Atmospheric Research 188 (2017) 100–106

Contents lists available at ScienceDirect

Atmospheric Research

journal homepage: www.elsevier.com/locate/atmosres

Application of recurrent neural networks for drought projections in California

J.A. Le [a], H.M. El-Askary [a,b,c,*], M. Allali [a], D.C. Struppa [a]

[a] Schmid College of Science and Technology, Chapman University, Orange, CA, USA
[b] Center of Excellence in Earth Systems Modeling & Observations, Chapman University, Orange, CA, USA
[c] Department of Environmental Sciences, Faculty of Science, Alexandria University, Moharem Bek, Alexandria, Egypt

ARTICLE INFO

Article history:
Received 21 March 2016
Received in revised form 6 December 2016
Accepted 4 January 2017
Available online 6 January 2017

Keywords:
Drought
Southern California
El Niño
Palmer *Z*-Index
Recurrent neural networks

ABSTRACT

We use recurrent neural networks (RNNs) to investigate the complex interactions between the long-term trend in dryness and a projected, short but intense, period of wetness due to the 2015–2016 El Niño. Although it was forecasted that this El Niño season would bring significant rainfall to the region, our long-term projections of the Palmer Z Index (PZI) showed a continuing drought trend, contrasting with the 1998–1999 El Niño event. RNN training considered PZI data during 1896–2006 that was validated against the 2006–2015 period to evaluate the potential of extreme precipitation forecast. We achieved a statistically significant correlation of 0.610 between forecasted and observed PZI on the validation set for a lead time of 1 month. This gives strong confidence to the forecasted precipitation indicator. The 2015–2016 El Niño season proved to be relatively weak as compared with the 1997–1998, with a peak PZI anomaly of 0.242 standard deviations below historical averages, continuing drought conditions.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

The 2015–2016 winter season witnesses two high-intensity regional/global phenomena namely the drought and El Niño. Beginning with late 2011, California has been facing its most intense and severe drought since historical recordings began in 1895 (Richman and Leslie, 2015). This drought is often compared to other significant droughts in California history, including the particularly long-lasting Dust Bowl drought of the late 1920s to early 1930s and the droughts in 1976–1977 and the late 1980s to early 1990s (Robeson, 2015). Griffin and Anchukaitis (2014) found that although the recent drought is not the longest drought in recorded history, it is the singular most extreme one when comparing rainfall deficits. Recently, over 70% of the state suffered extreme and exceptional drought where normally wet seasons of the yearly climate cycle have been underwhelming (Richman and Leslie, 2015; Robeson, 2015). The uniqueness of this drought season was confirmed by analyzing other drought predictors such as abnormal temperatures (Jeong et al., 2014; Shukla et al., 2015). Howitt et al. (2015) estimated the economic damage to the 2015 agriculture caused by the drought to be $1.8 billion, with a total statewide economic cost for the same period of $2.7 billion. This is a 23% increase compared to the $2.2 billion in losses incurred in 2014 due to similar drought conditions (Howitt et al., 2014). Furthermore, it is also estimated that as many as 21,000 agricultural and related jobs were lost in 2015, up 23% from 17,000 jobs lost in 2014. According to the Center for Watershed Sciences at Davis, an additional $2.8 billion and 21,400 jobs are projected to be lost due to drought in 2016 if conditions persist. Apart from socioeconomic losses, environmental losses are also anticipated. For example, as many as 58 million trees are in severe risk of dying in 2016 causing disastrous impacts on California ecology (Asner et al., 2015). Cook et al. (2015) suggests that the current drought trends could be the beginning of a larger drought taking place over the first half of the entire 21st century. On the other hand, the winter leading into 2016 was expected to bring heavy rain resulting from another powerful weather phenomenon, namely the El Niño Southern Oscillation (ENSO). ENSO affects tropical meteorological fields yet its influence is exerted by changing the largescale Walker circulation and associated convection and precipitation patterns (Slemr et al., 2016). ENSO is a periodic fluctuation in global climate with of a period between 2 and 7 years, and is the strongest throughout the boreal winter season of peak years (Capotondi et al., 2014). The ENSO effect on different regions of the globe is highly varied, but in California, strong El Niño seasons often manifest as periods of extreme and anomalous precipitation (Kahya and Dracup, 1993; El-Askary et al., 2004, 2012). A recent past El Niño event, during the winter of 1997–1998, led to extreme flooding in Los Angeles and caused multiple deaths and billions of dollars in flooding and related damages (Changnon, 1999). The (2015–2016) El Niño season has been projected by many to be abnormally strong (Hoell et al., 2016; Zhenya et al., 2015; Climate.gov, 2015). However, the exact strength of that season is still

* Corresponding author.
E-mail address: elaskary@chapman.edu (H.M. El-Askary).

http://dx.doi.org/10.1016/j.atmosres.2017.01.002
0169-8095/© 2017 Elsevier B.V. All rights reserved.

investigated in the context of the previous strong (1997–1998) El Niño season. It is noteworthy that El Niño drives on average only about 6% of the precipitation variability in California (Savtchenko et al., 2015). It is not guaranteed that El Niño has the potential to resolve the accumulated deficit of precipitation, which is presently equivalent to an entire year of precipitation. However, chances may improve in a strong El Niño that also coincides with the peak of the wet season in California (December–February). Moreover, there is a large geographical difference in the anticipated impact of El Niño on California (Piechota and Dracup 1996; Piechota et al., 1997). Being able to accurately gauge the El Niño intensity is important for agricultural, development, and public safety planning applications. Despite several record examples of strong El Niño seasons, we believe that there are complex interactions of ENSO effects with a drought as intense as the current one affecting California.

It is known that California prolonged drought resulted from a multi-year precipitation decline and anomalous warm temperatures, that in turns resulted from anomalously persisting high pressure in the East Pacific, which significantly changed the normally observed atmospheric circulation patterns. In this research we applied Artificial Neural Networks (ANNs) with recurrent topologies (RNNs), as the more natural topology for exploring time series data, in an attempt to project the local PZI for California Climate Divisions 6 & 7 which are California south coast drainage and southeast desert basins, respectively. Common applications of ANNs involve feed-forward networks trained with backpropagation (Cigizoglou et al., 2004). ANNs have been applied several times to rainfall and precipitation forecasting (Nastos et al., 2014; Bodri et al., 2000; Luck et al., 2000; Silverman et al., 2000; Sakellariou et al., 2004; Cigizoglou et al., 2004; Moustris et al., 2011). Badjate et al. (2009) explored RNNs for predicting precipitation and chaotic time series like sun spot occurrences. Different ANN topologies has been used in order to forecast extreme precipitation events and total precipitation levels, including recurrent Elman Networks (Maqsood et al., 2004) and Long-Term Short-Term Recurrent models (Nastos et al., 2014).

2. Data

This work aims at refining precipitation projections for California Climate Divisions 6 & 7, representing southern California; during the (2015–2016) El Niño season by means of RNNs exploiting the PZI. We are projecting on the regional PZI, which depicts moisture conditions for the current month, then compare the forecasts of the current season with what was observed and forecasted for the 1997–1998 season. Although precipitation data could be considered for our climate divisions, the raw monthly precipitation levels display a strong natural monthly intermittency that is not completely relevant to our study. While this monthly intermittency and variation is quantitatively legitimate, we wish to study and project longer-term trends. Therefore, we chose the PZI as a parameter that is not only more physically meaningful, but also is more resistant to monthly intermittency and spurious variation. We are not substituting PZI, a different drought indicator, for precipitation PZI is a broader hydrologic measure that integrates soil moisture helping in addressing hydrologic drought. We can further add that because of its multivariate origins, PZI is no doubt correlated with the causes of the precipitation anomalies, despite the drought not being the direct physical cause of precipitation anomalies in California.

2.1. Palmer Z-Index

The Palmer Z-Index is a memoryless drought index that measures short-term drought on a monthly scale based on ground data readings. Palmer (1965) established the widely adopted set of Palmer indices for drought analysis in research and industry applications. He proposed a system involving a weighted sum of various calculated and empirical factors, including soil moisture capacity, total precipitation and potential evapotranspiration, moisture recharge, run-off, and moisture loss. From these calculated or measured numbers, a weighted index (motivated by the relative contribution of each factor in drought severity) is calculated with respect to historical averages and behavior of each individual attribute. Higher Z values indicate moist conditions, and lower *Z*-values typically indicate drought conditions of short or long term. The PZI is calculated independently between each month, which makes it a suitable candidate for testing predictive models. The PZI computation begins with a climatic water balance using precipitation and temperature records (Szep et al., 2005). Since precipitation is used in PZI construction, and since their actual values vary together with a correlation of 0.70 and the anomalies with a correlation of 0.77, we believe that PZI can be viewed as a sensitive precipitation indicator reflecting more meaningful physical conditions than raw precipitation. The high correlation shows the general relationship of both parameters, averaging out the intermittent monthly effects in the raw precipitation data.

2.2. Data acquisition

The historical Z-indices for California Climate Divisions 6 and 7 were gathered from the Global Historical Climatology Network (GHCN) *nClimDiv* data set (ftp://ftp.ncdc.noaa.gov/pub/data/cirs/climdiv/), which gives per-climate division aggregates for raw and processed surface data measurements. Values in the *nClimDiv* data set are calculated using area-weighted averages of points on a 5 km-resolution grid overlaid across each division. This resolution is high enough to ensure sufficient spatial sampling, especially for the climate divisions being analyzed in this paper (Vose et al., 2014). Points are assigned climate data based on spatial interpolation of nearby stations, with topographic and network variability taken into account in the interpolation process. GHCN subjects the data to regular quality assurance reviews to ensure correctness. For California South Coast Drainage and South East Desert Basin climate divisions, data from a total of 526 and 184 stations respectively, are taken into account and aggregated. Aggregated station data are available from January 1895 to January 2016, on a month to month basis, giving 1452 total data points. The most recent 120 months of data are set aside for validation, and the remaining 1332 points are used for training purposes. From the *nClimDiv* data set, the gathered historical *Z*-indices were shifted to historical monthly averages and rescaled to have standard deviation equal to unity. This is done to remove yearly periodic components in mean and variance that might arise in the process of the Z-indices computation.

3. Materials and methods

3.1. Artificial neural networks

ANNs are a class of models roughly said to be inspired by "biological neural networks" – that is, the mechanisms and structure of neurons in an animal brain. Models in this family typically involve interconnected nodes, which output signals based on a weighted sum of input signals. Specialized algorithms for training an ANN involve selecting the proper set of weights to model the function or phenomenon in question. These algorithms process training sets of inputs with their known outputs. Different types of ANNs chiefly differ in the topology of the internal node graph. The traditional feed forward neural network can be thought of as a universal function approximator, and is shown to be able to approximate any function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ to arbitrary precision with respect to $L^p(\mu)$ performance criteria (Hornik, 1991). Feed-forward ANNs consist of layers of neurons which receive weighted inputs from the outputs of preceding layers. Feed-forward ANNs have seen great success in fields like pattern recognition. However, because feed-forward ANNs are inherently structured to approximate functions, they struggle in modeling dynamical systems and systems with an inherent temporal component; attempts to do so typically entail a large explosion of parameter. To obviate to this problem, the notion of RNN families was introduced (Hopfield, 1982; Elman, 1993). While traditional feed-forward ANNs

can be fully described as universal approximators of functions, RNN families can be described as universal approximators of dynamical processes, and it has been shown that RNN architectures can approximate arbitrary Turing machines in the same way feed-forward ANNs approximate functions (Hammer, 2000). RNNs add dynamics to traditional ANNs; specifically, they introduce an aspect of statefulness to a neural network: outputs from a given input can be influenced not only by the current input, but also by the residual state left behind from previous computations. This makes RNNs suitable for modeling time series and other such dynamical processes that explicitly depend on history. RNNs of different forms have been used in developing powerful language models (Sutskever et al., 2011) (Graves, 2013) (Mikolov and Zweig, 2012), video classification (Donahue et al., 2014), image captioning (Vinyals et al., 2014), video captioning (Venugopalan et al., 2015), visual question answering (Ren et al., 2015), image generation (Gregor et al., 2015), and drought forecasting (Morid et al., 2007).

3.2. Network and data shape

3.2.1. Network structure

We chose, as model for this study to be a recurrent neural network comprised of two fully-connected recurrent layers with forty nodes per layer. Each internal layer is fully-connected, with all outputs redirected to internal inputs. The output of each node is the result of the Rectified Linear Unit (ReLU) activation function applied to a weighted sum of both inputs from the previous layer and the previous outputs of the layer; these weights are the parameters to be trained for. Fig. 1 shows only a 3 input node architecture as an example for illustration purposes. However, in this work we have used 14 input nodes, where all hidden layers receive input from not only their current input, but also the previous outputs of every other node in the input layer (including itself). Every node receives input from every node in the previous layer and the most recent outputs of the same layer. The network is trained by picking the relative strengths and contributions of each connection (arrow), and also the initial output of the hidden layers.

For a layer with n inputs and m outputs, the output of node $j \in 1 \ldots m$ at time $t+1$, from an input vector $\mathbf{x}$, is:

$$y_j(t+1) = f\left(\sum_i^n w_{ij} x_j + \sum_s^m v_{sj} y_s(t)\right)$$

where $f(x)$, the Rectified Linear Unit activation function, is

$$f(x) = \begin{cases} 0, & x \le 0 \\ x, & x > 0 \end{cases}$$

w_{ij} is the matrix of weights of influences from the previous layer, and v_{sj} is the matrix of weights of influences from the previous activations of other nodes in that layer. By providing a non-linear activation function, we allow our network to exhibit non-linear behavior. Both weight matrices are trained parameters of the model, and $\boldsymbol{y}(t=0)$, the initial state of the network, is also a trained parameter; it is trained to create an adaptable initial condition from which prediction begins. For this network, the final output layer is a traditional feed-forward layer (that is, $\forall s,j \cdot v_{sj} = 0$) with the linear activation function $g(x) = x$.

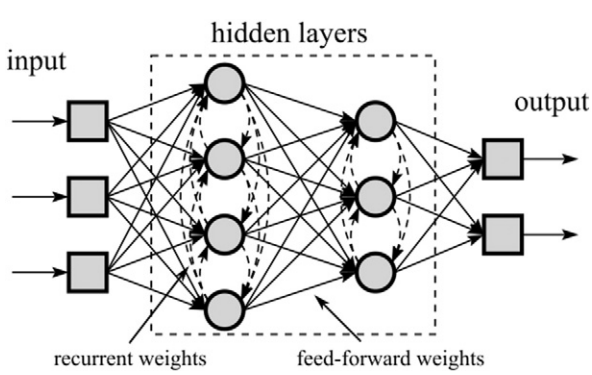

Fig. 1. Internal structure of the recurrent neural network topology with multiple hidden layers.

3.2.2. Long term projections

PZI data for each month is represented as a 14-element vector. The first twelve elements are indicator elements, representing the month of the data point where each component is either 0 or 1. The final two elements are the normalized, scaled PZIs from that month for both climate divisions. For training, data vectors are grouped together in contiguous samples of 48 months. Each 48-month group is paired with a single two element vector representing the two divisions' PZI data for the month right after the month of the final data vector in the group. Thus, the network used has 14 input nodes and 2 output nodes. In order to project several months into the future, the projection for the next month is joined together with a 12-element prefix indicating the next month and used to predict the month after. This allows the network to step forward several months into the future, despite its ability in providing projections for only the immediate following month. This technique uses the RNN as a directed continuous state non-Markov feedback generator resembling a continuous space sequence memorizer (Wood et al., 2009), in a similar manner as strategy explored extensively by Graves, et al. (2013) to generate curves and paths from training data. Hopfield (1982) explored the convergence of steady-states of this process.

3.3. Network and training methodology

To train the neural network, we implemented the well-known backpropagation through time (BPTT) algorithm in the Haskell programming language, training over contiguous samples of history (Mozer, 1995; Werbos, 1990). This approach gives the network the ability to be trained to act on influences of up to 4 years into the past (though short-term influences will be much stronger). BPTT is a gradient descent algorithm that optimizes the weight space by calculating the gradient of the error between model output and training data with respect to variations in weights and moving in the direction of negative gradient. By adjusting the step size, it is possible to tune our training process to be either quickly converging or potentially step past important local minima. For the model described in this paper, we chose a step size of 2×10^{-3}. To train on a contiguous sample of history, we ran the network over the entire length of the sample and we took the error value as the sum of the squared differences between outputs of the final network and known data for the next month into the future. From an initial network with randomly generated weights and $\boldsymbol{y}(t=0)$, the network is trained over a shuffled collection of contiguous samples. Each full pass over the training set is known as an epoch. Accuracy of the model is run against the 150-point validation set of the most recent 150 months of weather data.

3.3.1. Mitigating over-fitting

Through initial training runs, one sees that the network converges extremely quickly (within the two hundred epochs, typically), due to the small size of the network. The limit values tended to show a high correlation with training data set (typically, between 0.6 and 0.7 correlation). However, these limits were unsatisfactory when run against the validation set. This suggests that over-fitting is the main source of model failure. In over-fitting the data, the model fits its parameters to the noise of the training data, and not on the signal itself. In order to combat over-fitting, we utilized standard techniques, including the Gaussian noise injection into the training data (Zur et al., 2009), changing the choice of activation function from the sigmoidal logistic function to ReLU (Glorot et al., 2011; LeCun et al., 2015) and applying the dropout technique (Srivastava et al., 2014). With dropout, it is possible to effectively

train what are essentially several semi-dependent networks, but on top of a single concise network. The aggregation of these results is taken to be the final prediction of the model.

4. Results

4.1. Validation

With these improvements, the networks converge within one or two hundred epochs to consistently similar trained states. We also reduced the difference between the training and validation data correlations and achieved consistent results regardless of initial starting state. Despite underestimation in rare extreme cases, the network's output agrees strongly with observed data. Longer historical datasets are necessary for a better training, as the 1895–2005 historical record includes only a few El Niño events to analyze. Although we acknowledge the limitations of our effort, we consider the extracted results of the performed methodology quite satisfactory in forecasting such extreme values for the next year, based only on historical PZI time series. Fig. 2 shows the one, two, and three-month forward projections for the 2006–2015 validation set alongside observed values and the corresponding correlation plots. The black dashed lines correspond to perfect fit ($y = x$), while the red solid lines to the least-square fit. The total variation in y during the examined period for the 1 month ahead step is explained by the linear relationship between x and y represented by $y = 0.697x - 0.510$, (x) represents the observed PZI values and (y) represents the forecasted values. Examining the decline in predictive power as the network projects further into the future, we find that the model has statistically significant predictive power with correlation coefficients ranging from 0.610 to 0.434 for between one to three months ahead, validating it for confident predictions. It is noteworthy that for longer-term predictions, the developed RNN tends to overestimate, though in the correct direction of variance. Moreover, it is notable that the network validates better as it progresses down the timeline, using the first yew years to develop its internal state to gauge the current context.

For our trained networks we found that, after a period of time on the order of one year, the state from previous months are forgotten and the network settles into a periodic steady-state feedback cycle. However, the projected data can still be analyzed for results until the time when the residual influences of the past fade away. Furthermore, analysis of activation profiles of internal nodes for RNNs can be shown to yield physically insightful results which will be discussed in a future work.

4.2. Medium term predictions

Running the forward prediction method from the PZI records leading to January 2016, we found that PZI of the following month represents a drier than average record, despite the anticipated the wet late fall and early winter season (Hoell et al., 2016; Zhenya et al., 2015; Climate.gov, 2015), hence we expect a return to drought conditions. Fig. 3a shows a solid comparison between the powerful 1997–1998 El Niño season used here as a baseline when comparing with the 2015–2016 El Niño season. Fig. 3b shows the actual observed precipitation levels for the current season compared to precipitation for the 1997–

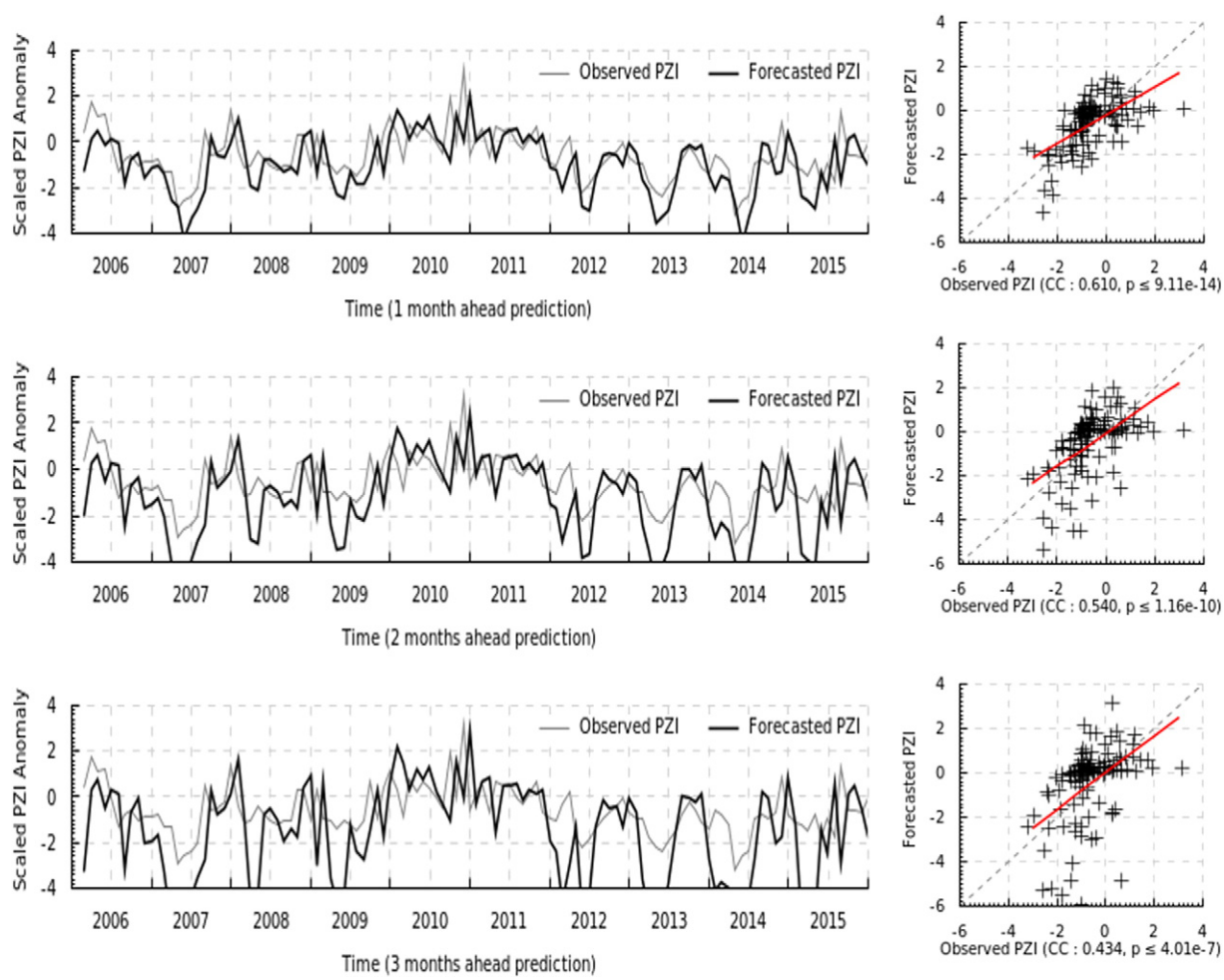

Fig. 2. Time series correlation plot between observed and forecasted PZI using RNN for climate divisions 6 & 7 with lead times of 1 month (top), 2 months (middle), 3 months (bottom), starting January 2006.

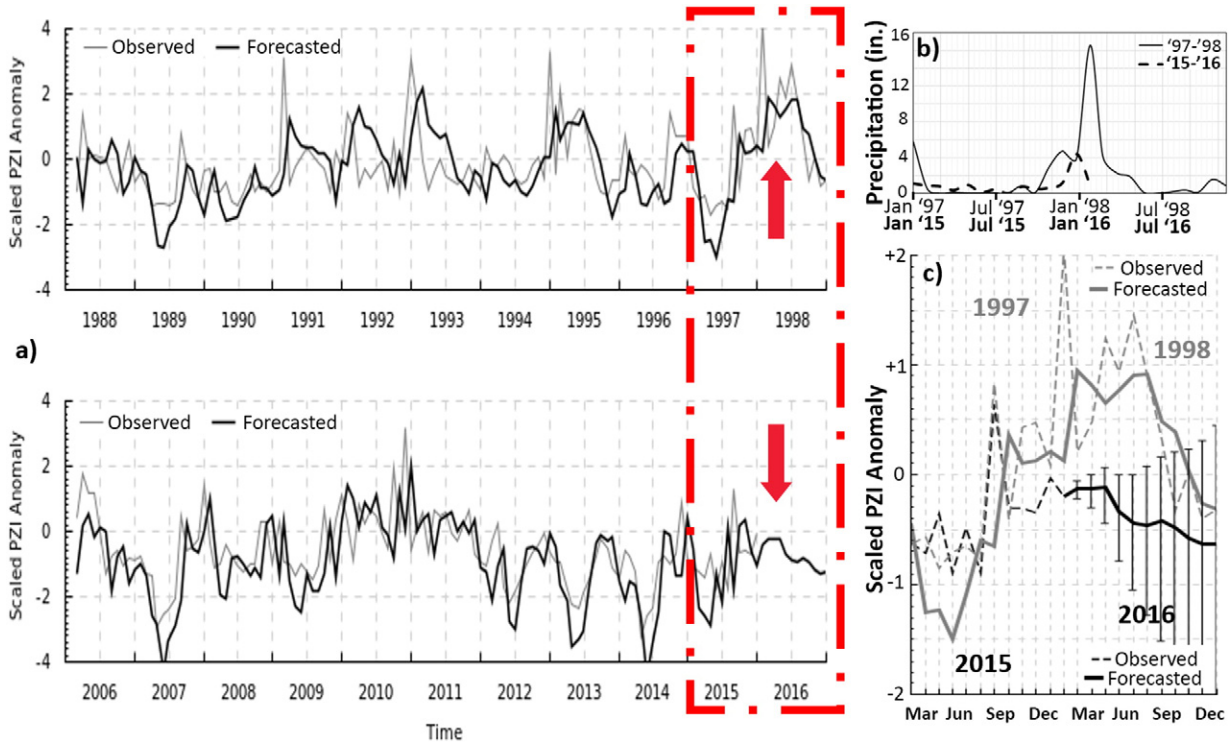

Fig. 3. (a) Comparing observed and forecasted PZI data for the two El Niño seasons in question (b) Observed precipitation totals for California Climate Division 6 for the two El Niño season in question, confirming the low-precipitation season that the model predicts (c) Detailed look at model outputs projections for PZI for the year 2016 compared to observed values and predictions for 1998, with model uncertainties.

1998 El Niño season which confirms the validity of our model's projections of a drier season associated with the 2015–2016 El Niño as compared to the 1997–1998 one.

The thick lines represent the model's direct output while the light lines show a measure of the model's uncertainty, calculated via a Monte Carlo process simulating stochastic noise in subsequent prediction steps to account for potential errors in the model (Fig. 3c). The grey dashed line is the baseline 1997–1998 El Niño season, superimposed over their respective months in the 2015–2016 season presented by the dark dashed lines. The PZI anomaly peaks at 0.242 standard deviations below the monthly average in May of 2016, and quickly sinks back to 0.924 standard deviations below monthly average by August of the same year. In total, 2016 will be a drier year than average with a −0.715 PZI anomaly for the California South Coast Drainage climate division.

The projected PZI for 2015–2016, indicates a much weaker season as compared to the 1997–1998. The data points themselves are contrasted with that several baseline El Niño seasons in Table 1. This season can be contrasted with the baseline 1997–1998 El Niño season, which saw a February 1998 that was 4.13 standard deviations above the average PZI for the month, and a 1998 that was 1.1 standard deviations higher than that of the average year. It can also be contrasted with the 1982–1983 El Niño season, which saw a peak anomaly of 2.22 standard deviations above the monthly average in April of 1983, and saw a 1983 that was 1.15 standard deviations above the annual average. With confidence, it can be concluded that the 2015–2016 season proves to be underwhelming in precipitation, and that drought conditions will persist past this winter season. The worst of the drought has apparently passed since 2013 with an annual anomaly of −1.30 and 2014 with −1.17 as compared to 2015 with −0.85, with a projected −0.715 annual anomaly for 2016, continuing a general trend of slow but steady emergence from the current drought season. While immediate predictions towards values one month into the future have strong predictive power, it cannot be assumed that longer term forward projections maintain the same predictive power.

Table 1
Historical El Niño Palmer Z Index levels (Anomalies in standard deviations).

Season	Peak anomaly	Peak anomaly month	Annual anomaly
1957–1958	3.03	April	0.5
1982–1983	2.22	April	1.15
1997–1998	**4.13**	**February**	**1.1**
2009–2010	1.1	January	0.45
2015–2016	**−0.242**	**May**	**−0.715** (less confident)

To clearly show the skill of our proposed method, the correlation between the observed and forecasted PZI anomaly, for past known weak to severe El Niño responses according to NOAA, is computed and presented in Table 2. P-values are also provided corresponding to the likelihood of the null hypothesis that the observed and the forecasted PZI are uncorrelated for the entire period under investigation. In other words very low P-values show that the correlation between forecasted and observed PZI is statistically significant. The model was intentionally trained on the entire history (all years), rather than, on known El Niño years for the purposes of drought projection. The main motivation of doing so is to be able to provide the model with more information on the behavior of the system in all situations. Having the model trained only on El Niño years, it would not be able to observe and learn from recurring phenomenon that do not normally occur on El Niño years. Training only on El Niño years would arbitrarily deny the model the chance to learn from these phenomenon, which might become significant in the specific season we are attempting to study. Moreover, by training on non- El Niño seasons, the model has the opportunity to distinguish between El Niño and Non-El Niño seasons and learn the degree of adjustment required in the context of the years leading into each season.

From the above table it is clear that model showed some skill towards the majority of El Niño years regardless of their strength. The correlation coefficients varied around 0.7 which is quite similar and slightly

Table 2
Observed (obs.) versus projected (proj.) PZI correlation coefficients with corresponding P-value for past El Niño events categorized as weak, moderate, strong and very strong showing the RNN model skill.

Weak	CC (obs. vs. proj.)/P value	Moderate	CC (obs. vs. proj.)/P value	Strong	CC (obs. vs. proj.)/P value	Very strong	CC (obs. vs. proj.)/P value
1953–54	0.608/0.0179	1951–52	0.822/0.0005	1957–58	0.837/0.0003	1982–83	0.623/0.0156
1958–59	0.837/0.0003	1963–64	0.892/0.0000	1965–66	0.869/0.0001	1997–98	0.870/0.0001
1968–69	0.556/0.0302	1986–87	0.873/0.0001	1972–73	0.694/0.0061		
1969–70	0.575/0.0253	1991–92	0.332/0.1459				
1976–77	0.483/0.0559	2002–03	0.892/0.0000				
1979–80	0.688/0.0067						
1994–95	0.584/0.0232						
2004–05	0.937/0.0000						
2006–07	0.880/0.0000						

higher than our validation data that was presented in Fig. 2. We did not include 2009–2010, 2015–2016 in this analysis as they are part of the validation dataset used in our forecast model.

5. Conclusions

This paper addressed the rationale of using PZI as a significant precipitation indictor to address the anticipated heavy rain over Southern California driven by the strong 2015–2016 El Niño season. By investigating many ANN models, utilizing proven effective RNN configurations and applying them to analyze over a century of monthly PZI data, it is shown with strong confidence that precipitation associated with the 2015–2016 El Niño season is currently and will continue to be weaker than that of the historic 1997–1998 El Niño season. From this, we anticipate that drought conditions will continue to persist (albeit at an alleviated level) beyond this winter. These forecasts are made with a model that is well tested with significant high correlations on a ten-year validation set, with p-values $< 10^{-6}$ for predictions up to three months into the future. Such projections are confirmed through current observed precipitation levels and PZI values for 2016 as compared to those of the 1997–1998 season.

Acknowledgments

The authors would like also to acknowledge the use of the Samueli Laboratory in Computational Sciences in the Schmid College of Science and Technology, Chapman University in data processing and analysis as well as the generous student support from Experian for the graduate program in Computational and Data Sciences. Our gratitude is also extended to the Kay Family Foundation Data Analytics Grant. The authors would like also to extend their appreciation to Prof. Thomas Piechota, vice president of research at Chapman University and to anonymous reviewers for the very helpful comments during the review process. The data used here were gathered from the Global Historical Climatology Network (GHCN) *nClimDiv* data set (ftp://ftp.ncdc.noaa.gov/pub/data/cirs/climdiv/).

References

Asner, G., Brodrick, P.G., Anderson, C.B., Vaughn, N., Knapp, D.E., Martin, R.E., 2015. Progressive forest canopy water loss during the 2012–2015 California drought. Proc. Natl. Acad. Sci. 2015, 249–255.
Badjate, S.L., Dudul, S.V., 2009. Multi step ahead prediction of north and south hemisphere sun spots chaotic time series using focused time lagged recurrent neural network model. WSEAS Trans. Inf. Sci. Appl. 6 (4), 684–693.
Bodri, L., Cermak, V., 2000. Prediction of extreme precipitation using a neural network; application to summer flood occurrence in Moravia. Adv. Eng. Softw. 31, 211–221.
Capotondi, A., Wittenberg, A., Newman, M., 2014. Understanding ENSO diversity. Bull. Am. Meteorol. Soc. 96, 921–938.
Changnon, S.A., 1999. Impacts of 1997–98 el Niño generated weather in the United States. Bull. Am. Meteorol. Soc. 80 (9):1819–1827. http://dx.doi.org/10.1175/1520-0477(1999)080<1819:IOENOG>2.0.CO;2.
Cigizoglou, H.K., Alp, M., 2004. Rainfall-Runoff Modelling Using three Neural Network Methods, Artificial Intelligence and Soft Computing - ICAISC 2004: 7th International Conference. pp. 166–171.
Climate.gov, October 2015. What to Expect this Winter: NOAA's Outlook Reveals What Conditions Are Favored across the US. (Available at: https://www.climate.gov/news-features/blogs/enso/what-expect-winter-noaa%E2%80%99s-outlook-reveals-what-conditions-are-favored).
Cook, B.I., Ault, T.R., Smerdon, J.E., 2015. Unprecedented 21st century drought risk in the American southwest and central plains. Sci. Adv. 1 (1). http://dx.doi.org/10.1126/sciadv.1400082.
Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T., 2014. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. pp. 647–655.
El-Askary, H., Sarkar, S., Chiu, L., Kafatos, M., El-Gahzawi, T., 2004. Rain gauge derived precipitation variability over Virginia and its relation with the el Niño southern oscillation (ENSO). Adv. Space Res. 33 (2004):338–342. http://dx.doi.org/10.1016/S0273-1177(03)00478-2.
El-Askary, H., Allali, M., Rakovski, C., Prasad, A., Kafatos, M., Struppa, D., 2012. Computational methods for climate data. WIREs Comput. Stat. 4:359–374. http://dx.doi.org/10.1002/wics.1213.
Elman, J.L., 1993. Learning and development in neural networks: the importance of starting small. Cognition 48 (1), 71–99.
Glorot, X., Bordes, A., Bengio, Y., 2011. Deep Sparse Rectifier Neural Networks. Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS-11). vol. 15, pp. 315–323.
Graves, A., 2013. Generating Sequences With Recurrent Neural Networks CoRR, abs/1308.0850.
Gregor, K., Danihelka, I., Graves, A., Wierstra, D., 2015. DRAW: A Recurrent Neural Network For Image Generation CoRR, abs/1502.04623.
Griffin, D., Anchukaitis, K.J., 2014. How unusual is the 2012–2014 California drought? Geophys. Res. Lett. 41, 9017–9023.
Hammer, B., 2000. On the approximation capability of recurrent neural networks. Neurocomputing 31 (1–4):107–123. http://dx.doi.org/10.1016/S0925-2312(99)00174-5.
Hoell, A., Hoerling, M., Eischeid, J., Wolter, K., Dole, R., Perlwitz, J., Xu, T., Cheng, L., 2016. Does el Niño intensity matter for California precipitation. Geophys. Res. Lett. 43: 819–825. http://dx.doi.org/10.1002/2015GL067102.
Hopfield, J.J., 1982. Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. 79 (8), 2554–2558.
Hornik, K., 1991. Approximation capabilities of multilayer feedforward networks. Neural Netw. 4 (2):251–257. http://dx.doi.org/10.1016/0893-6080(91)90009-T.
Howitt, R., Macewan, D., Medellin-azuara, J., 2014. Economic Analysis of the 2014 Drought for California Agriculture. Center for Watershed Sciences, University of California, Davis, California.
Howitt, R., Macewan, D., Medellin-azuara, J., 2015. Economic Analysis of the 2015 Drought for California Agriculture. Center for Watershed Sciences, University of California, Davis, California.
Jeong, D.I., Sushami, L., Khaliq, M.N., 2014. The role of temperature in drought projections over North America. Clim. Chang. 127.
Kahya, E., Dracup, J.A., 1993. U.S. streamflow patterns in relation to the el Niño southern oscillation. Water Resour. Res. 29, 2491–2503.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.
Luck, K.C., Ball, J.E., Sharma, A., 2000. A study of optimal model lag and spatial inputs to artificial neural networks for rainfal forecasting. J. Hydrol. 227 (1–4), 141–150.
Maqsood, I., Khan, M.R., Abraham, A., 2004. An ensemble of neural networks for weather forecasting. Neural Comput. & Applic. 13:112–122. http://dx.doi.org/10.1007/s00521-004-0413-4.
Mikolov, T., Zweig, G., 2012. Context dependent recurrent neural network language model, SLT. pp. 234–239.
Morid, S., Smakhtin, V., Bagherzadeh, K., 2007. Drought forecasting using artificial neural networks and time series of drought indices. Int. J. Climatol. 27 (15):2103–2111. http://dx.doi.org/10.1002/joc.1498.
Moustris, K.P., Larissi, I.K., Nastos, P.T., Paliatsos, A.G., 2011. Precipitation forecast using artificial neural networks in specific regions of Greece. Water Resour. Manag. 25, 1979–1993.
Mozer, M.C., 1995. A Focused Backpropagation Algorithm for Temporal Pattern Recognition. Backpropagation: Theory, Architectures, and Applications, pp. 137–169.
Nastos, P., Pallastos, A., Koukouletsos, K., Larissi, I.K., Moustris, K.P., 2014. Artificial neural networks modeling for forecasting and the maximum daily total precipitation at Athens, Greece. Atmos. Res. 144, 141–150.
Palmer, W.C., 1965. In U.S. weather bureau. Res. Pap. 45.
Piechota, T.C., Dracup, J.A., 1996. Drought and regional hydrologic variation in the United States: associations with the el Niño-southern oscillation. Water Resour. Res. 32 (5), 1359–1373.

Piechota, T.C., Dracup, J.A., Fovell, R.G., 1997. Western US streamflow and atmospheric circulation patterns during el Niño-southern oscillation. J. Hydrol. 201 (1):249–271. http://dx.doi.org/10.1016/S0022-1694(97)00043-7 (http://www.sciencedirect.com/science/article/pii/S0022169497000437 ISSN 0022-1694).

Ren, M., Kiros, R., Zemel, R.S., 2015. Image Question Answering: A Visual Semantic Embedding Model and a New Dataset CoRR, abs/1505.02074.

Richman, M., Leslie, L., 2015. Uniqueness and causes of the California drought. Procedia Comput. Sci. 6, 428–435.

Robeson, S.M., 2015. Revisiting the recent California drought as an extreme value. Geophys. Res. Lett. 42 (16), 6771–6779.

Sakellariou, N.K., Kambezidis, H.D., 2004. Prediction of the total rainfall amount during august and November in the Athens area, Greece. Fresenius Environ. Bull. 13 (3), 289–292.

Savtchenko, A.K., Huffman, G., Vollmer, B., 2015. Assessment of precipitation anomalies in California using TRMM and MERRA data. J. Geophys. Res. Atmos. 120 (16): 8206–8215. http://dx.doi.org/10.1002/2015JD023573.

Shukla, S., Safeeq, M., AghaKouchak, A., Guan, K., Funk, C., 2015. Temperature impacts on the water year 2014 drought in California. Geophys. Res. Lett. 42 (11).

Silverman, D., Dracup, J.A., 2000. Artificial neural networks and long-range precipitation predictions in California. J. Appl. Meteorol. 39 (1), 57–66.

Slemr, F., Brenninkmeijer, C.A., Rauthe-Schöch, A., Weigelt, A., Ebinghaus, R., Brunke, E.-G., Martin, L., Spain, T.G., O'Doherty, S., 2016. El Niño–southern oscillation influence on tropospheric mercury concentrations. Geophys. Res. Lett. 43. http://dx.doi.org/10.1002/2016GL067949.

Srivastava, N., Hinton, G., Krizhevsky, A., 2014. Dropout – a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.

Sutskever, I., Martens, J., Hinton, G.E., 2011. Generating text with recurrent neural networks, In Proceedings of the 28th International Conference on Machine Learning (ICML-11). pp. 1017–1024.

Szép, I.J., Mika, J., Dunkel, Z., 2005. Palmer drought severity index as soil moisture indicator: physical interpretation, statistical behaviour and relation to global climate. Phys. Chem. Earth, Parts A/B/C 30 (1–3), 231–243 (ISSN 1474-7065).

Venugopalan, S., Rohrbach, M., Donahue, J., Mooney, R.J., Darrell, T., Saenko, K., 2015. Sequence to Sequence - Video to Text CoRR, abs/1505.00487.

Vinyals, O., Toshev, A., Bengio, S., Erhan, D., 2014. Show and Tell: A Neural Image Caption Generator CoRR, abs/1411.4555.

Vose, R.S., Applequist, S., Squires, M., Durre, I., Menne, M.J., Williams Jr., C.N., Fenimore, C., Gleason, K., Arndt, D., 2014. Improved historical temperature and precipitation time series for U.S. climate divisions. J. Appl. Meteorol. Climatol. 53, 1232–1251.

Werbos, P.J., 1990. Backpropagation through time: what it does and how to do it. Proc. IEEE 78 (10):1550–1560. http://dx.doi.org/10.1109/5.58337.

Wood, F., Archambeau, C., Gasthaus, J., James, L., Teh, Y.W., 2009. A Stochastic Memoizer for Sequence Data. Proceedings of the 26th Annual International Conference on Machine Learning, pp. 1129–1136.

Zhenya, S., Qi, S., Ying, B., Xunqiang, Y., Fangli, Q., 2015. The prediction on the 2015/16 el Niño event from the perspective of FIO-ESM. Acta Oceanol. Sin. 34 (12):67–71. http://dx.doi.org/10.1007/s13131-015-0787-4.

Zur, R.M., Jiang, Y., Pesce, L.L., Drukker, K., 2009. Noise injection for training artificial neural networks: a comparison with weight decay and early stopping. Med. Phys. 36 (10):4810–4818. http://dx.doi.org/10.1118/1.3213517.