Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Improved estimation of terrestrial water loading deformation in GNSS vertical time series based on a hybrid machine learning method: A case study in Amazon River basin

Keke Xu
Beijing University of Civil Engineering and Architecture
Yifu LIU
Henan Polytechnic University
Xianglei Liu
125896565@qq.com

Beijing University of Civil Engineering and Architecture
Yongzhen Zhu
Henan Polytechnic University

Article

Keywords: TWLD, GNSS observations, TWS, GRACE, Deep learning

Posted Date: March 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8868371/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Improved estimation of terrestrial water loading deformation in GNSS vertical time series based on a hybrid machine learning method: A case study in Amazon River basin

Keke Xu[1,2*], Yifu Liu[2] , Xianglei Liu[1,*], and Yongzhen Zhu[2]

[1]*School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing, 102616, China*

[2]*School of Surveying and Land Information Engineering of Henan Polytechnic University, Jiaozuo, 454000, China.*

*Corresponding Author. *E-mail address*: xkk@hpu.edu.cn (K. Xu)

Abstract: Accurate estimation of terrestrial water load deformation (TWLD) from GNSS observations is of great significance to terrestrial water storage (TWS) inversion and tectonic deformation detection. However, it is difficult to separate the signals effectively from the complex and noisy GNSS observation series due to a variety of events and noises mixed together. We here present a hybrid estimation and prediction method for TWLD combining multi-scales decomposition and deep learning algorithm. Non-tidal atmospheric load and non-tidal ocean load are firstly corrected from the original GNSS time series. The residuals are then decomposed into main intrinsic mode components by variational mode decomposition algorithm, where grey wolf optimization and sample entropy is adopt to search for the optimal model parameters. Finally, every component is used as input to extract the data features suitable for BiLSTM model. We apply the method to estimate TWLD of the Amazon River Basin based on 28 GNSS sites. The results show that the hybrid model greatly enhance feature distinguishability and prediction accuracy of GNSS time series, with an average improvement of 73.20%, 26.87%, 73.40% and 51.14%, 7.83%, 59.55% in RMSE, R^2 and MAPE, compared to the single model of LSTM and BiLSTM. The TWS derived from GNSS agrees well with that of Gravity Recovery and Climate Experiment (GRACE) and Global Land Data Assimilation System (GLDAS), with the correlations of over 0.71 and 0.81, respectively. The joint TWS integrating GNSS, GRACE and GLDAS has a better consistency than single-source data, with the correlations of above 0.91. We find that the larger changes in the middle and lower reaches of the northeastern Amazon Basin, gradually tapering off towards the upstream areas near the Andes Mountains in the southwest. The significant variations area of water volume corresponds to the estuary and the downstream section of the Amazon River Basin. Moving southward from the aforementioned region, the magnitude of water storage changes gradually becomes more moderate. The research provides an important reference for the precise monitoring and efficient supervision of regional water resource.

Keywords：TWLD; GNSS observations;TWS; GRACE; Deep learning

1. Introduction

Crustal subsidence and uplift are the major causes of changes in the elevation of the earth's surface. The primary driving factors of these changes include atmospheric loading, terrestrial water storage (TWS), oceanic load disturbances, and human activities(Mimi *et al.*, 2024). Especially, TWS is the dominant influence on the vertical deformation(Tang *et al.*, 2024). This is referred to as terrestrial water load deformation (TWLD). TWLD is earth surface deformations caused by changes in the mass of terrestrial water bodies, including groundwater, lakes, rivers, and snow/ice cover(Li *et al.*, 2024). These deformations can be calculated through the vertical displacement of the Earth's surface, typically monitored and calculated using geodetic technologies such as the Global Navigation Satellite System (GNSS), Interferometric Synthetic Aperture Radar (InSAR) or Gravity Recovery and Climate Experiment (GRACE)(Woods *et al.*, 2024). Especially, GNSS can provide a difference from conventional hydrological or meteorological observation technology and have gradually become an important means and research approach for global water quality balance. Compared with the low spatiotemporal resolution of GRACE, GNSS has a temporal resolution of day and dense Continuous Operational Reference System (CORS) net. Under sufficient station coverage, GNSS data can be used to observe TWS changes with a resolution of tens of kilometers in a large area, and the significant water quality fluctuations can be tracked in time and the gap period of GRACE observation is filled. Although GNSS has a unique advantage in the study of TWS variations in local areas, which takes into account the advantages of high spatial and temporal resolution, GNSS time series is extremely complex and contain multiply sources information, such as non-tectonic, tectonic factors, quasi-periodicity, inter-annual variation signals and noises, characterized by randomness, non-periodicity, and instability, resulting in a serious impacts on accurate estimation of TWLD and dynamic detection of tectonic deformation(Métivier *et al.*, 2014). Therefore, TWLD model reconstruction is vital for applications in fields such as hydrology, geophysics and geodesy. It can better capture deformation characteristics caused by water mass changes, further obtain the accuracy TWS variations or detect the transient deformation by removing the influence of non-tectonic factors. Classic modeling methods usually regard seasonal signals as constant amplitude (Bevis and Brown, 2014, Wu *et al.*, 2015), which can no longer satisfy the non-stationary behavior of TWLD.

Recently, a number of methods have been proposed to detect the quasi-periodic variability of TWLD in geodetic time series. It can be divided into the following three categories.

(1) Mathematics methods, such as Least-Squares Wavelet Analysis (LSWA) (Ghaderpour and Pagiatakis, 2019), Least-Squares Spectral Analysis (LSSA) (Ghaderpour and Ghaderpour, 2020), Kalman filter(Didova *et al.*, 2016), wavelet decomposition (Bogusz and Janusz, 2015, Ghaderpour and Pagiatakis, 2019) and semi-parametric model-based Chebyshev polynomials (Bennett, 2008) and Wiener filter-based approaches(Klos *et al.*, 2019). Empirical Orthogonal Function (EOF)/Principal Component Analysis (PCA) (Dong *et al.*, 2006, Shen *et al.*, 2014)and Independent Component Analysis (ICA) (Hyvärinen and Oja, 1997), multichannel singular spectrum analysis (MSSA)(Chen *et al.*, 2013, Kondrashov and Ghil, 2006), Singular Value Thresholding (SVT)(Bao *et al.*, 2021), Kriged Kalman Filter(Liu *et al.*, 2017). These methods mainly emphasize time-frequency analyses and filtering. As a result, the filtered residuals may

still contain an artificial signal driven by colored noise and un- or mis- modeled signals. Little attention is paid to the separation of the TWLD signal in GNSS time series, and the interpretation of the estimated signals and how the frequencies and amplitudes of components change over time.

(2) Hydrological model. The Global Land Data Assimilation System (GLDAS) is a most widely used hydrological modeling system, jointly developed by NASA in the United States and JAXA in Japan. This system integrates data from multiple satellite remote sensing sources, ground observations, and meteorological models through land surface modeling and data assimilation techniques to provide a global dataset of land surface (including soil and vegetation) hydrological variables. Currently, the GLDAS system includes four traditional land surface models (LSM): Mosaic, CLM, Noah, and VIC. In this study, the changes in soil moisture content (sum of data from 4 soil layers), total canopy water storage changes, and snow water equivalent changes provided by the GLDAS-2.1 The TWS is usually estimated by the GLDAS model based on the accumulation of snow cover, tree crown water, and soil water at a depth of 0~2 m in the model. In fact, hydrological models or land surface model data are derived from satellite remote sensing and meteorological reanalysis data, subject to inconsistency of the final total amount because of the difference in output components of various products. In addition, most hydrological model products are barely calibrated resulting in a large uncertainty, and no human activity contributions are taken into account.

(3) Machine leaning methods. With rapid development of big data, more scholars have conducted extensive research on Model reconstruction and prediction based on machine learning, such as convolutional neural network (CNN), recurrent neural network (RNN), long short-term memory (LSTM)(Jing *et al.*, 2021, Junru *et al.*, 2021). Compared with traditional RNNs, LSTM solves the vanishing and exploding gradient problems through its gated mechanism and can better capture long-term dependencies(Ting *et al.*, 2023). However, unidirectional LSTM cannot encode information from future to past and only captures forward relationships within the observation series. In contrast, BiLSTM can extract features from both directions forward and backward enhancing the model's ability to capture local and global features in the observations series(Zhaoyi *et al.*, 2023). Simultaneously, Bi-directional LSTM (BiLSTM) considers future information, making the model more sensitive to patterns and dependencies within the observation series. Some researchers have adopted BiLSTM combined with an attention mechanism to more accurately predict changes in terrestrial water storage in typical regions(Wanqiu *et al.*, 2024).

Although the above methods have made significant progress in identification of TWLD and inversion of TWS, it has high computational cost due to the complex high-dimension of GNSS position time series. In addition, reconstruction and prediction model are susceptible from the complex nonlinear characteristics of the GNSS observations, especially for prediction of the interannual variation and modeling of the quasi-periodic signal. We here propose a hybrid machine learning method combining dimensionality reduction and deep learning, aiming to deal with the complex signals and noise effects and identifying TWLD signals. The structure of the paper is organized as follows: Section 2 elaborates on the methodology for improved

estimation of TWLD in GNSS vertical position time series. Section 3 presents the research results and validates the effectiveness of the proposed method by application of in Amazon River basin. In Section 4, we discuss the feasibility, practicality and reliability of the new method. Finally, conclusions are summarized in Section 5.

2. Methodology

In view of the spatially heterogeneous distribution and unsteady change of TWS signal, we adopt the multi-scales solution with Variational Mode Decomposition (VMD) and deep learning method with BiLSTM for estimating and predicting TWLD by decomposition and reconstruction of multi-component and multi-frequency/wavelength. The observations series are firstly decomposed into a series of spatiotemporal components with different resolutions from high to low frequency, which represent the fluctuation term, short period term and long trend term, respectively. And then every IMF component is used as input to extract the data features suitable for BiLSTM model.

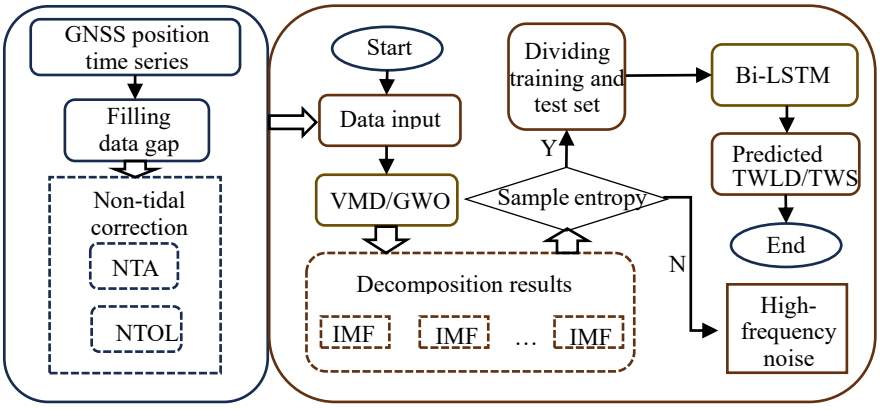


Fig.1 Technique flowchart for extracting TWLD from GNSS Data

2.1 Data reprocessing

For GNSS time series, the East and North components are primarily influenced by geological structural changes and are less affected by environmental loads, leading to lower sensitivity of GNSS time series horizontal components to non-geological signals(Lavoisiane *et al.*, 2021). Therefore, this paper mainly focuses on processing and analyzing the vertical GNSS time series. Specifically, seasonal variations in the vertical time series are mainly affected by hydrology, Non-Tidal Atmospheric Load (NTAL), Non-Tidal Ocean Load (NTOL) and noises, with minimal impact from trend components. The values of NTAL and NTOL for GNSS station correction are obtained from the Earth System Modeling Group at German Research Center for Geosciences (ESMGFZ). Using a sliding window approach with a window size of 8, the average values are calculated every 8 data points. NTAL effect on the vertical direction is about 5 mm, and the NTOL effect is about 2-3 mm. The modes decomposition is then adopted for the residuals. Effective signals synthesis is obtained from intrinsic mode function (IMF)

components by the sample entropy(SE) criteria. The missing data in GNSS station coordinate time series are restored by inter-station spatial correlations within the regional network based on Kriging interpolation with Kalman filtering (Liu *et al.*, 2017). The values exceeding triple the standard error are considered as outliers, and the time series with missing data for a year or more are removed. Linear fitting is employed to extract and remove the trend components.

2.2 Multi-scales decomposition

Given that the time series signal is decomposed into k IMF components, a constrained variational model is established as (Liu *et al.*, 2024, Li *et al.*, 2025):

$$\min_{\{u_k\},\{\omega_k\}}\left\{\sum_k\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-j\omega_k t}\right\|_2^2\right\} \quad \text{s.t.} \sum_k u_k = f \tag{1}$$

where k represents the number of intrinsic mode components, $\mathrm{u_k}$ is the set of all modes, ω_k is the actual center frequency of each mode, ∂_t represents the gradient operation, $\delta(t)$ represents the shock function, $j^2=-1$, t is the time, $e^{-j\omega_k t}$ is the estimated center frequency of each analytic signal, $\|\cdot\|$ represents finding the L_2 norm, s.t. representing constraints, $\sum$ is the sum of all model components.

To ensure the good convergence of the quadratic penalty under finite weights and the Lagrange multiplier's strict implementation of constraints for the optimal solution(Xi *et al.*, 2023), augmented Lagrange is introduced as(Liu *et al.*, 2024, Li *et al.*, 2025):

$$L(\{u_k\},\{\omega_k\},\lambda)=\alpha\sum_k\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-j\omega_k t}\right\|_2^2+\left\|f(t)-\sum_k u_k(t)\right\|_2^2+\left\langle\lambda(t),f(t)-\sum_k u_k(t)\right\rangle \tag{2}$$

where α is the quadratic penalty factor; $\lambda(t)$ is the Lagrange multiplier, $\langle*\rangle$ is the inner product operator. The original minimization problem is changed into the present one by using alternating direction multiplier algorithm to find the saddle point by iterative updating.

VMD cannot decompose time series adaptively, so two parameters k and α *are* required to set. Grey Wolf Optimization (GWO) is here adopt to search for the two optimal parameters. By simulating the behavior within a grey wolf pack, including the leading grey wolf γ, the deputy leading grey wolf β, and the ordinary grey wolves δ, the positions of grey wolves are continuously updated until a certain number of iterations is reached or a convergence condition is met by the equation:

$$\begin{cases} x_{\text{new}}^{\gamma} = x_{\text{old}}^{\gamma} - A\cdot d^{\gamma} \\ x_{\text{new}}^{\beta} = x_{\text{old}}^{\beta} - A\cdot d^{\beta} \\ x_{\text{new}}^{\delta} = x_{\text{old}}^{\delta} - A\cdot d^{\delta} \end{cases} \tag{3}$$

where x_{new} and x_{old} denotes the updated and the original position. A is a random number

typically within the range [0, 2]; $d^{\gamma}, d^{\beta}, d^{\delta}$ represents the distances from the current grey wolf to the leading grey wolf, deputy leading grey wolf, and ordinary grey wolves, respectively, with the details as follows,

(1) Initialize the GWO parameters: Set both the number of grey wolves and the maximum number of iterations to 10. Define the parameters k and α as individual grey wolves, with their respective value ranges set to [2, 12] and [1000, 3000].

(2) Calculate fitness: Decompose the time series using VMD under the positions of different grey wolf individuals. Determine the fitness of each position based on its permutation entropy and normalized mutual information entropy, and update the positions accordingly.

(3) Update grey wolf positions: This process simulates the cooperation and competition among grey wolves. The position update process is typically expressed as:

In the process of decomposition, the SE is adopt for describing the degree of randomness and irregularity of each component, further used to determine the information loss and uncertainty in the observations series. All low-frequency IMF components are retained according to SE values of less than 0.5, and synthesized into a filtered signals, effectively removing the noise from GNSS time series.(Dragomiretskiy and Zosso, 2014) . SE can be defined as,

$$\mathrm{SE} = \frac{-\ln\left[A(m,r)\right]}{-\ln\left[A(m+1,r)\right]} \tag{4}$$

where $A(m,r)$ represents the proportion of similar subsequences given the subsequence length m and the threshold r (typically set to 0.1 to 0.25 times the standard deviation of the subsequence).

2.3 Model reconstruction

A standard LSTM unit consists of a cell state and three main gates: the forget gate, the input gate, and the output gate(Chen *et al.*, 2023). The forget gate determines the extent to which past information is retained in the current state, the input gate controls the impact of new information on the current state, and the output gate regulates the extent to which the current state is output to the external environment. The calculations for these gates rely on activation functions (such as the sigmoid function) and operations like dot products. BiLSTM is a deep learning model commonly used in natural language processing and observations series modeling tasks. It extends the LSTM model by leveraging both past and future information for prediction(Seerangan *et al.*, 2024). The basic structure of BiLSTM is similar to that of LSTM. The information flow and update memory are controlled by the three main gates. However, unlike unidirectional LSTM, BiLSTM performs computations in both the forward and backward directions simultaneously at each time step. The operational mechanism is illustrated

in Fig.2.

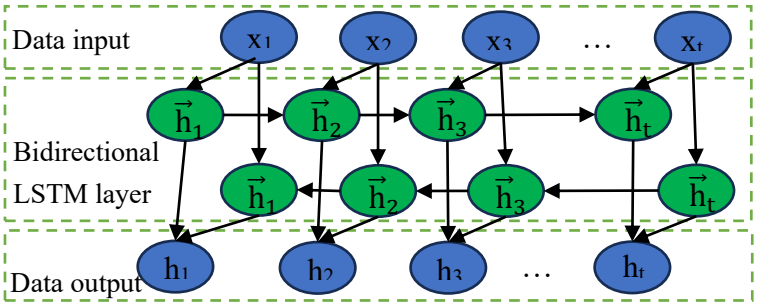


Fig.2 Prediction flow chart based on BiLSTM. $\vec{h}_i$(i=1,2,3…t) indicates the component of the hidden state output in the both directions; h_i(i=1,2,3…t) indicates the hidden state after stitching.

To evaluate the prediction performance, three loss functions, Root Mean Square Error (RMSE), Coefficient of Determination (R^2), and Mean Absolute Percentage Error (MAPE) are used to verify the effect of model. These three indexes directly reflect the error between the modeled value and the actual value by the equation (5-7).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y} - y_i)^2} \tag{5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y} - y_i)^2}{\sum_{i=1}^{n}(\bar{y} - y_i)^2} \tag{6}$$

$$MAPE = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{7}$$

where i denotes the coordinate epoch identifier. y_i represents the actual observation value of the ith sample. $\hat{y}_i$ denotes the predicted value of the ith sample. $\bar{y}$ is the mean of the actual observation values. N is the number of samples.

3. Results and analysis

The Amazon River Basin, located in South America, is one of the largest river basins in the world and is also considered as the largest freshwater river system on Earth(Miranda *et al.*, 2023). The basin spans nine South American countries, including Brazil, Peru, Colombia, Venezuela, Bolivia, Ecuador, Guyana, French Guiana, and Suriname. Covering an area of more than 7,000,000 square kilometers, the Amazon Basin accounts for about one-fifth of the world's freshwater resources. It consists of numerous tributaries and rivers, with the Amazon River being the main river of the basin. The Amazon River, stretching approximately 6,400 kilometers, is one of the largest rivers globally (Adams *et al.*, 2013, Schultz, 2013). The river flows through tropical rainforests, providing habitat for a wide variety of wildlife. Therefore,

studying changes in terrestrial water storage in the Amazon River Basin is of great importance for preserving biodiversity and ensuring the sustainability of the environment and development within the basin. The Amazon River Basin is densely populated with GNSS observation stations, providing abundant observational data for studying crustal deformation. The data used in this paper comes from 28 GNSS station time series provided by the Nevada Geodetic Laboratory (NGL)(Lavoisiane *et al.*, 2021). This dataset includes information such as the year, month, day, and observed values in three directions (East, North, and Up), as well as the latitude and longitude of each station. An overview of the study area and GNSS sites distribution is shown in Fig.3.

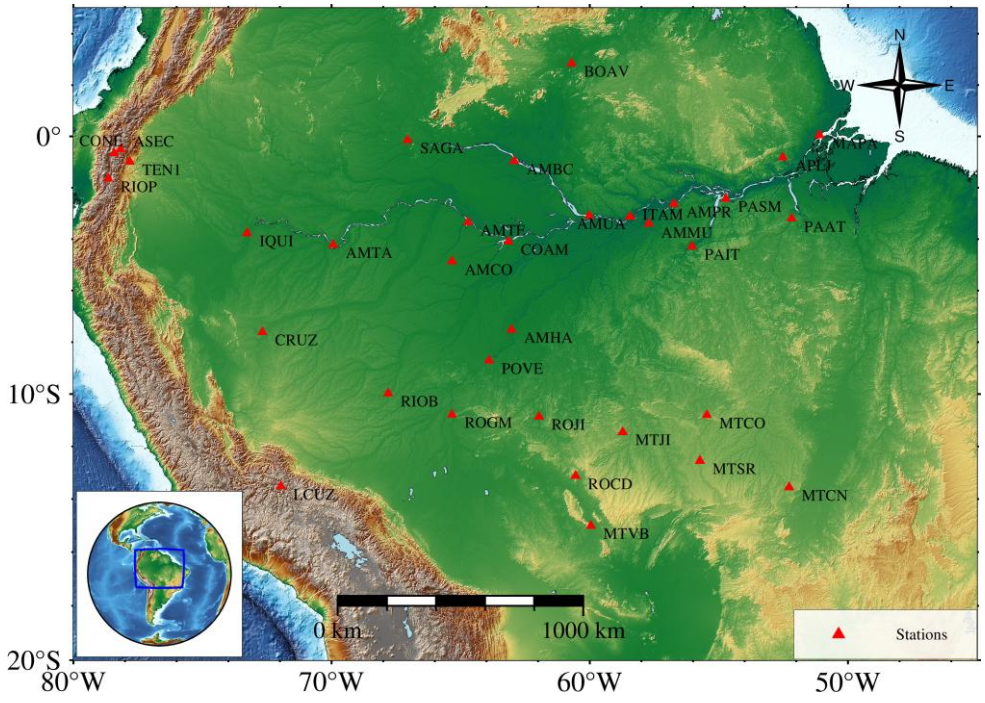


Fig.3 Overview of the Amazon River Basin and distribution of 34 GNSS reference stations

3.1 TWLD estimation results

All stations are decomposed with the improved VMD algorithm, 10 IMF components are obtained for each station. Taking AMBC station as an example, the concrete decomposition results are as shown in Fig.4. Low to high frequency sequence are from top to bottom. Low frequency sequences have obvious trend and low complexity, which are suitable for TWLD feature extraction and forward and backward integrated prediction with BILSTM model. The SE values for the first three components are 0.0063, 0.0548, and 0.2638, respectively. Therefore, IMF1-3 are identified as effective components and imported as input data into BiLSTM neural network model for further analysis. In contrast, the remaining seven IMF components are identified as high-frequency noise or interfering signals because their SE values exceed the set threshold.

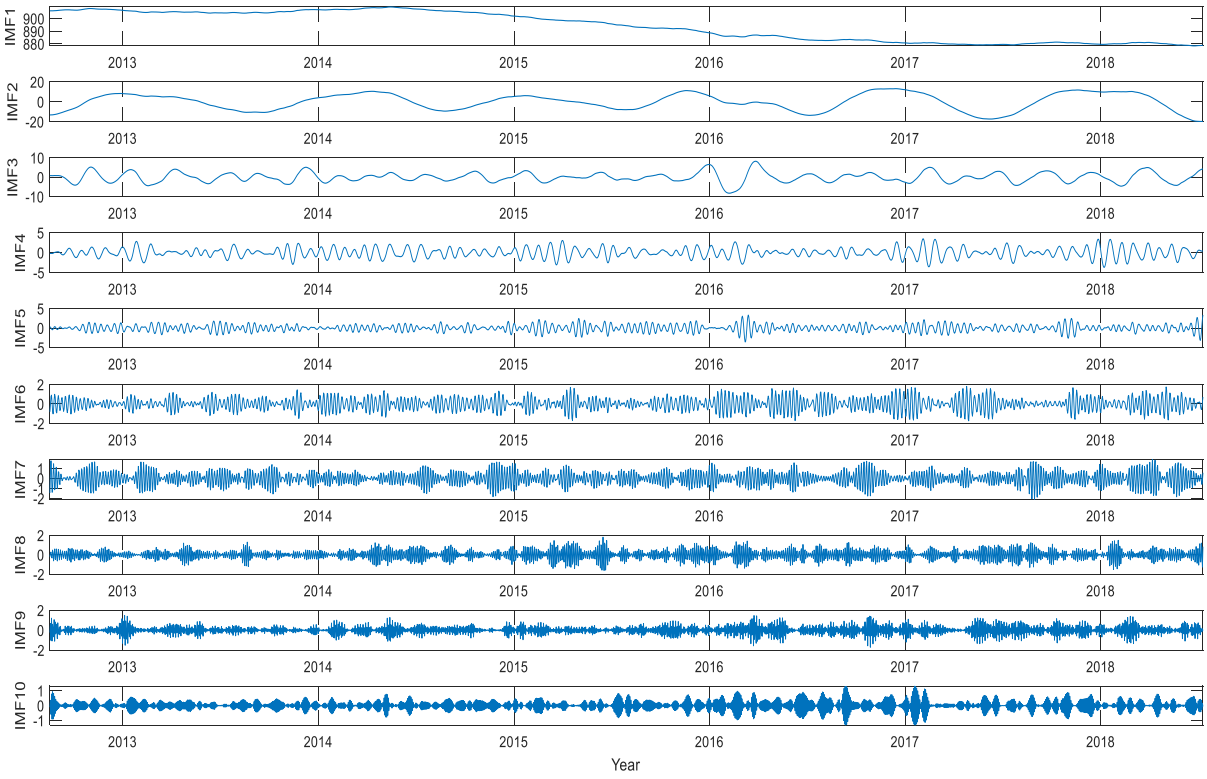


Fig.4. Decomposition result of AMBC site coordination time series

The extraction results of TWLD for six stations are shown in Fig.5. Obviously, after model processing, a significant amount of high-frequency noise is removed, the impact of outlier is reduced. Although there are localized data gaps of varying degrees existing in the original data, e.g., AMBC, AMHA, AMTA, AMPR stations suffers from the most severe data gaps. However, the processing method proposed in this study effectively mitigates this issue and the original trends and seasonality are well restored.

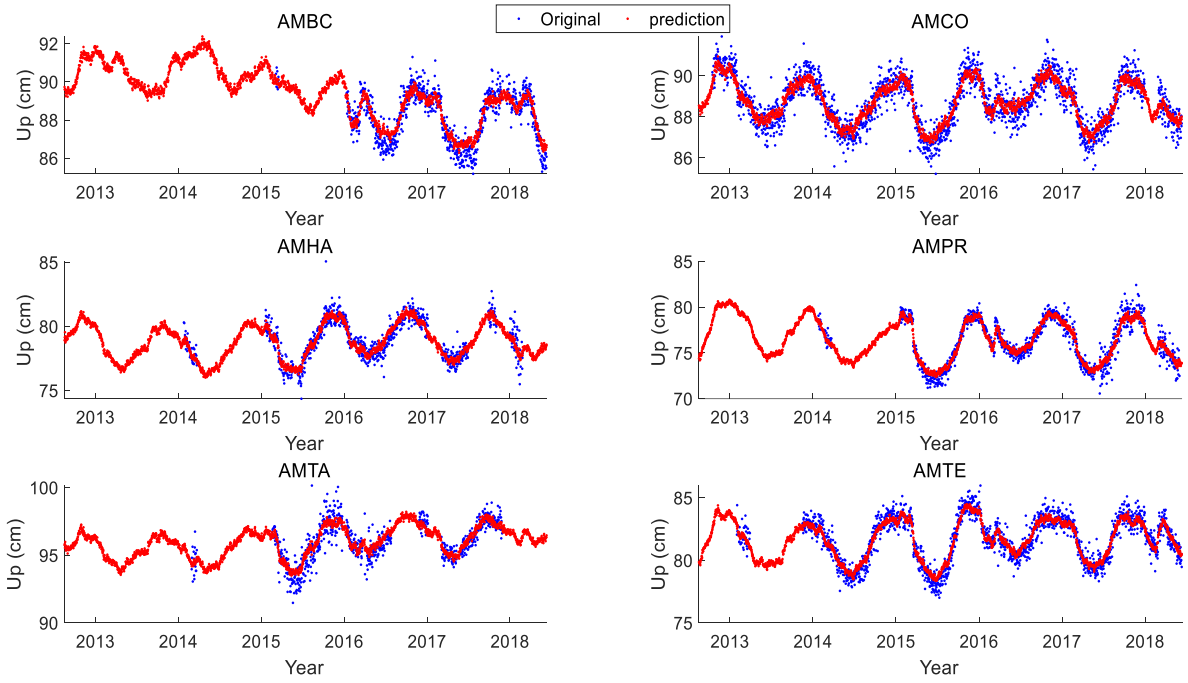


Fig.5 The prediction effects of neural network among the six out of the total 29 stations

The scatter plots of the deviation between the modeled values and the observed values among the six stations are shown in Fig.6. It exhibits a strong correlation and a low deviation. Obviously, the modeled values agree well with the observed values. It illustrates that the proposed integration strategy effectively reduces model training time and enhance feature distinguish ability. It preserves the annual periodic variations of GNSS time series and removes the influence of irrelevant information, illustrating a consistent improvement across all stations.

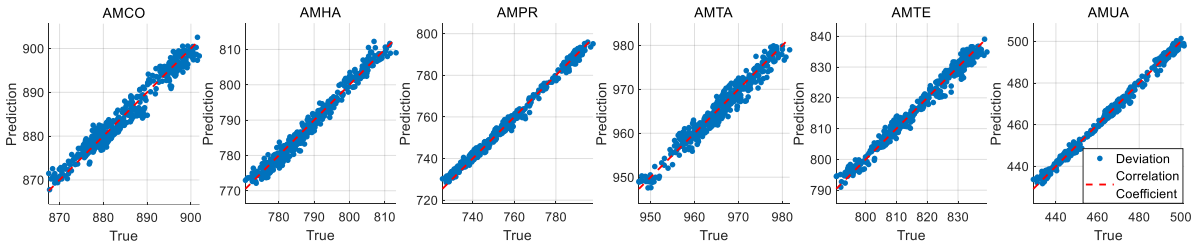


Fig.6 Scatter deviation diagram between the modeled value and the actual value

To quantitatively evaluate the reasonableness of the prediction results of VMD-BiLSTM, the RMSE, R^2, and MAPE of 28 stations are calculated, as shown in Tab.1. Compared with the single LSTM, BiLSTM model, VMD-BiLSTM shows a significant improvement, with the RMSE less than 3 mm, R^2 values are above 0.80. The MAPE values for all stations are within 2%. The average values of RMSE, MAE and MAPE is 1.495 mm, 0.971 and 1.337% respectively, while those of single LSTM and BiLSTM model are 5.576 mm, 0.780, 5.028% and 3.059 mm, 0.895, 3.307%, respectively.

Tab.1 Comparison of TWLD prediction results accuracy based on neural network

	LSTM			BiLSTM			VMD-BiLSTM		
Sites	RMSE/mm	R2	MAPE/%	RMSE/mm	R2	MAPE/%	RMSE/mm	R2	MAPE/%
AMBC	5.521	0.864	0.460	2.937	0.961	0.228	2.282	0.977	0.202
AMCO	4.211	0.813	0.404	2.962	0.908	0.281	1.849	0.964	0.165
AMHA	5.465	0.824	0.590	3.626	0.922	0.387	1.449	0.988	0.142
AMPR	9.046	0.825	1.021	5.439	0.937	0.608	1.571	0.995	0.154
AMTA	6.505	0.660	0.561	4.548	0.834	0.392	1.653	0.978	0.135
AMTE	6.737	0.814	0.711	4.182	0.928	0.437	1.746	0.987	0.166
AMUA	7.127	0.903	1.351	2.883	0.984	0.528	1.612	0.995	0.263
APLJ	7.015	0.570	0.611	1.702	0.975	0.141	1.914	0.968	0.152
ASEC	4.293	0.448	4.240	3.700	0.590	3.650	1.995	0.881	1.820
COAM	7.805	0.846	0.752	3.039	0.977	0.276	1.490	0.994	0.117
CONE	2.684	0.527	0.502	1.754	0.798	0.327	1.251	0.897	0.223
IQUI	4.302	0.821	3.059	3.231	0.899	2.264	1.304	0.984	0.825
ITAM	4.587	0.965	0.656	3.097	0.984	0.429	1.693	0.995	0.215
LCUZ	5.320	0.645	0.936	3.159	0.875	0.539	1.246	0.981	0.210
MAPA	6.881	0.447	2.301	5.411	0.658	1.811	1.313	0.980	0.415
MTCN	4.011	0.691	1.537	2.722	0.858	1.036	1.085	0.977	0.401
MTCO	5.984	0.659	2.242	2.431	0.944	0.898	1.234	0.986	0.452
MTJI	4.395	0.731	1.371	2.471	0.915	0.765	1.085	0.984	0.302
MTSR	5.278	0.755	0.690	3.945	0.863	0.521	1.399	0.983	0.168
MTVB	5.144	0.543	0.722	2.410	0.900	0.334	1.185	0.976	0.151
PAAT	5.444	0.783	0.868	2.791	0.943	0.441	1.510	0.983	0.223
PAIT	7.078	0.809	2.989	3.022	0.965	1.219	1.580	0.990	0.566
PASM	8.422	0.779	11.418	3.488	0.962	4.595	1.469	0.993	1.751
RIOB	4.386	0.668	0.585	1.975	0.933	0.259	1.262	0.973	0.161
ROCD	6.337	0.599	0.772	2.916	0.915	0.350	1.392	0.981	0.155
ROGM	6.317	0.694	0.692	2.040	0.968	0.204	1.774	0.976	0.185
ROJI	5.556	0.650	0.519	2.635	0.921	0.243	1.494	0.975	0.131
TEN1	2.987	0.357	0.249	2.108	0.680	0.175	1.599	0.816	0.132

The velocity of the modeled time series of all stations are shown in Fig.7. Compared to the traditional direct fitting method, the accuracy of velocity is improved 15.73%. It is evident that the uneven changes in TWS leading to the significant variations in vertical crustal deformation in the middle and lower Amazon River Basin. This observation aligns with the findings of previous research(Lavoisiane *et al.*, 2021), which indicate an overall subsidence trend in the northeast and southwest regions, an uplift trend in the southeast, and no significant changes in the northwest region.

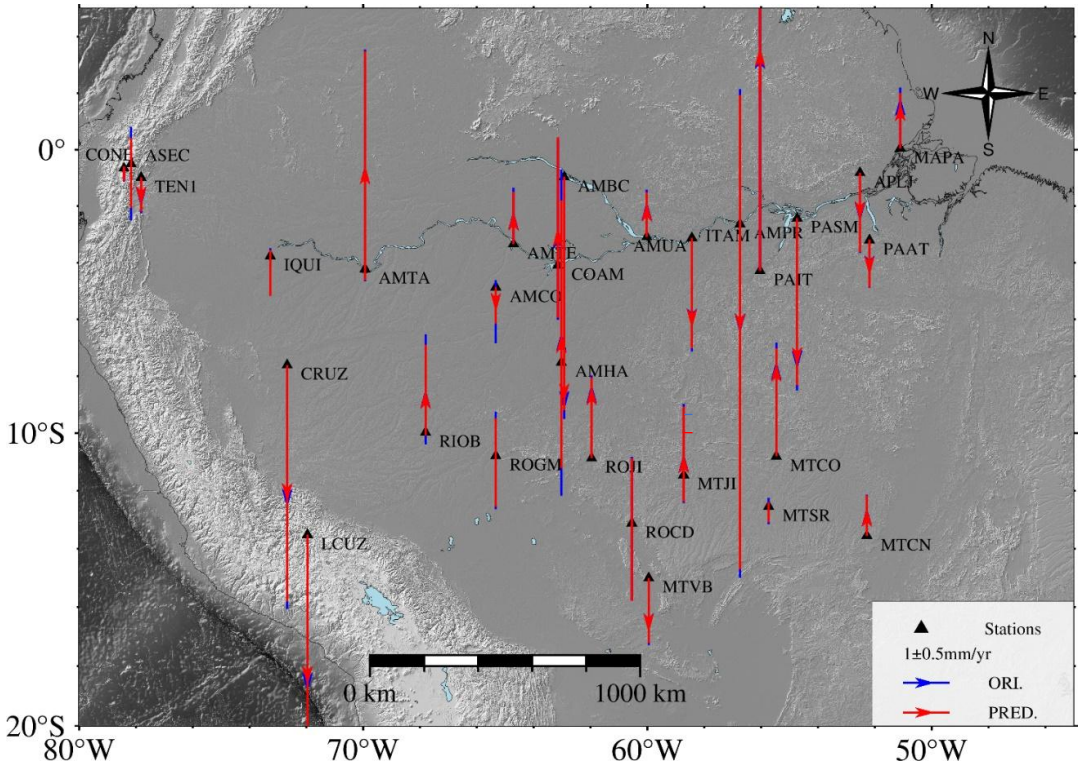


Fig.7 GNSS velocity field processed by VMD-BiLSTM neural network

3.2 TWS inversion results based on TWLD

The TWS changes are calculated based on Green's function constrained by TWLD(Jiang *et al.*, 2022), as shown in Fig.8. The change of water volume in the Amazon River basin show an obvious seasonal variation characteristic with a decrease trend in autumn and winter, and an increase trend in spring and summer. In addition, there is a small span of missing data between January and February 2018, which may be due to GNSS data integrity. EWH in the northeast changed sharply during 2012 and 2018, and gradually decreased in the south.

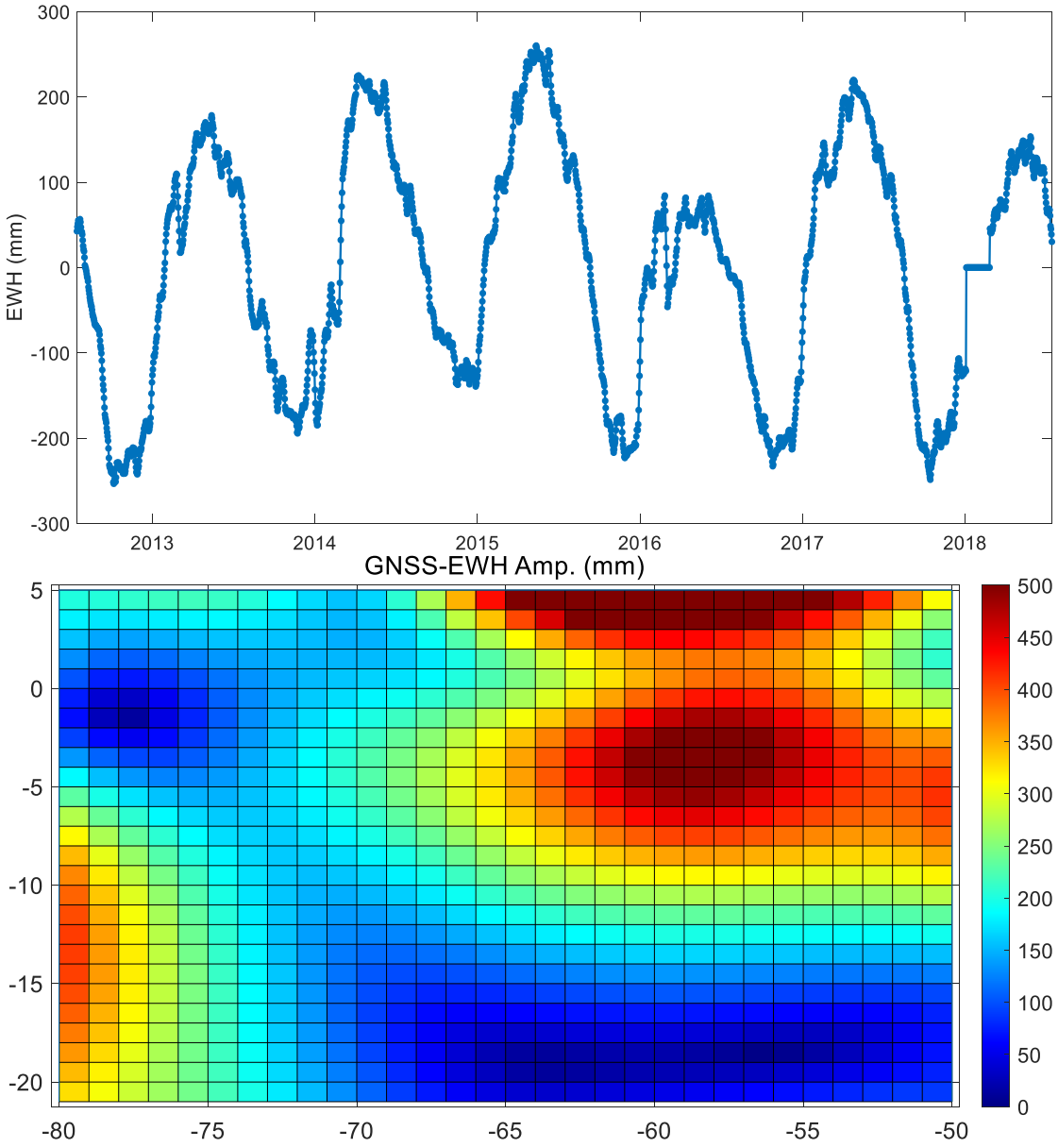


Fig.8 Spatio-temporal variations of TWS in the Amazon River basin

4. Discussion

To validate the reliability of the hybrid model, we compare the GNSS inversion results with GLDAS and GRACE data. Just like GNSS observations, GLDAS and GRACE data also contains errors, the effective signal contains long-term trend terms (linear or accelerated),

interannual, short-term (annual and semi-annual) oscillating term and trend variations. Therefore, all three types of data are used as input and processed using the same method VMD-Bi-LSTM. According to the criterion of SE value less than 0.5, the extracted valid IMF components quantity of GNSS, GLDAS and GRACE is 4, 8, 6, respectively, as shown in Fig.9. The first IMF of both GRACE and GLDAS exhibits an obvious interannual variation due to hydrology seasonal factors, and the second IMF shows a low-frequency trend change with reduction followed by increase, while the interannual and trend variation are mixed in both the first and second IMF components of GNSS.

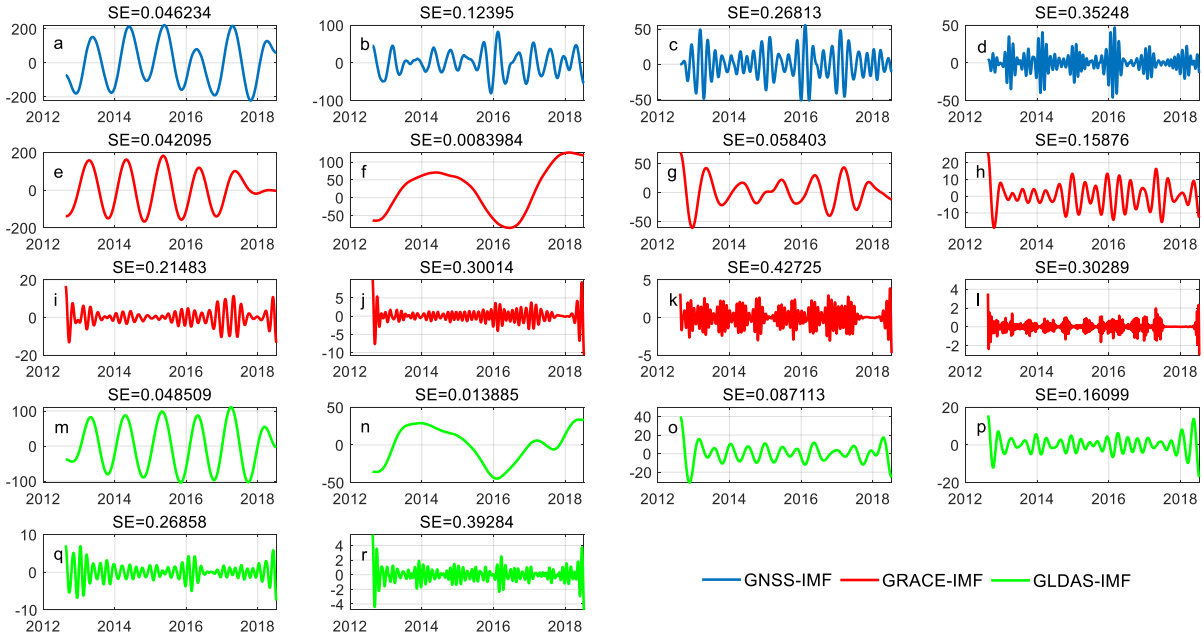


Fig.9 Effective components extracted from each data source using the improved VMD model

The valid components from each data source are composed as input to reconstruct TWS change series for BiLSTM model, as shown in Fig.10. It is seemed clearly that the amplitude of the inversion results from GNSS and GRACE are nearly identical, while the GLDAS results exhibit smaller amplitudes. However, all datasets display the similar seasonal patterns and a strong consistency.

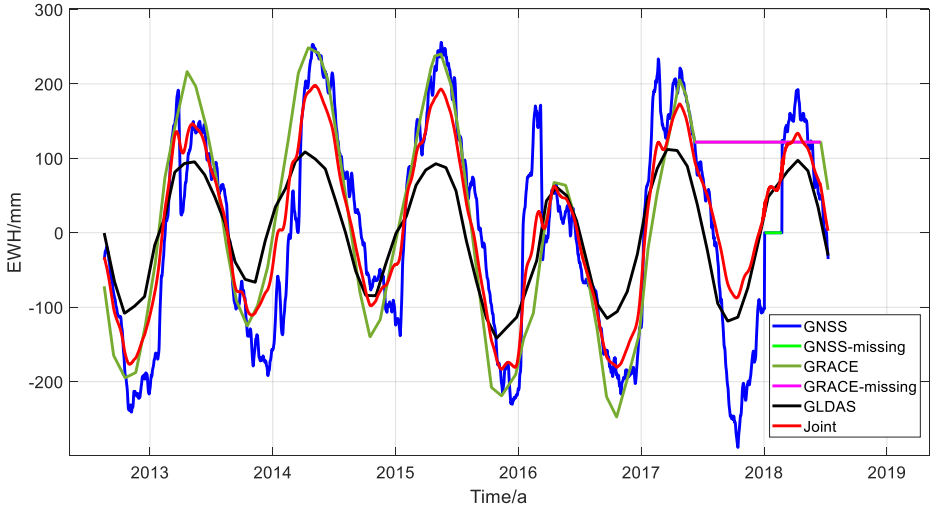


Fig.10 Comparison of inversion results for TWS change time series. The blue, green, black and red line represents GNSS, GRACE, GLDAS and joint GNSS, GRACE and GLDAS data inversion. The green and purple lines represent the data gaps in the GNSS and GRACE observations, respectively.

Spatial distribution of TWS from GRACE and GLDAS are shown in Fig.11. Compared to the other methods, GNSS inversion results show a slight difference in the southwest region of Amazon Basin, mainly due to the sparse GNSS station distribution, where only one station LCUZ. The spatial interpolation and smoothing effects may affect the signal and distort the information carried by the signal. There is a relatively dense distribution of GNSS stations in the northcentral and eastern part of the study area. As a result, TWS can be effectively inverted in these regions, where TWS from GNSS, GRACE and GLDAS exhibits an extremely similar change trends, with larger changes in the middle and lower reaches of the northeastern Amazon Basin, gradually tapering off towards the upstream areas near the Andes Mountains in the southwest. The significant variations area of water volume corresponds to the estuary or the downstream section of the Amazon River Basin. Moving southward from the aforementioned region, the magnitude of water storage changes gradually becomes more moderate.

It should be noted that integrating GNSS, GRACE and GLDAS can obtain a comprehensive and highly complementary product of TWS. It not only effectively mitigates the uncertainties of GNSS inversion results caused by the sparse stations distribution in certain areas, but also fill in the observations gap in geodetic time series. The correlation coefficient between the GNSS and GRACE, GLDAS reaches 0.71, 0.81, respectively. The joint result of three datasets maintains a high consistency with the single data inversion, with correlation coefficients of above 0.91. This indicates that our proposed method is not restricted to daily GNSS time-series, but it applies to a different geodetic time-series for TWS with different sampling rates.The joint inversion results of multi-sources data can provide an integrated product for the pros and cons of multi sources data.

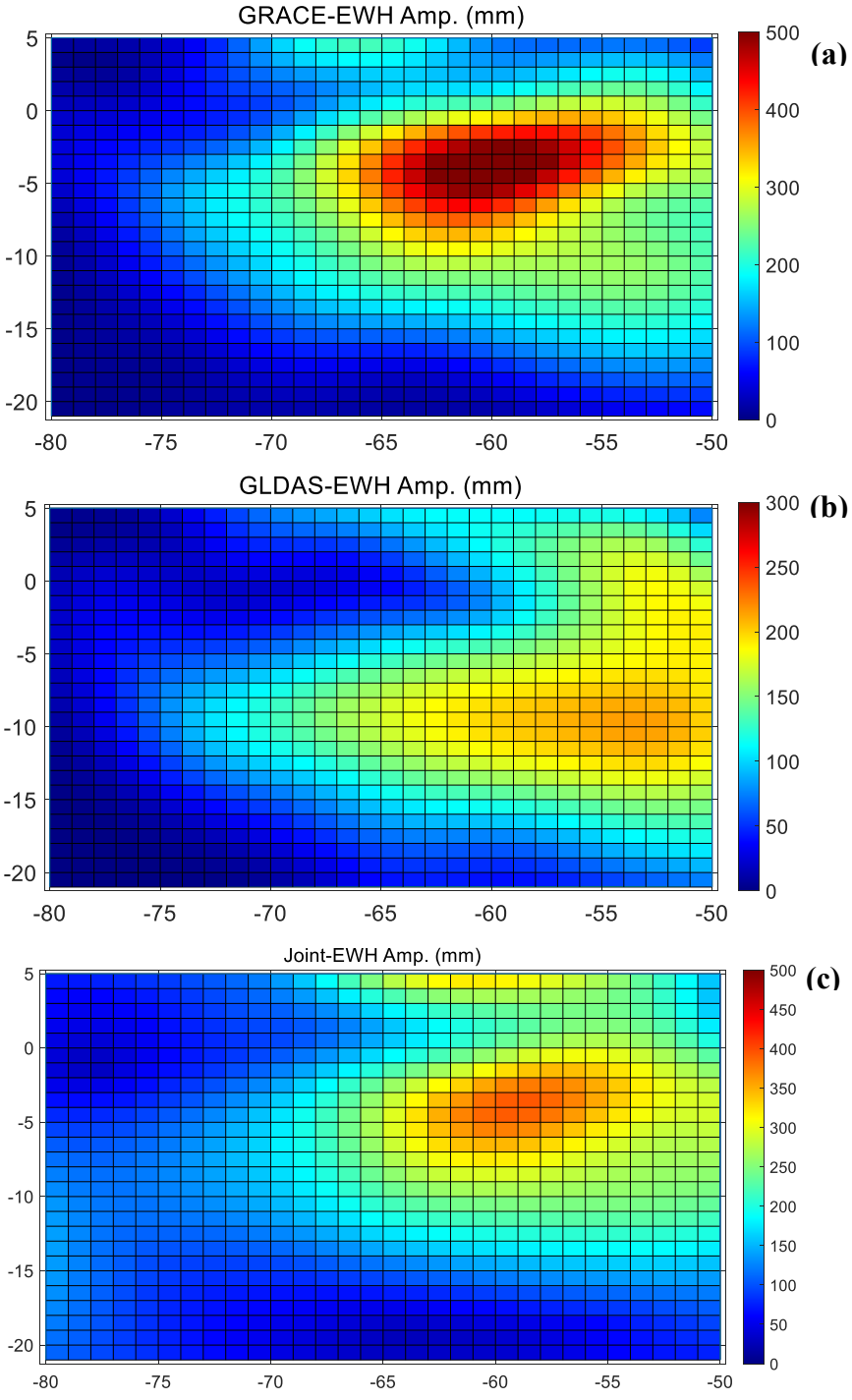


Fig.11 Spatial distribution of TWS inversion results.(a) , (b) and (c) is the annual amplitude distribution of EWH from GRACE, GLDAS and joint GNSS, GRACE and GLDAS data inversion

5. Conclusion

We propose a new hybrid method of integrating dimensionality reduction and deep learning for improving estimation of TWLD. It allows control over the number of modes generated during decomposition, and can control the number of basis functions and amplitude modulation

functions generated by adjusting the regularization parameters. This flexibility in meeting different signal decomposition needs avoids issues with unstable or excessive mode numbers, offering better control over the training results. Therefore, it is recommended as an alternative for precise model reconstruction the complex time-varying signals, estimating and signals extraction of satellite geodesy time series, especially in modeling the secular motion velocity, and recovering the large missing data. The experiments results show that the hybrid model significantly outperforms the single model, with an average improvement of 73.20%, 26.87%, 73.40% and 51.14%, 7.83%, 59.55% of RMSE, R^2 and MAPE compared to LSTM and BiLSTM. The joint inversion model integrating GNSS, GRACE and GLDAS data can give a more complete picture of the earth's water cycle, and obtain spatio-temporal distribution of TWS more continuously and reliably than single data inversion. To obtain more reliable TWS changes from GNSS observations, firstly, densely distributed GNSS stations will be necessary for providing extraordinary spatiotemporal resolution in constraining TWS changes, secondly, the multi-scales spatiotemporal decomposition and filtering algorithms should be introduced to separate the actual TWLD from GNSS observations. Only in this way, GNSS-based TWS inversition will have an extensive application prospect for robust characterization of hydrological extremes and improved operational flood forecasts or drought monitoring.

Availability of data and materials

Data will be made available on request.

Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This study is funded by the National Natural Science Foundation of China (42474039), and and the State Key Project of National Natural Science Foundation of China-Key projects of joint fund for regional innovation and development (U25A20783).

Authors' contributions

Keke Xu: Methodology, Conceptualization, Funding acquisition. Yifu Liu: Investigation, Data processing, Visualization, Conceptualization, Writing- original draft. Xianglei Liu: Revising- original draft, Editing, Formal analysis. Yongzhen Zhu: Data curation, Validation, Investigation.

Acknowledgements

We are grateful to the Center for Space Research (CSR), University of Texas for providing

GRACE mascon solutions (http://www2.csr.utexas.edu/grace).

Reference

Adams, D.K., Gutman, S.I., Holub, K.L. & Pereira, D.S., 2013. GNSS observations of deep convective time scales in the Amazon, *Geophysical Research Letters,* 40**,** 2818-2823.

Bao, Z., Chang, G., Zhang, L., Chen, G. & Zhang, S., 2021. Filling missing values of multi-station GNSS coordinate time series based on matrix completion, *Measurement,* 183**,** 109862.

Bennett, R.A., 2008. Instantaneous deformation from continuous GPS: contributions from quasi-periodic loads, *Geophysical Journal International*.

Bevis, M. & Brown, A., 2014. Trajectory models and reference frames for crustal motion geodesy, *Journal of Geodesy,* 88**,** 283-311.

Bogusz & Janusz, 2015. Geodetic aspects of GPS permanent station non-linearity studies, *Acta Geodynamica et Geomaterialia*.

Chen, H., Lu, T., Huang, J., He, X., Yu, K., Sun, X., Ma, X. & Huang, Z., 2023. An Improved VMD-LSTM Model for Time-Varying GNSS Time Series Prediction with Temporally Correlated Noise, *Remote Sensing,* 15**,** 3694-.

Chen, Q., van Dam, T., Sneeuw, N., Collilieux, X., Weigelt, M. & Rebischung, P., 2013. Singular spectrum analysis for modeling seasonal signals from GPS time series, *Journal of Geodynamics,* 72**,** 25-35.

Didova, O., Gunter, B., Riva, R., Klees, R. & Roese-Koerner, L., 2016. An approach for estimating time-variable rates from geodetic time series, *Journal of Geodesy,* 90**,** 1207-1221.

Dong, D., Fang, P., Bock, Y., Webb, F., Prawirodirdjo, L., Kedar, S. & Jamason, P., 2006. Spatiotemporal filtering using principal component analysis and Karhunen-Loeve expansion approaches for regional GPS network analysis, *Journal of Geophysical Research: Solid Earth,* 111**,** n/a-n/a.

Dragomiretskiy, K. & Zosso, D., 2014. Variational Mode Decomposition, *IEEE Transactions on Signal Processing,* 62**,** 531-544.

Ghaderpour, E. & Ghaderpour, S., 2020. Least-squares Spectral and Wavelet Analyses of V455 Andromedae Time Series: The Life After the Super-outburst, *Publications of the Astronomical Society of the Pacific,* 132**,** 114504.

Ghaderpour, E. & Pagiatakis, S.D., 2019. LSWAVE: a MATLAB software for the least-squares wavelet and cross-wavelet analyses, *GPS Solutions,* 23**,** 50.

Hyvärinen, A. & Oja, E., 1997. A Fast Fixed-Point Algorithm for Independent Component Analysis, *Neural Computation,* 9**,** 1483-1492.

Jiang, Z., Hsu, Y.J., Yuan, L., Feng, W., Yang, X. & Tang, M., 2022. GNSS2TWS: an open-source MATLAB-based tool for inferring daily terrestrial water storage changes using GNSS vertical data, *GPS Solutions,* 26**,** 1-7.

Jing, W., Guigen, N., Shengjun, G., Shuguang, W., Haiyang, L. & Xiaobing, R., 2021. Landslide Deformation Prediction Based on a GNSS Time Series Analysis and Recurrent Neural Network Model, *Remote Sensing,* 13**,** 1055-1055.

Junru, Y., Changsheng, Q., Qiqiang, C. & Jiantao, Q., 2021. Spatial-Spectral Network for Hyperspectral Image Classification: A 3-D CNN and Bi-LSTM Framework, *Remote Sensing,* 13**,** 2353-2353.

Klos, A., Bos, M.S., Fernandes, R.M.S. & Bogusz, J., 2019. Noise-Dependent Adaption of the Wiener Filter for the GPS Position Time Series, *Mathematical Geosciences,* 51**,** 53-73.

Kondrashov, D. & Ghil, M., 2006. Spatio-temporal filling of missing points in geophysical data sets, *Nonlin. Processes Geophys.,* 13**,** 151-159.

Lavoisiane, F., S., M.G., H., M.E., C., H.A.M., V., S.R. & N., C.J.M., 2021. Vertical Displacement Caused by Hydrological Influence in the Amazon Basin, *Journal of Geophysical Research: Solid Earth,* 126.

Li, J., Dai, L., Xu, K., Mei, X., Liu, Y., Shi, J. & Zhang, H., 2025. High-precision deformation monitoring and intelligent early warning for wellbore based on BDS/GNSS, *PLOS ONE,* 20**,** e0325913.

Li, X., Zhong, B., Chen, J., Li, J. & Wang, H., 2024. Investigation of 2020–2022 extreme floods and droughts in Sichuan Province of China based on joint inversion of GNSS and GRACE/GFO data, *Journal of Hydrology,* 632**,** 130868-.

Liu, N., Dai, W., Santerre, R. & Kuang, C., 2017. A MATLAB-based Kriged Kalman Filter software for interpolating missing data in GNSS coordinate time series, *GPS Solutions,* 22**,** 25.

Liu, Y., Xu, K., Guo, Z., Li, S. & Zhu, Y., 2024. Monitoring terrestrial water storage changes using GNSS vertical coordinate time series in Amazon River basin, *Scientific Reports,* 14**,** 24077.

Métivier, L., Collilieux, X., Lercier, D., Altamimi, Z. & Beauducel, F., 2014. Global coseismic deformations, GNSS time series analysis, and earthquake scaling laws, *Journal of Geophysical Research: Solid Earth,* 119**,** 9095-9109.

Mimi, P., Mahdi, M., Zhong, L., Zhuge, X., Zelong, G., Chaoying, Z. & Qinghao, L., 2024. Characterization and prediction of InSAR-derived ground motion with ICA-assisted LSTM model, *Remote Sensing of Environment,* 301**,** 113923-.

Miranda, P.M.A., Adams, D.K., Tomé, R., Fernandes, R. & Mateus, P., 2023. Optimizing Boundary Conditions in GNSS Tomography: A Continuous 7-Month Case Study in the Amazon, *Geophysical Research Letters,* 50.

Schultz, C., 2013. A new metric to help understand Amazon rainforest precipitation, *Eos, Transactions American Geophysical Union,* 94**,** 320-320.

Seerangan, K., Nandagopal, M., Nair, R.R., Periyasamy, S., Jhaveri, R.H., Balusamy, B. & Selvarajan, S., 2024. ERABiLNet: enhanced residual attention with bidirectional long short-term memory, *Scientific Reports,* 14**,** 20622-20622.

Shen, Y., Li, W., Xu, G. & Li, B., 2014. Spatiotemporal filtering of regional GNSS network's position time series with missing data using principle component analysis, *Journal of Geodesy,* 88**,** 1-12.

Tang, M., Yuan, L., Yang, X., Jiang, Z., Han, S.C. & You, W., 2024. Insights into water mass change in the Yangtze River Basin from the spectral integration of GNSS and GRACE observations, *Earth and Planetary Science Letters,* 644**,** 118929-118929.

Ting, X., Zhiqiang, D., Xiangwei, Z., Biyan, C. & Chengxin, R., 2023. LSTM-based short-term ionospheric TEC forecast model and positioning accuracy analysis, *GPS Solutions,* 27.

Wanqiu, L., Lifeng, B., Guobiao, Y., Fengwei, W., Qiuying, G., Jie, Z., Jinjie, Z., Zhiwei, W., Jingxue, B., Chengcheng, Z., Yulong, Z. & Shanbo, L., 2024. The analysis on groundwater storage variations from GRACE/GRACE-FO in recent 20 years driven by influencing factors and prediction in Shandong Province, China, *Scientific Reports,* 14.

Woods, K., Wallace, L.M., Williams, C.A., Hamling, I.J., Webb, S.C., Ito, Y., Palmer, N., Hino, R., Suzuki, S., Savage, M.K., Warren-Smith, E. & Mochizuki, K., 2024. Spatiotemporal Evolution of Slow Slip Events at

485 the Offshore Hikurangi Subduction Zone in 2019 Using GNSS, InSAR, and Seafloor Geodetic Data,
486 *Journal of Geophysical Research. Solid Earth,* 129**,** e2024JB029068-e022024JB029068.
487 Wu, H., Li, K., Shi, W., Clarke, K.C., Zhang, J. & Li, H., 2015. A wavelet-based hybrid approach to remove the
488 flicker noise and the white noise from GPS coordinate time series, *GPS Solutions,* 19**,** 511-523.
489 Xi, Z., Kailing, Y., Qin, L., Jingyu, W., Liang, Y. & Yu, L., 2023. Predicting carbon futures prices based on a
490 new hybrid machine learning: Comparative study of carbon prices in different periods, *Journal of*
491 *Environmental Management,* 346.
492 Zhaoyi, Z., Ying, L., Zhichen, L., Siwen, W., Hu, X. & Weihang, Z., 2023. Enhancing the reliability of shipborne
493 INS/GNSS integrated navigation system during abnormal sampling periods using Bi-LSTM and robust
494 CKF, *Ocean Engineering,* 288.

495