Precipitation forecasting based on deep learning strategy using empirical wavelet transform, Markov chain-incorporated long-short term memory network

Jie YANG (jie.yang@connect.um.edu.mo)
Chongqing Industry&Trade Polytechnic

Rui Tang
Kunming University of Science and Technology

Kun Lan
Quzhou University

Han Wang
The Chinese University of Hong Kong https://orcid.org/0000-0002-5002-3708

Lin Zhang
Yangtze Normal University

Simon Fong
University of Macau

Article

Keywords: precipitation forecasting, deep learning, empirical wavelet transform, Markov chain

Posted Date: April 12th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1517899/v1

Precipitation forecasting based on deep learning strategy using empirical wavelet transform, Markov chain-incorporated long-short term memory network

Abstract: Accurate and stable precipitation forecasting can better reflect the changing trend of climate and also provide timely and efficient environmental information for a management decision, as well as prevent the occurrence of floods or droughts. This study proposes a hybrid model for precipitation forecasting and demonstrates its efficiency. Firstly, the empirical wavelet transform (EWT) is introduced to decompose and pre-analysis hidden characteristics of the precipitation data. Secondly, The Long Short Term Memory (LSTM) network is improved in combination with the Markov Chain (MC) algorithm, thus providing more precise forecasting results for rainless and rainy months and mitigating any extreme and non-physical precipitation generation. Thirdly, the multi-step prediction is explored to improve the reliability and flexibility of rainfall. Monthly precipitation data is used as illustrative cases to verify the performance of the proposed model. Parallel experiments using non-decomposing models, other traditional machine learning approaches optimized by the mind evolution algorithm have been designed and conducted to compare with the proposed model. Results indicated that the proposed hybrid model can capture the nonlinear characteristics of the precipitation time series, thus provides more precise forecasting results.

Keywords: precipitation forecasting; deep learning; empirical wavelet transform; Markov chain

1. Introduction

Precipitation plays an essential role in hydrological research and meteorological disaster warning and premonitor. Due to the complexity and diversity of climatic conditions, the process of rainfall has a lot of ambiguity, uncertainty, and randomness. The effect of varying precipitation in both timing and amount can cause serious natural disasters, while atmospheric and water circulation is boosted in the cycles of global hydrological variation(Oki. and Kanae., 2006; Sun and Kim, 2016; Sun et al., 2018). For example, in Hong Kong SAR (Special Administration Region) and Macao SAR, heavy pouring in the summer is typically associated with the southwest monsoon, monsoon troughs, and tropical cyclones. Though less frequently, excessive rainfall can still occur during the cool season due to the land-sea breeze convergence and the passage of frontal systems, leading to landslips because both places have a hilly topography(Li and Lai, 2004). Therefore, accurate and reliable precipitation forecasting is of great significance for

agricultural development, climate analysis, monitoring, and warning of natural disasters such as floods, landslides, and mudslides.

Traditionally, the precipitation forecasting models can be classified into two main types: the physical prediction and the statistical prediction models(Nourani et al., 2011). Technically, physical-based models usually take multi-factors into consideration including atmospheric circulation, pressure belt, ocean current, and the Monsoon, etc., while statistical prediction models usually take advantage of the implicit information contained in the historical precipitation time series to forecast the future precipitation(Yang et al., 2016). Physically-based models can suffer from the drawbacks that physical relationships between different impact features are complex, and sometimes the lack of multi-factor data makes it difficult to complete physical modeling(Feng et al., 2021; Napolitano et al., 2011).

The main technologies of statistical simulating for precipitation forecasting can be classified into three types: the linear methods, the nonlinear methods, and the hybrid methods(Tao et al., 2017a). Linear methods, such as auto-regressive (AR) models, auto-regressive moving average (ARMA) models, and auto-regressive integrated moving average (ARIMA) models, capture hydrologic time series features in a limited way due to the usually nonlinear and unstable nature of the time series(PaoloBurlando. et al., 1993). Many stochastic methods dealing with precipitation lack attention to the retention of the actual marginal distribution and associated structure, thus affecting the predictive power of such models(Papalexiou, 2018; Papalexiou and Serinaldi, 2020). Simon et al. (Papalexiou et al., 2018)presented a fairly accurate statistical dimensionality reduction method named "Decomposition with Preserved Margins and Correlation" (DiPMaC) decomposition method. The proposed method can decompose time series from coarse scales to arbitrarily fine scales while reproducing the probability distribution and linear correlation structure of the fine-scale processes. Nonlinear methods such as support vector machines (SVM), support vector regression (SVR), gray prediction (GF), and artificial neural networks (ANN) are widely applied to deal with nonlinear and complex precipitation series(Yu et al., 2000). The traditional ANN models for rainfall forecasting include Feed-forward Neural Network (FNN), Back Propagation Neural Network (BPNN), Radial Basis Function Neural Network (RBFNN), and so on. The significant effect of these models is investigated in some cases. Santamaría-Bonfil et al. (Santamaría-Bonfil et al., 2016) developed an SVM model experimentally showing that the proposed model was more accurate than the persistence and auto-regressive models in medium and short-term time series forecasting. Abbot et al. (Abbot and Marohasy, 2014) used ANN to forecasts rainfall for sites in three geographically distinct regions in Queensland. He et al.(HE Hui et al., 2007) developed a prediction model for the monthly mean circulation field and monthly rainfall

at 500-hPa using the BP neural network approach. The prediction of monthly rainfall was formulated using the product of monthly dynamic extended range ensemble prediction to improve the downscaling interpretation of monthly climate elements in Guangxi through dynamic model products. However, the fitting error and prediction error of the established model are relatively large. Farajzadeh et al. (Farajzadeh et al., 2014) forecast rainfall in the Urmia basin of northern Iran by using a feed-forward neural network. Kisi O et al. (Kisi and Cimen, 2012) showed that their proposed SVM model was more accurate than the ANN models in precipitation forecasting. The hybrid methods are the combinations of some mainstream models and algorithms. Due to the good abilities of the hybrid models for the feature-extracting, these models have been widely applied in precipitation time series forecasting.

Recently, many hybrid prediction models in precipitation forecasting have been proposed and investigated, which mainly contain data pre-processing and forecast modeling. The main technologies for these two modeling types are the signal decomposition algorithms and the prediction algorithms(J. and H., 2018; Tao et al., 2017a).

The decomposition algorithms can effectively improve the prediction performance of the built models through decomposing the precipitation time series into several more stationary sub-layers(Sun et al., 2018). Among the decomposition algorithms, the Singular Spectrum Analysis (SSA), Wavelet Decomposition (WD), the Empirical Mode Decomposition (EMD), and the Ensemble Empirical Mode Decomposition (EEMD) are widely recognized and used in rainfall prediction. Kalteh et al. (Kalteh, 2017) provided a prediction model combined with SSA and ANN Ding et al. (Ding and Dong, 2015) presented the hybrid forecasting model using the WD, phase space reconstruction (PSR) and the ANN. Qi et al. (Ouyang and Lu, 2017) compared the performance of different decomposition algorithms such as WD, SSA, and EEMD. In general, the WD algorithms have excellent ability of time-frequency analysis, and the EMD and EEMD algorithms have the strengthened self-adaptive ability to remove the stochastic volatility. However, the aforementioned decomposition algorithms have some disadvantages: (1) the performance of the WD algorithm depends on the wavelet basis and decomposition levels highly; (2) the EMD and EEMD algorithms lack the strict mathematical theory(Liu et al., 2015a). To overcome these drawbacks, Gilles (Gilles, 2013) proposed the Empirical Wavelet Transform (EWT), which can extract a series of modes of a signal by using an appropriate wavelet to obtain better effective in the non-stationary signal processing(Shaari et al., 2018). In the study, the EWT is adopted to decompose the original precipitation time series.

The prediction algorithms are the core consist of the precipitation forecasting. In recent years, some new prediction algorithms have been proposed, among which, deep learning methods, such as the Deep Belief Network

(DBN)[45], the Convolutional Neural Network (CNN)[46], and the Recurrent Neural Network (RNN)[47], have been developed rapidly. Deep learning is widely used in many fields such as bioinformatics(Lan et al., 2018), speech recognition, visual object recognition, object detection, and many other domains such as drug discovery and genomics(LeCun et al., 2015). Compared with the shallow models, the deep learning models can extract the deep inherent features in data(Najafabadi et al., 2015). Based on the following forecasting studies, deep learning methods such as CNN can get more accurate precipitation estimates than NARR(North American Regional Reanalysis Project)(Liu et al., 2017). Zhang et al. (Zhang et al., 2017) built a DBN for the time series forecasting. Qiu(Qiu et al., 2017) used CNN for short-term rainfall forecasting tasks. Their results manifested that the proposed model was effective. Zhang et al. (Zhang et al., 2018a) forecasted rainfall using the multi-layer perceptron (MLP) combined with dynamic regional, only short-term (3 hours) time series data was researched. Tao et al. (Tao et al., 2017b) applied a deep neural network (DNN) to weather datasets and showed it as a potential tool for the feature fusion of time series problems. Grover et al. (Grover et al., 2015) presented a weather forecasting model based on a deep hybrid model and forecasts made at 6, 12, and 24 hours. However, they do not predict with the model more difficult weather data, such as rainfall dataset. Shi et al. (Xingjian Shi et al., 2017) proposed a convolutional long short-term memory (ConvLSTM) model for precipitation nowcasting based on spatiotemporal sequence forecasting problem. Tao et al. (Tao et al., 2017b) improved the performance of satellite rainfall retrievals by means of deep neural network approach, but rainfall times series data was not been researched. Therefore, the deep learning methods have not yet been widely used in precipitation forecasting. By considering that the precipitation time series often have long-term and short-term dependency, the Long Short Term Memory (LSTM) network(Xiaodan Zhu et al., 2015), a special kind of RNN, is employed to predict the decomposed sub-layer in this study.

Based on forecasting algorithms, some other technologies are also used in forecast modelings, such as the SVM model optimized by Particle Swarm Optimization (PSO) algorithm, ANN connection weights optimized by the social-based algorithm (SBA)(Ramezani et al., 2014) and mind evolutionary algorithm (MEA) algorithm(Liu et al., 2015b) and ANN model incorporated with Markov chains. In recent years, some prediction models are worked with Markov chains to get better predicting performance. Haidar(Haidar and Verma, 2017) proposed a hybrid Genetic Algorithm (GA) to select input features, network parameters to train neural network topologies in rainfall forecasting. Gui et al.(Gui and Shao, 2017) applied a Markov chain to predict the rainfall and described the changing trend of precipitation easily. Aksoy et al.(Aksoy and Dahamsheh, 2018) proposed a new model combined with a Markov chain to forecast

monthly precipitation in arid regions. In this paper, a novel hybrid precipitation prediction model is proposed based on the EWT, Markov chain and LSTM network. The model is composed of three steps as (a) the EWT is adopted to decompose the raw precipitation data into several sub-layers; (b) the LSTM network is incorporated with Markov chain (MC) to predict the decompose sub-layers; (c) the prediction results of each sub-layer are summarized to obtain the final results for the original precipitation series. The provided EWT-MC-LSTM model, compared with LSTM models and BPNN model with and without EWT, MC technologies, have been designed and conducted to apply to monthly precipitation data of the four regions to highlight the better performance. It has been demonstrated by the results of this study that the proposed hybrid model can provide an effective modeling approach to capture the nonlinear characteristics of monthly precipitation series, thus providing more satisfactory forecasting results.

2. Methodology

2.1. Empirical Wavelet Transform

The empirical wavelet transform (EWT), first introduced by Jerome Gilles(Gilles, 2013), can be defined as a set of band pass filters selected according to the spectral characteristics of a signal. The "spectral characteristics of a signal" is the different intrinsic modes of a time series. In order to determine the frequency ranges of the band pass filters, the Fourier spectrum of the signal is segmented. According to the literature(Peng et al., 2017), the EWT can effectively identify and extract a finite number of intrinsic modes of a precipitation time series. The EWT algorithm relies on robust preprocessing for peak detection, then performs spectrum segmentation based on detected maxima, and constructs a corresponding wavelet filter bank. The major steps of the EWT algorithm can be described as (a) extending the signal; (b) executing the Fourier transform; (c) extracting the boundaries; (d) building the filter bank; (e) extracting the sub-bands.

The computation of the EWT can be summarized as(Cao et al., 2016):

(1) Segment the Fourier spectrum of the original precipitation series into N contiguous segments. The limits between each segment can be defined as ω_n, where $\omega_0 = 0$ and $\omega_n = \pi$, respectively. Each segment can be defined as $\Lambda_n = [\omega_{n-1}, \omega_n]$. For each ω_n, a transition phase T_n with width $2\tau_n$ is used. $\tau_n = \gamma\omega_n$. The range of γ is defined as:

$$\gamma < min_n \left(\frac{\omega_{n+1}-\omega_n}{\omega_{n+1}+\omega_n}\right) \quad (1)$$

(2)Construct a series of empirical wavelets based on the Little-wood-Paley and Meyer's wavelets. For $\forall_n > 0$, the empirical scaling function and empirical wavelets are defined by Eqs. (2) and (3), respectively:

$$\hat{\phi}_n(\omega) = \begin{cases} 1 & |\omega| \leqslant (1-\gamma)\omega_n \\ cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_n}|\omega| - \omega_n + \tau_n\right)\right] & (1-\gamma)\omega_n \leqslant |\omega| \leqslant (1+\gamma)\omega_n \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$$\hat{\varphi}_n(\omega) = \begin{cases} 1 & \omega_n(1+\gamma) \leqslant |\omega| \leqslant (1-\gamma)\omega_{n+1} \\ cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_{n+1}}|\omega| - \omega_{n+1} + \tau_{n+1}\right)\right] & (1-\gamma)\omega_{n+1} \leqslant |\omega| \leqslant (1+\gamma)\omega_{n+1} \\ sin\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau}|\omega| - \omega_n + \tau_n\right)\right] & (1-\gamma)\omega_n \leqslant |\omega| \leqslant (1+\gamma)\omega_n \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The function β (x) is defined as:

$$\beta(x) = x^4(35 - 84x + 70x^2 - 20x^3) \tag{4}$$

The approximation coefficients $W_f^{\varepsilon}(0,t)$ are obtained by the inner products with the empirical scaling function as follows:

$$W_f^{\varepsilon}(0,t) = \langle f, \phi_1 \rangle = \int f(\tau)\overline{\phi_1(\tau - t)}d\tau \tag{5}$$

The detailed coefficients $W_f^{\varepsilon}(n,t)$ are obtained by the inner products with the empirical wavelets as follows:

$$W_f^{\varepsilon}(n,t) = \langle f, \psi_n \rangle = \int f(\tau)\overline{\psi_n(\tau - t)}d\tau \tag{6}$$

(3)Reconstruct the precipitation series. The reconstruction series and empirical modes are defined as:

$$f(t) = W_f^{\varepsilon}(0,t)^* \phi_1(t) + \sum_{n=1}^{N}[W_f^{\varepsilon}(n,t)^* \psi_n(t)] \tag{7}$$

$$x_0(t) = W_f^{\varepsilon}(0,t)^* \phi_1(t) \tag{8}$$

$$x_k(t) = W_f^{\varepsilon}(k,t)^* \psi_k(t) \tag{9}$$

2.2 Long-short term memory network

The long-short term memory network (LSTM) is a special kind of recurrent neural network (RNN) that is stable and powerful for solving long-term and short-term dependency problems. The key parameter of the LSTM network is the memory cell, which can memorize the temporal state. The LSTM network can add or remove information to the cell state through three controlling gates as input-, forget- and output gate. The calculation of the LSTM network is described as follows(Xingjian Shi et al., 2015): (1) When a new input comes, if the input gate is activated, the input information can be accumulated to the cell. (2) If the forget gate is activated, the past cell status can be forgotten in

the process. (3) The output gate can control whether the latest cell output can be propagated to the ultimate state. LSTM architecture is illustrated in Fig. 1.

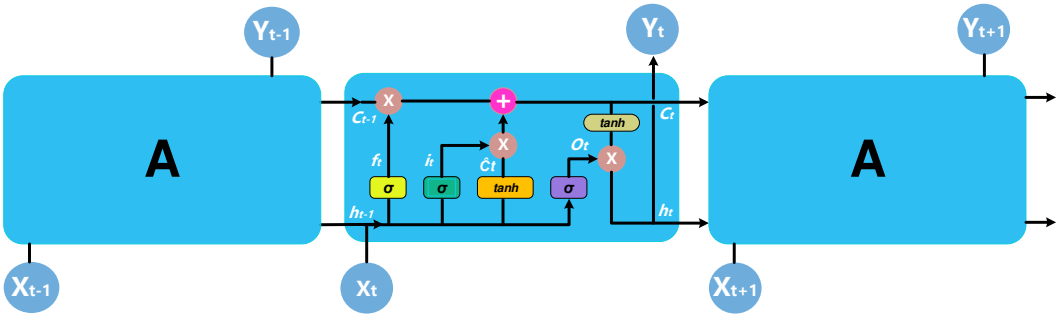


Fig. 1 The architecture of an LSTM model. Each line carries an entire vector, from the output of one node to the inputs of others. The circles represent pointwise operations, like vector addition, while the boxes are learned neural network layers. Lines merging denote concatenation, while a line forking denotes its content being copied and the copies going to different locations.

In the context of precipitation forecasting, $x=(x_1,x_2,\ldots,x_t)$ is the historical input data and $y=(y_1,y_2,\ldots,y_t)$ is the predicted data. The predicted precipitation series can be computed as:

$$i_t = \sigma(W_{ix}x_t + W_{im}m_{t-1} + W_{ic}c_{t-1} + b_i) \tag{10}$$

$$f_t = \sigma\left(W_{fx}x_t + W_{fm}m_{t-1} + W_{fc}c_{t-1} + b_f\right) \tag{11}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ tanh(W_{cx}x_t + W_{cm}m_{t-1} + b_c) \tag{12}$$

$$o_t = \sigma(W_{ox}x_t + W_{om}m_{t-1} + W_{oc}c_t + b_o) \tag{13}$$

$$m_t = o_t \circ tanh(c_t) \tag{14}$$

$$y_t = W_{ym}m_t + b_y \tag{15}$$

where i_t denotes the input gate, f_t denotes the forget gate, c_t denotes the activation vectors for each cell, o_t denotes the output gate, m_t denotes the activation vectors for each memory block, W denotes the weight matrices, b denotes the bias vectors, and '$\circ$ ' denotes the scalar product.

σ(·) is the standard logistic function as:

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{16}$$

g(·) is the centered logistic function as:

$$g(x) = \frac{4}{1+e^{-x}} - 2 \; x \in [-2,2] \tag{17}$$

h(·) is the centered logistic function as:

$$h(x) = \frac{2}{1+e^{-x}} - 1 \; x \in [-1,1] \tag{18}$$

2.3 Markov chain-incorporated model

A Markov chain is a stochastic process that transitions from one state to another. At any time, the chain can be in only one state. The set of all states is called the chain's state space. The process transitions from one state to another with a certain probability that is called the transition probability(Bohme et al., 2017). This probability depends only upon the current state rather than upon the path to the present state.

More formally, a Markov chain is a sequence of random variables $\{X_0, X_1, \ldots, X_n\}$ where $X_i$describes the state of the process at the time i. Given a set of states S = {1,2,…, N} for some N $\in \mathbb{N}$, the value of the random variables X_i are taken from S. The probability that the Markov chain starts out in state i is given by the initial distribution$P(X_0 = i)$.

The probability matrix $\boldsymbol{P}$ = (p_{ij}) specifies the transition rules. If |S| = N, then $\boldsymbol{P}$ is an $N \times N$ stochastic matrix where each entry is non-negative and the sum of each row is 1. The conditional probability p_{ij} defines the probability that the chain transitions to state j at time t + 1, given that it is in state i at time t,

$$p_{ij} = P(X_{t+1} = j | X_t = i) \tag{19}$$

A Markov chain is called time-homogeneous if the probability matrix (p_{ij}) does not depend on the time n. In other words, every time the chain is in state i, the probability of jumping to state j is the same. If a Markov chain is time homogeneous, then the vector π is called a stationary distribution of the Markov chain if for all $j \in S$ it satisfies

$$0 \leq \pi_j \leq 1 \tag{20}$$

$$1 = \sum_{i \in S} \pi_i \tag{21}$$

$$\pi_j = \sum_{i \in S} \pi_i p_{ij} \tag{22}$$

Informally, a Markov chain$\{X_0, X_1, \ldots, X_n\}$ is called rapidly mixing if X_n is "close" to the stationary distribution for a sufficiently low number of steps n. In other words, rapidly mixing chains approach the stationary distribution within a reasonable time—independent of the initial state.

2.4 The Hybrid EWT-MC-LSTM Model

The overall structure of the proposed hybrid EWT-MC-LSTM model is shown in Fig. 2. As can be seen in Fig. 2, the detailed procedure of the proposed model can be described in four steps:

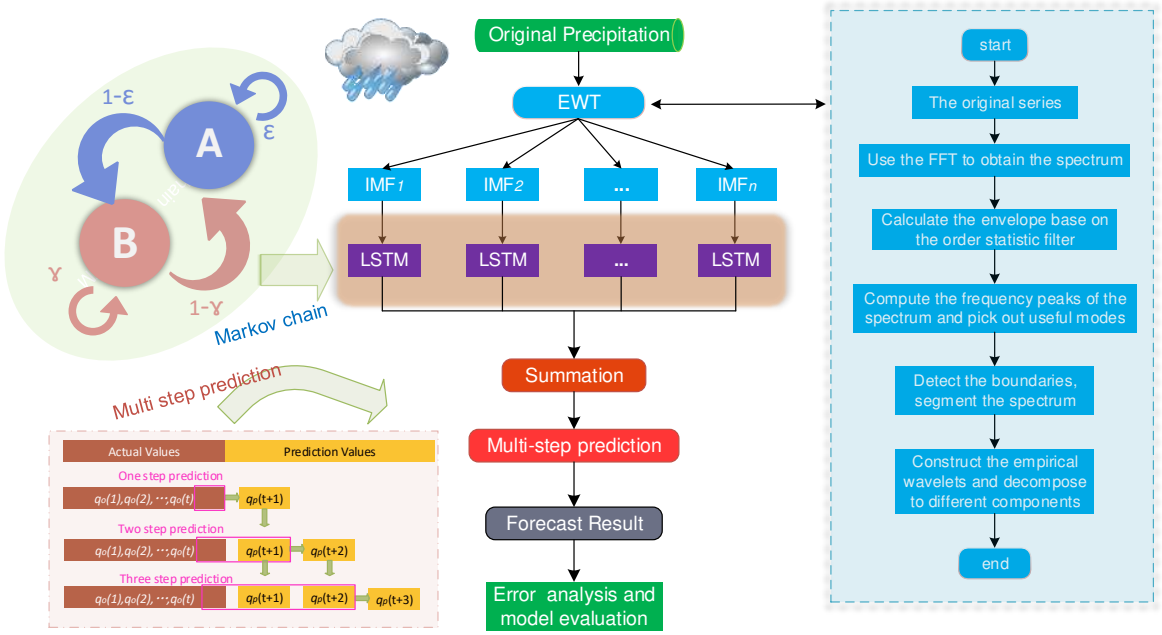


Fig. 2 The framework of the proposed EWT-MC-LSTM model. (IMF: Intrinsic Mode Functions, IMF is amplitude modulated-frequency modulated function, the aim of EWT method is employed to decompose the signal into several finite modes and identify and extract the different intrinsic modes of rainfall time series.)

Stage a. Decompose the original precipitation series into a sequence of modes by EWT. The decomposed modes that reflect the inner data characteristics were analyzed and selected to establish specific individual predictors.

Stage b. Establish the LSTM model incorporated with Markov chain technology for each sub series, including components decomposed by EWT.

Stage c. Use the multi-step method to predict decomposed components.

Stage d. Make the summation of all the sub prediction from Stage c and get the final forecast.

2.5. Model Performance Evaluation

To assess the prediction performance of the involved models, three error measures, which include the MAPE (Mean Absolute Percentage Error), the MAE (Mean Absolute Error) and the RMSE (Root Mean Square Error), are all utilized in the forecasting experiments.

These indexes can be defined as:

$$\mathrm{MAE}=\frac{1}{N}\sum_{i=1}^{N}\left|q_p(i)-q_o(i)\right| \tag{23}$$

$$\mathrm{MAPE}=\frac{1}{N}\sum_{i=1}^{N}\frac{\left|q_p(i)-q_o(i)\right|}{q_o(i)}\times 100\% \tag{24}$$

$$\text{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(q_p(i)-q_o(i))^2} \quad (25)$$

Where $q_p(i)$ and $q_o(i)$ denote the predicted and observed monthly precipitation series, respectively, and N denotes the length of data.

To compare the prediction performance of the involved models, three percentage error measures are used. They can be described as follows:

$$P_{MAE}=\left|\frac{(MAE_1-MAE_2)}{MAE_1}\right| \quad (26)$$

$$P_{MAPE}=\left|\frac{(MAPE_1-MAPE_2)}{MAPE_1}\right| \quad (27)$$

$$P_{RMSE}=\left|\frac{(RMSE_1-RMSE_2)}{RMSE_1}\right| \quad (28)$$

3. Study Area and Data Collection

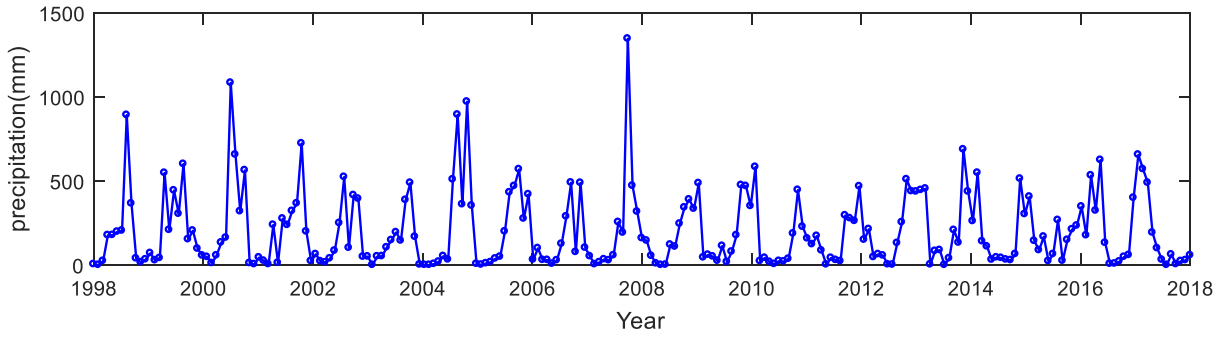
The monthly precipitation time series of four regions located on the southern coast and northwest of China, i.e., Hong Kong, Macao, Lanzhou, and Alashan League, are researched in this study.

Hong Kong and Macao are located in the subtropical zone of the northern hemisphere and the southern part of the Tropic of Cancer. They are backed by Eurasia and facing the Pacific Ocean, forming a maritime subtropical monsoon climate with four distinct rushes of warmth. On average, there are 5 to 6 typhoons per year bringing heavy rain. Located in the temperate zone, Lanzhou enjoys a semi-arid climate with hot summers and cold, dry winters. Alashan League is located in the western part of the Inner Mongolia autonomous region, which is a typical temperate continental climate.

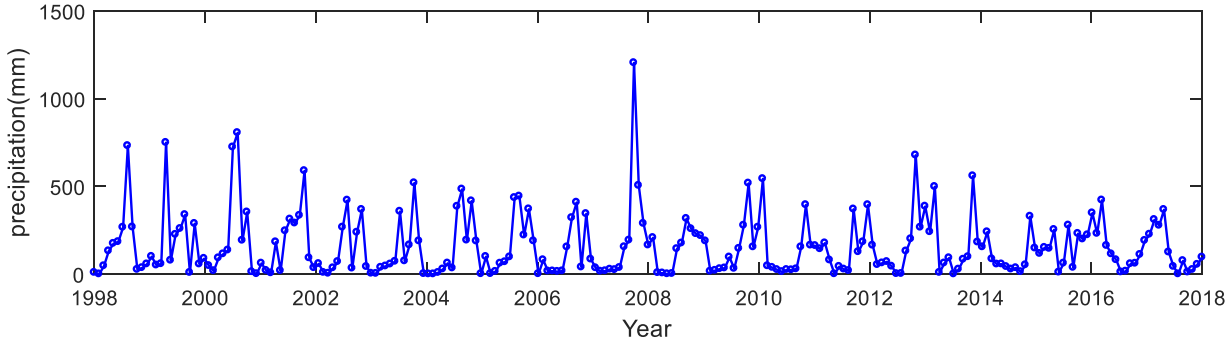
This study used data from the Hong Kong Observatory and the Macao Geophysical and Meteorological Bureau, and also the National Oceanic and Atmospheric Administration (NOAA). The statistical characteristics of the data are shown in Table 1. Table 1 lists the monthly basic statistics for these data sets, such as mean, standard deviation (St.Dev.), minimum (min), maximum (max), skewness (skew) and kurtosis (Kurt.). We replaced the zero value with a value close to zero(1e-6) and updated the kurtosis and skewness values to make the estimated statistics more meaningful. On the one hand, the monthly precipitation in the four regions fluctuates drastically. The precipitation from June to August is very high, and the average monthly precipitation can reach 501.81mm, but the monthly average precipitation in November, December and January is less than 50mm and the minimum value is less than 5mm. In Lanzhou and Alashan League, the rainfall in rainy months is also about 10

times higher than that in rainless months, although their maximal precipitation is almost 10 times lower than that in Hong Kong and Macao. Similarly, the standard deviation of the four regions is also highly variable throughout the year, the standard deviation of Hong Kong Station is 24.89mm in February and 297.63mm in June.

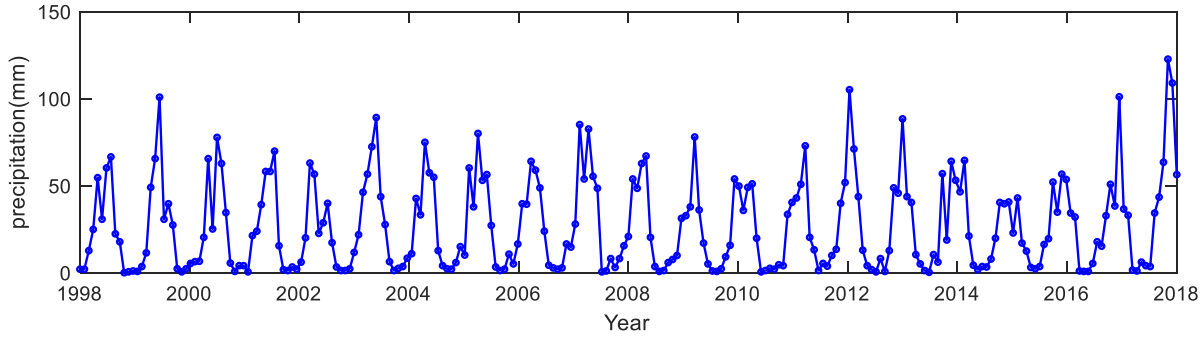
On the other hand, the average monthly precipitation in the region of Hong Kong and Macao varies widely. The average precipitation of Hong Kong in June is 456.7 mm, while the average precipitation in January is only



(a) Hong Kong



(b) Macau



(c) Lanzhou

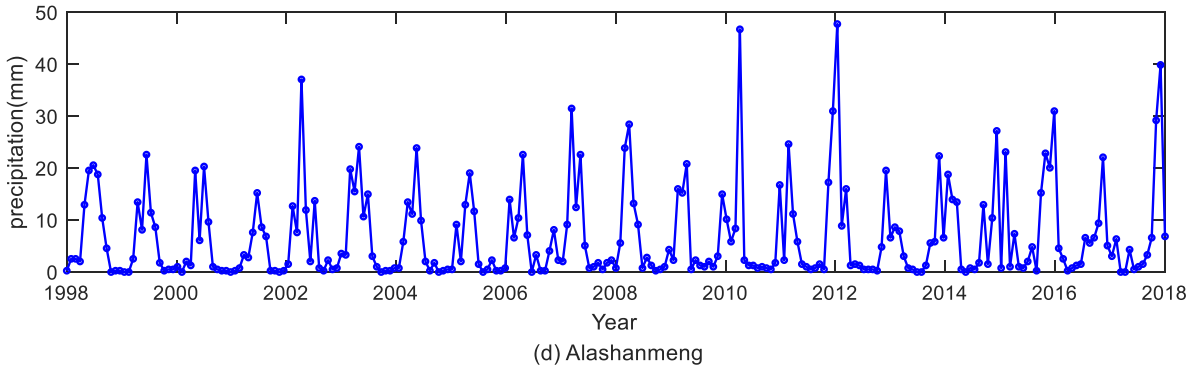


Fig. 3 Monthly precipitation series for the four regions.

24.7 mm. About 80% of the rainfall was recorded from May to September, June and August are usually the months with the most rainfall, while January and December have the least rainfall. There is little chance of rain because Lanzhou and Alashan League are located deep in the northwest inland and it is not easy to reach the warm and humid air flow in the ocean. Alashan League is drought, rainless, windy and sandy, cold winter and hot summer, four seasons climate characteristics are obvious. The annual average precipitation for Lanzhou is 327 mm, mainly from June to September, while the annual average precipitation for Alashan League is only 106 mm.

Four sets of original precipitation series, including 240 samples for each region, are shown in Fig. 3. The 1st-280th samples of precipitation series are used as the training dataset, while the leaving 281st-240th samples of each precipitation series are used as the testing dataset. Preprocessing the data set to prevent the presence of homogeneity and trends indicates that the precipitation at the four regions is uniform, and there is no trend at the 95% significance level. There is no missing data in the time series, so the time series do not need to process any missing data.

Table 1. Statistical characteristics of monthly total precipitation data recorded at Hong Kong, Macao, Lanzhou, and Alashan League meteorological stations

Region	Month	1	2	3	4	5	6	7	8	9	10	11	12
	Mean (mm)	36.70	25.61	75.39	155.43	311.37	501.81	360.71	435.98	315.99	106.76	47.48	31.23
	St. Dev. (mm)	57.62	24.89	63.68	121.61	147.24	297.63	155.36	214.45	182.38	139.14	37.19	26.23
Hong	*Cv*	14.24	7.43	1.13	4.50	0.41	2.49	-0.57	1.29	-0.62	9.91	-0.48	-0.88
Kong	*Cs*	3.58	2.33	1.40	1.79	1.14	1.61	-0.08	1.02	0.59	2.86	0.59	0.46
	Max (mm)	266.90	113.10	238.70	547.70	687.30	1346.10	656.40	971.30	723.00	624.40	131.30	88.30
	Min (mm)	0.10	0.10	17.50	12.40	162.00	144.70	76.90	143.30	87.90	2.30	0.40	0.10
Macao	Mean (mm)	36.92	29.57	73.14	156.52	286.50	368.47	258.70	311.02	241.29	82.29	41.84	32.40

	St.Dev. (mm)	24.09	58.03	160.42	147.71	248.82	172.28	137.81	162.77	86.69	30.55	29.52	0.00
	Cv	3.51	1.07	1.44	2.90	1.11	2.12	1.72	1.57	0.82	1.19	0.21	0.63
	Cs	13.40	1.47	1.62	9.76	1.30	5.85	4.28	3.25	-0.25	0.35	-1.21	-0.76
	Max (mm)	278.24	96.26	230.83	749.45	677.84	1204.02	805.40	731.22	588.62	287.64	99.43	92.02
	Min (mm)	0.01	0.00	12.86	18.64	77.67	74.26	33.02	151.05	8.24	1.03	0.09	0.03
Lanzhou	Mean (mm)	2.40	4.36	8.70	17.07	39.98	48.67	57.98	59.12	49.15	22.82	4.114	1.35
	St.Dev. (mm)	1.85	2.73	5.00	10.77	13.52	14.06	23.26	20.28	13.54	9.20	3.486	0.69
	Cv	1.75	1.22	0.30	2.78	-0.04	0.89	0.53	0.16	0.13	1.18	1.539	1.26
	Cs	3.97	0.94	-0.60	10.26	-1.05	0.68	-0.34	-0.20	-0.65	1.76	2.569	1.24
	Max (mm)	8.13	10.67	17.78	56.90	62.99	85.09	105.16	101.09	72.90	48.51	13.208	3.05
	Min (mm)	0.25	0.76	0.51	4.06	18.80	30.73	22.61	22.86	22.35	10.41	0.00	0.51
Alashan League	Mean (mm)	0.66	0.93	1.91	2.98	8.13	14.21	16.87	15.32	13.89	3.61	1.139	0.849
	St.Dev. (mm)	0.61	1.04	2.24	3.72	6.02	9.49	11.05	8.48	9.44	3.379	1.688	0.767
	Cv	1.30	1.74	1.77	2.07	0.48	0.80	1.11	0.06	2.37	1.549	2.968	1.089
	Cs	1.51	3.17	2.58	3.73	-1.13	0.35	1.69	-0.74	7.42	2.567	10.089	0.869
	Max (mm)	2.29	4.06	8.13	12.95	19.81	37.08	47.75	30.99	46.74	13.459	7.358	2.787
	Min (mm)	0.00	0.00	0.00	0.25	0.76	2.03	2.29	0.76	3.05	0.00	0.00	0.00

4. Model Construction and Development

4.1 Decomposing Using EWT

Graphical representations of the decomposed sub-series using EWT for the four regions are illustrated in Fig. 4. The original precipitation sequence is decomposed into six components, respectively. The EWT algorithm is applied to extract the feature sequence imf1 from the initial rainfall sequence, where imf1 is a low-frequency feature sequence that reflects the overall fluctuation of the data and correspondingly contributes more to the original time series and the frequencies of the remaining component sequences gradually increase, reflecting the local details of the rainfall sequence.

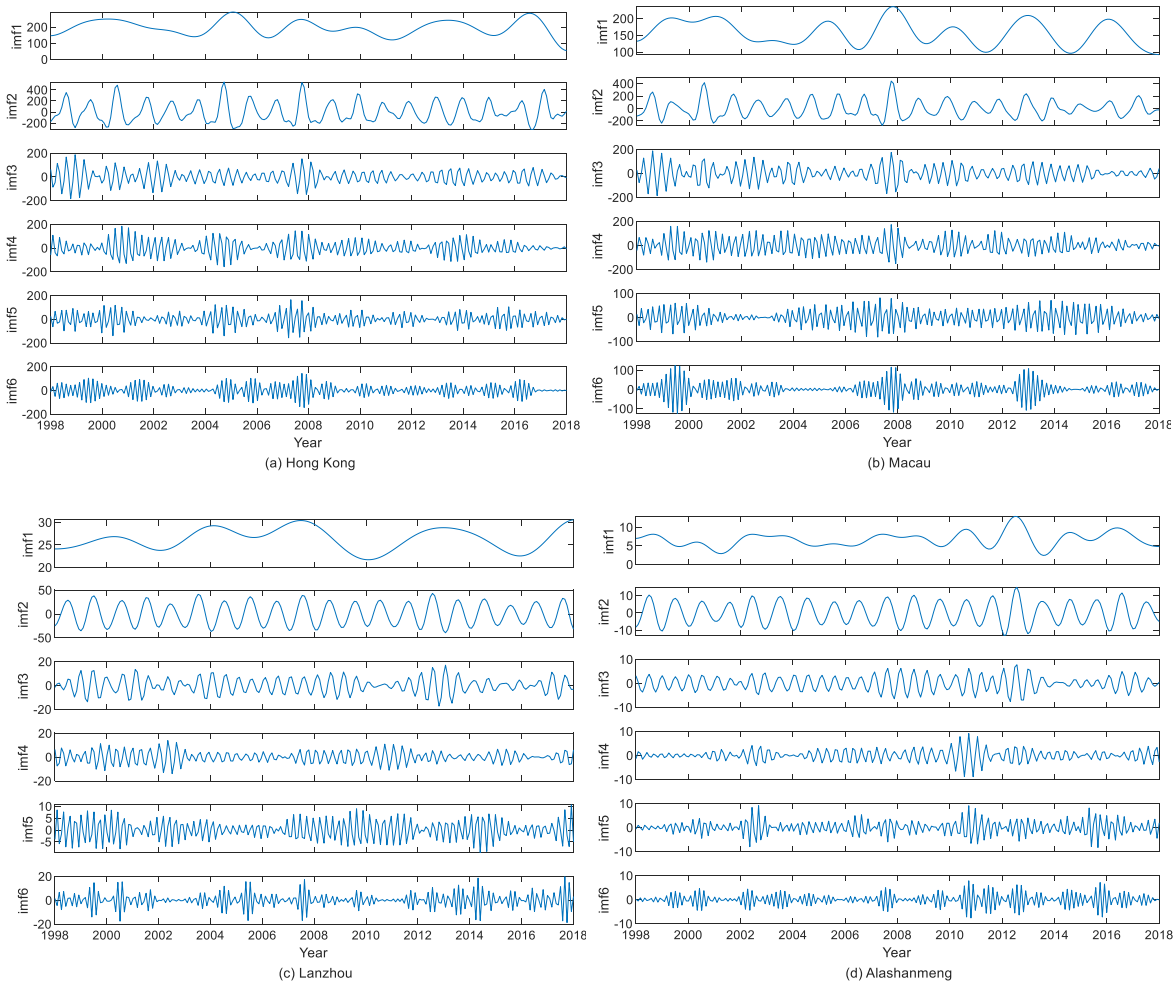


Fig 4. The EWT decomposed results of the original monthly precipitation series for the four regions

4.2 Markov chain-incorporated model

In the regular LSTM models, extreme values were generated for precipitation in special months such as winter months with little precipitation and, the negative values were all replaced with zero to avoid the physical unjustification. However, in summer with abundant rainfall, there may be extreme precipitation far beyond normal. In this case the observation of extremes is normal and anticipated. As an improvement to overcome this difficulty, the Markov chain (MC) was incorporated into the regular LSTM models. This provides the model the ability of forecasting months with, without and heavy and little rainfall[37]. This is called Markov chain-incorporated LSTM network (MC-LSTM) model. In the MC-LSTM model, it is first checked with the MC part of the model if precipitation is observed or not, or heavy in a particular month of the year, i.e., if a given month is rainless, normal or rainy. When

294 the MC part of the model forecasts a rainless month, a zero-precipitation is assigned to this particular month; when a
295 rainy month is forecasted by the MC part of the model, the LSTM part of the model is run to assign a larger value for
296 precipitation in the month.

297 A three-state (2/1/0) MC can ideally be used for forecasting the state of the month (rainless, normal and rainy),
298 where '2' stands for a rainy month, '1' for a normal and '0' for a rainless month. We simplified the classification into
299 three categories: rainy, normal, and rainless, based on the grading standard of the precipitation in the literature [37].
300 The main purpose of our rainfall grading is to facilitate the construction of a Markov transition probability matrix to
301 assist in improving the predicting performance of the LSTM networks. The transition probability matrix of the 2/1/0
302 Markov chain can be given as

303 $$P_{i,j} = \begin{bmatrix} P_{22} & P_{21} & P_{20} \\ P_{12} & P_{11} & P_{10} \\ P_{02} & P_{01} & P_{00} \end{bmatrix} \quad (29)$$

304 where P_{ij} shows the probability matrix of transition from a month with state i to a month with state j. The
305 probability of transition from state i to state j can be calculated as

306 $$P_{ij} = \frac{n_{ij}}{\sum_j n_{ij}}; i, j = 0,1 \quad (30)$$

307 Table 2. Transition probabilities of the 2/1/0 3-state Markov chain

Month	Hong Kong SAR			Macao SAR			Lanzhou			Alashan League		
	P10	P11	P12	P10	P11	P12	P10	P11	P12	P10	P11	P12
January	0.84	0.16	0.00	0.89	0.11	0.00	1.00	0.00	0.00	1.00	0.00	0.00
February	0.89	0.11	0.00	0.79	0.21	0.00	0.90	0.10	0.00	1.00	0.00	0.00
March	0.53	0.47	0.00	0.42	0.58	0.00	0.70	0.30	0.00	1.00	0.00	0.00
April	0.16	0.79	0.05	0.16	0.74	0.11	0.15	0.85	0.00	0.90	0.10	0.00
May	0.00	0.68	0.32	0.00	0.58	0.42	0.00	1.00	0.00	0.65	0.35	0.00
June	0.00	0.21	0.79	0.00	0.47	0.53	0.00	0.95	0.05	0.40	0.50	0.10
July	0.00	0.32	0.68	0.11	0.63	0.26	0.00	0.80	0.20	0.30	0.60	0.10
August	0.00	0.21	0.79	0.00	0.58	0.42	0.00	0.85	0.15	0.30	0.65	0.05
September	0.00	0.47	0.53	0.05	0.63	0.32	0.00	1.00	0.00	0.40	0.55	0.05
October	0.53	0.42	0.05	0.53	0.47	0.00	0.00	1.00	0.00	0.95	0.05	0.00
November	0.53	0.47	0.00	0.58	0.42	0.00	0.90	0.10	0.00	1.00	0.00	0.00
December	0.68	0.32	0.00	0.68	0.32	0.00	1.00	0.00	0.00	1.00	0.00	0.00

308

where n_{ij} is the total number of months with state j following a month with state i. The probabilities are determined for each month of the year because of seasonal effects. The sum of each row in the probability matrix satisfies

$$\sum_j P_{ij} = 1 \tag{31}$$

Thus, the number of parameters required is three, resulting in 36 parameters in total at a monthly basis, as shows in table 2.

Calibration of MC by using the calibration data set covers the determination of MC transition probabilities. Presented in table 2 are transition probabilities calculated considering the monthly total precipitation of each region. Months from May to September for Hong Kong means that this particular month has never been rainless during the calibration period, and the period from November to December has never been rainy. The same interpretation is valid for other stations.

4.3 Parameter Settings

After determining the optimal combination of input variables using the EWT method described in Section 3.2, the precipitation data are rescaled to [-1, 1]. To evaluate the performance of these models, RMSE, MAE and MAPE of the BPNN, MEA-BPNN and LSTM models with and without the process (EWT and MC) are analyzed and compared. MEA-BPNN(Wang et al., 2018) is a BP neural network optimized by mind evolution algorithm (MEA) model, which is compared as a predictor. In this study, MEA-BPNN is also shorthand for MEABP. The MEA is an evolutionary algorithm presented by Sun(Sun C et al., 1998), which emulates the activities of the human mind, and always perform better than GA in the convergent speed and the optimizing performance(Liu et al., 2015b). For the model with MEA-BPNN, the number of superior and temporary groups is defined as 5, and the size of superior/temporary groups is 20 according to Ref.(Wang et al., 2018). The max iteration is 500, and a value of more than 1000 for the maximum iteration number has also been tried for the MEA-BPNN model; results have shown that there is no significant improvement inaccuracy, but the experiment time is much longer. For the model with the LSTM network, the dimensions of the hidden layers are set as 50. The hyper-parameter a is set to 0.2, the dropout rate is set to 0.5 to prevent overfitting. The training procedure is stopped when the prediction accuracy does not improve for 10 iterations. Each experiment is repeated 20 times, and results with the best accuracy are presented.

5. Results and Discussion

The computations related to all of the models, including BPNN, MEA-BPNN, EWT-MEABP, EWT-MC-MEABP, LSTM, EWT-LSTM, and EWT-MC-LSTM, are implemented in the MATLAB environment in a computer with Intel core i7, 3.4-GHz CPU and 16 GB of RAM. The performance indices of the four regions are given in Table 3, and the promoting percentages of each model are shown in Fig. 5 to Fig. 8. Case 1, case 2, case 3 and case 4 in Fig. 5 to Fig. 8 respectively represents regions of Hong Kong, Macao, Lanzhou, and Alashan League.

Table 3. Forecasting performance of the models for the four regions

Model	Step1			Step2			Step3		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
Hong Kong									
BPNN	34.58	36.62	10.88	38.21	36.72	12.71	45.60	40.56	14.67
MEA-BPNN	32.22	34.61	6.12	31.94	36.64	6.91	33.02	38.33	12.67
EWT-MEA-BPNN	24.17	25.21	3.88	25.14	25.38	2.50	26.68	25.72	3.26
EWT-MEA-MC-BPNN	18.22	16.98	2.25	19.64	21.97	2.53	22.51	20.69	2.38
LSTM	28.78	25.81	3.52	31.69	28.99	3.67	27.27	30.06	4.55
EWT-LSTM	17.77	18.71	1.57	18.84	17.98	1.94	20.51	19.39	2.15
EWT-MC-LSTM	14.41	15.17	1.08	15.88	16.00	1.35	16.60	16.97	2.08
Macao									
BPNN	35.18	35.89	11.01	37.38	38.19	11.85	42.10	40.01	13.97
MEA-BPNN	30.88	34.15	6.03	26.68	36.44	8.82	33.12	37.66	10.67
EWT-MEA-BPNN	24.44	24.24	3.27	26.94	25.91	3.04	27.05	29.95	5.64
EWT-MEA-MC-BPNN	16.37	17.38	1.47	20.54	18.27	2.12	21.18	21.66	2.38
LSTM	29.92	26.38	3.33	32.01	28.33	3.68	27.97	21.34	5.71
EWT-LSTM	17.91	19.10	1.48	18.75	18.32	1.79	20.25	19.87	2.45
EWT-MC-LSTM	15.50	14.70	1.12	15.97	16.09	1.37	16.67	17.41	2.14
Lanzhou									
BPNN	10.88	8.64	1.28	11.43	9.17	1.42	12.67	9.47	1.39
MEA-BPNN	10.28	8.18	1.25	11.38	9.01	1.42	12.31	9.90	1.76
EWT-MEA-BPNN	8.39	6.92	0.92	8.78	7.63	1.18	10.53	9.36	1.54
EWT-MEA-MC-BPNN	8.38	6.70	0.91	8.83	7.21	1.12	10.29	8.38	1.30
LSTM	9.36	7.37	1.27	10.78	8.53	1.42	11.18	8.98	1.57
EWT-LSTM	7.66	6.30	1.13	8.28	7.10	1.16	8.64	7.23	1.15
EWT-MC-LSTM	7.29	5.69	0.88	7.48	5.80	0.87	8.05	6.48	0.83
Alashan League									
BPNN	3.21	2.42	5.02	3.72	2.67	5.49	3.96	3.03	5.92
MEA-BPNN	3.00	2.38	4.53	3.67	2.99	4.50	4.05	3.27	4.85
EWT-MEA-BPNN	2.63	1.38	1.75	2.48	1.64	2.14	2.89	1.92	2.36
EWT-MEA-MC-BPNN	2.22	1.34	1.63	2.44	1.62	1.95	2.52	1.91	2.21
LSTM	2.86	2.07	3.32	2.91	2.15	3.52	2.89	2.31	4.18
EWT-LSTM	2.22	1.51	1.63	2.45	1.71	1.82	2.67	2.05	2.17
EWT-MC-LSTM	1.97	1.28	1.45	2.38	1.57	1.80	2.69	2.01	1.93

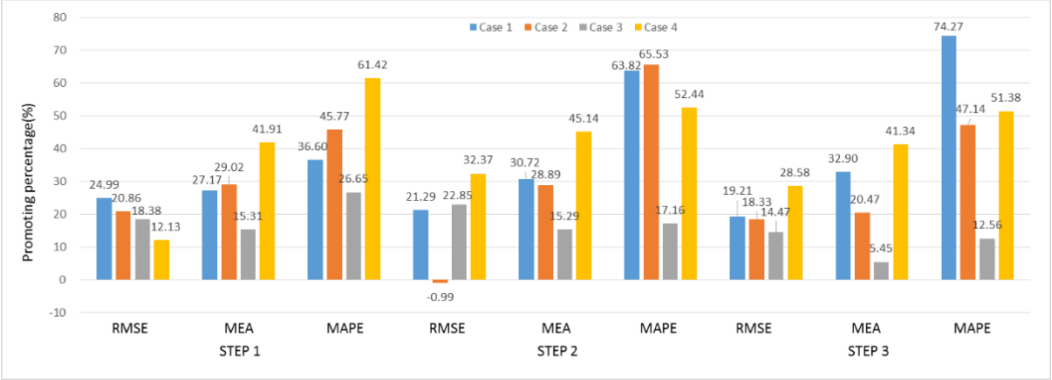
It is clear from table 3 that the proposed EWT-MC-LSTM model obtains the smallest RMSE, MAE and MAPE compared with the other six models. Comparison between the models with and

without the EWT, with and without MC algorithm, reveals that combining EWT and MC technologies with the BPNN model is an effective way to increase the forecasting accuracy. Detailed analyses are submitted to further evaluate the performance of these models. Though the proposed EWT-MC-LSTM model exhibits much better performance than the other models, the configuration is more complex on account of data pre-processing and weights update computing. Thus, the computation time required is much more than the single BPNN model without the decomposing and optimization processes.

From Fig. 5 (a)-(d), it can be found that:

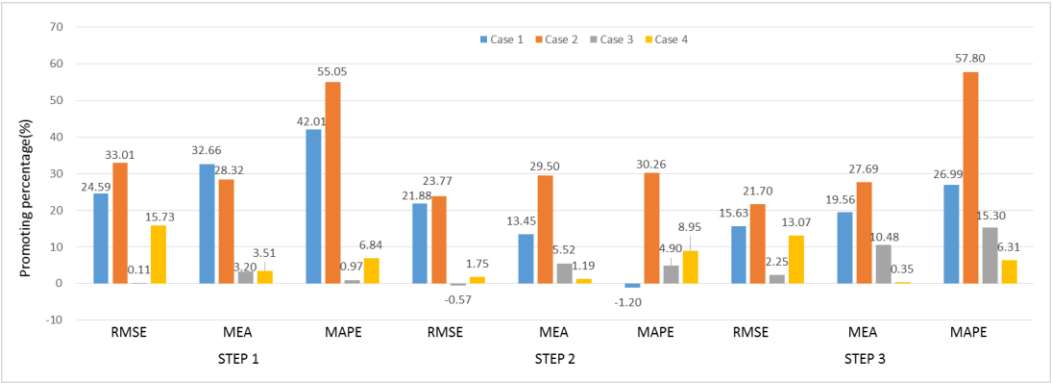
(a) Fig.5 (a) shows the promoting performance improvement of the EWT-MEABP model, incorporated with EWT decomposition technology, compared to the MEA-BPNN model in three step conditions. The prediction accuracy of the EWT-MEABP model is mostly higher than that of the MEA-BPNN model significantly. For example, in Macao, the promoting RMSE percentages of the MEA-BPNN model in the multi-step results by the EWT-MEABP model are 20.86%, -0.99% and 18.33%, respectively; the promoting MAE percentages of the MEA-BPNN model in the multi-step results by the EWT-MEA-BPNN model are 29.02%, 28.89%, and 20.47%, respectively; and the promoting MAPE percentages of the MEA-BPNN model in the multi-step results by the EWT-MEABP model are 45.77%, 65.53%, and 47.14%, respectively.

(b) Fig.5 (b) shows the promoting performance improvement of the EWT-MC-MEABP model, incorporated with the Markov chain, compared to the EWT-MEABP model in three step conditions. The prediction accuracy of the EWT-MC-MEABP model is also mostly higher than that of the EWT-MEABP model significantly. For example, in Hong Kong, the promoting RMSE percentages of the EWT-MEABP model in the multi-step results by the EWT-MC-MEABP model are 24.59%, 21.88%, and 15.63%, respectively; the promoting MAE percentages of the EWT-MEABP model in the multi-step results by the EWT-MC-MEABP model are 32.66%, 13.45%, and 19.56%, respectively; and the promoting MAPE percentages of the EWT-MEABP model in the multi-step results by the EWT-MC-MEABP model are 42.01%, -1.2% and 26.99%, respectively.



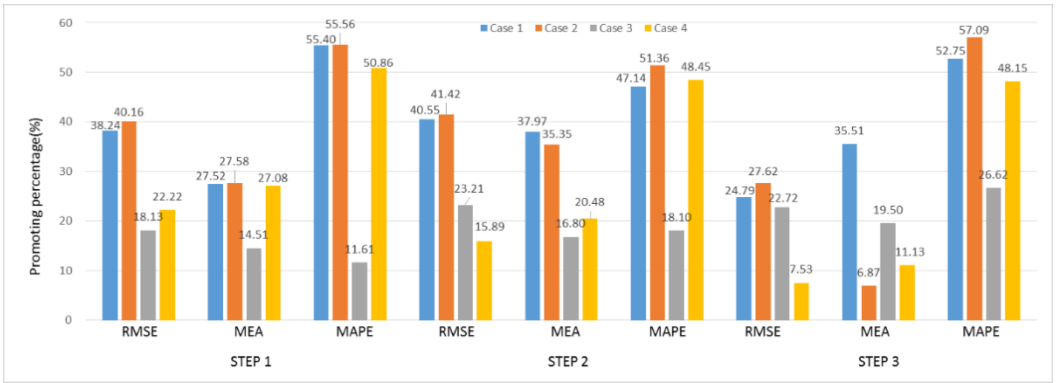
368

369 Fig. 5(a) The promoting percentages of the MEA-BPNN model by the EWT-MEABP model.



370

371 Fig. 5(b) The promoting percentages of the EWT-MEABP model by the EWT-MC-MEABP model.



372

373 Fig. 5(c) The promoting percentages of the LSTM model by the EWT-LSTM model.

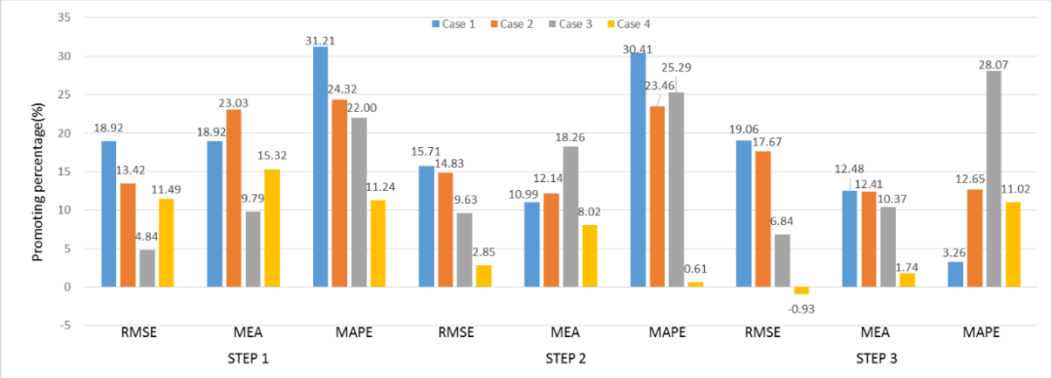


Fig. 5(d) The promoting percentages of the EWT-LSTM model by the EWT-MC-LSTM model.

(c) Fig.5 (c) shows the promoting performance improvement of the EWT-LSTM model, incorporated with EWT decomposition technology, compared to the LSTM model in three step conditions. The prediction accuracy of the EWT-LSTM model is higher than that of the LSTM model significantly. For example, in Lanzhou, the promoting RMSE percentages of the LSTM model in the multi-step results by the EWT-LSTM model are 18.13%, 23.21%, and 22.72%, respectively; the promoting MAE percentages of the LSTM model in the multi-step results by the EWT-LSTM model are 15.51%, 16.80%, and 19.5%, respectively; and the promoting MAPE percentages of the LSTM model in the multi-step results by the EWT-LSTM model are 11.61, 18.1% and 26.62%, respectively.

(d) Fig.5 (d) shows the promoting performance improvement of the EWT-MC-LSTM model, incorporated with Markov chain, compared to the EWT-LSTM model in three step conditions. The prediction accuracy of the EWT-MC-LSTM model is mostly higher than that of the EWT-LSTM model significantly. For example, in Alashan League, the promoting RMSE percentages of the EWT-LSTM model in the multi-step results by the EWT-MC-LSTM model are 11.49%, 9.63% and -0.93%, respectively; the promoting MAE percentages of the EWT-LSTM model in the multi-step results by the EWT-MC-LSTM model are 15.32%, 8.02%, and 1.74%, respectively; and the promoting MAPE percentages of the EWT-LSTM model in the multi-step results by the EWT-MC-LSTM model are 11.24%, 0.61%, and 11.02%, respectively. In addition, the average percentage improvement of figure 5 (b) is slightly higher than that of figure 5 (d); the difference between the two is that (b) uses MEABP as the classifier, while (d) uses LSTM. Compared with MEABP, LSTM can improve the classification accuracy by mining more implicit layer features through deep learning network structure. Since the LSTM model has better classification results than MEABP (as shown in Table 3), the percentage boost of

LSTM combined with MC does not increase linearly compared to the MEABP model. But the boosting percentage does not affect the prediction accuracy of the model. Also, from (b) and (d), we can still see that MC provides a boost to the prediction performance of the LSTM and MEABP classifiers.

(e) Among all the aforementioned precipitation prediction models, the EWT-MC-LSTM model has the best estimated performance. The reasons for such a significant forecasting promotion can be explained as: Firstly, compared to the models without EWT decomposed algorithms, the EWT has a better ability of mining more features of the precipitation series themselves due to its combination of the empirical ability and the wavelet computation. So the EWT can provide the deeper decomposing precipitation sub-layers for the forecasting computation. Secondly, compared to the models without MC incorporated, the models with MC technology have better performance in improving forecast accuracy on extreme values of precipitation series, such as precipitation in January or December, which precipitation often close to zero mm, but when it comes to May, June or July, the precipitation can reach 1000 mm (Hong Kong and Macao), 50 mm (Lanzhou) or 20 mm (Alashan League). Thirdly, in the proposed hybrid forecasting model, the LSTM network is adopted to forecast the precipitation sub-layers. The LSTM network is an improved model of the RNN network, which has all of the advantages of the deep learning RNN network but at the same time overcome the vanishing gradient problem of the RNN network, it has the strong nonlinear processing ability, and it is suitable for the non-stationary precipitation forecasting computation. Based on the successful combination of these upper algorithms, the proposed EWT-MC-LSTM model has satisfactory multi-step forecasting results.

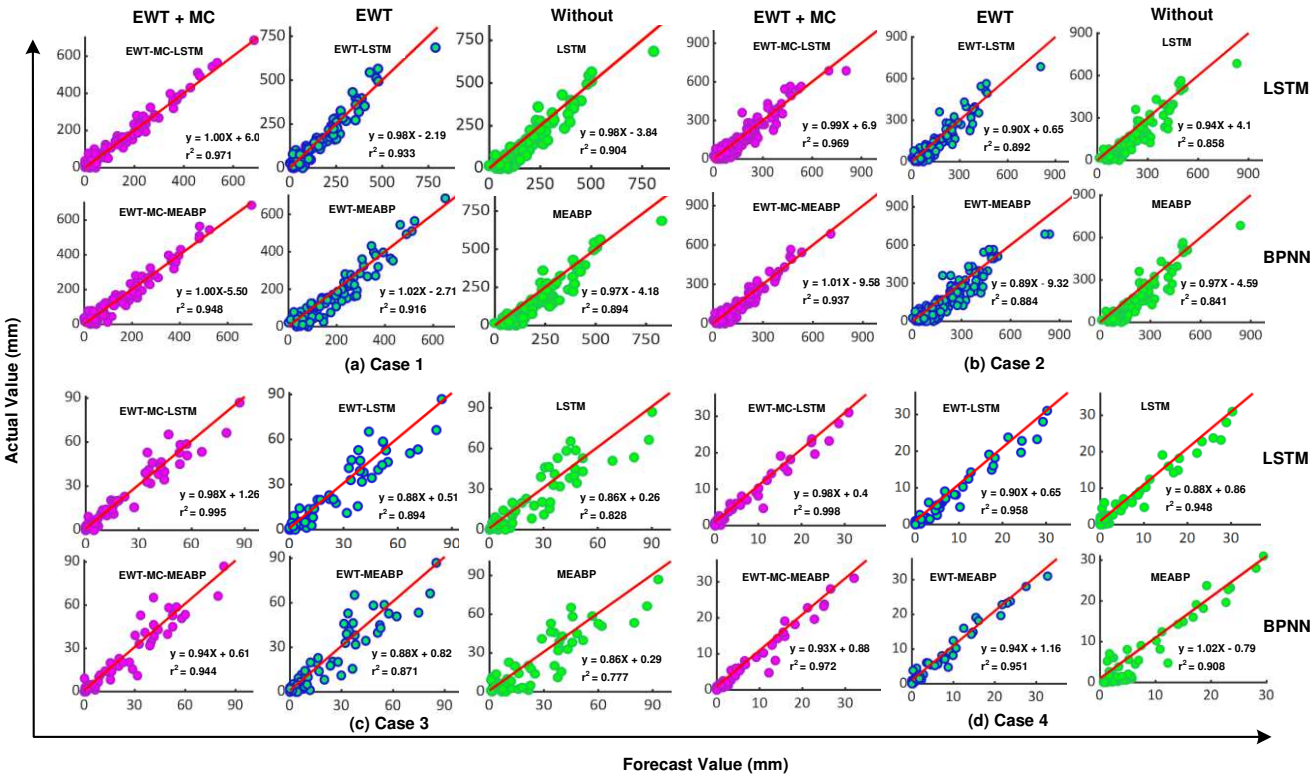


Fig. 6 Scatterplot of the forecasted and actual precipitation (mm) in the testing phase for the compared models by step one using the monthly data including the coefficient of determination (r^2) and a linear fit inserted in each panel for the four cases: (a) Hong Kong, (b) Macao, (c) Lanzhou and (d) Alashan League.

To further describe the effectiveness of the EWT and MC algorithm in enhancing the accuracy of LSTM and BPNN models in precipitation forecasting, test results of the six models for the four cases are illustrated in Fig. 6. Subgraphs (a-d) display a scatterplot showing the goodness-of-fit. Its correlation coefficient r is also shown to depict the extent of agreement between forecasted and observed precipitation. A higher value of r means that the predicted value is closer to the observed value. The 'Without' in Figure 6 refers to the individual prediction model without using EWT or MC. Taking the LSTM of Case1 as an example, 'EWT' and 'EWT+MC' represent the scatter plots of EWT+LSTM and EWT+MC+LSTM models, respectively. It can be seen that the scatters of the EWT-MC-LSTM models distribute a little more tightly around the linear fitting line compared with the other five models. Again, the EWT-MC-LSTM model is clearly better than the LSTM model without EWT and MC technologies in terms of r^2 (e.g., in case three: EWT-MC-LSTM≈0.995, EWT-LSTM≈0.894, LSTM ≈ 0.828) for Lanzhou. Overall, the effectiveness of EWT and MC in the forecasting of monthly rainfall time series is demonstrated; the proposed multi-

step EWT-MC-LSTM model convincingly outperforms the comparison models for all the study regions, confirmed by attaining the larger r^2 value.

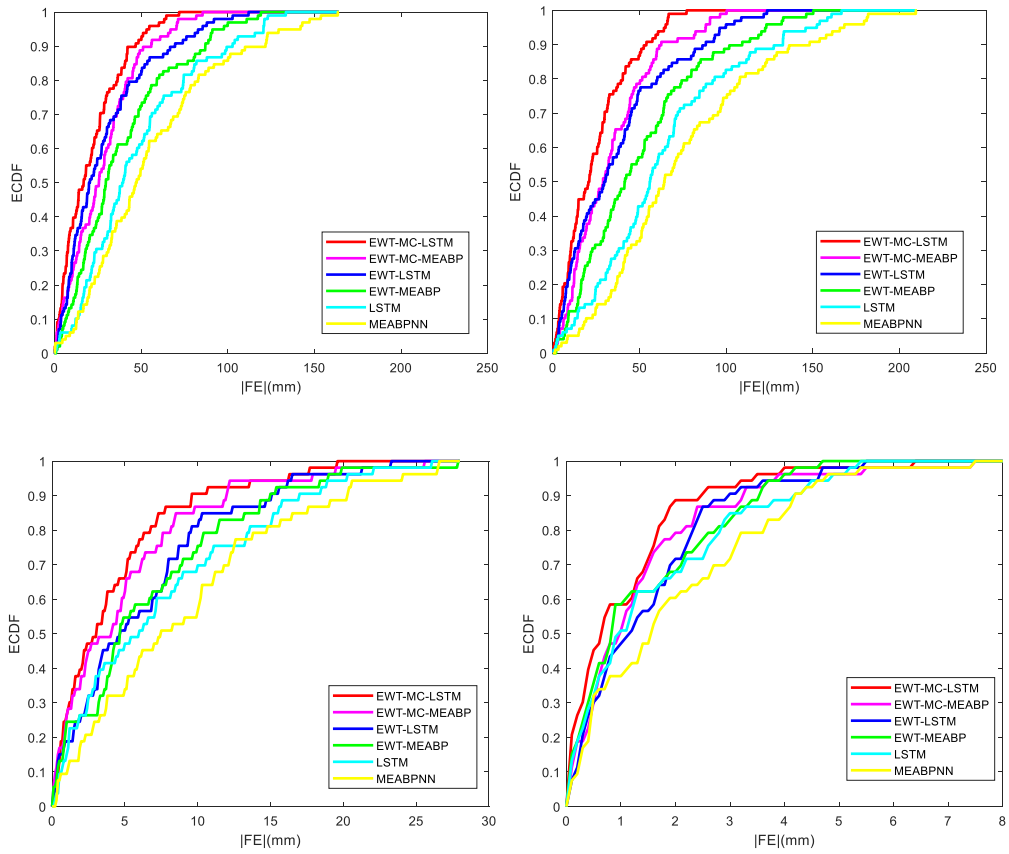


Fig.7 Empirical cumulative distribution function (ECDF) of the forecasted error |FE| (mm) generated by the proposed hybrid EWT-MC-LSTM model versus its counterpart models in step one for the four cases.

As with the compared methods shown in the studies (Mazuelas et al., 2011; Riahi-Madvar et al., 2021; Zhang et al., 2018b) , we show the probability distribution of the forecast errors through the empirical cumulative distribution function, which comes from the six compared methods mentioned in the paper. The percentage of the empirical cumulative distribution function (ECDF) was plotted at each region for different forecasting abilities in Fig. 7. According to this figure, the proposed EWT-MC-LSTM model was slightly better than the other five models for the four regions, and models incorporated with MC and EWT technologies were superior to the model only with EWT, and models incorporated with EWT technology performed better than the models without EWT method. Based on the percentage of errors in the bracket (0 to ±250 mm) for the Hong Kong and Macao, and bracket (0 to ±30 mm) for the Lanzhou, and bracket (0 to ±8 mm) for the Alashan League, Fig. 7 clearly confirms that the EWT-MC-LSTM method was the most responsive model in forecasting precipitation series data.

Fig. 8 compares boxplots of the six models in step one for the four regions. The outliers specified by '+' in every boxplot represent the extreme magnitudes of the forecasting error |FE| within the testing period (months) along with their upper quartile, median, and lower quartile values. The |FE| distribution of the EWT-MC-LSTM model shows a much lesser spread in case one, compared with the other five models. Similarly, the proposed EWT-MC-LSTM model again achieved good accuracy in terms of |FE| for Macao, Lanzhou, and Alashan League in relation to the counterpart models. By observing Fig. 11, the accuracy of the EWT-MC-LSTM model for each test region appeared to be better than the other five models.

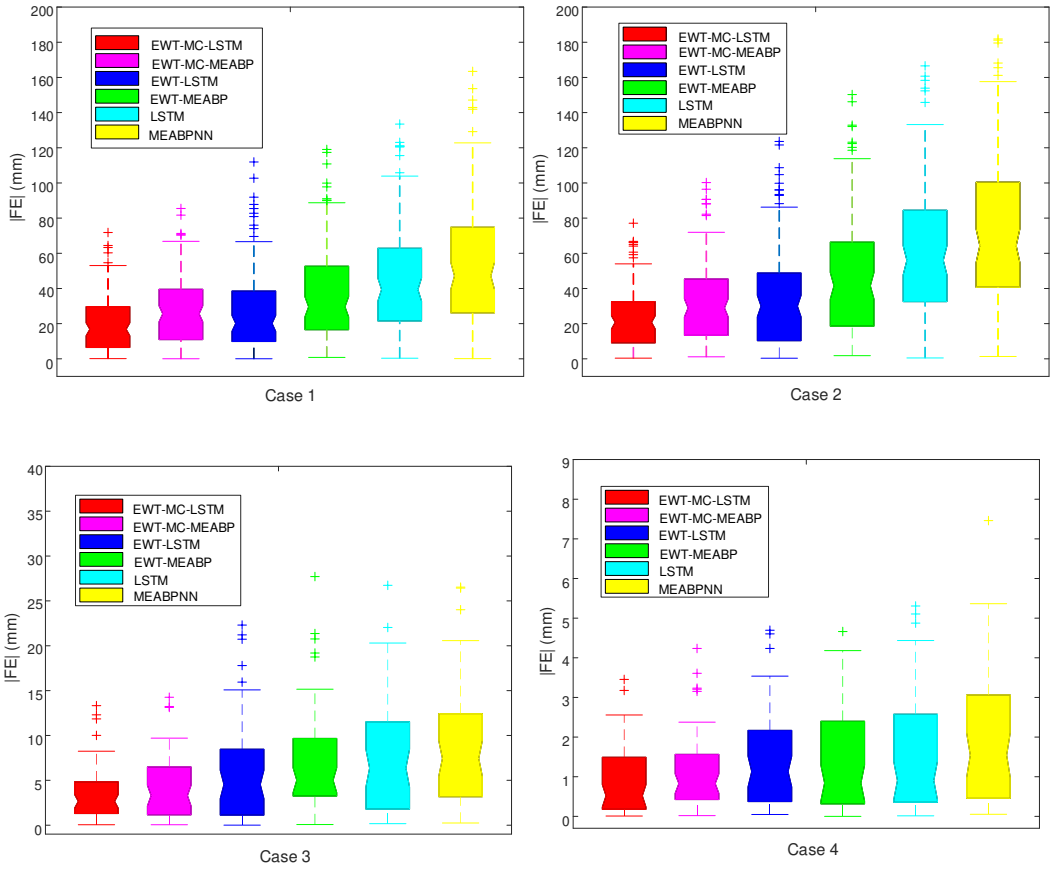


Fig.8 Box-plots of the forecasted error |FE| (mm) in step one for the proposed hybrid EWT-MC-LSTM model with counterpart models of monthly forecasted rainfall for the four cases.

Despite the superior error achieved by the proposed model, the following limitations remain in this study. First, since the proposed method is only for monthly rainfall, its applicability in other cases such as daily or sub-daily deserves further investigation. Second, the extreme tail characteristics of the rainfall data are not sufficiently considered, which affects the performance of the proposed model to some extent. Admittedly, the LSTM network can

learn distribution patterns in the data during the feature extraction process, but the interpretation of this aspect needs to be improved. Therefore, another focus for further research is to improve the accuracy and interpretability of deep learning model predictions by taking full account of the distribution characteristics of the rainfall data itself as well as the heavy-tailed scale.

6. Conclusions

This study develops a hybrid EWT-MC-LSTM model for precipitation time series forecasting. EWT is exploited to extract useful information and eliminate stochastic volatility from the original precipitation series. The deep learning LSTM network architecture incorporated with MC technology is employed as a predictor. To fully demonstrate the effectiveness of the proposed model, monthly data from four regions with different levels of discharge in every month of the twenty years are studied. Among the six models, including BPNN, MEA-BPNN, EWT-MEABP, EWT-MC-MEABP, LSTM, EWT-LSTM, and EWT-MC-LSTM, the EWT-MC-LSTM model performs the best according to five statistical indices (RMSE, MAE, MAPE, R^2, ECDF, and Box-plots). The following results have been concluded:

1. The EWT data preprocessing decomposition technique improves the prediction performance of the model, which demonstrated that EWT is suitable for data pre-analysis for precipitation time series.

2. Compared with the LSTM model, MC technology incorporated with LSTM deep learning network obtains better forecast results.

3. The proposed model is suitable for different magnitudes and fluctuations of monthly rainfall sequence prediction such as Hong Kong, Macao, Lanzhou, and Alashan League.

This study proposes a novel deep learning method based on long short-term memory network which incorporated with EWT as data preprocessing and MC technology, but the performance of deep learning with other network structures, such as Gate Recurrent Unit (GRU) and convolutional neural networks (CNN), have not been studied. More attention will be paid to the performances of different models for forecasting in future work.

Acknowledgments: We thank the reviewers and the Associate Editor for providing meticulous reviews, which helped us to significantly improve the initial manuscript. We thank National Oceanic and Atmospheric Administration (NOAA), Hong Kong Observatory and Macao Meteorological and Geophysical Bureau(SMG) for providing the information and dataset of precipitation.

References

Abbot, J., Marohasy, J., 2014. Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks. Atmospheric Research, 138: 166-178. DOI:10.1016/j.atmosres.2013.11.002

Aksoy, H., Dahamsheh, A., 2018. Markov chain-incorporated and synthetic data-supported conditional artificial neural network models for forecasting monthly precipitation in arid regions. Journal of Hydrology, 562: 758-779. DOI:10.1016/j.jhydrol.2018.05.030

Bohme, M., Pham, V.T., Roychoudhury, A., 2017. Coverage-based Greybox Fuzzing as Markov Chain. IEEE Transactions on Software Engineering: 1-1. DOI:10.1109/tse.2017.2785841

Cao, H., Fan, F., Zhou, K., He, Z., 2016. Wheel-bearing fault diagnosis of trains using empirical wavelet transform. Measurement, 82: 439-449. DOI:10.1016/j.measurement.2016.01.023

Ding, H., Dong, W., 2015. Chaotic feature analysis and forecasting of Liujiang River runoff. Soft Computing, 20(7): 2595-2609. DOI:10.1007/s00500-015-1661-1

Farajzadeh, J., Fakheri Fard, A., Lotfi, S., 2014. Modeling of monthly rainfall and runoff of Urmia lake basin using "feed-forward neural network" and "time series analysis" model. Water Resources and Industry, 7-8: 38-48. DOI:10.1016/j.wri.2014.10.003

Feng, Z.-k., Niu, W.-j., Tang, Z.-y., Xu, Y., Zhang, H.-r., 2021. Evolutionary artificial intelligence model via cooperation search algorithm and extreme learning machine for multiple scales nonstationary hydrological time series prediction. Journal of Hydrology, 595. DOI:10.1016/j.jhydrol.2021.126062

Gilles, J., 2013. Empirical Wavelet Transform. IEEE Transactions on Signal Processing, 61(16): 3999-4010. DOI:10.1109/tsp.2013.2265222

Grover, A., Kapoor, A., Horvitz, E., 2015. A Deep Hybrid Model for Weather Forecasting, Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15, pp. 379-386. DOI:10.1145/2783258.2783275

Gui, Y., Shao, J., 2017. Prediction of Precipitation Based on Weighted Markov Chain in Dangshan, Proceedings of the International Conference on High Performance Compilation, Computing and Communications - HP3C-2017, pp. 81-85. DOI:10.1145/3069593.3069617

Haidar, A., Verma, B., 2017. A novel approach for optimizing climate features and network parameters in rainfall forecasting. Soft Computing, 22(24): 8119-8130. DOI:10.1007/s00500-017-2756-7

HE Hui, JIN Long, QIN Zhi-nian, Li-jun, Y., 2007. Downscaling forecast for the monthly precipitation over Guangxi based on the BP neural network model. Journal of Tropical Meteorology: 1: 011.

J., S., H., W.S., 2018. Heavy Rainfall Forecasting Model Using Artificial Neural Network for Flood Prone Area. IT Convergence and Security: 68-76.

Kalteh, A.M., 2017. Enhanced Monthly Precipitation Forecasting Using Artificial Neural Network and Singular Spectrum Analysis Conjunction Models. INAE Letters, 2(3): 73-81. DOI:10.1007/s41403-017-0025-9

Kisi, O., Cimen, M., 2012. Precipitation forecasting by using wavelet-support vector machine conjunction model. Engineering Applications of Artificial Intelligence, 25(4): 783-792. DOI:10.1016/j.engappai.2011.11.003

Lan, K. et al., 2018. A Survey of Data Mining and Deep Learning in Bioinformatics. J Med Syst, 42(8): 139. DOI:10.1007/s10916-018-1003-9

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature, 521(7553): 436-44. DOI:10.1038/nature14539

Li, P.W., Lai, E.S.T., 2004. Short-range quantitative precipitation forecasting in Hong Kong. Journal of Hydrology, 288(1-2): 189-209. DOI:10.1016/j.jhydrol.2003.11.034

Liu, H., Tian, H.-q., Li, Y.-f., 2015a. Comparison of new hybrid FEEMD-MLP, FEEMD-ANFIS, Wavelet Packet-MLP and Wavelet Packet-ANFIS for wind speed predictions. Energy Conversion and Management, 89: 1-11. DOI:10.1016/j.enconman.2014.09.060

Liu, H., Tian, H., Liang, X., Li, Y., 2015b. New wind speed forecasting approaches using fast ensemble empirical model decomposition, genetic algorithm, Mind Evolutionary Algorithm and Artificial Neural Networks. Renewable Energy, 83: 1066-1075. DOI:10.1016/j.renene.2015.06.004

Liu, W. et al., 2017. A survey of deep neural network architectures and their applications. Neurocomputing, 234: 11-26. DOI:10.1016/j.neucom.2016.12.038

Mazuelas, S., Shen, Y., Win, M.Z., 2011. Belief Condensation Filter for Navigation in Harsh Environments, 2011 IEEE International Conference on Communications (ICC), pp. 1-6. DOI:10.1109/icc.2011.5963104

Najafabadi, M.M. et al., 2015. Deep learning applications and challenges in big data analytics. Journal of Big Data, 2(1). DOI:10.1186/s40537-014-0007-7

Napolitano, G., Serinaldi, F., See, L., 2011. Impact of EMD decomposition and random initialisation of weights in ANN hindcasting of daily stream flow series: An empirical examination. Journal of Hydrology, 406(3-4): 199-214. DOI:10.1016/j.jhydrol.2011.06.015

Nourani, V., Kisi, Ö., Komasi, M., 2011. Two hybrid Artificial Intelligence approaches for modeling rainfall–runoff process. Journal of Hydrology, 402(1-2): 41-59. DOI:10.1016/j.jhydrol.2011.03.002

Oki., T., Kanae., S., 2006. Global hydrological cycles and world water resources. Science: 313(5790): 1068-1072.

Ouyang, Q., Lu, W., 2017. Monthly Rainfall Forecasting Using Echo State Networks Coupled with Data Preprocessing Methods. Water Resources Management, 32(2): 659-674. DOI:10.1007/s11269-017-1832-1

PaoloBurlando., RenzoRosso., G.Cadavid., L., D.Salas., J., 1993. Forecasting of short-term rainfall using ARMA models. Journal of Hydrology: 144(1-4): 193-211.

Papalexiou, S.M., 2018. Unified theory for stochastic modelling of hydroclimatic processes: Preserving marginal distributions, correlation structures, and intermittency. Advances in Water Resources, 115.

Papalexiou, S.M., Markonis, Y., Lombardo, F., AghaKouchak, A., Foufoula-Georgiou, E., 2018. Precise Temporal Disaggregation Preserving Marginals and Correlations (DiPMaC) for Stationary and Nonstationary Processes. Water Resources Research, 54(10).

Papalexiou, S.M., Serinaldi, F., 2020. Random Fields Simplified: Preserving Marginal Distributions, Correlations, and Intermittency, With Applications From Rainfall to Humidity. Water Resources Research, 56(2).

Peng, T., Zhou, J., Zhang, C., Fu, W., 2017. Streamflow Forecasting Using Empirical Wavelet Transform and Artificial Neural Networks. Water, 9(6). DOI:10.3390/w9060406

Qiu, M. et al., 2017. A Short-Term Rainfall Prediction Model Using Multi-task Convolutional Neural Networks, 2017 IEEE International Conference on Data Mining (ICDM), pp. 395-404. DOI:10.1109/icdm.2017.49

Ramezani, F., Nikoo, M., Nikoo, M., 2014. Artificial neural network weights optimization based on social-based algorithm to realize sediment over the river. Soft Computing, 19(2): 375-387. DOI:10.1007/s00500-014-1258-0

Riahi-Madvar, H., Dehghani, M., Memarzadeh, R., Gharabaghi, B., 2021. Short to Long-Term Forecasting of River Flows by Heuristic Optimization Algorithms Hybridized with ANFIS. Water Resources Management. DOI:10.1007/s11269-020-02756-5

Santamaría-Bonfil, G., Reyes-Ballesteros, A., Gershenson, C., 2016. Wind speed forecasting for wind farms: A method based on support vector regression. Renewable Energy, 85: 790-809. DOI:10.1016/j.renene.2015.07.004

Shaari, M.A., Samsudin, R., Ilman, A.S., 2018. Forecasting Drought Using Modified Empirical Wavelet Transform-ARIMA with Fuzzy C-Means Clustering. Indonesian Journal of Electrical Engineering and Computer Science, 11(3). DOI:10.11591/ijeecs.v11.i3.pp1152-1161

Sun C, Sun Y, L, W., 1998. Mind-evolution-based machine learning: framework and the implementation of optimization, Proc. of IEEE Int. Conf. on Intelligent Engineering Systems, pp. 355-359.

Sun, M., Kim, G., 2016. Quantitative Monthly Precipitation Forecasting Using Cyclostationary Empirical Orthogonal Function and Canonical Correlation Analysis. Journal of Hydrologic Engineering, 21(1). DOI:10.1061/(asce)he.1943-5584.0001244

Sun, M., Li, X., Kim, G., 2018. Precipitation analysis and forecasting using singular spectrum analysis with artificial neural networks. Cluster Computing. DOI:10.1007/s10586-018-1713-2

Tao, L., He, X., Wang, R., 2017a. A Hybrid LSSVM Model with Empirical Mode Decomposition and Differential Evolution for Forecasting Monthly Precipitation. Journal of Hydrometeorology, 18(1): 159-176. DOI:10.1175/jhm-d-16-0109.1

Tao, Y., Gao, X., Ihler, A., Sorooshian, S., Hsu, K., 2017b. Precipitation Identification with Bispectral Satellite Information Using Deep Learning Approaches. Journal of Hydrometeorology, 18(5): 1271-1283. DOI:10.1175/jhm-d-16-0176.1

Wang, W., Tang, R., Li, C., Liu, P., Luo, L., 2018. A BP neural network model optimized by Mind Evolutionary Algorithm for predicting the ocean wave heights. Ocean Engineering, 162: 98-107. DOI:10.1016/j.oceaneng.2018.04.039

Xiaodan Zhu, Parinaz Sobhani, Guo, H., 2015. Long short-term memory over recursive structures, International Conference on Machine Learning, pp. 1604-1612.

Xingjian Shi, Zhihan Gao, Leonard Lausen, Hao Wang, Yeung, D.-Y., 2017. Deep learning for precipitation nowcasting A benchmark and a new model, Advances in Neural Information Processing Systems, pp. 5617-5627.

Xingjian Shi, Zhourong Chen, Hao Wang, Yeung, D.-Y., 2015. Convolutional LSTM network, A machine learning approach for precipitation nowcasting. Advances in neural information processing systems: 802-810.

Yang, M., Sang, Y.-F., Liu, C., Wang, Z., 2016. Discussion on the Choice of Decomposition Level for Wavelet Based Hydrological Time Series Modeling. Water, 8(5). DOI:10.3390/w8050197

Yu, P.-S., Chen, C.-J., Chen, S.-J., 2000. Application of gray and fuzzy methods for rainfall forecasting. Journal of Hydrologic Engineering: 5(4): 339-345.

Zhang, P., Jia, Y., Gao, J., Song, W., Leung, H.K.N., 2018a. Short-term Rainfall Forecasting Using Multi-layer Perceptron. IEEE Transactions on Big Data: 1-1. DOI:10.1109/tbdata.2018.2871151

Zhang, P., Zhang, L., Leung, H., Wang, J., 2017. A Deep-Learning Based Precipitation Forecasting Approach Using Multiple Environmental Factors, 2017 IEEE International Congress on Big Data (BigData Congress), pp. 193-200. DOI:10.1109/BigDataCongress.2017.34

Zhang, X., Wang, J., Li, J., Chen, W., Liu, C., 2018b. CRlncRC: a machine learning-based method for cancer-related long noncoding RNA identification using integrated features. BMC Med Genomics, 11(Suppl 6): 120. DOI:10.1186/s12920-018-0436-9