Journal of Environmental Management 283 (2021) 111979

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: http://www.elsevier.com/locate/jenvman

Research article

An improved SPEI drought forecasting approach using the long short-term memory neural network

Abhirup Dikshit [a], Biswajeet Pradhan [a,b,c,d,*], Alfredo Huete [a,e]

[a] Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), University of Technology Sydney, NSW, 2007, Australia
[b] Department of Energy and Mineral Resources Engineering, Sejong University, Choongmu-gwan, 209 Neungdong-ro, Gwangjin-gu, Seoul, 05006, South Korea
[c] Center of Excellence for Climate Change Research, King Abdulaziz University, P. O. Box 80234, Jeddah, 21589, Saudi Arabia
[d] Earth Observation Center, Institute of Climate Change, Universiti Kebangsaan Malaysia, 43600, UKM, Bangi, Selangor, Malaysia
[e] School of Life Sciences, Faculty of Science, University of Technology Sydney, Sydney, NSW, 2007, Australia

ARTICLE INFO

Keywords:
Standard precipitation evaporation index
New South Wales
Drought forecasting
Deep learning

ABSTRACT

Droughts are slow-moving natural hazards that gradually spread over large areas and capable of extending to continental scales, leading to severe socio-economic damage. A key challenge is developing accurate drought forecast model and understanding a models' capability to examine different drought characteristics. Traditionally, forecasting techniques have used various time-series approaches and machine learning models. However, the use of deep learning methods have not been tested extensively despite its potential to improve our understanding of drought characteristics. The present study uses a deep learning approach, specifically the Long Short-Term Memory (LSTM) to predict a commonly used drought measure, the Standard Precipitation Evaporation Index (SPEI) at two different time scales (SPEI 1, SPEI 3). The model was compared with other common machine learning method, Random Forests, Artificial Neural Networks and applied over the New South Wales (NSW) region of Australia, using hydro-meteorological variables as predictors. The drought index and predictor data were collected from the Climatic Research Unit (CRU) dataset spanning from 1901 to 2018. We analysed the LSTM forecasted results in terms of several drought characteristics (drought intensity, drought category, or spatial variation) to better understand how drought forecasting was improved. Evaluation of the drought intensity forecasting capabilities of the model were based on three different statistical metrics, Coefficient of Determination (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). The model achieved R^2 value of more than 0.99 for both SPEI 1 and SPEI 3 cases. The variation in drought category forecasted results were studied using a multi-class Receiver Operating Characteristic based Area under Curves (ROC-AUC) approach. The analysis revealed an AUC value of 0.83 and 0.82 for SPEI 1 and SPEI 3 respectively. The spatial variation between observed and forecasted values were analysed for the summer months of 2016–2018. The findings from the study show an improvement relative to machine learning models for a lead time of 1 month in terms of different drought characteristics. The results from this work can be used for drought mitigation purposes and different models need to be tested to further enhance our capabilities.

1. Introduction

Droughts are one of the most devastating natural hazards affecting various parts of the world. The phenomenon starts with the deficiency in rainfall, and it affects various aspects like stream-flow and soil moisture. The factors affecting or leading to droughts can be several ranging from meteorological parameters to climatic factors and even the effect of anthropogenic activities (Van Loon et al., 2016). Droughts can be broadly categorized as meteorological, which means the scarcity of rainfall beneath a certain truncation level; hydrological which refers to reduction in stream-flow; agricultural which leads to reduction in soil moisture content and ultimately crop yield; and socio-economic drought; which is the economic hardship faced by the people as a combination of all the above drought types. However, of late, several researchers have attempted to further categorize drought types, with Mishra and Singh (2010) suggesting to add ground water drought;

* Corresponding author. Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), University of Technology Sydney, NSW, 2007, Australia.
E-mail addresses: abhirup.dikshit@student.uts.edu.au (A. Dikshit), biswajeet.pradhan@uts.edu.au, Biswajeet24@gmail.com (B. Pradhan), alfredo.huete@uts.edu.au (A. Huete).

https://doi.org/10.1016/j.jenvman.2021.111979
Received 20 August 2020; Received in revised form 3 December 2020; Accepted 9 January 2021
Available online 19 January 2021

Vicente Serrano et al. (2020)0 suggesting environmental droughts as another category and a few suggesting to add ecological drought as a separate drought type (Slette et al., 2019). Among all these category types, one thing is for certain: that droughts are very complex and an international consensus on drought types is necessary.

Considering forecasting droughts, there are broadly three steps: i) defining a drought; ii) input data; and iii) models used. In terms of defining a drought, researchers have come up with several indices for different purposes which helps to understand various drought characteristics like onset, end, duration and intensity. These indices depend on the parameters being considered which could be a meteorological type, like Standard Precipitation Index (SPI) (McKee et al., 1993), derived from precipitation and typically used for meteorological droughts or Standard Precipitation Evaporation Index (SPEI) (Vicente Serrano et al. 2010, 2012), derived from precipitation and temperature and can be used for meteorological and/or hydrological droughts. Other drought indices type include derived from remote sensing products like Soil Adjusted Vegetation Index (SAVI) (Huete, 1988), which can be used to understand the vegetation aspects. Recently, Yihdego et al. (2019) presented a comprehensive list of the various drought indices used in the literature. As there have been over 150 drought indices developed, validating everyone and developing a common consensus is not feasible. However, lately there seems to be a growing consensus about the use of SPEI, primarily because of its use of rainfall and temperature while determining index and not only rainfall as is the case in SPI. Therefore, SPEI was used in the present study due to its growing acceptability and its ability to use both rainfall and temperature parameters for calculation. The values are further categorized into different levels of drought or non-drought conditions, which can be perceived as a reflection of the actual conditions (Hao et al., 2016).

Several researchers have attempted to predict droughts at different lead times, with the aim of increasing the forecasting capability at higher lead times. Another important aspect in drought studies is the input data and the variables being considered for any analysis type. The input data can be generally classified as ground-based or satellite based. The ground-based data could be further categorized as station-based and interpolated grids (Sun et al., 2018). The interpolated gridded datasets are based on gathering site-specific data from across the world and apply different interpolation grids to produce a global/continental scale map of certain drought affecting variables (Sun et al., 2018). Such datasets have the benefit of higher temporal resolution, which is crucial for drought studies. On the other hand, remote sensing based datasets suffer from lower temporal resolution, which are not ideal for forming a robust architecture for drought forecasting studies (Hao et al., 2018). Table 1 provides the advantages and limitations of the data types, along with suggested review articles of the various interpolated grids and remote sensing datasets.

Using the climatological dataset for drought studies has seen a rise as more and more datasets are being made available. Sun et al. (2018) reviewed 30 global precipitation datasets developed using various approaches and found that the use of any dataset depends on the study area and the type of study being performed. As precipitation is one of the key drought influencing factors and the present study is being conducted on a considerably large area, we have used one of the most popular and well-accepted global climatological dataset, namely Climate Research Unit (CRU TS v 4.03) spanning from 1901 to 2018 at a spatial resolution of 0.5° (Harris et al. 2019, 2020). Further, the usability of CRU dataset for drought studies has been explored by several studies, like Vicente-Serrano et al. (2012) compared the performance of four different drought indices using CRU at a global scale; Spinoni et al. (2019) used CRU to prepare a global database of meteorological droughts. For NSW, the viability of the dataset for drought studies has been examined and verified by Dikshit et al. (2020a). The use of variables to accurately forecast meteorological drought has seen utilization of several factors like temperature, evapotranspiration and other factors like sea surface temperatures and climatic indices. The inclusion of climatic variables have shown to improve the forecasting results at higher lead times (6–12 months) (Özger et al., 2012), however, as the present work forecasts only at a lead time of 1 month, the use of local climate variables can be considered sufficient as highlighted in previous works (Mishra and Singh 2011; Hao et al., 2018). The meteorological variables available from CRU dataset are precipitation, vapour pressure, cloud cover, potential evapotranspiration and temperature (mean, minimum and maximum).

And the final step depends on the forecasting model and the drought characteristics being analysed. The approaches used for drought forecasting can be classified into statistical (Deo and Sahin, 2015), physical (Hao et al., 2018) and hybrid (Wang et al., 2018) models. Statistical models analyses relationships among historical records, by considering various influencing factors as predictors. Physical based models involve the use of General Circulation Models (GCMs), which considers the physical processes between the atmosphere and land surface. Hybrid models involve the combination of both statistical and physical based models. In case of statistical models, various techniques like regression, time series analysis and machine learning approaches are used. The use of machine learning approaches, specifically, Artificial Neural Networks (ANN) have seen a rise, primarily due to the non-linear behaviour of droughts (Mishra and Singh, 2011). However, neural networks are incapable of dealing with non-stationarities in drought estimations and suffer from overfitting due to lag components involved in time series data (Alizadeh and Nikoo, 2018).Readers are referred to Mishra and Singh (2011); Hao et al. (2018) and Fung et al. (2019) for a more detailed understanding of the various approaches used for forecasting purposes, along with their advantages and limitations. Therefore, there is a growing consensus to improve the forecasting abilities and one way to achieve it is by the use of deep neural networks which has shown tremendous capabilities to outperform the traditional approaches. Various fields like speech recognition (Hinton et al., 2012), self-driving vehicles (Farabet et al., 2012), computer vision (Krizhevsky et al., 2012) and natural language processing (Collobert et al., 2011) have immensely benefitted. In the case of droughts, deep learning has been used to forecast sea surface temperatures using hybrid Long Short-Term Memory (LSTM) technique (Xiao et al., 2019); developing drought monitoring tool using deep feed forward neural network (Shen et al., 2019); drought index forecasting (SPI, SPEI) using LSTM (Poornima and Pushpalata, 2019). The study by Poornima and Pushplata (2019) used a

Table 1
Advantages and limitations of the data types.

Data Type	Advantages	Limitations	Popular products	Review Article
Ground based	- Long time series data - Good for analysing drought sensitive regions - Can cover large areas (continental to global scale)	- Manual Errors - Need to check the homogeneity and fill the missing values in case of site-specific data - Depends on the interpolation techniques used	- Meteorological Stations - Interpolated Grids (CRU, Global Precipitation Climatology Centre (GPCC))	Sun et al. (2018)
Remote Sensing	- Information of vegetation data - Closely monitor the changes in agriculture	- Not available for enough duration - Low temporal resolution	MODIS, Sentinel −1, 2, 3	West et al. (2020)

variety of local climatic variables to forecast SPEI and SPI for a temporal range of 1980–2013 for a single site in Hyderabad, India. The model was able to achieve accuracy of 97% and 99% for monthly SPEI and SPI, respectively. Their study used ground-based data and the temporal range used was 1980–2013. The use of relatively shorter time scale may hinder to adequately capture patterns and depict more reliable results. Therefore, the present study aims to use LSTM technique using global climatological dataset to understand the forecasting capabilities in terms of SPEI values and analysing the variation in terms of drought categories and spatial variation. The model was trained from 1901 to 2010 providing it with sufficient data to learn the relationship between drought index and the causative factors. Next, the model was validated from 2011 to 2018 and the results were analysed at 1 month lead time. In summary, the present work aims to achieve three main objectives: i) How well can deep learning models forecast meteorological drought?; ii) Understanding the spatial variation between observed and predicted values; and iii) Examining the variation in terms of drought categories as defined by index values. The flowchart of the study is illustrated in Fig. 1.

2. Study area

The area of interest for the present work is New South Wales (NSW) which is in the south-western part of Australia (Fig. 2). The region has a history of droughts that have led to huge economic loss in terms of agricultural production, water availability, and social distress (Pittock et al., 2015). The Bureau of Meteorology (BoM), Australia has classified the region into four different climatic zones, wherein most of the region suffers from hot dry summer and cold winters. The three hottest months or the summers typically start in December and ends in February. The region encompasses an area of around 800,400 km^2 with a population of 7,861,700 (ABS, 2018). Around 81% of the region falls under agricultural land, and 14% of the region is classified as protected areas. The most common land use by area is grazing native vegetation, which is roughly 44% of the state (ABARES, 2016). From 1900, the region has suffered from three major droughts [Federation Drought (1895–1902), World War II (1937–1945) and Millennium (2001–2010)] and several minor droughts (Dikshit et al., 2020b). Wittwer (2020) estimated the economic impact due to droughts from 2017 to 2019, and found that a total of 8.1 Billion$ was lost during this period. The Millennium Drought which is considered being the worst drought in history, had led to several social issues like enforcing water restrictions in major cities, an increase in electricity prices, and also a major contributor to bushfire events in 2003 and 2009 (Van Dijk et al., 2013). Further, the drought was considered as the leading factor for a clear reversal in water cycle intensification observed in previous years (Huntington, 2006). Also, the recent bushfires in 2019 have been found to be further aggravated due to the combination of drought conditions, dry vegetation and rise in temperature (Steffen et al., 2019; Nolan et al., 2020).

As explained in the introduction section, the understanding of the drought is quite varied and its definition for characteristics like onset and end is also different. As an example, BoM defines drought onset, when precipitation is below 10th percentile and as serious when precipitation is below 5% of observations, however a clear definition marking the cease of drought has not been defined (Heberger, 2012). However, the Department of Primary Industries, NSW uses Combined Drought Index (CDI) as a drought indicator, which combines meteorological, hydrological and agronomic definitions of drought using indexes for rainfall, soil water and plant growth and drought direction. Therefore, to avoid such confusion, we used the globally accepted SPEI drought index as an indicator to forecast drought index. In terms of climatic variation, the region has seen an increase in the intensity and frequency of hot days and heat waves in Australia, exacerbating drought conditions (Cai et al., 2012), with a decrease in rainfall since 1950 (Dey et al., 2019).

3. Drought index and data used

The data used for the present study is Climatic Research Unit (CRU TS v 4.03) dataset developed by the University of East Anglia at 0.5 ° × 0.5 ° spatial resolution from 1901 to 2018 (Harris et al. 2019, 2020). The dataset has been used for various purposes like climate variability, paleo-climatic (Nagavciuc et al., 2019) and agronomic studies (Renard et al., 2019). In total, the dataset provides ten different variables which can be either primary, secondary or derived. The variables used in the present study are primary variables that include precipitation and mean temperature; secondary variables are vapour pressure and cloud cover; and derived variables are potential evapotranspiration, minimum and maximum temperature.

3.1. Standard Precipitation Evaporation Index

The Standard Precipitation Evaporation Index (SPEI) is one of the most commonly used drought index for forecasting purposes after its introduction by Vincente Serrano et al. (2010). This is primarily due to its dependency on both rainfall and temperature data, unless like SPI, which only uses rainfall data. The calculation of SPEI includes determining “climatic water balance” which involves the use of rainfall and potential evapotranspiration. The calculation of SPEI can be conducted at various time scales, ranging from 1 month to 24 months, which

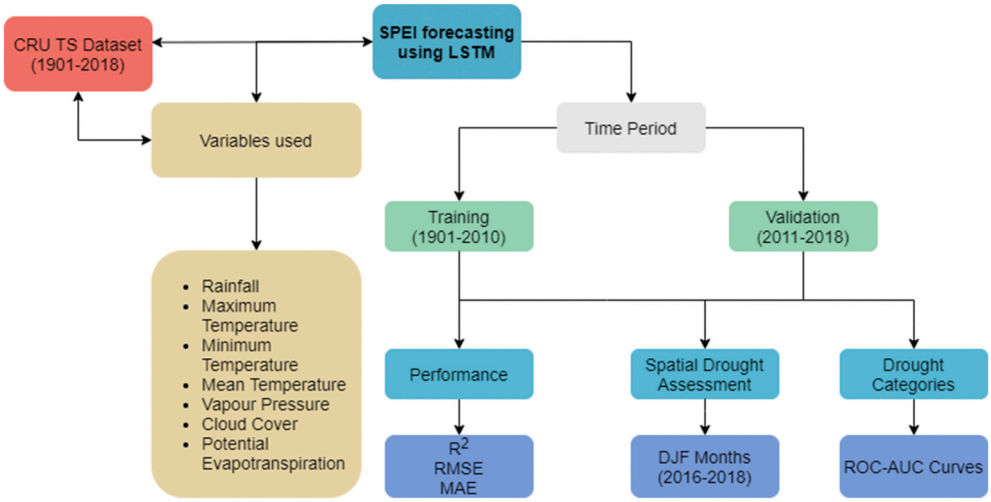

Fig. 1. Flowchart of the study conducted in this research.

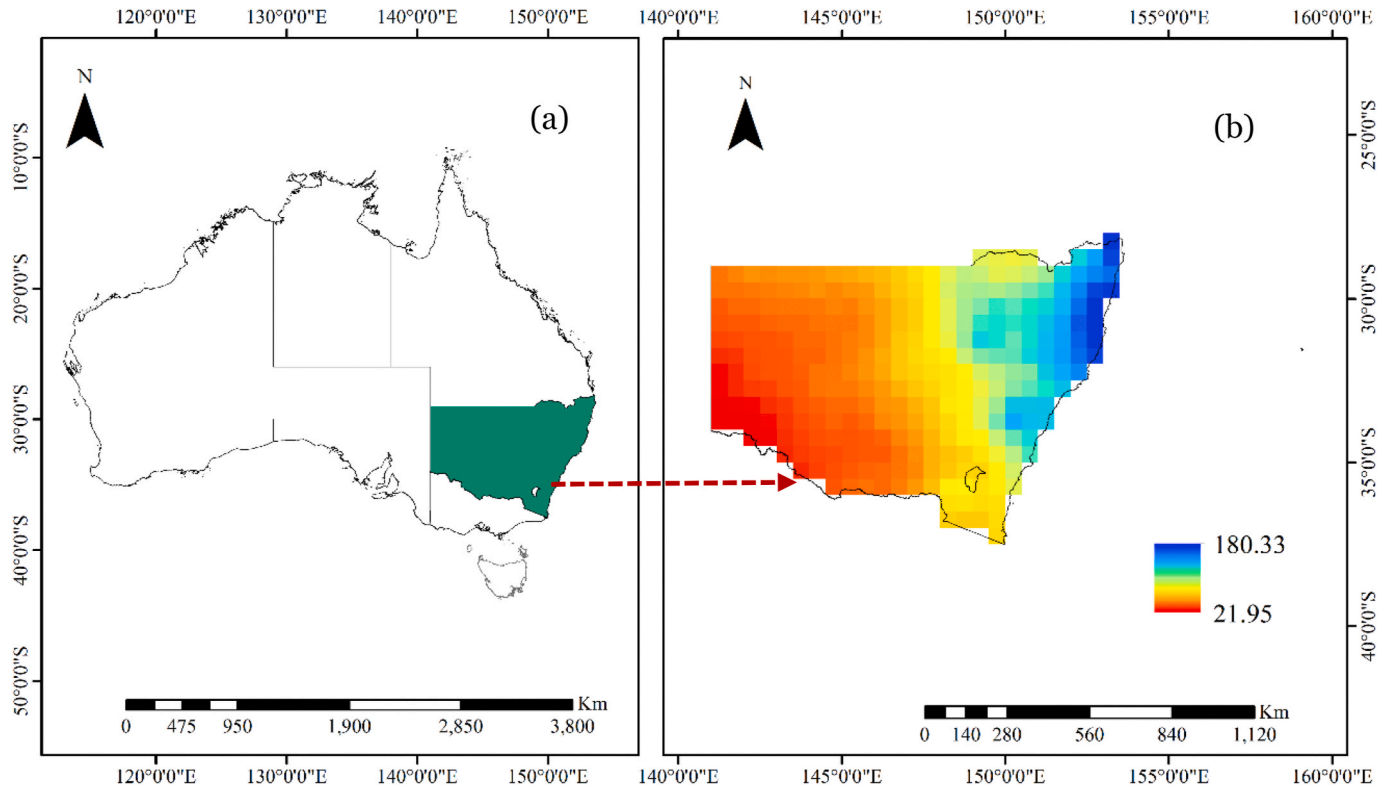

Fig. 2. Location of the study area and the long-term mean rainfall map based on the base period of 1961–1990 calculated from CRU-TS dataset.

depicts different drought type (Vicente Serrano et al., 2010). To calculate SPEI index at different time scales (n = 1, 3), the monthly climatic water balance series is aggregated with an n-month moving sum, i.e. the current monthly value and the previous n − 1 monthly values. Like, a 3-month accumulation data for January–February–March determines the index for the month of March (Vicente Serrano et al., 2010). The climatic water balance is computed at different time scales, and the resulting values are fitted to a log-logistic probability distribution to transform the original values to standardized units (Beguería et al., 2014). Generally, the shorter time scales (1–3 months) represent the meteorological drought, whereas 3–6 month time scale describes agricultural drought, while a larger time scale, 12–24 months is suitable to describe hydrological drought (Mishra and Singh, 2010). As meteorological droughts can be considered as the first step in drought evolution, the present study, forecasts SPEI at 2 different time scales (SPEI 1 and SPEI 3). Interested readers are referred to Vicente Serrano et al. (2010; 2012) for a detailed study about the calculation of SPEI drought index. The global SPEI database at different monthly scales using the CRU dataset can be accessed from https://spei.csic.es/database.html. Table 2 represents the various drought categories as per SPEI index values.

4. Recurrent Neural Networks and LSTM model

Deep learning as a distinct field has emerged to reduce human effort in traditional machine learning (ML) approaches for various tasks like feature extraction and regression purposes (LeCun et al., 2015). Typically, ML models have some level of human input which makes it difficult to understand complex situations and therefore, deep learning which does not involve human input became more prominent. Although, the concept of deep learning can be tracked back to 1950, it resurrected itself after defeating humans in the game of Go, which was one of the biggest achievement in the recent time (Silver et al., 2016). Further, the detailed review of deep learning, LeCun et al. (2015) gave new directions to various research fields and has been adopted in respective domains. There have been several milestones achieved in deep learning in the past decade, each achieving new feats in their respective fields (Schmidhuber, 2015). However, the traditional computer vision field is slightly different from geohazards or geosciences applications as the latter involves a dynamic component which is not the case in the former (Reichstein et al., 2019). Also, the availability of various types of data include remote sensing, atmospheric or climatic data has led researchers to use different approaches with a definite aim like forecasting or monitoring.

Table 2
Drought categories as per SPEI values (Rhee and Im, 2017).

SPEI Classifications	Categories
≤-2.0	Extremely Dry
−1.99~-1.5	Severely Dry
−1.49~-1.0	Moderately Dry
−0.99–0.99	Near Normal
1.0–1.49	Moderately Wet
1.5–1.99	Severely Wet
≥2.0	Extremely Wet

Of the various deep learning based approaches, Recurrent Neural Networks (RNN) is a neural network type which is used to understand non-stationary data like time series data. It can be considered as a series of interconnected networks for time series analysis and can be trained using back propagation based gradient descent algorithms (Williams and Zipser, 1989). The ability to consider both the current and preceding input data for mapping target vectors in RNN makes it useful compared to neural networks, which map target vectors by multiplying weights. Also, RNN has the ability to store an internal memory of previous inputs in the network, which allows it to recall key events that occurred several times in the past, which is key in studies like drought forecasting. The scenario where RNN fails is when stacking occurs leading to vanishing and exploding gradient problems (Bengio et al., 1994). This led to the introduction of long short-term memory (LSTM) (Hochreiter and Schmidhuber, 1997), which comprised of a cell capable of storing the values to be used at random intervals and three gates, viz., input, output

and forget gate, to control and adjust the cell state.

The structure of LSTM is like a chain as shown in Fig. 3, wherein the basic building block is a cell and its state is the key to the mode. There are three types of gate which determines the cell state, which includes an input, forget gate and an output gate. The gates analyse and control the amount of information it can pass through and are comprised of a sigmoid neural layer and point-wise multiplication operation (Olah, 2015). The working mechanism of the gates and information flow can be expressed using the following equations:

$$f_t = \sigma\left(W_f.[h_{t-1}, x_t] + b_f\right) \quad (1)$$

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \quad (2)$$

$$C_t^{'} = tanh(W_C.[h_{t-1}, x_t] + b_c) \quad (3)$$

$$C_t = f_t{*}C_{t-1} + i_t{*}C_t^{'} \quad (4)$$

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = \sigma_t{*}tanh(C_t) \quad (6)$$

Where x_t is the input vector at time t with σ being the activation function like *Sigmoid* or *ReLU*. W_f, W_i, W_C and W_o are the applied weights to concatenation of the new input x_t and output h_{t-1} from the previous cell, with b_f, b_i, b_c, and b_o being the corresponding bias (Xiao et al., 2019). f_t, i_t, and o_t are the outputs of three sigmoid functions σ, and the values range from 0 to 1. These control the information which are forgotten in the old cell state C_{t-1} and passed to the new cell C_t with the new information being $C_t^{'}$, with h_t being the output information from the cell. There are several variants of LSTM and interested readers can refer to Goodfellow et al. (2016).

The LSTM model used for SPEI forecasting, is depicted in Fig. 4. It consists of an input layer, one LSTM layers and one Dense layer (also known as fully connected layers). We have conducted several experiments and found that this architectural design achieves the best prediction performance. The input of the whole network is in 3D tensor form and expressed as [sample_size (1901–2010), time_steps, features_n (7)]. Sample_size is the training data, and is set to 2010. time_steps is the size of the time window (previous months) used to predict the SPEI. As there is no fixed rule for portioning the data, the most commonly used approach is to split the data into two sets (Mokhtarzad et al., 2017). Further, the amount of data in the training set has also no set rule, therefore, the present work uses 90% of the dataset as training (Deo et al., 2017b; Dikshit et al., 2020b). Therefore, training data is set to 2010 and the remaining as validation. The choice of the time steps was set to 20, based on a trial-and-error approach and running several experiments (time_step = 5, 10, 15, ….50). This means that the parameters from the past 19 months also including the 20th month parameters was used to predict the 21st month's SPEI. As we use SPEI values as predictors, based on multiple involved factors, we set the feature_n to the number of involved factors, i.e., 7. A dropout mechanism is applied to the inputs to help prevent over-fitting during training, which is empirically set to 0.25 (Xiao et al., 2019). For regression task, the input data is normalized to the range of [0 1] using the following algorithm:

$$X_{nor} = \frac{(Y_{max} - Y_{min})^{*}(X - X_{min})}{(X_{max} - X_{min})} + Y_{min} \quad (7)$$

Where X is the value to be normalized; X_{nor} is the normalized X value; $Y_{max} = 1$; $Y_{min} = 0$; X_{max} and X_{min} are the maximum and minimum value of each time series respectively. The LSTM deep neural network is implemented with Keras (Francois, 2015) utilizing TensorFlow 2.0 (GPU version) as the backend.

The design of the architecture was initially conducted for SPEI 1 case. Later, similar approach was used for SPEI 3 case, however, the change in metrics was minimal, and therefore, both SPEI 1 and SPEI 3 have the same architecture. After training the data, the predicted images for SPEI 1 SPEI 3 at one month lead time are generated to analyse the spatial variation. The analysis was conducted at different seasonal level to examine how the model performs, which would give more clarity about the model's capability. The performance of the model was examined using various approaches and the details are presented below.

4.1. Performance metrics

The performance metrics was based on three different statistical metrics for analysing the forecasted results at different lead times. The metrics used were Coefficient of Determination (R^2) and Root Mean Square Error Method (RMSE). The mathematical formulae to the metrics are:

$$R^2 = \frac{\sum_{i=1}^{N}(\hat{y}_i - \overline{y_i})}{\sum_{i=1}^{N}(y_i - \overline{y_i})^2}, \quad (8)$$

$$\overline{y_i} = \frac{1}{N}\sum_{i=1}^{N} y_i, \quad (9)$$

where, $\overline{y_i}$ is the mean value, y_i and $\hat{y}_i$ are observed and forecasted values, N being the number of data points.

$$RMSE = \sqrt{\frac{SSE}{N}}, \quad (10)$$

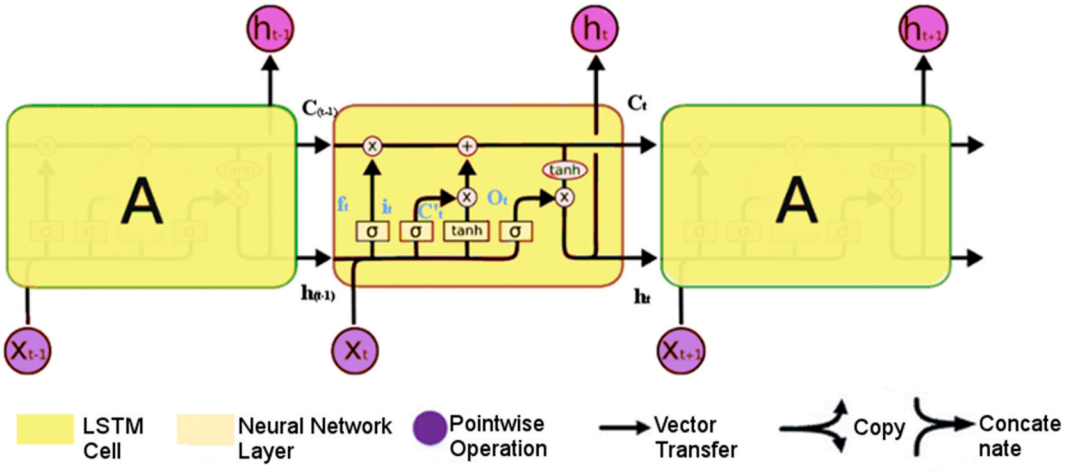

Fig. 3. Structure of LSTM network (Olah, 2015).

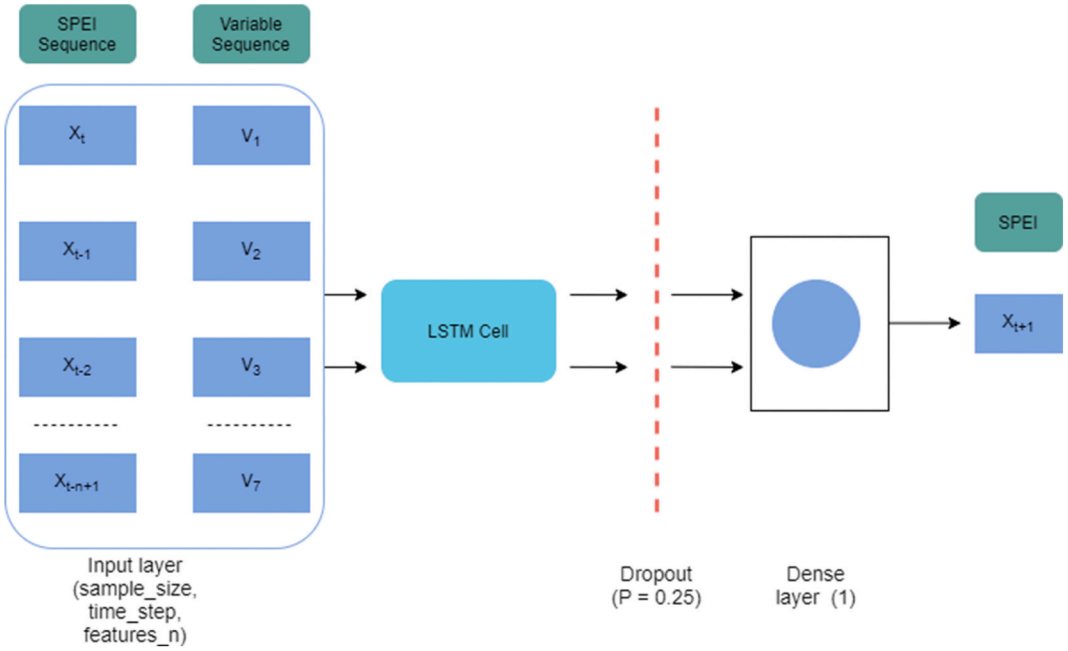

Fig. 4. Architecture of the LSTM network for SPEI forecasting.

$$SSE = \sum_{i=1}^{N} (\hat{y}_i - y_i)^2, \tag{11}$$

where, SSE refers to sum of squared errors.

RMSE is frequently used as an evaluation metric as it penalises large errors and is suitable for time series forecasting purposes. R^2 represents the extent of association between the observed and forecasted values. The value ranges from 0 to 1, where 1 indicates an exact match and 0 denotes no association. By contrast, a lower RMSE value depicts better performance. Mean Absolute Error (MAE) determines the average of absolute errors, analysing the degree of proximity of forecasted values with the observed values.

As the study also aimed to analyse the drought class (Table 2) of the forecasted results, a multi class Receiver Operating Characteristic based Area under Curve (ROC-AUC) approach was used. This approach determines the sensitivities and specificities at all the thresholds which are defined as per the response of the classifier for a test set which is followed by AUC computation using the trapezoid rule (Buda et al., 2018). The traditional use of these curves revolves around binary classification works, therefore, we implemented a multi-class ROC (Provost and Domingos, 2003) to address the various drought classes and the imbalances associated with it. We used the scikit-learn python package to implement the technique (Pedregosa et al., 2011). The methodology included the determination of statistical metrics for every class and thereby averaging the results. The average of the results can be done either using micro or macro averages. Generally, for a multi class classification problem with class imbalance case, as is the present study, micro-average is to be preferred (Van Ash, 2013). This is due to the principle behind computation of these averages. A micro average aggregates the contributions of all classes to compute the average metric, whereas a macro average would calculate the metric independently for each class and then compute the average. So, the methodology involves the use of a performance table as shown in Table 3.

The calculation of AUC for any class involves determining the Sensitivity and 1- Specificity. Sensitivity is defined as: $TP/TP+FN$; whereas Specificity is defined as: $TN/TN+FP$. The mathematical formulae for determining the micro-average AUC is (Pedregosa et al., 2011):

$$\text{Micro average} = \frac{2}{a(a-1)} \sum_{j=1}^{a} \sum_{y>x}^{a} (AUC(x|y) + AUC(y|x)) \tag{12}$$

where a is the number of classes and $AUC(x|y)$ is the AUC with class x as the positive class and class y as the negative class.

5. Results and discussion

At first, the statistical metrics of the regression aspect was computed as shown in Table 3. The results (Table 4) reveal excellent results during both the periods, thus affirming its superiority over traditional machine learning models based on previous works conducted in NSW. Deo and Şahin (2015) used Artificial Neural Networks (ANN) and climatic indices to forecast monthly SPEI at five NSW locations and achieved R^2 values of ~0.99. However, the present study did not use climatic indices as predictors and achieved similar results, thereby showing the benefits of using deep learning model over traditional machine learning models. Dikshit et al. (2020b) used Random Forests model to predict SPEI 1 and SPEI 3 for NSW region, and the R^2 value achieved was 0.73 and 0.76 respectively. Also, Deo et al. (2017a) used a multivariate adaptive regression splines (MARS) model to forecast SPI for five different regions of NSW, and the R^2 value achieved ranged from 0.971 to 0.987. The forecasted results are examined in two different ways: a) Spatio-temporal variation b) Variation in terms of drought categories. As

Table 3
Performance table for occurrences labelled with class label X (Van Ash, 2013).

	True label (X)	True not (X)
Predicted label (X)	True Positive (TP)	False Positive (FP)
Predicted not (X)	False Negative (FN)	True Negative (TN)

Table 4
Statistical metrics of LSTM model under training and validation period.

Predicting	Training			Validation		
	R^2	RMSE	MAE	R^2	RMSE	MAE
SPEI 1	0.998	0.013	0.012	0.996	0.018	0.01
SPEI 3	0.997	0.016	0.014	0.992	0.027	0.024

it is not feasible to depict the variation for every month, we highlight the spatio-temporal variation for the summer months (December–January – February) from 2016 to 2018. As summer months are a period of high temperature and low rainfall, and an analysis for this period can be considered as a good estimate of the forecasting abilities for other time periods, either based seasonally or on months. The variation at 1 month lead time between original and predicted SPEI 1 (Fig. 5) and SPEI 3 values for NSW region are illustrated in Fig. 6.

5.1. Spatio-temporal variation

5.1.1. SPEI 1 values

The observed SPEI 1 values during December 2015 depict very few regions with extremely dry conditions towards the south-west part, and the following month shows no drought conditions with the majority of the region depicting moderately to severely wet conditions. However, situations change in February 2016, wherein the spatial extent of drought increases and 37.7% of the area comes under drought. Now, when analysing the summer of 2017, the month of December depicts few regions (5.8%), particularly, the northern-part under drought, but the drought increases towards severely dry conditions in January engulfing the south-eastern region and then further intensifying in February with the drought regions more towards the central and northern part of NSW. Similarly, in the summer of 2018, the month of December showed no drought conditions, whereas January depicted areas with severe drought conditions, and the effect decreased in February. This would help to understand how the drought propagates within a region on a monthly scale. Now, on comparing with the predicted SPEI 1 maps, the spatial variation in terms of values and categories is similar to observed maps. Also, the number of pixels under drought conditions (SPEI < -1) is nearly 3%–5% more across the same months and under non drought conditions (SPEI>1) are less than the observed. This can be considered as a good step, as over prediction to a certain extent is good and could be helpful for policy makers. Similarly, for December 2017, the SPEI 1 values indicated no indication of drought and can be considered as near normal condition. However, in January 2018, the reduction in rainfall lead the western part of the area to come under severe drought. Also, almost 80% of the region was under some sort of drought category. The situation eases a little in February 2018, with the north and north-western part of the region under moderately dry condition. When comparing with the predicted images for the summer of 2018, the month of December leads to consistent results with the observed, and so is the case with the month of January. However, for February 2018, the area under drought is over predicted by 8% with some pixels depicting severely dry conditions.

5.1.2. SPEI 3 values

For SPEI 3 case, the index value for December 2015 depicted the western part to be under severe drought (5.4%), with the conditions easing in the following month, with no region under drought. However, the month of February depicted very few pixels under drought (1.9%) in the northern part. The month of December in 2016 depicts most of the region as under near-normal conditions, whereas certain regions exhibited moderately dry conditions in the northern part, with few pixels highlighting severely dry conditions. The drought intensity for severely dry condition was less, and generally, the area was not under drought influence. Thereafter, January 2017 highlighted more drought areas with more regions exhibiting moderately-dry and severely-dry conditions, especially towards the south-eastern part. Further, in February 2017 the drought conditions expand to more areas, with more regions depicting moderately and severely dry conditions. Now, when we compare this with the predicted images, the range of index values as per drought class is generally same across summer of 2017, however the number of pixels under drought were over estimated by 4.5% in December 2016, under predicted in January and February 2017 by 3.7% and 5.6% respectively. Also, when analysing in terms of clusters for drought pixels, the variation is not significant enough and follows same trend.

The index values during December 2017 depict near normal conditions across the state, which leads to moderately dry conditions for January 2018, especially in the northern part of NSW and the drought intensifies to severe category and moves to further west in February 2018. On comparing it with the predicted images, the month of December 2017 shows similar conditions, and as the area was in near normal conditions, the variation in index value per pixel is not important. Further, for January 2018 the number of pixels under drought is more than observed, but the values have been over predicted by 6.4%, with a few of the pixels (2.4%) depicting moderately dry conditions, when the observed did not depict any drought. Similar is the condition for February 2018, which was under predicted by 3.6%.

To examine the importance of the LSTM architecture, annual rainfall and annual mean temperature anomaly maps are shown in Fig. 7(a) and (b), with the baseline period as 1961–1990. For forecasting purposes, ML models learn uniform weights, whereas LSTM models learn variable weights across time steps. As the figure suggests, a significant variation in rainfall and temperature anomalies is observed during the summer periods of 2016–2018. This phenomenon necessitates the use of decay over weights across periods. Hence, the use of LSTM is encouraged to learn decayed weights. The forget gate in LSTM ensures that the model can effectively capture the decay weighted lag–lead sequence relationship without the vanishing gradient problem.

Further, on examining the variation between SPEI 1 and SPEI 3, an interpretation study was conducted to understand the difference among the weights learned by variables. The results indicate that in case of SPEI 1, the most dominant factor was rainfall, however, temperature was most dominant for SPEI 3. The influence of vapour pressure and cloud cover were relatively similar for both the scenarios, but were more influential than PET. This suggests that cloud cover and vapour pressure also play a key role, and their inclusion can improve forecasting results. Further, in order to examine the spatial variation between variables and drought index, a Convolutional Neural Network based LSTM (CNN-LSTM) architecture, would we well suited. The recent study by Ham et al. (2019) to forecast El Niño/Southern Oscillation (ENSO) index used such architecture to identify the hotspots of the variables, along with forecasting at multi-year lead times.

5.2. Drought classification

The above discussion reflects how the drought movement occurs at monthly scales and also during different summer seasons. The key to understand the difference is the spatial variability depicting the index values and also the drought category it represents. When we analysed the observed and predicted images, there was minor variation in the index values, but represented different drought classes. For SPEI 1 case, the observed value of February 2017 showed a maximum index value of -1.99 which is the borderline for severely dry condition, however, for the same period, the predicted value depicted maximum value of -2.08, which comes under severely dry condition. Similar was the case for February 2018, wherein the predicted value depicted a higher drought class as compared to the observed value. This was also the case for SPEI 3 condition, December 2016, observed as severely dry condition whereas predicted fell under moderately dry condition. Therefore, there can be few scenarios where statistics may not be able to reveal drought class, due to the proximity between them.

The variation was analysed by computing the micro average of ROC based AUC curves. The results reveal that under both SPEI 1 and SPEI 3; the model achieved a value of 0.83 and 0.82, respectively (Fig. 8). This proximity along the borderline of drought category has led to slightly lower values of AUC curves. However, when considering all the classes, the micro average from the present study outperforms the previous studies. This suggests that even though the statistical metrics provide excellent results, the understanding of droughts cannot be based only on

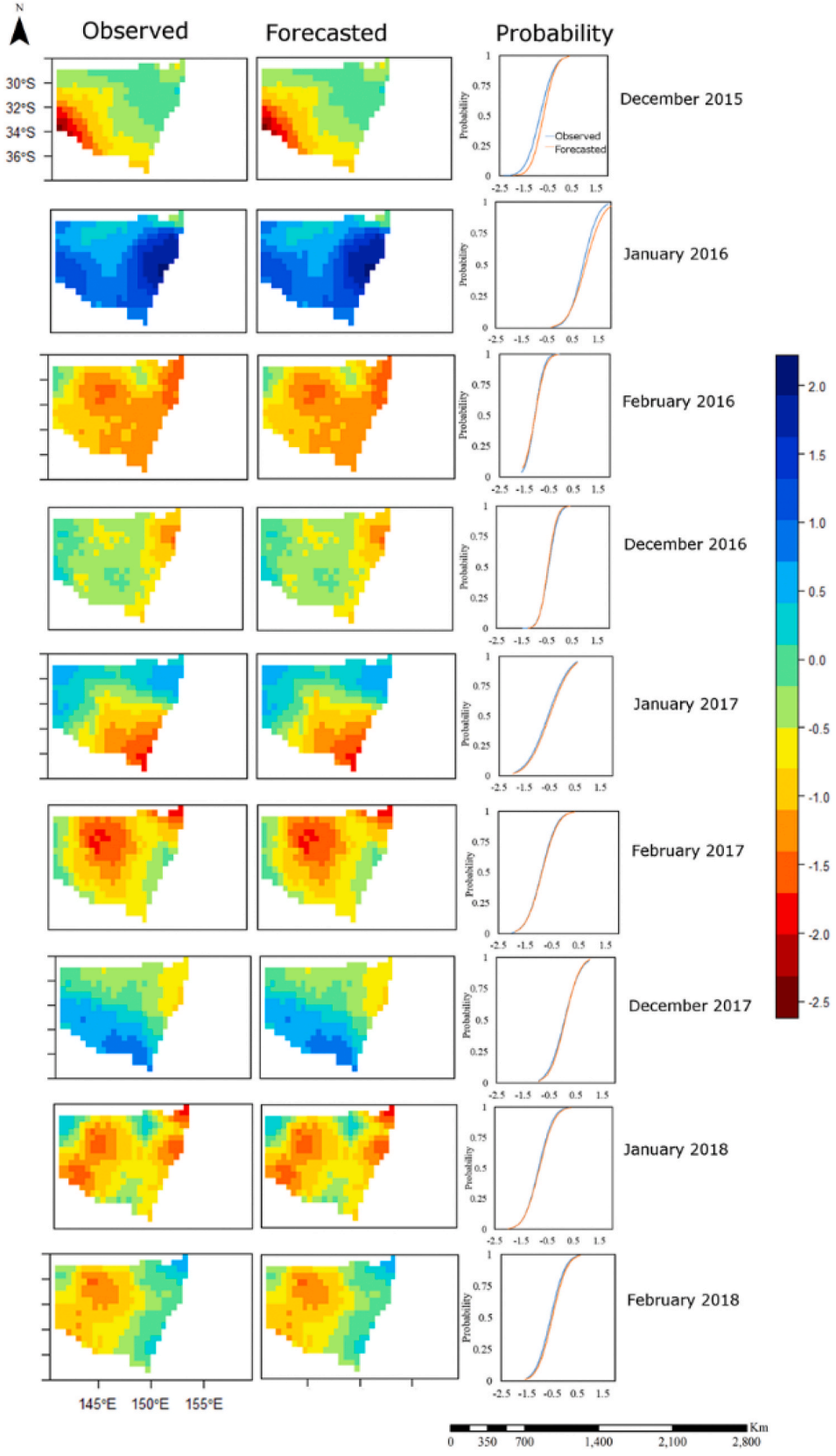

Fig. 5. Spatial variation of SPEI 1 among the DJF (December–January – February) months of 2016–2018. The first column depicts the observed SPEI 1 values, second column depicts the predicted index values. In the third column depict the probability (cumulative distribution function) plots of observed and forecasted values – of SPEI 1.

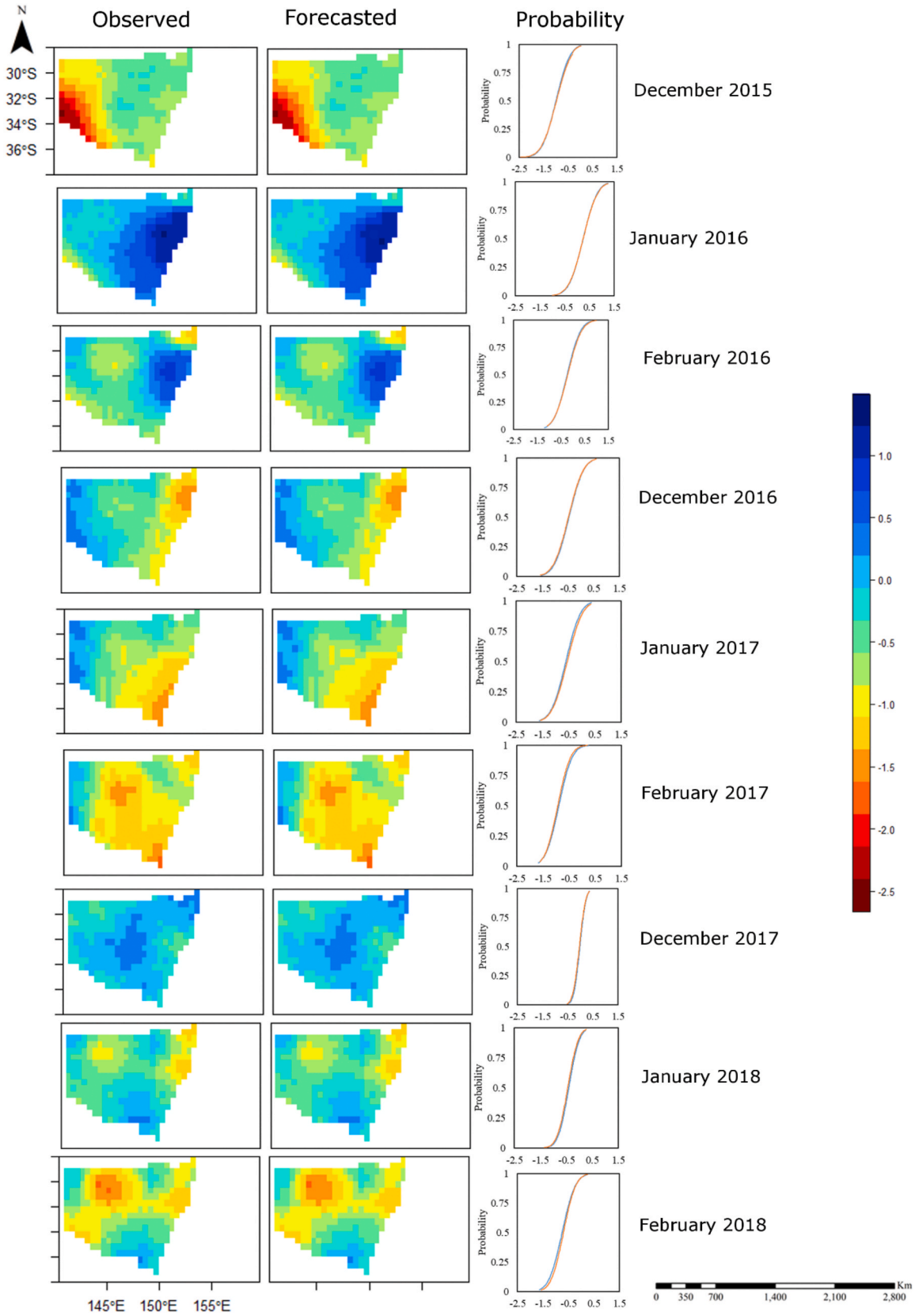

Fig. 6. Spatial variation of SPEI 3 among the DJF (December–January – February) months of 2016–2018. The first column depicts the observed SPEI 3 values, second column depicts the predicted SPEI 3 values. Legends represent drought index values. The third column depicts the probability (cumulative distribution function) plots of observed and forecasted SPEI 3 values.

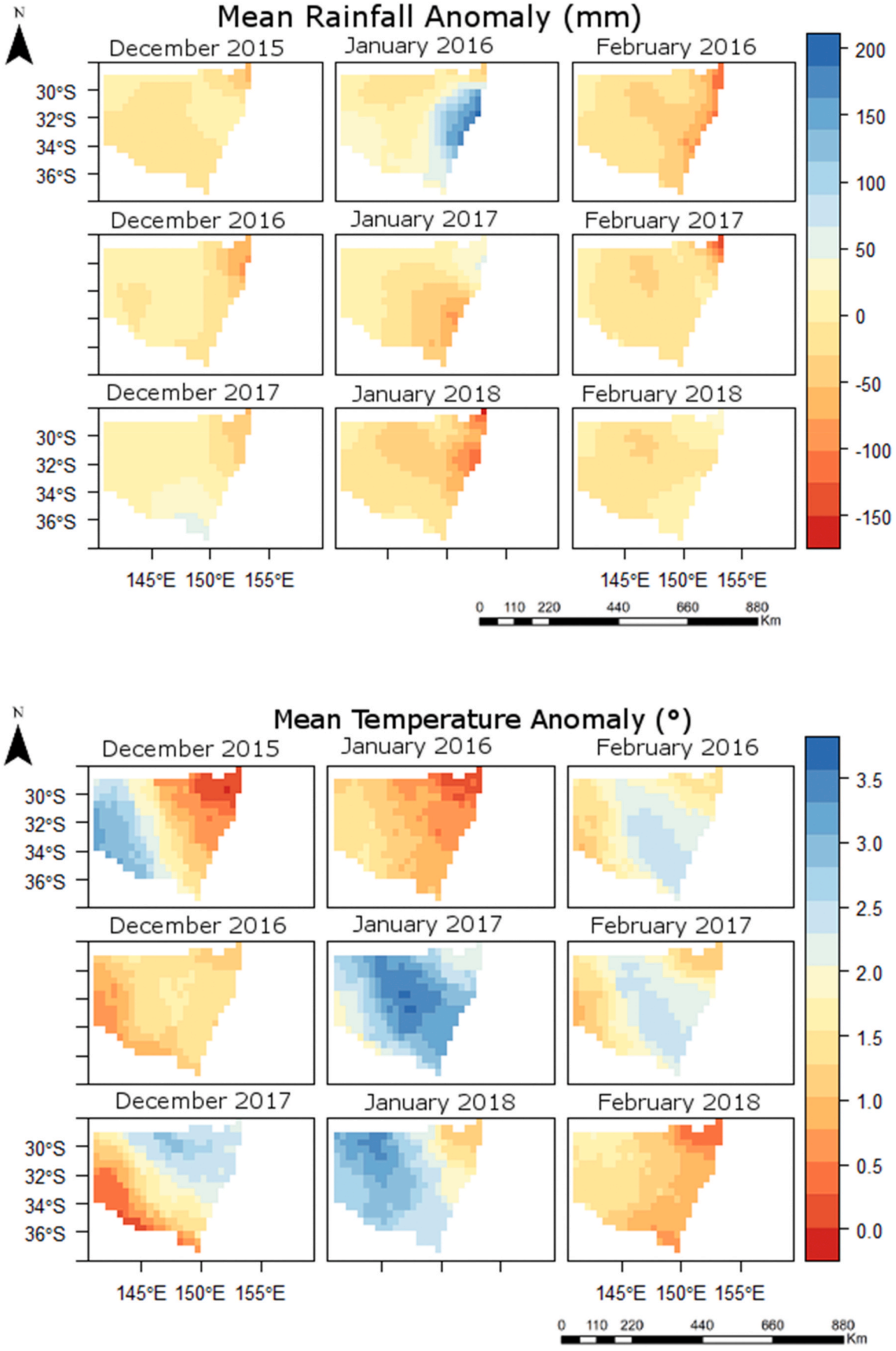

Fig. 7. Spatial anomaly maps of (a) rainfall; and (b) mean temperature during the summer months of 2016–2018.

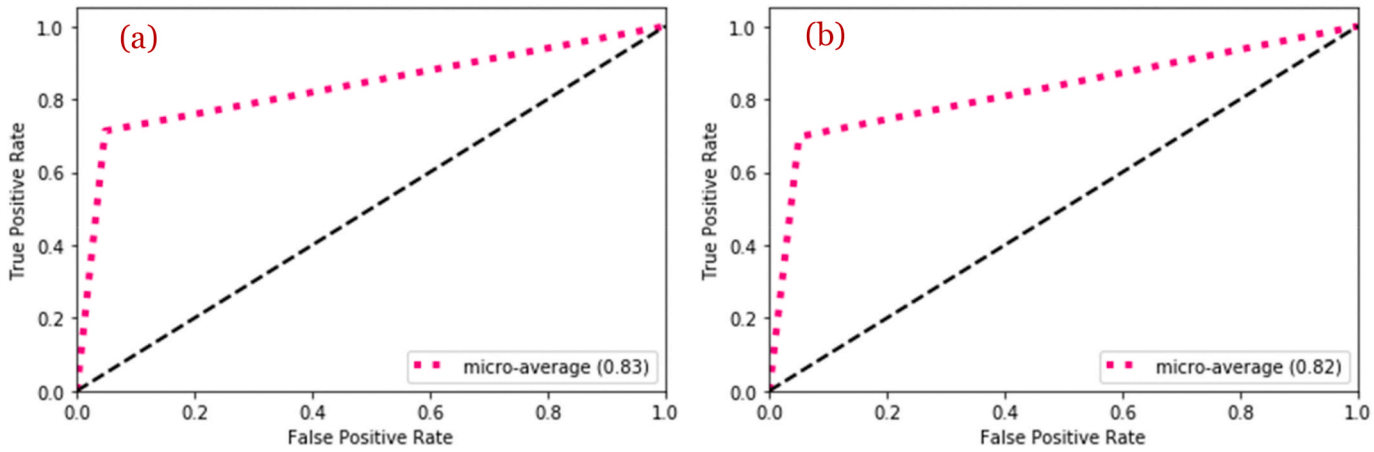

Fig. 8. ROC-AUC curves of different drought categorization classes (a) SPEI 1, and (b) SPEI 3.6. Conclusion.

such metrics, instead it should be based on the objective of the study. Moreover, the results achieved in the present work can be considered quite acceptable given that a formidable mitigation strategy for 1 month lead time can be developed.

Droughts are one of the most destructive hazards causing severe economic and social distributions. One of the most effective ways to understand droughts is to improve the existing forecasting ability. As the onset of drought is not clearly defined, forecasting of droughts make a compelling argument to improve the models providing better mitigation strategies. The present work uses a deep learning approach, namely LSTM, which has proven to be more effective for forecasting purposes compared to traditional machine learning approaches. The study forecasts SPEI drought index for New South Wales region determined using the global climatological dataset (CRU TS v 4.03) using seven different meteorological variables also collected from CRU dataset. The predictor and variables were collected from 1901 to 2018 of which 1901–2010 were used as training data and then validated from 2011 to 2018. Apart from understanding the variation in pixel values, it is equally important to analyse under different drought characteristics, therefore a multi class ROC-AUC curves was prepared to understand the changes in terms of drought classes. In the present work, we focussed towards examining the model in terms of drought categories, which could be essential for mitigation purposes, especially when studying for a large area. This is the first study in the use of a deep learning approach utilizing a global climatological dataset for drought forecasting. We strongly believe that the LSTM model has the forecasting capability to forecast droughts across different dataset type and drought indices, subsequent studies would be conducted to assert the findings. The findings from the study are as follows:

- The performance metrics for both SPEI 1 and SPEI 3 depicted excellent results highlighting its significance over other models. The R^2 value for SPEI 1 and SPEI 3 achieved values of and 0.998 and 0.996 respectively.
- However, the statistical metrics may not always reflect the variation and therefore a spatial analysis for the summer of 2017 and 2018 was conducted to examine how the index values vary at pixel level and across the summer seasons.
- The micro average value of ROC-AUC curves depicted value of 0.83 and 0.82 for SPEI 1 and SPEI 3 case, respectively. This could be a reflective of the threshold levels of drought categories, as few months depicted index values at the borderline. However, more stacked deep neural networks model would be built in the future, to improve our findings, especially for pixels representing values at the threshold point.

The applicability of the LSTM architecture needs to be tested in different climatic conditions to examine how well it captures the interrelationship between variables and drought index. The future works would also look towards involving climatic variables to forecast at longer time scales using deep learning techniques. It's well accepted that the use of deep neural networks would provide better forecasting results (Reichstein et al., 2019), the important aspect would be to interpret the model and provide a more in-depth explanation of the results (Dikshit et al., 2020c). The results from this study are useful for drought mitigation purposes, bushfires and for policy makers. The future study would look towards improving the LSTM architecture and use of different drought indices at both short-term and long-term drought scales, which would further enhance our understanding.

Funding

This research is funded by the Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of engineering and IT, University of Technology Sydney.

Credit author statement

Abhirup Dikshit: Conceptualization, Methodology, Modelling, Writing original draft. Biswajeet Pradhan and Alfredo Huete: Supervision, Validation, Visualization, Review & Editing, Funding.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

ABARES, 2016. Land Use of Australia 2010–11. ABARES, Canberra. May.
ABS, 2018. Regional Population by Age and Sex, Australia, 2017, Cat. No. 0 Australian Bureau of Statistics, Canberra, p. 3235.
Alizadeh, M.R., Nikoo, M.R., 2018. A fusion-based methodology for meteorological drought estimation using remote sensing data. Remote Sens. Environ. 211, 229–247.
Beguería, S., Vicente-Serrano, S.M., Reig, F., Latorre, B., 2014. Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. Int. J. Climatol. 34 (10), 3001–3023.
Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. IEEE Trans. Neural Network. 5, 157–166.
Buda, M., Maki, A., Mazurowski, M.A., 2018. A systematic study of the class imbalance problem in convolutional neural networks. Neural Network. 106, 249–259.
Cai, W., Purich, A., Cowan, T., et al., 2014. Did climate change–induced rainfall trends contribute to the Australian Millennium Drought? J. Clim. 27, 3145–3316.
Collobert, R., et al., 2011. Natural language processing (almost) from scratch. J. Mach. Learn. Res. 12, 2493–2537.

Deo, R.C., Şahin, M., 2015. Application of the artificial neural network model for prediction of monthly standardized precipitation and evapotranspiration index using hydrometeorological parameters and climate indices in eastern Australia. Atmos. Res. 161, 65–81.
Deo, R.C., Kisi, O., Singh, V.P., 2017a. Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model. Atmos. Res. 184, 149–175.
Deo, R.C., Tiwari, M.K., Adamowski, J.F., Quilty, J.M., 2017b. Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model. Stoch. Environ. Res. Risk Assess. 31, 1211–1240.
Dey, R., Lewis, S.C., Arblaster, J.M., Abram, N.J., 2019. A review of past and projected changes in Australia's rainfall. Wiley Interdiscip. Rev.: Clim. Change 10, e577.
Dikshit, A., Pradhan, B., Alamri, A.M., 2020a. Temporal hydrological drought index forecasting for New South Wales, Australia using machine learning approaches. Atmosphere 11 (6), 585.
Dikshit, A., Pradhan, B., Alamri, A.M., 2020b. Short-term spatio-temporal drought forecasting using random Forests model at New South Wales, Australia. Appl. Sci. 10, 4254.
Dikshit, A., Pradhan, B., Alamri, A.M., 2020c. Pathways and challenges of the application of artificial intelligence to geohazards modelling. Gondwana Res. https://doi.org/10.1016/j.gr.2020.08.007.
Farabet, C., Couprie, C., Najman, L., LeCun, Y., 2012. Scene parsing with multiscale feature learning, purity trees, and optimal covers. In: Proc. International Conference on Machine Learning. http://arxiv.org/abs/1202.2160.
Francois, C., 2015. Keras. Github. https://github.com/keras-team/keras.
Fung, K.F., Huang, Y.F., Koo, C.H., Soh, Y.W., 2019. Drought Forecasting: A Review of Modelling Approaches 2007–2017. J. Water Clim, Chang.
Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning.
Ham, Y., Kim, J., Luo, J., 2019. Deep learning for multi-year ENSO forecasts. Nature 573, 568–572.
Hao, Z., Hao, F., Xia, Y., et al., 2016. A statistical method for categorical drought prediction based on NLDAS-2. J. Appl. Meteor. Climatol. 55 (4), 1049–1061.
Hao, Z., Singh, V.P., Xia, Y., 2018. Seasonal drought prediction: advances, challenges, and future prospects. Rev. Geophys. 56 (1), 108–141.
Harris, I.C., 2019. CRU TS v4.03: Climatic Research Unit (CRU) Time-Series (TS) Version 4.03 of High-Resolution Gridded Data of Month-By-Month Variation in Climate. Centre for Environmental Data Analysis (CEDA). https://doi.org/10.5285/10d3e3640f004c578403419aac167d82. Jan. 1901- Dec. 2018.
Harris, I., Osborn, T.J., Jones, P., Lister, D., 2020. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. Sci. Data 7, 109.
Heberger, M., 2012. Australia's millennium drought: impacts and responses. In: The World's Water. Springer, pp. 97–125.
Hinton, G.E., Deng, L., Yu, D., et al., 2012. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Process. Mag. 29 (6), 82–97.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780.
Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Rem. Sens. Environ. 25 (3), 295–309.
Huntington, T., 2006. Evidence for intensification of the global water cycle: review and synthesis. J. Hydrol. 319 (1—4), 83–95.
Krizhevsky, A., Sutskever, I., Hinton, G., 2012. ImageNet classification with deep convolutional neural networks. Proc. Adv. Neural Inf. Process. Syst. 25, 1090–1098.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.
McKee, T.B., Doesken, N.J., Kleist, J., 1993. The relationship of drought frequency and duration to time scales. In: Proceedings of Proceedings of the 8th Conference on Applied Climatology, pp. 179–183.
Mishra, A.K., Singh, V.P., 2010. A review of drought concepts. J. Hydrol. 391, 202–216.
Mishra, A.K., Singh, V.P., 2011. Drought modeling–A review. J. Hydrol. 403, 157–175.
Mokhtarzad, M., Eskandari, F., Vanjani, N.J., Arabasadi, A., 2017. Drought forecasting by ANN, ANFIS, and SVM and comparison of the models. Environ. Earth Sci. 76, 729.
Nagavciuc, V., Ionita, M., Perșoiu, A., et al., 2019. Stable oxygen isotopes in Romanian oak tree rings record summer droughts and associated large-scale circulation patterns over Europe. Clim. Dynam. 52, 6557–6568.
Nolan, R.H., Boer, M.M., Collins, L., et al., 2020. Causes and consequences of eastern Australia's 2019-20 season of mega-fires. Global Change Biol. 26 (3), 1039–1041.
Olah, C., 2015. Understanding LSTM Networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs/.
Özger, M., Mishra, A.K., Singh, V.P., 2012. Long lead time drought forecasting using a wavelet and fuzzy logic combination model: a case study in Texas. J. Hydrometeorol. 13 (1), 284–297. https://doi.org/10.1175/JHM-D-10-05007.1. In this issue.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al., 2011. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.
Pittock, J., Grafton, R.Q., Williams, J., 2015. The MurrayDarling Basin Plan fails to deal adequately with climate change. Water. J. Aust. Water Assoc. 42 (6), 28–34.
Poornima, S., Pushpalatha, M., 2019. Drought prediction based on SPI and SPEI with varying timescales using LSTM recurrent neural network. Soft Comput. 23, 8399–8412.
Provost, F., Domingos, P., 2003. Tree induction for probability-based ranking. Mach. Learn. 52 (3), 199–215.
Reichstein, M., Camps-Valls, G., Stevens, B., et al., 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204.
Renard, D., Tilman, D., 2019. National food production stabilized by crop diversity. Nature 571, 257-26.
Rhee, J., Im, J., 2017. Meteorological drought forecasting for ungauged areas based on machine learning: using long-range climate forecast and remote sensing data. Agric. For. Meteorol. 237–238, 105–112.
Schmidhuber, J., 2015. Deep learning in neural networks: an overview. Neural Network. 61, 85–117.
Shen, R., Huang, A., Li, B., Guo, J., 2019. Construction of a drought monitoring model using deep learning based on multi-source remote sensing data. Int. J. Appl. Earth Obs. Geoinf. 79, 48–57.
Silver, D., et al., 2016. Mastering the game of Go with deep neural networks and tree search. Nature 529, 484–489.
Slette, I.J., Post, A.K., Awad, M., et al., 2019. How ecologists define drought, and why we should do better. Global Change Biol. 25 (10), 3193–3200.
Spinoni, J., Barbosa, P., De Jager, A., et al., 2019. A new global database of meteorological drought events from 1951 to 2016. J. Hydrol.-Reg. Stud. 22, 100593.
Steffen, W., Hughes, L., Mulling, G., et al., 2019. Dangerous Summer: Escalating Bushfire, Heat and Drought Risk. Available:. https://www.climatecouncil.org.au/wp-content/uploads/2019/12/report-dangerous-summer_V5.pdf.
Sun, Q., Miao, C., Duan, Q., et al., 2018. A review of global precipitation data sets: data sources, estimation, and intercomparisons. Rev. Geophys. 56 (1), 79–107.
Van Asch, V., 2013. Macro-and Micro-averaged Evaluation Measures [[basic Draft]]. CLiPS, Belgium, p. 49.
Van Dijk, A.I., Beck, H.E., Crosbie, R.S., et al., 2013. The millennium drought in southeast Australia (2001–2009): natural and human causes and implications for water resources, ecosystems, economy, and society. Water Resour. Res. 49, 1040–1057.
Van Loon, A.F., Gleeson, T., Clark, J., et al., 2016. Drought in the anthropocene. Nat. Geosci. 9, 89.
Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., 2010. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. J. Clim. 23, 1696–1718.
Vicente-Serrano, S.M., Begueria, S., Lorenzo-Lacruz, J., et al., 2012. Performance of drought indices for ecological, agricultural, and hydrological applications. Earth Interact. 16 (10), 1–27.
Vicente-Serrano, S.M., Quiring, S.M., Peña-Gallardo, M., et al., 2020. A review of environmental droughts: increased risk under global warming? Earth-Sci. Rev. 201, 102953. https://doi.org/10.1016/j.earscirev.2019.102953. In this issue.
Wang, D., Borthwick, A.G., He, H., Wang, Y., et al., 2018. A hybrid wavelet de-noising and Rank-Set Pair Analysis approach for forecasting hydro-meteorological time series. Environ. Res. 160, 269–281.
West, H., Quinn, N., Horswell, M., 2020. Remote sensing for drought monitoring & impact assessment: progress, past challenges and future opportunities. Rem. Sens. Environ. 232, 111291.
Williams, R.J., Zipser, D., 1989. A learning algorithm for continually running fully recurrent neural networks. Neural Comput. 1, 270–280.
Wittwer, G., 2020. Estimating the Regional Economic Impacts of the 2017 to 2019 Drought on NSW and the Rest of Australia. Victoria University, Centre of Policy Studies/IMPACT Centre, 2020.
Xiao, C., Chen, N., Hu, C., et al., 2019. Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost combination approach. Rem. Sens. Environ. 223, 111358.
Yihdego, Y., Vaheddoost, B., Al-Weshah, R.A., 2019. Drought indices and indicators revisited. Arab. J. Geosci. 12, 69.