



Development of wavelet-based machine learning models for predicting long-term rainfall from sunspots and ENSO

Yuan-Chien Lin^{1,2,3} · Tzu-Han Weng¹

Received: 1 August 2023 / Accepted: 15 November 2023 / Published online: 21 December 2023
© The Author(s) 2023

Abstract

The variations in rainfall and its spatial and temporal distribution in wet and dry seasons have increased substantially globally owing to the effect of climate change. These disparities can lead to droughts and severe water shortages, as exemplified by the unprecedented drought in Taiwan in 2021, which is considered the worst in 50 years. From a broader perspective, the overall climate and water resources on Earth are influenced by factors, such as the El Niño phenomenon and solar activity. Accordingly, this study examines the relationship between rainfall and planetary- or large-scale influencing factors, such as sunspots and the El Niño-Southern Oscillation. Additionally, rainfall patterns under various conditions are predicted using machine learning models combined with wavelet analysis. These models use 60-years historical data to build models, and the Bayesian network model exhibited the best overall prediction accuracy (85.7%), with sunspots emerging as the most influential factor. The novel findings of this study strongly confirmed that the relationship between sunspot and local rainfall patterns can serve as a valuable reference for water resources management and planning by relevant organizations.

Keywords Planetary scale · Sunspots · Rainfall pattern · Machine learning · Water resources · Bayesian network

Introduction

In recent years, the impact of climate change and the natural terrain characteristics of high mountains and rapid water in several areas globally have resulted in the uneven rainfall distribution (in time and space), and the agricultural, industrial, and economic developments with the growing population density globally have resulted in the unfavorable utilization of water resources and groundwater (Allan et al. 2013; Misra 2014; Ritzema and Van Loon-Steensma 2018; Varis and Vakkilainen 2001). Moreover, changes in the spatial and temporal characteristics of rainfall have resulted in several extreme hydrological events and environmental problems. Although the annual average rainfall of Taiwan is

approximately 2500 mm, which is 2.6 times higher than the average annual rainfall of the world, Taiwan is still regarded as a water-poor region (Narvaez et al. 2022; Shiau and Hsiao 2012). The main rainfall seasons in Taiwan are concentrated in the Meiyu season, from May to June, and the typhoon season, from July to September, (Chen and Chen 2003; Yim et al. 2015). Particularly, there is a significant difference in the rainfall behavior in the wet and dry seasons. Therefore, slight changes in the rainfall during the rainy season will directly or indirectly affect the risk of drought and flood disasters in Taiwan (Chen et al. 2009; De Silva and Kawasaki 2018).

In meteorology, the scale refers to the size and duration of the weather system. Generally, the more commonly used scale is the scale of the atmospheric weather system defined by Orlanski (1975), and it can be divided into small scales with a horizontal range of less than 2 km, mesoscale ranging from 2 to 2000 km, and large scale above 2000 km. The large scale can be further subdivided into the comprehensive scale of more than 2000 km with a life span of approximately 2–10 days (e.g., frigid cyclones, extratropical cyclones, and jet streams) and planetary-scale weather systems with a horizontal scale of more than 10,000 km and a life span of more than 10 days.

✉ Yuan-Chien Lin
yclin@ncu.edu.tw

¹ Department of Civil Engineering, National Central University, No. 300, Zhongda Rd., Zhongli District, Taoyuan 32001, Taiwan

² Research Center for Hazard Mitigation and Prevention, National Central University, Taoyuan 32001, Taiwan

³ Graduate Institute of Environmental Engineering, National Central University, Taoyuan 32001, Taiwan

It is essential to scale up the discussion of the changing characteristics of climate and water resources to a relatively macroscopic spatio-temporal scale, and the scale corresponding to meteorological data can be at the mesoscopic scale or above the planetary scale (Barthel and Banzhaf 2016). Existing studies have demonstrated that the climate phenomenon of earth may be affected by the outer space of the earth and solar activity (Marsh and Svensmark 2003).

Sunspots are temporary phenomena on the photosphere of the sun, typically representing the strength of the solar activity, and they appear darker than the surrounding areas under visible light. The strong magnetic field activity of the sun suppresses convection, resulting in a relatively low surface temperature (approximately 3000–4500 K), and the darker areas are important indicators of solar activity. When sunspots are active, they affect the magnetic field of the Earth, resulting in bad weather and may even damage electronic products and electrical appliances (Cowling 1933). Another important large-scale climate factor is the El Niño-Southern Oscillation (ENSO), and when it occurs, it causes climate anomalies globally, such as a sharp drop in rainfall, severe drought, and forest fires (e.g., the forest fires in Indonesia, India, and Australia during the El Niño year). In contrast, ENSO may also cause an increase in rainfall (e.g., the increase in rainfall in the eastern Pacific during the El Niño year) and mild winter conditions (e.g., Canada experienced mild winter conditions during the El Niño year) (Yeh et al. 2009), and these factors may influence macroscopic spatio-temporal scale hydrological characteristics.

Currently, it is well known that the spatial and temporal characteristics of rainfall varies significantly in many areas (Cristiano et al. 2017; Kao et al. 2013). In recent years, several studies have analyzed factors that potentially affect rainfall to explore the relationship between long-term rainfall characteristics and potential macroscopic influencing factors, such as sunspots (Ananthakrishnan and Parthasarathy 1984; Bhattacharyya and Narasimha 2005; Seleshi et al. 1994). For example, by analyzing the time-frequency correlation between sunspots and annual rainfall through cross wavelet transform (XWT), studies have reported a significant correlation between sunspots and regional rainfall phenomena at a frequency cycle of 8–12 years (Nazari-Sharabian and Karakouzian 2020; Thomas and Abraham 2022). However, only few studies have attempted to construct prediction models to correlate this relationship. In addition, some studies have employed the commonly used indexes of ENSO, such as southern oscillation index (SOI) and Multivariate ENSO Index to determine the relationship between ENSO and the average rainfall of an area (Indeje et al. 2000; Kuo et al. 2010; Vladimiro and Guido 2018). For example, studies have reported that the average annual rainfall in Australia increases with an increase in the intensity of the anti-El Niño phenomenon, and vice versa (García-García

and Ummenhofer 2015). In recent years, some large-scale studies have been conducted on the impact of the El Niño phenomenon on the climate and water resources of Taiwan (Jiang et al. 2003; Lee et al. 2020). For example, some studies have discussed the varying characteristics of different seasonal rainfall or spring rainfall, or the characteristics of typhoon intensity in the Northwest Pacific (Chu 2004; Wang et al. 2020), and some studies have explored the long-term changes in the rainfall characteristics of Taiwan through sediment core drilling (Chen et al. 2019). These studies demonstrated that factors on macroscopic space and time scales may affect the characteristics of large-scale or periodic regional water resources and the overall medium-term and long-term trends, including possible seasonal or interannual changes in abundance and drought and the impact on drought and flood disasters.

Therefore, using actual ground weather observation data in Taiwan combined with remote sensing data at planetary scales, this study extracted features and investigated the relationship between rainfall in Taiwan and the factors that may induce changes in hydrological and water resources (e.g., sunspot and ENSO) using wavelet signal analysis method. In addition, machine learning methods were employed to predict and classify rainfall amount, and a set of rainfall-water resources warning system was established. We believe that the findings of this study will provide a better understanding of the rainfall mechanism in Taiwan and provide relevant government agencies with a more accurate grasp of large-scale rainfall fluctuation trends, and serve as a reference for water resources management policy in Taiwan.

Study area and data

The main study area of this study is the main island of Taiwan. Taiwan is located on the west side of the Pacific Ocean, and it is surrounded by the ocean. It is an island in the island arc of East Asia, and its longitude is from 120 to 122° East, and the latitude is from 22 to 25° North. The climate of Taiwan is between the tropical monsoon climate and the subtropical monsoon climate. The average annual rainfall is 2515 mm, which is 2.6 times higher than the average rainfall of the world. The main rainfall seasons are concentrated in the rainy season, from May to June, and the typhoon season from July to September.

To obtain data on the meteorological factors, this study selected 16 stations from the Central Weather Bureau (CWB) in Taiwan, and analyzed the data for 60 years, from 1960 to 2020, using the monthly monitoring data of meteorological factors, such as rainfall and humidity. The spatial distribution of the measuring stations is shown in Fig. 1, and the detailed information is presented in Table 1. The observation data of the sunspot number (SSN) and El Niño-SOI

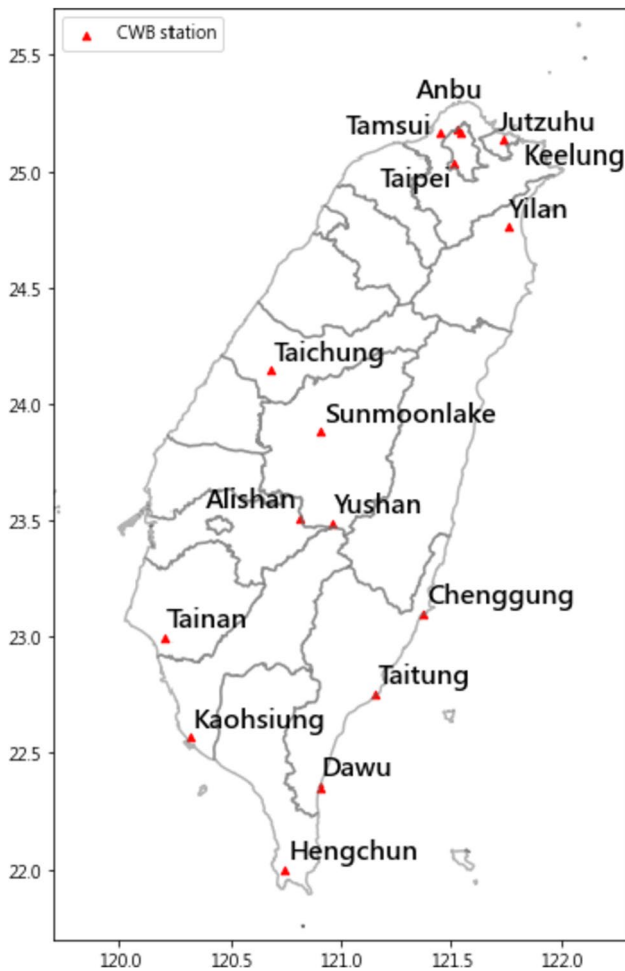


Fig. 1 Map showing the spatial distribution of the local weather stations (manual observation) of the Central Weather Bureau (CWB), Taiwan

were obtained from the U.S. Space Weather Prediction Center (SWPC) of the U.S. National Oceanic and Atmospheric Administration (NOAA) and the National Center for Environmental Information (NCEI), and data from January 1, 1960 to December 31, 2020 (a total of 60 years) were collected and analyzed.

Methodology

Study framework

The detailed process of this research is mainly divided into three parts. The first part involves the data collection and arrangement, including meteorological data, SSN, and El Niño-SOI data for 60 years from 1960 to 2020. The second part is the wavelet analysis and feature extraction. This part involves the disassembly of the collected data using the continuous wavelet transform (CWT) and inverse wavelet

Table 1 Basic information of the data from the local weather stations (manual observation) of CWB, Taiwan

Station name	Start and end time	Observation items used
Tamsui	1942/10/01 ~	RH01
Anbu	1943/01/01 ~	PP01
Taipei	1896/08/11 ~	
Jutzuhu	1943/01/01 ~	
Jilung	1946/10/01 ~	
Yilan	1935/12/06 ~	
Tainan	1897/01/01 ~	
Kaohsiung	2022/01/24 ~	
Taichung	1897/01/01 ~	
Alishan	1933/04/01 ~	
Dawu	1940/01/01 ~	
Yushan	1943/10/01 ~	
Hengchun	1896/11/20 ~	
Chengkung	1940/01/01 ~	
Sunmoonlake	1941/11/01 ~	
Taitung	1901/01/01 ~	

transform methods for the signal filtering of different frequencies of the signal. Subsequently, the correlation between the meteorological factors of each station and the SSN and SOI was investigated using wavelet coherence analysis. The complete spatial distribution was estimated and mapped using the radial basis function network (RBFN), and the rainfall amount under different conditions were predicted and classified, and compared to each other using several machine learning models, including Naive Bayes, decision tree, and random forest models.

Wavelet signal analysis method

There are several types of signal analysis techniques, and among these techniques, wavelet analysis technology has emerged as one of the important tools for signal analysis and processing in recent years (Adamowski and Chan 2011; Boggess and Narcowich 2015; Kuo et al. 2010; Thomas and Abraham 2022; Yu and Lin 2015). When monitoring dynamic signals, the data of the signal in the frequency domain can be analyzed using Fourier-based analysis. However, the analyzed signal can only be disassembled and analyzed under the premise of linearity and stability, which is extremely unfavorable when observing abnormal signals. The wavelet method is a multi-resolution analysis that converts the time domain and the frequency domain for continuous signals. Compared to other signal processing techniques, it can better localize time features and can analyze non-stationary signals. The main purpose of the wavelet method is to decompose the time series through the signal and express it as a function of several frequency combinations (Boggess

and Narcowich 2015); thus, both extreme and short-term events can be discussed.

The wavelet analysis techniques used in this study include the CWT, XWT, and wavelet coherence (WTC) analysis. Using the MATLAB toolbox developed by Grinsted et al. (2004), the main theories are described as follows:

CWT

The CWT is a function used to decompose a continuous time function and transform it into several wavelets. Compared to the Fourier transform analysis, CWT can: construct a time domain with localized time–frequency signals in the frequency domain; investigate the dynamic characteristics of the signal through the changes in the signal under different magnifications; understand the irregular periodic changes of the signal and analyze the nonlinear signal change period; and display the non-stationary signal using the characteristics of the time–frequency structure. As expressed in Eq. (1), the wavelet sequence can be obtained after expansion and translation through the mother wavelet function (i.e., Morlet Wavelet; $\varphi(t)$), where s is the scale factor related to frequency and τ is the shift factor related to time.

$$WT_x(s, t) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} x(t) \varphi\left(\frac{t-\tau}{s}\right) dt \quad (s \in \mathbb{R}^+, \tau \in \mathbb{R}) \quad (1)$$

XWT

For two time series x_n and y_n , its XWT is defined as $W^{XY} = W^X W^{Y*}$, where $*$ represents a conjugate complex number. We can further define $|W^{XY}|$ as the cross wavelet power, and its complex argument $\arg(W^{XY})$ is the local relative phase of x_n and y_n in the time–frequency space. The theoretical distribution of the cross wavelet power of the background power spectrum of the two time series is (Torrence and Compo 1998):

$$D\left(\frac{|W_n^X(s)W_n^{Y*}(s)|}{\sigma_X \sigma_Y} < p\right) = \frac{Z_\nu(p)}{\nu} \sqrt{P_k^X P_k^Y} \quad (2)$$

Here, P_k^X and P_k^Y are the theoretical distribution of the cross wavelet power of two time series with background power spectra, $Z_\nu(p)$ is the confidence level of the probability p of the square root of the probability density function multiplied by two χ^2 distributions. For more detailed theoretical derivation, please refer to the following reference (Grinsted et al. 2004; Torrence and Compo 1998).

WTC

The cross wavelet power obtained through XWT can show some areas with common higher power, and WTC is an analysis method to further understand the correlation of the two time series in the time–frequency domain. According to Torrence and Compo (1998), we defined the WTC of two time series as expressed in Eq. (3).

$$R_n^2(s) = \frac{|S(s^{-1}W_n^{XY}(s))|^2}{S(s^{-1}|W_n^X(s)|^2) \cdot S(s^{-1}|W_n^Y(s)|^2)} \quad (3)$$

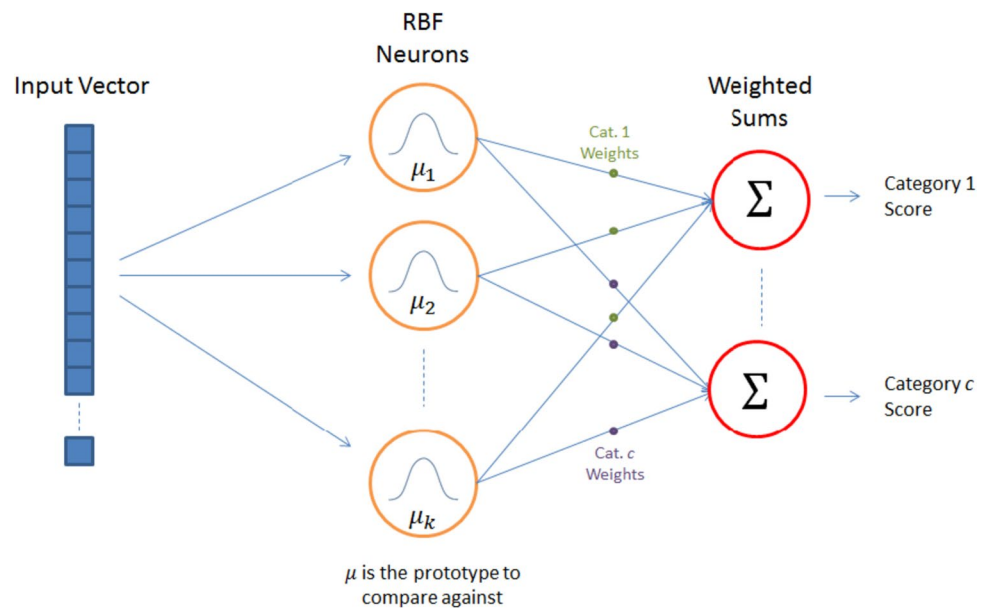
This definition is very close to the definition of the traditional correlation coefficient, which is well known, so WTC can be regarded as the correlation coefficient of two time series in the time–frequency domain, as this will be easier to understand in terms of interpretation. In Eq. (3), S is the smoothing operator when calculating the weighted moving average, and $W_n^X(s)$ and $W_n^Y(s)$ are the wavelet coefficients of X and Y , respectively. Owing to the wordcount limitations, please refer to the following reference for detailed definitions (Grinsted et al. 2004; Torrence and Compo 1998).

RBFN

In this study, the correlation coefficients of signals between the local meteorological data and SSN and El Nino-SOI were calculated using the WTC analysis, and through the spatial estimation, the coherence spatial distribution of certain frequency band with higher coherence was estimated for the whole Taiwan to better understand the changing trend in the WTC.

Compared to traditional methods, the training process of neural network models, which emerged in recent years, has the advantages of speed and simplicity, and can better handle nonlinear and complex systems. Accordingly, neural network models based on different architectures have been constructed. The RBFN (Chen et al. 2018; Chen et al. 1991; Ding et al. 2018; Orr 1996; Toit 2008) is a neural network for supervised learning, and it exhibits a three-layer feed-forward network architecture, as shown in Figure 2. The construction process of this model is faster and simpler than those of other multi-layer deep learning models. The input layer is composed of perceptual units, which are mapped to the output layer in a nonlinear manner, and the hidden layer is subjected to linear and nonlinear transformations to accelerate network operations. The RBFN estimates nonlinear data into a high-dimensional space through the radial basis function of the hidden layer, making it linearly separable. Currently, RBFN is commonly used in operations, such as

Fig. 2 Schematic of the architecture of the radial basis function network (RBFN)



nonlinear function approximation, time series analysis, spatial estimation, and system modeling.

Machine learning models

With the rapid development of technology and the ease of obtaining data, the development of AI and big data analysis technology can be considered to be relatively mature. In recent years, machine learning and deep learning methods have been applied in several studies for classification and prediction. Therefore, this study employed four machine learning classifiers to predict and classify rainfall amount under different conditions.

Naive Bayes classifier

The simple Bayesian classifier was developed based on the Bayes’ theorem (Lin et al. 2015; Rish 2001; Webb et al. 2010), which is mathematically expressed in Eq. (4). The simple Bayesian classifier is based on the Bayesian classifier with the assumption of conditional independence, and it exhibits a fast and simple operation.

It is difficult to consider the independence of conditions from reality, so in 1997, Domingos and Pazzani proposed that when the assumption of independence of conditions is not established, as the loss function in its classifier is a zero–one loss function, it impact on the accuracy of the model is slight (Domingos and Pazzani 1997).

$$P(A|B) = \frac{P(A)P(A|B)}{P(B)} \tag{4}$$

As Bayes’ theorem derives event probability based on conditional probability, the simple Bayesian classifier is suitable for processing discrete attribute data, such as gender. However, most observation data in real life are continuous data, so the accuracy of the model can be improved by appropriately discretizing continuous data.

Decision tree classifier

Decision tree is often applied to data classification, and it is a nonlinear data classification method (Myles et al. 2004). The structure of this model is similar to that of a tree, starting from the root and branching out from the children nodes under different conditions until the end. Each child node represents a specific category or object, and will eventually be classified into a certain category to complete the classification. The basic concept of classification and regression tree (CART) is to use a recursive method to binary cut a large amount of complex data, and divide the data into two sub-data sets at each node.

When the two sets of data are classified into the same category, or when the nodes can no longer be divided, the construction of the decision tree is completed. For the basis of decision tree classification nodes, the commonly used index is Gini, and it can be calculated as follows:

$$\text{Gini}(\text{node}) = 1 - \sum p(i|t)^2 \tag{5}$$

When all the data are classified consistently, the Gini Index is 0; when the Gini Index is larger, it indicates that the data categories of this node are scattered. Therefore, CART aims to reduce the Gini Index value to complete the classification of each node, and divide the nodes according

to this principle, so that the data property category of each node is closer.

Random forest classifier

The machine learning model architecture of random forest was proposed by Leo Breiman (Fig. 3) (Belgiu and Drăguț, 2016; Breiman 2001). The random forest model creates multiple decision trees in a random manner, and each decision tree is independent of each other. All the decision trees vote to determine the output category. The random forest is more stable than a single decision tree under multiple variables or characteristics (Breiman 2001). To form multiple trees with differences, first, it is essential to randomly resample the training set into n groups of subsets with a Bootstrap, and randomly select m variables. This is to select the minimum Gini Index segmentation method to generate multiple CART trees with differences, and then select the best classification result by combining the classification results of n decision trees.

Through the random forest classifier, we can analyze the category data and achieve a high accuracy, while suppressing the noise of the training samples. Therefore, this study utilized humidity, SSN, SOI, and the wavelet analysis results as the classification factors, and the local rainfall anomalies of each station are divided into five grades according to their size: extremely high, high, medium, low, and extremely low. The random forest was used to train the rainfall observation data over the years: 80% of the data is the training set, and 20% is the testing set, which was used to predict and classify the rainfall amount under different large-scale factors.

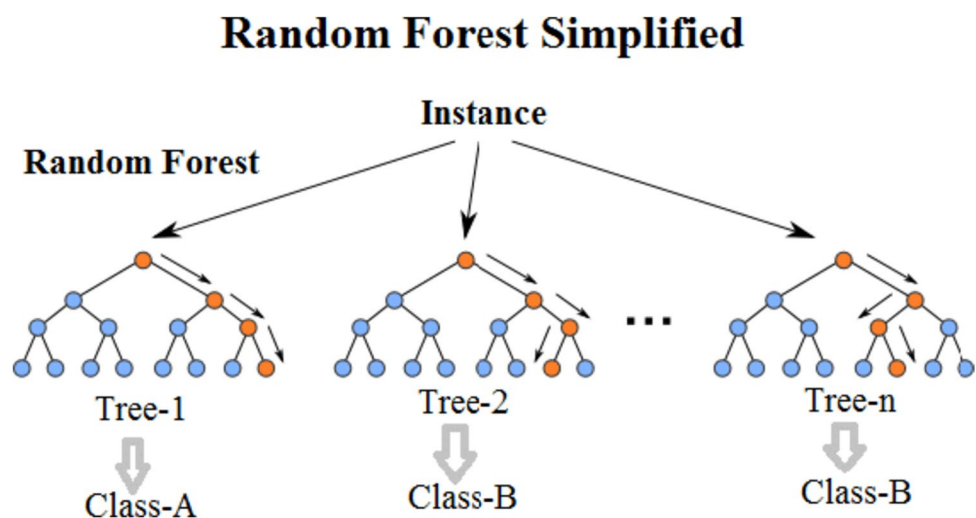
Bayesian networks

Bayesian networks (BN) are probabilistic graphical models used to represent and analyze causal relationships among

factors. This concept combines Bayesian theorem with graphical modeling, also known as probabilistic networks. This model not only intuitively represents complex relationships among factors through graphs and network structures but also enables easy understanding and visualization. Furthermore, it effectively handles uncertainty by conducting probability assessments. Additionally, one significant feature of the model is its flexibility in incorporating different control variables or datasets, and dynamically updating them to obtain conditional probabilities of target factors. Typically, BNs are highly accurate in simulating and predicting causal relationships between variables, making them widely applicable in various fields, including medical diagnosis and risk assessment. Overall, BNs are efficient, reliable, and flexible data analysis tools that play a crucial role in solving complex problems, making them suitable as decision support aids (Ben 2007). The BN model is a type of directed acyclic graph, where the causal relationships between variables are represented by directed edges and nodes. Conditional probability represents the probability of an event occurring given that another event has already occurred.

GeNIe is a graphical BN modeling and analysis tool developed by the Decision Systems Laboratory (DSLAB) at the University of California, Los Angeles (UCLA). It is used to establish models of uncertainty and causal relationships between variables, and it provides a comprehensive interface for users to easily build and analyze BN models. With GeNIe, users can not only construct BN models among variables but can also calculate conditional probabilities, enabling their use in applications, such as probability prediction and decision analysis for future events. This study aimed to construct a BN model using GeNIe to analyze the relationship between rainfall patterns in different regions of Taiwan and various variables. The graphical interface and features of GeNIe were utilized in this study to construct a comprehensive BN model that captures the uncertainty and

Fig. 3 Schematic of the random forest process (Breiman 2001)



causal dependencies among the variables. By incorporating data on rainfall patterns and other relevant factors, the BN model will provide insights into the probabilistic relationships between these variables and enable predictions and analysis related to rainfall in different regions of Taiwan. This study leveraged on the capabilities GeNIe to calculate conditional probabilities and facilitate decision-making and analysis based on the constructed BN model.

Model evaluation index

Confusion Matrix is one of the methods commonly used to evaluate the quality of a model. Visual supervision is the classification result of learning. If the target is a binary classification item, the confusion matrix is a 2×2 square matrix:

$$\begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix}$$

In its two-dimensional square matrix, the category of behavior prediction classification is listed as the category of actual classification, and the elements in the matrix are True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN). The Accuracy, Precision, Recall and F1-score can be calculated using the elements in the confusion matrix.

Accuracy:

$$\text{Accuracy} = \frac{TP + TN}{P + N} \quad (6)$$

Precision:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

Recall:

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{TP}{P} \quad (8)$$

F1-score:

$$\text{F1 score} = \frac{2TP}{2TP + FP + FN} \quad (9)$$

In addition to the confusion matrix, cross-validation can also be applied to evaluate the pros and cons of machine learning models. In this study, we randomly cut the data into multiple small subsets, a part of which is the training set and the other part is the testing set. The main purpose of this was to prevent the training process from significantly relying on a specific training and test set data, resulting in deviation or overfitting. This study utilized the K-fold cross-validation (K-fold CV) to randomly divide the data into k subsets, one of which is the test set and the rest is the training set. After running, a validation error is obtained. The

action was repeated until each sub-set data was used as the test set data, and the process was stopped. The action was executed for a total of k times, and k validation errors were generated. Lastly, the average value of the k validation errors was used as an indicator to determine whether the model is good or bad (Browne 2000).

Results

Analysis of the meteorological and large-scale factors

SSN

Research on the variation in sunspot cycles contributes to a deeper understanding of solar activity and its impact on Earth. In this study, data on SSN were collected and compiled from SWPC and NCEI for 60 years (from January 1, 1960 to December 31, 2020). Figure 4 illustrates the time series of the SSN at different time scales, reflecting the variation in the sunspot cycles and characteristics, such as maximum and minimum values. The image revealed that the SSNs followed an approximately 11-year solar cycle, where each cycle spans from the minimum value to the next minimum value.

Furthermore, the time series plot confirmed that the SSNs entered a period of low activity since 2014. The maximum values during this period are relatively lower compared to those of previous cycles, and this low-activity phase persisted until the end of 2019. Additionally, the SSNs during this period remained consistently low, indicating weaker solar activity compared to that of the previous 50 years. However, from 2020, the SSNs gradually increased, indicating a gradual resurgence of solar activity. Based on the observed variation in the cycle, it is expected that the SSNs will continue to increase in the coming years, leading to a stronger solar activity level.

SOI

It is essential to investigate the ENSO cycle to understand global weather patterns and climate change trends. A clearer understanding of the characteristics and underlying physical mechanisms of ENSO variability can be achieved by combining various methods and techniques. In this study, the data for a period of 60 years from January 1, 1960 to December 31, 2020 were collected and compiled from SWPC and NCEI. Figure 5 reveals that El Niño and La Niña events exhibit a periodicity of approximately 2–7 years over the years. Several studies have reported a significant correlation between the ENSO variability and the SOI as a measure of the ENSO state. Additionally, the figure reveals an

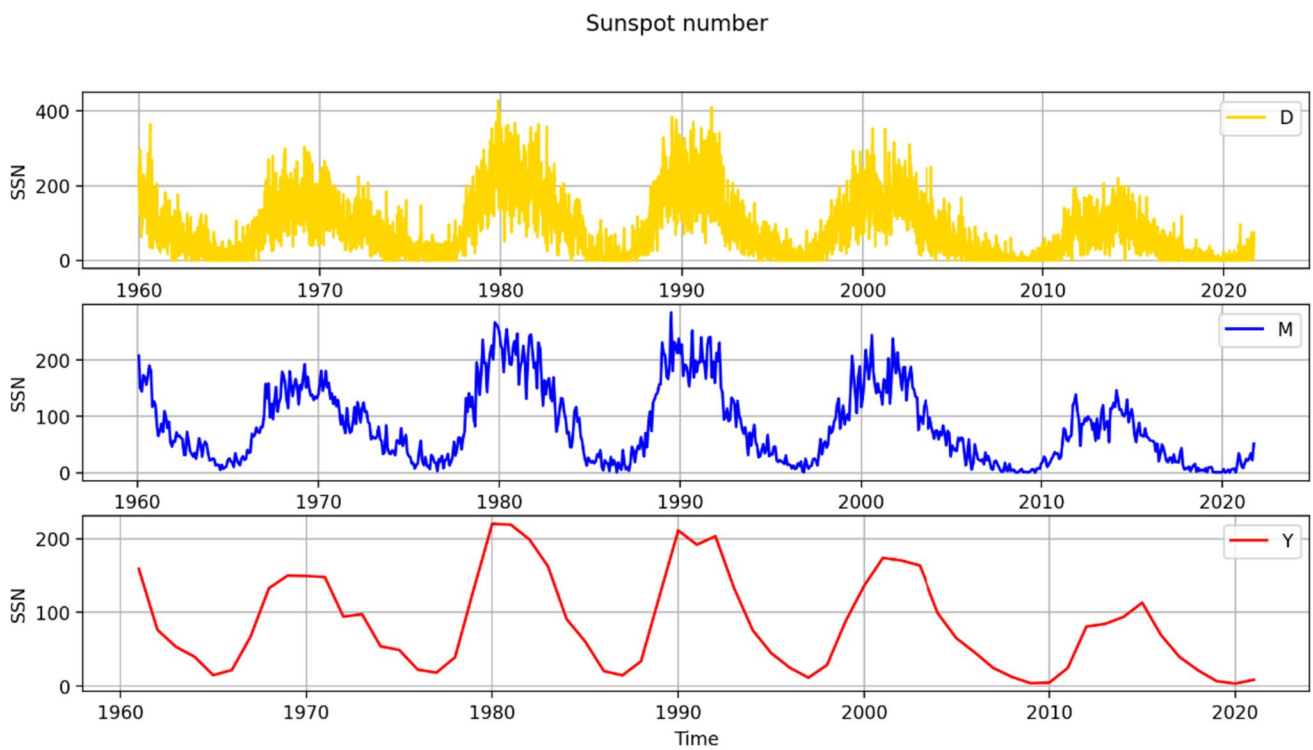


Fig. 4 Time series of the average sunspot numbers (SSN) at different time scales

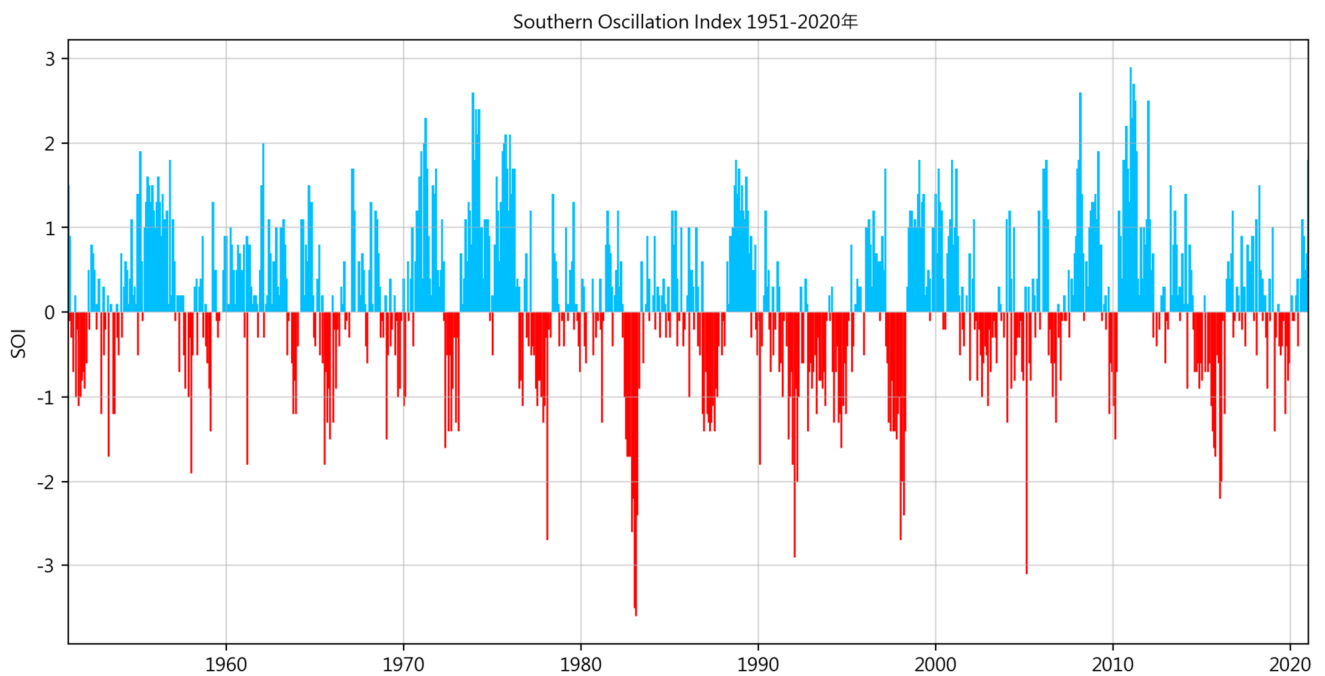


Fig. 5 Time series of the southern oscillation index (SOI) from 1951 to 2020

increasing amplitude of SOI fluctuations in recent years, reflecting increased uncertainty and variability in the global climate. In addition, a declining trend in the SOI since 1980

was observed, with values falling below -3 in 1983–1984 and in 2007–2008. Furthermore, significant fluctuations in the SOI was observed from 2010 to 2018, with strong

El Niño and La Niña events, which significantly impacted global climate and the environment.

Classification of rainfall levels

This study aimed to identify factors that may influence potential rainfall water resources and long-term rainfall trend changes using various research methods, and to achieve accurate predictions. However, the uneven spatial and temporal distribution of rainfall in Taiwan results in significant differences in the rainfall seasonality and amount in the northern and southern regions. Therefore, by calculating the anomaly values and appropriately classifying rainfall amounts in each region, the study not only eliminated seasonal effects for an improved identification of rainfall anomalies but also enhanced the accuracy of the classification

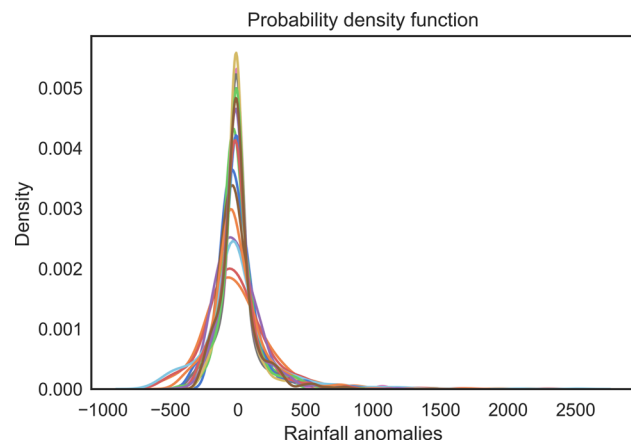


Fig. 6 Probability density function of the rainfall anomalies at each station

models. Additionally, models were employed to predict the monthly rainfall levels in Taiwan and provides specific rainfall warnings. This is expected to serve as a reference for future agricultural water use and water resource management policies in Taiwan.

Given the significant differences in the rainfall amounts across different regions in Taiwan, the probability distribution of rainfall in each location can be described using probability density functions. The area under the curve represents the probability of obtaining a certain value within a specific range of the random variable. Furthermore, the classification of the rainfall amounts using cumulative probabilities enables the quantification of rainfall frequency and intensity in each region, providing a more accurate description of the distribution of the frequency and intensity of a particular phenomenon, without limitation to a single region. Consequently, this enables comparisons among different regions. Therefore, in this study, the monthly average rainfall and humidity at each weather station were extracted from the historical data obtained from the Central Weather Bureau, and the anomalies of each weather station are calculated. Figure 6 shows the probability density function curve of the rainfall anomalies, and based on the curve and its corresponding cumulative probability, the rainfall anomalies were divided into five equal levels: very high, high, medium, low, and very low.

Wavelet signal analysis

CWT

Figure 7 shows the continuous wavelet spectrum graphs of the SSN and SOI from 1960 to 2020, where the vertical axis is the period, the horizontal axis is the time,

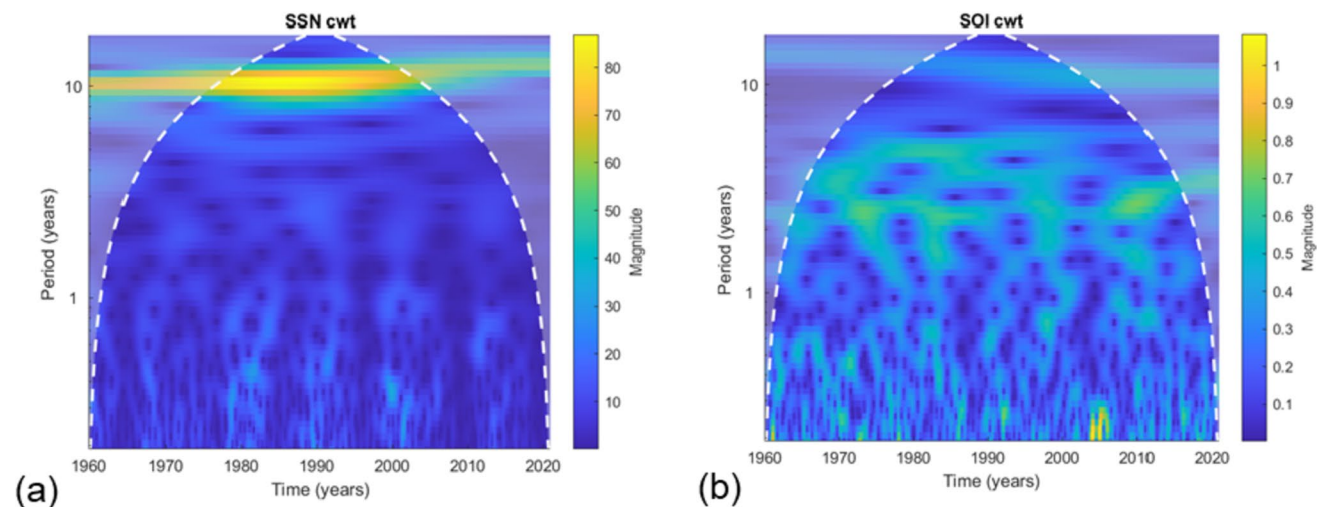


Fig. 7 Continuous wavelet spectrogram of the **a** SSN and **b** El Niño-SOI from 1960 to 2020

and the temporal scale is month. The yellow area in the figure indicates that the time period exhibits a relatively higher magnitude of variation degree (i.e., amplitude of the signal) in this time-period, and the blue area indicates a smaller magnitude of variation.

The continuous wavelet spectrogram indicated that sunspots have a significant cycle of 10–12 years, which is consistent with the actual observation data and the findings of previous studies (Nazari-Sharabian and Karakouzian 2020; Thomas and Abraham 2022), which reported that the number of sunspots significantly changes according to a long-term periodic cycle of approximately 11 years. Furthermore, SOI exhibited a relatively higher amplitude between a periodic cycle of 2–8 years, which is closely related to the cycle of the ENSO phenomenon (An and Wang 2000; Jin and Liu 2021). In addition, two notable yellow blocks were observed in the wavelet transform spectrum of the SOI between 2000 and 2010 in the high-frequency/low-period band, indicating the short-term occurrence of strong El Nino or La Nina events during this time period (Fig. 7b).

Figure 8 shows the continuous wavelet spectrum of the rainfall at Keelung, Taichung, and Kaohsiung stations from 1960 to 2020. A significant strong amplitude was observed in the annual periodic cycle in the wavelet transform spectrum of rainfall. As there is rainfall in the Keelung area all year round, the influence of seasonal rainfall in this region is not as significant as the influence in other regions (i.e., the wet and dry seasons are not distinct). In addition, the amplitude in the annual periodic cycle is smaller than those of the other two stations located in central and southern Taiwan.

WTC

The correlation between two time series in the time–frequency domain can be further understood using WTC analysis, as it enables the assessment of the correlation between the two at different time–frequency scales. Figures 9 and 10 show the WTC spectra of SSN vs rainfall and SOI vs rainfall at different demo stations, respectively.

In the WTC time–frequency spectrum, the horizontal axis is the time and the vertical axis is the frequency domain. The

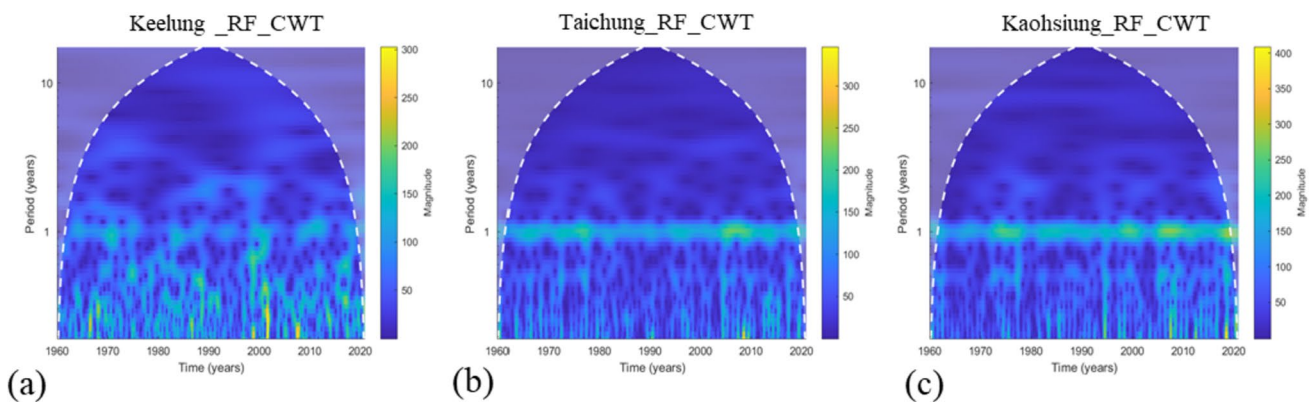


Fig. 8 Continuous wavelet spectrum map of rainfall from 1960 to 2020: **a** Keelung station, **b** Taichung station and **c** Kaohsiung station

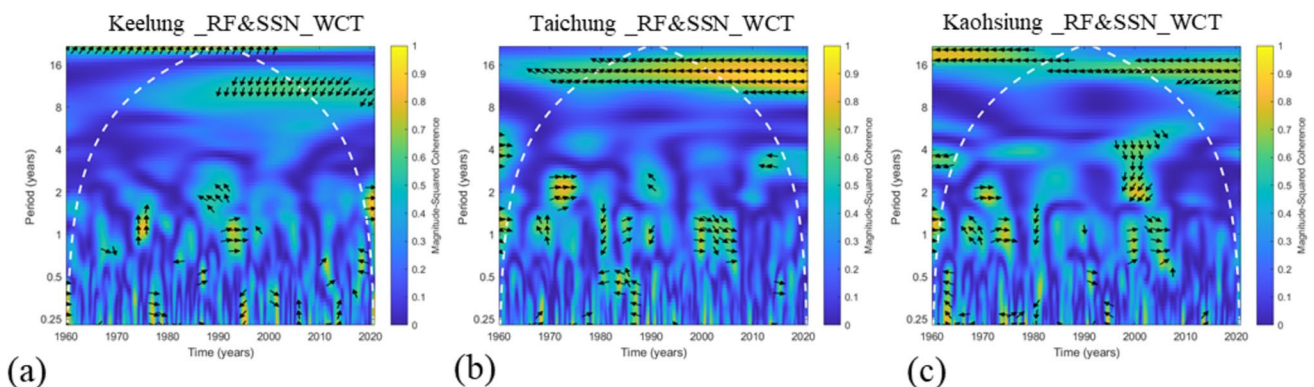


Fig. 9 Time–frequency diagram of the wavelet correlation between SSNs and rainfall at different stations: **a** Keelung station, **b** Taichung station and **c** Kaohsiung station

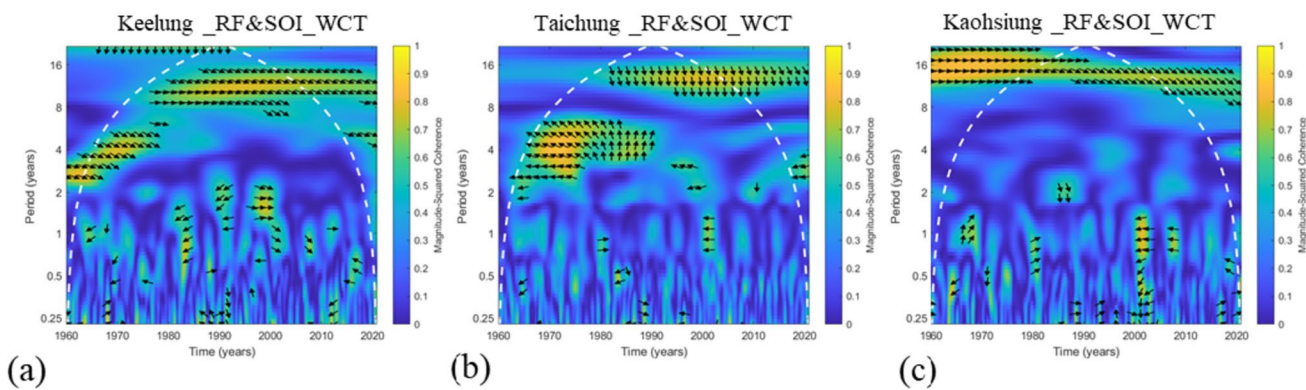


Fig. 10 Time–frequency diagram of the wavelet correlation between SOI and rainfall at different stations: **a** Keelung station, **b** Taichung station and **c** Kaohsiung station

closer the color of the spectrum is to yellow, the higher the correlation between the two, whereas the closer the color is to blue, the lower the correlation between the two. The arrow in the image is the phase angle, and the direction of the arrow indicates positive or negative correlation. The arrow pointing to the right is the forward phase angle, and that pointing to the left is the reverse phase angle. The phase angle can reveal the temporal lags between each other under the identified periods.

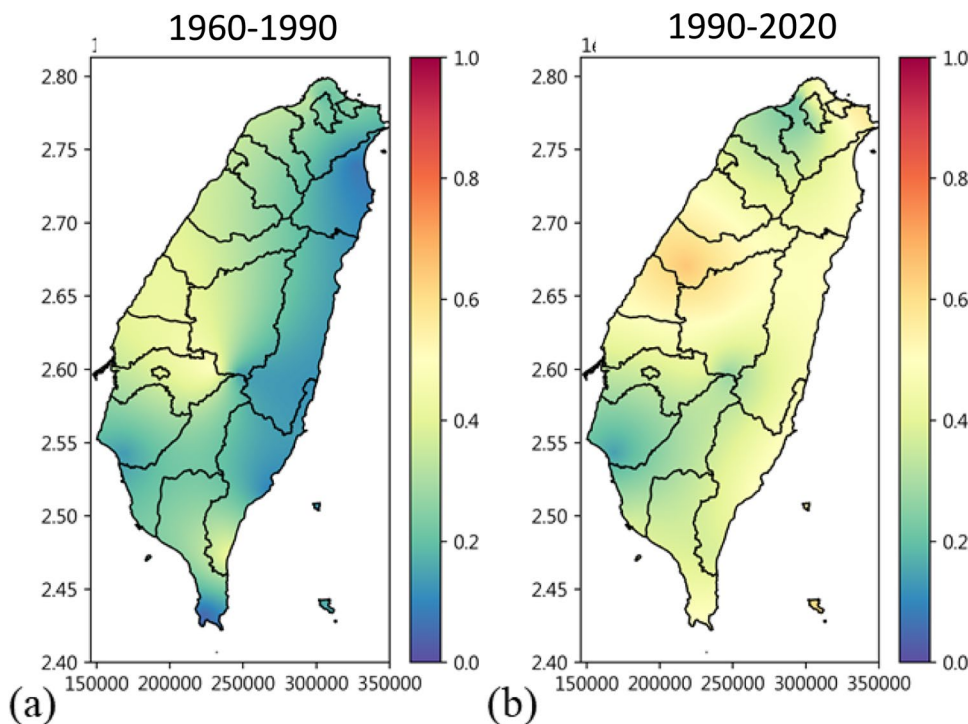
The frequency spectrum of the three measuring stations not only exhibited a high negative correlation in the frequency interval of 10–12 years but also exhibited a high correlation in the interval of 2–8 years. In addition, the degree of correlation has been increasing since 1990. The

time–frequency WTC spectrum revealed that there was a significant correlation between rainfall and SSN and SOI in the frequency intervals of 10–12 years and 2–8 years, and this correlation increased over time. These wavelet results can then be used and applied as important features for future rainfall prediction models.

Spatial distribution map of the mean WTC

To understand the changing trend and spatial distribution of the correlation between sunspots and rainfall, this study estimated the yearly mean WTC of the entire Taiwan in the frequency interval of 10–12 years using RBFN (Fig. 11). With 1990 as the boundary, the spatial estimation map

Fig. 11 Spatial estimation map of the mean WTC between SSN and rainfall in the frequency range of 10–12 years from **a** 1960–1990 and **b** 1990–2020



revealed that the correlation between sunspots and rainfall in Taiwan has increased significantly in the recent 30 years, and the areas where this phenomenon was most significant areas have moved northwards and eastwards from the original Chiayi and Yunlin areas to the Taichung-Changhua and Yilan-Hualien areas.

Rainfall prediction machine learning models

The analysis results indicated that both sunspots and ENSO effect are potential factors influencing the spatio-temporal rainfall and hydrological characteristics. Therefore, models were constructed based on the information on the sunspots and ENSO to predict the monthly rainfall level for all the CWB rainfall stations across Taiwan. These models can not only accurately predict the rainfall level of the current month but can also provide a rainfall warning signal for the government.

In this study, humidity, the inverse wavelet transform of SSN (ssn_icwt10to12), SOI, and the WTC analysis results of the 10–12 years frequency range (wcoh10_12) were used as the model input factors, and the rainfall anomalies of each station were divided into five levels (extremely high, high, medium, low, and very low). Eighty percent of the data was used as the training set, and 20% was used as the testing set. For different factors, the models performed Monte-Carlo cross-validation for prediction and classification of the rainfall levels, and optimized the parameter selection. The accuracy of the testing set of the four prediction models is shown in Table 2, for all CWB stations, and the evaluation index

Table 2 Accuracy of the testing set of the four prediction models

	Naïve Bayesian	Random forest	Decision tree	Bayesian network
Tamsui	0.833	0.841	0.818	0.877
Anbu	0.872	0.873	0.855	0.796
Taipei	0.811	0.805	0.773	0.825
Juzuhu	0.899	0.886	0.873	0.817
Keelung	0.776	0.764	0.705	0.798
Yilan	0.888	0.895	0.882	0.832
Tainan	0.758	0.505	0.409	0.878
Kaohsiung	0.765	0.527	0.473	0.906
Taichung	0.656	0.541	0.482	0.818
Alishan	0.647	0.555	0.505	0.899
Dawu	0.609	0.677	0.595	0.873
Yushan	0.809	0.768	0.736	0.922
Hengchun	0.738	0.477	0.432	0.908
Chenggung	0.885	0.850	0.832	0.844
Sunmoonlake	0.711	0.700	0.682	0.872
Taitung	0.662	0.677	0.573	0.842
Average	0.770	0.709	0.664	0.857

Table 3 Evaluation index of the classification prediction of the random forest model for Hengchun Station as example

	Precision	Recall	F1-score
Training set	0.881	0.873	0.873
Testing set	0.846	0.856	0.862

for the classification prediction of the random forest model is shown in Table 3. Among the four prediction models, the BN model exhibited the overall highest accuracy with a mean value of 0.857, and this was followed by the simple Bayesian classifier model (mean value of 0.770), random forest model (mean value of 0.709), and decision tree model (mean value of 0.664). The confusion matrix of the classification prediction of the random forest model for Hengchun station as example is shown in Fig. 12. Sunspots exhibited the highest feature importance, which was followed by the wavelet correlation coefficient between sunspots and rainfall (Fig. 13). Due to the unique confusion matrix and feature importance results for each station, we are limited by the article’s length and cannot present all the results. Therefore, we have chosen Taichung station as an example to showcase in Fig. 13. Generally, the feature importance results for various stations are quite similar, with sunspots having the highest feature importance in the 10–12 year wavelet feature band. This result confirmed the strong correlation between sunspots and long-term rainfall properties. In addition, improved prediction results were observed in relatively

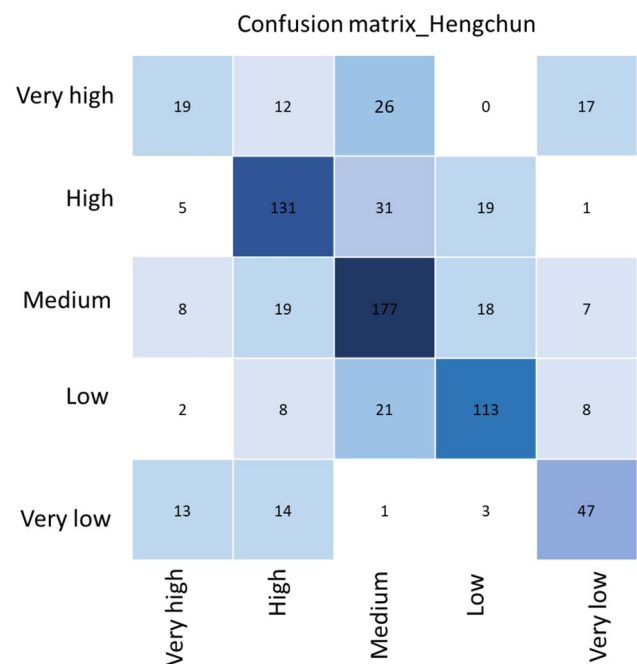


Fig. 12 Confusion matrix of the classification prediction of the random forest model for Hengchun Station as example

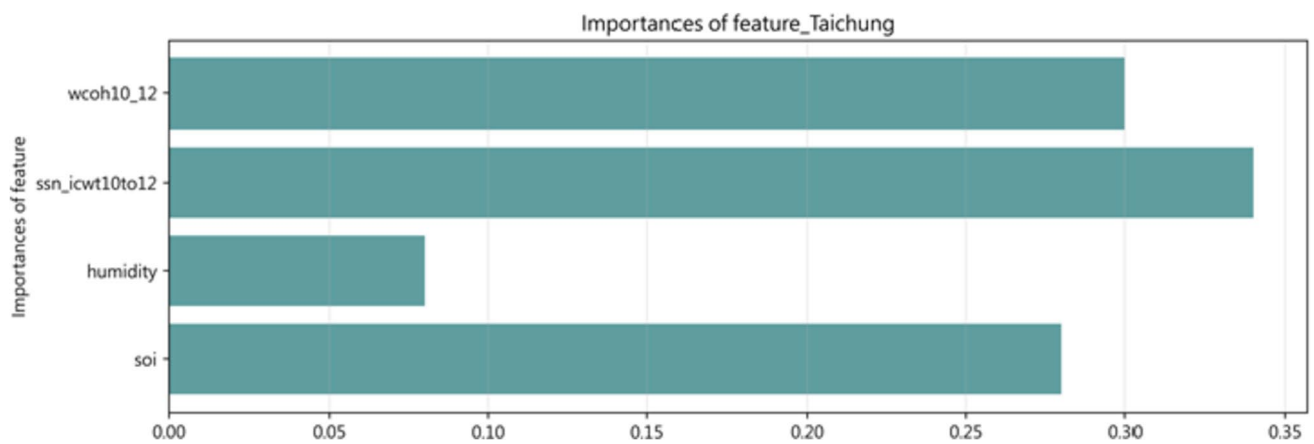


Fig. 13 Feature importance for the Taichung station as example

rainy areas (Table 2), and most of these areas are areas with higher sunspot and rainfall WTC in Fig. 11. This may be attributed to the fact that the rainfall in other regions may be affected by extreme rainfall factors, such as typhoons, and the factors of rainfall caused by typhoons cannot be directly presented in this model and linked with sunspot and ENSO effect.

Discussion

In the past, traditional research on the relationship between sunspots and rainfall was mostly limited to determining correlations. Only in recent years have studies begun to employ time-frequency analysis methods, such as wavelet analysis (Bhattacharyya and Narasimha 2005; Nazari-Sharabian and Karakouzian 2020; Thomas and Abraham 2022), to investigate their relationship in the time and frequency domains. However, most of these studies also stopped at correlation analysis. This study stands out as one of the few to combine innovative approaches by integrating wavelet time-frequency analysis and machine learning models while considering the sunspot and ENSO effect to establish a comprehensive water resource forecasting model. It has demonstrated fairly accurate predictions in different regions, which is highly beneficial for long-term water resource management. This is particularly valuable for Taiwan, where rainfall distribution is highly uneven due to topographical factors and is susceptible to the impacts of climate change, resulting in droughts and floods.

Furthermore, this study analyzes and models data over a long time scale (60 years) and thoroughly explores the spatiotemporal distribution changes of SSN to rainfall in the region, rather than only one station. It includes the distribution of monitoring stations across various spatial locations in Taiwan and their changes over time. Additionally, this

study differs from traditional black-box machine learning models. Instead, it utilizes Bayesian Networks and employs the Feature Importance method to analyze and confirm that sunspots make the most significant contribution to rainfall prediction in specific wavelet feature bands.

Owing to the effects of climate change and the primary rainfall periods of Taiwan (which occurs during the Meiyu season from May to June and the typhoon season from July to September), notable variations are observed in rainfall and its spatial and temporal distribution in the wet and dry seasons (Chen and Chen 2003; Lee et al. 2020). If the precipitation received during the rainy season and typhoon season of the previous year is insufficient, it can result in drought conditions and potentially lead to a severe water shortage crisis (Lin et al. 2021). From a macroscopic time scale perspective, climate change or slight changes in the overall rainfall in the annual rainfall season may directly or indirectly affect Taiwan's drought or flood disasters. In addition, from a macroscopic spatial scale perspective, the El Niño phenomenon and solar activities may have a certain impact on the overall climate and water resources of the earth (García-García and Ummenhofer 2015; Thomas and Abraham 2022; Vladimiro and Guido 2018). In the past, several studies have explored the spatio-temporal characteristics of water resources at the regional scale, such as the variation of catchment areas or alluvial fans at shorter time scales within a few years (Chen et al. 2013; Tsai et al. 2014); however, only few studies have explored the characteristics of water resource variability at the macroscopic spatio-temporal scales.

Compared to the spatial and temporal characteristics of water resources at the regional scale investigated in previous studies, this study considered a macroscopic spatio-temporal scale using actual observation data and combining the data at macroscopic time and planetary scales. In this study, the variation characteristics of factors in different time-frequency domains were extracted using the wavelet

signal analysis method, after which the method was utilized to explore the relationship between rainfall in Taiwan and the factors that may affect the characteristics of hydrological and water resource changes. The time–frequency WTC spectrum revealed a high degree of negative correlation between rainfall and sunspots in the frequency range of 10–12 years, which is similar to the conclusion of previous studies (Nazari-Sharabian and Karakouzian 2020; Thomas and Abraham 2022). In contrast, there was a higher correlation between rainfall and ENSO in the frequency range of 2–8 years, which is also consistent with the conclusion of previous studies (An and Wang 2000; Jin and Liu 2021). In addition, the results revealed that the correlation has been increasing since 1990. Particularly, the correlation between sunspots and rainfall in Taiwan has increased significantly in the last 30 years, and the area's most significantly affected by this have moved northwards and eastwards from the original Chiayi and Yunlin areas to the Taichung-Changhua and Yilan-Hualien areas. This suggests a greater relationship between sunspots and local regional rainfall characteristics in recent years, and also in more areas, which are yet to be discovered.

Among the four prediction models, the BN model exhibited the highest accuracy (mean value of 0.857), which was followed by the simple Bayesian classifier (mean value of 0.770), random forest (mean value of 0.709), and decision tree (mean value of 0.664) models. In addition, sunspots exhibited the highest feature importance, which was followed by the wavelet correlation coefficient between sunspots and rainfall. These results indicated that the constructed models can not only accurately predict the rainfall level of the month and give a warning signal for rainfall but can also serve as a reference for the future agricultural water use and water resource management guidelines in Taiwan.

Conclusion

This study employed a novel data-driven approach to investigate the time–frequency relationships between sunspot and long-term local rainfall amount, and constructed machine learning prediction models to confirm the effect of solar activity on long-term local rainfall patterns. The results demonstrated that improved prediction results were achieved in relatively rainy areas. This could be attributed to the possible influence of extreme weather events, such as typhoons, on rainfall in other areas, and the inability to directly incorporate specific rainfall patterns associated with typhoons into this model to establish a connection with sunspot activity and the ENSO effect. In addition, the results revealed that the relationship between sunspots and local rainfall in Taiwan is increasing yearly, regardless of space

or time; particularly, this was more notable in 1990, which was set as the boundary.

These results indicated that rainfall behavior (except extreme rainfall caused by typhoon) can be described using sunspots and ENSO effect, and this will be beneficial for providing water resource management. Although the impact of direct sunspots on rainfall may not be significant, the wavelet extraction indicated that it is one of the most influential features, and its impact exceeded that of humidity, which is typically believed to exert the greatest impact on rainfall. Therefore, we recommend that indicators of planetary-scale solar activity, such as sunspots, should be incorporated in future long-term water resource management or predictions, as it has been confirmed in this study as an important factor influencing regional rainfall on larger timescales, and is even more significant than the El Niño phenomenon and humidity.

Acknowledgements We are grateful for the support from the National Science and Technology Council (Ministry of Science and Technology), Taiwan, Grant Nos. 109-2636-E-008-008, 110-2636-E-008-006, and 111-2636-E-008-014 (Young Scholar Fellowship Program), and MOST 108-2638-E-008-001-MY2 (Shackleton Program Grant). We are also thankful for the Python programming language and its modules, which were used as a powerful tool in our data analysis.

Funding The funding was provided by National Science and Technology Council (Ministry of Science and Technology), Taiwan, Grant Nos. 109-2636-E-008-008, 110-2636-E-008-006, and 111-2636-E-008-014 (Young Scholar Fellowship Program), and MOST 108-2638-E-008-001-MY2 (Shackleton Program Grant)

Data availability The data supporting the findings of this study are available upon reasonable request from the corresponding author.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval Not applicable.

Informed consent The authors confirm independence from the sponsors; the content of the article has not been influenced by the sponsors. The authors declare their consent to participate and publish in this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adamowski J, Chan HF (2011) A wavelet neural network conjunction model for groundwater level forecasting. *J Hydrol* 407(1):28–40. <https://doi.org/10.1016/j.jhydrol.2011.06.013>
- Allan C, Xia J, Pahl-Wostl C (2013) Climate change and water security: challenges for adaptive water management. *Curr Opin Environ Sustain* 5(6):625–632. <https://doi.org/10.1016/j.cosust.2013.09.004>
- An S-I, Wang B (2000) Interdecadal change of the structure of the ENSO mode and its impact on the ENSO frequency. *J Clim* 13(12):2044–2055. [https://doi.org/10.1175/1520-0442\(2000\)013%3c2044:ICOTSO%3e2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013%3c2044:ICOTSO%3e2.0.CO;2)
- Ananthakrishnan R, Parthasarathy B (1984) Indian rainfall in relation to the sunspot cycle: 1871–1978. *J Climatol* 4(2):149–169. <https://doi.org/10.1002/joc.3370040205>
- Barthel R, Banzhaf S (2016) Groundwater and surface water interaction at the regional-scale—a review with focus on regional integrated models. *Water Resour Manag* 30(1):1–32
- Belgiu M, Drăguț L (2016) Random forest in remote sensing: a review of applications and future directions. *ISPRS J Photogramm Remote Sens* 114:24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Ben G (2007) Bayesian networks. *Encyclopedia of statistics in quality and reliability*. Wiley, New York
- Bhattacharyya S, Narasimha R (2005) Possible association between Indian monsoon rainfall and solar activity. *Geophys Res Lett*. <https://doi.org/10.1029/2004GL021044>
- Bogges A, Narcowich FJ (2015) A first course in wavelets with Fourier analysis. Wiley, Hoboken
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32. <https://doi.org/10.1023/A:1010933404324>
- Browne MW (2000) Cross-validation methods. *J Math Psychol* 44(1):108–132
- Chen C-S, Chen Y-L (2003) The rainfall characteristics of Taiwan. *Mon Weather Rev* 131(7):1323–1341. [https://doi.org/10.1175/1520-0493\(2003\)131%3c1323:TRCOT%3e2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131%3c1323:TRCOT%3e2.0.CO;2)
- Chen S, Cowan CF, Grant PM (1991) Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Trans Neural Netw* 2(2):302–309
- Chen S-T, Kuo C-C, Yu P-S (2009) Historical trends and variability of meteorological droughts in Taiwan / Tendances historiques et variabilité des sécheresses météorologiques à Taiwan. *Hydrol Sci J* 54(3):430–441. <https://doi.org/10.1623/hysj.54.3.430>
- Chen Y-C, Chang K-T, Chiu Y-J, Lau S-M, Lee H-Y (2013) Quantifying rainfall controls on catchment-scale landslide erosion in Taiwan. *Earth Surf Proc Land* 38(4):372–382. <https://doi.org/10.1002/esp.3284>
- Chen C, Li Y, Zhao N, Guo B, Mou N (2018) Least squares compactly supported radial basis function for digital terrain model interpolation from airborne lidar point clouds. *Remote Sens* 10(4):587
- Chen J, Li T, Nan Q, Shi X, Liu Y, Jiang B, Zou J, Selvaraj K, Li D, Li C (2019) Mid-late Holocene rainfall variation in Taiwan: a high-resolution multi-proxy record unravels the dual influence of the asian monsoon and ENSO. *Palaeogeogr Palaeoclimatol Palaeoecol* 516:139–151. <https://doi.org/10.1016/j.palaeo.2018.11.026>
- Chu P-S (2004) ENSO and tropical cyclone activity. *Hurric Typhoons Past Present Potential* 297:332
- Cowling TG (1933) The magnetic field of sunspots. *Mon Not R Astron Soc* 94:39–48
- Cristiano E, ten Veldhuis MC, van de Giesen N (2017) Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas – a review. *Hydrol Earth Syst Sci* 21(7):3859–3878. <https://doi.org/10.5194/hess-21-3859-2017>
- De Silva M, Kawasaki A (2018) Socioeconomic vulnerability to disaster risk: a case study of flood and drought impact in a rural Sri Lankan community. *Ecol Econ* 152:131–140
- Ding Q, Wang Y, Zhuang D (2018) Comparison of the common spatial interpolation methods used to analyze potentially toxic elements surrounding mining regions. *J Environ Manag* 212:23–31. <https://doi.org/10.1016/j.jenvman.2018.01.074>
- Domingos P, Pazzani M (1997) On the optimality of the simple Bayesian classifier under zero-one loss. *Mach Learn* 29(2):103–130
- García-García D, Ummenhofer CC (2015) Multidecadal variability of the continental precipitation annual amplitude driven by AMO and ENSO. *Geophys Res Lett* 42(2):526–535
- Grinsted A, Moore JC, Jevrejeva S (2004) Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Process Geophys* 11(5/6):561–566
- Indeje M, Semazzi FHM, Ogallo LJ (2000) ENSO signals in East African rainfall seasons. *Int J Climatol* 20(1):19–46. [https://doi.org/10.1002/\(SICI\)1097-0088\(200001\)20:1%3c19::AID-JOC449%3e3.0.CO;2-0](https://doi.org/10.1002/(SICI)1097-0088(200001)20:1%3c19::AID-JOC449%3e3.0.CO;2-0)
- Jiang Z, Tai-Jen Chen G, Wu M-C (2003) Large-scale circulation patterns associated with heavy spring rain events over Taiwan in strong ENSO and non-ENSO years. *Mon Weather Rev* 131(8):1769–1782. <https://doi.org/10.1175/2561.1>
- Jin Y, Liu Z (2021) A theory of the spring persistence barrier on ENSO. Part I: the role of ENSO period. *J Clim* 34(6):2145–2155. <https://doi.org/10.1175/JCLI-D-20-0540.1>
- Kao S-C, Kume T, Komatsu H, Liang W-L (2013) Spatial and temporal variations in rainfall characteristics in mountainous and lowland areas in Taiwan. *Hydrol Process* 27(18):2651–2658. <https://doi.org/10.1002/hyp.9416>
- Kuo C-C, Gan TY, Yu P-S (2010) Wavelet analysis on the variability, teleconnectivity, and predictability of the seasonal rainfall of Taiwan. *Mon Weather Rev* 138(1):162–175. <https://doi.org/10.1175/2009MWR2718.1>
- Lee JH, Yang C-Y, Julien PY (2020) Taiwanese rainfall variability associated with large-scale climate phenomena. *Adv Water Resour* 135:103462. <https://doi.org/10.1016/j.advwatres.2019.103462>
- Lin Y-C, Kuo E-D, Chi W-J (2021) Analysis of meteorological drought resilience and risk assessment of groundwater using signal analysis method. *Water Resour Manage* 35(1):179–197. <https://doi.org/10.1007/s11269-020-02718-x>
- Lin Y, Wang J, Zou R (2015) An improved naïve bayes classifier method in public opinion analysis. In Wong WE (Ed.) *Proceedings of the 4th international conference on computer engineering and networks: CENet2014*, Springer International Publishing, Cham, pp 219–225
- Marsh N, Svensmark H (2003) Solar Influence on Earth's Climate. In: Chian ACL, Cairns IH, Gabriel SB, Goedbloed JP, Hada T, Leubner M, Nocera L, Stening R, Toffoletto F, Uberoi C, Valdivia JA, Villante U, Wu CC, Yan Y (eds) *Advances in space environment research*, vol I. Springer, Netherlands, Dordrecht, pp 317–325
- Misra AK (2014) Climate change and challenges of water and food security. *Int J Sustain Built Environ* 3(1):153–165. <https://doi.org/10.1016/j.ijsbe.2014.04.006>
- Myles AJ, Feudale RN, Liu Y, Woody NA, Brown SD (2004) An introduction to decision tree modeling. *J Chemom* 18(6):275–285. <https://doi.org/10.1002/cem.873>
- Narvaez L, Janzen S, Eberle C, Sebesvari Z (2022) Technical Report: Taiwan drought. In: *Interconnected Disaster Risks 2021/2022*. United Nations University - Institute for Environment and Human Security (UNU-EHS)
- Nazari-Sharabian M, Karakouzian M (2020) Relationship between sunspot numbers and mean annual precipitation: application of cross-wavelet transform—a case study. *Journal* 3(1):67–78
- Orlanski I (1975) A rational subdivision of scales for atmospheric processes. *Bull Am Meteor Soc* 56(5):527–530

- Orr MJ (1996) Introduction to radial basis function networks. In: Technical Report, center for cognitive science, University of Edinburgh
- Rish I (2001) An empirical study of the naive Bayes classifier. In: Paper presented at the IJCAI 2001 workshop on empirical methods in artificial intelligence
- Ritzema HP, Van Loon-Steensma JM (2018) Coping with climate change in a densely populated delta: a paradigm shift in flood and water management in The Netherlands. *Irrig Drain* 67(S1):52–65. <https://doi.org/10.1002/ird.2128>
- Seleshi Y, Demarée GR, Delleur JW (1994) Sunspot numbers as a possible indicator of annual rainfall at Addis Ababa. *Ethiop Int J Climatol* 14(8):911–923. <https://doi.org/10.1002/joc.3370140807>
- Shiau JT, Hsiao YY (2012) Water-deficit-based drought risk assessments in Taiwan. *Nat Hazards* 64(1):237–257
- Thomas E, Abraham NP (2022) Relationship between sunspot number and seasonal rainfall over Kerala using wavelet analysis. *J Atmos Solar Terr Phys* 240:105943. <https://doi.org/10.1016/j.jastp.2022.105943>
- Toit WD (2008) Radial basis function interpolation
- Torrence C, Compo GP (1998) A practical guide to wavelet analysis. *Bull Am Meteor Soc* 79(1):61–78. [https://doi.org/10.1175/1520-0477\(1998\)079%3c0061:apgtwa%3e2.0.co;2](https://doi.org/10.1175/1520-0477(1998)079%3c0061:apgtwa%3e2.0.co;2)
- Tsai M-J, Abraham RJ, Mount NJ, Chang F-J (2014) Including spatial distribution in a data-driven rainfall-runoff model to improve reservoir inflow forecasting in Taiwan. *Hydrol Process* 28(3):1055–1070. <https://doi.org/10.1002/hyp.9559>
- Varis O, Vakkilainen P (2001) China's 8 challenges to water resources management in the first quarter of the twenty first Century. *Geomorphology* 41(2):93–104. [https://doi.org/10.1016/S0169-555X\(01\)00107-6](https://doi.org/10.1016/S0169-555X(01)00107-6)
- Vladimiro T, Guido W (2018) Seasonal rainfall patterns classification, relationship to ENSO and rainfall trends in Ecuador. *Int J Climatol* 38(4):1808–1819. <https://doi.org/10.1002/joc.5297>
- Wang L, Yang Z, Gu X, Li J (2020) Linkages between tropical cyclones and extreme precipitation over China and the role of ENSO. *Int J Disaster Risk Sci* 11(4):538–553. <https://doi.org/10.1007/s13753-020-00285-8>
- Webb GI, Keogh E, Mikkilainen R (2010) Naïve Bayes. *Encycl Mach Learn* 15:713–714
- Yeh S-W, Kug J-S, Dewitte B, Kwon M-H, Kirtman BP, Jin F-F (2009) El Niño in a changing climate. *Nature* 461(7263):511–514
- Yim S-Y, Wang B, Xing W, Lu M-M (2015) Prediction of Meiyu rainfall in Taiwan by multi-lead physical-empirical models. *Clim Dyn* 44(11):3033–3042. <https://doi.org/10.1007/s00382-014-2340-0>
- Yu H-L, Lin Y-C (2015) Analysis of space-time non-stationary patterns of rainfall-groundwater interactions by integrating empirical orthogonal function and cross wavelet transform methods. *J Hydrol* 525:585–597. <https://doi.org/10.1016/j.jhydrol.2015.03.057>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.