

Article

Dynamic Neuro-Fuzzy Systems for Forecasting El Niño Southern Oscillation (ENSO) Using Oceanic and Continental Climate Parameters as Inputs

Ming Ze Lee ¹, Fatemeh Mekanik ² and Amin Talei ^{1,*}

¹ Discipline of Civil Engineering, School of Engineering, Monash University Malaysia, Subang Jaya 47500, Selangor, Malaysia

² School of Engineering, Swinburne University of Technology, Hawthorn, VIC 3122, Australia

* Correspondence: amin.talei@monash.edu

Abstract: El Niño Southern Oscillation is one of the significant phenomena that drives global climate variability, showing a relationship with extreme events. Reliable forecasting of ENSO phases can minimize the risks in many critical areas, including water supply, food security, health, and public safety on a global scale. This study develops an ENSO forecasting model using the dynamic evolving neural fuzzy inference system (DENFIS), an artificial intelligence-based data-driven algorithm. To forecast ENSO phases for 1, 2, and 3 months ahead, 42 years (1979–2021) of monthly data of 25 oceanic and continental climatic variables and ENSO-characterizing indices are used. The dataset includes 12 El Niño and 14 La Niña events, of which the latest 2 El Niño and 4 La Niña events are reserved for testing while the remaining data are used for training the model. The potential input variables to the model are short-listed using a cross-correlation analysis. Then a systematic input selection procedure is conducted to identify the best input combinations for the model. The results of this study show that the best performing combination of such climate variables could achieve up to 78.57% accuracy in predicting short-term ENSO phases (up to 3 months ahead). Heat content at 0 to 300 m of central equatorial Pacific shows promising performance in forecasting ENSO phases. Moreover, DENFIS was found to be a reliable tool for forecasting ENSO events using multiple oceanic and continental climate variables.

Keywords: ENSO; climate parameters; fuzzy inference systems; DENFIS



Citation: Lee, M.Z.; Mekanik, F.; Talei, A. Dynamic Neuro-Fuzzy Systems for Forecasting El Niño Southern Oscillation (ENSO) Using Oceanic and Continental Climate Parameters as Inputs. *J. Mar. Sci. Eng.* **2022**, *10*, 1161. <https://doi.org/10.3390/jmse10081161>

Academic Editor: Anders Jensen Knudby

Received: 15 July 2022

Accepted: 16 August 2022

Published: 22 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

El Niño-Southern Oscillation (ENSO) is defined as periodic variations in winds and Sea Surface Temperature (SST) due to ocean-atmosphere coupled feedbacks [1,2]. The term oscillation describes the shifting between El Niño and La Niña conditions occurring every few years [3]. In the neutral phase, trade winds displace warm water from the eastern Pacific to the western Pacific while upwelling colder subsurface water in the east [4]. The resulted east-west SST contrast reinforces the east-west pressure difference, further driving the trade winds [4]. The equatorial easterly winds gather water vapor and contribute to the convective rainfall over the western Pacific warm pool. During La Niña (El Niño), the trade winds strengthen (weaken) and increase (reduce) the eastern cooling and atmospheric convection. It is a significant driver of precipitation variability far beyond the tropical Pacific through atmospheric teleconnections [5–7].

During extreme ENSO events, the large-scale changes can elevate the likelihood of extreme global weather events such as cyclones, drought, and intense rainfall. For example, the 2015–2016 El Niño event led to major hydrological crises over eastern and southern Africa, where 29 million people faced food insecurity [8]. Moreover, the hydrological crises, such as extreme drought conditions, were followed by extensive crop failures [9]. Furthermore, studies have shown that the 2015–2016 El Niño event triggered a series

of disease outbreaks and casualties in affected areas [10]. Moreover, this event brought environmental disruptions such as floods to China, causing mass coral bleaching events and forest fires in the Amazon [11–13]. Therefore, understanding and predicting such impactful events are necessary, especially when they have been projected to increase in frequency due to climate warming [2].

Previous research has established that SST and surface wind stress are dominated by upper ocean heat content and characterized by a propagating mode during ENSO events [14]. The spatial and temporal effects of SST anomalies on convective rainfall have been the focus of many studies [15]. Teleconnection with other climatic variables such as air temperature and pressure have been studied mainly on a global scale. With the use of remote sensing data, some studies have shown that, during El Niño, air temperature is higher in most of the tropics and is strongly correlated with SST [15]. Using statistical assessment, Gershunov and Barnett (1998) found the modulation of ENSO affects the Pacific Decadal Oscillation (PDO), a North Pacific climate variability pattern. El Niño (La Niña) patterns are solid and consistent under a positive (negative) phase of PDO [16,17]. Based on retrospective research, researchers introduced several indices to describe the phase and strength of ENSO events using the aforementioned forcing variables (i.e., SST, wind, heat content, and subsurface temperature) and global multivariate patterns [18].

Since the 1980s, several models have been developed and employed to forecast the ENSO phenomenon. ENSO forecast models can be categorized into three types: dynamic-coupled models, statistical models, and hybrid-coupled models. Dynamic-coupled models generally outperform statistical models [19]. Most statistical models are linear models that do not reasonably describe the non-linear features of SST and surface wind anomalies in the Pacific Ocean [20]. However, statistical models still play an essential role in forecasting ENSO events. For instance, Graham et al. [21] used a linear statistical model to predict SST for 7–16 months lead time using near-global surface level pressure (SLP). This model correctly forecasted 13 out of 17 events; however, the authors concluded that their model showed poor performance at tracking the rise of SST at the beginning of El Niño, especially in spring [21]. This significant challenge in forecasting before or during spring is associated with the “spring predictability barrier” (SPB) and has been encountered by both dynamic and statistical models [14,21–23]. Due to the SPB, the models contain initial errors and exhibit prominent error growth that is noise-induced and seasonal [23–25].

Most of the major meteorological centers, such as the U.S. Climate Prediction Center (CPC) and the European Centre for Medium-Range Weather Forecasts (ECMWF), have developed dynamical seasonal forecasts with comprehensive ocean-atmosphere coupled general circulation models (CGCMs), using an ensemble approach [26,27]. Jin et al. [22] investigated the overall performance of ENSO prediction with ten coupled GCMs developed by meteorological centers. Data of the Niño 3.4 index were used at different lead times to examine the accuracy of stimulated variability. The authors concluded that the models’ errors in the simulation of SST were significant, as the forecast performance strongly depended on the season and ENSO phase and intensity. Neutral phase periods were worse to predict than strong El Niño phases. A similar evaluation was made by Barnston et al. [28] using 20 prediction models (12 dynamics, 8 statistical) for forecasting SST in Niño 3.4 region. They reported that the SST forecast for 1981–2010 yielded a correlation coefficient of 0.65 between predicted and observed values. They found that the predictions for the 1990s gave a slightly lower correlation coefficient of 0.6 compared to the one yielded for the 30-year prediction [28]. In some studies, forecast models are developed by combining dynamic and statistical approaches [29]. Using Genetic Algorithms (GA), Hong et al. [29] constructed a dynamic-statistical forecasting model for SST in Niño 3.4 region (5° N– 5° S, 120° W– 170° W) using historical data. However, the derived prediction equations were significantly dependent on initial values and the long-term forecasts (i.e., exceeding 5 months ahead) deviated significantly. Tao and Duan [30], and Tao et al. [31] combined an intermediate-complexity ENSO model (ICM) with a non-linear forcing singular vector as an approach to suppress

the initial errors and errors within ICM. As a result, the authors extended the skillful predictions up to a lead time of 12 months.

Although the representation of ENSO indices in models has shown considerable advancement in forecasting ENSO events during the past decade, there is still room for improvement. The systemic errors shown in the projected ENSO events suggest the deficiencies of existing models' forecasting ability, especially with their dependency on the phase, season, and intensity of ENSO. With the advancements in big data analysis and artificial intelligence (AI), the advantage of AI-based modelling techniques should be taken to improve forecasting problems such as ENSO predictions.

Among AI techniques, artificial neural networks (ANN) have appeared as a powerful tool for modelling non-linear and complex problems. ANN consists of layers of computing nodes, known as neurons, imitating neurons in a biological brain. ANNs can identify the associations between inputs and outputs by learning it through training using input-output data samples. Mu et al. [32] designed a multivariate atmosphere-ocean coupled model using graphical neural networks for ENSO prediction. The authors evaluated the model performance for ENSO index forecasting and achieved correlation coefficients above 0.5 for Niño 3.4 index up to 18 months lead time. Moreover, Ham, Kim, and Luo (2019) utilized the convolutional neural network (CNN) model to predict ENSO indices. The authors concluded that the CNN was superior to almost all dynamical and statistical models (with correlation coefficients exceeding 0.5) for Niño 3.4 predictions up to 17 months lead time. The excellent performance of CNN is consistent with the results found by Zhou and Zhang [33], where high correlations were achieved for Niño 3.4 predictions up to a 17-month lead time. In this study, the authors used a hybrid model combining the principal oscillation patterns (POP) analysis with a CNN-based technique known as the long short-term memory (LSTM) algorithm. In both the above-discussed studies, the authors demonstrated that CNNs outperformed other models as they could distinguish deterministic behavior chaos from random noise [19,33,34]. These recent studies have indicated the potential of using ANNs for ENSO prediction.

The other family of AI-based modeling techniques is known as neuro-fuzzy systems (NFS), which combines the connectionist structure of ANNs with the reasoning capabilities of fuzzy systems. Perhaps this group's most well-known and widely practiced algorithm is the adaptive network-based fuzzy inference system (ANFIS) [35], which employs Takagi-Sugeno fuzzy inference system. ANFIS has been successfully used in several hydrological modelling applications, including rainfall-runoff modelling [36], river stage forecasting [37], rainfall forecasting [38], and evapotranspiration simulation [39]. Nguyen et al. [40] investigated using the ANFIS model to constitute SST anomalies as input variables to predict precipitation index and evapotranspiration-precipitation index for ENSO-induced drought forecasting. The authors obtained promising results with a correlation coefficient of up to 0.75 between the observed and forecasted standardized evapotranspiration-precipitation index (SEPI). The other well-studied NFS algorithm is the dynamic evolving neural fuzzy inference system (DENFIS), which has been successfully used in a wide range of hydrological modelling applications due to its adaptability, including rainfall-runoff modelling, runoff forecasting, reference evapotranspiration modelling, and river water level forecasting [41–43]. However, no studies have been reported on applying DENFIS in predicting ENSO indices or events to our best knowledge.

From the above literature review, it can be inferred that almost all applications of dynamic-coupled, statistical, and data-driven models in predicting ENSO have been focused on predicting the ENSO indices (e.g., Niño 3.4) rather than the ENSO events. Perhaps this is because the definition of ENSO events could differ in each country or region. On the other hand, several research studies have shown reliable results using NFS algorithms in hydrological modeling and forecasting. However, very few research studies focus on forecasting ENSO indices or events using NFS algorithms such as DENFIS. To address these knowledge gaps, this study aims to directly predict ENSO events up to 3-months ahead using a data-driven approach, namely the DENFIS model. In this study, climatic variables

and ENSO-characterizing indices are utilized as inputs for the model to predict the ENSO events (i.e., neural phase, La Niña, or El Niño). The results of this study will address the effectiveness of the proposed forecast model and the selected variables in accurately predicting ENSO events.

2. Materials and Methods

2.1. Data Description

In this study, 42 years of monthly climatic data and indices of the Pacific Ocean and Indian Ocean (1979–2021) were utilized as potential input variables for the model, as shown in Table 1. There are 25 variables, including heat content anomalies for different regions, Trade Wind Index, and months of moving average for the Bivariate El Niño Southern Oscillation Timeseries (BEST) index values. In Table 1, variables are denoted as X1 to X25 considering their regions and average values over certain months. The data was retrieved from National Oceanic and Atmospheric Administration (NOAA) (<https://www.cpc.ncep.noaa.gov/data/indices/>) (accessed on 10 January 2022) and Climate Prediction Centre (CPC) (<https://psl.noaa.gov/enso/dashboard.neut.html>) (accessed on 10 January 2022). Most of the data from NOAA represent in situ measurement, especially for SST, and thus are considered one of the best available direct records of ENSO conditions [44]. In addition, most variables are evaluated in NOAA's ENSO alert system as advisory criteria to characterize the duration of ongoing ENSO events.

These indices and climatic variables were selected to characterize ENSO mode. The major established features of ENSO depend on large-scale spatial distributions of SST and sea level pressure (SLP) [18]. A see-saw pattern is observed for SLP and SST. Walker circulation system results in the difference of SST and SLP over western and eastern Pacific. For example, during a La Niña event, strong Walker circulation causes upwelling of cool water in the east of the Pacific with low air surface pressure. It amplifies warming in the western Pacific with high sea level pressure, while the converse is true for El Niño. Hence, SST and SLP over the western, eastern, and central Pacific Ocean are used to characterize the ENSO mode. SST anomalies in these Pacific regions, computed with different operational definitions and for various regions, include Niño 3, Niño 3.4, Oceanic Niño Index (ONI), and Trans-Niño Index (TNI) [45]. Heat content for the upper 300 m depth of the Pacific was also used as a climatic indicator. The southern oscillation index (SOI) was included to describe sea level pressure fluctuations between the western and eastern tropical Pacific [46]. The wind-driven ocean dynamics of the Walker circulation can also be characterized by the Trade Wind Index, a near-surface wind index across the Pacific. During events, positive SST anomalies in the equatorial are accompanied by anticyclonic anomalous circulations of 200 mb winds [47]. The formation of rainfall and convective rainfall induced by the ENSO phases are also characterized. In this regard, El Niño Southern Precipitation Index (ESPI), El Niño precipitation index (EI), and La Niña precipitation index (LI) were used to measure the rainfall anomalies induced [48]. Moreover, outgoing longwave radiation (OLR) anomalies, which measure radiation at the top of the atmosphere, are used to characterize the cloudiness. This is because clouds formed by the SST anomalies will capture outgoing infrared radiation and lower values of OLR.

Besides climatic variables associated with the phenomenon, ENSO events are also linked to other modes. Quasi-biennial oscillation (QBO) is the oscillation of downward propagating easterly or westerly zonal winds in the equatorial stratosphere [49]. The zonal winds oscillate from east to west for around 28 months [50]. El Niño (La Niña) development is hypothesized to be associated with the east (west) phase of QBO. This has been observed from the circulation anomalies and regional pressure matching consistently with ENSO trends. Hence, this model used the QBO wind index with equivalent pressure of 30 hPa (QBO30) and 50 hPa (QBO50) to understand the possible underlying relationships. The Indian Ocean Dipole (IOD), oscillation of SST in the Indian Ocean, is also linked to ENSO. The warming and cooling of the eastern part of the Indian Ocean correspond to the western Pacific, which is one of the forcing effects hypothesized by Pinault [51].

Similarly, Pacific Decadal Oscillation (PDO) defines SST oscillation between the central north and eastern pacific. The warming and cooling are hypothesized as remote forcing by ENSO [52]. As their indicators, dipole mode index (DMI) and Pacific Decadal Oscillation index (PDO) describe oscillation of IOD and PDO, respectively [53,54]. Next, the geopotential height, defined as the height of a pressure surface above mean sea level, is strongly influenced by ENSO. It refers to a level of the atmosphere above sea level at which a specified atmospheric pressure (pressure surface) is constant. Since cold air is denser than warm air, the pressure surface is lower in colder air masses, and the contrary is true. Hence, warm (cold) SST in the eastern Pacific developed in El Niño (La Niña) simulates above (below) average geopotential height, which gives a positive (negative) Pacific North American index (PNA). Lastly, a composite index such as multivariate ENSO index (MEI) and bivariate ENSO index (BEST) comprising the mathematical operation of the aforementioned variables are used to characterize ENSO [55]. In order to demonstrate dynamic forecasting, the moving average of monthly values are computed considering the past months only. For example, BEST composite index with 3 months average was calculated by averaging the past 3 months' values until the current month to forecast one month ahead in this model. A similar approach was conducted for the 3-month average of ONI, and the 5-month average of BEST and TNI.

Besides climatic variables and indicators, historical records of ENSO monthly events are considered for training and testing. These historical ENSO events were defined using the NOAA's criteria. A threshold value of the Oceanic Niño Index (ONI) was used as the defining criteria. The value of ONI meeting the threshold of +0.5 °C for five consecutive months defines an El Niño event, whereas −0.5 °C for a La Niña event [56]. The past event records were expressed so that only three possible values represent each ENSO phase. The value of 0 means the neutral phase of ENSO, whereas the value of 1 defines El Niño and the value of 2 represents La Niña. There is a total of 12 El Niño and 14 La Niña events among the 512 months of historical data. In this study, 400 out of 512 months (10 El Niño and 10 La Niña events) were used as training data, while the remaining 112 months (2 El Niño and 4 La Niña events) were used as testing data. Based on trials, this distribution of training and testing proportion gives the best result.

Table 1. Input variables used in this study with their description and corresponding source.

Ref	Input Variables	Description	Datasets	Source
X1	Niño3	Average SST anomalies average over 5° S–5° N and 150°–90° W	HadISST1	NOAA PSL [57]
X2	Niño3.4	Average SST anomalies average over 5° S–5° N and 170°–120° W	HadISST1	NOAA PSL [57]
X3	SOI	Normalized pressure difference between Tahiti (equatorial of Pacific) and Darwin (east of Pacific), which then standardized itself	CRU	NOAA PSL/CRU [46]
X4	DMI	Difference of anomalous SST between western equatorial Indian Ocean (50°–70° E and 10° S–10° N) and south equatorial Indian Ocean (90°–110° E and 10° S–10° N)	HadISST1.1 [55]	NOAA PSL
X5	ONI	Past 3 months moving average of Niño 3.4 index based on centered 30-year base period updated every 5 years	ERSSTv5 and ERSSTv3 [58]	NOAA PSL
X6–X8	BEST	Standardized sum of SOI and Niño 3.4 index in moving average of past 1, 3, and 5 months, respectively.	HadISST1.1 [57]	NOAA PSL [55]
X9	MEI v2	Computed as leading principal component time series using the empirical orthogonal function (EOF) of standardized anomalies of sea level pressure, sea surface temperature, zonal and meridional wind component, and outgoing longwave over (30° S–30° N and 100° E–70° W)	NOAA CDR and JRA-55 global reanalysis [59]	NOAA PSL
X10	TNI	Standardized difference between Niño1+2 and Niño 4 with past 5-month moving average	HadISST1 [57]	NOAA PSL [45]
X11	PDO	Standardized principal component time series using EOF of SST anomalies over North Pacific (poleward of 20° N)	HadISST1.1 and ERSSTv5 [57,58]	NOAA PSL

Table 1. Cont.

Ref	Input Variables	Description	Datasets	Source
X12	PNA	Rotated Principal Component Analysis (RPCA) based on anomalies of geopotential height fields at 500 mb over 20°–90° N	CDAS [60]	NOAA CPC
X13	OLR	Anomalies of the outgoing long wave over central equatorial Pacific (160° E–160° W)	CDAS/Reanalysis [60]	NOAA CPC
X14–X16	Heat Content	Pacific integrated temperature anomalies at 0 to 300 m over 3 regions, 160° E–80° W, 130° E–80° W, and 180°–100° W	GODAS [61]	NOAA CPC
X17	200 mb Wind	Zonal average wind anomalies over 2.5° S–2.5° N and 165°–100° W at the altitude of air pressure 200 millibars equivalence	CDAS/Reanalysis [60]	NOAA PSL
X18–X20	850 mb Trade Wind Index	Zonal average wind anomalies over 3 regions over 5° S–5° N, southwest pacific (135° E–180° W), south central pacific (175° W–140° W) and southeast pacific (135° E–120° W) at the altitude of air pressure 850 millibars equivalence	CDAS/Reanalysis [60]	NOAA CPC
X21	ESPI	Normalized sum of precipitation index EI and LI	GPCP v2.2 [48]	NOAA PSL
X22	EI	Rainfall anomalies over eastern Pacific, 10° S–10° N and 160° E–100° W	GPCP v2.2 [48]	NOAA PSL
X23	LI	Rainfall anomalies over Maritime Continent, 10° S–10° N and 90° E–150° E	GPCP v2.2 [48]	NOAA PSL
X24	QBO50	Lower stratospheric, downward propagating zonally average wind at the equator with equivalent pressure of 50 hPa	CDAS/Reanalysis [60]	NOAA CPC
X25	QBO30	Lower stratospheric, downward propagating zonally average wind at the equator with equivalent pressure of 30 hPa	CDAS/Reanalysis [60]	NOAA CPC

2.2. Model Description

DENFIS inherits a similar structure to evolving fuzzy neural network (EFuNN) [62]. In this network, the first layer represents input variables, while the second layer implements the fuzzification process of transforming crisp values (non-linguistic) into fuzzy values (linguistic) [63]. The fuzzy quantification of each input variable space is made using membership functions. These functions describe the degree of membership of the datapoint with the linguistic variables. However, in DENFIS, the first-order Takagi-Sugeno fuzzy rules are employed where a weighted linear combination of crisp inputs is used to generate the outputs instead of producing fuzzy values. The samples of rules in such a system are denoted in Equations (1) and (2):

$$\text{IF } (x \text{ is } A_1) \text{ AND } (y \text{ is } B_1) \text{ THEN } (f_1 = p_1x + q_1y + r_1) \tag{1}$$

$$\text{IF } (x \text{ is } A_2) \text{ AND } (y \text{ is } B_2) \text{ THEN } (f_2 = p_2x + q_2y + r_2) \tag{2}$$

where A_1, A_2 and B_1, B_2 are membership values of input variables x and y , respectively; p_1, q_1, r_1 , and p_2, q_2, r_2 are parameters of output functions f_1 and f_2 , respectively. In DENFIS, the fuzzy inference rules are formed using evolving clustering method (ECM) [62]. Clusters are formed by partitioning the scatters of input space and dynamically updating based on new input. When new input is introduced, a new cluster will be created, or the existing cluster center is repositioned, depending on the maximum cluster radius inserted, denoted as D_{thr} . In DENFIS, the Gaussian membership functions are formulated based on the Euclidean distance of data point to the cluster centres. In the learning process, DENFIS employs a linear least-square estimator (LSE) to create and update linear functions, as shown in Equations (1) and (2). The fuzzy rules are dynamically created and updated in the DENFIS online model through learning based on errors in estimating output. The details of this clustering process can be found in a study by Chang et al. [64].

2.3. Input Data Selection and Model Development

Input selection analysis is required to identify the potential informative input variables that can contribute to capturing the desired output variable. One of the well-known

methods for input selection is correlation analysis, where the correlation coefficient (CC) can be calculated as follows:

$$CC(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \times \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \tag{3}$$

where x and y define x -variable and y -variable; \bar{x} and \bar{y} are average values of x and y , respectively; and n is the number of data points calculated between ENSO events as desired output and climatic variables and ENSO-characterizing indices as inputs. The correlation coefficient is computed using monthly time series for input and output variables. As was stated earlier, the monthly ENSO event data come in values of 0, 1, and 2, representing neutral, El Niño, and La Niña, respectively. Since this study is focused on forecasting ENSO events for 1, 2, and 3 months ahead, the corresponding time series is shifted accordingly to represent such forecasting time leads. The CC values are then compared relatively to identify the best possible set of inputs. The CC values for 25 potential input variables are summarized in Table 2. Based on the CC values, 19 variables are selected as potential inputs. The variables with correlation values lower than 0.1 are excluded for model development and highlighted in Table 2. The excluded variables are DMI, PNA, trade wind index for 135° W–120° W, EI, QBO30, and QBO50. However, this study conducts a sensitivity analysis with the excluded variables to evaluate their forecast possibilities despite their low CC values.

Table 2. Pearson correlation coefficient values between the monthly time series of ENSO phases and input variables.

Forecasting Lead Time	Variables					BEST			MEI	TNI	PDO	PNA	OLR
	Niño 3	Niño 3.4	SOI	DMI	ONI	1-Month Average	3-Month Average	5-Month Average					
1-month	0.230	0.327	0.289	0.040 *	0.321	0.353	0.342	0.307	0.305	0.295	0.266	0.071 *	0.178
2-month	0.216	0.304	0.263	0.060 *	0.287	0.323	0.303	0.258	0.292	0.246	0.265	0.094	0.159
3-month	0.192	0.270	0.261	0.094	0.240	0.300	0.256	0.205	0.265	0.207	0.248	0.097	0.144
	Heat content			200 mb Wind	850 mb Trade Wind			ESPI	EI (ESPI)	LI (ESPI)	QBO50	QBO30	
	130° E–80° W	160° E–80° W	180° E–100° W		135° E–180° W	175° W–140° W	135° W–120° W						
1-month	0.286	0.340	0.364	0.222	0.317	0.200	0.006 *	0.185	0.102	0.268	0.044 *	0.009 *	
2-month	0.314	0.358	0.382	0.188	0.296	0.169	0.012 *	0.177	0.102	0.249	0.017 *	0.008 *	
3-month	0.328	0.357	0.379	0.163	0.290	0.174	0.003 *	0.173	0.088	0.257	0.003 *	0.019 *	

* Marks the results that are statistically insignificant at a 95% confidence level.

In model development, all possible combinations of n number of variables are fed into the model where $n = 1, 2, 3, \dots, k$. In this study, the maximum value of $k = 7$ for the number of variables is considered as the subsequent increase of input number does not improve the forecast abilities. It is worth mentioning that the desired outputs for each input set are 1, 2, and 3 months ahead of ENSO events, denoted by Output ($t + 1$), Output ($t + 2$), and Output ($t + 3$), respectively. Since the result of the model in forecasting ENSO events is a time series of real numbers, the values need to be rounded to either 0, 1, or 2 (Neutral: 0; El Niño: 1; La Niña: 2) before performance evaluation.

DENFIS model requires user input only for independent parameters, D_{thr} and m . D_{thr} is the clustering parameter that determines the maximum size of clusters in the ECM algorithm and, consequently, the number of rules. The initial value, m , defines the smallest number of initial rules to be created. Based on trials, $D_{thr} = 0.1$ and $m = 3$ were found to give the best results. This is aligned with values reported in the literature using DENFIS in rainfall-runoff modelling studies [43,64,65].

2.4. Performance Criterion

To evaluate the model performance, the simulated outputs by the model are compared with the observed outputs. The performance is calculated by the ratio of correctly forecasted

events to the total number of predicted events. The performance criterion is called overall accuracy percentage and can be calculated by the below equation:

$$A (\%) = \frac{c}{t} \times 100 \quad (4)$$

where c is the number of months with correctly predicted ENSO phase and t is the total number of prediction months. In this study, t defines the total of 112 months of the testing data set. Then, the accuracy of predicting the El Niño and La Niña phases solely was calculated using the same approach with c_e as the number of correctly forecasted ENSO events, and t_e is the total number of predicted ENSO events. The top 1% highest overall accuracy of each combination of n variables is recorded.

Instead of selecting input combinations with the highest overall accuracy, the set of the top 1% combinations is taken to assess the model accuracy. As a result, the standard deviation of overall accuracy among the top 1% of n variables ($n = 1$ to 7) combinations is found to be between 0.52% and 2.52%. The combinations of input variables may produce coincidental results instead of persistent ones; this could be attributed to the random noise in ENSO features and adaptability of the input variables. Hence, the frequency of variables in the top 1% combinations and the overall accuracy loss when a specific variable is removed from a combination are also computed to find consistently high-performing variables.

3. Results and Discussion

3.1. Input Selection

The frequency of variables is presented as a frequency distribution in Figure 1 for 1-month ahead prediction. The frequency of variables varies, especially when the number of variables in a combination increases from 3 to 5. Hence, variables that frequently appear in combinations with different lengths (i.e., $n = 1$ to 7) are recommended for the next stage. The BEST 5-months averaged index (X8), MEI (X9), and heat content (X14–X16) were found to have high frequency across different numbers of variables (n) and forecasting lead times (i.e., 1–3 months ahead). The frequency of SOI (X3) and PNA (X12) appearance in input combinations becomes significant for the longest forecasting lead time of 3-months. The overall accuracy loss, however, is presented as a box plot. For example, Figure 2 shows the spread of accuracy loss in the top 1% combinations with six variables for forecasting ENSO events one month ahead, Output ($t + 1$). The box plots indicate the 25th percentile (Q_1), median, and 75th percentile (Q_3) statistics. The lower and upper whiskers indicate $Q_1 - 1.5$ (IQR) and $Q_3 + 1.5$ (IQR), respectively, while IQR is the interquartile range, $Q_3 - Q_1$. Based on the box plots, variables with lower IQR and median were considered the recommended input set. A negative accuracy difference means accuracy loss, while a positive accuracy difference means accuracy gain. This analysis showed that removing one variable in combinations with 4, 5, and 6 variables may not cause accuracy loss and may even increase the accuracy. This can be seen as the upper whisker extends to the boxplot's positive side and extends more as the number of variables increases. Then, the final justification was made by comparing the top 1% of input combinations as certain variables might perform better with a specific variable in the combination. The frequency of duo and trio of variables in the combinations were also recorded to evaluate the possible relationship between variables. The processes specified were repeated for up to six variables in combination for forecasting Output ($t + 2$) and Output ($t + 3$) to find the recommended set of variables.

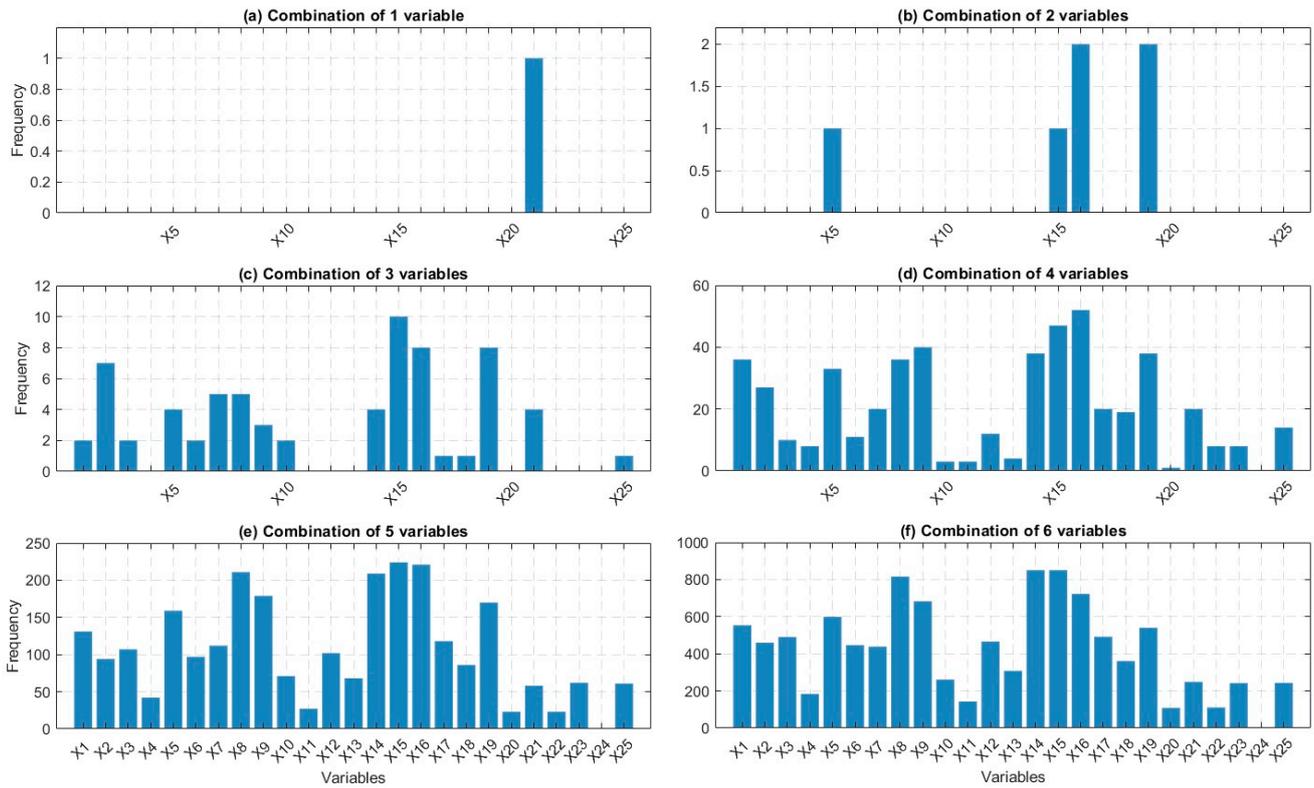


Figure 1. Frequency distribution of variables in the top 1% input combinations of (a) 1, (b) 2, (c) 3, (d) 4, (e) 5, and (f) 6 input variables in forecasting Output ($t + 1$).

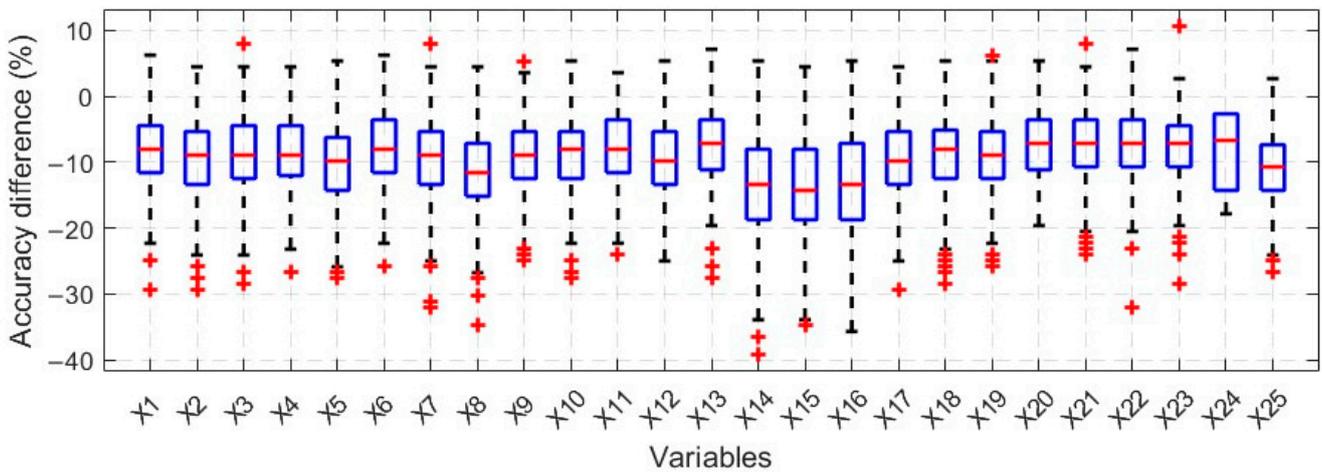


Figure 2. Box plots of accuracy loss for variables in the top 1% input combinations with six variables in forecasting Output ($t + 1$).

Lastly, a cross-correlation analysis is conducted to inspect the inter-association between variables in characterizing the ENSO. The results of this analysis are presented in Figure 3 using a color-coded demonstration of the correlation coefficient between different variables. As it can be seen, the SST characterizing variables such as Niño indices, ONI, BEST indices, and MEI are strongly correlated with one another. OLR and ESPI are correlated to the cluster of SST variables as well. Moreover, another two clusters of variables are found to be inter-correlated including wind variables (200 mb wind & Trade Wind Index) and heat content variables. Indices such as TNI, PDO, PNA, QBO, and DMI demonstrated weak linear relationship (low correlation) with other variables. These observations suggest the potential mutual information or dependence between some variables. The results of this

stage could aid in selecting the most informative variables as inputs to the model while avoiding the ones with potentially repeated information.

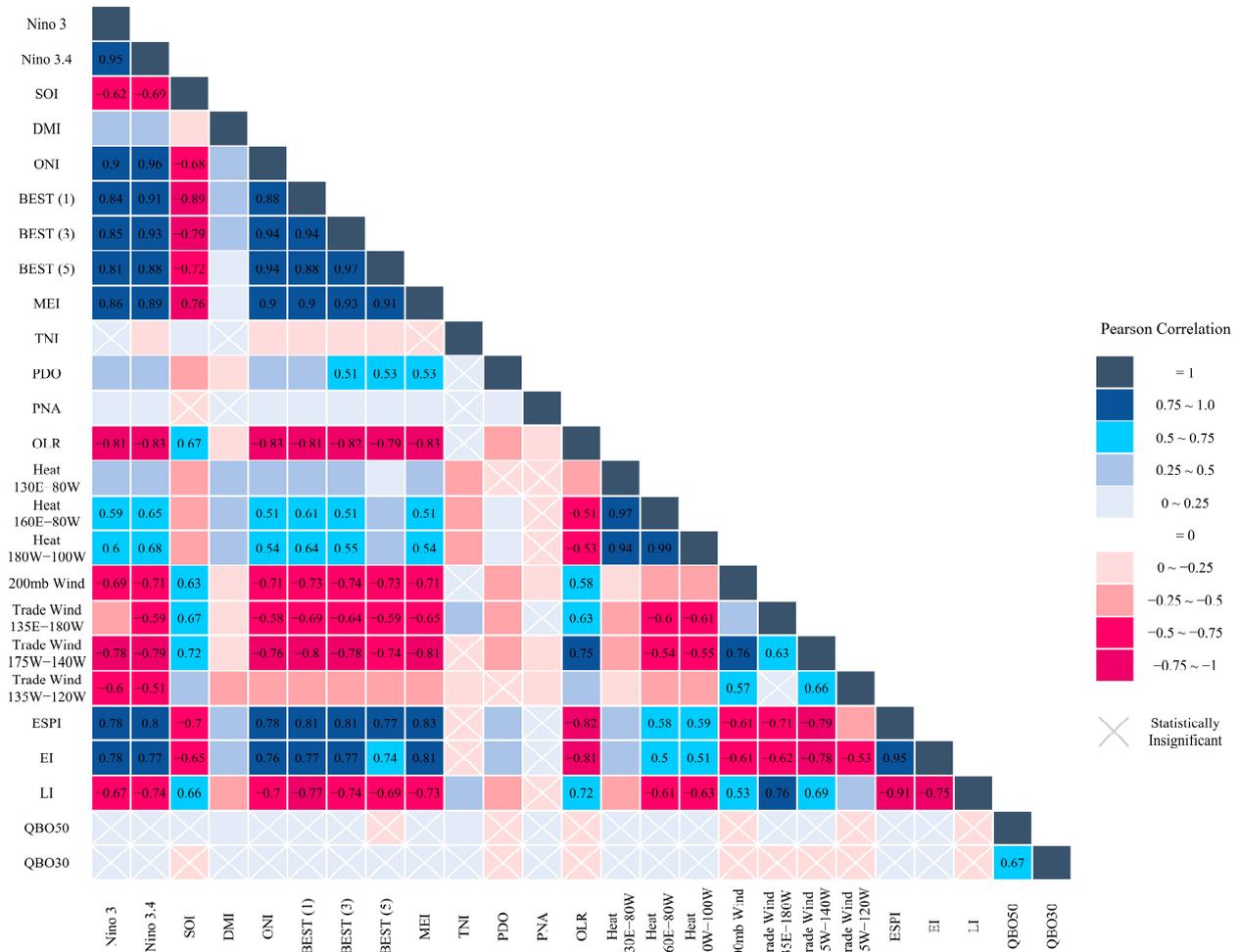


Figure 3. Correlation matrix between the input variables with cross marks indicating statistical insignificance in 95% confidence level.

3.2. Forecasting of the ENSO Event up to 3-Months Ahead

The DENFIS model was trained and validated using the input variables. The averaged overall accuracy in the top 5 combinations of n number of variables are presented for Output ($t + 1$), Output ($t + 2$), and Output ($t + 3$), as shown in Figure 4. The top 5 combinations achieving the highest accuracy are tabulated in Table 3. The highest accuracy for Output ($t + 1$), Output ($t + 2$), and Output ($t + 3$) are 78.57%, 74.11%, and 71.43%, respectively. The spread of overall accuracy in the top 1% of each combination of n variables is investigated. The standard deviation of accuracy in the top 1% decreases as the n , the number of variables in combination converges from 5 to 7. The gradual increase in the number of input variables from 1 to 6 shows a gradual improvement in overall accuracy for all forecasting time leads. However, a further increase over six input variables does not elevate the highest accuracy in the model. This is because the redundant variables are attributed to the excessive complexity of DENFIS architecture. In the combination of a large number of variables (e.g., 5, 6, and 7), the variables within the combinations complement each other in forecasting the desired output, achieving higher overall accuracy. The specific information or signals from ENSO characterizing variables are increasingly contributed to the simulated output as the number of variables increases. At the same time, repeated information might be introduced, causing the saturation of accuracy gain. Based on an observation of the boxplots for n (i.e.,

$n = 1$ to 7) variables, the medians were found to increase as n increased. The increased medians indicate a reduced accuracy loss (negative values) in the spread. The upper whiskers extend to the positive side for a combination of six variables, as shown in Figure 2. It indicates that removing a variable in input combinations may increase or decrease the accuracy of the prediction. Similar diagrams are also developed for other n values not shown here. It was observed that the frequency and its spread toward the positive side increase when the number of variables, n , increases to 7. This finding also suggests that the highest accuracy of the model has reached a saturation level. Moreover, the larger spread of accuracy difference (seen as extended upper whiskers) indicates larger uncertainties in removing variables; hence, an unreliable approach for selecting them. Therefore, the frequency of variables appearing in the top 1% combinations may be a better factor in selecting variables for combinations with a larger n .

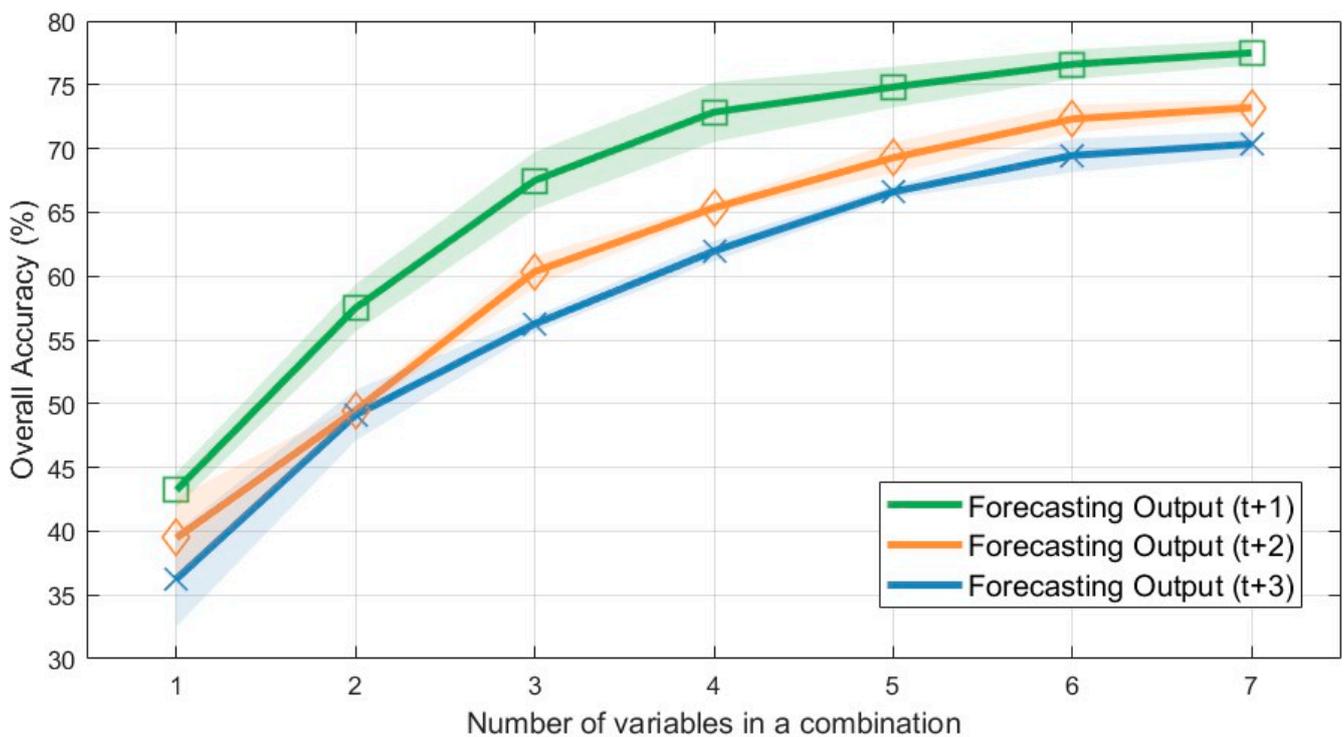


Figure 4. Plots of the top 5 averaged overall accuracy for forecasting ENSO phases for 1, 2, and 3-month ahead, Output (t + 1), Output (t + 2), Output (t + 3), respectively, versus the number of variables in combination with shaded error bar signifying one standard deviation.

The event accuracy was found to be comparable to or better than the overall accuracy with a range of 77.36% to 88.68% for the top 5 combinations showing a larger spread than overall accuracy. This may be due to the limited events in the testing stage. The event accuracy of the top 1% combination gives standard deviations of 1.7 to 2.4 in forecasting ENSO events for 1, 2, and 3-month ahead. The overall accuracy of Output (t + 1) has a relatively steepest gradient and the highest value. Both infer the variables' dependency on the recent changes to correctly forecasted events. This is aligned with other models as there are larger uncertainties introduced before the precedent phases. A study using a statistical model showed that ENSO events were better predicted using stronger SSTs [22]. This is also true in this study since relatively high anomalies are observed near the mature phase of the ENSO event, which provides better prediction to the model [66]. The values of observed and simulated events for the sample of the first 46 months in the testing data are plotted in Figure 5. The observed and simulated values are presented in the form of the ENSO phases, i.e., 0 for neutral, 1 for El Niño, and 2 for La Niña. For a better presentation of the results, both raw and rounded outputs are plotted against observed values in Figure 5.

Table 3. Top 5 highest accuracy combinations of variables grouped separately for each month lag and their corresponding accuracy.

Forecast	Rank	Variables Combination	Overall Accuracy
Output (t + 1)	1	DMI, ONI, BEST (1 month), MEI, Heat Content (160° E–80° W), Trade Wind (175° W–140° W)	78.57%
	2	Niño 3.4, DMI, MEI, Heat Content (130° E–80° W), Heat Content (180° E–100° W), Trade Wind (175° W–140° W)	76.79%
	3	Niño 3, BEST (3 months), BEST (5 months), Heat Content (130° E–80° W), Heat Content (160° E–80° W), ESPI	75.89%
	4	BEST (1 month), BEST (3 months), BEST (5 months), MEI, Heat Content (130° E–80° W), Heat Content (160° E–80° W)	75.89%
	5	BEST (5 months), MEI, TNI, Heat Content (130° E–80° W), Heat Content (160° E–80° W)	75.89%
Output (t + 2)	1	BEST (1 month), MEI, TNI, PNA, Heat Content (130° E–80° W), Trade Wind (175° W–140° W)	74.11%
	2	Niño 3, SOI, BEST (5 months), PNA, Heat Content (130° E–80° W), 200 mb wind	72.32%
	3	Niño 3, BEST (1 month), MEI, TNI, PNA, (160° E–80° W)	72.32%
	4	Niño 3, ONI, MEI, PNA, Heat Content (180° E–100° W), Trade Wind (175° W–140° W)	71.43%
	5	Niño 3, BEST (1 month), MEI, PNA, Heat Content (180° E–100° W), 200 mb wind	71.43%
Output (t + 3)	1	SOI, ONI, MEI, PDO, Heat Content (130° E–80° W), Heat Content (180° E–100° W)	71.43%
	2	Niño 3, SOI, BEST (3 months), MEI, PDO, Heat Content (130° E–80° W)	69.64%
	3	SOI, DMI, BEST (3 months), Heat Content (180° E–100° W), 200 mb wind, Trade Wind (135° W–180° W)	69.64%
	4	Niño 3, Niño 3.4, BEST (3 months), BEST (5 months), PNA, Heat Content (130° E–80° W)	68.75%
	5	Niño 3, BEST (1 month), MEI, PNA, Heat Content (130° E–80° W), Trade Wind (175° W–140° W)	67.86%

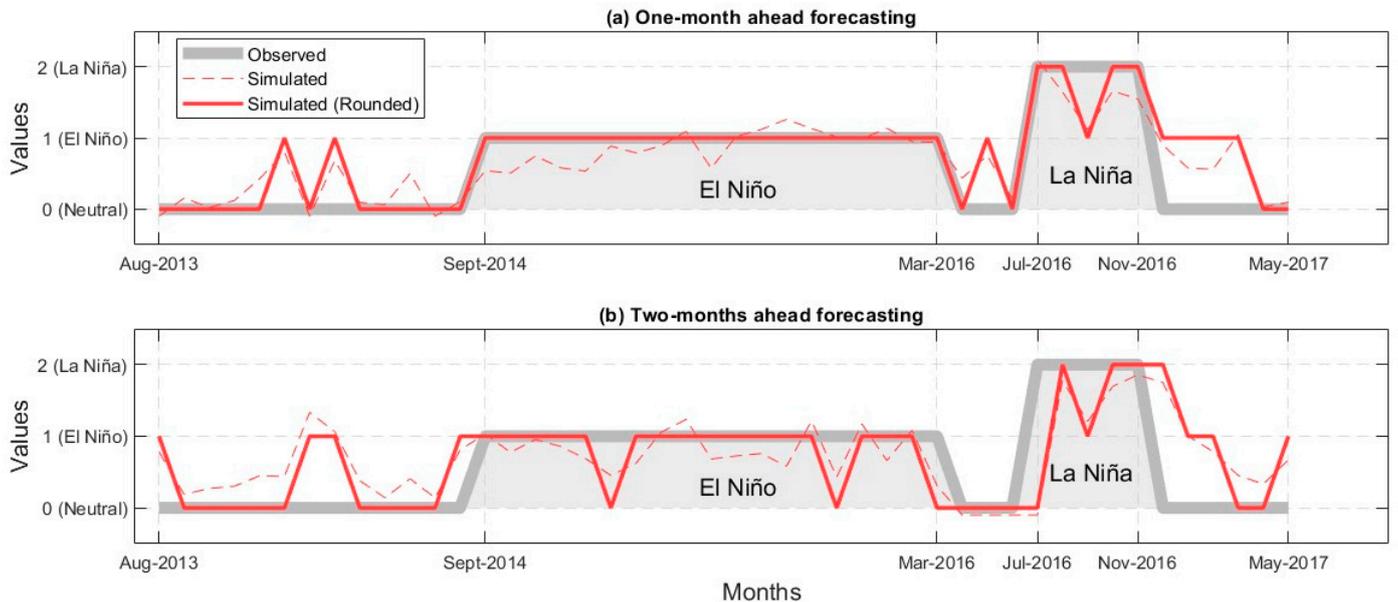


Figure 5. Plots of observed versus (a) one month, and (b) two-month ahead forecasted values representing El Niño and La Niña events from August 2013 to May 2017.

Based on simulated phases using best input combinations for Output (t + 2) in Figure 5b, major false positives were observed at the on-set and decay of the ENSO phase. The occurrence includes early and delayed prediction during a non-neutral phase (Figure 5b) in the on-set and decay in September 2014 and March 2016, respectively. The delay increases in forecasting Output (t + 3) due to further differences in lead times between input anomalies and Output (t + 3). Different combinations of input variables were tried, but no significant improvement in overall accuracy was observed. It is worth mentioning

that the overall accuracy of the proposed model of this study is not directly comparable to the results available in the literature as those studies forecast the climate variables of SST or indices; the results of those models are generally reported in terms of the correlation coefficient between observed and simulated values [19,21,28–30,32,33]. For example, Graham et al. [21] achieved a maximum correlation coefficient of 0.6 for the predicted ENSO indices. In this study, the correlation values between non-rounded simulated and observed phases were calculated as a comparative indicator. The correlation coefficients in predicting Output ($t + 1$), Output ($t + 2$), and Output ($t + 3$) were 0.75, 0.62, and 0.6, respectively. As stated earlier, these results are not directly comparable with the literature as the present study predicts the ENSO phases, not the indices, and the intensities of such events were not part of this study's scope. Despite the promising accuracy attained, the results have shown limitations in predicting larger lead times (months ahead). However, the proposed DENFIS model could be a reliable and quick forecasting asset for short-term forecasting and early warning systems.

Overall, the accuracy deficiency can be explained by two aspects: forecast skills of the model and event-capturing abilities of input variables. Regarding forecasting skills, simulated phases using the best combinations are presented in Figure 5. Besides errors in predicting on-set and decay of phase, false positive predictions were observed within an event. It was found that 37.5% of errors in forecasting Output ($t + 1$) happened during the on-set and decay phase of ENSO events. An example of the errors is that El Niño month was incorrectly predicted amidst the La Niña months (see Figure 5a) between July and November 2016. A similar case was observed during the neutral months (between August 2013 and September 2014) and La Niña months (between July and November 2016). The oscillation between phases generally follows a gradual change (typically with observable variations on a monthly timescale), and abrupt shifts between phases have not been commonly seen in historical data. Such sudden phase shifts in simulated data could be attributed to the weakening of specific signals represented by input variables of that combination. Second, inter-El Niño differences and seasonality may introduce uncertainties in climate anomalies. One of such characteristics is the spring predictability barrier (SPB), where SST errors are observed to be relatively larger in the spring months (April–May–June) [67]. The possible cause is hypothesized to be the weak ocean-atmosphere coupling in the eastern Pacific during the spring [24]. Hence, during these months, the model finds SST anomalies difficult to detect and forecast. This was observed in forecasting Output ($t + 1$) using the best combination of variables. Similar situations were perceived in forecasting Output ($t + 2$) and Output ($t + 3$), but in the prediction of the months after spring due to the shift in lead time. The errors of forecasting solely in spring months for Output ($t + 1$), Output ($t + 2$), and Output ($t + 3$) were observed to be around 24.1%, 32.1%, and 37%, respectively. Besides SPB, inter-El Niño differences may also contribute to the errors. Recent studies have shown that El Niños can be classified into eastern-Pacific (EP) and central-Pacific (CP) types [66]. EP-El Niño was found in the center of the SST anomaly located in the eastern equatorial Pacific, whereas CP-El Niño has most of its surface wind, SST, and subsurface anomalies confined in the central Pacific. Furthermore, studies have shown that there are different timing for triggering SST anomalies in El Niño events [66,68]. To investigate this aspect, El Niño events during the testing and training stage were then determined to address these issues. Based on the EP/CP-index method of Kao and Yu (2009), there are 9 CP-El Niño and 3 EP-El Niño events within this study's dataset. The presented results account solely for 2 CP-El Niño when forecasting Output ($t + 1$) in the testing data, with overall and event accuracy of 78.6% and 83%, respectively. Thus, an investigation was made with another trial using 60% of data for training (6 CP and 2 EP) and 40% for testing (3 CP and 1 EP). The overall and event accuracy for forecasting Output ($t + 1$) were found to be 54.6% and 49.2%, respectively. The stimulated ENSO phases showed correct prediction for the whole EP events, but false warnings were provided in other CP events. However, the decay and on-set of phases were found to be predicted one month earlier or later. This may reflect the insufficiency of the training data. Furthermore, the overall and event

accuracy in forecasting Output (t + 3) was 45.2% and 43.4%, respectively, with increased false positives amidst the events. Overall, the results showed good performance in the short-term prediction of the CP-El and EP-El Niño events.

3.3. Sensitivity Analysis of Climatic Parameters

Table 4 presents the recommended set of variables corresponding to each number of variables in a combination. These sets of input variables were selected to show persistent promising accuracy in the model. There were several combinations of variables attaining similar accuracy, and the top 1% combinations prevailed with significant differences up to 2–3% from what is presented in Table 3. It may be coincidental; hence, they are evaluated based on frequency distribution, accuracy loss addressed for each variable, the computed accuracy of the model, and their cross-correlation with other variables. Some of the variables in the combinations are interchangeable (denoted by “/”), meaning they can be replaced by other options. The combination of variables recommended here has comparable forecasting performance relative to the top combinations presented in Table 3.

Table 4. Set of 6 recommended variables to forecast ENSO events up to 3 months ahead sorted based on the priority (1 stands for the highest while 6 for the lowest).

Number (Priority)	Recommended Combination of Variables for Forecasting *		
	Output (t + 1)	Output (t + 2)	Output (t + 3)
1	Heat (160° E–80° W)/(130° E–80° W)	Heat (130° E–80° W)/(160° E–80° W)/(180° W–100° W)	Heat (130° E–80° W)/(160° E–80° W)
2	Heat (130° E–80° W)/(180° W–100° W)/ONI/Niño 3/Niño 3.4/BEST (5 months/3 months)	Trade Wind (175° W–140° W)/200 mb wind/LI/ESPI	MEI/BEST (5 months)
3	Trade Wind (175° W–140° W)/200 mb wind/ESPI/QBO30	PNA	PNA/PDO
4	MEI/BEST (5 months)	MEI/BEST (5 months)	SOI
5	DMI/TNI	SOI/BEST (3 months/1 month)	BEST (5 months/3 months/1 month)/TNI/DMI/Heat (160° E–80° W)
6	BEST (5 months/3 months/1 month)/SOI	Niño 3/TNI	BEST (5 months/3 months/1 month)/TNI/DMI/Heat (160° E–80° W)

* The notation “/” means the variables are interchangeable.

Among all the variables, heat content in different regions was observed to be the most common and well-performing variable across combinations with a different number of variables and different forecasting lead times. The median of accuracy loss and the frequency for the heat content was the highest compared to other variables. In forecasting Output (t + 1), heat content anomalies in region 160° E–80° W were recommended over other regions. On the other hand, heat anomalies in region 130° E–80° W were favorable for forecasting ENSO events for two and three months ahead. Heat anomalies in region 130° E–80° W appear to give the highest accuracy as a single variable for forecasts for two and three months ahead. Studies have classified El Niño into two different types of time lag. In El Niño events with a negative time lag, SST anomalies form in the eastern Pacific and extend westward beyond 160° E, reaching maturity [68]. In El Niño events with a positive time lag, SST anomalies form in the central Pacific and join in the eastern basin, reaching maturity at 100° W [68]. Hence, it is inferred that the central Pacific region 160° E–80° W captures both SST anomalies well during offsetting these cases. However, region 130° E–80° W covers a larger zone of the Pacific, introducing more uncertainties in short-term predictions. Region 180° W–80° W, however, concentrating in the central Pacific may capture lesser information and forecast solely the maturity of CP-El Niño well [68]. Although heat content with different regions was strongly cross-correlated with $>|\pm 0.95|$ correlation values, the trio of heat regions 130° E–80° W, 160° E–80° W, and BEST (5-month) were found to have the highest frequency among other trio of variables in forecasting Output (t + 1). To avoid repeated information, variables characterizing sea surface temperature were recommended in replacement for one of the two heat content variables in a combination.

The trade wind index for the south central Pacific region (175° W– 140° W) is observed mainly in the short-term forecast (one-month lead time). Relatively weaker 200 mb wind anomalies are proposed as a replacement since both wind components are interlinked in the circulation. The accuracy loss and frequency of the trade wind index decrease as the forecasting lead time increases. However, the SOI index, which describes sea level pressure, is found to be a good input in forecasting ENSO events for all studied lead times, especially two and three-months ahead forecasts. This finding was also confirmed in composite indices, where the presence of SOI in those indices enhanced the forecasting skills for 2 and 3-months ahead forecasts. MEI and BEST indices were computed using SOI as part of their components, reflecting their capabilities in forecasting 1–3 months ahead. It is further aligned with the results of linear statistical models using sea level pressure and wind anomalies [21]. The authors suggested that the wind anomalies that show consistent features from one event to another are not responsible for generating temperature anomalies. The analysis of their models concluded that SLP anomalies provide SST development information up to 7 to 16 months lead time, whereas wind models only do well for 1 to 3 months forecasting lead time. Kao and Yu [66] also suggested that the SST anomalies are the main forcing factors of EP and CP-El Niño. Hence, it can be inferred that high trade wind anomalies are generally observed in the central Pacific near the maturity phase instead of acting as a forcing factor for SST. Therefore, models may perform relatively better in the South-Central Pacific region, where both CP and EP types of El Niño wind anomalies are present. Regarding SLP, previous studies have suggested that tropical SST variations can force extratropical SLP variability, while the SLP are precursors of SST variability [69]. Hence, the variation of SST may be captured and characterized better by SLP compared to trade winds of antecedent months.

The next most repeated variables across Table 4 are the BEST indices with a 3-month and 5-month moving average, followed by ONI with a 3-month moving average of Niño 3.4. The moving average of variables is usually used in the statistical model of predicting SST to reduce the uncertainty of abrupt changes. The superior performance from these moving average indices originates from the information in averaged antecedent monthly values. This information provides better forecasting abilities in larger lead times where recent information is absent. Hence, the moving average variables perform better than the monthly value of Niño 3.4, Niño 3, and BEST index when forecasting 2–3 months ahead. In forecasting Output ($t + 1$), Niño 3.4 and Niño 3 indices are proposed to be replaced with ONI as they address the similar region of SST anomalies with different spatial computations. Niño 3.4 and Niño 3 were interchangeable in the input combinations as each one may capture an event that the other one lost to capture. This is supported by studies showing Niño 3.4 could not capture EP-La Niña due to their general lower intensity and spread of anomalies that are not within the region [66]. MEI was present in the top 1% combinations and can replace BEST indices as both constitute the SLP and SST components.

All 25 variables were considered in the input selection process during the model validation. It was found that the previously excluded variables, DMI, PNA, EI, and QBO30, appeared in the top 1% combinations. Although these variables showed relatively low correlation values with the desired output, their presence in the top high-performing input combinations suggests their potential non-linear associations with the output that have been informative for the model. However, due to their low frequency, it could be inferred that these variables played the role of a supplementary variable in the input combination. Interestingly, DMI, PNA, and QBO30 had lower cross-correlation values with other variables suggesting their different information compared to other variables. PNA performed well in forecasting two and three months ahead with a high frequency of appearance and accuracy loss in removal. The frequency of duo of PNA and heat anomalies in region 130° E– 80° W prevails the highest performance in forecasting Output ($t + 3$), whereas duo of BEST (5-month averaged) and heat anomalies in region 130° E– 80° W for Output ($t + 2$) and Output ($t + 1$) performed very well. TNI was another observed variable with relatively lower cross-correlation values with the output. TNIs have been used for

classifying CP-El Niño [18]; however, in the present study, the model's performance in using TNI was not promising. DMI, QBO30, and OLR showed considerable high accuracy loss but low frequency. It may suggest their insufficient characterizing abilities in forecasting ENSO events. Since OLR describes only the central pacific, uncertainties may be introduced in the long term.

Similarly, ESPI, EI, and LI precipitation index have a relatively low frequency. However, the ESPI precipitation index showed the highest overall accuracy in forecasting Output (t + 1) as a single variable but was insignificant for larger forecasting lead times. One possible explanation is that the precipitation may require a shorter time scale than the resolved 1-month average to better characterize the phases [21]. The monthly anomalies may have averaged the predictive characteristics of the index. Among all the variables, the QBO50 wind index had the least appearance.

Overall, SOI and PNA showed reasonable forecasting abilities (increased frequency and accuracy loss) as the forecasting lead time increased. LI and QBO30 showed distinctive decreased spread and increased accuracy loss, but their frequency decreased as the forecasting lead time increased. The frequency of the trio of variables was insignificant, but the frequency of the trio increases when progressing from Output (t + 1) to Output (t + 3). The highest frequency of trio combination was observed in MEI, PNA, heat content (130° E–80° W), and BEST (1 month averaged)/SOI forecasting Output (t + 3) trios.

4. Conclusions

In this study, the forecasting ability of AI-based data-driven technique, dynamic evolving neural fuzzy inference system (DENFIS), in forecasting ENSO events on a monthly timescale up to 3-months ahead using climatic variables and ENSO-characterizing indices is investigated. Furthermore, the model performance is evaluated using its accuracy in predicting ENSO events.

The following can be concluded in this study:

(1) The proposed model achieved accuracy up to 78.57% in forecasting ENSO phases one month ahead, Output (t + 1). The input combination that gave such an accuracy consisted of DMI, ONI, BEST (1 month averaged), MEI, heat content in region (160° E–80° W), and trade wind index in the central Pacific (175° W–140° W). The accuracy decays to 71.43% when forecasting 3 months ahead, Output (t + 3). The other successful variables in predicting ENSO events were SOI, BEST index with 3 and 5 averaged months values, PNA, and heat content at region (130° E–80° W). Trade wind index for central Pacific prevails SOI in forecasting Output (t + 1), while SOI performs better in larger forecasting lead times. Overall, heat content at three specified regions showed the best performance across all variables. It is inferred that the region of 160° E–80° W could describe the characteristics of most types of ENSO events.

(2) DENFIS model was observed to perform well in capturing events under the uncertainty compounded by both seasonality (SGB) and inter-difference of the ENSO events. It is inferred that the model can describe the non-linear features of variables. However, the false positive predictions indicate the model's weakness in replicating the gradual changes of the ENSO events, as most of the errors were attributed to abrupt changes in the forecast. The on-set and decay of ENSO phases were also wrongly predicted, especially for longer forecasting lead times. Overall, the model shows the potential as an early warning tool in deciding the phases of ENSO.

(3) As recommendations, pre-processing data such as averaging antecedent values could be investigated to achieve better accuracy. Sensitivity analysis can be conducted to find the model's best-customized heat content and trade wind region. Moreover, the model may be specifically trained for a different type of ENSO events to address the inter-difference of ENSO events and produce better accuracy, especially in longer forecasting lead times.

Author Contributions: Conceptualization, F.M. and A.T.; methodology, A.T., F.M. and M.Z.L.; software, M.Z.L. and A.T.; validation, A.T., F.M. and M.Z.L.; formal analysis, M.Z.L.; investigation,

M.Z.L. and A.T.; resources, A.T.; data curation, M.Z.L.; writing—original draft preparation, M.Z.L.; writing—review and editing, A.T. and F.M.; visualization, M.Z.L.; supervision, A.T.; project administration, A.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bjerknes, J. A possible response of the atmospheric Hadley circulation to equatorial anomalies of ocean temperature. *Tellus* **1966**, *18*, 820–829. [\[CrossRef\]](#)
2. Santoso, A.; McPhaden, M.J.; Cai, W. The defining characteristics of ENSO extremes and the strong 2015/2016 El Niño. *Rev. Geophys.* **2017**, *55*, 1079–1129. [\[CrossRef\]](#)
3. Wang, C.; Deser, C.; Yu, J.Y.; DiNezio, P.; Clement, A. El Niño and southern oscillation (ENSO): A review. *Coral Reefs East. Trop. Pac.* **2017**, *8*, 85–106.
4. Cai, W.; Santoso, A.; Wang, G.; Yeh, S.W.; An, S.I.; Cobb, K.M.; Collins, M.; Guilyardi, E.; Jin, F.F.; Kug, J.S. ENSO and greenhouse warming. *Nat. Clim. Change* **2015**, *5*, 849–859. [\[CrossRef\]](#)
5. Allan, R.; Lindesay, J.; Parker, D. El Niño southern oscillation and climatic variability. *Oceanogr. Lit. Rev.* **1997**, *6*, 555.
6. Hoerling, M.P.; Kumar, A.; Zhong, M. El Niño, La Niña, and the nonlinearity of their teleconnections. *J. Clim.* **1997**, *10*, 1769–1786. [\[CrossRef\]](#)
7. Okumura, Y.M.; Deser, C. Asymmetry in the duration of El Niño and La Niña. *J. Clim.* **2010**, *23*, 5826–5843. [\[CrossRef\]](#)
8. Sazib, N.; Mladenova, I.E.; Bolten, J.D. Assessing the impact of ENSO on agriculture over Africa using earth observation data. *Front. Sustain. Food Syst.* **2020**, *4*, 509914. [\[CrossRef\]](#)
9. Nobre, G.G.; Muis, S.; Veldkamp, T.I.; Ward, P.J. Achieving the reduction of disaster risk by better predicting impacts of El Niño and La Niña. *Prog. Disaster Sci.* **2019**, *2*, 100022. [\[CrossRef\]](#)
10. Anyamba, A.; Chretien, J.P.; Britch, S.C.; Soebiyanto, R.P.; Small, J.L.; Jepsen, R.; Forshey, B.M.; Sanchez, J.L.; Smith, R.D.; Harris, R. Global disease outbreaks associated with the 2015–2016 El Niño event. *Sci. Rep.* **2019**, *9*, 1930. [\[CrossRef\]](#)
11. Barkley, H.C.; Cohen, A.L.; Mollica, N.R.; Brainard, R.E.; Rivera, H.E.; DeCarlo, T.M.; Lohmann, G.P.; Drenkard, E.J.; Alpert, A.E.; Young, C.W. Repeat bleaching of a central Pacific coral reef over the past six decades (1960–2016). *Commun. Biol.* **2018**, *1*, 177. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Barni, P.E.; Rego, A.C.M.; Silva, F.d.C.F.; Lopes, R.A.S.; Xaud, H.A.M.; Xaud, M.R.; Barbosa, R.I.; Fearnside, P.M. Logging Amazon forest increased the severity and spread of fires during the 2015–2016 El Niño. *For. Ecol. Manag.* **2021**, *500*, 119652. [\[CrossRef\]](#)
13. Ma, F.; Ye, A.; You, J.; Duan, Q. 2015–2016 floods and droughts in China, and its response to the strong El Niño. *Sci. Total Environ.* **2018**, *627*, 1473–1484. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Latif, M.; Anderson, D.; Barnett, T.; Cane, M.; Kleeman, R.; Leetmaa, A.; O'Brien, J.; Rosati, A.; Schneider, E. A review of the predictability and prediction of ENSO. *J. Geophys. Res. Ocean.* **1998**, *103*, 14375–14393. [\[CrossRef\]](#)
15. Plisnier, P.; Serneels, S.; Lambin, E. Impact of ENSO on East African ecosystems: A multivariate analysis based on climate and remote sensing data. *Glob. Ecol. Biogeogr.* **2000**, *9*, 481–497. [\[CrossRef\]](#)
16. Gershunov, A.; Barnett, T.P. Interdecadal modulation of ENSO teleconnections. *Bull. Am. Meteorol. Soc.* **1998**, *79*, 2715–2726. [\[CrossRef\]](#)
17. Whitfield, P.H.; Moore, R.; Fleming, S.W.; Zawadzki, A. Pacific decadal oscillation and the hydroclimatology of western Canada—Review and prospects. *Can. Water Resour. J.* **2010**, *35*, 1–28. [\[CrossRef\]](#)
18. Norel, M.; Kałczyński, M.; Pińskwar, I.; Krawiec, K.; Kundzewicz, Z.W. Climate Variability Indices—A Guided Tour. *Geosciences* **2021**, *11*, 128. [\[CrossRef\]](#)
19. Ham, Y.G.; Kim, J.H.; Luo, J.J. Deep learning for multi-year ENSO forecasts. *Nature* **2019**, *573*, 568–572. [\[CrossRef\]](#)
20. Wu, A.; Hsieh, W.W.; Tang, B. Neural network forecasts of the tropical Pacific sea surface temperatures. *Neural Netw.* **2006**, *19*, 145–154. [\[CrossRef\]](#)
21. Graham, N.E.; Michaelsen, J.; Barnett, T.P. An investigation of the El Niño–Southern Oscillation cycle with statistical models: 2. Model results. *J. Geophys. Res. Ocean.* **1987**, *92*, 14271–14289. [\[CrossRef\]](#)
22. Jin, E.K.; Kinter, J.L.; Wang, B.; Park, C.K.; Kang, I.S.; Kirtman, B.; Kug, J.S.; Kumar, A.; Luo, J.J.; Schemm, J. Current status of ENSO prediction skill in coupled ocean–atmosphere models. *Clim. Dyn.* **2008**, *31*, 647–664. [\[CrossRef\]](#)
23. Duan, W.; Zhao, P.; Hu, J.; Xu, H. The role of non-linear forcing singular vector tendency error in causing the “spring predictability barrier” for ENSO. *J. Meteorol. Res.* **2016**, *30*, 853–866. [\[CrossRef\]](#)
24. Webster, P. The annual cycle and the predictability of the tropical coupled ocean–atmosphere system. *Meteorol. Atmos. Phys.* **1995**, *56*, 33–55. [\[CrossRef\]](#)
25. Mu, M.; Xu, H.; Duan, W. A kind of initial errors related to “spring predictability barrier” for El Niño events in Zebiak–Cane model. *Geophys. Res. Lett.* **2007**, *34*, L03709. [\[CrossRef\]](#)
26. Kanamitsu, M.; Kumar, A.; Juang, H.M.H.; Schemm, J.K.; Wang, W.; Yang, F.; Hong, S.Y.; Peng, P.; Chen, W.; Moorthi, S. NCEP dynamical seasonal forecast system 2000. *Bull. Am. Meteorol. Soc.* **2002**, *83*, 1019–1038. [\[CrossRef\]](#)

27. Mason, S.J.; Goddard, L.; Graham, N.E.; Yulaeva, E.; Sun, L.; Arkin, P.A. The IRI seasonal climate prediction system and the 1997/1998 El Niño event. *Bull. Am. Meteorol. Soc.* **1999**, *80*, 1853–1874. [[CrossRef](#)]
28. Barnston, A.G.; Tippet, M.K.; L'Heureux, M.L.; Li, S.; DeWitt, D.G. Skill of real-time seasonal ENSO model predictions during 2002–2011: Is our capability increasing? *Bull. Am. Meteorol. Soc.* **2012**, *93*, 631–651. [[CrossRef](#)]
29. Hong, M.; Zhang, R.; Wang, D.; Feng, M.; Wang, Z.; Singh, V.P. Reconstruction of a dynamical–statistical forecasting model of the ENSO index based on the improved self-memorization principle. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **2015**, *101*, 14–26. [[CrossRef](#)]
30. Tao, L.; Duan, W. Using a non-linear forcing singular vector approach to reduce model error effects in ENSO forecasting. *Weather Forecast.* **2019**, *34*, 1321–1342.
31. Tao, L.; Duan, W.; Vannitsem, S. Improving forecasts of El Niño diversity: A non-linear forcing singular vector approach. *Clim. Dyn.* **2020**, *55*, 739–754. [[CrossRef](#)]
32. Mu, B.; Qin, B.; Yuan, S. ENSO-ASC 1.0. 0: ENSO deep learning forecast model with a multivariate air–sea coupler. *Geosci. Model Dev.* **2021**, *14*, 6977–6999. [[CrossRef](#)]
33. Zhou, L.; Zhang, R.H. A Hybrid Neural Network Model for ENSO Prediction in Combination with Principal Oscillation Pattern Analyses. *Adv. Atmos. Sci.* **2022**, *39*, 889–902. [[CrossRef](#)]
34. Tangang, F.; Hsieh, W.; Tang, B. Forecasting the equatorial Pacific sea surface temperatures by neural network models. *Clim. Dyn.* **1997**, *13*, 135–147. [[CrossRef](#)]
35. Jang, J.-S. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Man Cybern.* **1993**, *23*, 665–685. [[CrossRef](#)]
36. Nayak, P.C.; Sudheer, K.; Rangan, D.; Ramasastri, K. A neuro-fuzzy computing technique for modeling hydrological time series. *J. Hydrol.* **2004**, *291*, 52–66. [[CrossRef](#)]
37. Nawaz, N.; Harun, S.; Talei, A. Application of adaptive network-based fuzzy inference system (ANFIS) for river stage prediction in a tropical catchment. *Appl. Mech. Mater.* **2015**, *735*, 195–199. [[CrossRef](#)]
38. Mekanik, F.; Imteaz, M.; Talei, A. Seasonal rainfall forecasting by adaptive network-based fuzzy inference system (ANFIS) using large scale climate signals. *Clim. Dyn.* **2016**, *46*, 3097–3111. [[CrossRef](#)]
39. Shiri, J.; Sadraddini, A.A.; Nazemi, A.H.; Kisi, O.; Marti, P.; Fard, A.F.; Landaras, G. Evaluation of different data management scenarios for estimating daily reference evapotranspiration. *Hydrol. Res.* **2013**, *44*, 1058–1070. [[CrossRef](#)]
40. Nguyen, L.B.; Li, Q.F.; Ngoc, T.A.; Hiramatsu, K. Adaptive neuro-fuzzy inference system for drought forecasting in the Cai River basin in Vietnam. *J. Fac. Agric. Kyushu Univ.* **2015**, *60*, 405–415. [[CrossRef](#)]
41. Adnan, R.M.; Liang, Z.; Parmar, K.S.; Soni, K.; Kisi, O. Modeling monthly streamflow in mountainous basin by MARS, GMDH-NN and DENFIS using hydroclimatic data. *Neural Comput. Appl.* **2021**, *33*, 2853–2871. [[CrossRef](#)]
42. Nguyen, P.K.T.; Chua, L.H.C.; Talei, A.; Chai, Q.H. Water level forecasting using neuro-fuzzy models with local learning. *Neural Comput. Appl.* **2018**, *30*, 1877–1887. [[CrossRef](#)]
43. Talei, A.; Chua, L.H.C.; Quek, C.; Jansson, P.E. Runoff forecasting using a Takagi–Sugeno neuro-fuzzy model with online learning. *J. Hydrol.* **2013**, *488*, 17–32. [[CrossRef](#)]
44. Gergis, J.L.; Fowler, A.M. Classification of synchronous oceanic and atmospheric El Niño–Southern Oscillation (ENSO) events for palaeoclimate reconstruction. *Int. J. Climatol. A* **2005**, *25*, 1541–1565. [[CrossRef](#)]
45. Trenberth, K.E.; Stepaniak, D.P. Indices of el Niño evolution. *J. Clim.* **2001**, *14*, 1697–1701. [[CrossRef](#)]
46. Ropelewski, C.F.; Jones, P.D. An extension of the Tahiti–Darwin southern oscillation index. *Mon. Weather Rev.* **1987**, *115*, 2161–2165. [[CrossRef](#)]
47. Arkin, P.A. The relationship between interannual variability in the 200 mb tropical wind field and the Southern Oscillation. *Mon. Weather Rev.* **1982**, *110*, 1393–1404. [[CrossRef](#)]
48. Curtis, S.; Adler, R. ENSO indices based on patterns of satellite-derived precipitation. *J. Clim.* **2000**, *13*, 2786–2793. [[CrossRef](#)]
49. Gray, W.M.; Sheaffer, J.D.; Knaff, J.A. Influence of the stratospheric QBO on ENSO variability. *J. Meteorol. Soc. Japan. Ser. II* **1992**, *70*, 975–995. [[CrossRef](#)]
50. Tian, E.W.; Su, H.; Tian, B.; Jiang, J.H. Interannual variations of water vapor in the tropical upper troposphere and the lower and middle stratosphere and their connections to ENSO and QBO. *Atmos. Chem. Phys.* **2019**, *19*, 9913–9926. [[CrossRef](#)]
51. Pinault, J.L. Resonance of baroclinic waves in the tropical oceans: The Indian Ocean and the far western Pacific. *Dyn. Atmos. Ocean.* **2020**, *89*, 101119. [[CrossRef](#)]
52. Schneider, N.; Cornuelle, B.D. The forcing of the Pacific decadal oscillation. *J. Clim.* **2005**, *18*, 4355–4373. [[CrossRef](#)]
53. Mantua, N.J.; Hare, S.R.; Zhang, Y.; Wallace, J.M.; Francis, R.C. A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Am. Meteorol. Soc.* **1997**, *78*, 1069–1080. [[CrossRef](#)]
54. Saji, N.; Yamagata, T. Possible impacts of Indian Ocean dipole mode events on global climate. *Clim. Res.* **2003**, *25*, 151–169. [[CrossRef](#)]
55. Smith, C.A.; Sardeshmukh, P.D. The effect of ENSO on the intraseasonal variance of surface temperatures in winter. *Int. J. Climatol. A* **2000**, *20*, 1543–1557. [[CrossRef](#)]
56. Mekanik, F.; Imteaz, M. Variability of cool seasonal rainfall associated with Indo-Pacific climate modes: Case study of Victoria, Australia. *J. Water Clim. Change* **2018**, *9*, 584–597. [[CrossRef](#)]

57. Rayner, N.; Parker, D.E.; Horton, E.; Folland, C.K.; Alexander, L.V.; Rowell, D.; Kent, E.C.; Kaplan, A. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res. Atmos.* **2003**, *108*, 4407. [[CrossRef](#)]
58. Huang, B.; Thorne, P.W.; Banzon, V.F.; Boyer, T.; Chepurin, G.; Lawrimore, J.H.; Menne, M.J.; Smith, T.M.; Vose, R.S.; Zhang, H.-M. Extended reconstructed sea surface temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *J. Clim.* **2017**, *30*, 8179–8205. [[CrossRef](#)]
59. Kobayashi, S.; Ota, Y.; Harada, Y.; Ebita, A.; Moriya, M.; Onoda, H.; Onogi, K.; Kamahori, H.; Kobayashi, C.; Endo, H. The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteorol. Soc. Japan. Ser. II* **2015**, *93*, 5–48. [[CrossRef](#)]
60. Kalnay, E.; Kanamitsu, M.; Kistler, R.; Collins, W.; Deaven, D.; Gandin, L.; Iredell, M.; Saha, S.; White, G.; Woollen, J. The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **1996**, *77*, 437–472. [[CrossRef](#)]
61. Behringer, D.W.; Ji, M.; Leetmaa, A. An improved coupled model for ENSO prediction and implications for ocean initialization. Part I: The ocean data assimilation system. *Mon. Weather Rev.* **1998**, *126*, 1013–1021. [[CrossRef](#)]
62. Kasabov, N.K.; Song, Q. DENFIS: Dynamic evolving neural-fuzzy inference system and its application for time-series prediction. *IEEE Trans. Fuzzy Syst.* **2002**, *10*, 144–154. [[CrossRef](#)]
63. Mamdani, E. Applications of fuzzy algorithms for simple dynamic plant. *Proc. IEEE* **1974**, *121*, 1585–1588. [[CrossRef](#)]
64. Chang, T.K.; Talei, A.; Chua, L.H.; Alaghmand, S. The impact of training data sequence on the performance of neuro-fuzzy rainfall-runoff models with online learning. *Water* **2018**, *11*, 52. [[CrossRef](#)]
65. Kwin, C.T.; Talei, A.; Alaghmand, S.; Chua, L.H. Rainfall-runoff modeling using dynamic evolving neural fuzzy inference system with online learning. *Procedia Eng.* **2016**, *154*, 1103–1109. [[CrossRef](#)]
66. Kao, H.Y.; Yu, J.Y. Contrasting eastern-Pacific and central-Pacific types of ENSO. *J. Clim.* **2009**, *22*, 615–632. [[CrossRef](#)]
67. Duan, W.; Wei, C. The spring predictability barrier for ENSO predictions and its possible mechanism: Results from a fully coupled model. *Int. J. Climatol.* **2013**, *33*, 1280–1292. [[CrossRef](#)]
68. Pinault, J.L. Long wave resonance in tropical oceans and implications on climate: The Pacific Ocean. *Pure Appl. Geophys.* **2016**, *173*, 2119–2145. [[CrossRef](#)]
69. Yu, J.Y.; Kim, S.T. Relationships between extratropical sea level pressure variations and the central Pacific and eastern Pacific types of ENSO. *J. Clim.* **2011**, *24*, 708–720. [[CrossRef](#)]