ORIGINAL ARTICLE

Open Access

Inversion of surface vegetation water content based on GNSS-IR and MODIS data fusion

Yalong Pan[1], Chao Ren[1,2]*, Yueji Liang[1,2], Zhigang Zhang[1] and Yajie Shi[1]

Abstract

Obtaining high-precision, long-term sequences of vegetation water content (VWC) is of great significance for assessing surface vegetation growth, soil moisture, and fire risk. In recent years, the global navigation satellite system-interferometric reflection (GNSS-IR) has become a new type of remote sensing technology with low cost, all-weather capability, and a high temporal resolution. It has been widely used in the fields of snow depth, sea level, soil moisture content, and vegetation water content. The normalized microwave reflectance index (NMRI) based on GNSS-IR technology has been proven to be effective in monitoring changes in VWC. This paper considers the advantages and disadvantages of remote sensing technology and GNSS-IR technology in estimating VWC. A point-surface fusion method of GNSS-IR and MODIS data based on the GA–BP neural network is proposed to improve the accuracy of VWC estimation. The vegetation index products (NDVI, GPP, LAI) and the NMRI that unified the temporal and spatial resolution were used as the input and output data of the training model, and the GA–BP neural network was used for training and modeling. Finally, a spatially continuous NMRI product was generated. Taking a particular area of the United States as a research object, experiments show that (1) a neural network can realize the effective fusion of GNSS-IR and MODIS products. By comparing the GA–BP neural network, BP neural network, and multiple linear regression (MLR), the three models fusion effect. The results show that the GA–BP neural network has the best modeling effect, and the *r* and RMSE between the model estimation result and the reference value are 0.778 and 0.0332, respectively; this network is followed by the BP neural network, in which the *r* and RMSE are 0.746 and 0.0465, respectively. MLR has the poorest effect, with *r* and RMSE values of 0.500 and 0.0516, respectively. (2) The spatiotemporal variation in the 16 days/500 m resolution NMRI product obtained by GA–BP neural network fusion is consistent with that in the experimental area. Through the testing of GNSS stations that did not participate in the modeling, the *r* between the estimated value of the NMRI and the reference value is greater than 0.87, and the RMSE is less than 0.049. Therefore, the method proposed in this paper is optional and effective. The spatially continuous NMRI products obtained by fusion can reflect the changes in VWC in the experimental area more intuitively.

Keywords: Vegetation water content, GNSS-IR, MODIS, GA–BP, Fusion, Precision analysis

Introduction

Surface vegetation is one of the critical components of terrestrial ecosystems, and the water content in the vegetation canopy is approximately 40–80% (Elvidge 1990),

playing an essential role in soil formation and environmental changes. Vegetation water content (VWC) is one of the main factors controlling plant photosynthesis, respiration, primary productivity, and biomass. It plays an essential role in plant physiological status, vegetation function, drought, and fire risk assessment (Peñuelas et al. 1993). Because a plant is covered with soil, its water content will affect the monitoring of soil moisture, and the correct estimation of VWC can improve the accuracy

*Correspondence: renchao@glut.edu.cn
[1] College of Geomatics and Geoinformation, Guilin University
of Technology, Guilin 541006, China
Full list of author information is available at the end of the article

of soil moisture inversion (Jin et al. 2017b). Therefore, obtaining high-precision, long-term sequences of VWC is of considerable significance to the study of surface vegetation and soil moisture.

In 1980, the European Space Agency (ESA) first proposed that GPS L-band signals could be used as ocean scatterometers (Hall 1988). Subsequently, in 1993, ESA's Martin-Neria first introduced the concept of PARIS and, for the first time, used GPS-reflected signals to achieve ocean height measurements (Martin-Neira 1993). Since then, foreign scholars have realized that GPS-reflected signals might serve as a new remote sensing method (Liu et al. 2007). Therefore, two new GNSS remote sensing technologies have been proposed: Global Navigation Satellite System-Reflection (GNSS-R) remote sensing technology and Global Navigation Satellite System-Interferometric Reflection (GNSS-IR) remote sensing technology. However, GNSS-R technology needs to receive both direct and reflected signals from the satellite at the same time. Therefore, it requires an instrument with both a left-handed circularly polarized (LHCP) antenna and a right-handed circularly polarized (RHCP) antenna. Therefore, GNSS-R technology has higher requirements for hardware and a higher cost; thus, it has limited its development and promotion to a certain extent. GNSS-IR technology requires only an ordinary geodetic receiver to perform experiments and has the characteristics of being low cost, useful in all weather, high accuracy, and high spatiotemporal resolution.

GNSS-IR technology was first proposed by Professor Larson of the University of Colorado. Larson and others separated the direct and reflected components from the GPS signal-to-noise ratio (SNR) observations and studied the reflected components and the surrounding environment of the station. The results show that the amplitude of the reflected components can reflect the overall change in the surrounding soil moisture (Larson et al. 2008a). Subsequent research found that the reflected signal could well reflect the evolution of soil moisture within 5 cm of the soil surface. There is a certain correlation between the reflected signal phase, satellite height angle change, and soil moisture content; however, when the soil moisture content is less than 10%, the correlation weakens (Larson et al. 2008b, 2010). Considering that GNSS-IR technology can realize the estimation of surface environmental parameters, Larson and others extended GNSS-IR technology to snow depth monitoring and provided a series of results for the development of GNSS-IR technology (Larson et al. 2009; Larson and Nievinski 2013; Larson 2016). Since then, GNSS-IR technology has been widely used to monitor changes in the surface environment, such as soil moisture (Larson 2016; Liang et al. 2019; Ren et al. 2019), sea level (Larson et al. 2013; Jin

et al. 2017a), and snow depth (Zhang et al. 2017, 2018). In terms of vegetation detection, Small and others used the GPS noise statistic MP_1 root mean square (RMS) for the first time to qualitatively estimate the growth of plants. They noted that the SNR data in the reflected signal could reflect the growth of vegetation (Small et al. 2010). Chew and others used a model for soil moisture inversion to quantitatively analyze VWC and SNR and the actual reflection surface height. The results show that when VWC does not exceed 1.5 kg/m^2, the vegetation water content and SNR amplitude have a linear relationship (Chew et al. 2014). Chen and others proposed a method based on the amplitude and frequency analysis of the interference pattern of the SNR to eliminate the influence of VWC in soil moisture retrieval and achieved excellent results (Chen et al. 2016). In terms of VWC, Larson and others defined a daily VWC metric based on the amplitude of the reflected signal based on the relationship between the SNR amplitude and VWC, which is the normalized microwave reflection index (NMRI) (Larson and Small 2014). In 2004, verification was performed at four grassland sites in Montana, and the results showed that under the conditions of similar vegetation and climate, the NMRI and VWC have a strong correlation, and the VWC can be more accurately retrieved by the NMRI (Small et al. 2014). At present, the Plate Boundary Observatory (PBO) in the United States provides daily NMRI data. However, because the NMRI data are based on GPS stations, only VWC changes within 1000 m^2 around the station are monitored. Additionally, the interval between GNSS stations of the PBO observation network is large, and spatial continuity cannot be achieved. Therefore, the application of the NMRI product is further limited.

In recent years, with the rapid development of remote sensing technology and imaging spectroscopy technology, it is relatively easy to obtain large-scale and long-term VWC data using remote sensing technology. Due to the different sensors, the current remote sensing technology can be divided into optical remote sensing and microwave remote sensing. Among them, the normalized difference vegetation index (NDVI), leaf area index (LAI), gross primary productivity (GPP) and other vegetation indexes provided by MODIS satellites belonging to optical remote sensing are widely used in VWC research (Tucker et al. 2005). Although these vegetation indexes have a high spatial resolution, due to the defects of optical sensors, the vegetation index images are easily affected by clouds and smog, resulting in the loss of information. Additionally, the NDVI is primarily regarded as an indicator of vegetation greenness. These indexes are used to infer the biomass and LAI and other vegetation indexes (Gutman and Ignatov 1998; Paruelo et al. 1997; Wylie et al. 2002). Because environmental

factors such as plant type and hydrometeorology have different effects on VWC and "greenness," the NDVI cannot be used to accurately estimate the VWC. At present, microwave remote sensing has also been used to successfully estimate VWC (Brakke et al. 1981; Owe et al. 2001). Although microwave remote sensing is not affected by clouds and fog, its spatial resolution is low. Some scholars have proposed a method of fusing the effects of optical remote sensing images and microwave remote sensing (Dasgupta and Qu 2006). However, due to the significant difference in the spatial resolution between the two, the accuracy and spatial resolution of the fusion results are weak in practical applications. Therefore, using remote sensing technology to estimate VWC has certain flaws.

At present, how to make full use of multisource data to achieve the fusion of multisource data has become a research hotspot. Multisource data fusion has a complex nonlinear relationship, and machine learning technology can better solve nonlinear problems. Therefore, many scholars have conducted multisource data fusion and research using machine learning technology and have achieved good results (Li et al. 2017; Xu et al. 2018; Yuan et al. 2020). In summary, the NMRI products based on GNSS-IR can better reflect the changes in VWC. The vegetation index products based on remote sensing technology have the characteristics of being large-scale and having a long-term series. In this paper, a fusion inversion method of GNSS-IR and remote sensing data based on the GA−BP neural network is proposed. The spatial resolution of optical remote sensing is better than that of microwave remote sensing. Therefore, this article selects MODIS vegetation index products for experiments. The GA−BP neural network is used to achieve effective fusion of the two and fully exploit their advantages. Finally, a spatially continuous NMRI product map is established. To compensate for the limitations of the original NMRI product space, this product is used to better estimate VWC.

Model principle and methodology
GNSS-IR and NMRI principles
The core observational measurement of GNSS-IR is SNR data. SNR is a composite signal of direct component A_d and reflected component A_r. The commonly used receivers are currently designed to be right-handed circularly polarized, resulting in $A_d \gg A_r$. Because A_d is much larger than A_r, the direct signal A_d is a trend term. It determines the overall change in SNR. The reflected signal A_d appears as a local periodic fluctuation. Because A_d and A_r differ significantly, they can be separated by using low-order polynomials. There is a fixed frequency sine (cosine) function relationship between the separated

reflection components A_r and $\sin(\theta)$ (Chew et al. 2013). The reflection component can be expressed as follows:

$$A_r = A \cos \left(\frac{4\pi h}{\lambda} \sin(\theta) + \varphi \right) \tag{1}$$

In the formula, θ, λ, and h represent the satellite height angle, carrier wavelength, and vertical distance from the antenna phase center to the reflection point, respectively, and A and ϕ represent the amplitude and phase of the reflection component, respectively. If $t = \sin(\theta)$ and $f = 2h/\lambda$, then Equation (1) can be expressed as follows:

$$A_r = A \cos \left(2\pi f t + \varphi \right) \tag{2}$$

The NMRI is a comprehensive index used to evaluate the amplitude change of the reflected signal. The core is to calculate the RMS of the pseudo-range multipath index MP_1 on the L1 carrier. MP1 is defined as (Estey and Meertens 1999):

$$MP_1 = P_1 - \frac{f_1^2 + f_2^2}{f_1^2 - f_2^2} \lambda_1 \varphi_1 + \frac{2f_2^2}{f_1^2 - f_2^2} \lambda_2 \varphi_2 \tag{3}$$

In the equation, P_1 is the pseudo-range observation on the L_1 carrier; f_1 and f_2 are the carrier frequencies of L_1 and L_2, $f_1 = 1575.42$ MHz, $f_2 = 1227.6$ MHz; λ_1 and λ_2 are the carrier wavelengths of L_1 and L_2, $\lambda_1 = 0.19$ m, and $\lambda_2 = 0.24$ m; ϕ_1 and ϕ_2 are the L_1 and L_2 carrier phase observations. The calculation of the NMRI is based on the RMS value of MP_1, and its calculation method is as follows:

$$NMRI = \frac{\max\left(RMS_{MP_1}\right) - RMS_{MP_1}}{\max\left(RMS_{MP_1}\right)}. \tag{4}$$

In the equation, RMS_{MP_1} is the RMS value of MP_1 on a single day, and $\max\left(RMS_{MP_1}\right)$ is the average value of RMS_{MP_1} in the top 5% after the RMS_{MP_1} values are ranked from large to small.

GA−BP neural network
Considering the time and space complexity of GNSS-IR and MODIS data, it is difficult to achieve effective fusion using only linear methods. Therefore, this article attempts to use the artificial neural network (ANN) of the typical BP neural network model for experimental analysis. The BP neural network was proposed by McClelland and Rumelhart (Rummelhart et al. 1986; Mcclland et al. 1986). It is a multilayer feedforward neural network that can better handle nonlinear problems. Its main structure is composed of an input layer, a hidden layer, and an output layer. Each layer is formed by several neurons, and the output value of each neuron is determined by the input value, action function, and threshold. The learning

process of the BP neural network includes two processes: forward information propagation and error backward propagation. In the forward propagation process, the input information is transmitted from the input layer through the hidden layer to the output layer, and the output value and the expected value are compared. If there is an error, the error is propagated back along the original connection path, and the weights of the neurons in each layer are modified layer by layer to reduce the error. This cycle is repeated until the output results meet the accuracy requirements (Alpaydin 2004). The BP neural network topology is shown in Fig. 1.

In Fig. 1, x_1, x_2, ..., x_n is the input value of the BP neural network; y_m is the output value of the neural network; w_{ij} is the connection weight of input layer neurons and hidden layer neurons; w_{jk} is the connection weight of hidden layer neurons and output layer neurons.

Although the BP neural network can better address nonlinear problems, it still has flaws (Srinivas and Deb 1994). The first is related to the BP neural network modeling process. To reduce network errors and improve accuracy, the number of hidden layer neurons must be appropriately selected. However, there is no specific calculation method for the optimal number of hidden layer neurons. Second, the initial weights and thresholds of the BP neural network are randomly generated. Each node and weight of the BP neural network will affect the output. Therefore, the adaptation process and the global approximation process are time-consuming, resulting in a slow convergence of the network. Finally, the BP neural network uses a gradient descent method, which has local optimization problems.

To solve the shortcomings of the above BP neural network, this paper uses the genetic algorithm (GA) to optimize the BP neural network. The GA is a parallel random search optimization method formed by simulating genetic mechanisms and biological evolution in nature (Holland 1975). This method is based on the sample fitness function and selects, intersects, and mutates the initial population to guide learning and determine the search direction. Because it uses the population to search, it can use random methods to find the optimal solution in multiple regions of the global solution space, making it particularly suitable for large-scale parallel processing (Goldberg 1989). Considering the defects of the BP neural network and the advantages of the GA, the two are combined to construct the GA–BP neural network algorithm. First, the GA is used to optimize the weights and thresholds of the BP network to reduce the range of weights and thresholds. Subsequently, the optimized weights and thresholds are inputted to the BP neural network and solved accurately to improve the training accuracy and speed of the BP neural network. The GA–BP model algorithm flow is shown in Fig. 2.

Methodology

The methodologies of GNSS-IR and MODIS fusion inversion based on the GA–BP neural network are as follows:

1. Study area setting. Select an area where the GNSS stations are evenly distributed, the terrain is diverse, and the climate change is significant;
2. Data collection. Among them, the NMRI data came from PBO H2O (https://gnss-h2o.Jpl.nasa.gov/index.php), and the three MODIS vegetation index products came from the Google Earth Engine (GEE) (https://developers.google.com/earth-engine/datasets/). Detailed information on each product is shown in Table 1.
3. Data preprocessing and data set construction. First, we discuss the effect of GNSS-IR and MODIS fusion in the short term. First, we extract the corresponding vegetation index values from the three vegetation images based on the latitude, longitude, and date of the 37 GNSS stations and establish a raw dataset consisting of the NMRI, NDVI, GPP, LAI, Lat, and Lon. Subsequently, the correlation between the NMRI and three vegetation indexes was analyzed to verify whether all three indexes could be included in modeling. Finally, the time resolution of NMRI, GPP, and LAI was unified to 16 days through averaging, and a modeling dataset with 16 days/500 m resolution was finally established. In the modeling dataset, the NDVI, GPP, LAI, LAT, and LON are the input of modeling and the output of NMRI modeling.
4. Establish a training model. In the modeling dataset, the data of 34 stations are selected as the model input sequence and use the corresponding NMRI as the model output sequence. Follow the steps in Fig. 2 of

Fig. 1 BP neural network topology
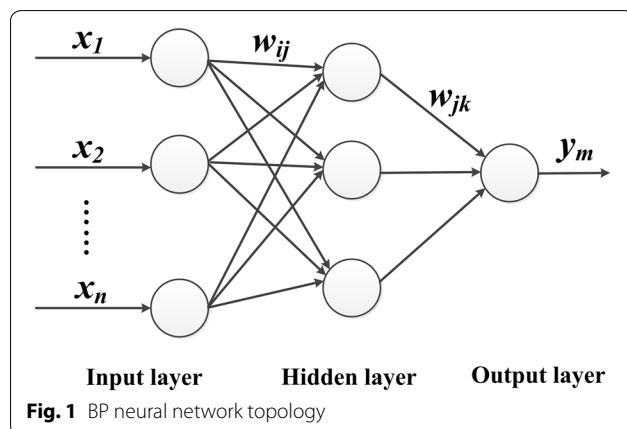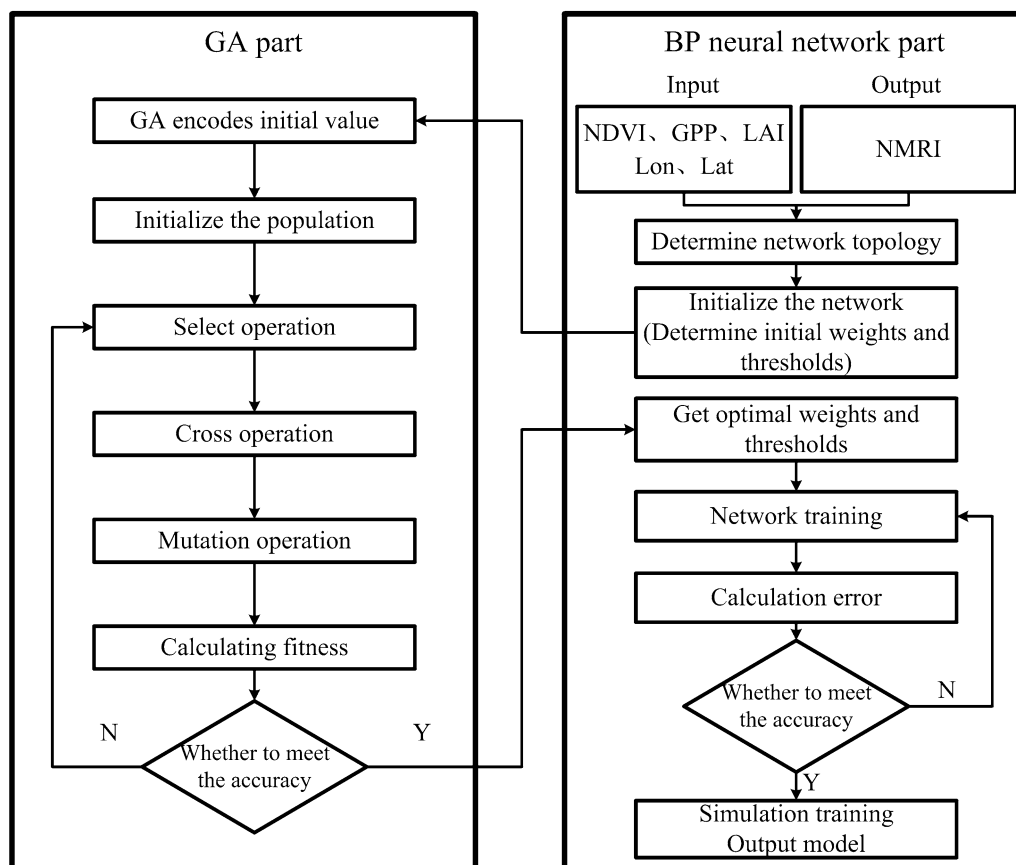
Fig. 2 GA–BP model algorithm flow

Table 1 Product details

Index	Resolution	Product	Time
NMRI	Daily/station	PBO H$_2$O	2010.7.78–2010.10.16
NDVI	16 days/500 m	MOD13A1	2010.7.28–2010.10.16
GPP	8 days/500 m	MOD17A2H	2010.7.78–2010.10.16
LAI	4 days/500 m	MCD15A3H	2010.7.78–2010.10.16

"GA–BP neural network" section to train the GA–BP model.

5. Model testing and accuracy inspection. Extract the NDVI, GPP, LAI, latitude and longitude of each pixel in the vegetation index image of the predicted date. These data are input into the trained model, and finally, a spatially continuous 16 days/500 m resolution NMRI product is generated. Finally, the accuracy test and evaluation are carried out through GNSS stations that are not involved in modeling.

The inversion process is shown in Fig. 3.

Study area

PBO is currently the only observing network based on the GNSS-IR principle. This network is based on GNSS-IR technology, and the daily estimates of vegetation moisture content, soil moisture, and snow depth are widely used in GNSS-IR technology research. Figure 4 is a schematic diagram of GNSS stations deployed by PBO in the western United States (30° N–50° N, 95° W–125° W). GNSS-IR and MODIS products have significant differences in temporal and spatial resolution. Therefore, it is considered that if the distance between GNSS stations is large and the distribution of the stations is sparse, it may result in poor local modeling results. The terrain around the station, regional climate change, and other factors may affect the effectiveness of the model. Therefore, this paper considers the above three factors comprehensively and selects the blue area (39° N–44.5° N, 111° W–114° W) as the experimental area. The GNSS station distribution in this area is relatively uniform, the terrain is diverse, and climate change is significant, which can better train the model. The DEM map and Landsat image of this area are shown in Fig. 5.
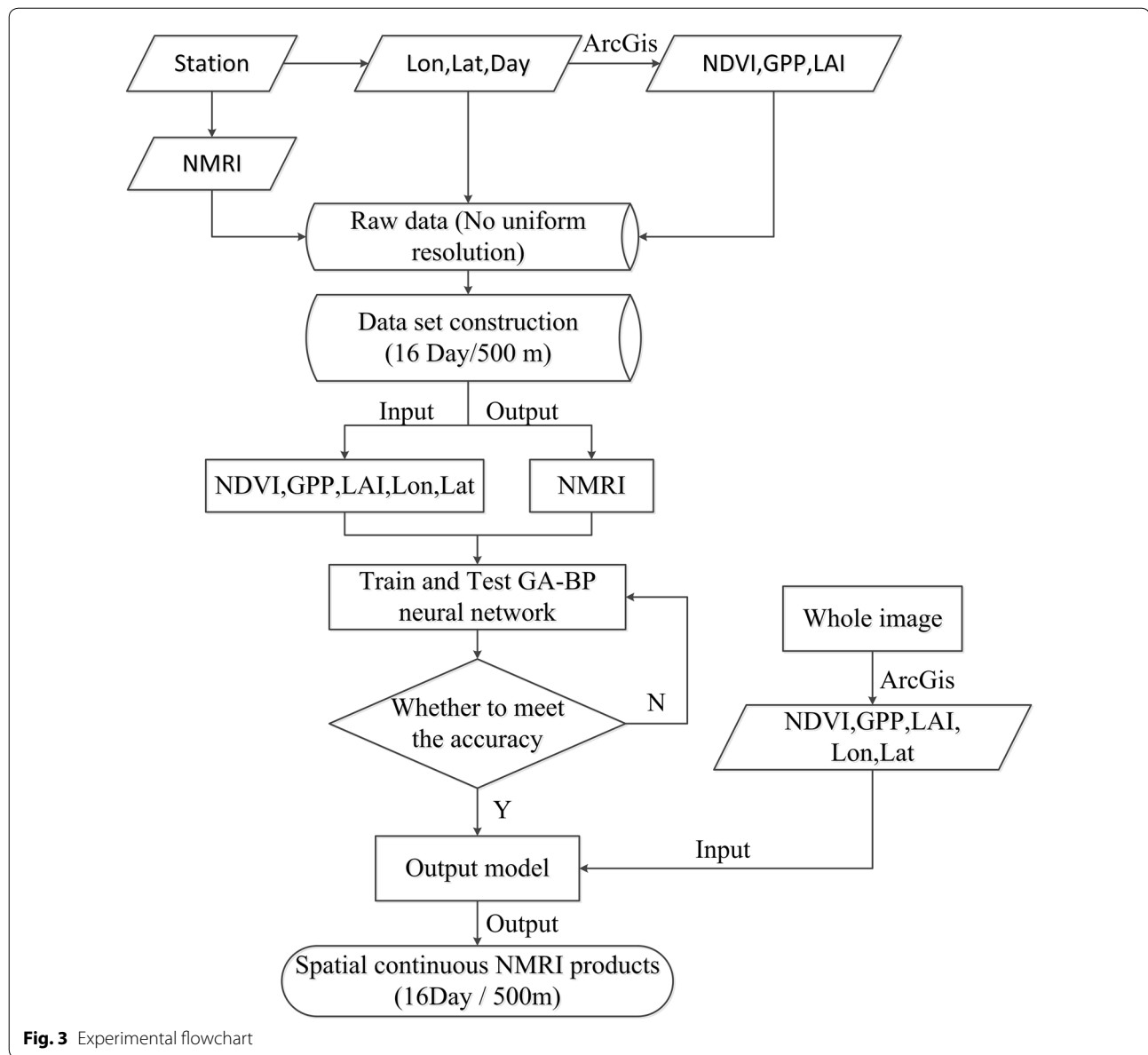
Fig. 3 Experimental flowchart

Figure 5b shows that there are 37 stations in this area, with different distribution locations and densities. Among them, 34 circles represent the stations involved in modeling. Three triangles (P112, P354, P359) represent the stations that were not included in the modeling, as they were used as later verification data. With reference to Figs. 4 and 5 and analysis of existing data, the upper half of the study area is southeast Idaho, and the lower half is northern Utah; Idaho and Utah both have a temperate continental climate. Because it is far from the ocean and blocked by the terrain, the humid air mass is difficult to reach, so it is dry and less rainy. It has the characteristics of hot and humid summers and cold and dry winters. Among them, the blue box

in Fig. 5b is the famous Great Salt Lake Desert in the United States, and the red box is the Great Salt Lake. The area contains more minerals, so the vegetation is thinner. Therefore, the VWC of the region selected in this paper will have significant differences in time and space to verify the accuracy of the inversion results later.

Results and analysis

In this paper, the correlation coefficients (r) of the NMRI, NDVI, GPP, and LAI vegetation indexes from 37 stations in the experimental area from July 28, 2010, to October 16, 2010, are calculated using the linear regression principle, and the results are shown in Fig. 6.
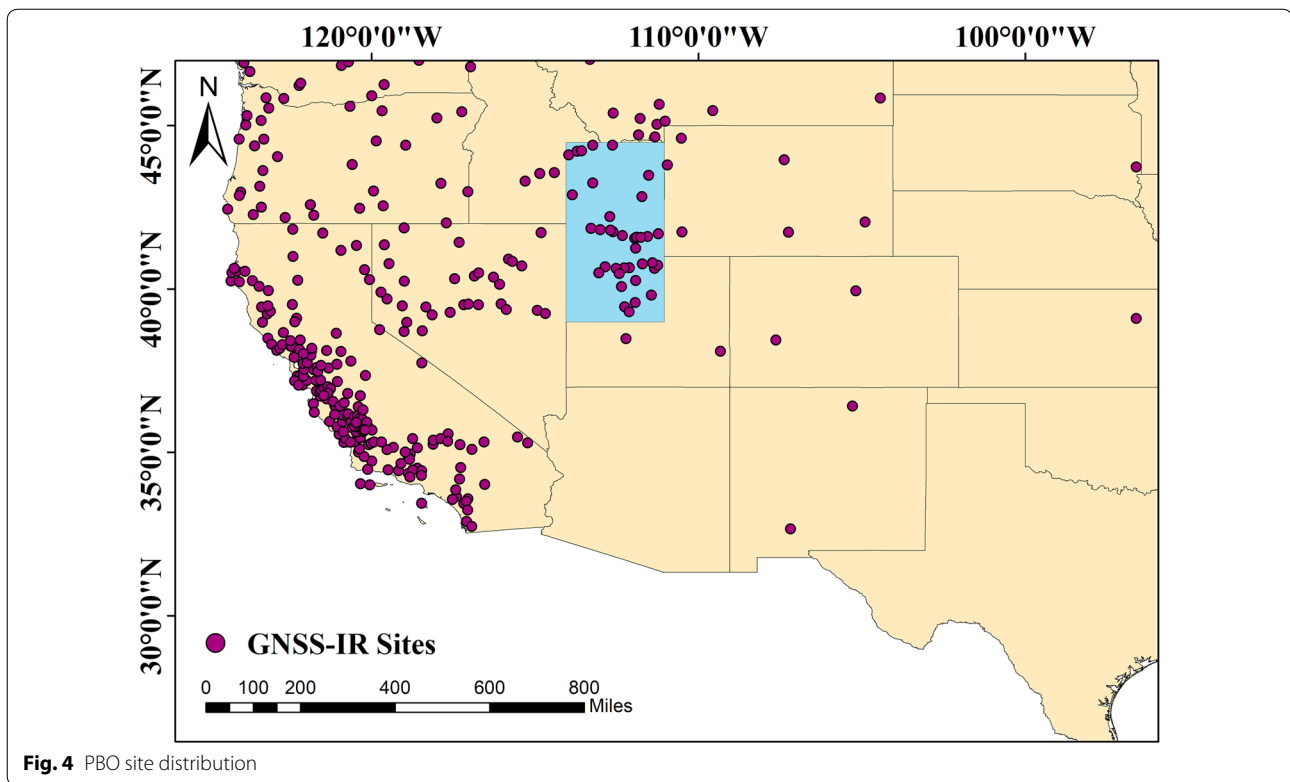
Fig. 4 PBO site distribution

It can be seen from Fig. 6 that there are 30 stations whose correlation coefficient between the NMRI and NDVI exceeded 0.6, accounting for 81% of the total stations. There are 29 stations whose correlation coefficient between the NMRI and GPP exceeded 0.6, accounting for 78% of the total stations. There are 25 stations with a correlation coefficient between the NMRI and LAI greater than 0.6, accounting for 68% of the total stations. The NMRI, NDVI, GPP, and LAI all have strong correlations, and the selected NDVI, GPP, and LAI can participate in modeling.

To better verify the modeling effect of the GA–BP model, this paper compares the GA–BP neural network, BP neural network, and multivariable linear regression model (MLR). First, determine the number of hidden neurons in the model according to the empirical formula $m = \sqrt{n + l} + a$. In the formula, m is the number of hidden layer nodes; n and l are the input layer and output layer nodes, respectively; and a is a positive integer between 1 and 10. Because the number of neurons in the input layer is five (i.e., the NDVI, GPP, LAI, Lon, and Lat) and the number of neurons in the output layer is one (i.e., the NMRI), the value range of the neurons in the hidden layer is m = [3, 14]. Subsequently, this paper uses a cross-validation method that is randomly run 500 times to select the best number of hidden layers. The results are shown in Table 2.

Table 2 shows that the nine hidden layer neurons appear most often. The average root mean square error (RMSE) value of the test at this time is relatively small. Subsequently, this paper randomly shuffles the dataset established in step 3 and divides it into training and test sets and inputs three models at the same time, running 50 times in a loop. Compare the modeling effects of the three models by testing whether the r and RMSE of the set are less than the threshold. The results are shown in Table 3.

From the table above, we can see that, after using the three models of BP, GA–BP, and MLR, the number of r values of the test set that are greater than 0.6 is 22, 23, and 13, respectively; the number of RMSEs in the test set that are less than the threshold is 35, 35, and 34, respectively. Based on the analysis of the numbers of r and RMSE that are less than the threshold, we can see that both the BP neural network and the GA–BP neural network are better than the MRL result, and the GA–BP neural network is slightly better than the BP neural network. To further compare the modeling effects of the three models, this paper uses K-fold cross-validation to further analyze the models. First, a part of the data is separated from the dataset established in step 3 as a test set, and the remaining data are divided into ten parts as a training set. Subsequently, the training model was repeated ten times, nine training datasets were

Pan *et al. Satell Navig*   (2020) 1:21

Page 8 of 15
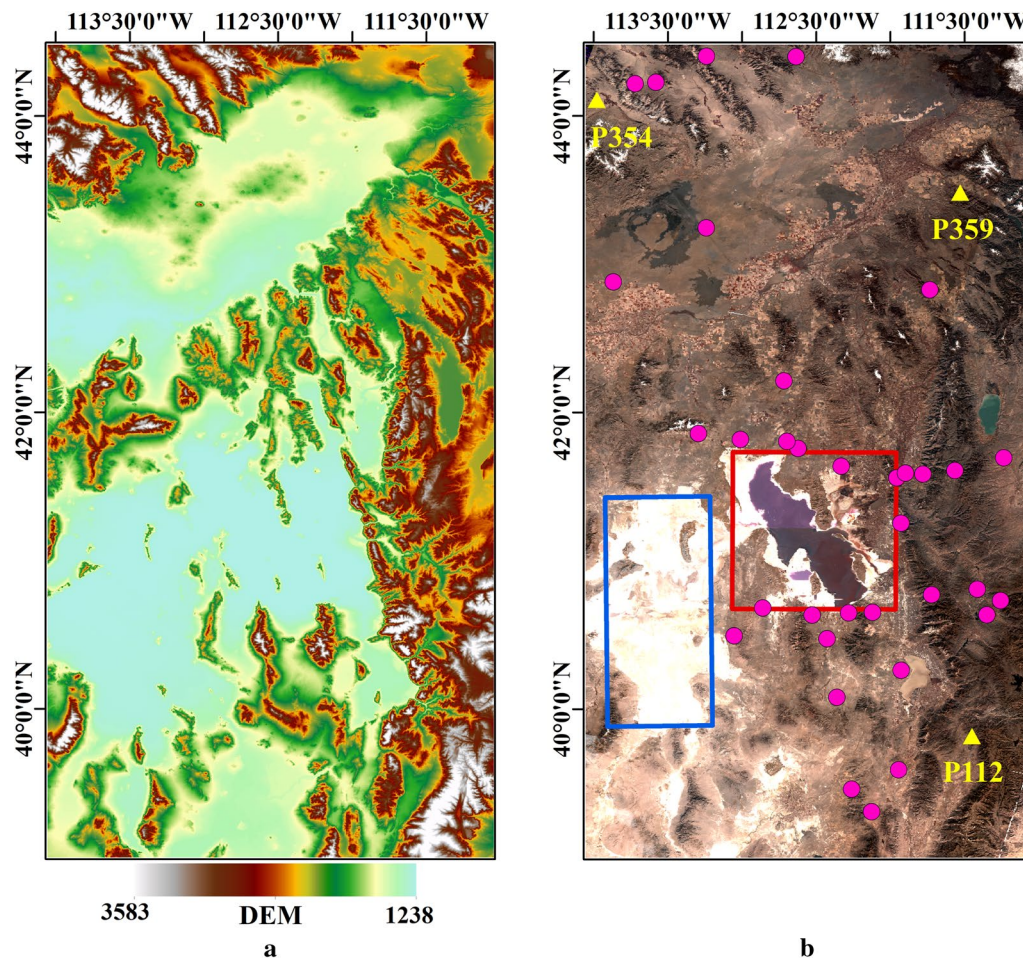
Fig. 5 DEM map and satellite map of the study area: **a** DEM map; **b** Landsat map

selected for training each time, and the remaining dataset was used for validation. The test data were used to test the model effect. Finally, the test results obtained by running the three models ten times were averaged. The results are shown in Fig. 7.

We can see from Fig. 7 that, among the three models, the GA−BP neural network has the best modeling effect, and the r and RMSE values of the estimated results are 77.8% and 0.0332, respectively. Following the BP neural network, the estimated r and RMSE values are 74.6% and 0.0465, respectively. The MLR model has the worst modeling effect, with r and RMSE estimated at 50% and 0.0516, respectively. This result is consistent with the results shown in Table 3. Based on the above results, both the BP neural network and the GA−BP neural network are relatively stable and better than the MLR results. However, GA−BP neural network modeling works better. Therefore, the GA−BP neural network adopted in this paper is used for modeling. In this paper, the established dataset is randomly divided into 70%, 15%, and

15%, which are used as the training, confirmation, and test sets, respectively. According to the algorithm flow of Fig. 2, the GA−BP model is trained, and the model with the best training effect is selected after multiple pieces of training. Figure 8 shows the change in the reciprocal of sum-squared errors of the genetic algorithm and the change in the fitness function. The model training accuracy is shown in Fig. 9.

It can be seen from Fig. 8 that when the number of iterations is less than 10, the sum-squared errors and fitness functions tend to be stable. The GA finds the gene chain with the smallest error energy as much as possible at the global level, reducing the possibility that the simple BP neural network is easy to oscillate and does not converge (or locally converge). It can be seen from Fig. 9 that the r values of the training and testing parts have reached 0.86 and 0.64, respectively, and the overall modeling accuracy has reached 0.79. The test part RMSE was calculated as 0.031. Therefore, the trained model is feasible and effective. Thus, using the model trained above to perform
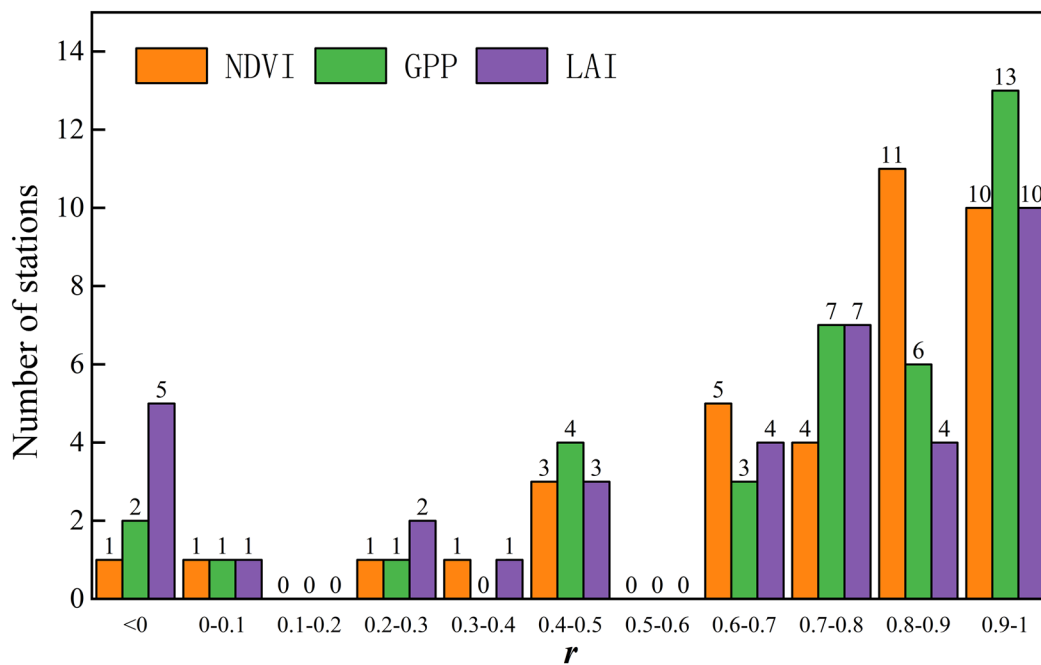
Fig. 6 Correlation analysis between NMRI, NDVI, GPP, and LAI
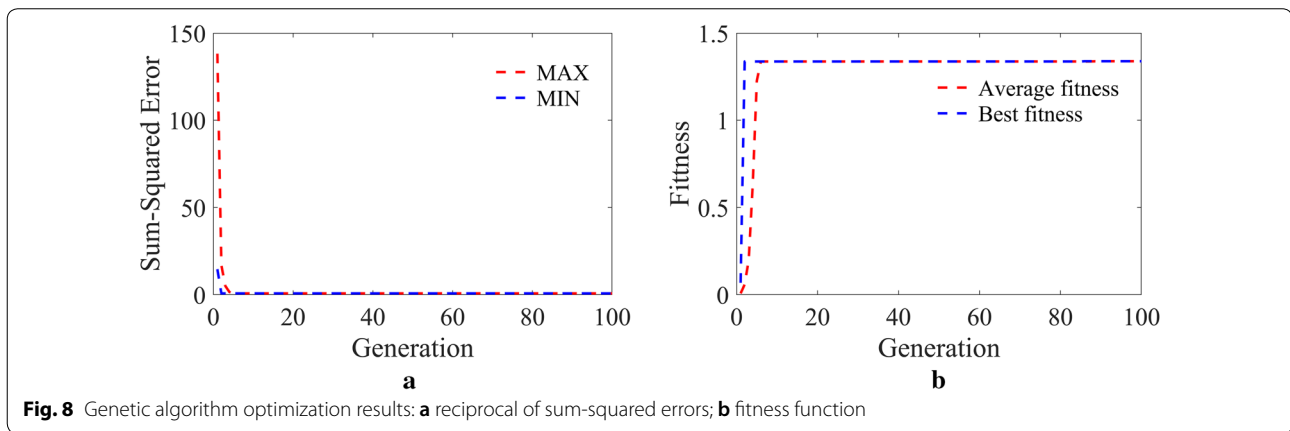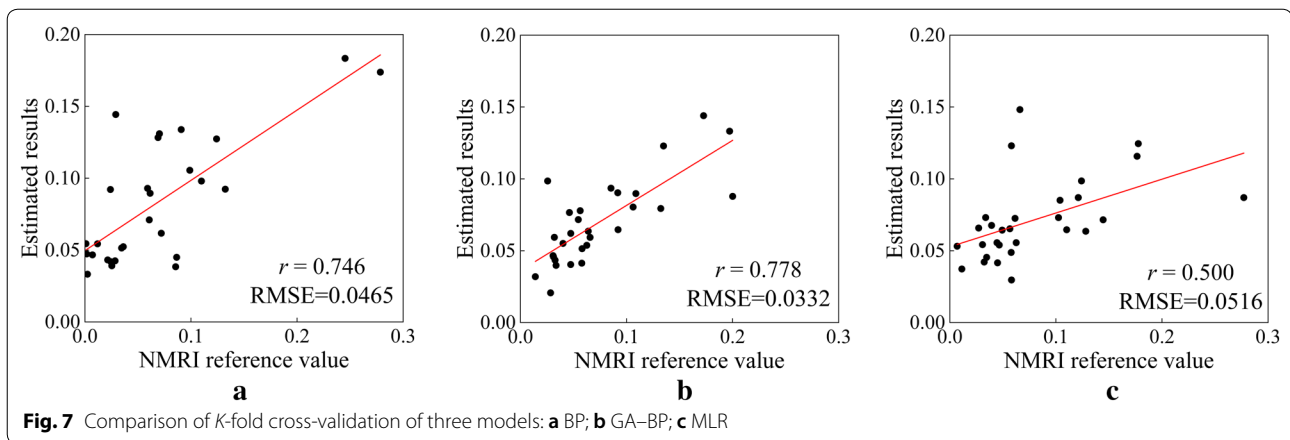
Table 2 Hidden layer number selection

Number of hidden layers	Frequency	RMSE average	Number of hidden layers	Frequency	RMSE average
3	33	0.0394	9	56	0.0400
4	27	0.0392	10	55	0.0404
5	28	0.0415	11	51	0.0397
6	38	0.0400	12	47	0.0412
7	39	0.0405	13	45	0.0400
8	43	0.0413	14	38	0.0390

Table 3 Comparison of running results of three models

Model	r		RMSE threshold (0.005)		
	< 0.6	> 0.6	Less than	Greater than	Less than a threshold percentage (%)
BP	28	22	35	15	70
GA–BP	27	23	35	15	70
MLR	37	13	34	16	68

inversion, six spatially continuous 16 days/500 m resolution NMRI product maps were obtained. Figure 10 shows six spatially continuous NMRI product maps. Figure 11 shows the NDVI, GPP, and LAI images on July 28 and October 16.

By comparing the NDVI, GPP, and LAI images at the same time, it was found that the spatial distribution of the NMRI in the test area was relatively consistent with the NDVI, GPP, and LAI. From the spatial distribution of the NMRI in Fig. 10, it can be seen that the NMRI values in the eastern and northwestern mountains of the experimental area are high, and the NMRI values in the Great Salt Lake and Great Salt Lake Desert areas in the southwest are low. This result is consistent with the geographical characteristics of the experimental area described in "Study area" section. From the time distribution, comparing Figs. 10 and 11, we found that as the weather became colder, the vegetation growth in most areas of the experimental area decreased significantly, and the water content of the vegetation decreased significantly. This change agreed with the climate characteristics of cold and dry winters in the experimental area. Through

Fig. 7 Comparison of *K*-fold cross-validation of three models: **a** BP; **b** GA–BP; **c** MLR

Fig. 8 Genetic algorithm optimization results: **a** reciprocal of sum-squared errors; **b** fitness function

the consistency of the NMRI spatial distribution with the NDVI, GPP, and LAI in Figs. 10 and 11, the feasibility of point-surface fusion was preliminarily verified. To further verify the accuracy of spatially continuous NMRI products, the NMRI values of three non-participating stations in the six spatially continuous NMRI images obtained by modeling were extracted and compared with the reference values of the NMRI provided by PBO H2O. This paper uses the three indicators of *r*, RMSE, and maximum inversion error (Max) for accuracy analysis. The results are shown in Fig. 12.

It can be seen from Fig. 12 that the *r* between the inversion value and the actual value of the three stations is higher than 0.87, which shows that there is a strong correlation between the inversion results and the actual value; for the RMSE, P359 is the largest among the three stations, reaching 0.049. The other two stations are less than 0.03; for MAX, P359 is the largest among the three stations, reaching 0.082, and the other two stations are less than 0.046. It can be seen that the errors obtained at the three stations are small, and there are no gross

errors; thus, the inversion results obtained are valid and accurate.

It can be seen from Figs. 10 and 11 that there are still some inconsistencies in the generated spatially continuous NMRI map. For example, in the southwestern part of the map, this region has a low resolution of vegetation moisture content, and the NMRI value is relatively close, which cannot accurately reflect the changes in vegetation moisture content in this region. In the images in Fig. 10e, f, it is obvious that there is an overestimation of the vegetation water content in this area. Combining Fig. 5b, we see there is no PBO H2O station in this area. Therefore, this may be because no GNSS stations from this area are included in modeling. Ultimately, the effect and accuracy of local modeling were affected, resulting in an abnormal NMRI value in this area. Therefore, there may be a large relationship between the fusion effect and the distribution of GNSS stations. For regions with GNSS distribution, it is feasible and effective to use GA–BP neural network to establish a point-plane fusion model to finally obtain a spatially continuous NMRI product.
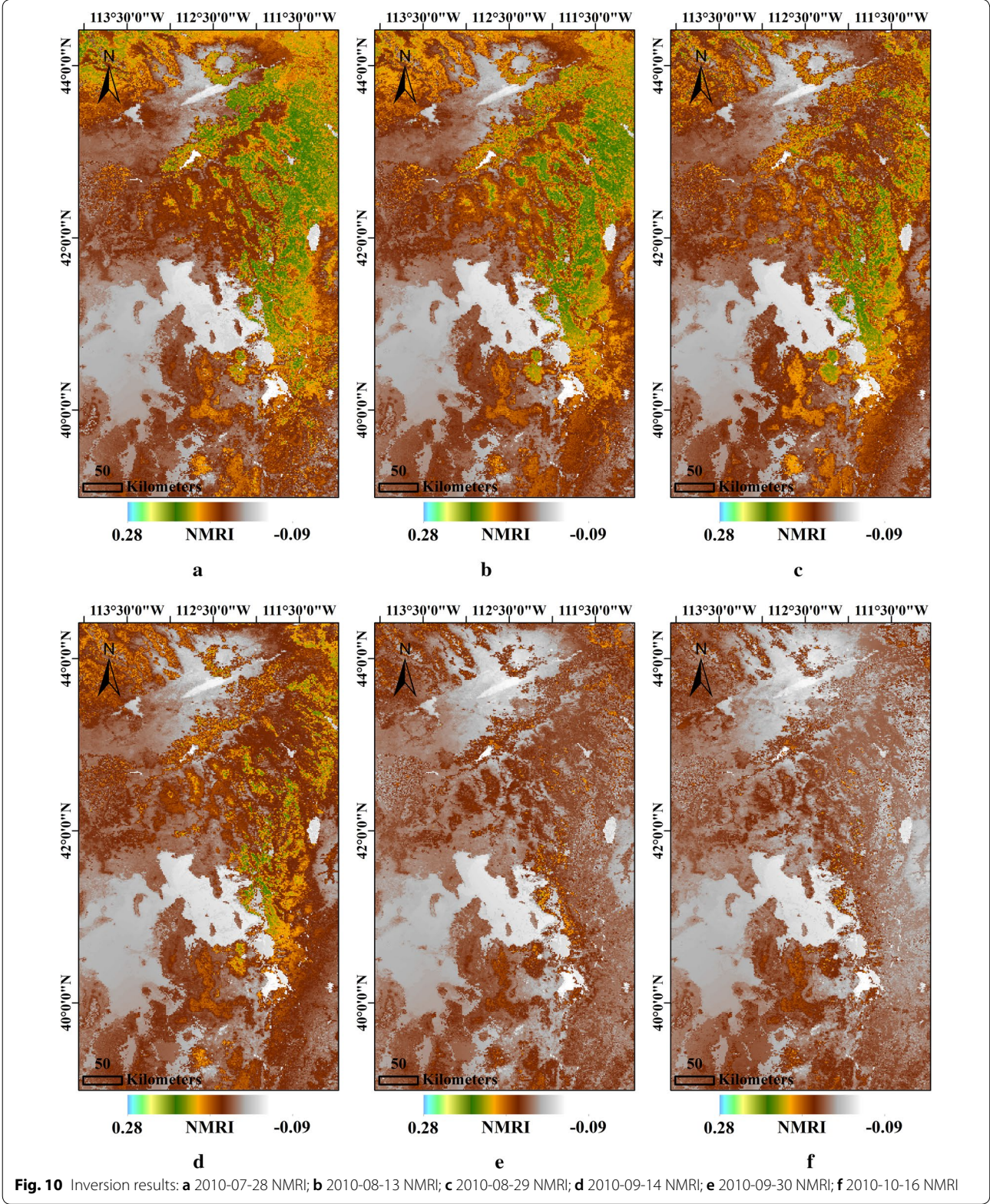
Fig. 9 Model training accuracy
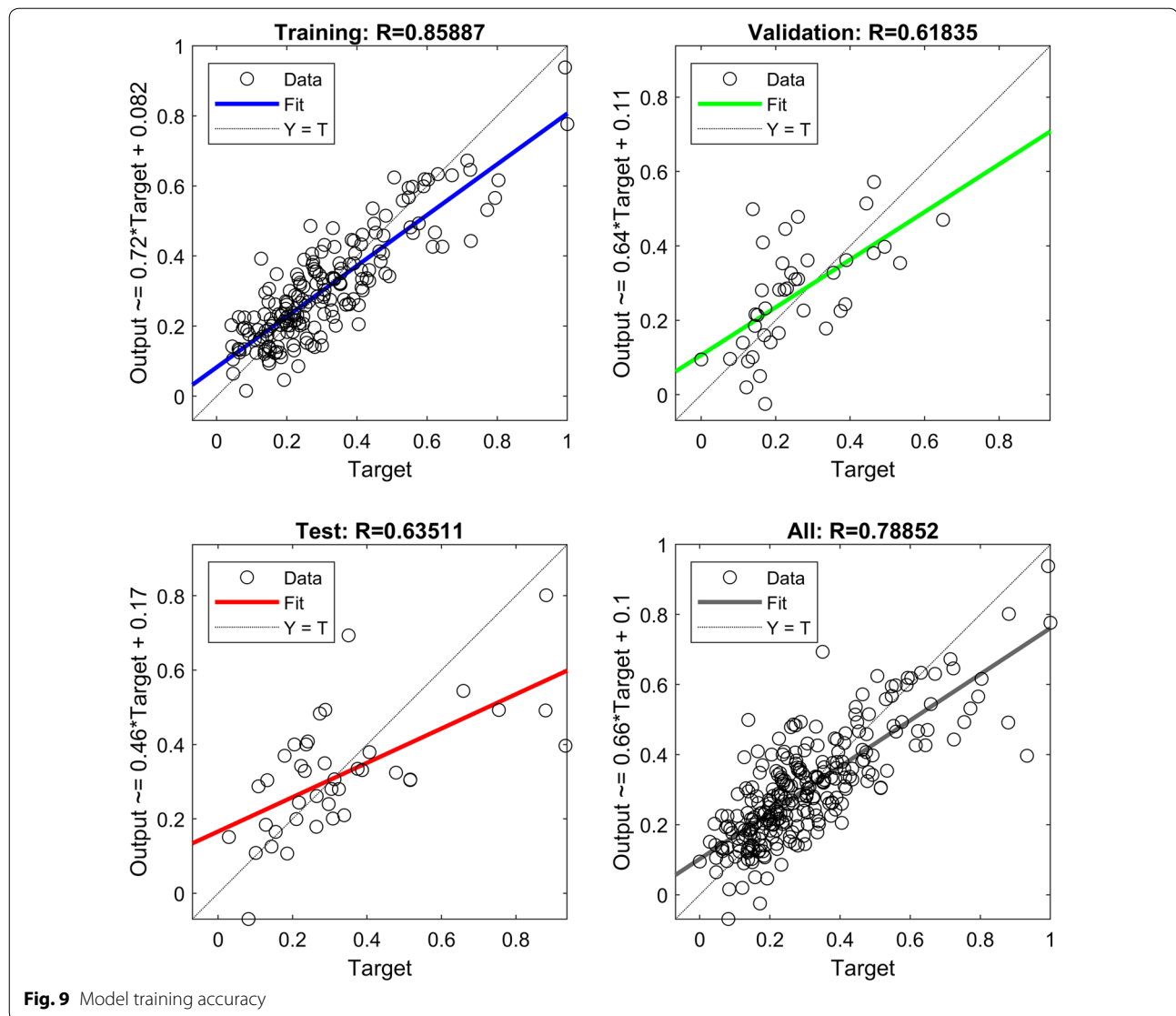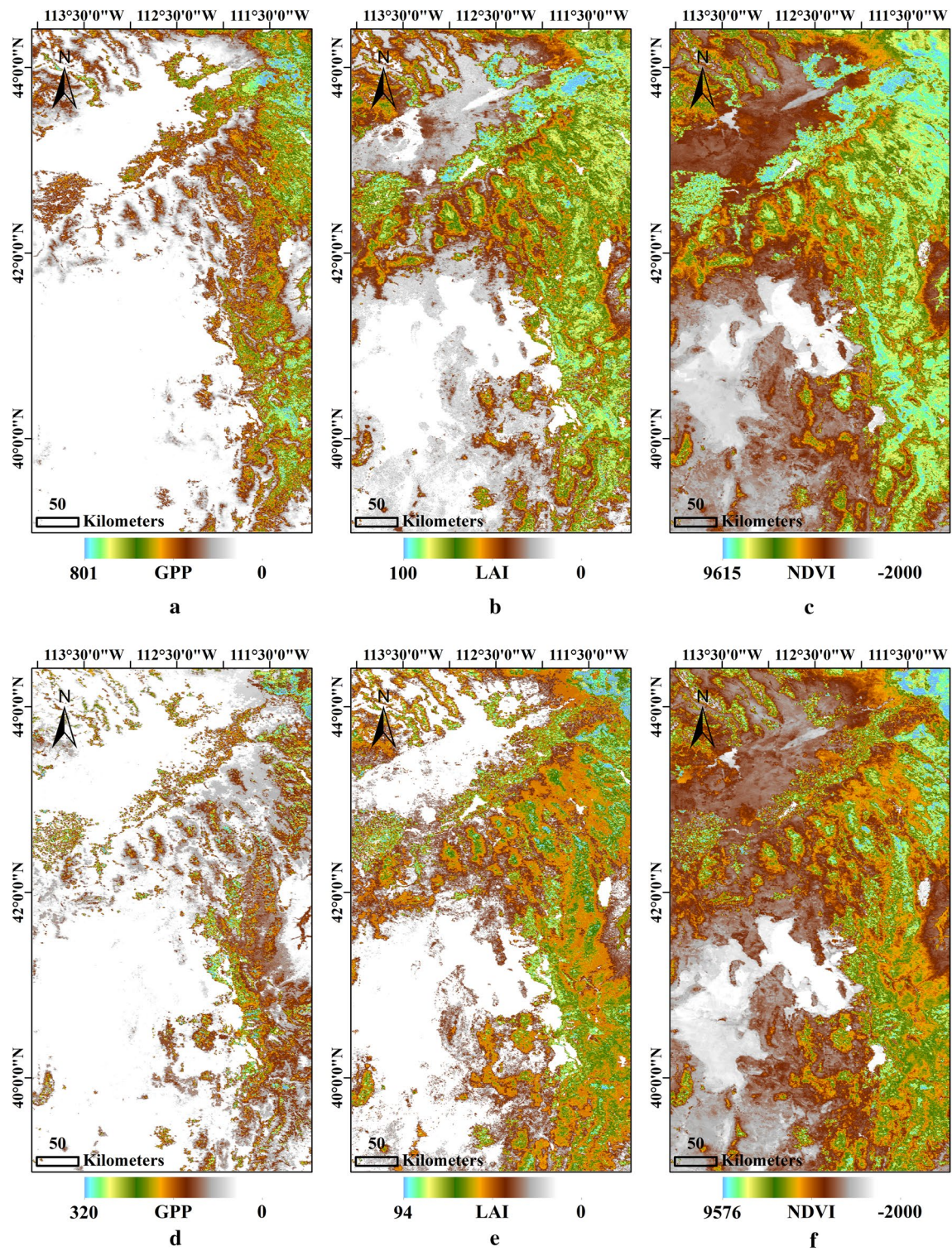
Conclusion and future research

The accurate and long-term monitoring of VWC is of great significance to environmental scientific research. Given the current limitations of using GNSS-IR technology and remote sensing technology to monitor VWC changes. Based on the idea of multisource data fusion, this paper proposes a method of point-surface fusion using neural networks to integrate MODIS three vegetation indexes: the NDVI, GPP, LAI, and NMRI products based on GNSS-IR. Theoretical analysis and experimental results show that (1) the NDVI, GPP, and LAI all have strong correlations with the NMRI. (2) We compared the effect of point-surface fusion of three models: the GA–BP neural network, BP neural network, and MLR. The results show that the GA–BP neural network has the best modeling effect, and

the *r* and RMSE values between the model estimation result and the reference value are 0.778 and 0.0332, respectively; it was followed by the BP neural network, with *r* and RMSE values of 0.746 and 0.0465, respectively; MLR had the poorest effect, with *r* and RMSE values of 0.500 and 0.0516, respectively. Therefore, the GA–BP neural network can be used to achieve a productive fusion of GNSS-IR and MODIS data. (3) The spatiotemporal variation in the 16 days/500 m resolution NMRI product obtained by point-surface fusion is consistent with that in the experimental area. The estimated results of the six spatially continuous NMRI products were extracted and compared with the reference values. The *r* values of both were higher than 0.87. The RMSE values were less than 0.049, and the MAX values were less than 0.082.

Fig. 10 Inversion results: **a** 2010-07-28 NMRI; **b** 2010-08-13 NMRI; **c** 2010-08-29 NMRI; **d** 2010-09-14 NMRI; **e** 2010-09-30 NMRI; **f** 2010-10-16 NMRI
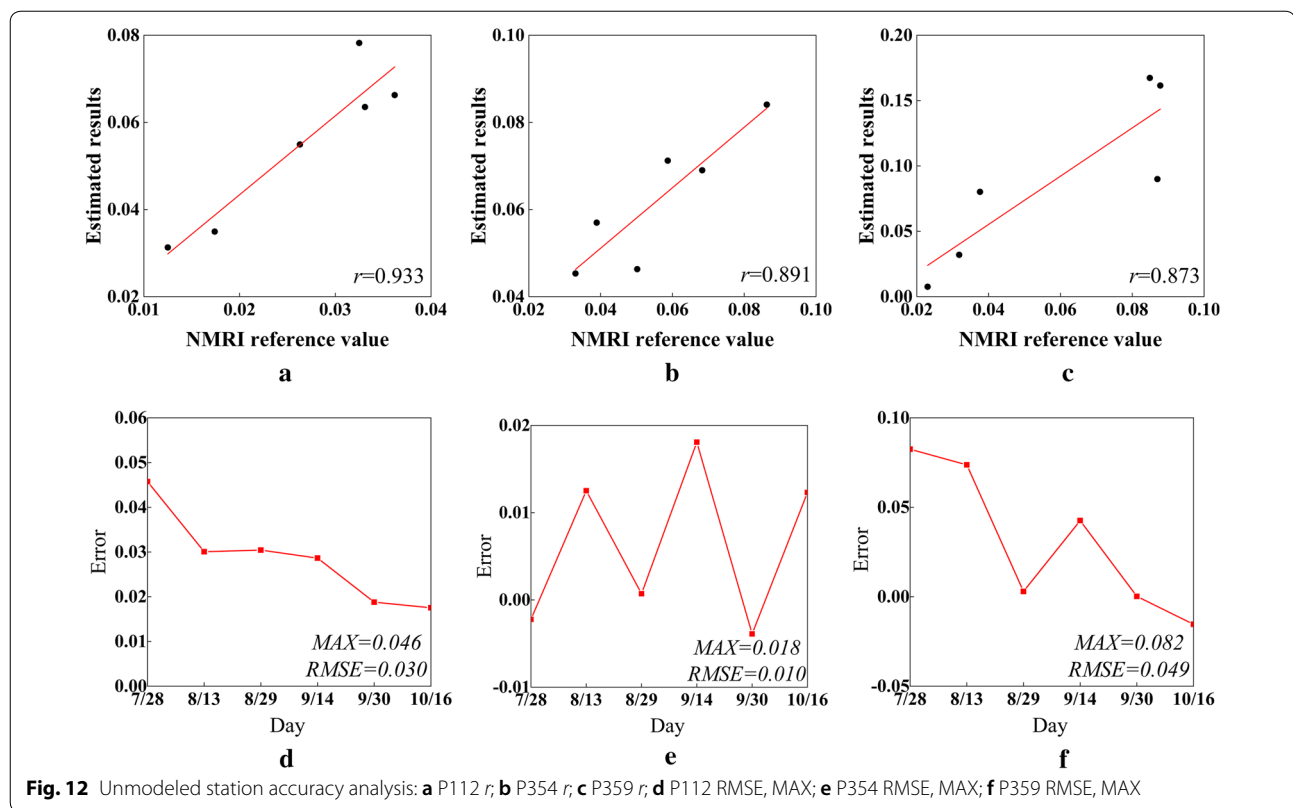
Fig. 11 Vegetation index products: **a** 2010-07-28 GPP; **b** 2010-07-28 LAI.; **c** 2010-07-28 NDVI; **d** 2010-10-16 GPP; **e** 2010-10-16 LAI; **f** 2010-10-16 NDVI

Fig. 12 Unmodeled station accuracy analysis: **a** P112 *r*; **b** P354 *r*; **c** P359 *r*; **d** P112 RMSE, MAX; **e** P354 RMSE, MAX; **f** P359 RMSE, MAX

In summary, this paper proposes that the GNSS-IR and MODIS point-surface fusion model based on the GA−BP neural network is feasible and effective, and the resulting spatially continuous NMRI products can be used to better, and more intuitively, represent the VWC changes in the region. It was found through experiments that the distribution distance between GNSS stations in the experimental area had a great impact on the modeling results. If the number of GNSS stations in the study area is small or the distance between the stations is considerable, it will affect the modeling effect of the local area. Therefore, to further verify the method of this paper, the impact of modeling after encrypting GNSS stations by the technique of spatial interpolation will be discussed in depth later. In addition, more VWC-related remote sensing products will be incorporated for modeling.

Authors' contributions
LY and RC proposed the idea; PY has carried on the thesis experiment part; ZZ and SY assisted in carrying out the simulation. All authors read and approved the final manuscript.

Availability of data and materials
In this experiment, the NMRI data came from the California Institute of Technology Jet Propulsion Laboratory GNSS H2O project (https://www.jpl.nasa.gov/). The three MODIS vegetation index products came from Google Earth Engine Data Catalog (https://developers.google.com/earth-engine/datasets/).

Competing interests
The authors declare that they have no competing interests.

Author details
[1] College of Geomatics and Geoinformation, Guilin University of Technology, Guilin 541006, China. [2] Research Center of Precise Engineering Surveying, Guangxi Key Laboratory of Spatial Information and Geomatics, Guilin 541006, China.

References
Alpaydin, E. (2004). *Introduction to machine learning (adaptive computation and machine learning)* (pp. 91–111). Berlin: The MIT Press.
Brakke, T. W., Kanemasu, E. T., Steiner, J. L., et al. (1981). Microwave radar response to canopy moisture, leaf-area index, and dry weight of wheat, corn, and sorghum. *Remote Sensing of Environment, 11,* 207–220.

Pan *et al. Satell Navig*   (2020) 1:21

Page 15 of 15

Chen, Q., Won, D., Akos, D. M., et al. (2016). Vegetation sensing using GPS interferometric reflectometry: Experimental results with a horizontally polarized antenna. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9*(10), 4771–4780.

Chew, C. C., Small, E. E., Larson, K. M., et al. (2013). Effects of near-surface soil moisture on GPS SNR data: Development of a retrieval algorithm for soil moisture. *IEEE Transactions on Geoscience and Remote Sensing, 52*(1), 537–543.

Chew, C. C., Small, E. E., Larson, K. M., et al. (2014). Vegetation sensing using GPS-interferometric reflectometry: Theoretical effects of canopy parameters on signal-to-noise ratio data. *IEEE Transactions on Geoscience and Remote Sensing, 53*(5), 2755–2764.

Dasgupta, S., & Qu, J. J. (2006). Combining MODIS-and AMSR-E-based vegetation moisture retrievals for improved fire risk monitoring. In *Remote sensing and modeling of ecosystems for sustainability III* (Vol. 6298, p. 62981B). New York: International Society for Optics and Photonics.

Elvidge, C. D. (1990). Visible and near infrared reflectance characteristics of dry plant materials. *Remote Sensing, 11*(10), 1775–1795.

Estey, L. H., & Meertens, C. M. (1999). TEQC: The multi-purpose toolkit for GPS/GLONASS data. *GPS Solutions, 3*(1), 42–49.

Goldberg, D. E. (1989). Genetic algorithms in search, optimization and machine learning. *NN Schraudolph and Journal, 3*(1), 301–302.

Gutman, G., & Ignatov, A. (1998). The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models. *International Journal of Remote Sensing, 19*(8), 1533–1543.

Hall, C.D., & Cordey, R. A. (1988). Multistatic Scatterometry. *International Geoscience and Remote Sensing Symposium, Remote Sensing: Moving Toward the 21st Century, 1*, 561–562.

Holland, J. H. (1975). Adaptation in natural and artificial systems. *Ann Arbor, 6*(2), 126–137.

Jin, S. G., Qian, X. D., & Wu, X. (2017a). Sea level change from BeiDou Navigation Satellite System-Reflectometry (BDS-R): First results and evaluation. *Global and Planetary Change, 149,* 20–25.

Jin, S. G., Zhang, Q. G., & Qian, X. D. (2017b). New progress and application prospects of global navigation satellite system reflectometry (GNSS+R). *Acta Geodaetica et Cartographica Sinica, 46*(10), 1389–1398.

Larson, K. M. (2016). GPS interferometric reflectometry: Applications to surface soil moisture, snow depth, and vegetation water content in the western United States. *Wiley Interdisciplinary Reviews: Water, 3,* 775–787.

Larson, K. M., Braun, J. J., Small, E. E., et al. (2010). GPS multipath and its relation to near-surface soil moisture content. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 3*(1), 91–99.

Larson, K. M., Gutmann, E. D., Zavorotny, V. U., et al. (2009). Can we measure snow depth with GPS receivers? *Geophysical Research Letters*. https://doi.org/10.1029/2009GL039430.

Larson, K. M., Lofgren, J. S., & Haas, R. (2013). Coastal sea level measurements using a single geodetic GPS receiver. *Advances in Space Research, 51*(8), 1301–1310.

Larson, K. M., & Nievinski, F. G. (2013). GPS snow sensing: Results from the earth scope plate boundary observatory. *GPS Solutions, 17*(1), 41–52.

Larson, K. M., & Small, E. E. (2014). Normalized microwave reflection index: A vegetation measurement derived from GPS networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7*(5), 1501–1511.

Larson, K. M., Small, E. E., Gutmann, E., et al. (2008a). Using GPS multipath to measure soil moisture fluctuations: Initial results. *GPS Solutions, 12*(3), 173–177.

Larson, K. M., Small, E. E., Gutmann, E. D., et al. (2008b). Use of GPS receivers as a soil moisture network for water cycle studies. *Geophysical Research Letters, 35*(24), 581–599.

Li, T. W., Shen, H. F., Yuan, Q. Q., et al. (2017). Estimating ground-level $PM_{2.5}$ by fusing satellite and station observations: A geo-intelligent deep learning approach. *Geophysical Research Letters, 44*(23), 11–985.

Liang, Y. J., Ren, C., Wang, H. Y., et al. (2019). Research on soil moisture inversion method based on GA–BP neural network model. *International Journal of Remote Sensing, 40*(5–6), 2087–2103.

Liu, J. N., Shao, L. J., & Zhang, X. X. (2007). Advances in GNSS-R studies and key technologies. *Geomatics and Information Science of Wuhan University, 32*(11), 955–960.

Martin-Neira, M. (1993). A passive reflectometry and interferometry system (PARIS): Application to ocean altimetry. *ESA Journal, 17*(4), 331–355.

Mcclland, J. L., Rumelhart, D. E., & PDP Research Group. (1986). *Parallel distributed processing, volume 1: Explorations in the microstructure of cognition: Foundations* (pp. 282–317). Berlin: The MIT Press.

Owe, M., De Jeu, R., & Walker, J. (2001). A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index. *IEEE Transactions on Geoscience and Remote Sensing, 39*(8), 1643–1654.

Paruelo, J. M., Epstein, H. E., Lauenroth, W. K., et al. (1997). ANPP estimates from NDVI for the central grassland region of the United States. *Ecology, 78*(3), 953–958.

Peñuelas, J., Filella, I., Biel, C., et al. (1993). The reflectance at the 950–970 nm region as an indicator of plant water status. *International Journal of Remote Sensing, 14*(10), 1887–1905.

Ren, C., Liang, Y. J., Lu, X. J., et al. (2019). Research on the soil moisture sliding estimation method using the LS-SVM based on multi-satellite fusion. *International Journal of Remote Sensing, 40*(5–6), 2104–2119.

Rummelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. *Nature, 323*(2), 318–362.

Small, E. E., Larson, K. M., & Braun, J. J. (2010). Sensing vegetation growth with reflected GPS signals. *Geophysical Research Letters, 37*(12), L12401.

Small, E. E., Larson, K. M., & Smith, W. K. (2014). Normalized microwave reflection index: Validation of vegetation water content estimates from Montana grasslands. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7*(5), 1512–1521.

Srinivas, N., & Deb, K. (1994). Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation, 2*(3), 221–248.

Tucker, C. J., Pinzon, J. E., Brown, M. E., et al. (2005). An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *International Journal of Remote Sensing, 26*(20), 4485–4498.

Wylie, B. K., Meyer, D. J., Tieszen, L. L., et al. (2002). Satellite mapping of surface biophysical parameters at the biome scale over the North American grasslands: A case study. *Remote Sensing of Environment, 79*(2–3), 266–278.

Xu, H. Z., Yuan, Q. Q., Li, T., et al. (2018). Quality improvement of satellite soil moisture products by fusing with in-situ measurements and GNSS-R estimates in the Western continental US. *Remote Sensing, 10*(9), 1351.

Yuan, Q. Q., Xu, H. Z., Li, T. W., et al. (2020). Estimating surface soil moisture from satellite observations using a generalized regression neural network trained on sparse ground-based measurements in the continental US. *Journal of Hydrology, 580,* 124351.

Zhang, S. C., Dai, K. N., Nan, Y., et al. (2018). Preliminary research on GNSS-MR for Snow depth. *Geomatics and Information Science of Wuhan University, 43*(2), 11.

Zhang, S. C., Wang, X., & Zhang, Q. (2017). Avoiding errors attributable to topography in GPS-IR snow depth retrievals. *Advances in Space Research, 59*(6), 1663–1669.

Publisher's Note