Journal of Hydrology 615 (2022) 128706

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Dynamic-LSTM hybrid models to improve seasonal drought predictions over China

Zhiyong Wu [a,b,*], Hao Yin [a], Hai He [a], Yuan Li [a]

[a] College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China
[b] Yangtze Institute for Conservation and Development, Hohai University, Nanjing 210098, China

ARTICLE INFO

This manuscript was handled by Marco Borga, Editor-in-Chief

Keywords:
Seasonal drought prediction
Hybrid models
LSTM

ABSTRACT

Accurate drought prediction is essential for drought resilience and water resources management. The skill of seasonal drought prediction from dynamical and statistical models has limitations. This study combines dynamical models and deep learning to construct hybrid (dynamical-statistical) models. We use the random forest to extract typical grids based on the geopotential height, sea-level pressure, and 2-m temperature. The long short-term memory (LSTM) is used to construct the statistical models, with atmospheric variables as predictors and the 3-month Standardized Precipitation Index (SPI3) as the predictand. The hindcasts of atmospheric variables from the European Centre for Medium-Range Weather Forecasts (ECMWF) SEAS5 model are processed as predictors to force the statistical models. The hybrid model, constructed using the dynamical models and the LSTM, is named dynamic-LSTM ("D-LSTM"). The results suggest that the LSTM models are of information-extraction capability and robustness. When the lead time exceeds one month, the prediction skills are significantly improved by the D-LSTM models, especially in the East, Northwest, Southwest, and Tibet. In most regions, the D-LSTM models are more skillful across all seasons for lead times exceeding 30 days and are reliable in predicting droughts in spring and summer when the ECMWF SEAS5 loses skills at the seasonal scale. Furthermore, the D-LSTM models are more accurate in drought onset prediction.

1. Introduction

Droughts are natural disasters that occur under all climate regimes (Liu et al., 2016). Global chronically drought-prone areas have increased significantly, from 16.19 % between 1902 and 1949 to 41.09 % between 1950 and 2008 (Wang et al., 2014). According to the Food and Agriculture Organization (FAO) of the United Nations, developing countries suffered $29 billion in agriculture owing to drought disasters between 2005 and 2015 (Conforti et al., 2018). Therefore, accurate drought prediction is vital for improving drought resilience, managing water resources, and reducing drought losses.

With a better understanding of the climate system and improved computational capabilities, predicting droughts using general circulation models (GCMs) has become a fundamental approach (Hao et al., 2018). Seasonal drought indices, such as the Standardized Precipitation Index (SPI) and the Standardized Precipitation Evapotranspiration Index (SPEI), can be calculated from GCM precipitation and temperature outputs (Lang et al., 2020; Ma et al., 2015; Ujeneza and Abiodun, 2015). The advantage of GCMs is that they are based on physical principles and can capture the nonlinear changes in the ocean, atmosphere, and on land (Bader et al., 2008). Significant advances have been made in drought prediction in recent years, with the widespread use of post-processing and multi-model ensembles (Li et al., 2020; Mo and Lyon, 2015). Nevertheless GCMs are limited by chaotic properties, initial value errors, and model biases, which lead to uncertainties in the forecasted precipitation with a lead time of more than two weeks (Li et al., 2017; Merryfield et al., 2020). Statistical methods, less computationally demanding, usually first identify a range of predictors from historical hydroclimatic observations (reanalysis), which are then input into statistical models to predict drought. However, unclear physical mechanisms and unstable statistical relationships are always the problems of statistical methods (Hao et al., 2018). To combine the advantages of dynamical and statistical models, many researchers have attempted to construct hybrid (dynamical-statistical) models (Li et al., 2018a; Madhukumar et al., 2021).

Hybrid models mainly involve two steps, first, calibration to correct the bias (and ensemble spread) of GCMs forecasts, and second, merging forecasts from multiple sources (Madadgar et al., 2016; Ribeiro and

* Corresponding author at: College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China.
E-mail address: wzyhhu@gmail.com (Z. Wu).

https://doi.org/10.1016/j.jhydrol.2022.128706
Received 28 March 2022; Received in revised form 16 October 2022; Accepted 17 October 2022
Available online 9 November 2022

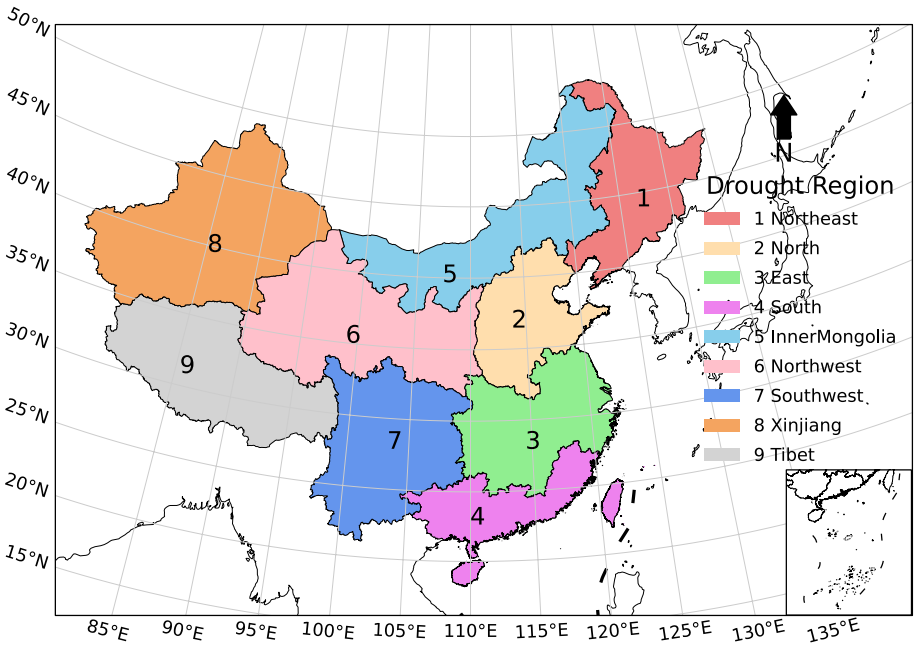

Fig. 1. 9 drought regions over China.

Pires, 2016; Yan et al., 2017). Previous efforts have focused on combining forecasts from multiple dynamical and statistical models, achieving excellent results (Hao et al., 2016; Madadgar et al., 2016; Wang et al., 2012). Nevertheless, some variables, such as geopotential height (gh) and temperature, yield higher skill scores with long lead times and are often used to analyze atmospheric circulation anomalies during droughts but can be used for prediction (Kuswanto et al., 2019; Liu et al., 2018). For seasonal drought, large-scale circulation is an important mechanism, and the role of atmospheric anomalies varies between droughts (Lhotka et al., 2020; Li et al., 2018b; Okumura et al., 2017; Zhuang et al., 2020). Thus, some studies have aimed to extract information from atmospheric circulation patterns to predict drought events. For example, Costa-Cabral et al. (2016) were able to project meteorological drought in California using the North Pacific High (NPH) sea-level pressure anomaly; in northeastern Europe, more than 65 % of drought events were detected one month in advance by Lavaysse et al. (2018); and Richardson et al. (2020) explored the potential benefits of using weather pattern predictions to improve sub-seasonal forecasts of meteorological drought. In these studies, selecting an appropriate method to obtain key information from atmospheric variables was critical.

With the explosive growth of observation (and reanalysis) data and the considerable increases in computing capabilities, climate prediction with deep learning (DL) has become an active area of research (Reichstein et al., 2019). The most significant advantage of DL is to mine information from "big data" for future prediction. In the past few years, DL has helped improve climate predictions significantly and assisted in interpreting the underlying mechanisms (Pan et al., 2020; Yoo-Geun et al., 2019). Two state-of-the-art, widely adopted architectures are convolutional neural network (CNN) for image tasks and recurrent neural network (RNN) for sequence learning tasks such as time series modeling (Shen, 2018). A CNN and transfer learning are used to predict the East Asian summer monsoon (EASM) in Tang and Duan (2021), and the anomaly correlation is higher than that of each of the current state-of-the-art dynamic models. Furthermore, Andersson et al. (2021) adopted the U-net, a CNN variant that takes images as input and produces images as output, to realize the seasonal Arctic sea ice probabilistic forecasting. As an improved model of ordinary RNNs for time series prediction, long short-term memory (LSTM) has been applied in drought prediction widely (Poornima and Pushpalatha, 2019; Xu et al., 2022; Zhang et al.). The previous studies based on LSTM are mainly through two methods. One is to use the drought indices (e.g., SPI, SPEI, etc.) or combine them with related hydro-meteorological variables such as temperature, humidity, and wind speed as predictors to derive the drought indices for the next moment or period (Poornima and Pushpalatha, 2019; Wu et al., 2021; Xu et al., 2022; Xu et al., 2018b). The second is to find the teleconnection between antecedent climatic drivers such as the Southern Oscillation Index (SOI), Pacific Decadal Oscillation Index (PDO), Indian Ocean Dipole (IOD), and drought indices for prediction at a future time (Dikshit and Pradhan, 2021; Dikshit et al., 2021a). Although the above studies have predicted drought indices at various time scales (e.g., SPEI1, SPEI3, SPEI12, etc.), the lead time of many studies is 1 month or 2 months. For example, the parameters from the past 19 months (also including the 20th-month parameters) were used to predict the 21st month's SPEI in Dikshit et al. (2021b); the lead time in Xu et al. (2022) is 1 month, and 1-to-2 month; Danandeh Mehr et al. (2022) forecasted the SPEI3 and SPEI6 over the lead time of one month. It may be related to the LSTM model construction. Specifically, the sequence length in LSTM determines the length of previously stored information, correspondingly in drought prediction, which is how long the past information is used to predict the current drought. As for LSTM, single-step prediction is easier and more accurate than multi-step prediction. Fortunately, Xu et al. (2018b) and Dikshit et al. (2021a) used recursive and sequence-to-series (seq2seq) methods to make drought predictions ranging from 1 month to 5 months and 1 month to 12 months lead times, respectively.

China has long been affected by drought due to its geographical location and climatic regime (Ayantobo et al., 2017), and in recent years, extreme droughts have become increasingly frequent and severe (Chen et al., 2018; Jin et al., 2013; Ma et al., 2020). While progress has been made in drought prediction in China using various models (Xu et al., 2018a; Xu et al., 2018b; Zhang et al., 2019; Zhu et al., 2020), there have been few attempts to employ DL to predict seasonal drought in all regions of China, for the reason that drought mechanisms vary between different regions and across spatial–temporal scales. In addition, few studies have input the contemporaneous atmospheric variables from the dynamical models to force the LSTM to make the daily prediction for the next three months. This method differs from Xu et al. (2018b) and Dikshit et al. (2021a) and matches the daily timestep used for the operational drought monitoring.

Table 1
Data used in this study.

Data type	Variable name	Pressure level (hPa)	Abbreviation	Data source	Unit
Atmospheric variables (resolution: 5°×5°, coverage: 10°S-90°N and 30°E-180°E)	2-m temperature Sea-level pressure Geopotential height	surface surface 200/500/850	t2m slp gh200/gh500/ gh850	**Reanalysis**: ERA5 reanalysis **Hindcast**:ECMWF SEAS5 (25 members)	K Pa m**2 s**-2
Precipitation (resolution 0.5°×0.5°, coverage: China)	Total precipitation	surface	tp	**Observation**: CMA **Hindcast**: ECMWF SEAS5 (25 members) /CFSv2 (4 members)/CMCC35 (50 members)	mm

Note. All the hindcast data used in this study are the ensemble average results of all members. That is, the outputs of SEAS5, CFSv2, and CMCC35 are obtained by averaging 25, 50, and 4 members, respectively.

In this study, we combine dynamical models and DL to construct hybrid models. We use the random forest (RF) to extract typical grids from atmospheric circulation variables. LSTM is used as the statistical model to find the relationships between atmospheric variables and meteorological drought. The atmospheric variables forecasted by the dynamical models are then processed to force the drought-prediction models. The remainder of this paper is structured as follows: Section 2 introduces the data and methodology. Section 3 presents the prediction results; we will discuss the study limitations and future work in Section 4. The conclusions are presented in Section 5.

2. Data and methodology

2.1. Data

2.1.1. Observations

Observed daily precipitation amounts from the Climate Data Centre (CDC) of the National Meteorological Information Centre, China Meteorological Administration (CMA), are used to provide 0.5° × 0.5° gridded data for the period 1962–2016 (https://data.cma.cn/data/detail/dataCode/SURF_CLI_CHN_PRE_DAY_GRID_0.5.html). The gridded precipitation data are area-averaged to regional precipitation over the 9 drought regions (Fig. 1). More details of the drought regions can be found in Wu et al. (2011).

2.1.2. GCM hindcast data

We focus on seasonal drought prediction for the next 90 days, corresponding to the "seasonal scale" discussed throughout this paper. For this, data for the atmospheric variables (in 90-day blocks) are downloaded from the ECMWF SEAS5 hindcast dataset for 1993–2016 (https://cds.climate.copernicus.eu/cdsapp#!/dataset/seasonal-original-single-levels), including gh200, gh500, gh850, t2m, and slp. For comparison, we select hindcast precipitation from the ECMWF SEAS5, NCEP coupled forecast system model version 2 (CFSv2), and Euro-Mediterranean Center for Climate Change - Seasonal Prediction System, Version 3.5 (CMCC35) (Table 1) during the same period over China from the CDS (https://cds.climate.copernicus.eu/cdsapp#!/dataset/seasonal-original-single-levels).

2.2. Methodology

2.2.1. Daily-updated 3-month standardized precipitation Index (SPI3)

Since rainfall varies significantly in different regions, the concept of drought varies. In order to assess drought better, the World Meteorological Organization (WMO) recommends using the SPI to monitor the drought. SPI is a standardized index that reflects the probability of given observed precipitation relative to the long-term climatic conditions of the area. Negative and positive SPI values represent precipitation deficit and surplus, respectively. SPI can be applied to different time scales (e.g., 3, 6, and 12 months) to analyze the impact of drought on various water resource demands.

We select SPI3 as the seasonal drought index in this study, calculated for the 9 regions using area-averaged precipitation. Traditionally, the SPI3 varies on a monthly scale, reflecting cumulative precipitation over the past three months; however, a monthly timescale does not sufficiently reflect the evolution of drought onset, persistence, and relief, presenting limitations for meeting the requirements for real-time monitoring and prediction for effective management. Therefore, we follow the WMO recommendation to adjust the sliding time window of the SPI3 to 90 days (Svoboda et al., 2012). Thus, we obtain information on the daily variation of meteorological droughts while the mathematical treatment remained unchanged. Full calculation details are given by Liu et al. (2018).

2.2.2. Predictors construction

Hart and Grumm (2001) first used standardized anomalies to

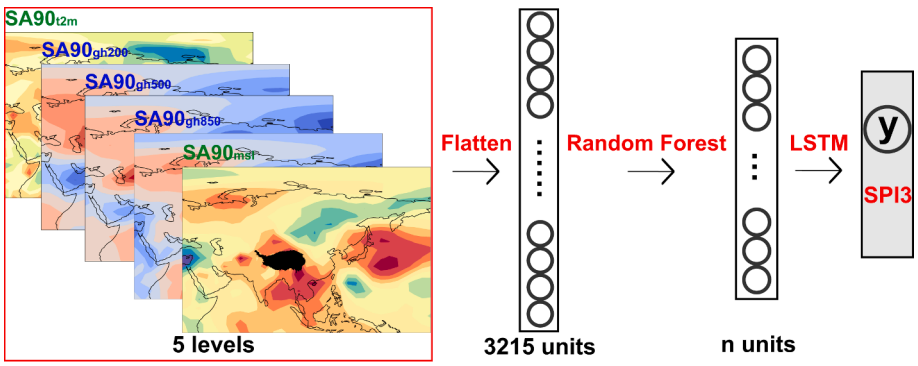

Fig. 2. Model structure. $SA90_{gh200}$, $SA90_{gh500}$, $SA90_{gh850}$, $SA90_{t2m}$, and $SA90_{slp}$ are the 90-day standardized anomalies of gh200, gh500, gh850, t2m, and slp, respectively. The grids over Tibetan Plateau on gh850 and slp are masked, so the total number of flattened units is 3215.

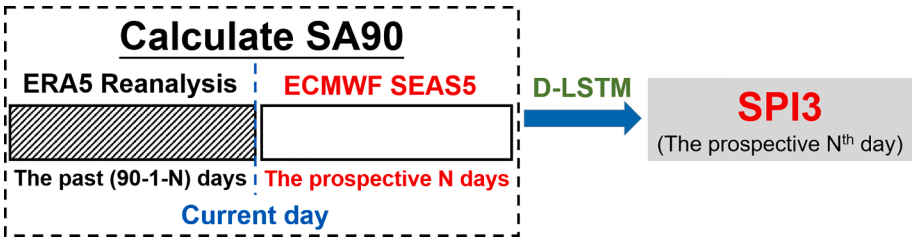

Fig. 3. Schematic representation of the calculation process for the prospective Nth-day SPI3. N represents the prospective Nth day, changing with the lead time.

delineate weather-scale events, and this approach has been carried out in weather analysis and forecasting (Duan et al., 2014; Liu et al., 2017). Evidence from these studies has demonstrated the superiority of anomaly in many aspects (Qian et al., 2021). Here, we calculate 90-day standardized anomalies (SA90) using the following equation:

$$\mathrm{SA90} = \frac{X_{90} - \mu_{90}}{\sigma_{90}} \quad (1)$$

Where X_{90} represents the original variable, and μ_{90} and σ_{90} are the climatological mean and standard deviation, respectively. Explicitly speaking, X_{90} is the 90-day running mean, whose value is located on the last day of the running window. The two climatological variables (μ_{90}, σ_{90}) are further computed based on the 90-day running mean. The climatic background period in this study is the 30 years before each hindcast year, so the background is sliding. For example, X_{90} for 1 April 2000 represented the 90-day average for the period from 3 January 2000 to 1 April 2000; and μ_{90} and σ_{90} are the 90-day averages and standard deviation for the period from 3 January 1970 to 1 April 1999. So we obtain the SA90 for all considered atmospheric variables (i.e., $\mathrm{SA90_{gh200}}$, $\mathrm{SA90_{gh500}}$, $\mathrm{SA90_{gh850}}$, $\mathrm{SA90_{t2m}}$, and $\mathrm{SA90_{slp}}$).

2.2.3. Drought prediction model

We construct one LSTM model for each region using 22-year data for training, 8-year data for validation, and 1-year data for test. There are a total of 9 × 24 models for 9 regions in 24 hindcast years (1993–2016). Each model adopts the daily 5-layer (gh200, gh500, gh850, t2m, slp) SA90 dataset as the predictors and the SPI3 of the day as the predictand. The grids over the Tibetan Plateau on gh850 and slp are masked before constructing the model. Over and near large orography, slp, and gh850 readings are unphysical. The readings underneath the orography are the results of extrapolation from dynamical models and hence are artificial.

The model structure is shown in Fig. 2. First, we flatten the five-layer SA90 into a column of 3215 data points (the input neurons), extract typical data through the RF, and then produce the output SPI3 data by forcing the LSTM. The details of the RF and LSTM training, validation, and test stages are in Supplements 3–5.

The models constructed above incorporate the contemporaneous statistical relationships between the predictors and the SPI3 but cannot predict the future. During the training, validation, and test periods, the models are forced with ERA5 data. For prediction, we first calculate SA90 using ERA5 data for the past (90-1-N) days and the current day, ECMWF SEAS5 outputs for the prospective N days, and then force the LSTM models to predict the Nth-day SPI3, as shown in Fig. 3. Therefore, the drought prediction models we construct are hybrid (dynamical-statistical) models, which we name dynamic-LSTM ("D-LSTM").

2.2.4. GCM hindcast precipitation

The SPI3 calculated from the SEAS5, CFSv2, and CMCC35 hindcast precipitation datasets are compared with the hybrid models. For prediction, we use observed precipitation for the past (90-1-N) days and the current day, GCM outputs for the prospective N days, and then calculate the Nth-day SPI3.

2.2.5. Skill scores

Root-mean-square error (RMSE) is adopted as the loss function to describe the error between the model output and the observed data, and for the inverse calculation of the LSTM models, as follows:

Table 2
SPI3 simulation skills of LSTM for the training/validation/test Set.

		Training				Validation				Test			
		$\mathrm{LSTM_1}$	……	$\mathrm{LSTM_{24}}$	$\mathrm{LSTM_{mean}}$	$\mathrm{LSTM_1}$	……	$\mathrm{LSTM_{24}}$	$\mathrm{LSTM_{mean}}$	$\mathrm{LSTM_1}$	……	$\mathrm{LSTM_{24}}$	$\mathrm{LSTM_{mean}}$
Northeast	R	0.90	……	0.93	0.91	0.83	……	0.82	**0.85**	0.79	……	0.93	0.82
	RMSE	0.55		0.64	0.50	0.54		0.61	0.58	0.43		1.08	0.71
North	R	0.90		0.93	0.91	0.83		0.82	**0.85**	0.79		0.93	0.82
	RMSE	0.45		0.33	0.46	0.57		0.51	0.55	0.46		0.57	**0.59**
East	R	0.92		0.89	**0.92**	0.86		0.83	0.83	0.88		0.71	0.72
	RMSE	0.47		0.59	**0.44**	0.47		0.56	**0.53**	0.37		1.41	0.67
South	R	0.81		0.87	0.88	0.78		0.78	0.77	0.87		0.81	0.81
	RMSE	0.63		0.75	0.48	0.65		0.63	0.64	0.35		1.46	0.72
InnerMongolia	R	0.79		0.90	0.87	0.81		0.76	0.83	0.89		0.85	**0.85**
	RMSE	0.56		0.57	0.53	0.57		0.65	0.62	0.76		1.00	**0.75**
Northwest	R	0.91		0.89	0.89	0.91		0.84	0.81	0.93		0.50	0.73
	RMSE	0.42		0.43	0.49	0.49		0.57	0.59	0.41		1.11	**0.75**
Southwest	R	0.91		0.88	**0.86**	0.79		0.87	0.81	0.70		0.93	0.77
	RMSE	0.42		0.44	**0.57**	0.61		0.60	0.60	0.79		0.86	**0.75**
Tibet	R	0.80		0.92	**0.86**	0.71		0.83	0.79	0.92		0.83	0.77
	RMSE	0.58		0.39	**0.57**	0.66		0.68	0.64	0.33		0.46	0.70
Xinjiang	R	0.88		0.85	0.88	0.87		0.71	**0.74**	0.73		0.78	**0.65**
	RMSE	0.50		0.57	0.55	0.61		0.77	**0.66**	0.51		0.82	**0.75**

Note. LSTM = long short-term memory; R = correlation coefficient; RMSE = root mean square error. LSTM1 indicates the first LSTM (1963–1984 for training, 1985–1992 for validation, and 1993 for test); LSTM2 indicates the second LSTM (1964–1985 for training, 1986–1993 for validation, and 1994 for test), and so on; LSTMmean indicate the mean values of the 24 models.
The bold and underlined red (black) values indicate the best (worst) statistics for the corresponding dataset.

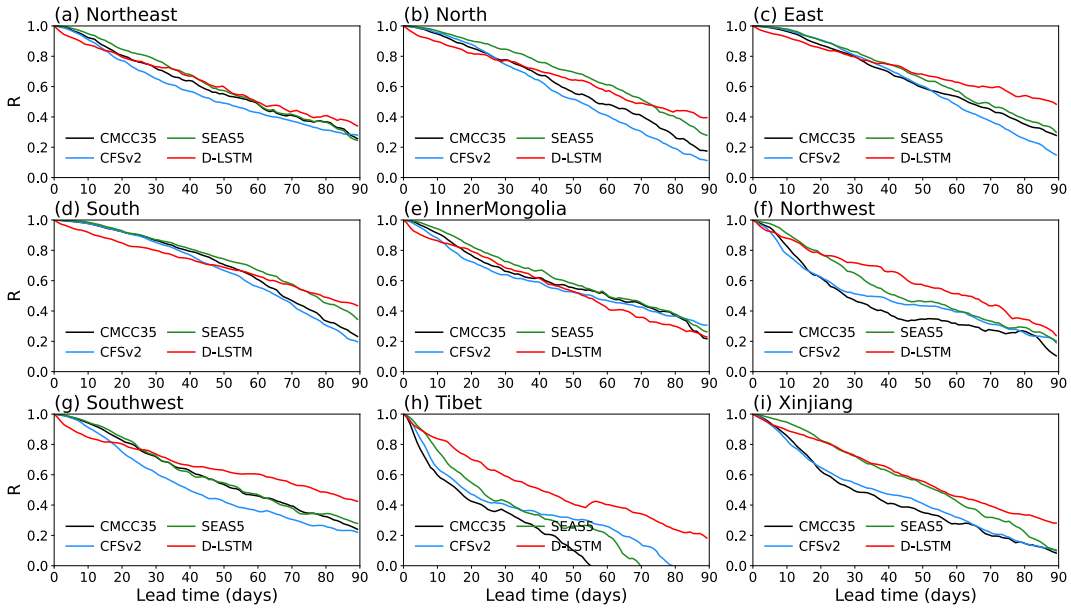

Fig. 4. Correlation between predicted SPI3 and observed values for 4 models at different lead times in 9 regions. Black, blue, green, and red curves represent the CMCC-SPS3.5, CFSv2, ECMWF SEAS5, and D-LSTM outputs. The correlation coefficients all pass the 99% significance test. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - x_i)^2}{n}} \tag{2}$$

where x_i represents the observed value, y_i represents the predictand, and n represents the sample size.

The correlation coefficient (R) is used as the performance evaluation function for the LSTM to measure the correlation between the model output and the expected values:

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

where x_i is the observed value, y_i is the predictand, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ (the sample mean), and analogously for $\bar{y}$; and n represents the sample size.

3. Results

3.1. Simulated SPI3 in training, validation, and test periods

As introduced previously in Sect. 2.2.3, we construct models based on 22-year training data, 8-year validation data, and 1-year test data. As shown by the two skill scores in Table 2, the mean values of the R are all over 0.86, and the averages of the RMSE are all below 0.57 in the training set, which indicates that the LSTM can extract information from atmospheric variables and capture the nonlinear atmospheric processes. In addition, the two skill scores are slightly worse in the validation and test set, which means the robustness of the models. However, the LSTM models perform much better in the North and East than in the Xinjiang and Tibet, which may reflect the different drought mechanisms in these regions.

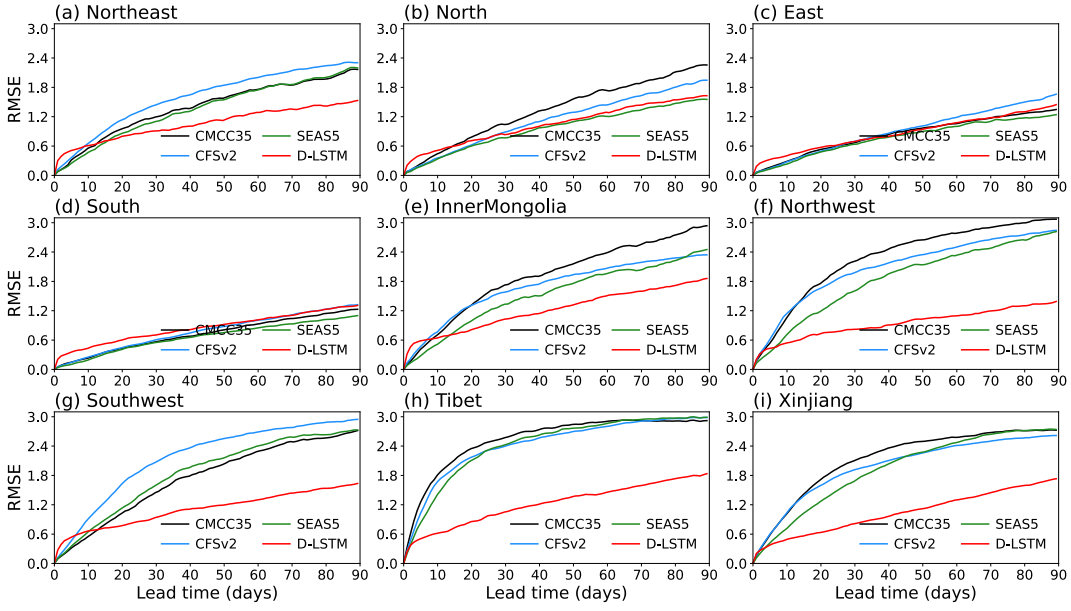

Fig. 5. The same as Fig. 4, but for RMSE.

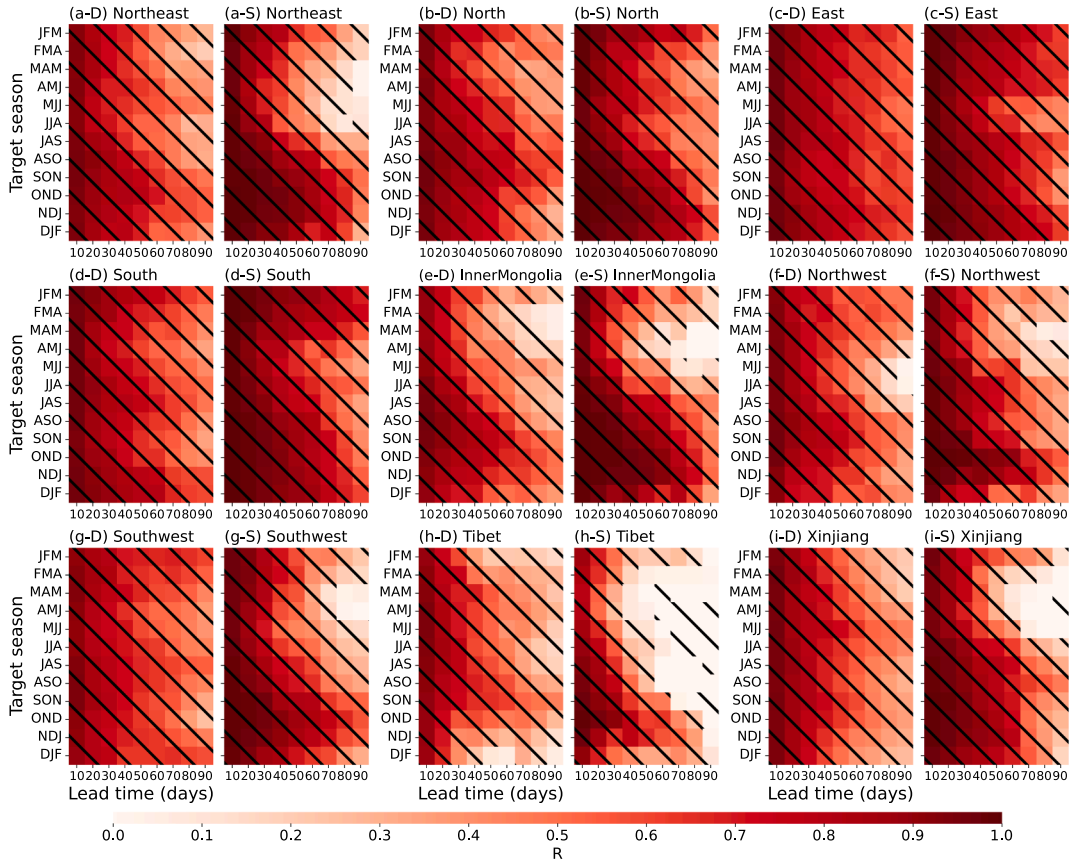

Fig. 6. Comparison of the correlation of the D-LSTM and the SEAS5 models in different target seasons in the 9 regions. The title "(a-D) Northeast" represents the Fig. a of the D-LSTM in the Northeast; the title '(a-S) Northeast' represents the Fig. a of the SEAS5 in the Northeast, and so on. The shaded areas represent the correlation coefficients that pass the 99 % significance test. Horizontal coordinates indicate different lead times (days); the first 10 days include the predictions of the days 1 to 10, the second 10 days include days 11 to 20, and so on. Vertical coordinates are target seasons. JFM represents January, February, and March; FMA represents February, March, April, etc.

3.2. Seasonal drought prediction

Between 1993 and 2016, ECMWF SEAS5 hindcasts data for the first day of each month, as do the D-LSTM models. When we predict the SPI3 in each year, the LSTM model tested in the year is used. Finally, we generated 4 sets of seasonal drought prediction products (SPI3) based on the CMCC35, CFSv2, SEAS5, and D-LSTM.

Fig. 4 and Fig. 5 show the correlation and RMSE results of the SPI3 between 1993 and 2016. The strength of the skill scores of all 4 models decreases as the lead time increases. At one month or less, the prediction skills of the SEAS5 outputs are stronger than that of the D-LSTM models, indicating that dynamic models are more accurate in simulating precipitation with short lead times, which is consistent with other analyses (Lang et al., 2014; Li et al., 2017). However, when the lead time exceeds 30 days, the D-LSTM models outperform the dynamic models in most regions. The RMSE has apparent differences between the four models for the two skill scores in all regions. Maybe the reasons are that we choose RMSE as the loss function to train the LSTM models and the four models have varying skills for the regions. While in the North, South, and Inner Mongolia, the prediction skills of the D-LSTM models are almost the same as the dynamical models. We also find that the skill scores are different between the regions. For example, the outputs from both the dynamical and the D-LSTM models in the Tibet and Xinjiang are worse than the others. Contrary to this, the dynamical and the D-LSTM models both perform well in the East and South.

3.3. Prediction skills across all seasons

According to Fig. 4 and Fig. 5, the SEAS5 models are the best for seasonal drought prediction among the three dynamical models. So we compare the prediction skills of the SEAS5 and the D-LSTM for all target seasons, and the results are shown in Fig. 6 and Fig. 7. Both models underperform in the first half of the year compared with the second half. The D-LSTM models perform much better for summer than the SEAS5 in the Northeast, Southwest, and Tibet, but does not perform well for the winter. Nevertheless, the D-LSTM models are more skillful across all seasons in other regions for lead times exceeding 30 days. In addition, the D-LSTM models are still reliable in spring and summer, when the SEAS5 often loses skills.

3.4. Drought onset prediction

For seasonal drought prediction, the capability of the models to determine whether drought will occur in the future is essential. Here, given a SPI3 time series, the drought condition is defined as the SPI3 below a specified threshold (−0.8 in this study, according to Ma et al. (2015) and Svoboda et al. (2002)). If the daily-observed SPI3 and predicted SPI3 on the same day are both below −0.8, the drought prediction is correct. So we calculate the accuracy of drought prediction as follows:

$$\text{Accuracy} = \frac{n}{N} \tag{4}$$

where N represents the total number of the observed drought days (SPI3 < -0.8) at a specific lead time (1st, 2nd, ⋯, or the 90th day), and n

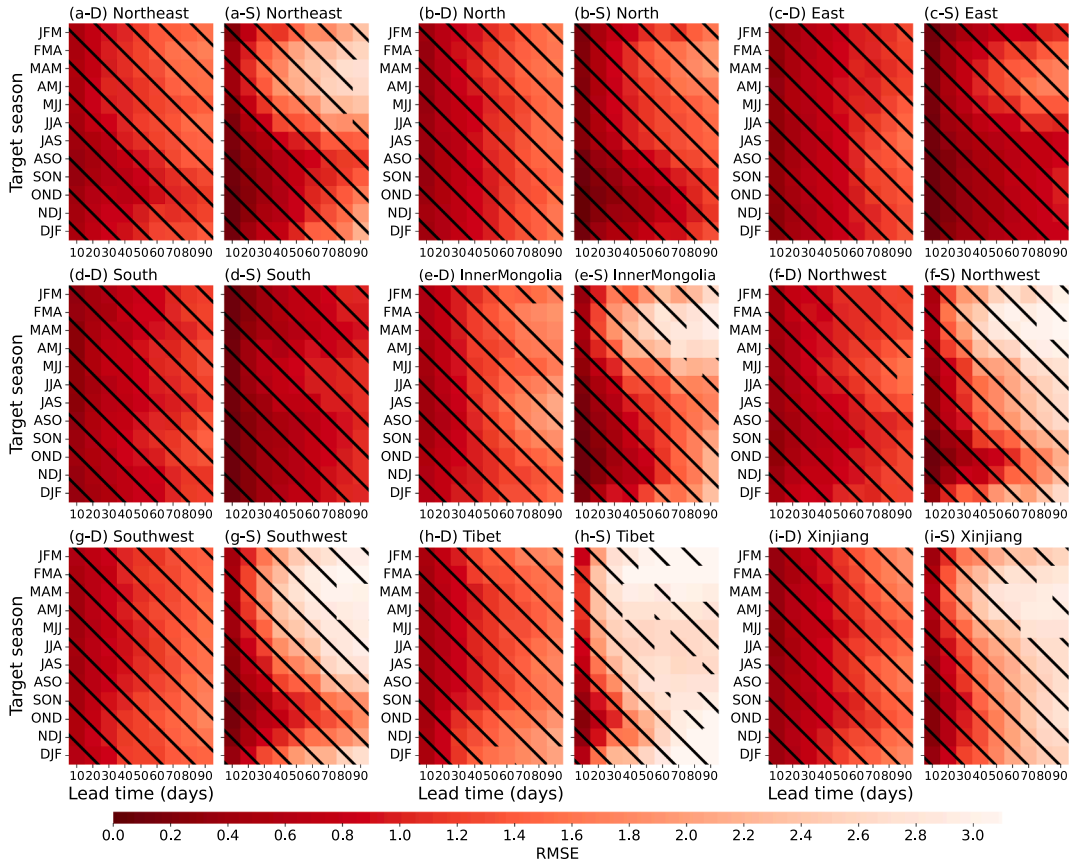

Fig. 7. The same as Fig. 6, but for RMSE.

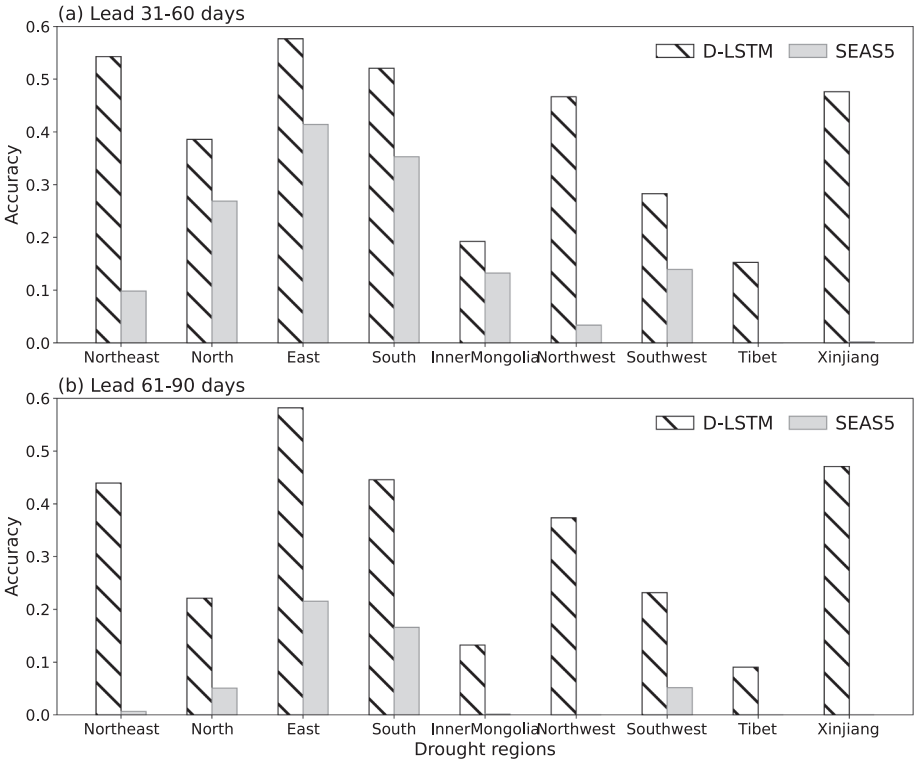

Fig. 8. Bar charts of the accuracy of drought prediction in 9 regions at different lead times. a (b) includes the predictions of the days 31 to 60 (61–90). The hatched (grey) bars represent the predictions of the D-LSTM (SEAS5).

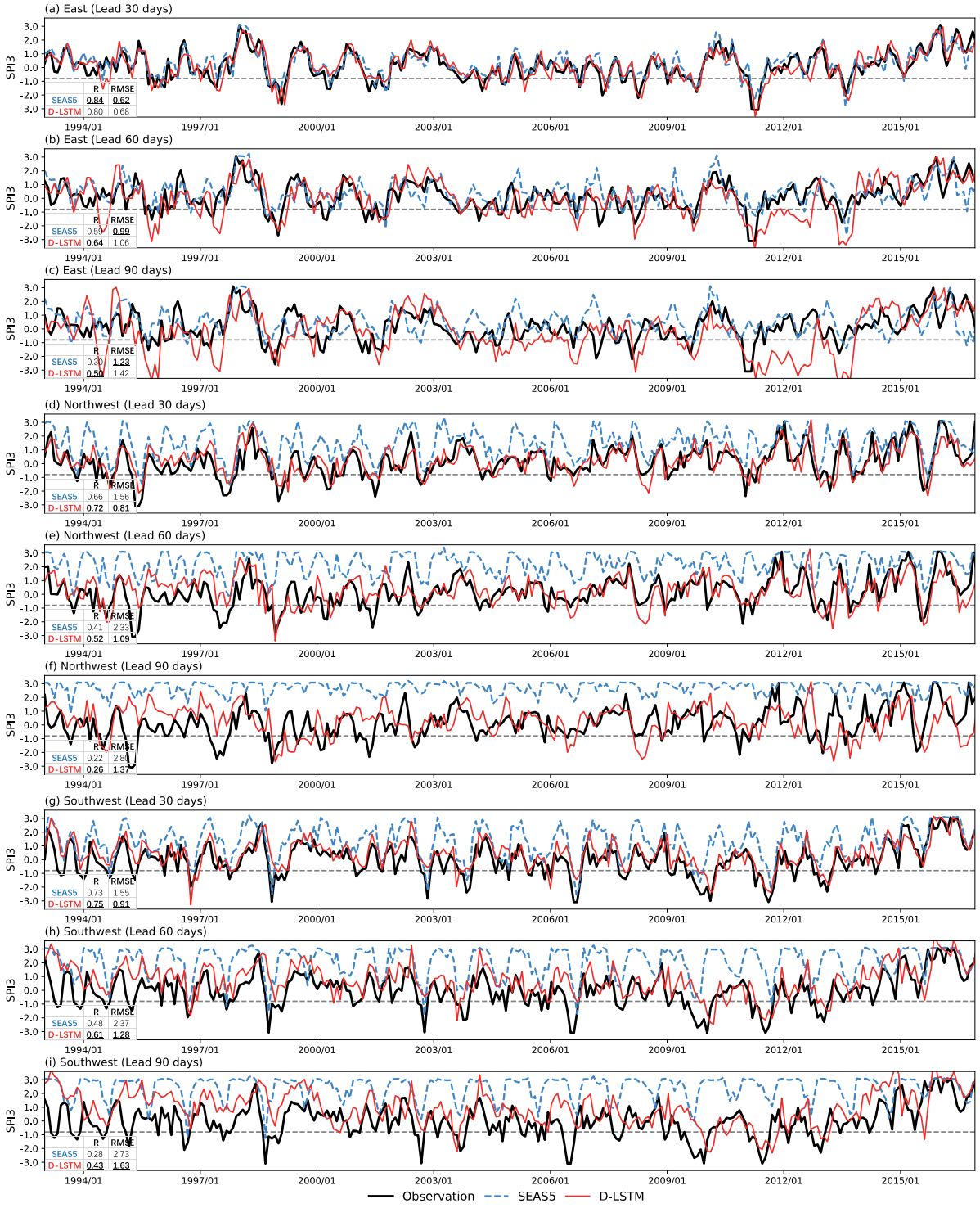

Fig. 9. The predicted SPI3 of the D-LSTM and the SEAS5 in the East, Northwest, and Southwest at different lead times. The black, blue, and red curves represent the observation, the SEAS5, and the D-LSTM, respectively. The grey dotted curves represent the SPI3 = −0.8. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

represents the total number of the days correctly predicted.

From Fig. 8, it is clear that the D-LSTM models outperform the SEAS5 in predicting droughts when the lead time is over 30 days, especially after 60 days, which means the D-LSTM allows for better prediction of drought onset at the seasonal scale. Compared to SEAS5, the decrease in accuracy of the D-LSTM models is smaller when the lead time is extended from 31–60 days to 61–90 days. These results are consistent with the low RMSE of the D-LSTM models, as shown in Fig. 5.

To find out why the D-LSTM models are more accurate in predicting drought onset, especially in the East, Northwest, and Southwest, we plot the time series of the predicted SPI values on the 30th, 60th, and 90th lead days, respectively, shown in Fig. 9. The outputs from the D-LSTM always match the observed time series as the lead time increases. However, the SEAS5 tends to predict more precipitation when the lead time is over 30 days, contributing to positive values of SPI3. As a result, it is hard for the SEAS5 to determine the occurrence of droughts at the

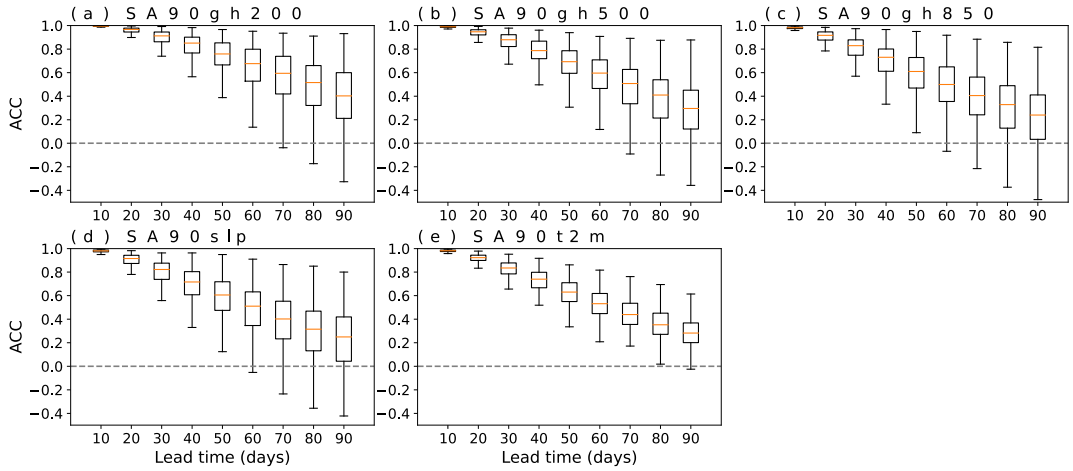

Fig. 10. Boxplots of ACC for all atmospheric variables (30°–180° E, 10°S–90° N) with increasing lead times. A box plot's top to bottom five lines represent the minimum, first quartile, median, third quartile, and maximum data values, respectively.

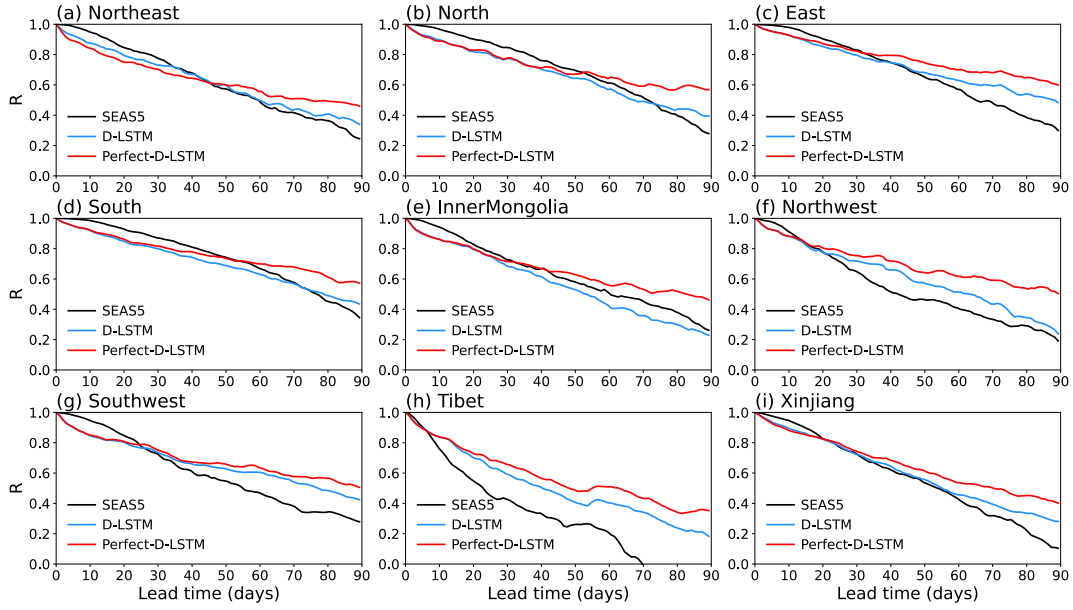

Fig. 11. The correlation of 3 models predicting the SPI3 with different lead times in all regions. The black, blue, and red curves represent SEAS5, D-LSTM, and Perfect-D-LSTM predictions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

seasonal scale.

4. Discussion

4.1. Impact of forecast error from the ECMWF SEAS5 dataset

4.1.1. Anomaly correlation coefficients (ACC) of forecasted atmospheric variables

We use the hindcast atmospheric circulation dataset from the SEAS5 as predictors to force the D-LSTM. Therefore, the models contain two kinds of error, first from the LSTM and second from the GCM. This section focuses on the impact of the second error source. Fig. 10 shows that the ACC of all variables decreases with increasing lead time, which indicates that the error from the SEAS5 dataset introduces lots of uncertainties to the prediction.

4.1.2. Perfect drought prediction

We replace all atmospheric data for the prospective 90 days with reanalysis data to re-force the D-LSTM models, which we name 'Perfect-D-LSTM' (Fig. 11 and Fig. 12). The correlation and RMSE of the Perfect-D-LSTM are all improved in all regions. Actually, the performances of the perfect models indicate the upper limit of prediction abilities of the LSTM for the reason that the inputs are reanalysis data. Therefore, the red curves are consistent with the simulation skills of the LSTM in Table 2. Post-processing GCM precipitation outputs have also been included in the dynamical-statistical prediction methods (Schepen et al., 2016), so it is necessary to correct the forecasted atmospheric variables accordingly.

4.2. Comparison between the LSTM model and other machine learning models

As we know, the artificial neural network (ANN) is one of the classic machine learning models. Many researchers have used this model to improve precipitation and drought prediction (Belayneh et al., 2014; da Paz et al., 2011; Sigaroodi et al., 2014). Can the same results be obtained by replacing LSTM with ANN according to the same processing steps for our study? Furthermore, since the LSTM model is an ANN, we take linear

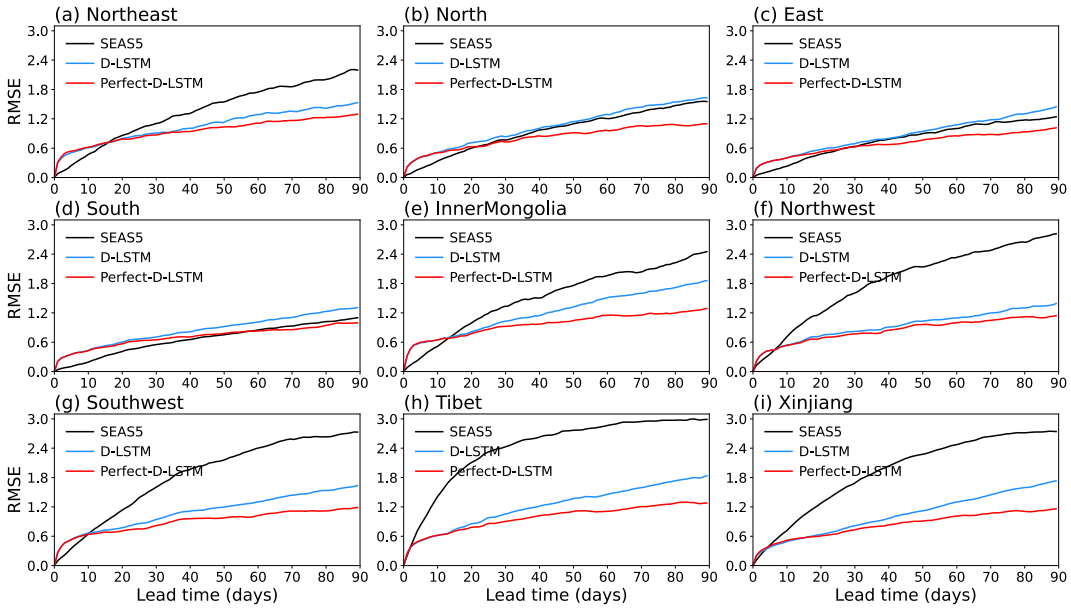

Fig. 12. The same as Fig. 11, but for RMSE.

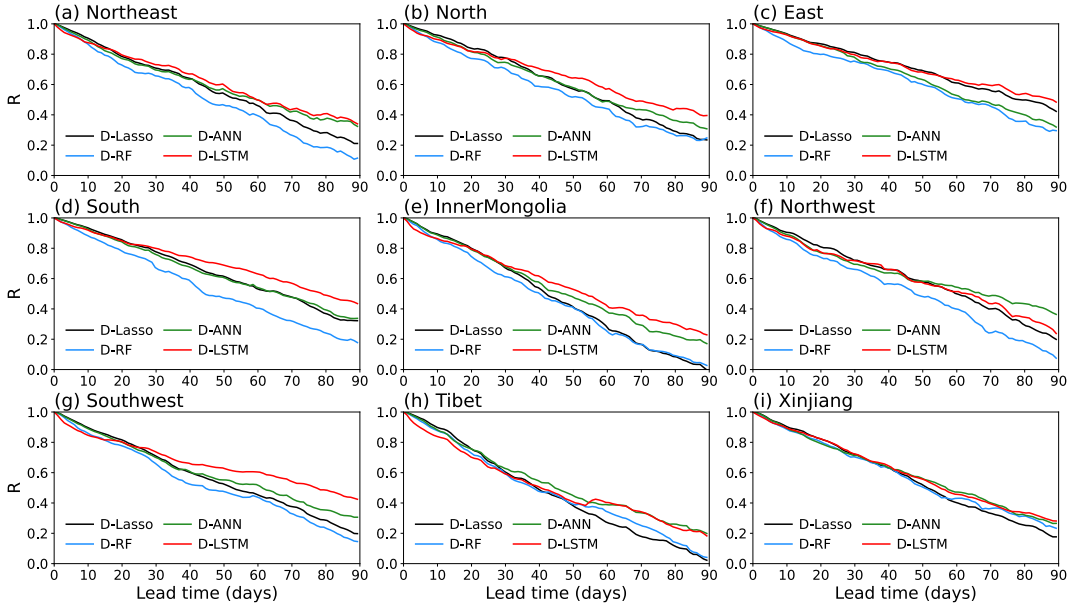

Fig. 13. Comparison of the correlation between the D-Lasso, D-RF, D-ANN, and D-LSTM models.

models and tree-based models into consideration. We choose Least Absolute Shrinkage and Selection Operator (Lasso) and Random Forest Regressor in this study. Following section 2.2.3, we name the models "D-ANN", "D-Lasso", and "D-RF", respectively.

From Fig. 13, the D-LSTM models have better or the equivalent performance than the other models in most regions, except the Northwest and Xinjiang. The RMSE shown in Fig. 14 indicates that the D-LSTM models are more skillful. It is obvious that D-LSTM models make slight improvements in predicting SPI3 compared to the other models. Actually, in machine learning, a little increase in accuracy is often an essential improvement for machine learning models. Another point of view is that data and features determine the upper limit of machine learning performances, and various models/algorithms are used to approach this upper limit (Zheng et al., 2021).

5. Conclusions

Effective drought prediction is essential for drought resilience and water resources management. Nevertheless, the dynamic and traditional statistical models have limitations in seasonal drought prediction. This study combines the ECMWF SEAS5 and the LSTM to construct drought-prediction models over China at the seasonal scale. The capability and robustness of the LSTM models to extract information between atmospheric variables and meteorological drought is evaluated, and the extent to which the prediction skills could be improved and the effectiveness of predicting drought onset are investigated.

The results suggest that the accuracy of the simulated SPI3 by the LSTM models in training, validation, and test set using ERA5 reanalysis data is high. Both the correlation and RMSE are significantly improved by the D-LSTM when the lead time exceeds 30 days over China. In most regions, the D-LSTM models are more skillful across all seasons and

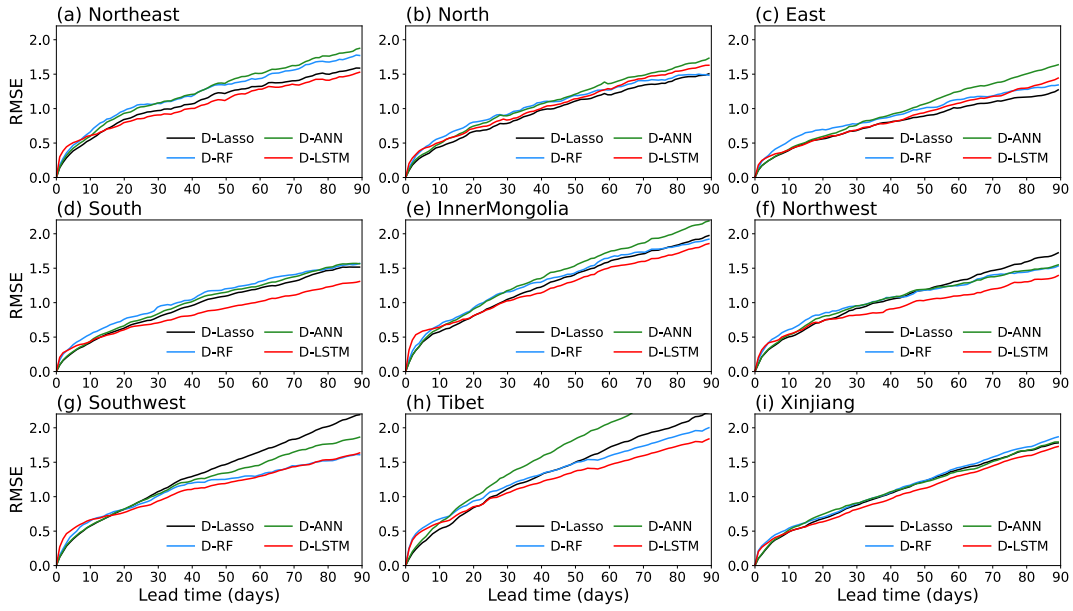

Fig. 14. The same as Fig. 13, but for RMSE.

perform well in spring and summer when the SEAS5 often loses predictability. In addition, the D-LSTM models have better skills in predicting the drought onset at the seasonal scale.

In the future, the post-processing of the forecasted atmospheric variables will be considered. We should note that the drought-prediction models constructed in this study cannot predict the spatial distribution of the SPI3 because the precipitation data are processed as the regional average. To reflect more details of the drought, other deep learning models, such as the Conv-LSTM or U-net model, are prospected to solve this problem. Last but not least, we investigate the capability of the D-LSTM models in predicting the drought onset by transforming the numerical SPI3 trajectories into a categorical series using −0.8 as a threshold in Section 3.4. We did not consider the opportunity of constructing a classification model to perform this task rather than post-processing the outputs of the D-LSTM in a regression model. Many articles have focused on drought class prediction with excellent results (Moreira et al., 2018; Moreira, 2016; Moreira et al., 2008; Niaz et al., 2021; Paulo and Pereira, 2007). Therefore, it is expected that we will try to train a deep learning classifier which may generate higher performances in future researches.

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.

Data availability

Data will be made available on request.

Acknowledgments

This work is supported by NSFC-MWR-CTGC Joint Yangtze River Water Science Research Project (No. U2240225); National Key Research and Development Program of China (No. 2017YFC1502403).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2022.128706.

References

Andersson, T.R., et al., 2021. Seasonal Arctic sea ice forecasting with probabilistic deep learning. Nat. Commun. 12 (1), 1–12.
Ayantobo, O.O., Li, Y., Song, S., Yao, N., 2017. Spatial comparability of drought characteristics and related return periods in mainland China over 1961–2013. J. Hydrol. 550, 549–567. https://doi.org/10.1016/j.jhydrol.2017.05.019.
Bader, D. et al., 2008. Climate models: an assessment of strengths and limitations.
Belayneh, A., Adamowski, J., Khalil, B., Ozga-Zielinski, B., 2014. Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. J. Hydrol. 508, 418–429. https://doi.org/10.1016/j.jhydrol.2013.10.052.
Chen, S., Zhang, L., Liu, X., Guo, M., She, D., 2018. The use of SPEI and TVDI to assess temporal-spatial variations in drought conditions in the middle and lower reaches of the Yangtze River Basin, China. Adv. Meteorol. 2018, 9362041. https://doi.org/10.1155/2018/9362041.
Conforti, P., Ahmed, S., Markova, G., 2018. Impact of disasters and crises on agriculture and food security, 2017. http://www.fao.org/3/I8656EN/i8656en.pdf.
Costa-Cabral, M., et al., 2016. Projecting and forecasting winter precipitation extremes and meteorological drought in California using the North Pacific high sea level pressure anomaly. J. Clim. 29 (13), 5009–5026. https://doi.org/10.1175/JCLI-D-15-0525.1.
da Paz, A.R., Uvo, C., Bravo, J., Collischonn, W., da Rocha, H.R., 2011. Seasonal precipitation forecast based on artificial neural networks, Computational methods for agricultural research: Advances and applications. IGI Global 326–354. https://doi.org/10.4018/978-1-61692-871-1.ch016.
Danandeh Mehr, A., Rikhtehgar Ghiasi, A., Yaseen, Z.M., Sorman, A.U., Abualigah, L., 2022. A novel intelligent deep learning predictive model for meteorological drought forecasting. J. Ambient Intell. Hum. Comput. 1–15.
Dikshit, A., Pradhan, B., 2021. Interpretable and explainable AI (XAI) model for spatial drought prediction. Sci. Total Environ. 801, 149797.
Dikshit, A., Pradhan, B., Alamri, A.M., 2021a. Long lead time drought forecasting using lagged climate variables and a stacked long short-term memory model. Sci. Total Environ. 755, 142638.
Dikshit, A., Pradhan, B., Huete, A., 2021b. An improved SPEI drought forecasting approach using the long short-term memory neural network. J. Environ. Manage. 283, 111979.
Duan, W., et al., 2014. Anomalous atmospheric events leading to Kyushu's flash floods, July 11–14, 2012. Nat. Hazards 73 (3), 1255–1267.
Hao, Z., et al., 2016. A statistical method for categorical drought prediction based on NLDAS-2. J. Appl. Meteorol. Climatol. 55 (4), 1049–1061. https://doi.org/10.1175/JAMC-D-15-0200.1.
Hao, Z., Singh, V.P., Xia, Y., 2018. Seasonal drought prediction: advances, challenges, and future prospects. Rev. Geophys. 56 (1), 108–141. https://doi.org/10.1002/2016RG000549.
Hart, R.E., Grumm, R.H., 2001. Using normalized climatological anomalies to rank synoptic-scale events objectively. Mon. Weather Rev. 129 (9), 2426–2442. https://doi.org/10.1175/1520-0493(2001)129<2426:UNCATR>2.0.CO;2.
Jin, D., Guan, Z., Tang, W., 2013. The extreme drought event during winter-spring of 2011 in East China: combined influences of teleconnection in midhigh latitudes and thermal forcing in maritime continent region. J. Clim. 26 (20), 8210–8222. https://doi.org/10.1175/JCLI-D-12-00652.1.

Kuswanto, H., Yuliatin, I.L., Khoiri, H.A., 2019. Statistical downscaling to predict drought events using high resolution satellite based geopotential data. IOP Publishing 052040. https://doi.org/10.1088/1757-899X/546/5/052040.
Lang, Y., et al., 2014. Evaluating skill of seasonal precipitation and temperature predictions of NCEP CFSv2 forecasts over 17 hydroclimatic regions in China. J. Hydrometeorol. 15 (4), 1546–1559. https://doi.org/10.1175/JHM-D-13-0208.1.
Lang, Y., Luo, L., Ye, A., Duan, Q., 2020. Do CFSv2 seasonal forecasts help improve the forecast of meteorological drought over mainland China? Water 12 (7), 2010. https://doi.org/10.3390/w12072010.
Lavaysse, C., Vogt, J., Toreti, A., Carrera, M.L., Pappenberger, F., 2018. On the use of weather regimes to forecast meteorological drought over Europe. Nat. Hazards Earth Syst. Sci. 18 (12), 3297–3309. https://doi.org/10.5194/nhess-18-3297-2018.
Lhotka, O. et al., 2020. Atmospheric circulation as a factor contributing to increasing drought severity in central Europe. J. Geophys. Res.: Atmos., 125(18): e2019JD032269.
Li, Y., et al., 2020. Post-processing sub-seasonal precipitation forecasts at various spatiotemporal scales across China during boreal summer monsoon. J. Hydrol. 125742. https://doi.org/10.1016/j.jhydrol.2020.125742.
Li, X., Li, D., Li, X., Chen, L., 2018b. Prolonged seasonal drought events over northern China and their possible causes. Int. J. Climatol. 38 (13), 4802–4817. https://doi.org/10.1002/joc.5697.
Li, G., Ma, X., Yang, H., 2018a. A hybrid model for monthly precipitation time series forecasting based on variational mode decomposition with extreme learning machine. Information 9 (7). https://doi.org/10.3390/info9070177.
Li, Q., Yang, S., Wu, T., Liu, X., 2017. Subseasonal dynamical prediction of East Asian cold surges. Weather Forecast. 32 (4) https://doi.org/10.1175/WAF-D-16-0209.1.
Liu, Z., Törnros, T., Menzel, L., 2016. A probabilistic prediction network for hydrological drought identification and environmental flow assessment. Water Resour. Res. 52 (8), 6243–6262.
Liu, Z., Lu, G., He, H., Wu, Z., He, J., 2017. Anomalous features of water vapor transport during severe summer and early fall droughts in southwest China. Water 9 (4), 244.
Liu, Z., Lu, G., He, H., He, J., 2018. A conceptual prediction model for seasonal drought processes using atmospheric and oceanic standardized anomalies: application to regional drought processes in China. Hydrol. Earth Syst. Sci. 22 (1), 529–546. https://doi.org/10.5194/hess-22-529-2018.
Ma, F., Yuan, X., ye, A., 2015. Seasonal drought predictability and forecast skill over China. J. Geophys. Res.: Atmos. 120, n/a-n/a. https://doi.org/10.1002/2015JD023185.
Ma, S., Zhu, C., Liu, J., 2020. Combined impacts of warm central equatorial Pacific Sea surface temperatures and anthropogenic warming on the 2019 severe drought in East China. Adv. Atmos. Sci. 37 (11), 1149–1163. https://doi.org/10.1007/s00376-020-0077-8.
Madadgar, S., et al., 2016. A hybrid statistical-dynamical framework for meteorological drought prediction: Application to the southwestern United States. Water Resour. Res. 52 (7), 5095–5110. https://doi.org/10.1002/2015WR018547.
Madhukumar, N., Wang, E., Zhang, Y.F., Xiang, W., 2021. Consensus forecast of rainfall using hybrid climate learning model. IEEE Internet Things J. 8 (9), 7270–7278. https://doi.org/10.1109/JIOT.2020.3040736.
Merryfield, W.J., et al., 2020. Current and emerging developments in subseasonal to decadal prediction. Bull. Am. Meteorol. Soc. 101 (6), E869–E896. https://doi.org/10.1175/BAMS-D-19-0037.1.
Mo, K.C., Lyon, B., 2015. Global meteorological drought prediction using the North American multi-model ensemble. J. Hydrometeorol. 16 (3), 1409–1424. https://doi.org/10.1175/JHM-D-14-0192.1.
Moreira, E.E., 2016. SPI drought class prediction using log-linear models applied to wet and dry seasons. Phys. Chem. Earth Parts A/B/C 94, 136–145. https://doi.org/10.1016/j.pce.2015.10.019.
Moreira, E.E., Coelho, C.A., Paulo, A.A., Pereira, L.S., Mexia, J.T., 2008. SPI-based drought category prediction using loglinear models. J. Hydrol. 354 (1–4), 116–130.
Moreira, E., Russo, A., Trigo, R.M., 2018. Monthly prediction of drought classes using log-linear models under the influence of NAO for early-warning of drought and water management. Water 10 (1). https://doi.org/10.3390/w10010065.
Niaz, R., Almazah, M.M.A., Zhang, X., Hussain, I., Faisal, M., 2021. Prediction for various drought classes using spatiotemporal categorical sequences. Complexity 2021, 7145168. https://doi.org/10.1155/2021/7145168.
Okumura, Y.M., DiNezio, P., Deser, C., 2017. Evolving impacts of multiyear La Niña events on atmospheric circulation and US drought. Geophys. Res. Lett. 44 (22), 11–614. https://doi.org/10.1002/2017GL075034.
Pan, B., et al., 2020. Improving seasonal forecast using probabilistic deep learning. arXiv: Geophysics.
Paulo, A.A., Pereira, L.S., 2007. Prediction of SPI drought class transitions using Markov chains. Water Resour. Manage. 21 (10), 1813–1827. https://doi.org/10.1007/s11269-006-9129-9.
Poornima, S., Pushpalatha, M., 2019. Drought prediction based on SPI and SPEI with varying timescales using LSTM recurrent neural network. Soft. Comput. 23 (18), 8399–8412.
Qian, W., Du, J., Ai, Y., 2021. A review: Anomaly-based versus full-field-based weather analysis and forecasting. Bull. Am. Meteorol. Soc. 102 (4), E849–E870. https://doi.org/10.1175/BAMS-D-19-0297.1.
Reichstein, M., et al., 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566 (7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1.
Ribeiro, A.F.S., Pires, C.A.L., 2016. Seasonal drought predictability in Portugal using statistical–dynamical techniques. Phys. Chem. Earth, Parts A/B/C 94, 155–166. https://doi.org/10.1016/j.pce.2015.04.003.
Richardson, D., Fowler, H., Kilsby, C., Neal, R., Dankers, R., 2020. Improving sub-seasonal forecast skill of meteorological drought: a weather pattern approach. Nat. Hazards Earth Syst. Sci. 20, 107–124. https://doi.org/10.5194/nhess-20-107-2020.
Schepen, A., Wang, Q.J., Robertson, D.E., 2016. Application to post-processing of meteorological seasonal forecasting. In: Duan, Q. (Ed.), Handbook of Hydrometeorological Ensemble Forecasting. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp. 1–29. https://doi.org/10.1007/978-3-642-40457-3_18-1.
Shen, C., 2018. A transdisciplinary review of deep learning research and its relevance for water resources scientists. Water Resour. Res. 54 (11), 8558–8593.
Sigaroodi, S.K., Chen, Q., Ebrahimi, S., Nazari, A., Choobin, B., 2014. Long-term precipitation forecast for drought relief using atmospheric circulation factors: a study on the Maharloo Basin in Iran. Hydrol. Earth Syst. Sci. 18 (5), 1995–2006. https://doi.org/10.5194/hess-18-1995-2014.
Svoboda, M., et al., 2002. The drought monitor. Bull. Am. Meteorol. Soc. 83 (8), 1181–1190. https://doi.org/10.1175/1520-0477-83.8.1181.
Svoboda, M., Hayes, M., Wood, D., 2012. Standardized Precipitation Index User Guide. World Meteorological Organization Geneva, Switzerland.
Tang, Y., Duan, A., 2021. Using deep learning to predict the East Asian summer monsoon. Environ. Res. Lett. 16 (12), 124006.
Ujeneza, E.L., Abiodun, B.J., 2015. Drought regimes in Southern Africa and how well GCMs simulate them. Clim. Dyn. 44 (5–6), 1595–1609. https://doi.org/10.1007/s00382-014-2325-z.
Wang, Q., et al., 2014. Temporal-spatial characteristics of severe drought events and their impact on agriculture on a global scale. Quat. Int. 349, 10–21. https://doi.org/10.1016/j.quaint.2014.06.021.
Wang, Q.J., Schepen, A., Robertson, D.E., 2012. Merging seasonal rainfall forecasts from multiple statistical models through Bayesian model averaging. J. Clim. 25 (16), 5524–5537. https://doi.org/10.1175/JCLI-D-11-00386.1.
Wu, X., et al., 2021. The development of a hybrid wavelet-ARIMA-LSTM model for precipitation amounts and drought analysis. Atmosphere 12 (1), 74.
Wu, Z.-Y., Lu, G.-H., Wen, L., Lin, C.A., 2011. Reconstructing and analyzing China's fifty-nine year (1951–2009) drought history using hydrological model simulation. Hydrol. Earth Syst. Sci. 15 (9), 2881–2894. https://doi.org/10.5194/hess-15-2881-2011.
Xu, L., Chen, N., Zhang, X., 2018a. A comparison of large-scale climate signals and the North American Multi-Model Ensemble (NMME) for drought prediction in China. J. Hydrol. 557, 378–390. https://doi.org/10.1016/j.jhydrol.2017.12.044.
Xu, L., Chen, N., Zhang, X., Chen, Z., 2018b. An evaluation of statistical, NMME and hybrid models for drought prediction in China. J. Hydrol. 566, 235–249.
Xu, D., Zhang, Q., Ding, Y., Zhang, D., 2022. Application of a hybrid ARIMA-LSTM model based on the SPEI for drought forecasting. Environ. Sci. Pollut. Res. 29 (3), 4128–4144.
Zhang, N., Dai, X., Ehsan, M.A., Deksissa, T., Development of a Drought Prediction System Based on Long Short-Term Memory Networks (LSTM). Springer, pp. 142-153.
Yan, H., Moradkhani, H., Zarekarizi, M., 2017. A probabilistic drought forecasting framework: A combined dynamical and statistical approach. J. Hydrol. 548, 291–304. https://doi.org/10.1016/j.jhydrol.2017.03.004.
Yoo-Geun, H., Jeong-Hwan, K., Jing-Jia, L., 2019. Deep learning for multi-year ENSO forecasts. Nature 573 (7775). https://doi.org/10.1038/s41586-019-1559-7.
Zhang, R., Chen, Z.-Y., Xu, L.-J., Ou, C.-Q., 2019. Meteorological drought forecasting based on a statistical model with machine learning techniques in Shaanxi province, China. Sci. Total Environ. 665, 338–346. https://doi.org/10.1016/j.scitotenv.2019.01.431.
Zheng, K., et al., 2021. A machine learning model for photorespiration response to multi-factors. Horticulturae 7 (8), 207. https://doi.org/10.3390/horticulturae7080207.
Zhu, Q. et al., 2020. Drought prediction using in situ and remote sensing products with SVM over the Xiang River Basin, China. Nat. Hazards. 10.1007/s11069-020-04394-x.
Zhuang, Y., Fu, R., Wang, H., 2020. Large-scale atmospheric circulation patterns associated with US Great Plains warm season droughts revealed by self-organizing maps. J. Geophys. Res. Atmos. 125 https://doi.org/10.1029/2019JD031460.