Date: 18 July 2017

Using Network Theory and Machine Learning to predict El Niño

Peter Nooteboom[1]

[1]Institute for Marine and Atmospheric Research Utrecht (IMAU), Utrecht University, The Netherlands

Correspondence to: Peter Nooteboom (p.d.nooteboom@students.uu.nl)

Abstract. This article aims at improving the El Niño Southern Oscillation (ENSO) prediction, which is currently not reliable more than six months ahead. Topological properties of Climate Networks of both a Zebiak–Cane-type model and observations are described and a hybrid model is introduced for ENSO prediction. This hybrid model combines Autoregressive Integrated Moving Average and an Artificial Neural Network. The predictions of the hybrid model improve the CFSv2 ensemble prediction by the National Centers for Environmental Prediction (NCEP), for predictions up to six months ahead. Moreover, the addition of a network variable as input of the prediction model, results in a twelve month lead time prediction with a comparable skill to the shorter lead time predictions.

1 Introduction

Approximately every four years, the sea surface temperature (SST) is higher than average in the eastern equatorial Pacific (Dijkstra, 2006; Dijkstra and Burges, 2002). This phenomenon is called an El Niño and is likely to occur in boreal winter. Bjerknes (1969) found the equatorial zonal winds have an eastward anomaly during such an event. Ocean-atmosphere interactions seemed to be an important feature for the development of the SST in this area and an important part of the El Niño Southern Oscillation (ENSO) dynamics. ENSO is seen as one of the most important drivers of climate variability (Ludescher et al., 2014) and has teleconnections worldwide.

So far, both statistical and dynamical models are used to predict ENSO (Chen et al., 2004; Yeh et al., 2009; Fedorov et al., 2003). However, El Niño events are not predicted well enough up to six months ahead. Specifically, the prediction from spring is found to be difficult (Goddard et al., 2001). Some theories indicate this is due to the chaotic, yet deterministic behavior (Jin et al., 1994; Tziperman et al., 1994). Others point out the importance of atmospheric noise, acting as a high frequency forcing sustaining a damped oscillation (Moore and Kleeman, 1999).

Table 1. List of abbreviations

AI	Artificial Intelligence	ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average	CN	Climate Network
ENSO	El Niño Southern Oscillation	EOF	Empirical Orthogonal function
ML	Machine Learning	NCEP	National Centers for Environmental Prediction
NON	Network of Networks	NRMSE	Normalized Root Mean Squared Error
PC	Principal Component	RNN	Recurrent Neural Network
SC	Seasonal Cycle	SSH	Sea Surface Height
SST	Sea Surface Temperature	SVM	Support Vector Machine
TDNN	Time Delay Neural Network	VAR	Vector Autoregressive model
WWB	Westerly Wind Burst	WWV	Warm Water Volume
ZC	Zebiak Cane		

Recently, attempts have been made to improve the ENSO prediction skill beyond this spring-predictability boundary, by using machine learning (ML) (Wu et al., 2006), also combined with network techniques (Feng et al., 2016). So far, ML showed to be a promising tool, outperforming conventional methods in more disciplines of physics (Hush, 2017). Besides, it did a better job than humans in a game of Go, which is difficult for Artificial Intelligence (AI) since it requires intuition and creative thinking (Silver et al., 2016). As the amount of data in the climate sciences is increasing, ML methods are becoming more interesting to apply it to.

In this article, a hybrid model is introduced for ENSO prediction. The model combines the statistical method Autoregressive Integrated Moving Average (ARIMA) and the AI based Artificial Neural Networks (ANN). First, ARIMA is a linear model, including a part of the forecast depending on the history. ANN is applied to predict the residual that is left after the ARIMA forecast, due to the nonlinear processes (Wu et al., 2006). Briefly, ANN is a system of linked neurons that describes an optimized function from one or more input variables to the output variable(s). Generally, two different approaches can be considered to applying the ANN. The first is to use a complicated ANN structure with a lot of different input variables (or attributes). This approach is situated on the deep learning part of the ML spectrum and is believed to filter the important information from the attributes itself. The deep learning approach requires a lot of input data and is computationally intensive. Therefore, simpler ANN structures are used in this article. However, techniques will have to be applied in order to reduce the amount of input variables and select the important ones, to make the problem appropriate for the simpler ANN. This reduction and selection problem can be tackled in many ways, being crucial for the prediction. In this article, the focus will be on the attributes that

are predictive according to the known ENSO mechanisms.

Moreover, complex networks are recently found to be an efficient way to represent spatial-temporal information (Tsonis et al., 2006; Steinhaeuser et al., 2012; Fountalis et al., 2015) and can be used as an attribute reduction technique. These Climate Networks (CN) are in general constructed by linking spatial-temporal points that are significantly associated with each other according to some measure. It has been demonstrated that relationships exist between topological properties of CNs and nontrivial statistical properties of the underlying dynamical system (Donner et al., 2010), also for ENSO specifically (Gozolchiani et al., 2011, 2008). Besides, CNs already appeared to be a useful tool for ENSO prediction, by considering a warning of El Niño if a network variable exceeds some critical value (Ludescher et al., 2014; Meng et al., 2016).
In this article however, multiple CNs of the pacific area are constructed and implemented in the hybrid model. This prevents the prediction method from being sensitive to some arbitrary chosen critical value. Besides, with information from additional variables, a machine could recognize whether a strong or weak signal from a network variable should be taken into account, reducing the amount of false positives and false negatives. The network variables are chosen as attributes, such that a conventional physical quantity is related, but spatial information is conserved. Moreover, the current literature with respect to El Niño dynamics and network theory will be supplemented with a network of networks (NON) approach and the dynamics of the interacting networks (Donges et al., 2011) are investigated. Because of the importance of ocean-atmosphere coupling in the ENSO dynamics, such a variable can contain interesting properties of the system.

Section 2 describes the ZC model and the methods considering both the CNs and ML. In section 3, the network methods are first applied to the ZC model. Second, the resulting attributes from observations are presented. These attributes, among which a network variable, are applied in the hybrid prediction model in section 4, which discusses the prediction results. Finally, section 5 summarizes and includes a discussion.

2 The Zebiak Cane-type model and methods

This section first describes the ZC model, including its ENSO mechanisms. Then atmospheric noise, which is not considered as an 'ENSO process' itself, is incorporated into the model and its effects are discussed. Second, methods are explained considering both the construction of CNs and describing their topological properties. Finally, the methods concerning the ML are explained, including the hybrid model.

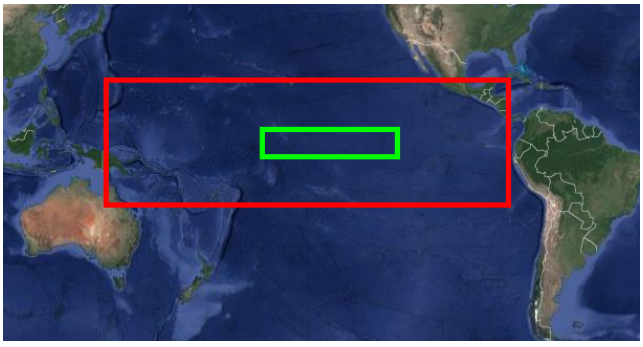


Figure 1. Pacific area (red rectangle) from 140 to 280 $^\circ$ E and -20 to 20 $^\circ$ N and the NINO3.4 area (green rectangle) from 170 to 120 $^\circ$ W and -5 to 5 $^\circ$ N

2.1 Zebiak Cane model with atmospheric noise

The ZC model represents the coupled ocean-atmosphere flow on an equatorial β-plane in the Pacific area (figure 1), with planetary vorticity gradient β_0. The basin has a zonal length L and is meridionally unbounded. The ocean component consists of a shallow water layer with mean depth H and constant density ρ. This layer contains an embedded well-mixed layer with mean depth H_1, such that $H = H_1 + H_2$. The deep ocean has density $\rho + \Delta\rho$ and is assumed to be at rest. Only long wave motions are considered in this model.

The shallow water flow is governed by (Von Der Heydt et al., 2011):

$$\frac{\partial u}{\partial t} + a_m u - \beta_0 y v + g' \frac{\partial h}{\partial x} = \frac{\tau^x}{\rho H}$$

$$\beta_0 y u + g' \frac{\partial h}{\partial y} = 0$$

$$\frac{\partial h}{\partial t} + a_m h + c_0^2 \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) = 0$$

Here u, v are respectively the zonal an meridional velocity, h the thickness of the layer, τ^x the zonal wind stress, $g' = \frac{g\Delta\rho}{\rho}$ is the reduced gravity, $c_0 = \sqrt{g'H}$ the phase speed of the first oceanic baroclinic Kelvin mode and a_m a linear damping coefficient. The boundary conditions in the East ($x = L$) and West ($x = 0$) of the basin are:

$$\int_{-\infty}^{\infty} u(0,y,t)dy = 0$$

$$u(L,y,t) = 0$$

The equations for the surface layer velocities are:

$$a_s u_s - \beta_0 y v_s = \frac{H_2}{H} \frac{\tau^x}{\rho H_1}$$

$$a_s v_s - \beta_0 y u_s = 0$$

a_s is a linear damping coefficient. The evolution of sea surface temperature T is described by:

$$\frac{\partial T}{\partial t} + \alpha_T (T - T_0) + \frac{w_1}{H_1}\Theta(w_1)(T - T_s(h)) + u_1\frac{\partial T}{\partial x} + v_1\frac{\partial T}{\partial y} = 0 \tag{1}$$

In this equation, α_T is a linear damping coefficient, $u_1 = u_s + u$, $v_1 = v_s + v$ and $w_1 = w_s + w$. T_0 is the radiation equilibrium temperature and Θ the Heaviside function to assure cooling/heating only occurs by upwelling and not by downwelling. Subsurface temperature T_s depends on the vertical temperature distribution, therefore on the thermocline depth h by:

$$T_s(h) = T_{s0} + (T_0 - T_{s0})\tanh\left(\frac{h + h_0}{\hat{H}}\right) \tag{2}$$

Here $\hat{H}$ and h_0 control the steepness and the offset of the T_s profile. T_{s0} is the characteristic temperature of the water that is being upwelled into the surface layer.

The ocean component of the model is coupled to a Gill atmosphere model with horizontal velocities (u_a, v_a), geopotential height ϕ and linear damping coefficient A_m. The atmosphere is driven by heat fluxes from the ocean, depending linearly on the anomaly of the sea surface temperature T with respect to T_0, with proportionality constant α_0. The atmosphere is governed by:

$$\begin{aligned}
\frac{\partial u_a}{\partial t} + A_m u_a - \beta_0 y v_a - \frac{\partial \phi}{\partial x} &= 0 \\
\frac{\partial v_a}{\partial t} + A_m v_a + \beta_0 y u_a - \frac{\partial \phi}{\partial y} &= 0 \\
\frac{\partial \phi}{\partial t} + A_m \phi - c_a^2\left(\frac{\partial u_a}{\partial x} + \frac{\partial v_a}{\partial y}\right) &= \alpha_T (T - T_0)
\end{aligned}$$

Here c_a is the phase speed of the atmospheric Kelvin wave.

The zonal wind stress τ^x is build up from a coupled and an external part:

$$\tau^x = \tau^x_{ext} + \tau^x_c \tag{3}$$

The external part τ^x_{ext} is independent of the coupling between the atmosphere and ocean. It represents an easterly wind stress due to the Hadley circulation. It is assumed to be zonally constant and depends on latitude according to:

$$\tau^x_{ext} = -\tau_0 e^{-\frac{1}{2}\left(\frac{y}{L_a}\right)} \tag{4}$$

Here $\tau_0 \sim 0.01\ Pa$, L_a the atmospheric Rossby deformation radius. The coupled part of the zonal wind stress is proportional to the zonal wind field: $\tau^x_c = \gamma_\tau u_a$. The meridional component of the wind stress τ^y is neglected in this model.

This γ_τ is proportional to the coupling strength, which parameter determines the coupling between ocean and atmosphere (or wind stress per temperature anomaly) and is defined by:

$$\mu = \frac{\alpha_T \Delta T L}{c_a^3} \tag{5}$$

Here ΔT is the typical horizontal temperature difference across the basin.

Without any included noise, a temperature anomaly damps out to a constant value and a stationary state if $\mu < 3$. However, if the coupling strength exceeds a critical value of $\mu_c = 3$, a supercritical Hopf bifurcation occurs. The system does not damp out, but an oscillation is sustained with a period of approximately four years.

Three positive feedbacks related to the thermocline depth, upwelling and zonal advection, cause a positive SST anomaly to grow more (Dijkstra, 2006). However, the oscillatory behavior associated with ENSO is caused by negative delayed feedbacks.

The 'classical delayed oscillator' paradigm assumes this negative feedback is caused by waves through geostrophic adjustment, controlling the thermocline depth.

A complementary, different view is the 'recharge/discharge oscillator' (Jin, 1997), also regarding oceanic waves excited through oceanic adjustment. The waves excited to preserving the Sverdrup balance, are responsible for a transport of warm surface water to higher latitudes, discharging the warm water in the tropical Pacific. The thermocline depth is raised, resulting in more cooling of SST.

The warm water volume (WWV) is generally used to capture how much the tropical Pacific is 'charged.' Specifically, this is the water volume above the $20\ ^\circ C$ isotherm between 120° to 280° East and -5° to 5° North.

Besides the ENSO processes discussed so far, the area is also suspect to other processes from the atmosphere, which are considered as noise. An important example of atmospheric noise are the so called westerly wind bursts (WWB). These are related to the Madden-Julian oscillation (Madden and Julian, 1994). The WWB is a strong westerly anomaly in the zonal wind field, occurring every forty to fifty days and lasting approximately a week.

To represent the atmospheric noise in the model, we obtained a residual of the wind stress. We used the extended reconstructed sea surface temperature (ERSST) over the Pacific for the period 1978-2004 and the Florida State University pseudo-wind-stress data for the same period. First the part of the wind stress anomalies linearly related to the SST anomalies, are subtracted from the wind stress anomalies. Then the residual is projected on its empirical orthogonal function (EOF).

The six EOF components that explain most variability were considered to describe the spatial patterns of the red noise and their principal components (PC) were used to construct the time series of red noise (Feng, 2015). These time series are modeled as an independent AR1 process by:

$$x_{t+1} = ax_t + \sigma\epsilon_t \tag{6}$$

a is the lag-1 autocorrelation of each PC and $\sigma\epsilon_t$ is the white noise with variance σ. Every timestep is one week, since weekly data are considered.

The effect of the noise on the model depends on whether the model is in the super- or sub-critical regime (i.e whether μ above or below μ_c). If $\mu < \mu_c$, the noise excites the ENSO mode, causing irregular oscillations. In the supercritical regime, the cycle of approximately four years is still present, but the noise causes a higher amplitude of ENSO variability.

2.2 Network variables

2.2.1 Climate network construction

Different methods can be used to construct CNs. Here only the Pearson correlation will be considered for the construction of two different types of CN. The Pearson correlation between variables $p_i,\ p_j$ associated with two points on the grid $i,\ j \in [0, \cdots, N]$ is defined as:

$$R(i,j) = \frac{\sum_{k=1}^{n} p_i(t_k) p_j(t_k)}{\sqrt{\left(\sum_{k=1}^{n} p_i^2(t_k)\right)\left(\sum_{k=1}^{n} p_j^2(t_k)\right)}} \tag{7}$$

Here p_i is a vector of size n of the time series. The data is detrended before the correlation is calculated.

In the first method an undirected and unweighted network is constructed. The links are stored in a symmetric adjacency matrix A, where $A_{ij} = 1$ if node i is connected to node j and zero otherwise. It is defined by:

$$A_{ij} = \Theta\left(|R_{ij}| - \epsilon\right) - \delta_{ij} \tag{8}$$

Here R_{ij} is the Pearson correlation between node i and j. ϵ is the threshold value. If the Pearson correlation exceeds this threshold, the two points will be linked. Θ denotes the Heaviside function. δ_{ij} is the Kronecker delta function, implemented to prevent connection of nodes with themselves.

The second method creates a weighted, undirected network. The cross-correlation $C_{ij}(\Delta t)$ between nodes $i,\ j$ is considered: The Pearson correlation with a lag Δt in time. Then the weights between the nodes are calculated by:

$$W_{ij} = \frac{\max_{\Delta t}(C_{ij}) - \text{mean}(C_{ij})}{\text{std}(C_{ij})} \tag{9}$$

Here $\max_{\Delta t}()$ denotes the maximum, std() the standard deviation and mean() the mean value over all Δt that are considered.

2.2.2 Topological properties

Some methods are explained in this section, capturing the information about the topology of both types of constructed networks.

The first is the local degree d_i of node i in the CN:

$$d_i = \sum_j A_{ij} \tag{10}$$

In the case of the unweighted network, the degree is equal to the amount of nodes that are connected to node i.

Second, the cross clustering contains information about the interaction between two networks. The cross clustering of a node is the probability two connected nodes in the other network are also connected to each other. The local cross clustering c_v of node v is defined as:

$$c_v = \frac{1}{k_v\,(k_v - 1)} \sum_{p \neq q} A_{vp} A_{pq} A_{qv} \tag{11}$$

Here p, q are the nodes in the other network and k_v denotes the cross degree of node v. That is the amount of cross links node v has with the other network. The global clustering coefficient is the average of the local clustering coefficient over all nodes of a network.

Third, percolation theory is considered, describing the connectivity of different clusters in a network. It has been found before, the connectivity of some CN increases when approaching an El Niño and decreases afterwards (Rodríguez-Méndez et al., 2016), as local correlations between points increase and decrease. At such a percolation-like transition, the addition of only a few links can cause a considerable part of the network to become connected.

Before the percolation transition, small cluster sizes will form. Therefore the variable c_s will warn for the transition. This is the fraction of nodes that are part of a cluster of (generally small) size s, normalized by the amount of nodes in the network.

Moreover, Δ can be used to capture the percolation-like transition (Meng et al., 2016). Links are added to a network one by one, adding the link with the largest weight first (equation 9). At every step T that a link is added, the size of the largest cluster $S_1(T)$ is calculated. At the point of the transition, $S_1(T)$ increases rapidly. The size of this jump is Δ:

$$\Delta = \max\left[S_1(2) - S_1(1), \ldots, S_1(T+1) - S_1(T), \ldots\right] \tag{12}$$

Finally, the algebraic connectivity (λ_2) is considered, determining the diffusivity of information within a network. Suppose any node i is represented by a value ψ_i in an undirected network. In our case ψ_i is directly related to the time series. How much ψ_i changes by diffusion depends on the values of nodes j it is connected to (Newman, 2010):

$$\frac{d\psi_i}{dt} = C \sum_j A_{ij}\,(\psi_j - \psi_i) \tag{13}$$

Here C is the diffusion constant and A the adjacency matrix. By separating the sum in equation 13, this can be rewritten as:

$$\frac{d\psi_i}{dt} = C \sum_j (A_{ij} - \delta_{ij} d_i)\, \psi_j \tag{14}$$

δ_{ij} is the Kronecker delta function and d_i the degree of node i. In matrix notation this reduces to:

$$\frac{d\psi_i}{dt} = C\,(\mathbf{A} - \mathbf{D})\,\psi \tag{15}$$

Here $\mathbf{D}$ with $\mathbf{D_{ij}} = \delta_{\mathbf{ij}}\mathbf{d_i}$ is a square matrix that contains the degrees at the diagonal and zero elsewhere. Now we define the graph Laplacian of the network as:

$$\mathbf{L} = \mathbf{A} - \mathbf{D} \tag{16}$$

As you can see, equation 15 reduces to the diffusion equation, but with the graph Laplacian matrix $\mathbf{L}$ in stead of ∇^2. By calculating the eigenvalues of the Laplacian matrix $\lambda_1, \cdots, \lambda_n$ with $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$, we can determine the diffusion within the network. Since the matrix is symmetric, the eigenvalues are real. Moreover, the smallest eigenvalue λ_1 is always zero and no eigenvalues are negative. This means the network will decay to a stable solution. The second smallest eigenvalue, called the algebraic connectivity, is of particular interest. In general, $\lambda_2 > 0$ if the network is fully connected.

2.3 Machine Learning

Up to now the methods are explained to investigate CNs of both the ZC model and observations. This section first discusses some possibilities of predicting ENSO with ML and explains the eventual applied prediction scheme: The hybrid model. The ARIMA part of the prediction model is illustrated. Then the problem of attribute selection is discussed. Finally the ANN part of the hybrid model is explained, where the selected attributes will be applied to.

An ANN should be a well-suited AI method for the ENSO prediction problem, because of the nonlinearity of the system. A feed-forward ANN (figure 2) will describe a nonlinear function from the currently known input variables to the NINO3.4 index in the future. However, the skill of the prediction will improve if this function also depends on the history. In general this can be done by adapting the general feed forward ANN structure, by implementing loops within the ANN to obtain internal memory. This is called a Recurrent Neural Network (RNN).

Another method is the Time Delay Neural Network (TDNN), where lags of attributes are implemented as input.

However, both methods increase the complexity of the problem. Using RNN will increase the amount of possible ANN structures. Using TDNN will rise the question which lags of variables will have to be implemented, such that no noise, but only complementary information is added. Besides the

TDNN will increase the amount attributes, therefore the ANN structure will have to be more complicated to solve the problem.

To overcome these problems, the hybrid model (Valenzuela et al., 2008) will be applied to predict ENSO. Previous work showed the results of a hybrid model are in general more stable and reduce the risk of a bad prediction, compared to a single prediction method (Hibon and Evgeniou, 2005). The problem is split into two parts: An ARIMA part depending on history and a nonlinear part making use of feed-forward ANN. The hybrid model represents the observation Z_t at time t by:

$$Z_t = Y_t + N_t \tag{17}$$

Here Y_t is the linear part and N_t the nonlinear part. Let $\tilde{Y}_t$ be the prediction of the part Y_t using ARIMA. Then $Z_t - \tilde{Y}_t$ is the residual between the ARIMA prediction and the observed value.
In the second part the residual $Z_t - \tilde{Y}_t$ will be predicted by the feedforward ANN:

$$\tilde{N}_t = f\left(x_1(t), \cdots, x_N(t)\right) \tag{18}$$

Here f is a nonlinear function of the N attributes $x_1(t), \cdots, x_n(t)$ and $\tilde{N}_t$ the prediction of residual $Z_t - \tilde{Y}_t$ at time t. Notice the nonlinear function f does not depend on history. Therefore, a longer lead time prediction of the model could have a better prediction skill. This results in the final prediction $\tilde{Z}_t$ of the hybrid model:

$$\tilde{Z}_t = \tilde{Y}_t + \tilde{N}_t \tag{19}$$

This scheme describes a 'supervised' model, implying the predictant is 'known.' This known quantity is the NINO3.4 index. The standard procedure for supervised learning is to optimize the ML method on a 'training set' to define an optimal model, which predicts ENSO up to a prediction time τ ahead. This function will then be tested on a test set. In this article generally a training set of $80\ \%$ and an test set of $20\ \%$ of the total time series is used.

The data set can be represented by a $T \times N$ matrix, where T represents the length of the time series and each time $t = 1,\ldots,T$ has a set of N attributes $x_1(t),\ldots,x_N(t)$. Note that, since we are predicting time series, for any training set $[t_i^{train}, t_f^{train}]$ and test set $[t_i^{test}, t_f^{test}]$, must hold $t_i^{test} > t_f^{train}$ (where $t_i^{train}, t_f^{train}, t_i^{test}, t_f^{test} \in [1, T]$).

2.3.1 Autoregressive Integrated Moving Average

First, the training set is used to optimize an ARIMA(p,1,1) process for the NINO3.4 time series. The standard method maximizing the log likelihood function is used to fit $\alpha_1, \cdots, \alpha_p, \beta_1$ and define the ARIMA(p,1,1) process for some time series X_t:

$$X_t = X_{t-1} + \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \beta_1 \varepsilon_{t-1} \tag{20}$$

Here ε_t is the residual between the ARIMA prediction and observation at time t. Only $p \leq 13$ is considered. Making a prediction using the ARIMA model k steps ahead, implies equation 20 is

applied k times.

For a $k > 1$ month lead ARIMA prediction, the real residual ε_{t-1} cannot be determined, because the observed value of X_{t-1} is in the future. Therefore we will have to work with the difference between the k month lead prediction of the hybrid model and the ARIMA prediction as residual.

Besides, an ARIMA(p,1,0) model will be considered, implying that $\beta_1 = 0$ in equation 20 and no moving average part is considered.

2.3.2 Attribute selection

If the ARIMA prediction is performed, other variables than the NINO3.4 index itself are added to the attribute set, to be implemented in the ANN. Deciding which of the variables to add, is not a straightforward problem, yet crucial for the eventual prediction. Sometimes a pair of two variables can be compatible in the prediction, but perform poor when applied alone. Other pairs can be redundant and cover important information alone, but solely noise is included when used together (Guyon and Elisseeff, 2003). Adding a variable to the attribute set and see if it improves prediction, can only conclude whether it improves prediction with respect to the old attribute set, not whether the variable is predictive in itself. To determining the attribute set, it is considered which variables physically represent a certain mechanism that is important for the ENSO prediction. Besides, it is tested whether the prediction skill is reduced if a variable is dropped out of the attribute set.

Moreover, the attributes should be selected at optimal lead times. Two methods will help to decide which variables can improve the prediction, although these methods are not necessarily decisive in determining the most optimal attribute set. First, correlation between the predictor and predictant is a commonly used measure for attribute selection. Therefore the cross-correlation is calculated for the attributes to show the predictability of a time series, using again the Pearson correlation from equation 7:

$$R_\tau(p,q) = \max_\tau \left(\frac{\sum_{k=1}^{n} p(t_k) q(t_k - \tau)}{\sqrt{\left(\sum_{k=1}^{n} p^2(t_k)\right)\left(\sum_{k=1}^{n} q^2(t_k - \tau)\right)}} \right) \tag{21}$$

Here p is the predictor, q is the predictant and lag $\tau \leq 64$ weeks such that no information too far in the past is considered.

However, the effect of a variable on ENSO at a short lead time, increases the cross-correlation at a longer lead time, due to the effect of autocorrelation (Runge, 2014). To solve this autocorrelation problem, variables can be tested for Granger causality. In general, a first signal Granger causes a second signal, if the first has a statistically significant effect on the second variable, taking past values of the second signal into account as regressors. Note Granger causality is not the same as a 'true' causality. To test if a signal x_t Granger causes a second time series y_t at some lag τ, an F-test is performed to test the significance of the parameter b_τ in the sum of squares optimized linear

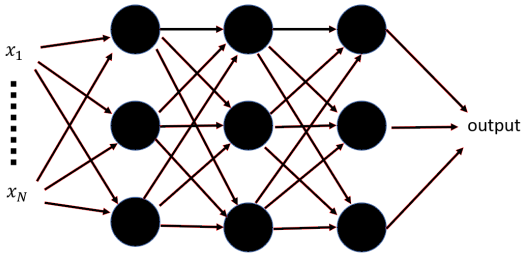


Figure 2. $3 \times 3 \times 3$ feed-forward artificial neural network. This implies it consists of three layers with each three neurons, without any internal loops.

regression equation:

$$y_t = a_0 + a_1 y_{t-1} + ... + a_n y_{t-\tau} + b_1 x_{t-1} + ... + b_\tau x_{t-\tau} + \varepsilon_t \tag{22}$$

Here ε_t is the error at time t. The p-value of this hypothesis is considered for different lags τ. A low p-value means the null hypothesis that x_t does not Granger cause y_t is rejected at a lower significance level: The variable is more likely to Granger cause ENSO.

Notice these methods give us merely an idea of which variables can be used for the prediction at different lead times. These methods are linear, while the attributes will be used in a nonlinear method.

2.3.3 Artificial Neural Network

Finally, the $T \times N$ dataset with selected attributes is used to predict the residual between the ARIMA forecast and the observations in an ANN. Besides using the NINO3.4 sequence itself, the additional attributes can be applied to add important information and improve the prediction.

Generally, a feed-forward ANN has a structure without loops, similar to figure 2. The input variables are linearized and projected to the first layer neurons according to (Bishop, 2013):

$$z_j = h\left(\sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}\right) \tag{23}$$

Here z_j is the value of the j-th neuron of the layer, $w_{ji}^{(1)}$ is the weight between input x_i from neuron i to neuron j, where the (1) denotes the first layer. $w_{j0}^{(1)}$ is referred to as the bias. h is the nonlinear activation function and is generally a sigmoidal function: $h(a) = \frac{1}{1+\exp(-a)}$. The activation function is important to incorporate the nonlinearity in the optimization scheme. Without activation function, a linear function will still exist from the input to output variables.

These z_j can again be used as input for a second layer, which can be used for a third layer etc.. Eventually this leads to some output which can be compared with the time series that must be predicted. Using a backward propagating technique, the squared error $\sum_t (y_t - \hat{y}_t)^2$ between the residual we are predicting y_t and the output of the ANN $\hat{y}_t$, will be minimized over the weights. Initially, some

random distribution of weights is used. Generally, this backward propagating technique will always decrease the squared error if the activation function h is chosen to be a sigmoidal function.

3 Analysis of network properties and other attributes attributes

In this section, the resulting topological properties of CNs are analysed. Firstly, the network variables from the ZC model are explained. Secondly, this results in a network variable of observations containing interesting properties, which will be explained along with all attributes that are applied in the hybrid model in section 4.

3.1 ZC model results

Weekly spatial-temporal data on a 31×30 grid in the pacific area are obtained for forty-five years from the ZC model, to construct the CNs. The first five years are not considered, such that the system does not depend on the initial conditions of the model.

Generally, a sliding window approach is used to calculate the network variables. This implies the network is calculated for some sequence in time, which is slided four weeks ahead every time step. For the ZC model, either the thermocline network (from h), SST network (from T), wind network (from τ^x) or a combination of these are considered for CN construction. The following network variables are discussed in this section:

- Skewness of the SST degree distribution
- Cross clustering of the zonal wind and temperature network
- Zonal skewness in the degree field of the thermocline network
- Two variables describing a percolation-like transtition
- Algebraic connectivity of the zonal wind and thermocline network

3.1.1 Skewness of the SST degree distribution

Determining how strong noise can excite the ENSO mechanisms in the sub-critical case, or determining whether the feedbacks sustain an oscillation in the supercritical state, could provide information to increase the prediction skill. Feng (2015) has found a network variable which relates to the coupling strength (μ).

First an undirected, unweighted network is constructed of the SST field, with a threshold of $\epsilon = 0.5$. The skewness of the probability distribution for some node in the network to have a certain degree (S_d), decreases monotonically over increasing coupling strength μ (see figure 3).

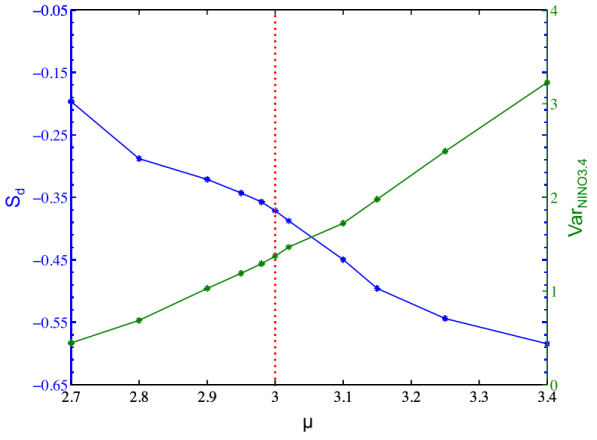


Figure 3. S_d (in blue) and the variance of the NINO3.4 index (in green) for different coupling strength μ in the ZC model from Feng (2015). The vertical dotted red line denotes the critical coupling strength μ_c

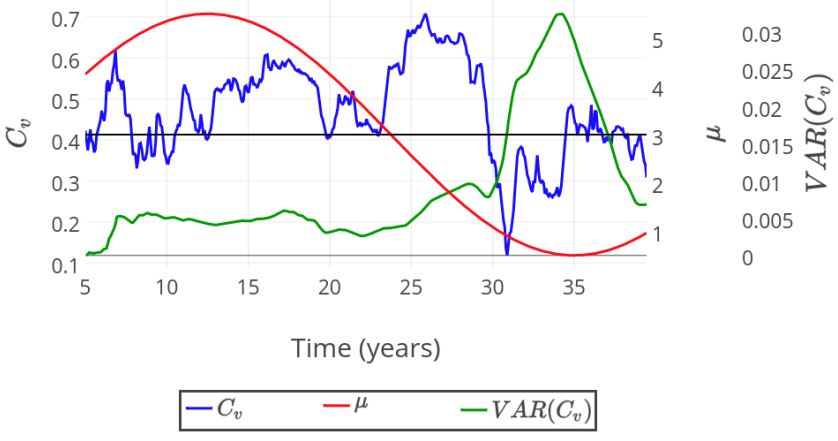


Figure 4. Global cross clustering between the SST and wind network in blue and its variance in green. The coupling strength μ defined as a sinusoid around $\mu_c = 3$ with an amplitude of 0.25 in red. The sliding window is applied with a window of five years.

3.1.2 Cross clustering of the zonal wind and temperature network

Although S_d relates to the climate stability and coupling strength, it does not inform whether the system is in either the supercritical or sub-critical state. A NON variable has a related property:
385 The global cross clustering (C_v) between the SST and wind network. In this case, the global cross clustering coefficient is a measure of the amount of triangles in the networks, containing one wind node and two SST nodes.

In figure 4 this cross clustering is calculated, when coupling strength μ changes in time. In the beginning the system is in the supercritical state, the state is sub-critical after $\mu_c = 3$ is crossed. When the critical point is passed, the noise has a larger influence on local correlations. This causes triangles to break and the variance of the cross clustering coefficient to increase. The cross clustering C_v is a diagnostic network variable, informing in which state the system is.

3.1.3 Zonal skewness in the degree field of the thermocline network

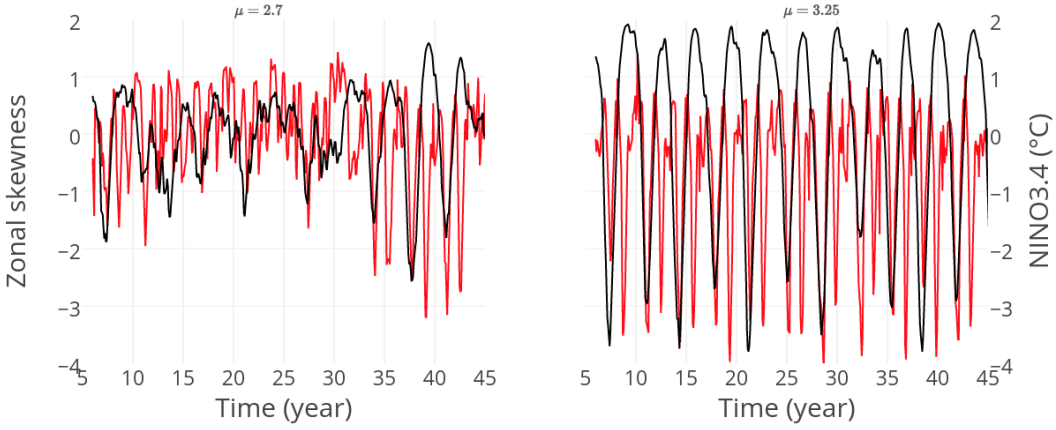


Figure 5. Zonal skewness of the degree field with $\epsilon = 0.6$ in red, NINO3.4 in black. Left the sub-critical and right the supercritical case.

From the classical view of the oscillatory behavior of ENSO, the waves should contain memory of the system, because of their negative delayed feedback. From this point of view, the changing structure of the thermocline network is of interest when predicting ENSO. Calculating the network with threshold $\epsilon = 0.6$, a zonal pattern in the change of the network close to the equator can be observed during an ENSO cycle. Generally, the degree field is quite symmetric, but when ENSO turns either from upward to downward, or from downward to upward, the degree of the nodes in the east decreases. This is at the peak El Niño or La Niña.

To capture this zonal asymmetry around the equator with a variable, the zonal skewness of the degree field will be used between $-7°$ to $7°$ North. The higher the skewness, the more the degree will be located west of the basin. If the skewness is close to zero, the degree is symmetrically distributed over the basin. If it is low, most of the degree is situated in the east. The skewness will show a negative peak when the sign of the first ENSO derivative changes (figure 5).
In the supercritical case $\mu = 3.25$ this effect is indeed observed. Nevertheless, in the sub-critical case, the pattern is only visible once ENSO shows a clear oscillation (around year 32).

3.1.4 Percolation theory providing a warning precursor

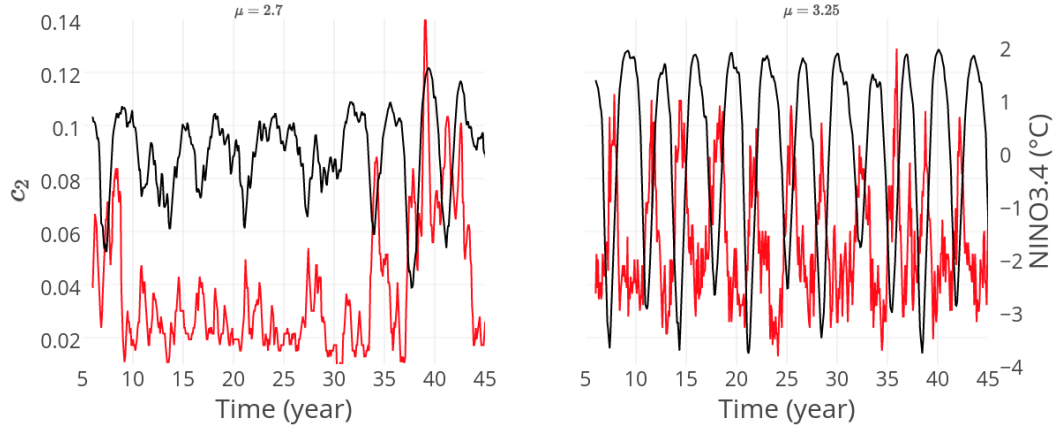


Figure 6. c_2 of the thermocline network in red. NINO3.4 in black. Left the sub-critical case with threshold $\epsilon = 0.99999$ and right the supercritical case with $\epsilon = 0.999$.

For the ZC model, c_2 (the amount of clusters of size two) of the thermocline network is found to indicate a percolation transition of the network (figure 6). c_2 increases approximately one to two years before an El Niño event. This is mainly clear in the supercritical case. In the sub-critical case, a clear warning of an event occurs when the oscillation of ENSO is more clear and the El Niños are stronger.

Δ shows the same behavior as c_2 (see figure 7).

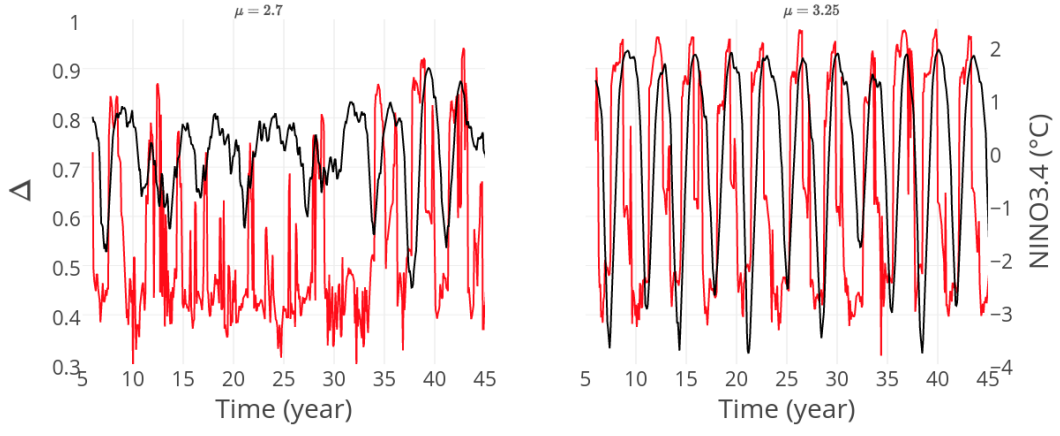


Figure 7. Δ in red and ENSO in black. Left the sub-critical and right the supercritical case.

Although Δ is not sensitive to a chosen threshold, it tends to show a warning closer to an event, compared to c_2.

3.1.5 Algebraic connectivity of the zonal wind and thermocline network

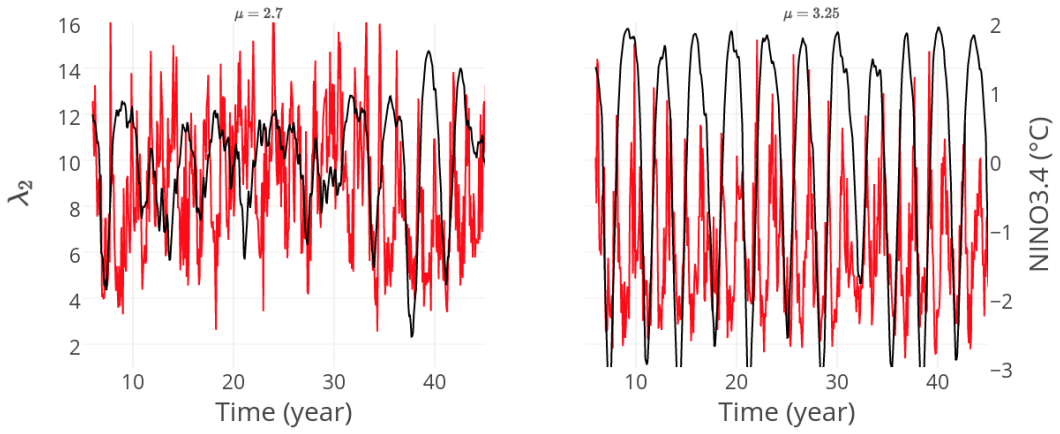


Figure 8. Algebraic connectivity λ_2 in red and ENSO in black. Left the sub-critical and right the supercritical case.

If the feedbacks between two variables in the system are stronger, the diffusion of information between the networks associated to these variables could increase (see section 2.2.2). The algebraic connectivity of a NON might capture the amount of interaction between them. λ_2 of a network constructed from the thermocline and wind network shows the diffusion in this network is larger before and after an event (figure 8). In the supercritical case ($\mu = 3.25$), clearly peaks are observed in λ_2 before and after an El Niño, implying the spread of information in the network is the highest at these moments. In the sub-critical case ($\mu = 2.7$) λ_2 is noisy. Nevertheless, if some clear oscillations are excited around the year 35, again the peaks before and after El Niño can be observed.

3.2 Attributes from observations

Up to now it is demonstrated how CNs can describe properties of the dynamical system that is associated with ENSO. The variables from the CNs could be used to predict El Niño. Although the network variables show interesting behavior when applied to the model in the previous section, this is not always the case for the observations. This section describes which variables, among which network variables, are implemented in the hybrid model and the selection of these attributes at different lead times.

To quantify ENSO, the NINO3.4 index is used: The three month running mean of the SST between $170°$ to $120°$ West and $-5°$ to $5°$ North.

First, from the recharge/discharge oscillator point of view, the WWV shows great potential for the prediction of ENSO (Bosc and Delcroix, 2008; Bunge and Clarke, 2014). Therefore it is used in the attribute set.

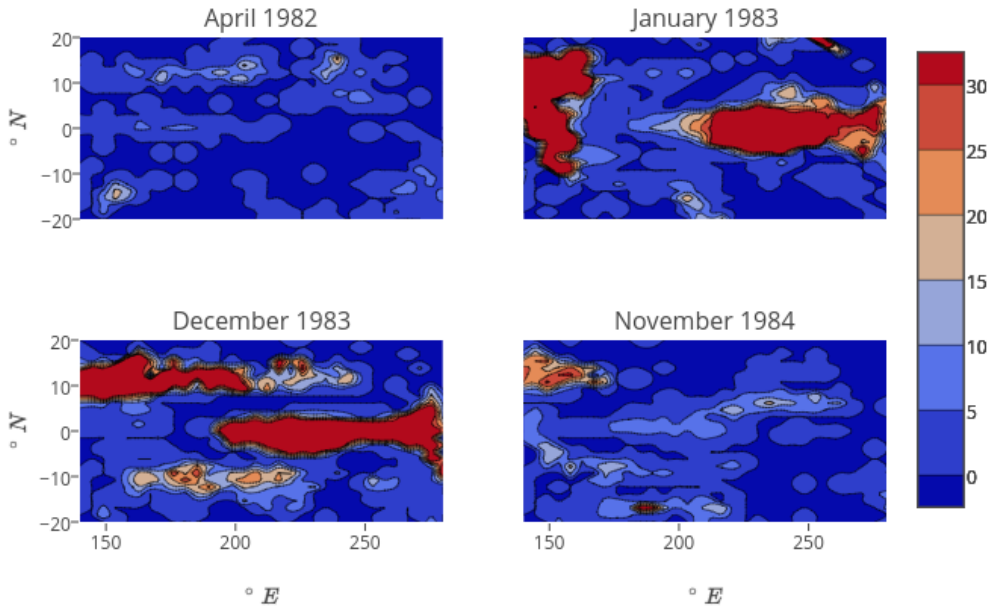


Figure 9. Evolution of the local degree of the SSH network with threshold $\epsilon = 0.9$ during the 1982 El Niño.

The second attribute considers a WWV related network variable. For the CN construction, the

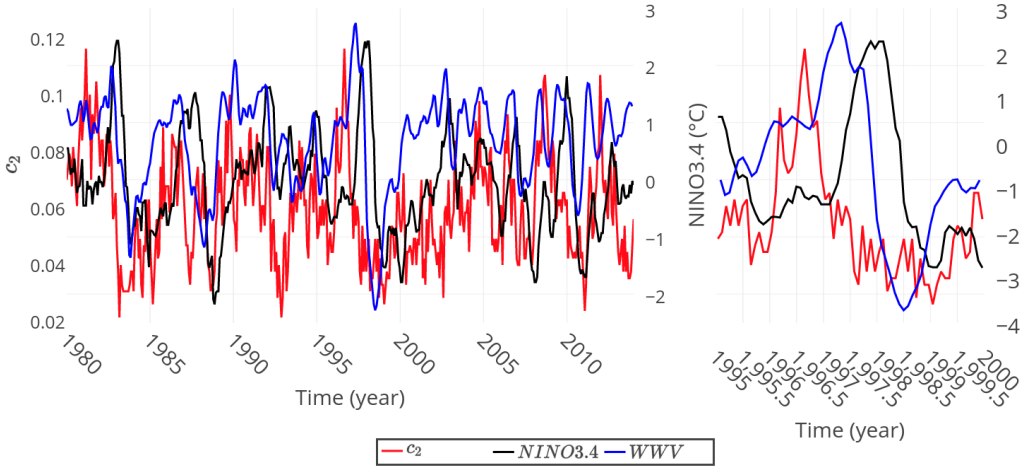


Figure 10. The WWV and c_2 of the SSH network with $\epsilon = 0.9$. On the left for the whole time series and on the right only during the 1997 El Niño where a warning is visible of the WWV and c_2. c_2 gives a warning almost a year before El Niño, while WWV warns almost seven months ahead.

sea surface height (SSH) is used from the weekly ECMWF reanalysed dataset from 1979 to 2014 between $140°$ to $280°$ East and $-20°$ to $20°$ North. The SSH is by approximation proportional to the thermocline depth (Rebert et al., 1985), thus related to propagating waves and strength of the thermocline feedback. A grid of 27 latitude points and 30 longitude points in the Pacific area is used with a threshold $\epsilon = 0.9$, such that the network is not fully connected. During an El Niño event, the link density of this network increases in the warm pool and the cold tongue specifically (figure 9), causing a percolation-like transition. As discussed in the previous section, c_2 gives an early warning of this event. This variable allows us to extend the lead time of the WWV.

Third, atmospheric noise from the WWB are a limitation for the prediction of ENSO (Moore and

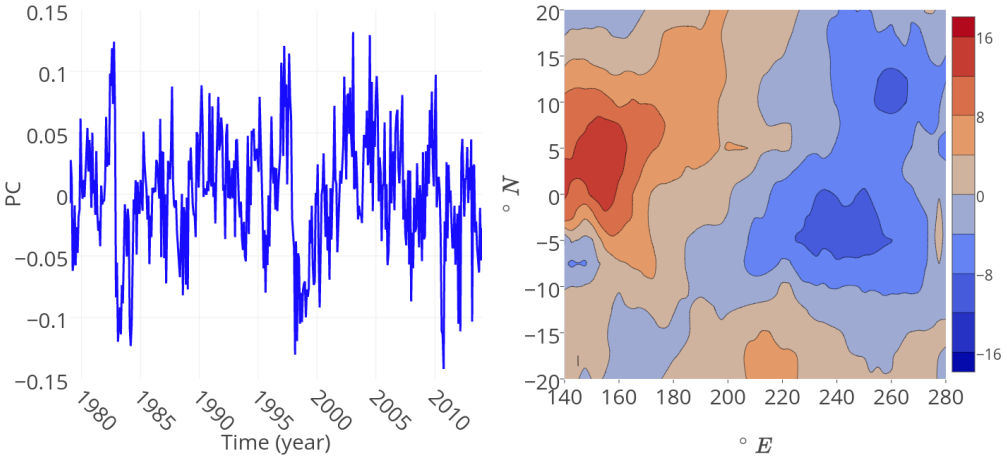


Figure 11. Left the second PC and right the second EOF of the residual of the wind stress associated with the WWB's.

Kleeman, 1999; Latif et al., 1988). To obtain a variable related to the WWB, the linear effect of the SST is subtracted from the zonal component of the wind stress. The second principle component (PC_2), explaining $8\ \%$ of the variance, is associated with these WWB's. In figure 11, the PC and its EOF are presented. Clearly the peaks in the PC are visible before the great El Niño events of 1982 and 1997. Thereby, the EOF has the typical WWB structure, being positive west from the dateline and negative east. The HadISST dataset of the Hadley center has been used for the SST and the NCEP/NCAR Reanalysis dataset for the wind stress, to compute the principle component of the residual of the wind stress for the period 1980-2014.

Finally, the attribute set does not contain any information about the seasonal cycle (SC) yet. The phase locking of an El Niño event is something typical to ENSO. Therefore a sinusoid with the period of a year is used as attribute, to see if it can improve the prediction skill.

To determine at which lead time the different attributes should be applied, the cross-correlation

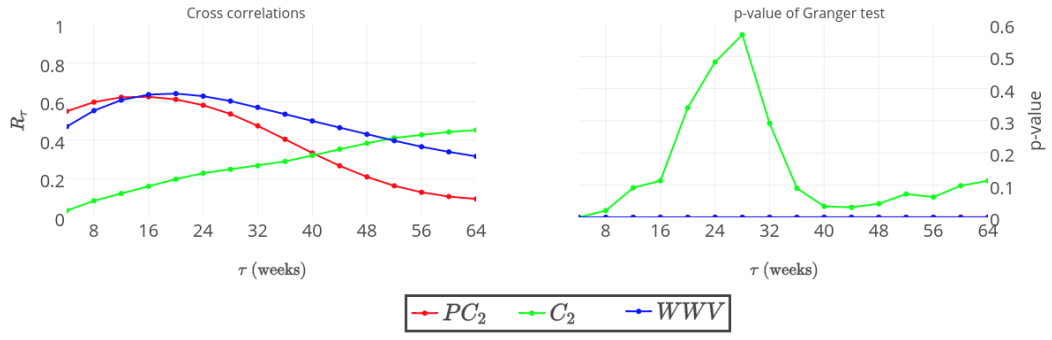


Figure 12. Left the cross correlation of the PC_2, WWV and c_2 with respect to ENSO. On the right the p-value of the Wiener-Granger hypothesis test. A low p-value means the null hypothesis that the time series causes ENSO in the Granger sense is not rejected up to a low significance level. The p-value of the PC_2 and WWV is almost zeros for all lags

and the p-value of the Granger test are considered (figure 12). The cross-correlations of PC_2 and the WWV show clear peaks at respectively 12 and 20 weeks, although the cross-correlation only increases up to the 64 weeks.

Mainly for c_2, the Granger causality test with respect to ENSO shows interesting properties. While the cross correlation increases for a higher lag, the p-value of the Granger causality test has a local minimum at a lag of approximately 44 weeks. This indicates c_2 can be more predictive at this order of lead times. The p-value of the WWV and PC_2 is close to zero at all lags, implying that the effect of autocorrelation is not large and the variable will be more predictive at a shorter lead time.

To summarize, we are interested in the variables that represent a specific physical mechanism related to the prediction of ENSO, in order to define the attribute sets. Both c_2 and the WWV are related to the recharge/discharge mechanism. PC_2 is related to the atmospheric noise from WWBs. The seasonal cycle (SC) is related to the phase locking of El Niño events in Boreal winter.

Moreover, the hybrid model allows us to implement different variables in the attribute set at different lead times. Therefore, the cross correlations and Wiener-Granger causality are taken into account to determine which attribute is more optimal at some lead time. This implies c_2 could replace the WWV at lead times of more than 40 weeks.

4 Results of the predictions

This section presents the results of predictions from the hybrid model. The skill with ANN structures up to three hidden layers is investigated. If possible, a simple structure is preferred to reduce the risk

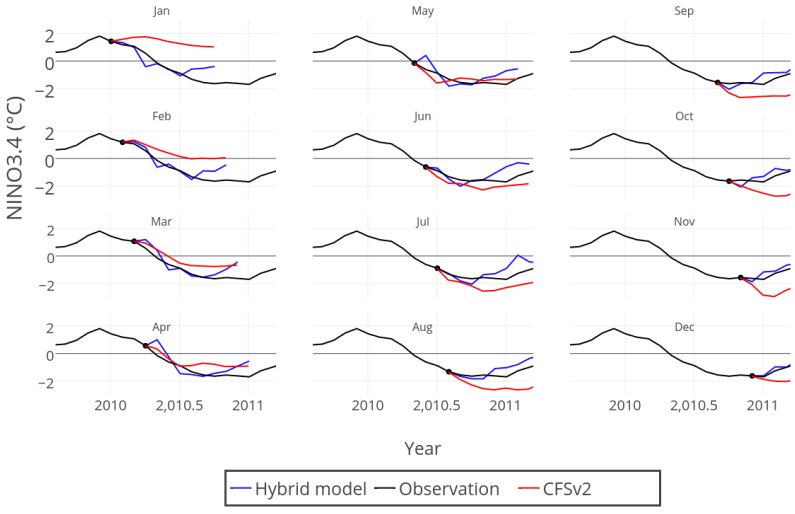


Figure 13. Prediction from all months during the year 2010 up to nine months ahead. Blue is the hybrid model prediction with attributes the three month running mean of WWV, PC_2 and SC. The black line is the observation. Red is the mean of the CFSv2 ensemble prediction.

of overfitting. The predictions are compared to several predictions of the CFSv2 model ensemble of NCEP (from their website: http://www.cpc.ncep.noaa.gov/). First, a comparison between both predictions is made for the year 2010 (figure 13). Moreover, several lead time predictions are shown and compared to the available CFSv2 lead time predictions. Besides, the seasonal bias of the hybrid model and CFSv2 ensemble are compared. Finally, a recent forecast is made and it is shown how the hybrid model predicts the development of ENSO the coming year.

From now on, the Normalized Root Mean Squared Error (NRMSE) is used to indicate the skill of predictor y^A to predict y^B. A low NRMSE indicates the prediction skill is better. The NRMSE is defined as:

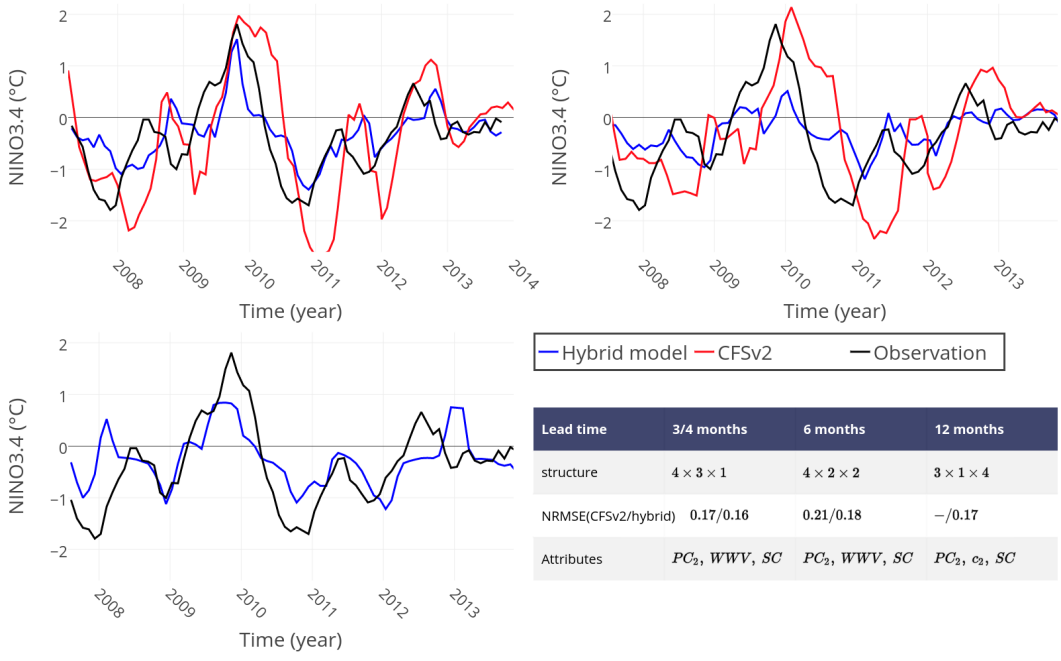
$$NRMSE(y^A, y^B) = \frac{1}{\max_k(y^A, y^B) - \min_k(y^A, y^B)} \sqrt{\frac{\sum_t (y^A(t) - y^P(t))^2}{n}} \tag{24}$$

Here t are the time steps of the time series and n is the number of points in the time series.

4.1 The year 2010

The year 2010 is a recent example of an under-performing CFSv2 ensemble. Especially in January, all members of the ensemble overestimate the NINO3.4 index, resulting in an overestimation of the ensemble mean (see figure 13). In the figure, a hybrid model is used to predict the same period, with ARIMA(12,1,1) and a $2 \times 1 \times 1$ ANN structure with the three month running mean of the WWV, PC_2, the SC and ENSO itself as attributes. Taking the three month running mean from the attributes

is found to increase the prediction skill, because a three month running mean is predicted as well. A $2 \times 1 \times 1$ structure means a feed-forward structure with three layers of respectively two, one and one neuron. This ANN structure is found to be the best performing structure at a three month lead time prediction. It will probably not even be the most optimal ANN structure at other lead times.



Lead time	3/4 months	6 months	12 months
structure	$4 \times 3 \times 1$	$4 \times 2 \times 2$	$3 \times 1 \times 4$
NRMSE(CFSv2/hybrid)	$0.17/0.16$	$0.21/0.18$	$-/0.17$
Attributes	PC_2, WWV, SC	PC_2, WWV, SC	PC_2, e_2, SC

Figure 14. Predictions of the CFSv2 ensemble (red) and the hybrid model, including an ARIMA(12,1,0) model (blue), compared to the observations (black). From an ensemble of eighty-four different ANN structures, some structures resulting in a low NRMSE are presented. Upper left the three month lead time prediction of CFSv2 (red) and four month lead time prediction of the hybrid model. On the upper right the six month lead time predictions. Down left the twelve month lead prediction of the hybrid model. The CFSv2 ensemble does not have a twelve month lead prediction. The table contains informations about the NRMSE of all predictions, ANN structures and attributes used in the hybrid model predictions.

4.2 Lead time predictions

Considering the three, six and twelve month lead time predictions, both the three and six month lead time prediction of the CFSv2 ensemble show some lag and amplification of the real NINO3.4 index (figure 14).

Comparing the three month lead prediction of the CFSv2 ensemble with the four month lead prediction of the hybrid model, the lag of the prediction is less and the amplification is not as large in the hybrid model. While the lead time of the hybrid model is one month longer, the prediction skill is

better in terms of NRMSE.

The prediction skill of the hybrid model decreases at a six month lead compared to the four month lead time prediction. Thereby the lag and amplification of the CFSv2 prediction increase. Although the hybrid does not suffer as much from the lag, it underestimates the El Niño event of 2010. In terms of NRMSE the hybrid model obtains a better prediction skill.

To perform a twelve month lead prediction, the attributes from the shorter lead time predictions are found to be insufficient. However, c_2 of the SSH network has shown to be predictive at this lead time, according to the Granger causality and cross-correlation. Therefore the WWV is replaced by c_2 for this prediction, which is related to the same mechanism. In terms of NRMSE, the twelve month lead prediction even improves the six month lead prediction of the hybrid model. On average the prediction does not contain a lag in this period.

To show the results of figure 14 can be generalized, the mean of the predictions with the nine lowest NRMSE of an ensemble with eighty-four different ANN structures is considered (figure 15). The eighty-four different structures, are all structures up to three hidden layers with up to four neurons. This ensemble does not differ much from the best predictions of figure 14: The spread of the ensemble remains limited, although it is a bit larger in the six and twelve month lead prediction compared to the four month lead prediction.

To test the robustness of these results, a series of cross-validations has been performed. Several percentage splits have been chosen (65-35, 70-30, 75-25 and 80-20), but 200 different initial times of the test set t_i^{test} are randomly chosen between March 1985 and December 2014. This implies that $t_i^{test} > t_f^{train}$ is not necessarily satisfied anymore. If the results for different training and test sets do not deviate much, it is evidence the model also generalizes to an arbitrary training and test set. The cross validation results of the hybrid models of figure 14 are presented in figure 16.

At all three the prediction lead times, the peaks coincide at the same NRMSE for different training-test set ratios. Therefore the different sizes of training and test sets do not seem to influence the result. However, the spread of the peaks increases when the prediction lead time increases. That is why the model results differ more between different parts of time at a higher lead time. Interestingly, at the four and six month lead time predictions, the average NRMSE is lower than the NRMSE of the prediction of figure 14. This implies the predictions at other parts of the set are even better than the prediction shown in figure 14.

4.3 Bias

The hybrid model results in a prediction that is much closer to unbiased, compared to the CFSv2 model ensemble (figure 17). This means the hybrid model underestimates as much as it overestimates. Notice the spring predictability barrier for the CFSv2 ensemble in figure 17, for predictions

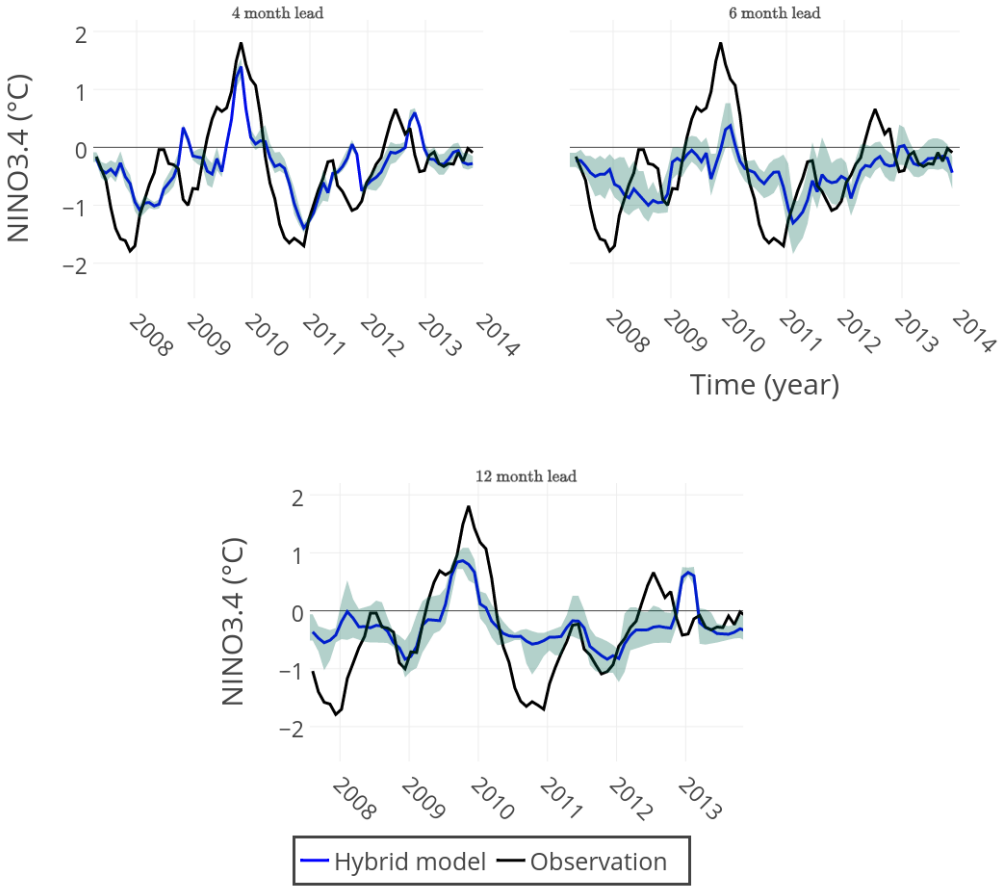


Figure 15. The same predictions as figure 14, but the average of the nine ANN structures with the lowest NRMSE in an ensemble of eighty-four structures is taken for the hybrid model prediction. The shaded blue area denotes the spread of the nine predictions. The NRMSE of the ensemble mean predictions are respectively 0.15, 0.18, 0.17

starting before spring at longer lead times. The predictions of the ensemble from spring, underestimate ENSO on average.

4.4 Recent prediction: From May 2017

Finally, a prediction is made for the coming year in figure 18. Different hybrid models are used, such that ANN structures are chosen that are founs to be optimal at the different lead times. For the predictions up to five months, the attributes WWV, PC_2 and SC are used from 1980 until present. For the twelve month lead prediction, the WWV is replaced by c_2 again. This time it is computed from the Salto/Duacs dataset. Therefore, only a dataset from 1993 until present has been used to train the model and perform the twelve month lead prediction.

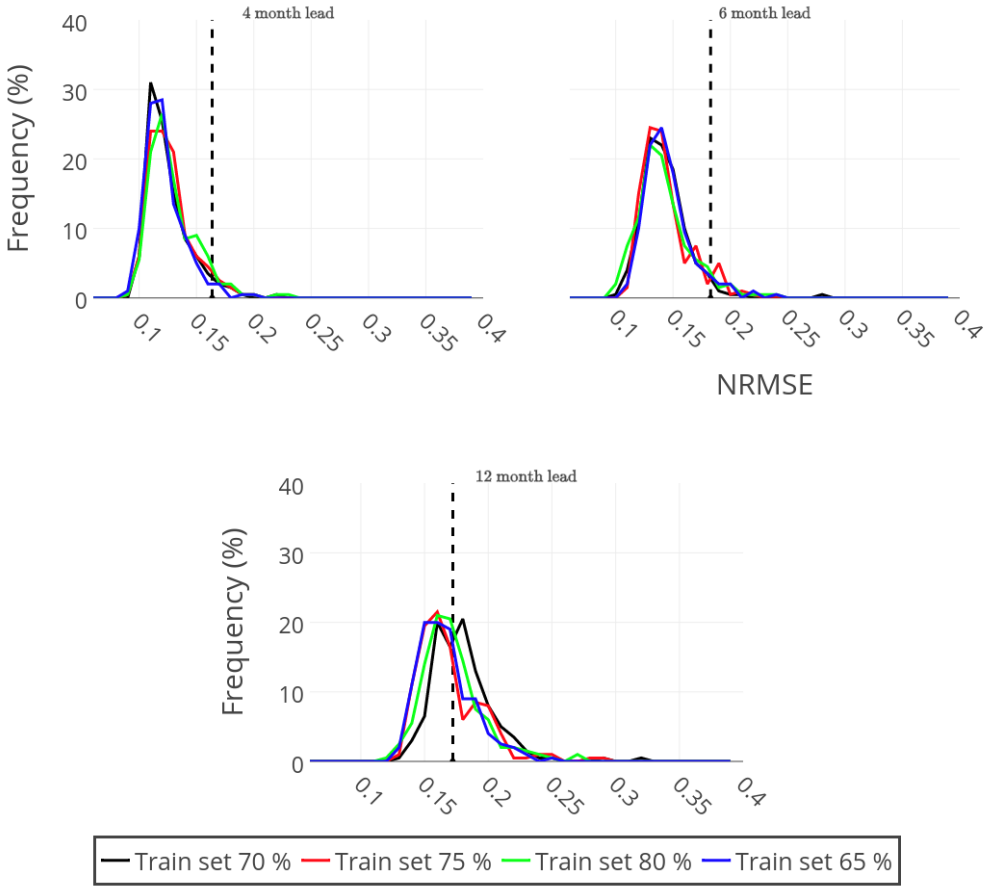


Figure 16. Cross validations of the four, six and twelve month lead predictions of hybrid models from figure 14. The vertical dashed line denotes the NRMSE of the predictions of figure 14.

Interestingly, as can be seen in figure 18, the hybrid model typically predicts much lower ENSO development than the CFSv2 ensemble. The uncertainty of the CFSv2 ensemble is 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) for the coming 9 months. The hybrid models predict development to a strong La Niña (NINO3.4 index lower than -1.5) the coming year. From the time of writing, only time will tell which prediction is better.

5 Summary and Discussion

In this article, it is demonstrated how network theory can describe different properties of the ZC model. This resulted in several variables: A network variable which relates to the parameter controlling the strength of the ocean-atmosphere coupling (S_d), a NON variable that distinguishes between

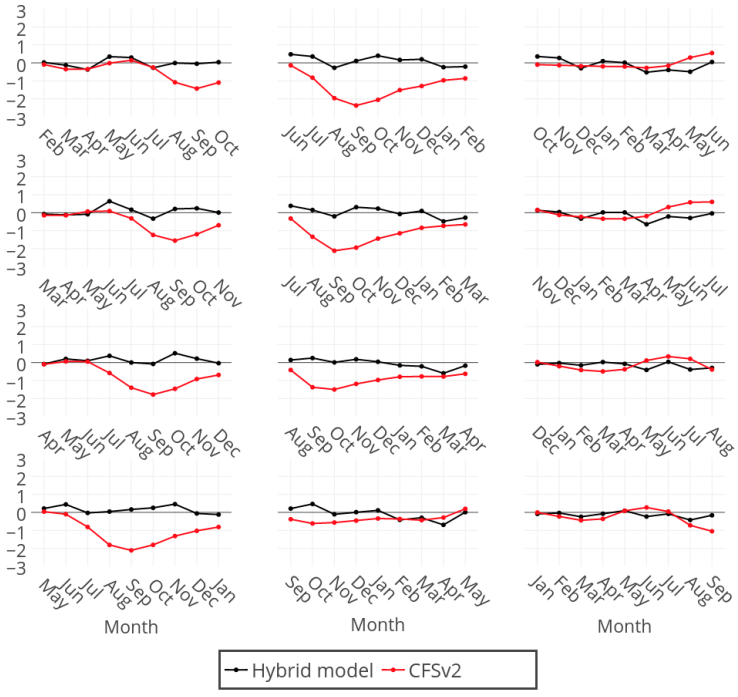


Figure 17. The average bias from all initial months of different predictions. The bias of the CFSv2 ensemble over the period 1982-2009 and of the hybrid model over the period 2007-2014.

super- and sub-critical state of the model (cross clustering between wind and SST network), a variable related to the waves (the zonal skewness in the degree field of the thermocline network), two variables anticipating a percolation-like transition (Δ, c_2) and finally the diffusivity of information in the wind and SST network (λ_2).

Of these network variables, c_2 constructed from observations, being the amount of clusters of size two in a SSH network, is found to provide a warning of a percolation-like transition in the SSH network. This percolation-like transition coincides with an El Niño event. This variable relates to the WWV and the recharge/discharge mechanism, but extends the prediction lead time of the WWV when applied in the prediction scheme.

Furthermore, apart from both 'recharge/discharge' related quantities, the PC_2 and SC improve the prediction skill, representing respectively the WWBs and the phase locking of ENSO.

Moreover, the hybrid model is introduced as method to predicting ENSO. The flexibility of implementing different variables at different lead times, allows the model to improve the CFSv2 ensemble at short lead times (up to six months). Furthermore, it had a better prediction result than all members of the CFSv2 ensemble in January 2010. Thereby the predictions of the hybrid model do not have a systematic bias, as is the case for the CFSv2 ensemble.

Including the network variable c_2 obtained a twelve month lead time prediction comparable to the predictions at shorter lead times. This prediction shows a step towards beating the spring predictabil-

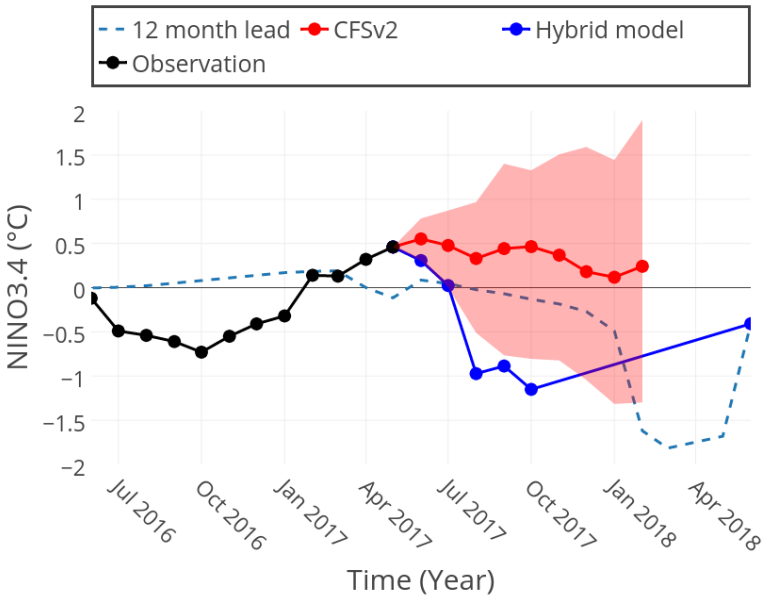


Figure 18. In black the observations until May 2017. Red is the CFSv2 ensemble prediction mean and the shaded area is the spread of the ensemble. The hybrid model prediction in blue is given by predictions from hybrid models found to be most optimal at the different lead times. The dashed blue line is the twelve month lead time prediction.

ity barrier.

Using ML had the advantage of recognizing the early warning signal of c_2 as either a false or true positive. Therefore it can be a more reliable method then considering a warning when the signal exceeds a certain threshold (Ludescher et al., 2014). Moreover, the early signal from the network variable is not only used to predict an El Niño event, but the development of ENSO, as the hybrid model provides a regression of the NINO3.4 index.

ML serves as a tool which is able to recognize important, but subtle patterns. Something the conventional statistical and dynamical models fail to do in the chaotic system.

In the end, the predictions from May 2017 are discussed. By the time of writing, this is the prediction for the coming year. The CFSv2 ensemble mean predicts neutral conditions the coming nine months, with the spread between different members ranging from a strong El Niño to a moderate La Niña. The hybrid model predicts moderate to strong La Niña conditions for the coming year.

Although the results of the methods are promising, some adaptations to the methods might still improve predictions. Several network variables resulted in a clear signal in the ZC model, but not necessarily for the observations. Perhaps the naked eye, the cross-correlation and a Granger causality test are not enough, to determining the suitability of a variable in the observations. Testing all

possible attribute sets in the prediction scheme and comparing results costs time. As a solution, the nonlinear methods 'lagged mutual information' and 'transfer entropy' can be considered techniques to select variables. After all, the attributes are applied in the nonlinear part of the prediction scheme. Consequently, more variables might be found to increase the prediction skill.

Even though the currently applied network measures showed interesting properties, different CN construction methods can still be interesting to apply. The Pearson correlation is a simple, effective method to define links between nodes. However, different properties of CNs could be found, when using mutual information instead. Moreover, the effect of spatial distance between nodes can be investigated and corrected for (Berezin et al., 2012).

Besides, we have limited ourselves to networks within the Pacific area itself. As ENSO is an important mode in the whole climate system, the area used for CN construction might as well be extended. More specifically, it can be interesting to include the Indian Ocean in the CN construction. Evidence is found that a cold SST in the West of the Indian Ocean is related to a WWB a few months later (Wieners et al., 2016). This could result in a variable related to WWBs, but increasing the lead compared to PC_2, which is comparable to c_2 increasing the lead compared of the WWV.

Using ARIMA as simple, yet effective statistical method to apply in the first step of the scheme, the hybrid model shows promising results. However, the exact reason how this model works, remains a topic of investigation. The ARIMA prediction could be related to the linear wave dynamics. It can be interesting to replace the ARIMA part of the scheme, by a dynamical model accounting for these linear wave dynamics. For the same reason Vector Autoregression (VAR) can be used instead of ARIMA. Being a multivariate generalization of an autoregressive model, this can determine the linear effect of other variables on ENSO. The moving average part is not implemented anymore in this case.

Next to investigation of the exact reason the hybrid model works, some adaptations could still improve the prediction scheme. For example, it is assumed the linear and nonlinear part of the model are additive (see equation 17). This is not necessarily the case for the real system (Khashei and Bijari, 2011). Besides, the current model does not take into account possible nonlinear effects from the history, since the ANN describes a nonlinear function which does not depend on the history.

Moreover, a simple ANN structure is chosen in the hybrid model if possible, to reduce the risk of overfitting. However, the ANN could be replaced by a support vector machine (SVM) (Pai and Lin, 2005). In contrast to the backward propagating algorithm of the ANN, the optimization scheme of a SVM does not only search for an optimal solution in terms of least squares minimization. It also searches a solution that is simple.

Appendix

Table 2. List of frequently used mathematical symbols

A	Adjacency matrix (section 2.2.1)
c_2	Amount of clusters of size two in the thermocline network (section 3.2)
c_v	Local cross clustering (section 2.2.2)
C_v	Global cross clustering between SST and wind network (section 3.1.2)
d_i	Degree of a node i in a network (section 2.2.2)
Δ	Variable showing a percolation transition (equation 12)
h	Thermocline depth
λ_2	Algebraic connectivity (section 3.1.5)
μ	Coupling strength between the ocean and atmosphere (equation 5)
PC_2	Second principle component of the residual of the wind stress (section 3.2)
$NRMSE(y^A, y^B)$	Normalized Root Mean Squared Error between predictor y^A and predictant y^B (equation 24)
$R(i,j)$	Pearson correlation (equation 7)
$R_\tau(p,q)$	Cross-correlation (equation 21)
S_d	Skewness of the degree distribution of the SST network (section 3.1.1)
τ^x	Zonal wind stress in the Zebiak Cane model (equation 3)
W_{ij}	Weight between two nodes $i,\ j$ in a network (equation 9)

Acknowledgements. I would like to sincerely thank Henk Dijkstra, Emilio Hernández-García, Qing Yi Feng and Cristóbal López for their supervision, support and suggestions for this research. Moreover, I would like to thank the *Institute for Marine and Atmospheric Research Utrecht* (IMAU) and the *Instituto de Física Interdisciplinar y Sistemas Complejos* (IFISC), for providing the facilities to make this research possible.

References

Berezin, Y., Gozolchiani, a., Guez, O., and Havlin, S.: Stability of Climate Networks with Time, Scientific Reports, 2, 1–8, doi:10.1038/srep00666, 2012.

Bishop, C. M.: Pattern Recognition and Machine Learning, vol. 53, doi:10.1017/CBO9781107415324.004, 2013.

Bjerknes, J.: Atmospheric Teleconnections From The Equatorial Pacific, Monthly Weather Review, 97, 163–172, doi:10.1175/1520-0493(1969)097<0163:ATFTEP>2.3.CO;2, 1969.

Bosc, C. and Delcroix, T.: Observed equatorial Rossby waves and ENSO-related warm water volume changes in the equatorial Pacific Ocean, Journal of Geophysical Research, 113, 1–14, doi:10.1029/2007JC004613, 2008.

Bunge, L. and Clarke, A. J.: On the Warm Water Volume and Its Changing Relationship with ENSO, Journal of Physical Oceanography, 44, 1372–1385, doi:10.1175/JPO-D-13-062.1, http://journals.ametsoc.org/doi/abs/10.1175/JPO-D-13-062.1, 2014.

Chen, D., Cane, M. a., Kaplan, A., Zebiak, S. E., and Huang, D.: Predictability of El Niño over the past 148 years., Nature, 428, 733–736, doi:10.1038/nature02439, 2004.

Dijkstra, H. A.: The ENSO phenomenon: theory and mechanisms, Advances in Geosciences, 6, 3–15, doi:10.5194/adgeo-6-3-2006, 2006.

Dijkstra, H. A. and Burges, G.: International Symposium on Fluid Dynamics, Annual Review Fluid Mechanics, 34, 531–558, 2002.

Donges, J. F., Schultz, H. C. H., Marwan, N., Zou, Y., and Kurths, J.: Investigating the topology of interacting networks: Theory and application to coupled climate subnetworks, European Physical Journal B, 84, 635–651, doi:10.1140/epjb/e2011-10795-8, 2011.

Donner, R. V., Zou, Y., Donges, J. F., Marwan, N., and Kurths, J.: Recurrence networks - a novel paradigm for nonlinear time series analysis, New Journal of Physics, 12, 33 025, doi:10.1088/1367-2630/12/3/033025, http://arxiv.org/abs/0908.3447, 2010.

Fedorov, A. V., Harper, S. L., Philander, S. G., Winter, B., and Wittenberg, A.: How predictable is El Niño?, Bulletin of the American Meteorological Society, 84, 911–919, doi:10.1175/BAMS-84-7-911, 2003.

Feng, Q. Y.: A complex network approach to understand climate variability, Ph.D. thesis, Utrecht University, 2015.

Feng, Q. Y., Vasile, R., Segond, M., Gozolchiani, A., Wang, Y., Abel, M., Havlin, S., Bunde, A., and Dijkstra, H. A.: ClimateLearn: A machine-learning approach for climate prediction using network measures, Geoscientific Model Development Disucssions, pp. 1–18, doi:10.5194/gmd-2015-273, 2016.

Fountalis, I., Bracco, A., and Dovrolis, C.: ENSO in CMIP5 simulations: network connectivity from the recent past to the twenty-third century, Climate Dynamics, 45, 511–538, doi:10.1007/s00382-014-2412-1, http://dx.doi.org/10.1007/s00382-014-2412-1, 2015.

Goddard, L., Mason, S., Zebiak, S., Ropelewski, C., Basher, R., and Cane, M.: Current Approaches to, International Journal of Climatology, 21, 1111–1152, doi:10.1080/002017401300076036, 2001.

Gozolchiani, a., Yamasaki, K., Gazit, O., and Havlin, S.: Pattern of climate network blinking links follows El Niño events, EPL (Europhysics Letters), 83, 28 005, doi:10.1209/0295-5075/83/28005, 2008.

Gozolchiani, A., Havlin, S., and Yamasaki, K.: Emergence of El Ni??o as an autonomous component in the climate network, Physical Review Letters, 107, 1–5, doi:10.1103/PhysRevLett.107.148501, 2011.

Guyon, I. and Elisseeff, A.: An Introduction to Variable and Feature Selection, Journal of Machine Learning Research, 3, 1157–1182, doi:10.1016/j.aca.2011.07.027, 2003.

Hibon, M. and Evgeniou, T.: To combine or not to combine: Selecting among forecasts and their combinations, International Journal of Forecasting, 21, 15–24, doi:10.1016/j.ijforecast.2004.05.002, 2005.

Hush, M. R.: Machine learning for quantum physics, Science, 355, 2017.

Jin, F.-F.: An Equatorial Ocean Recharge Paradigm for ENSO. Part II: A Stripped-Down Coupled Model, Journal of the Atmospheric Sciences, 54, 830–847, doi:10.1175/1520-0469(1997)054<0830:AEORPF>2.0.CO;2, http://journals.ametsoc.org/doi/abs/10.1175/1520-0469(1997)054%3C0830:AEORPF%3E2.0.CO;2, 1997.

Jin, F.-F., Neelin, D. J., and Ghil, M.: El Niño on the Devil's staircase: Annual Subharmonic Steps to Chaos, Science, 264, 70–72, doi:10.1126/science.264.5155.70, 1994.

Khashei, M. and Bijari, M.: A novel hybridization of artificial neural networks and ARIMA models for time series forecasting, Applied Soft Computing Journal, 11, 2664–2675, doi:10.1016/j.asoc.2010.10.015, http://dx.doi.org/10.1016/j.asoc.2010.10.015, 2011.

Latif, M., Biercamp, J., and von Storch, H.: The response of a Coupled Ocean-Atmosphere General Circulation Model to Wind Bursts, Journal of the Atmospheric Sciences, 45, 1988.

Ludescher, J., Gozolchiani, A., Bogachev, M. I., Bunde, A., Havlin, S., and Schellnhuber, H. J.: Very early warning of next El Niño., Proceedings of the National Academy of Sciences of the United States of America, 111, 2064–6, doi:10.1073/pnas.1323058111, http://www.ncbi.nlm.nih.gov/pubmed/24516172, 2014.

Madden, R. a. and Julian, P. R.: Observations of the 40–50-Day Tropical Oscillation—A Review, doi:10.1175/1520-0493(1994)122<0814:OOTDTO>2.0.CO;2, 1994.

Meng, J., Fan, J., Ashkenazy, Y., and Havlin, S.: Percolation framework to describe El Ni\˜no conditions, pp. 1–15, doi:10.1063/1.4975766, http://arxiv.org/abs/1611.06056, 2016.

Moore, A. M. and Kleeman, R.: Stochastic forcing of ENSO by the intraseasonal oscillation, Journal of Climate, 12, 1199–1220, doi:10.1175/1520-0442(1999)012<1199:SFOEBT>2.0.CO;2, 1999.

Newman, M.: Networks An introduction, vol. 6, Oxford university press, Oxford, doi:10.1017/S1062798700004543, 2010.

Pai, P.-F. and Lin, C.-S.: A hybrid ARIMA and support vector machines model in stock price forecasting, Omega, 33, 497–505, doi:10.1016/j.omega.2004.07.024, www.elsevier.com/locate/omega, 2005.

Rebert, J. P., Donguy, J. R., Eldin, G., and Wyrtki, K.: Relations between sea level, thermocline depth, heat content, and dynamic height in the tropical Pacific Ocean, Journal of Geophysical Research, 90, 11 719, doi:10.1029/JC090iC06p11719, http://doi.wiley.com/10.1029/JC090iC06p11719, 1985.

Rodríguez-Méndez, V., Eguíluz M, V. M., Hernández-García, E., and Ramasco, J. J.: Percolation-based precursors of transitions in extended systems, Scientific Reports, 6, 29 552, doi:10.1038/srep29552, http://www.nature.com/articles/srep29552, 2016.

Runge, J. G.: Detecting and Quantifying Causal Interactions from Time Series of Complex Systems, Ph.D. thesis, Humboldt-Universität zu Berlin, http://edoc.hu-berlin.de/dissertationen/runge-jakob-2014-08-05/PDF/runge.pdf, 2014.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., and Hassabis, D.: Mastering the game of Go with deep neural networks and tree search, Nature, 529, 484–489, doi:10.1038/nature16961, http://dx.doi.org/10.1038/nature16961, 2016.

Steinhaeuser, K., Ganguly, A. R., and Chawla, N. V.: Multivariate and multiscale dependence in the global climate system revealed through complex networks, Climate Dynamics, 39, 889–895, doi:10.1007/s00382-011-1135-9, 2012.

Tsonis, A. A., Swanson, K. L., and Roebber, P. J.: What do networks have to do with climate?, Bulletin of the American Meteorological Society, 87, 585–595, doi:10.1175/BAMS-87-5-585, 2006.

Tziperman, E., Stone, L., Cane, M. a., and Jarosh, H.: El nino chaos: Overlapping of resonances between the seasonal cycle and the pacific ocean-atmosphere oscillator, Science, 264, 72–74, doi:10.1126/science.264.5155.72, 1994.

Valenzuela, O., Rojas, I., Rojas, F., Pomares, H., Herrera, L. J., Guillen, A., Marquez, L., and Pasadas, M.: Hybridization of intelligent techniques and ARIMA models for time series prediction, Fuzzy Sets and Systems, 159, 821–845, doi:10.1016/j.fss.2007.11.003, 2008.

Von Der Heydt, A. S., Nnafie, A., and Dijkstra, H. A.: Cold tongue/Warm pool and ENSO dynamics in the Pliocene, Climate of the Past, 7, 903–915, doi:10.5194/cp-7-903-2011, 2011.

Wieners, C. E., de Ruijter, W. P., Ridderinkhof, W., von der Heydt, A. S., and Dijkstra, H. A.: Coherent tropical Indo-Pacific interannual climate variability, Journal of Climate, 29, 4269–4291, doi:10.1175/JCLI-D-15-0262.1, 2016.

Wu, A., Hsieh, W. W., and Tang, B.: Neural network forecasts of the tropical Pacific sea surface temperatures, Neural Networks, 19, 145–154, doi:10.1016/j.neunet.2006.01.004, 2006.

Yeh, S.-W., Kug, J.-S., Dewitte, B., Kwon, M.-H., Kirtman, B. P., and Jin, F.-F.: El Niño in a changing climate, Nature, 461, 511–514, doi:10.1038/nature08316, http://www.nature.com/doifinder/10.1038/nature08316, 2009.