The Reconstruction and Extension of Terrestrial Water Storages Based On A Combined Prediction Model

Erhao Meng
Xi'an University of Technology
Shengzhi Huang (huangshengzhi7788@126.com)
Xi'an University of Technology https://orcid.org/0000-0001-7592-5268
Qiang Huang
Xi'an University of Technology
Linyin Cheng
University of Arkansas Fayetteville
Wei Fang
Xi'an University of Technology

Research Article

Keywords: GRACE, Total water storage anomalies, Support vector machine, Combined prediction model, Input selection strategy

DOI: https://doi.org/10.21203/rs.3.rs-531840/v1

The reconstruction and extension of terrestrial water storages based on a combined prediction model

Erhao Meng[a,b], Shengzhi Huang[a]*[1], Qiang Huang[a], Linyin Cheng[b] and Wei Fang[a]

[a] State Key Laboratory Base of Eco-Hydraulic Engineering in Arid Area, Xi'an University of Technology, Xi'an 710048, China

[b] Geosciences Department, University of Arkansas, 216C Gearhart Hall, 72701, USA

*Corresponding author at: State Key Laboratory Base of Eco-Hydraulic Engineering in Arid Area, Xi'an University of Technology, Xi'an 710048, China. Tel.: +86 29 82312801; fax: +86 29 82312797. E-mail Address: huangshengzhi7788@126.com.

Abstract The monthly changes in total water storage (△TWS) can be employed for drought and flood monitoring and early warning and can be obtained from the total water storage anomalies (TWSA) of the Gravity Recovery and Climate Experiment (GRACE). However, the relatively short GRACE time series limits its further wide application. To this end, a combined prediction (CP) model including Support Vector Machine (SVM) and Artificial Neural Network (ANN) was proposed in this study for the reconstruction and extension of monthly TWSA from 1960 to 2012. Moreover, an innovative input selection strategy is proposed to build a monthly TWSA prediction model, in which the partial correlation algorithm is used to select the best input variables from candidate input variables. These candidate input variables include streamflow, precipitation, evaporation, and soil moisture storage (SMS). The Yunnan province, a typical humid area in China, was selected as a case study. The results show that: (1) The innovative input selection strategy effectively improves the simulation ability of the model, especially when the candidate input variables influence each other; (2) The performance of the CP model using the innovative input selection strategy is best; (3) The monthly △TWS obtained from the extension of TWSA recorded five of the seven extreme meteorological drought events in Yunnan Province from 1961 to 2001, therefore, the reliability of the expanded TWSA is better than GLDAS TWSA. Generally, the findings of this study showed that the CP model using an innovative input selection strategy is a useful and powerful tool for monthly TWSA prediction.

Keywords: GRACE; Total water storage anomalies; Support vector machine; Combined prediction model; Input selection strategy

1. Introduction

The Gravity Recovery and Climate Experiment (GRACE) satellite, designed and developed by the National Aeronautics and Space Administration (NASA) and the German Aerospace Center (DLR), was launched on March 17, 2002. The monthly changes in the Earth's gravity field observed by the GRACE mission are related to changes in the total water storage (△TWS), including surface water and groundwater, soil moisture, and snow/ice (Wahr et al., 1998). Many studies have verified it using situ data, model data, and remotely sensed data (Doell et al., 2014; Long et al., 2015). In recent years, GRACE △TWS data have been used widely in many fields of science. Now, GRACE is a strong tool for scientists in a few science fields (Wouters et al., 2014), and it has been employed to estimate individual or multiple water storages for different applications by several studies (Andrew et al., 2017; Eicker et al., 2016 and Reager et al., 2015).

GRACE △TWS has proven to be an important tool for drought and flood monitoring in the hydrological field (Long et al., 2014; Zhang et al., 2015). The GRACE △TWS at different time intervals can be obtained from the GRACE total water storage anomalies (TWSA). Although GRACE △TWS has been successfully used in hydrological science, it still has limitations in its further application in drought and flood monitoring. GRACE's monthly TWSA data cover the period from 2002 to 2016, and there are 13 months of data loss (Famiglietti et al., 2013). In the past ten years, some studies have been conducted using GRACE△TWS for drought and flood monitoring (Zhang et al., 2015). However, TWSA data before the GRACE launch is

65 also necessary to understand the long-term trends of △TWS. In addition, there may
66 not be reliable TWSA data between the period of the decaying orbit of current GRACE
67 satellites and the launch of the GRACE Follow-On Mission in 2018. Thus, there is an
68 urgent need to develop a model for expand GRACE data with high accuracy.
69 Pan integrated several hydrological variable date records of the water cycle and used
70 data assimilation technology to establish a water cycle data record set for 32 major river
71 basins around the world from 1984 to 2006 (Pan et al., 2012). In this data set, TWS
72 outside of the GRACE period is provided by the Variable Infiltration Capacity (VIC)
73 model, which does not include simulations of changes in groundwater reserves, so the
74 data may be missing the interaction between surface water and groundwater systems
75 information (Liang et al., 1994). de Linage used a simple statistical model to predict
76 TWSA data over Amazon by examining the anomalous changes in sea surface
77 temperature (SST) from the equatorial central Pacific (Niño 4) and tropical North
78 Atlantic (TNAI) (de Linage et al., 2013). Finally, the R^2 between TWSA and a
79 combination of Niño 4 and TNAI was 0.43 in the overall Amazon basin. Please note
80 that due to the location and number of observing stations, it is not always possible to
81 obtain complete water level data. Moreover, surface models (for example, VIC) usually
82 ignore the interrelationship between surface water and groundwater, which may make
83 the total water storage obtained by the model lose the accuracy. Therefore, data-driven
84 models (such as ANN, SVM) that can obtain satisfactory results using available data
85 can be used to extend TWSA. Long used monthly precipitation, monthly average
86 temperature, and soil moisture storage (SMS) to construct an ANN model, and made a

long-term forecast of TWSA of the Yunnan-Guizhou Plateau from 1979 to 2002 (Long et al., 2014). However, it is well known that a large amount of data is required to build a reliable and powerful ANN model, but the available monthly observation value of TWSA is 159 months. Therefore, more research should be conducted to establish a robust model for TWSA prediction.

Support vector machines (SVM) are particularly suitable for modeling small sample data sequences (Cortes et al., 1995). It is based on the Vapnik-Chervonenkis (VC) dimensional theory and the principle of structural risk minimization. In theory, it can overcome local convergence and achieve a global optimal solution. The SVM has been widely used in streamflow prediction (Huang et al., 2014; Meng et al., 2019 and Meng et al., 2021). However, due to the short length of the available monthly TWSA, using a single model for TWSA expansion will lead to uncertainty and instability. ANN has been employed for the TWSA extension in the karst plateau of Southwest China (Long et al., 2014). At the same time, SVM has a strong ability for small sample data prediction. Therefore, in this paper, a combined prediction (CP) model (including SVM and ANN) is constructed for monthly TWSA extension. Meanwhile, hydrology variables (including streamflow, precipitation, evaporation, and SMS) were employed as potential explanatory variables of monthly TWSA. However, there is a correlation between these potential explanatory variables. In this paper, a flexible input selection strategy that can eliminate the interdependence between potential explanatory variables is established to find the best combination of explanatory variables for monthly TWSA expansion.

Yunnan province, the study area, is one of the wettest regions in China, with a forest coverage rate as high as 50% (Han et al., 2019). Most of GRACE's research on TWSA focuses on the changes in groundwater reserves in the southwestern region and the karst plateau, as well as the monitoring of droughts and floods, but rarely pays attention to Yunnan Province (Long et al., 2014; Huang et al., 2019 and Long et al., 2015). Especially at the end of 2009, Yunnan province experienced a once-in-a-hundred-year drought, which caused huge losses to local production and life. Therefore, it is of great significance to predict TWSA with higher prediction accuracy.

The objectives of this study are to (1) check the performance of the flexible input selection strategy for monthly TWSA reconstruction extension, (2) built an optimal combination of explanatory variables for monthly TWSA reconstruction and extension, (3) explore the performance of the ANN, SVM, and the CP model for monthly TWSA extension in Yunnan province using the optimal combination of explanatory variables. The extension of long-term TWSA data is of great valuable for understanding the impact of climate change on the hydrological cycle and provides inspiration for water resource management in Yunnan province.

2. Materials and Methods

2.1 Study region

Yunnan Province (97.31°-106.11°E, 21.8°-29.15°N) covers an area of 394,100 km2 (Fig. 1). The climate of Yunnan belongs to the subtropical plateau monsoon type, with clear distinctions between wet and dry seasons, and the temperature varies significantly with the terrain. Yunnan Province has an average rainfall of 1258 mm for many years

and has abundant water resources. There are as many as 908 rivers with a runoff area of more than 100 square kilometers (Jiang et al., 2018).

2.2 GRACE data, Climate data, Streamflow data, Soil moisture data

In this study, the research data GRACE was got from The University of Texas Centre for Space Research (CSR) at http://isdc.gfz-potsdam.de/grace-isdc/. The monthly GRACE data lost 13 months of values, and these missing values were filled in with the month values on both sides of the missing data through mean interpolation.

There are 27 meteorological stations in the province. The monthly precipitation and potential evaporation data (from January 1960 to December 2010) of these stations can be downloaded from the China National Meteorological Information Center (http://data.cma.cn/).

In addition, monthly streamflow data (from January 1960 to December 2010) was collected from the hydrologic manual. Strictly control data quality during release.

In this study, rasterized soil moisture data from 1960 to 2010 was obtained from the VIC model product. This production is obtained by the VIC model using the meteorological factors of the China Meteorological Administration (CMA). The monthly soil moisture data is obtained by averaging the rasterized soil moisture data in each hydrological station.

2.3 Methods

2.3.1 Artificial neural network (ANN)

ANN is an intelligent algorithm designed by imitating the working mechanism of the human brain. It consists of three parts: input layer, hidden layers, and output layer. ANN

also needs weights and activation functions to simulate biological neural networks. The activation functions commonly used in the hydrology field are tangent, logistic, and linear. Detailed theoretical information about ANN can be found in (Zhang et al., 2014 and Fang et al., 2018). In this paper, the ANN is developed with three layers. The sigmoid function was employed as the transfer function, and the optimal number of neurons in the hidden layer is 5 after several trying.

2.3.2 Support vector machine (SVM)

Cortes and Vapnik (1995) proposed the support vector machine (SVM) which is an efficient machine learning tool for time series forecasting. SVM uses a kernel function to map the input factors of a complex nonlinear problem to a high-dimensional space to simplify the complex problem, thereby transforming the complex nonlinear problem into a linear problem. SVM is built based on statistical learning theory, which can help SVM to obtain the global optimal solution. Some many papers and books have described the theory of SVM in detail (Yoon et al., 2016; Carrier et al., 2013; Ch et al., 2013 and He et al., 2014), the description of SVM is omitted. The kernel function is the key for SVM to solve complex prediction problems, so, it's important to choose a suitable kernel function. The kernel function includes linear, sigmoid, Radial Basis Function (RBF), and so on. In this paper, RBF is employed as the kernel function:

$$K\left(\mathrm{x}, \mathrm{x}_j\right) = \exp\left(-\| x - x_j \|^2 / 2\sigma^2\right) \quad (1)$$

where σ represents the Gaussian noise level of standard deviation.

2.3.3 Combined Prediction model

Meanwhile, the combined prediction (CP) model was employed to predict monthly

TWSA. The combined prediction can be expressed as follows:

$$Y = w_1 * X_1 + w_2 * X_2 \quad (2)$$

where Y is the prediction value of TWSA using combined prediction, X_1 is the prediction values of SVM, X_2 is the prediction value of ANN, w_1 is the weight of the prediction value of SVM, w_2 is the weight of the prediction value of ANN. In this paper, the values of w_1 and w_2 are obtained by the Genetic algorithm (GA). The detail of the combined prediction model is shown in Fig. 2.

2.3.4 Innovative input variable selection strategy

To select the optimal input variable for monthly SWA prediction, an innovative input selection strategy is proposed. The detail of this strategy is demonstrated as follows:

1. The 13 lags of streamflow $\{S_t, S_{t-1},\ldots,S_{t-12}\}$, precipitation $\{P_t, P_{t-1},\ldots,P_{t-12}\}$, evaporation $\{E_t, E_{t-1},\ldots,E_{t-12}\}$ and SMS $\{I_t, I_{t-1},\ldots,I_{t-12}\}$ will be used as candidate input variables.
2. The correlation coefficient between $\{P_t, P_{t-1},\ldots,P_{t-12}\}$, $\{E_t, E_{t-1},\ldots,E_{t-12}\}$ and output variables is calculated in MATLAB. Suppose $\{P_t, P_{t-1}, P_{t-3}, P_{t-5}\}$ and $\{E_{t-1}, E_{t-2}, E_{t-6}\}$ are the optimal input combinations of rainfall and evaporation.
3. To get optimal input variable, a partial correlation algorithm is employed to calculate the partial correlation algorithm between streamflow $\{S_t, S_{t-1},\ldots,S_{t-12}\}$, SMS$\{I_t, I_{t-1},\ldots,I_{t-12}\}$ and output variables. This algorithm can eliminate the influence of precipitation and evaporation on streamflow and SMS, the algorithm is as follows:

196 $$r_{x_i y \cdot x_j} = \frac{r_{x_i y} - r_{x_i x_j} r_{y x_j}}{\sqrt{(1 - r^2_{x_i x_j})(1 - r^2_{y x_j})}} \quad (3)$$

197 where $r_{x_i y}$ is the correlation coefficient between x_i ({ P_t, P_{t-1}, P_{t-3},P_{t-5}}, {E_{t-1}, E_{t-2}, E_{t-

198 $_6$}) and output variables y, $r_{x_i x_j}$ is the correlation coefficient between x_i ({P_t, P_{t-1}, P_{t-

199 $_3$,P_{t-5}}, {E_{t-1}, E_{t-2}, E_{t-6}}) and x_j ({ S_t, S_{t-1},…,S_{t-12}}, {I_t, I_{t-1},…,I_{t-12}}), $r_{y x_j}$ is the correlation

200 coefficient between x_j ({ S_t, S_{t-1},…,S_{t-12}}, {I_t, I_{t-1},…,I_{t-12}}) and output variables. And

201 $r_{x_i y \cdot x_j}$ is the partial correlation coefficient between x_j ({ S_t, S_{t-1},…,S_{t-12}}, {I_t, I_{t-1},…,I_{t-}

202 $_{12}$})and output variables excluding the influence of x_i ({ P_t, P_{t-1}, P_{t-3},P_{t-5}} and {E_{t-1}, E_{t-}

203 $_2$, E_{t-6}}).

204 2.3.5 Performance assessment principles

205 The general standard proposed by Moriasi et al. (2007) was introduced to evaluate

206 the predictive performance of the model. The standard recommends using the Nash-

207 Sutcliffe efficiency (NSE) coefficient, the percentage of deviation (PBIAS) and the

208 ratio of the root mean square error (RMSE) to the standard deviation of the observations

209 (RSR). The NSE is defined as:

210 $$NSE = 1 - \frac{\sum_{i=1}^{n} (z_i - z_i^*)^2}{\sum_{i=1}^{n} (z_i - \bar{z})^2} \quad (4)$$

211 where $\bar{z}$ is the average of the observations; z_i and z_i^* represent the i-th observation

212 and predicted value, respectively; n is the length of observation.

213 The PBIAS is as Eq. (5).

214 $$PBIAS = \frac{\sum_{i=1}^{n} (z_i - z_i^*) \times 100}{\sum_{i=1}^{n} z_i} \% \quad (5)$$

PBIAS can be used to calculate how far the predicted value differs from the observed value.

The RSR is shown in Eq. (6), where the RMSE between the predicted value and the observed value is standardized based on the standard deviation of the observed value. The value range of this indicator is from zero to positive infinity, with zero being the best.

$$RSR = \frac{RMSE}{STDEV} = \frac{\sqrt{\sum_{i=1}^{n}(z_i - z_i^*)^2}}{\sqrt{\sum_{i=1}^{n}(z_i - \bar{z})^2}} \quad (6)$$

The algorithms used in the proposed framework all can be easily implemented with the help of Matlab software or other sciences software.

3. Results and discussion

3.1 The performance of SVM and ANN using correlation coefficient algorithm

In this paper, runoff, precipitation, evaporation, and SMS data are set to fourteen scenarios (Table 1) to find the optimal combination of input variables. The performance of SVM and ANN will be explored in fourteen scenarios. Meanwhile, two different input variable selection strategies are used to obtain the best input scenario for TWSA reconstruction and extension.

The performance of SVM and ANN in 14 scenarios using correlation coefficient algorithm (CCA) are shown in Table 2. Table 2 shows that in Scenario 1 to Scenario 4, SVM and ANN have the best performance in Scenario 4 (NSE 85.16% of SVM, NSE 81.33% of ANN). It is proved that SMS is the main factor affecting TWSA in Yunnan province (Han et. al 2019). At the same time, in Scenario 4, the performance of SVM

is better than that of ANN, which proves that the predictive ability of SVM is better than ANN under a small number of samples. Among the 14 scenarios, the performance of SVM in scenario 10 is best, and the same results can be found for the performance of ANN (NSE 89.00% of SVM, NSE 87.77% of ANN). Interesting results indicate that evaporation also affected the TSWA in Yunnan province.

3.2 The performance of SVM and ANN using the innovative input selection strategy

However, CCA does not consider the interaction between input variables when calculating the correlation coefficient between input variables and output variables, which may mislead the determination of the best combination of input variables. Thus, an innovative input selection strategy is proposed to eliminate the interaction among input variables. Table.3 and Fig. 3 show the performance of SVM and ANN using innovative input selection strategy in fourteen scenarios.

The results in Table. 3 and Fig. 3 show that in scenario 4 to 14, the performance of SVM and ANN is improved after using the innovative input selection strategy. Table. 3 and Fig. 3 show that among the fourteen scenarios, the best performance of SVM appears in scenario 13 (NSE is 93.06%), and the best performance of ANN also appears in scenario 13 (NSE is 92.98%). The NSE values of SVM and ANN increased by 5.33% and 6.70%, respectively. The results prove that when the input variables are not independent, the innovative input variable selection strategy can improve the predictive ability of the model.

Scenario 13 includes precipitation, evaporation, and SMS. The water balance formula in the basin system is $P\text{-}R\text{-}E=\pm\triangle S$, in which P is precipitation, R is streamflow,

E is evaporation and △S is water storage of the basin. At the same time, TWSA is the interannual variation of water storage (△S). Therefore, P, R, and E are the main factors affecting TWSA in Yunnan province.

3.3 The performance of the combined prediction

Table 3 and Figure 3 show that in scheme 13, the performance of SVM is better than that of ANN. However, in scheme 13, the performance of ANN is the best among the 14 schemes. Therefore, in order to make full use of the advantages of support vector machines and artificial neural networks, a combined forecasting method is proposed. The performance of CP and other models is shown in Fig. 4. The results in Fig. 4 show that the performance of CP in scenario 13 is better than that of SVM (the NSE value is increase by 1.26%). Me, CP has the best performance among the five models under scenario 13 (NSE is 94.25%, PBISA is -32.83, RSR is 0.24). At the same time, Long et al. made predictions on the TWSA in the three regions of the Yunnan-Guizhou Plateau, and the prediction accuracy was 91%, 83%, and 76%, respectively. In this study, the NSE value of TWSA extension using CP model and innovative selection strategy is 94.25%, and the prediction accuracy of the model has been improved. The results prove that the performance of CP in the reconstruction and expansion of TWSA is better than that of a single model.

The extension values of monthly △TWS (1962~2001) obtained by CP using the innovative selection strategy in Yunnan province are shown in Fig. 5. The results in Fig. 5 show that the monthly value of △TWS is the highest from June to August of the year, while the monthly value of △TWS is lower from November to April of the year.

Meanwhile, the precipitation in Yunan Province is mainly concentrated from June to August, accounting for 60% of the annual precipitation, while the precipitation from November to April only accounts for 10-20% of the annual precipitation. The annual change of the extended monthly △TWS values is highly consistent with the annual change of precipitation in Yunnan Province, which proves the extension of TWSA is reliable.

The annual △TWS value obtained by CP using the innovative selection strategy and the annual △TWS value obtained by the Global Land Data Assimilation System (GLDAS) are shown in Fig. 6. In the history of Yunnan province, there are seven extreme meteorological drought events during 1961~2001, that are 1962/1963, 1968/1969, 1978/1979, 1980, 1983, 1987, and 1992 (Zheng et al., 2017). The results in Fig. 6 show that the annual △TWS values obtained by CP catch five of seven extreme meteorological drought events from 1961 to 2001. However, the annual △TWS value obtained from GLDAS only captured three of seven extreme meteorological drought events that occurred between 1961 and 2001. Moreover, the GLDAS △TWS value only captures the beginning of two extreme meteorological drought events in 1968/1969 and 1978/1979 and failed to capture the water shortage in the middle and late stages of these two drought events. At the same time, the annual △TWS values obtained by the CP model also captured the two catastrophic floods that occurred in Yunnan in 1991 and 1998, but the GLDAS △TWS value did not capture the two catastrophic floods. Therefore, the accuracy of the TWSA extension obtained through the CP model is higher than the TWSA of the GLDAS product obtained through the

physical model. Yunnan Province is located in a typical karst landform area, where surface water and groundwater are closely connected. However, physical models may ignore the interrelationship between surface water and groundwater, which may cause distortion of the total water storage data they generate.

Figure 7(a) shows the correlation analysis between the monthly △TWS obtained through the CP model and the monthly normalized vegetation index (NDVI) during the verification period. As shown in Fig. 7(a), there is a significant negative correlation between NDVI and △TWS. Fig. 7(b) shows the monthly NDVI anomaly value and the monthly observation value of △TWS during the training period and the monthly △TWS extension value during the verification period. As shown in Figure 7(b), during the training period, the observed value of monthly △TWS is highly consistent with the monthly NDVI anomaly value. At the same time, the monthly △TWS is also highly consistent with the monthly NDVI anomaly value during the verification period. The results in Figure 7 prove that the TWSA extension value obtained through the CP model and the innovative input selection strategy has reliable accuracy. Therefore, in Yunnan Province, the reliability of the TWSA expansion value obtained through the CP model is higher than that of the GLDAS TWSA product. The results further proved that the CP model using an innovative selection strategy has high accuracy for the extension and reconstruction of TWSA in Yunnan Province.

4. Conclusions

The prediction accuracy and ability of the SVM, ANN, and CP using the innovative selection strategy were explored by predicting monthly TWSA series in Yunnan

province in the present study. Meanwhile, fourteen scenarios of input variables were explored to select an optimal input scenario for TWSA prediction. The performance of the SVM was superior to the ANN due to its power prediction ability for a small sample set. Of the fourteen scenarios, the performances of SVM and ANN using CCA in scenario 10 (including evaporation and SMS) were optimal, and the performance of SVM was better than that of ANN. The performances of SVM and ANN using the innovative selection strategy were all better than that of SVM and ANN using CCA in scenarios 5 to 14. In scenario 13 (including precipitation, evaporation, and SMS), SVM using the innovative selection strategy performed best among fourteen scenarios (with NSE 93.08%), and the same result was found in the ANN (with NSE 92.13%). Therefore, input variables including precipitation, evaporation, and SMS are the optimal input variables for TWSA prediction. Moreover, in scenario 13, the performance of CP was superior to SVM and ANN using the innovative selection strategy (with NSE 94.25%). Thus, CP with the innovative selection strategy has good stability and high prediction accuracy for TWSA prediction. Therefore, this study provides a superior choice for monthly TWSA prediction. TWSA is a significant tool for the monitor drought and flood, providing a model with high prediction accuracy is extremely meaningful for disaster prevention and mitigation.

Acknowledgements

This study was jointly funded by the National Key Research and Development Program of China (grant number 2017YFC0405900), the National Natural Science Foundation of China (grant number 51709221), the Planning Project of Science and

346 Technology of Water Resources of Shaanxi (grant numbers 2015slkj-27 and 2017slkj-
347 19), the China Scholarship Council (grant number 201908610170), the Open Research
348 Fund of State Key Laboratory of Simulation and Regulation of Water Cycle in River
349 Basin (China Institute of Water Resources and Hydropower Research, grant number
350 IWHR-SKL-KF201803) and the Doctorate Innovation Funding of Xi'an University of
351 Technology (grant number 310-252072007).

352

353 **Ethics declarations**

354 **Conflict of interest**

355 The authors have no conflicts of interest to declare.

356

357 **Ethical Approval**

358 Not applicable.

359

360 **Consent to Participate**

361 Not applicable.

362

363 **Consent to Publish**

364 Not applicable.

365

366 **Authors Contributions**

367 Conceptualization: S. H; E. M. Methodology: E. M; W. F. Formal analysis and

investigation: E. M; L. C. Writing - original draft preparation: E. M. Writing - review and editing: E. M. Supervision: Q. H.

Competing Interests

None

Available of data and materials

Authors agree with data transparency and undertake to provide any required data and material.

References

Andrew, R.; H. Guan; O. Batelaan. Estimation of grace water storage components by temporal decomposition. *Journal of Hydrology.* **2017**, 552, 341-50.

Cortes, C.; V. Vapnik. Support-vector networks. *Machine learning.* **1995**, 20, 273-97.

Carrier, C.; A. Kalra; S. Ahmad. Using paleo reconstructions to improve streamflow forecast lead time in the western u nited s tates. *JAWRA Journal of the American Water Resources Association.* **2013**, 49, 1351-66.

Ch, S.; N. Anand; B. K. Panigrahi; S. Mathur. Streamflow forecasting by svm with quantum behaved particle swarm optimization. *Neurocomputing.* **2013**, 101, 18-23.

de Linage, C.; J. Famiglietti; J. Randerson. Forecasting terrestrial water storage changes in the amazon basin using atlantic and pacific sea surface temperatures. *Hydrology & Earth System Sciences Discussions.* **2013**, 10, 12453–83.

Doell, P.; H. Mueller Schmied; C. Schuh; F. T. Portmann; A. Eicker. Global‐scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and grace satellites. *Water Resources Research.* **2014**, 50, 5698-720.

Eicker, A.; E. Forootan; A. Springer; L. Longuevergne; J. Kusche. Does grace see the terrestrial water cycle “intensifying”? *Journal of Geophysical Research: Atmospheres.* **2016**, 121, 733-45.

Famiglietti, J. S.; M. Rodell. Water in the balance. *Science.* **2013**, 340, 1300-01.

Fang, W.; S. Huang; Q. Huang; G. Huang; E. Meng; J. Luan. Reference evapotranspiration forecasting based on local meteorological and global climate information screened by partial mutual information. *Journal of Hydrology.* **2018**, 561, 764-79.

He, Z.; X. Wen; H. Liu; J. Du. A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *Journal of Hydrology.* **2014**, 509, 379-86.

Huang, S.; J. Chang; Q. Huang; Y. Chen. Monthly streamflow prediction using modified emd-based support vector machine. *Journal of Hydrology.* **2014**, 511, 764-75.

Han, Z.; S. Huang; Q. Huang; G. Leng; H. Wang; L. He; W. Fang; P. Li. Assessing grace-based terrestrial water storage anomalies dynamics at multi-timescales and their correlations with teleconnection factors in yunnan province, china. *Journal of Hydrology.* **2019**, 574, 836-50.

Huang, Z.; P. J.-F. Yeh; Y. Pan; J. J. Jiao; H. Gong; X. Li; A. Güntner; Y. Zhu; C. Zhang; L. Zheng. Detection of large-scale groundwater storage variability over the karstic regions in southwest china. *Journal of hydrology.* **2019**, 569, 409-22.

Jiang, Z.; R. Li; A. Li; C. Ji. Runoff forecast uncertainty considered load adjustment model of cascade hydropower stations and its application. *Energy.* **2018**, 158, 693-708.

Liang, X.; D. P. Lettenmaier; E. F. Wood; S. J. Burges. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research: Atmospheres.* **1994**, 99, 14415-28.

Long, D.; Y. Shen; A. Sun; Y. Hong; L. Longuevergne; Y. Yang; B. Li; L. Chen. Drought and flood monitoring for a large karst plateau in southwest china using extended grace data. *Remote Sensing of Environment.* **2014**, 155, 145-60.

Long, D.; Y. Yang; Y. Wada; Y. Hong; W. Liang; Y. Chen; B. Yong; A. Hou; J. Wei; L. Chen. Deriving scaling factors using a global hydrological model to restore grace total water storage changes for china's yangtze river basin. *Remote Sensing of Environment.* **2015**, 168, 177-93.

Long, D.; L. Longuevergne; B. R. Scanlon. Global analysis of approaches for deriving total water storage changes from grace satellites. *Water Resources Research.* **2015**, 51, 2574-94.

Moriasi, D. N.; J. G. Arnold; M. W. Van Liew; R. L. Bingner; R. D. Harmel; T. L. Veith. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE.* **2007**, 50, 885-900.

Meng, E.; S. Huang; Q. Huang; W. Fang; L. Wu; L. Wang. A robust method for non-stationary streamflow prediction based on improved emd-svm model. *Journal of hydrology.* **2019**, 568, 462-78.

Meng, E; Huang, S; Huang, Q; Fang, W; Wang, H; Leng, G; Wang, L; Liang, H. A Hybrid VMD-SVM

Model for Practical Streamflow Prediction Using an Innovative Input Selection Framework. Water Resources Management, **2021**, 35(4): 1321-1337.

Pan, M.; A. K. Sahoo; T. J. Troy; R. K. Vinukollu; J. Sheffield; E. F. Wood. Multisource estimation of long-term terrestrial water budget for major global river basins. *Journal of Climate.* **2012**, 25, 3191-206.

Reager, J. T.; A. C. Thomas; E. A. Sproles; M. Rodell; H. K. Beaudoing; B. Li; J. S. Famiglietti. Assimilation of grace terrestrial water storage observations into a land surface model for the assessment of regional flood potential. *Remote Sensing.* **2015**, 7, 14663-79.

Wahr, J.; M. Molenaar; F. Bryan. Time variability of the earth's gravity field: Hydrological and oceanic effects and their possible detection using grace. *Journal of Geophysical Research: Solid Earth.* **1998**, 103, 30205-29.

Wouters, B.; J. A. Bonin; D. P. Chambers; R. E. Riva; I. Sasgen; J. Wahr. Grace, time-varying gravity, earth system dynamics and climate change. *Rep Prog Phys.* **2014**, 77, 116801. 10.1088/0034-4885/77/11/116801. https://www.ncbi.nlm.nih.gov/pubmed/25360582.

Yoon, H.; Y. Hyun; K. Ha; K.-K. Lee; G.-B. Kim. A method to improve the stability and accuracy of ann-and svm-based time series models for long-term groundwater level predictions. *Computers & Geosciences.* **2016**, 90, 144-55.

Zhang, X.-J.; Q. Tang; M. Pan; Y. Tang. A long-term land surface hydrologic fluxes and states dataset for china. *Journal of Hydrometeorology.* **2014**, 15, 2067-84.

Zhang, Z.; B. Chao; J. Chen; C. R. Wilson. Terrestrial water storage anomalies of yangtze river basin droughts observed by grace and connections with enso. *Global and Planetary Change.* **2015**, 126, 35-45.

Zheng, J; Wei, H; Yan, C; Zhou, J. Study on Meteorological Extreme-Drought Index for Yunnan Province. Plateau Meteorology, 2017, 36(4): 1039-1051.

Figures

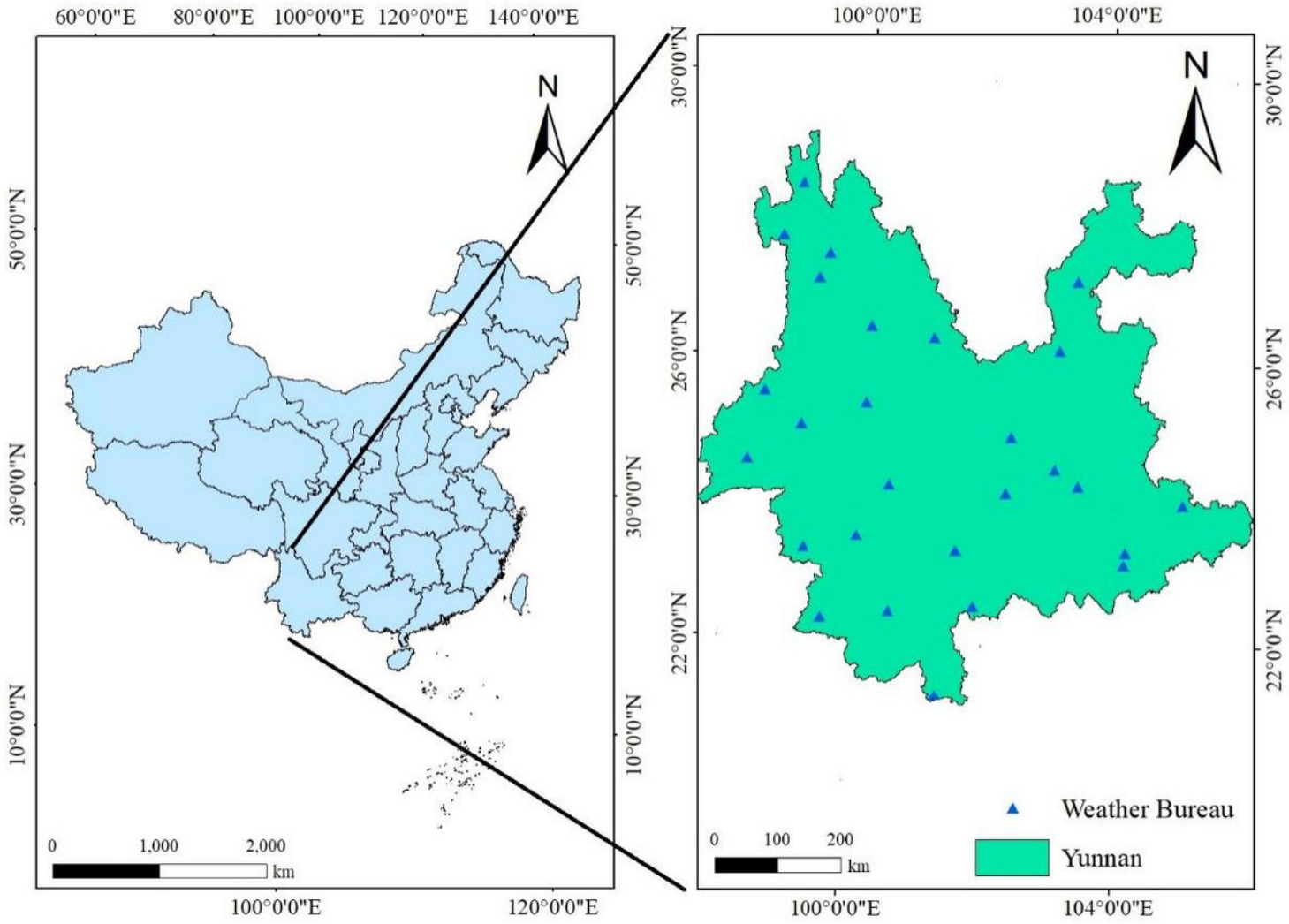

Figure 1

Location of the Yunnan province.

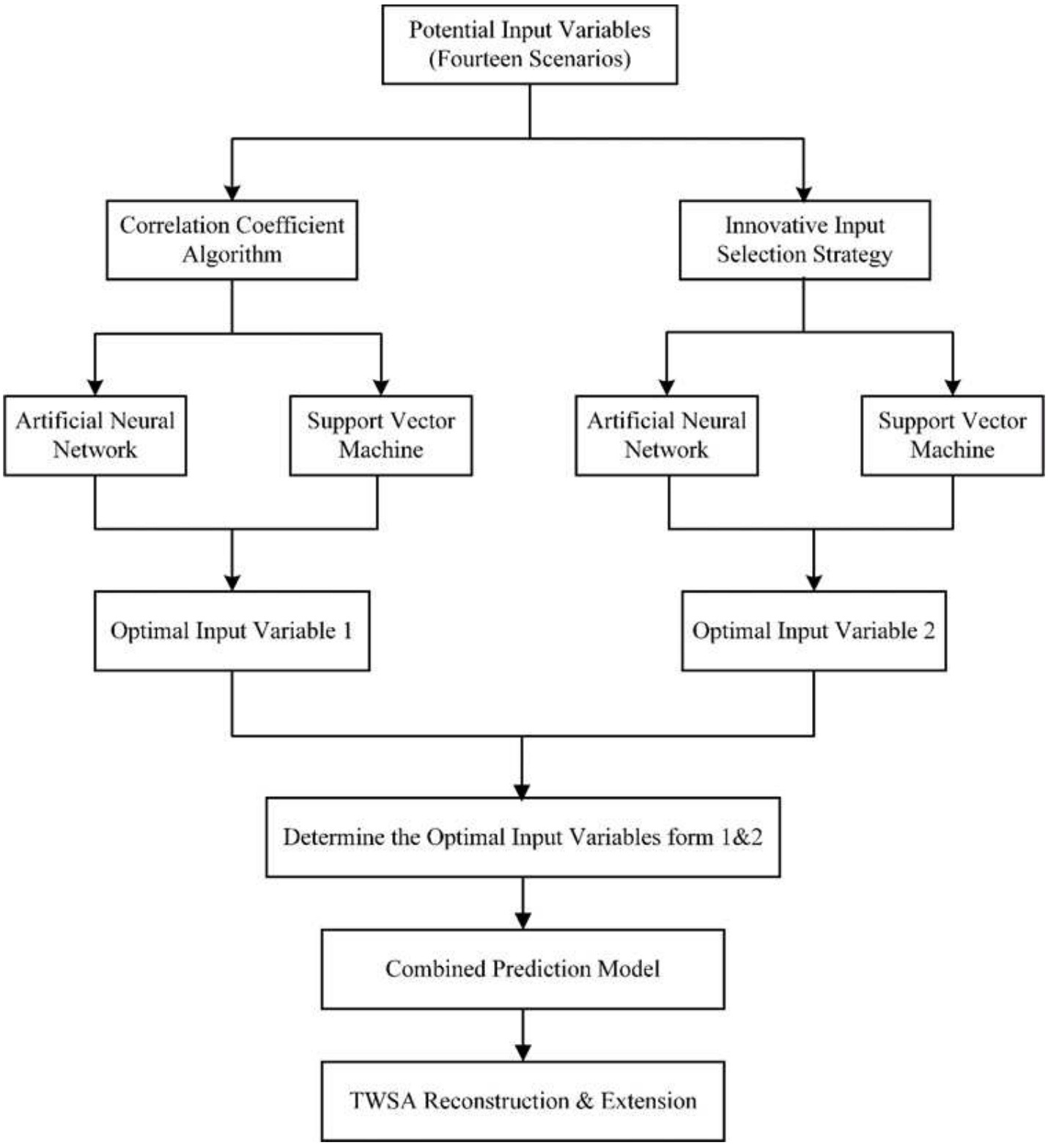

Figure 2

Flow chart of the Combined Prediction Model.

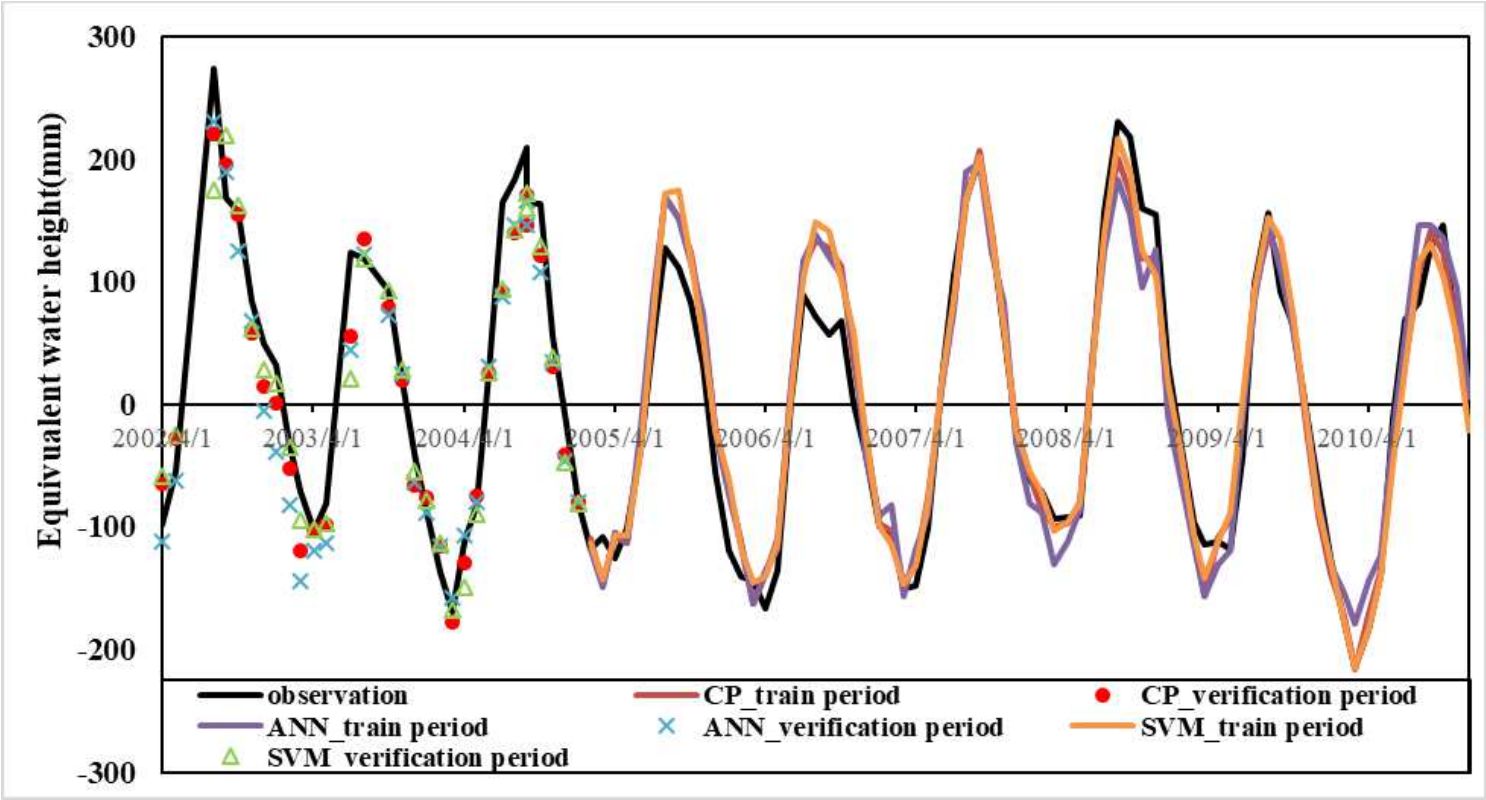

Figure 3

The performance of combined prediction (CP), ANN and SVM of TWSA using the innovative input selection strategy in test period and verification period.

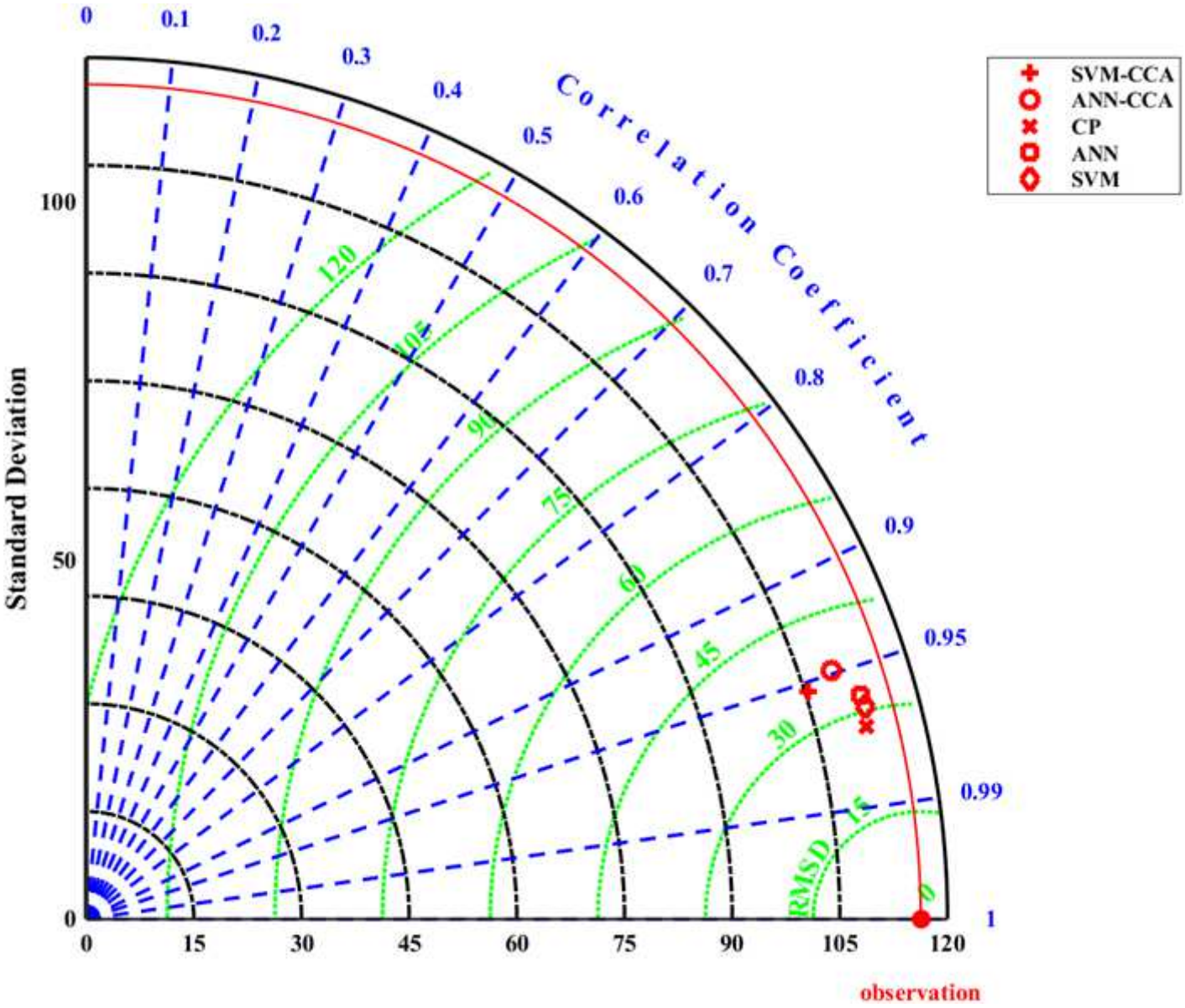

Figure 4

Taylor diagram of prediction by SVM and ANN using correlation coefficient algorithm (CCA), and combined prediction (CP), SVM, ANN using the innovative input selection strategy in verification period in Yunnan province where blue contours represent Pearson correlation coefficient; green contours represent centered RMS error in the simulated field; and black contours represent standard deviation of the simulated pattern.

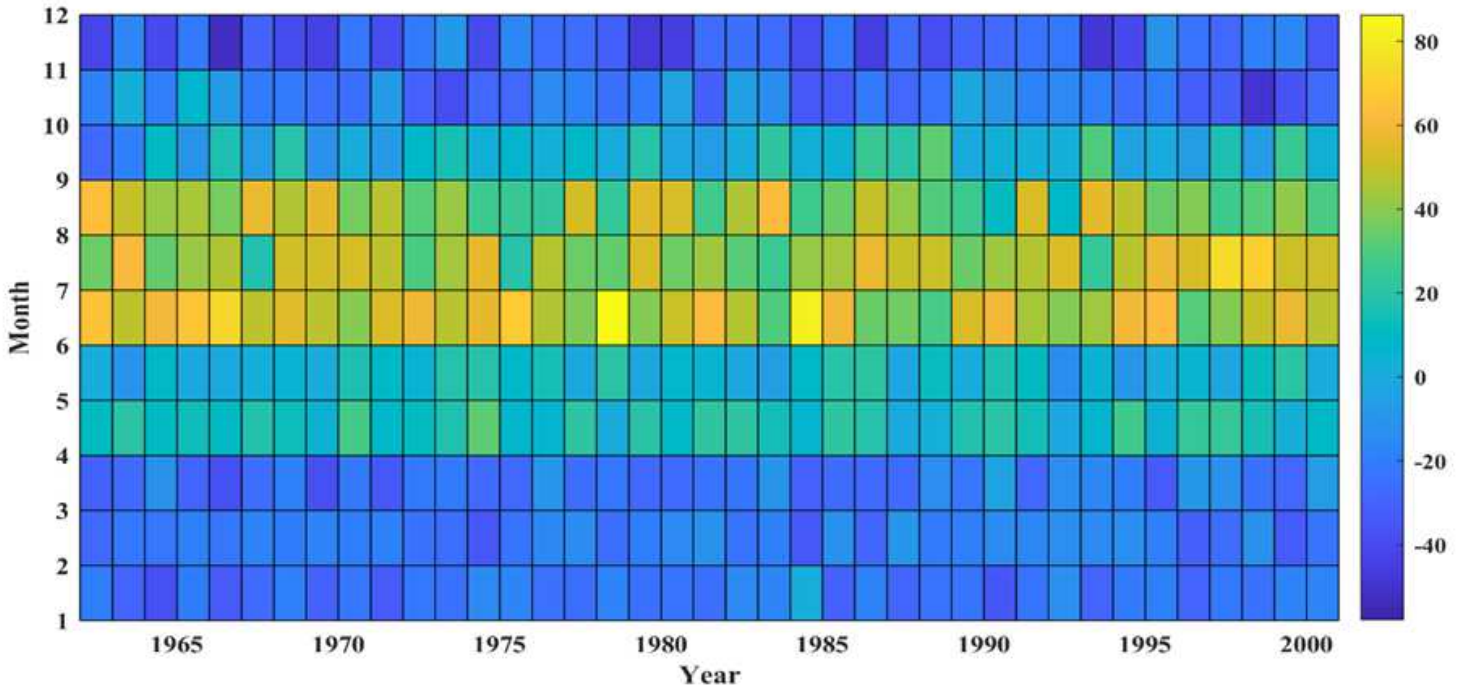

Figure 5

The monthly ΔTWS values obtained by CP using the innovative input selection strategy in Yunnan province.

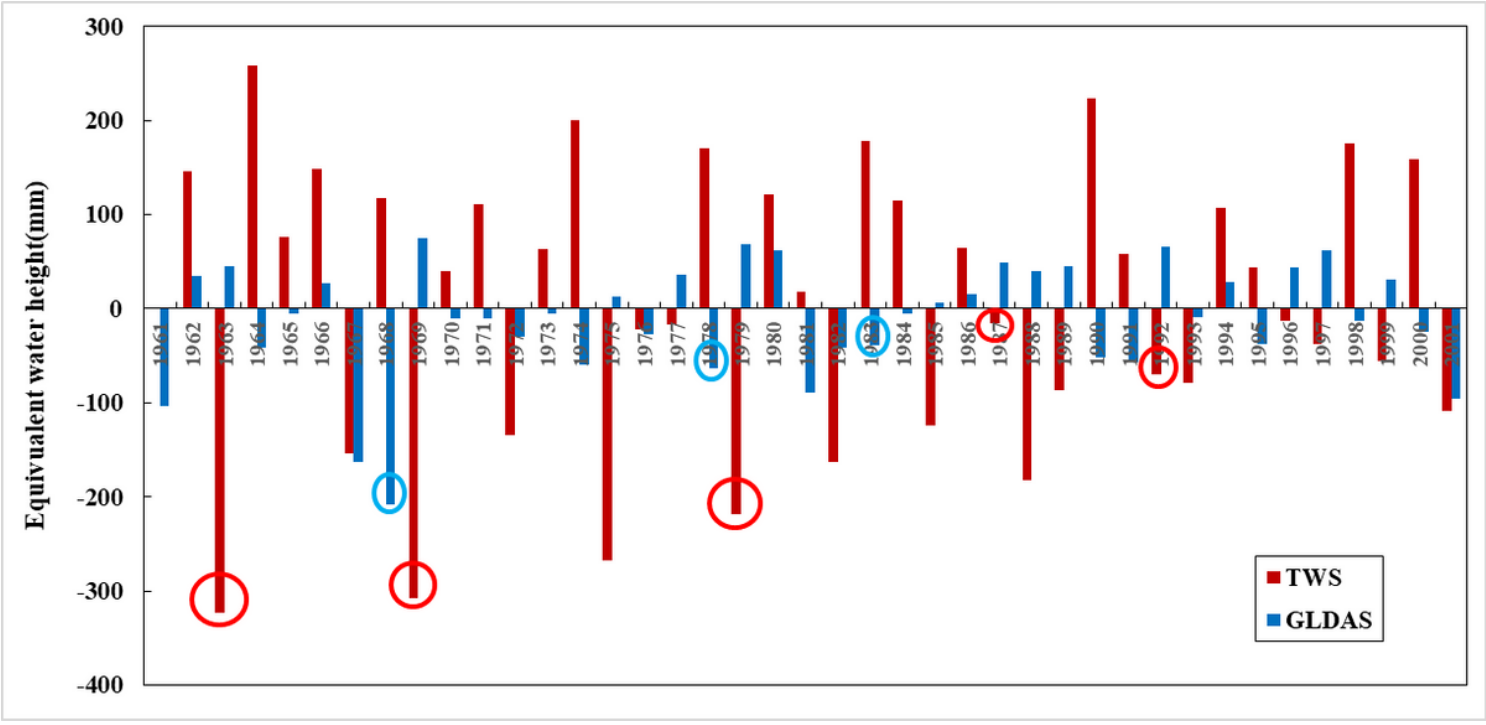

Figure 6

The annual ΔTWS values obtained by CP model using the innovative input selection strategy and GLDAS in Yunnan province

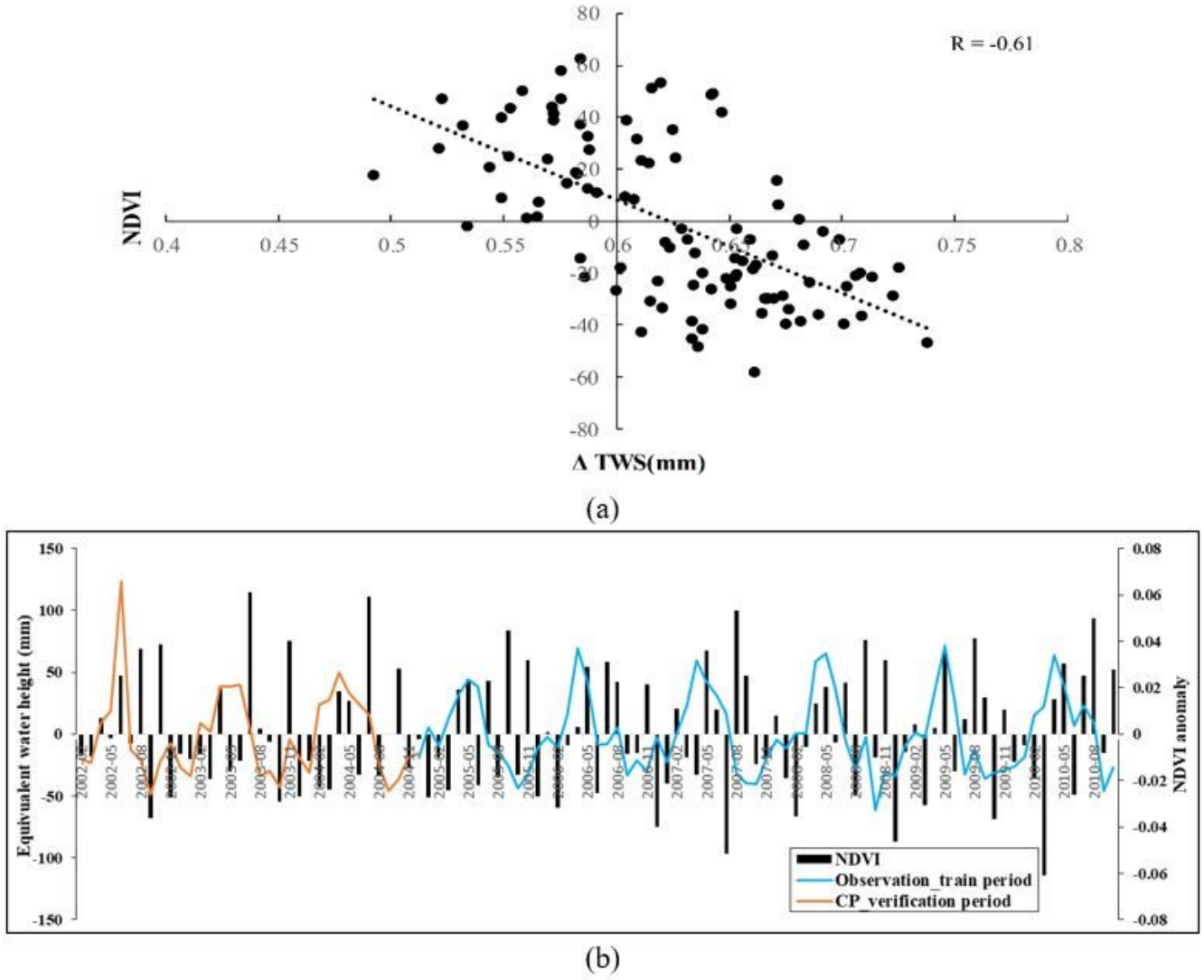

Figure 7

(a) The correlation analysis between monthly NDVI and monthly ΔTWS by CP model using the innovative input selection strategy in verification period; (b) The monthly NDVI anomaly and monthly ΔTWS.