootboom et al. [73] who combined the use of network quantities with a thorough search for attributes based on the physical mechanism behind ENSO. A two-step methodology was used which resulted in a hybrid model for ENSO prediction. In a first step, a classical Autoregressive Integrated Moving Average (ARIMA) linear statistical method [74] is optimized to perform a linear forecast using past NINO3.4 values. Specifically, ARIMA(12,1,0) and ARIMA(12,1,1) were implemented, which means that the NINO3.4 values in the 12 months previous to the start of the prediction were used. The linear prediction was far from perfect, and then an ANN was trained from single-time attributes to forecast the residuals between the linear prediction and the true NINO3.4 values. The sum of the linear forecast and the nonlinear ANN prediction completes the final hybrid model forecast. In Hibon and Evgeniou [75], it is shown that, compared to a single prediction method, this hybrid methodology is more stable and reduces the risk of a bad prediction. This is probably due to the fact that long memory is taken into account, but not in the ANN part, which remains then relatively simple with respect to inputs and can then be more efficiently trained.

To motivate the choice of the attributes in the ANN, Nootboom et al. [73] used the ZC model [36]. In this model, the physical mechanisms of ENSO are clearly represented and it can be used for extensive testing of different attributes, specially network-based ones which contain correlations and spatial information. Several interesting network variables, such as the cross clustering and an eigenvalue quantifying the coupling between wind and SST networks, were determined from an analysis of the ZC model. More importantly, it revealed the importance of the dynamics of the thermocline, which can be quantified in properties of the thermocline-depth network or the related sea surface height (SSH) network. Also the zonal skewness in the degree field of the thermocline network and two variables related to a percolation-like transition [69, 70], namely the temporal increment in the size of the largest connected cluster and the fraction of nodes in clusters of size two  $c_2$  (see previous subsection) in the SSH network, had good prediction properties. These variables taken from the SSH network are related to the warm water volume (WWV, the integrated volume above the 20°C isotherm between 5°N–5°S and 120–280°E), which was also tested as input in the ANN forecast, and contain information on the physics of the recharge/discharge mechanism of ENSO [76]. It turns out that  $c_2$  performs better than WWV when used in long-lead-time predictions.

Furthermore, apart from these “recharge/discharge” related quantities, a sinusoidal seasonal cycle (SC), introducing



information needed for the phase locking of ENSO, and the second principal component ( $PC_2$ ) of a modified zonal component of the wind stress, which carries information on westerly wind bursts, were included as predictors. A single sinusoid is not a complete representation of the annual solar forcing, but it gives to the algorithm the phase information necessary to lock El Niño events to the annual cycle. The hybrid model improves on the CFSv2 ensemble at short lead times (up to 6 months) and it had also a better prediction result than all members of the CFSv2 ensemble in the case study of January 2010 [73]. From now on, the Normalized Root Mean Squared Error (NRMSE) is used to indicate the skill of prediction within the test set:

$$\begin{aligned} \text{NRMSE}(y^A, y^B) & \quad (2) \\ &= \frac{1}{\max(y^A, y^B) - \min(y^A, y^B)} \sqrt{\frac{\sum_{t_1^{\text{test}} \leq t_k \leq t_n^{\text{test}}} (y_k^A - y_k^B)^2}{n}}. \end{aligned}$$

Here  $y_k^A$ ,  $y_k^B$  are respectively the NINO3.4 index and its prediction at time  $t_k$  in the test set.  $n$  is the number of points in the test set. A low NRMSE indicates the prediction skill is better.

For short lead times, the hybrid model was used with the WWV,  $PC_2$ , SC and NINO3.4 itself as attributes. A temporal shift can be seen in the CFSv2 ensemble NINO3.4 results, both for the 3- and 6-months lead-time prediction (Figure 5). The hybrid model predictions used ARIMA(12,1,0) for the linear part, and the eighty-four possible ANN structures with three hidden layers with up to four neurons each were tested. Figure 5 shows the results from the structures giving the lowest NRMSE.

The prediction skill of the hybrid model decreased at a 6-months lead, while the shift and amplification of the CFSv2 prediction increased. Although the hybrid model did not suffer as much from the shift, at this lead time it underestimated (or missed) the El Niño event of 2010. In terms of NRMSE the hybrid model still obtained a better prediction skill than the CFSv2 (Figures 5A,B). The attributes from the shorter lead time

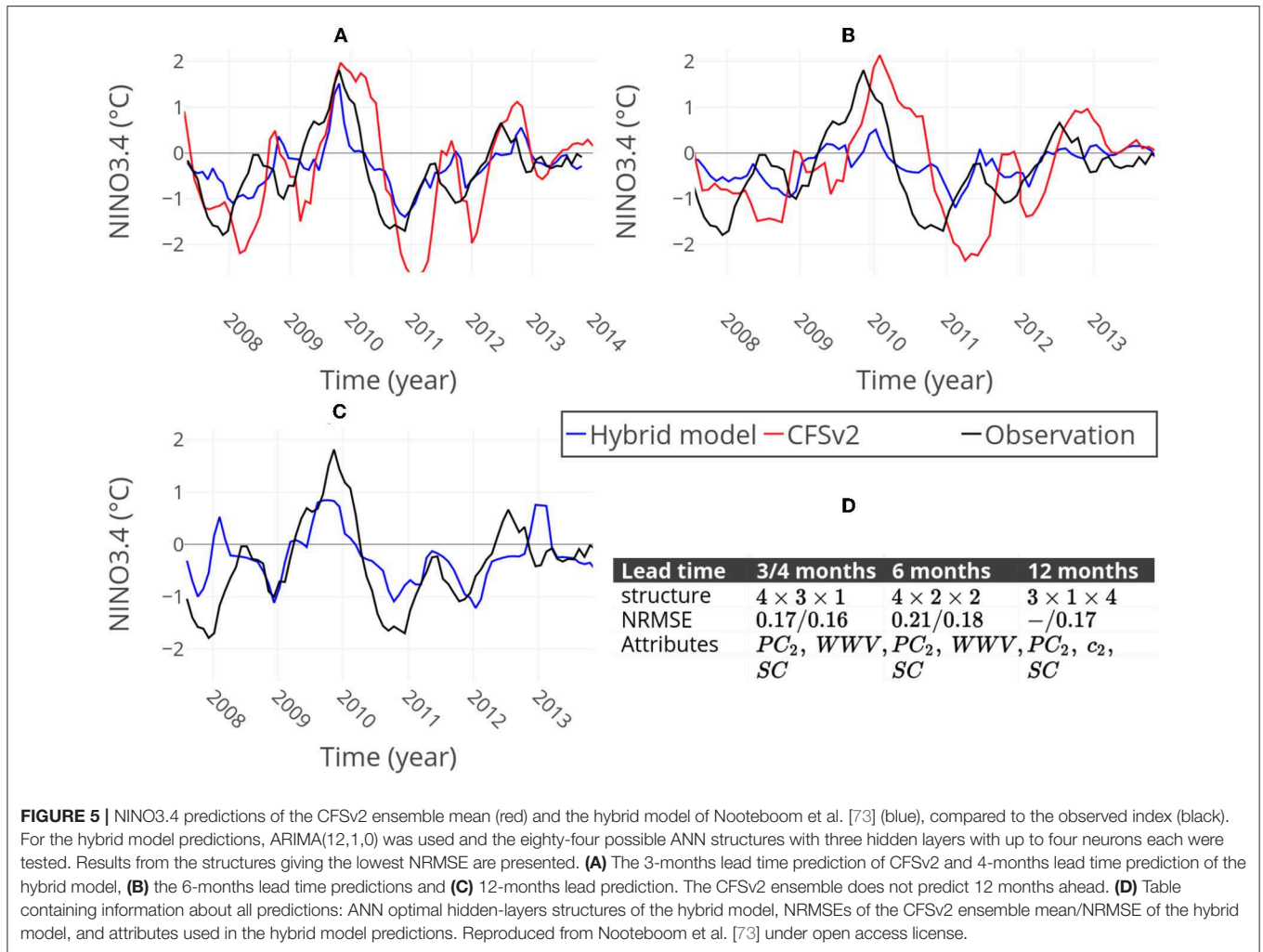
predictions were found to be insufficient for the 12-months-lead prediction. However,  $c_2$  of the SSH network was predictive at this lead time and hence the WWV was replaced by  $c_2$ . The 12-months lead time prediction of the hybrid model even improved the 6-months lead time prediction. On average the prediction did not contain a shift for this lead time (Figure 5C).

A prediction was made in Nooteboom et al. [73] for the year 2018, starting in May 2017 (Figure 6A). Different hybrid models were used at different lead times, always with ARIMA(12,1,0). The training set was from 1980 until May 2017 and the ANN structures used are the optimal ones at different lead times. For the predictions up to 5 months, the attributes WWV,  $PC_2$ , and SC were used whereas for the 12 months lead time prediction, the WWV was replaced by  $c_2$ . Here  $c_2$  was computed from the SSALTO/DUACS altimetry dataset<sup>3</sup>, which starts from 1993, and thus leads to a reduced training set. The hybrid model typically predicted much lower Pacific temperatures than the CFSv2 ensemble and was much closer to the eventual observations (black curve in Figure 6A). The uncertainty of the CFSv2 ensemble was large, since the spread of predictions is between a strong El Niño (NINO3.4 index between 1.5 and 2) and a moderate La Niña (NINO3.4 index between  $-1$  and  $-1.5^\circ\text{C}$ ) for the following 9 months, being the ensemble average prediction close to a neutral state. The hybrid model of Nooteboom et al. [73] predicted development of a strong La Niña (NINO3.4 index lower than  $-1.5^\circ\text{C}$ ) the coming year. There was indeed a La Niña event in 2017/2018, although the NINO3.4 index remained above  $-1^\circ\text{C}$ . A new prediction starting from December 2018 with the hybrid model is presented in Figure 6B, indicating the weak El Niño 2018–2019 to end by June 2019.

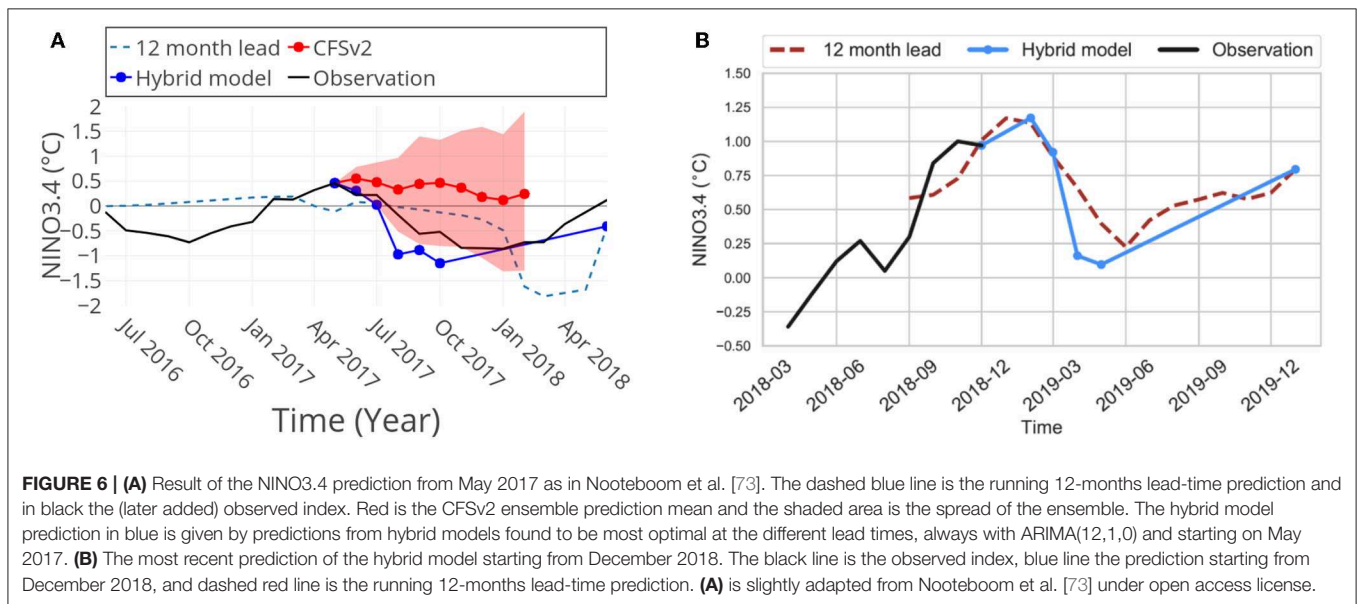
### 3.4. Prediction Uncertainty

In contrast to ensemble predictions of dynamical models, the proposed simple ANN-models lack the ability to estimate the predictive uncertainty. For instance, the ensemble spread of 84

<sup>3</sup><http://marine.copernicus.eu>



**FIGURE 5 |** NINO3.4 predictions of the CFSv2 ensemble mean (red) and the hybrid model of Nootboom et al. [73] (blue), compared to the observed index (black). For the hybrid model predictions, ARIMA(12,1,0) was used and the eighty-four possible ANN structures with three hidden layers with up to four neurons each were tested. Results from the structures giving the lowest NRMSE are presented. (A) The 3-months lead time prediction of CFSv2 and 4-months lead time prediction of the hybrid model, (B) the 6-months lead time predictions and (C) 12-months lead time prediction. The CFSv2 ensemble does not predict 12 months ahead. (D) Table containing information about all predictions: ANN optimal hidden-layers structures of the hybrid model, NRMSEs of the CFSv2 ensemble mean/NRMSE of the hybrid model, and attributes used in the hybrid model predictions. Reproduced from Nootboom et al. [73] under open access license.



**FIGURE 6 |** (A) Result of the NINO3.4 prediction from May 2017 as in Nootboom et al. [73]. The dashed blue line is the running 12-months lead-time prediction and in black the (later added) observed index. Red is the CFSv2 ensemble prediction mean and the shaded area is the spread of the ensemble. The hybrid model prediction in blue is given by predictions from hybrid models found to be most optimal at the different lead times, always with ARIMA(12,1,0) and starting on May 2017. (B) The most recent prediction of the hybrid model starting from December 2018. The black line is the observed index, blue line the prediction starting from December 2018, and dashed red line is the running 12-months lead-time prediction. (A) is slightly adapted from Nootboom et al. [73] under open access license.



ANNs in Figure 8 of Nooteboom et al. [73] does not encompass most of the observed NINO3.4 index values. Hence, although the 84 ANN models had different architectures and initial weights (but were trained on the same training data set), the trained models predicted nearly the same NINO3.4 index values. Bootstrap-aggregating (bagging) methods [77] can be used to obtain a larger and more realistic ensemble spread. Another approach to better estimate the predictive uncertainty for neural network models is the so-called Bayesian neural networks (BNN). Here, all weights of the network have a distribution that can be learned by Bayesian inference. A first application to ENSO prediction in combination with a recurrent neural network (RNN) architecture is shown in McDermott and Wikle [78]. Unfortunately, the authors just present results for a short time period between 2015 and 2016. A comprehensive analysis of the application of BNNs for ENSO prediction is still lacking.

The endeavor of training a BNN is a far from trivial task. A simpler approach to estimate uncertainties in the prediction of ENSO is the application of the so-called Deep Ensembles (DEs) as presented in Lakshminarayanan et al. [79]. These DEs consist of multiple feed-forward neural network models that have two output neurons to predict the mean and the standard deviation of a Gaussian distribution. Instead of choosing the weights that minimize the mean-squared error, the models are trained by minimizing the negative log-likelihood of a Gaussian distribution with the predicted mean  $\hat{\mu}$  and variance  $\hat{\sigma}^2$ , given the observation  $y$ , i.e.,

$$-\log P(y|\hat{\mu}, \hat{\sigma}^2) = \frac{1}{2} \log \hat{\sigma}^2 + \frac{(y - \hat{\mu})^2}{2\hat{\sigma}^2} + \text{constant}, \quad (3)$$

The final prediction for the variable and its uncertainty is obtained by combining the Gaussian distributions from all members of the ensemble. In plain words, if the model does not find strong relations between predictor variables and the predicted variable in the training data, it is still able to optimize the negative log-likelihood to some extent by increasing  $\hat{\sigma}$ . Therefore, it is less prone to be overconfident about any weak relationship in the data.

Here, we give an example [80] of the application of this method for the prediction of the NINO3.4 index. For this, a DE was trained to predict the future values of the 3-month running mean NINO3.4 index. To keep the example simple, the NINO3.4 index, WWV and SC were used as input variables, where for each variable the past 12 months were included in the feature set. Hence, each ANN had 36 inputs. Each ensemble member had one hidden layer with 16 neurons with a Rectified Linear Unit as activation function. The output neuron for the mean was equipped with a linear activation function and the output neuron for the standard deviation with a softplus function ( $f(x) = \log(1 + e^x)$ ). To avoid overfitting, various regularization techniques were applied (early stopping, Gaussian noise to the inputs, dropout and  $L_1$  as well as  $L_2$  penalty terms). The training/validation period was set to be 1981–2002 and the test period 2002–2018. The training/validation data was further divided into 5 segments. One ensemble member was trained on 4 segments and validated on the remaining one to check for

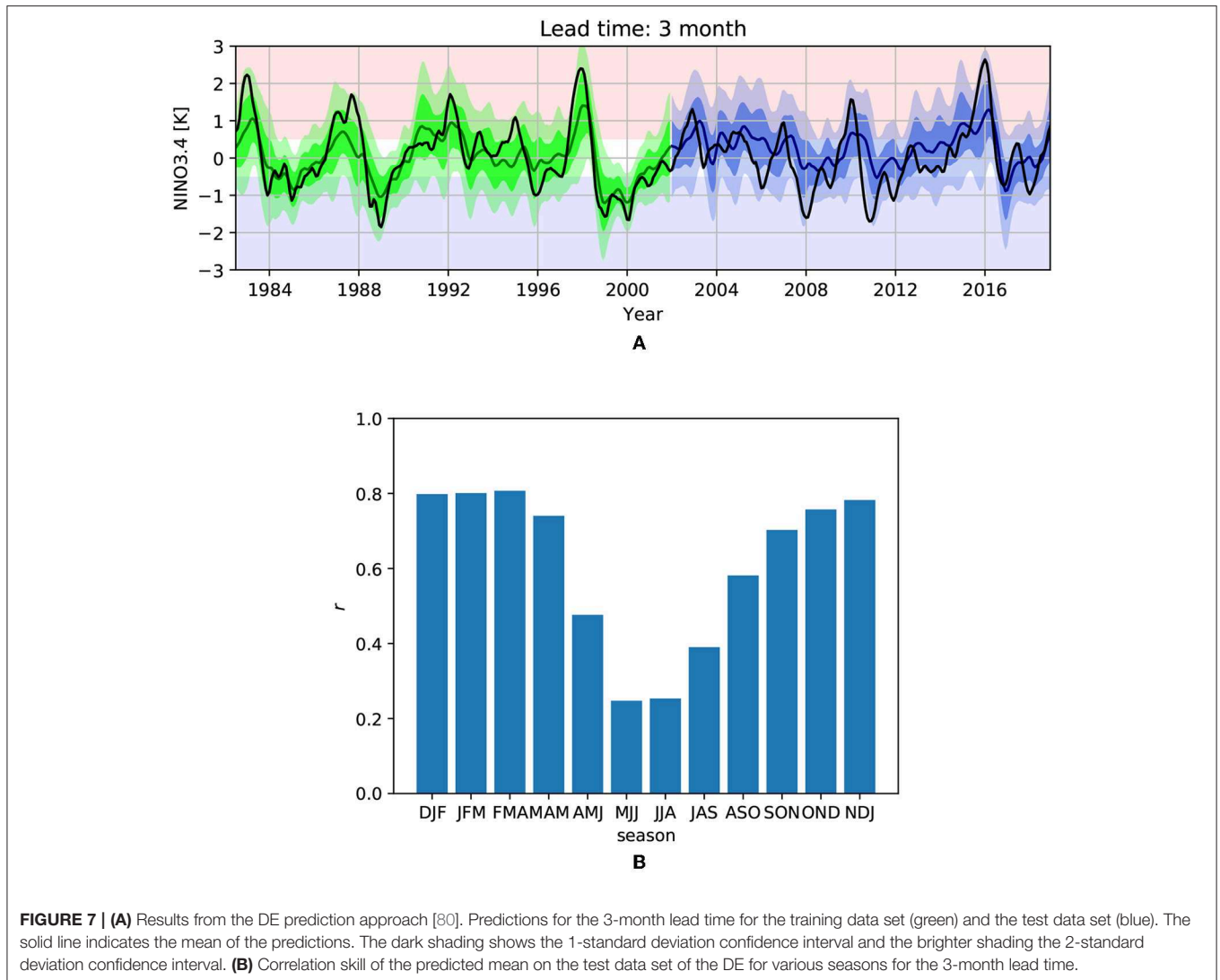
overfitting. This was repeated until each segment was one time the validation data set. Therefore, the DE had in total 5 ensemble members. Here, lead time was defined as in Barnston et al. [21] being the time that passed between the last date of the initial period and the first date of the target period.

Exemplary results for a 3-month lead-time prediction are shown in Figure 7A. In contrast to Nooteboom et al. [73], the confidence intervals of the prediction using the test data (blue line and shadings) could incorporate actual observation (black line) to a good extend. In fact, 55% were incorporated in the 1-standard deviation and 91% were inside the 2-standard deviation interval for the predictions on the test data. This indicated that such a prediction model could estimate the predictive uncertainty to a good extend. Interestingly, the predicted uncertainty had a seasonal cycle with lower uncertainties during boreal summer and higher uncertainties during boreal winter. This fitted the observations of the NINO3.4 values that follow the (same) seasonal cycle. The correlation skill on the test data set between the predicted mean and the observed NINO3.4 index is shown in Figure 7B. The relatively low skill values during the seasons AMJ to JAS indicates the spring predictability barrier. The overall correlation skill of the predicted mean on the test data set was 0.65 and the overall root-mean-square error 0.68.

## 4. DISCUSSION AND OUTLOOK

Machine Learning techniques are potentially useful to improve the skill of El Niño predictions. The choice of attributes is crucial for the degree of improvement. We have highlighted here the use of network science based attributes and the benefits of using physical knowledge to select them. Network variables provide global information on the building of correlations which occur when approaching an El Niño event, and knowledge of the physical mechanisms behind ENSO helps in determining which variables store relevant memory of the dynamics, and help to overcome the spring predictability barrier. Several network variables resulted in a clear success when applied to the ZC model [73], but not necessarily when predicting the real climatic phenomena. Work on the systematic identification of good attributes needs to be continued.

Most of the ANN studies to predict El Niño used simple architectures with a single hidden layer. Recently deeper architectures have been successfully tested [72, 73]. Nevertheless, a very complex ANN architecture will face the problem of overfitting, since the available time series are not very long and the number of parameters to optimize grows rapidly with ANN complexity. Probably, what makes El Niño prediction so challenging is that every event looks somehow different [17], and we still lack enough data to systematize these differences. Most of the methods aimed for a prediction model being most optimal in terms of least squares minimization. However, it could be interesting to put larger weight at predicting the extreme events in the optimization scheme. For example, the 6-months lead predictions of Nooteboom et al. [73] hybrid model missed the 2010 El Niño event (cf. Figure 5). Apart from this, it is important to investigate the exact reason why the



hybrid model [73] provides such a good skill for a one-year lead time.

Despite the positive findings in applying ANNs for the ENSO prediction in work of the British Columbia group [58] or of Nootboom et al. [73], the application of neural networks to ENSO prediction is still surrounded by inconsistent, non-transparent and unfavorable practices. Whereas, Tangang et al. [54] defined lead time as in Barnston et al. [21], i.e., as the time between the latest observed date and the first date of the target period, Wu et al. [58] defined lead time as the time from the center of the period of the latest predictors to the center of the target period. We suggest to use the definition of lead time as given in Barnston et al. [21], as also applied in Feng et al. [72] or Nootboom et al. [73], in future research.

Furthermore, the problem of ENSO prediction is limited by a very low amount of data. Since 1980 there have been just 3–4 major El Niño (and a similar number of major La Niña) events. This little amount of data makes neural networks extremely susceptible to overfitting. To avoid this, it is necessary

to regularize neural networks using methods such as Gaussian Noise layers, Dropout, Early Stopping,  $L_1$  or  $L_2$  penalty terms. Another problem that can arise due to the low amount of data is, that accidentally a signal in a variable exists in the training and the test data set, making the researcher confident that the model is a good generalization of the system. However, as the failure of the UBC-NNET model in the [21] study indicates, one has to be careful and not to put too much trust into the neural network predictions on ENSO considering the low amount of data. We advice not to use any variables as input to the neural network that do not have a justified reason to be a predictor for ENSO (e.g., the 9th leading EOF of the SSTA used in Wu et al. [58]).

The low amount of data, e.g., just three major El Niño events occurred since 1980, can make an educated choice of predictors very beneficial for the forecast model. This is because the ML-model cannot distinguish between relevant (deterministic concurrence) and non-relevant (random concurrence) information in a relatively large predictor

data set when the amount of training data is low. In general, if rather vague variables are used, there should be a method such as the  $L_1$ -penalty term, also called Lasso (least absolute shrinkage and selection operator), that is able to perform a feature selection and regularization [81].

Finally, past studies often did not provide the codes that they used for their results. This makes it increasingly difficult for the reader to build upon previous work and check the work for mistakes. Nowadays online platforms exist that make it easily possible to share code in a public repository and we advise that this should be the standard for any future research in the ML-ENSO prediction. To develop the idea of public available codes mentioned above one step further, we want to motivate that it would be very beneficial for this community to work together on a public repository that provides a framework for new investigations. All definitions, i.e., the lead time, as well as the used data sources with the applied preprocessing should be incorporated in this framework. Such a framework would lead to more transparency, prevent inconsistency between different research efforts as well as foster collaboration. A starting point for this endeavor could be the repository ClimateLearn published for the study of Feng et al. [72] on GitHub (<https://github.com/Ambrosys/climatelearn>).

## REFERENCES

- Bishop CM. *Pattern Recognition and Machine Learning*. New York, NY: Springer (2006).
- Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J*. (2015) **13**:8–17. doi: 10.1016/j.csbj.2014.11.005
- Heaton JB, Polson NG, Witte JH. Deep learning for finance: deep portfolios. *Appl Stochast Models Business Indust.* (2017) **33**:3–12. doi: 10.1002/asmb.2209
- McCoy JT, Auret L. Machine learning applications in minerals processing: a review. *Miner Eng.* (2019) **132**:95–109. doi: 10.1016/j.mineng.2018.12.004
- Blackwell WJ, Chen FW. *Neural Networks in Atmospheric Remote Sensing*. Boston, MA: Artech House (2009).
- Haupt SE, Pasini A, Marzban C. *Artificial Intelligence Methods in the Environmental Sciences*. Berlin: Springer Verlag (2009).
- Hsieh WW. *Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels*. Cambridge: Cambridge University Press (2009).
- Schneider T, Lan S, Stuart A, Teixeira J. Earth system modeling 2.0: a blueprint for models that learn from observations and targeted high-resolution simulations. *Geophys Res Lett.* (2017) **44**:12,396–417. doi: 10.1002/2017GL076101
- Dueben PD, Bauer P. Challenges and design choices for global weather and climate models based on machine learning. *Geosci Model Dev.* (2018) **11**:3999–4009. doi: 10.5194/gmd-11-3999-2018
- Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N, et al. Deep learning and process understanding for data-driven Earth system science. *Nature.* (2019) **566**:195–204. doi: 10.1038/s41586-019-0912-1
- O’Gorman PA, Dwyer JG. Using machine learning to parameterize moist convection: potential for modeling of climate, climate change, and extreme events. *J Adv Model Earth Syst.* (2018) **10**:2548–63. doi: 10.1029/2018MS001351
- Anderson GJ, Lucas DD. Machine learning predictions of a multiresolution climate model ensemble. *Geophys Res Lett.* (2018) **45**:4273–80. doi: 10.1029/2018GL077049
- Scher S. Toward data-driven weather and climate forecasting: approximating a simple general circulation model with deep learning. *Geophys Res Lett.* (2018) **45**:12,616–22. doi: 10.1029/2018GL080704

## AUTHOR CONTRIBUTIONS

The review was lead by HD. All authors contributed to writing of the paper.

## FUNDING

The paper originated from a visit of HD to IFISC in January 2019 and was funded by the University of the Balearic Islands. EH-G and CL were supported by the Spanish Research Agency, through grant MDM-2017-0711 from the Maria de Maeztu Program for Units of Excellence in R&D. HD also acknowledges support from the Netherlands Earth System Science Centre (NESSC), financially supported by the Ministry of Education, Culture and Science (OCW), grant no. 024.002.001.

## ACKNOWLEDGMENTS

We thank Peter Nootboom and Qingyi Feng (IMAU, Utrecht University, The Netherlands) for their excellent work on ML-based ENSO prediction and for helping with several figures for this paper.

- Preisendorfer RW. *Principal Component Analysis in Meteorology and Oceanography*. Amsterdam: Elsevier (1988).
- Rayner N, Parker D, Horton E, Folland C, Alexander L, Rowell D, et al. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J Geophys Res.* (2003) **108**:4407. doi: 10.1029/2002JD002670
- Adams RM, Chen CC, McCarl BA, Weiher RF. The economic consequences of ENSO events for agriculture. *Clim Res.* (1999) **13**:165–72.
- McPhaden MJ, Timmermann A, Widlansky MJ, Balmaseda MA, Stockdale TN. The curious case of the EL Niño that never happened: a perspective from 40 years of progress in climate research and forecasting. *Bull Amer Meteor Soc.* (2015) **96**:1647–65. doi: 10.1175/BAMS-D-14-00089.1
- Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, et al. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quart J Roy Meteorol Soc.* (2011) **137**:553–97. doi: 10.1002/qj.828
- Latif M. Dynamics of interdecadal variability in coupled ocean-atmosphere models. *J Climate.* (1998) **11**:602–24.
- Chen D, Cane MA. El Niño prediction and predictability. *J Comput Phys.* (2008) **227**:3625–40. doi: 10.1016/j.jcp.2007.05.014
- Barnston AG, Tippett MK, L’Heureux ML, Li S, DeWitt DG. Skill of real-time seasonal ENSO model predictions during 2002–11: is our capability increasing? *Bull Amer Meteor Soc.* (2012) **93**:631–51. doi: 10.1175/BAMS-D-11-00111.1
- Timmermann A, An SI, Kug JS, Jin FF, Cai W, Capotondi A, et al. El Niño–southern oscillation complexity. *Nature.* (2018) **559**:535–45. doi: 10.1038/s41586-018-0252-6
- Tang Y, Zhang RH, Liu T, Duan W, Yang D, Zheng F, et al. Progress in ENSO prediction and predictability study. *Natl Sci Rev.* (2018) **5**:826–39. doi: 10.1093/nsr/nwy105
- Saha S, Moorthi S, Wu X, Wang J, Nadiga S, Tripp P, et al. The NCEP climate forecast system version 2. *J Clim.* (2014) **27**:2185–208. doi: 10.1175/JCLI-D-12-00823.1
- L’Heureux ML, Takahashi K, Watkins AB, Barnston AG, Becker EJ, Di Liberto TE, et al. Observing and predicting the 2015/16 El Niño. *Bull Amer Meteor Soc.* (2017) **98**:1363–82. doi: 10.1175/BAMS-D-16-0009.1

26. Huang B, Thorne PW, Banzon VF, Boyer T, Chepurin G, Lawrimore JH, et al. Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): upgrades, validations, and intercomparisons. *J Clim.* (2017) **30**:8179–205. doi: 10.1175/JCLI-D-16-0836.1
27. McPhaden MJ. Tropical pacific ocean heat content variations and ENSO persistence barriers. *Geophys Res Lett.* (2003) **30**:2705–9. doi: 10.1029/2003GL016872
28. Federov A, Harper S, Philander S, Winter B, Wittenberg A. How predictable is El Niño? *Bull Amer Meteor Soc.* (2003) **84**:911–9. doi: 10.1175/BAMS-84-7-911
29. Tziperman E, Stone L, Cane MA, Jarosh H. El Niño chaos: overlapping of resonances between the seasonal cycle and the Pacific ocean-atmosphere oscillator. *Science.* (1994) **264**:72–74.
30. Jin FF, Neelin JD, Ghil M. El Niño on the devil's staircase: annual subharmonic steps to chaos. *Science.* (1994) **264**:70–2.
31. Lian T, Chen D, Tang Y, Wu Q. Effects of westerly wind bursts on El Niño: a new perspective. *Geophys Res Lett.* (2014) **41**:3522–7. doi: 10.1002/2014GL059989
32. Roulston M, Neelin JD. The response of an ENSO model to climate noise, weather noise and intraseasonal forcing. *Geophys Res Lett.* (2000) **27**:3723–6. doi: 10.1029/2000GL011941
33. Eisenman I, Yu L, Tziperman E. Westerly wind bursts: ENSO's tail rather than the dog? *J Climate.* (2005) **18**:5224–38. doi: 10.1175/JCLI3588.1
34. Webster PJ. The annual cycle and the predictability of the tropical coupled ocean-atmosphere system. *Meteor Atmos Phys.* (1995) **56**:33–55.
35. Latif M, Barnett TP, Cane MA, Flügel M, Graham NE, von Storch H, et al. A review of ENSO prediction studies. *Clim Dynam.* (1994) **9**:167–79.
36. Zebiak SE, Cane MA. A model El Niño-Southern oscillation. *Mon Wea Rev.* (1987) **115**:2262–78.
37. Mu M, Duan W, Wang B. Season-dependent dynamics of nonlinear optimal error growth and ENSO predictability in a theoretical model. *J Geophys Res.* (2007) **112**:D10113. doi: 10.1029/2005JD006981
38. Duan W, Liu X, Zhu K, Mu M. Exploring the initial errors that cause a significant “spring predictability barrier” for El Niño events. *J Geophys Res.* (2009) **114**:C04022. doi: 10.1029/2008JC004925
39. Yu Y, Mu M, Duan W. Does model parameter error cause a significant “spring predictability barrier” for El Niño events in the Zebiak-Cane model? *J Climate.* (2012) **25**:1263–77. doi: 10.1175/2011JCLI4022.1
40. Horii T, Ueki I, Hanawa K. Breakdown of ENSO predictors in the 2000s: decadal changes of recharge/discharge-SST phase relation and atmospheric intraseasonal forcing. *Geophys Res Lett.* (2012) **39**:L10707. doi: 10.1029/2012GL051740
41. Chen D, Lian T, Fu C, Cane MA, Tang Y, Murtugudde R, et al. Strong influence of westerly wind bursts on El Niño diversity. *Nat Geosci.* (2015) **8**:339. doi: 10.1038/ngeo2399
42. Timmermann A, Jin FF, Abshagen J. A nonlinear theory of El Niño bursting. *J Atmospheric Sci.* (2003) **60**:165–76. doi: 10.1175/1520-0469(2003)060<0152:ANTFEN>2.0.CO;2
43. Guckenheimer J, Timmermann A, Dijkstra H, Roberts A. (Un)predictability of strong El Niño events. *Dynam Statist Climate Syst.* (2017) **2**:2399–412. doi: 10.1093/climsys/dzx004
44. Russell SJ, Norvig P. *Artificial Intelligence: A Modern Approach, 2nd Edn.* Upper Saddle River, NJ: Pearson Education (2003).
45. Affenzeller M, Wagner S, Winkler S, Beham A. *Genetic Algorithms and Genetic Programming: Modern Concepts and Practical Applications.* Boca Raton, FL: Chapman and Hall/CRC (2009).
46. Lukoševičius M, Jaeger H. Reservoir computing approaches to recurrent neural network training. *Comput Sci Rev.* (2009) **3**:127–49. doi: 10.1016/j.cosrev.2009.03.005
47. Álvarez A, Vélaz P, Orfila A, Vizoso G, Tintoré J. Evolutionary computation for climate and ocean forecasting: “El Niño forecasting.” In: Fiemming NC, Vallergera S, Pinardi N, Behrens HWA, Manzella G, Prandle D, et al., editors. *Operational Oceanography: Implementation at the European and Regional Scales, Vol. 66 of Elsevier Oceanography Series.* Amsterdam: Elsevier (2002). p. 489–94. Available online at: <http://www.sciencedirect.com/science/article/pii/S0422989402800551>
48. De Falco I, Della Cioppa A, Tarantino E. A genetic programming system for time series prediction and its application to El Niño forecast. In: Hoffmann F, Köppen M, Klawonn F, Roy R, editors. *Soft Computing: Methodologies and Applications.* Berlin; Heidelberg: Springer (2005). p. 151–62.
49. Lima AR, Cannon AJ, Hsieh WW. Nonlinear regression in environmental sciences using extreme learning machines: a comparative evaluation. *Environ Model Softw.* (2015) **73**:175–88. doi: 10.1016/j.envsoft.2015.08.002
50. Cybenko G. Approximation by superpositions of a sigmoidal function. *Math Control Signals Syst.* (1989) **2**:303–14.
51. Tangang FT, Hsieh WW, Tang B. Forecasting the equatorial Pacific sea surface temperatures by neural network models. *Climate Dynam.* (1997) **13**:135–47.
52. Tangang FT, Hsieh WW, Tang B. Forecasting regional sea surface temperatures in the tropical Pacific by neural network models, with wind stress and sea level pressure as predictors. *J Geophys Res Oceans.* (1998) **103**:7511–22.
53. Maas O, Boulanger JP, Thiria S. Use of neural networks for predictions using time series: Illustration with the El Niño Southern oscillation phenomenon. *Neurocomputing.* (2000) **30**:53–8. doi: 10.1016/S0925-2312(99)00142-3
54. Tangang FT, Tang B, Monahan AH, Hsieh WW. Forecasting ENSO events: a Neural Network-Extended EOF approach. *J Climate.* (1998) **11**:29–41.
55. Meinen CS, McPhaden MJ. Observations of warm water volume changes in the equatorial pacific and their relationship to El Niño and La Niña. *J Climate.* (2000) **13**:3551–9. doi: 10.1175/1520-0442(2000)013<3551:OOWWVC>2.0.CO;2
56. Yuval. Neural network training for prediction of climatological time series, regularized by minimization of the generalized cross-validation function. *Month Weather Rev.* (2000) **128**:1456–73. doi: 10.1175/1520-0493(2000)128<1456:NNTFPO>2.0.CO;2
57. Yuval. Enhancement and error estimation of neural network prediction of Niño-3.4 SST anomalies. *J Climate.* (2001) **14**:2150–63. doi: 10.1175/1520-0442(2001)014<2150:EAEON>2.0.CO;2
58. Wu A, Hsieh WW, Tang B. Neural network forecasts of the tropical Pacific sea surface temperatures. *Neural Netw.* (2006) **19**:145–54. doi: 10.1016/j.neunet.2006.01.004
59. Baawain MS, Nour MH, El-Din AG, El-Din MG. El Niño southern-oscillation prediction using southern oscillation index and Niño3 as onset indicators: application of artificial neural networks. *J Environ Eng Sci.* (2005) **4**:113–21. doi: 10.1139/s04-047
60. Tang Y. Hybrid coupled models of the tropical Pacific. I: interannual variability. *Clim Dyn.* (2002) **19**:331–42. doi: 10.1007/s00382-002-0230-3
61. Tang Y, Hsieh WW. Hybrid coupled models of the tropical Pacific – II ENSO prediction. *Clim Dynam.* (2002) **19**:343–53. doi: 10.1007/s00382-002-0231-2
62. Tsonis AA, Swanson KL, Roebber PJ. What do networks have to do with climate? *Bull Am Meteorol Soc.* (2006) **87**:585–95. doi: 10.1175/BAMS-87-5-585
63. Donges JF, Zou Y, Marwan N, Kurths J. Complex networks in climate dynamics. *Eur Phys J Spec Top.* (2009) **174**:157–79.
64. Gozolchiani A, Havlin S, Yamasaki K. Emergence of El Niño as an autonomous component in the climate network. *Phys Rev Lett.* (2011) **107**:148501. doi: 10.1103/PhysRevLett.107.148501
65. Dijkstra HA, Hernández-García E, Masoller C, Barreiro M. *Networks in Climate.* Cambridge: Cambridge University Press (2019).
66. Ludescher J, Gozolchiani A, Bogachev MI, Bunde A, Havlin S, Schellnhuber HJ. Improved El Niño forecasting by cooperativity detection. *Proc Natl Acad Sci USA.* (2013) **110**:11742–5. doi: 10.1073/pnas.1309353110
67. Ludescher J, Gozolchiani A, Bogachev MI, Bunde A, Havlin S, Schellnhuber HJ. Very early warning of next El Niño. *Proc Natl Acad Sci USA.* (2014) **111**:2064–6. doi: 10.1073/pnas.1323058111
68. Stauffer D, Aharony A. *Introduction to Percolation Theory, 2nd Edn.* Philadelphia: Taylor and Francis Inc. (1994).
69. Rodríguez-Méndez V, Eguíluz M VM, Hernández-García E, Ramasco JJ. Percolation-based precursors of transitions in extended systems. *Sci Rep.* (2016) **6**:29552. doi: 10.1038/srep29552
70. Meng J, Fan J, Ashkenazy Y, Havlin S. Percolation framework to describe El Niño conditions. *Chaos.* (2016) **27**:1–15. doi: 10.1038/srep30993



71. Muscoloni A, Thomas JM, Ciucci S, Bianconi G, Cannistraci CV. Machine learning meets complex networks via coalescent embedding in the hyperbolic space. *Nat Commun.* (2017) **8**:1615. doi: 10.1038/s41467-017-01825-5
72. Feng QY, Vasile R, Segond M, Gozolchiani A, Wang Y, Abel M, et al. ClimateLearn: a machine-learning approach for climate prediction using network measures. *Geosci Model Dev Discuss.* (2016). doi: 10.5194/gmd-2015-273
73. Nootboom PD, Feng QY, López C, Hernández-García E, Dijkstra HA. Using network theory and machine learning to predict El Niño. *Earth Syst Dynam.* (2018) **9**:969–83. doi: 10.5194/esd-9-969-2018
74. Shumway RH, Stoffer DS. *Time Series Analysis and Its Applications. 4th Edn.* New York, NY: Springer (2017).
75. Hibon M, Evgeniou T. To combine or not to combine: selecting among forecasts and their combinations. *Int J Forecast.* (2005) **21**:15–24.
76. Jin FF. An equatorial ocean recharge paradigm for ENSO. Part II: a stripped-down coupled model. *J Atmos Sci.* (1997) **54**:830–47.
77. Kotsiantis SB. Bagging and boosting variants for handling classifications problems: a survey. *Knowledge Eng Rev.* (2014) **29**:78–100. doi: 10.1017/S0269888913000313
78. McDermott PL, Wikle CK. Bayesian recurrent neural network models for forecasting and quantifying uncertainty in spatial-temporal data. *Entropy.* (2019) **21**:184. doi: 10.3390/e21020184
79. Lakshminarayanan B, Pritzel A, Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. *Advances in Neural Information Processing Systems 30.* Curran Associates, Inc. Long Beach, CA (2017). p. 6402–13. Available online at: <http://papers.nips.cc/paper/7219-simple-and-scalable-predictive-uncertainty-estimation-using-deep-ensembles.pdf>.
80. Petersik P. *Machine Learning in El Niño Prediction.* MSc thesis. Utrecht University (2019).
81. Tibshirani R. Regression shrinkage and selection via the lasso. *J Roy Stat Soc Ser B.* (1996) **58**:267–88.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Dijkstra, Petersik, Hernández-García and López. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.