

Long short-term memory exploitation of satellite gravimetry to infer floods

Omid Memarian Sorkhabi^a, Joseph Awange^{b,*} 

^a Department of Geography and Earth Sciences, Aberystwyth University, Aberystwyth SY23 3DB, United Kingdom

^b Department of Land Surveying and Geo-Informatics, the Hong Kong Polytechnic University, Hong Kong Special Administrative Region of China

ARTICLE INFO

Keywords:

Flood
Satellite Gravimetry
Deep learning
GRACE
Downscaling

ABSTRACT

Flood forecasting is a vital segment of disaster risk management in that it contributes to the prediction of the magnitude, occurrence, duration and timing of floods. Owing to the nonlinear nature of atmospheric phenomena, however, forecasting becomes a challenging task that requires a multifaceted approach involving various sensors. Indeed, there exist compounding evidence that flood processes would benefit from use of various sensors. One such sensor is the Gravity Recovery and Climate Experiment (GRACE) and its follow-on (GRACE-FO), which provides Total Water Storage (TWS) products that could potentially be useful for flood monitoring and forecasting. However, GRACE/GRACE-FO's coarse spatial resolution of 300 km remains a bottleneck to the full exploitation of its products for flood studies and management. Herein, a deep learning Long Short-Term Memory (LSTM) method with high learning capability that optimizes the hyperparameters is proposed to downscale the coarse GRACE/GRACE-FO TWS products (from 300 km to 55 km). Its spatial and temporal learning is subjected to three different training scenarios (i.e., 60 %, 70 % and 85 %), where the one with least root-mean-square-errors (RMSE) is selected as the best-case scenario. The proposed LSTM deep learning approach is tested based on the 2019 Lorestan flood in Iran, where the results show that it successfully models the spatio-temporal behavior of TWS changes with its long-term and short-term memory capabilities. In March and April 2019, heavy precipitation caused a significant increase in TWS changes, approximately 40 ± 2 cm. This is captured by the LSTM-downscaled products but not the coarse GRACE/GRACE-FO TWS changes. Furthermore, the LSTM downscaled GRACE-FO TWS for the period after 2018 shows a strong and statistically significant mean correlation (above 0.70 at the 95 % confidence level) with both river discharge and precipitation. The original GRACE-FO on the other hand shows a correlation of 0.40, indicating the superiority of the LSTM-derived GRACE-FO's TWS changes. The coarse resolution of the GRACE satellite is a major cause of low correlation, which improves after downscaling. LSTM thus has the potential of downscaling GRACE products, providing data that are useful for flood process, management and studies.

1. Introduction

Flood forecasting is increasingly becoming a vital task in light of climate extremes. It is essential for protecting lives as it enables emergency alerts to be issued, reducing economic losses through formation of better preparation and response strategies, managing water resources effectively, and adapting to climate change challenges. Since the atmospheric phenomena are nonlinear in nature, where very large changes may occur within a short period of time, the integration of advanced modeling techniques to enhance the effectiveness of mitigating the impacts of flooding is becoming a necessity. Such integration would require use of various sensors to exploit their benefits. Although there are various sensors available for flood monitoring, including

Sentinel-1 radar images, which can penetrate clouds, each sensor has its own limitations, such as satellite pass frequency and challenges with backscatter in desert areas (Martinis et al., 2018). The Gravity Recovery And Climate Experiment (GRACE) satellites launched in 2002 (Tapley et al., 2019) and its follow-on (GRACE-FO) launched in 2018 (Pascolini-Campbell et al., 2021) have emerged as sensors that could be of potential benefit to flood forecasting as they provide a more comprehensive understanding by measuring TWS changes, which can play a strategic role in monitoring extreme and large-scale hydrological events. Since its launch in 2002, the GRACE satellites have been able to measure total water storage changes (TWS; surface, groundwater, soil moisture, vegetation water, and ice; see e.g., Awange and Kiema, 2019; Awange, 2022), which are useful for flood studies (e.g., Reager et al., 2014;

* Corresponding author.

E-mail address: joseph.awange@polyu.edu.hk (J. Awange).

<https://doi.org/10.1016/j.jag.2025.104562>

Received 20 September 2024; Received in revised form 21 March 2025; Accepted 22 April 2025

Available online 3 May 2025

1569-8432/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Gouweleeuw et al., 2018; Gupta and Dhanya, 2020; Khorrami et al., 2022, Zhang et al., 2023). That TWS changes is vital for flood monitoring is demonstrated by the development of its index employed to study floods (see, e.g., Reager and Famiglietti, 2009; Idowu and Zhou, 2019). Examples of the potential use of GRACE in flood detection include the works of Gupta and Dhanya (2020) who studied the Peninsular Indian River basins and developed flood potential Index that achieved 90 % accuracy in mild, moderate, and severe floods, and Gouweleeuw et al. (2018) who employed it to track major flood events in the Ganges–Brahmaputra Delta. In addition, local flood scale observations have been found to be well consistent with the United States of America (USA)’s flood potential index from 2003 to 2012 (Molodtsova et al., 2016). Also, USA’s Missouri River discharge has been studied using TWS changes, which indicated the flood potential of the river basin 5 to 11 months before its occurrence in 2011 (e.g., Reager et al., 2014). Chinnasamy (2017) studied the TWS changes with discharge for Koshi basin, where TWS changes were directly related to the monthly discharge. Shah and Mishra (2021) examined changes in TWS as a potential for detecting the Indian floods and found that an increase or decrease in TWS changes was a sign of potential floods.

Unfortunately, despite the advantages of GRACE satellites and the many efforts made over the past two decades in predicting floods using TWS changes as shown in the plethora of studies above, there are limitations to studying stored water changes (see e.g., Verma and Katpatal, 2020; Arshad et al., 2022) due to the coarse GRACE/GRACE-FO’s spatial resolution (approximately 300 km), which hinders its full exploitation to better understand TWS variability at localized levels (Hu et al., 2024; Ma et al., 2024; Yu et al., 2021; Xu et al., 2021). Downscaling of GRACE’s TWS products, therefore, is a necessity if its potential is to be fully exploited. In this regard, Vishwakarma et al. (2021) studied TWS downscaling with partial least squares with a RMSE of about 30 mm and found that statistical downscaling was successful. Further, machine learning methods (MLM) have been more successful in the TWS downscaling process where they have provided high correlation with independent sensors (see e.g., Seyoum et al., 2019; Chen et al., 2019; Zhong et al., 2021; Khorrami et al., 2023). Compared to classical methods such as recurrent neural network (RNN), MLM are designed to predict, and are thus more suitable where big data sets are to be used based on their input and output adjustment axis (He et al., 2021). MLM trains by identifying input features and output changes, and reduce prediction errors (Xie et al., 2022a). Deep learning, a subset of ML, has demonstrated considerable success in the fields of object detection and flood prediction. For instance, it has recently been used to reconstruct TWS changes of the Nile and Yangtze River basins by Wang et al. (2023, 2025). This powerful method has shown its effectiveness in accurately identifying objects and predicting floods, showcasing its potential across diverse applications (see e.g., Wu et al., 2018; Wu et al., 2020, 2021).

Even with the power of MLM above, the impact and effectiveness of TWS changes for flood monitoring are still controversial, and may have many uncertainties (Landerer and Swenson, 2012; Boergens et al., 2022). This is due to the fact that the chain of the precipitation process remains precisely unknown (see e.g., Khain et al., 2020), where meteorological parameters reduce the probability of flood prediction due to their dynamicity and sudden changes. To reduce uncertainty, therefore, the long-term and short-term behavior of TWS changes needs to be analyzed in such a way that its significant changes reveal its potential for flood prediction with appropriate accuracy. Long Short-Term Memory (LSTM) is one of the deep learning methods that considers long-term and short-term memories in its prediction (Sherstinsky, 2020). Unlike the traditional RNN in which content is updated at every step of the way, LSTM network decides on the preservation of the current memory through the introduction of gateways (Lin et al., 2021). Intuitively, if the LSTM unit detects an important feature in the input sequence of the initial steps, it easily transmits this information over a long distance, thus receiving and maintaining such potential long-term dependencies (Wu et al., 2020). Unfortunately, a simple RNN network cannot learn

such a distance connection, which is a big disadvantage (Su et al., 2020). LSTM has been successful in various fields of forecasting, e.g., Le et al. (2019) employed it to predict the flooding of the Da River basin achieving a flow rate forecasting with a Nash-Sutcliffe Efficiency (NSE) of 0.99. Other examples are reported in the works of Widiyari et al., (2018), Li et al. (2021), Man et al. (2022), Wang et al. (2021) and Xiong et al. (2022).

Although deep learning methods have been presented as a possible solution to address the spatial resolution problem of GRACE as discussed above, they have limitations in that those based on Long Short Term Memory (LSTM) rely on exterior data that at times may not be available or have to be modelled. Moreover, employing LSTM to downscale TWS using exterior variables such as precipitation comes with unknown uncertainties that limits its operation because the accuracy of the parameters involved are not known on one hand, while on the other hand, some of the parameters have to be extracted from models making the true downscaled TWS products unclear (e.g., Yang et al., 2019; Xiong et al., 2022). Here, we propose the LSTM method where the novelty lies in the fact that it employs optimization of the hyperparameters, thereby providing products suitable for flood monitoring unlike previous studies. LSTM is programmed to learn the short- and long-term behavior of TWS changes, which has more than 20 years of data and subjected to the case study of the 2019 Lorestan floods in Iran to assess its functionality. The downscaled GRACE-FO spatial resolution TWS products are investigated under three different scenarios (60–85 % training sets) to assess how they enhance both the spatial and temporal resolution of the data, addressing the limitations of previous studies. The study also examines how, unlike traditional methods, LSTM effectively localizes the signals, reducing spectral leakage and signal attenuation on one hand and on the other hand, explores how weak signals are amplified, providing more accurate and reliable results compared to prior models.

2. Long-Short term Memory: Optimized Framework

2.1. Algorithm

Long Short-Term Memory (LSTM) network is a special type of Recurrent Neural Network (RNN) that solves the long-term memory problem experienced by the RNN network. A common LSTM consists of three gates (see Fig. 1a): input, forget, and output, which control the flow of information and specify the data that is important to retain in the sequence, as well as the data that is less important and should be deleted (Man et al., 2022). The input gate is in-built to update the values (information) in the cell state while the forget gate is responsible for controlling the flow of information from the previous time step, and determines whether the memory information is used from the previous time step or not, and if something has to be entered from the previous time step. The Output gate ultimately decides what the next hidden state will be (Li et al., 2021).

In addition to these three gates, there is a memory cell, abbreviated as C. One input is hidden memory or (h) and the other input is X_t , which produces two outputs; C_t and h_t . h_t is divided into two parts, i.e., the part that is moved to the next time step and that which is used in the current time step if it needs to generate output (Song et al., 2019). Fig. 1 b) shows the LSTM cell that passes the important information along the sequence chain to provide the desired output (Wu et al., 2020).

The basic concept of the LSTM network is the cell state and its associated gates; the cell state acts as a freeway that carries information along the sequence chain, i.e., the network memory. Gates store information in a cell state (Kao et al., 2020), and are the various neural networks that decide what information enters the cell state, as well as learn what information should be stored or forgotten during network training (Gohar et al., 2022).

If the network is not able to understand recent history, long-term memory plays a stabilizing role and completes the prediction by looking back (Pang et al., 2020). One solution proposed for conditional

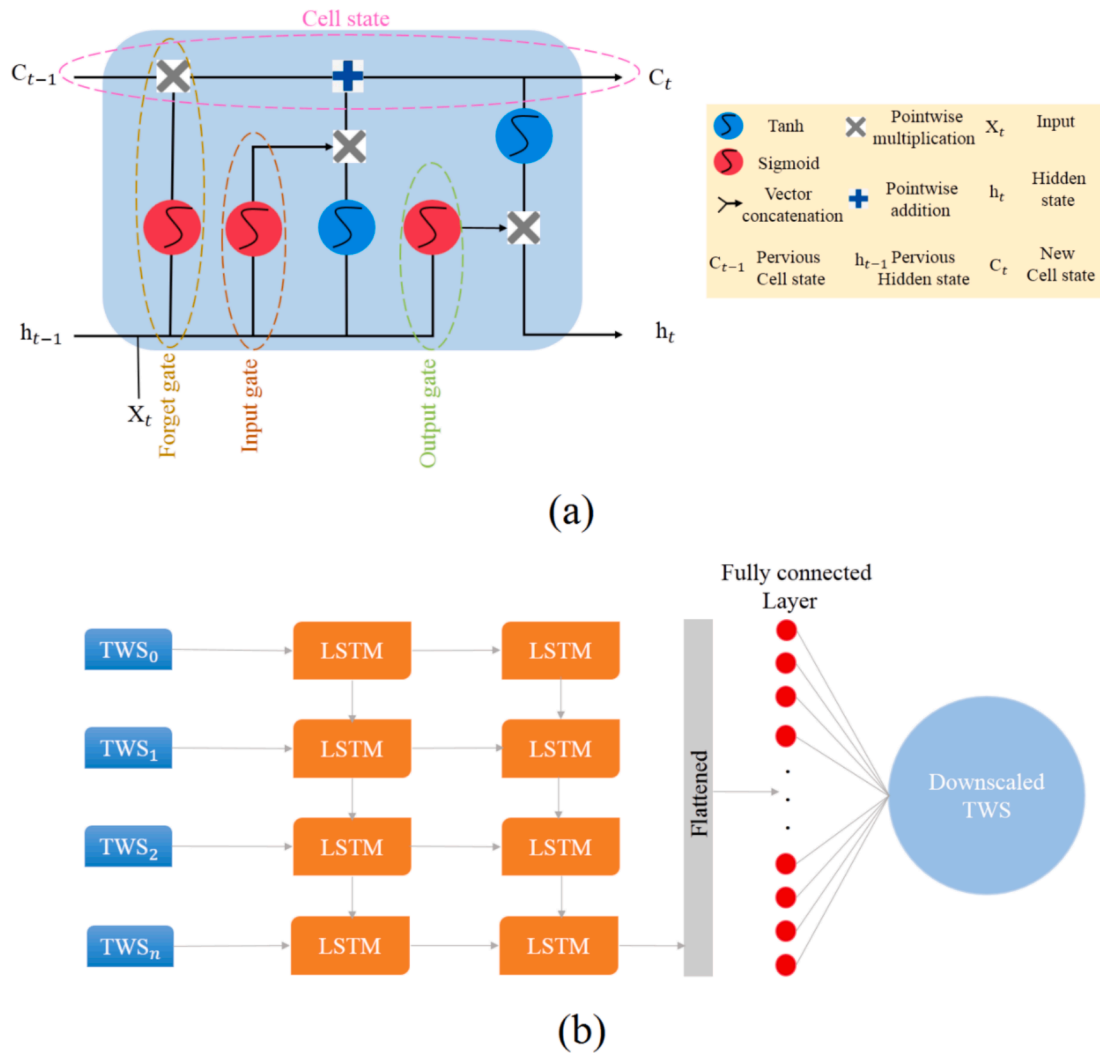


Fig. 1. A) LSTM cell featuring three gates: input, output, and forget gates. These gates play a pivotal role in storing information in a cell state, and b), LSTM architecture used for downscaling TWS changes, which is input into the LSTM at various time intervals. After being flattened, it proceeds to enter the fully connected layer.

models is to inject noise into the predictions made by the network before feeding them to the next time step (Alhussein et al., 2020). This strengthens the network against unexpected inputs. Nevertheless, effective solutions are less important than better memory. Long-Short Term Memory, or LSTM in short, is a RNN architecture designed to store and access information better than the traditional RNN version.

2.2. Optimization of hyperparameters

In deep learning, optimizing the network architecture and hyperparameters is crucial for achieving robust performance. For the LSTM model employed in this study, hyperparameters such as the number of hidden units, batch size, learning rate, and dropout rate are carefully selected based on preliminary experiments and cross-validation. The optimization is performed using Adaptive Moment Estimation (ADAM; Sakinah et al., 2019), which helps in adjusting the learning rate dynamically to ensure efficient convergence.

To determine the most effective (optimum) hyperparameter settings, multiple configurations are tested. The learning rate is varied across 0.001, 0.005, and 0.01 to assess its impact on model convergence and accuracy. A dropout rate of 0.3 provides the best balance between overfitting prevention and model performance, as lower dropout rates (e.g., 0.2) lead to slight overfitting, while higher dropout rates (e.g., 0.5)

cause underfitting and degraded performance. Additionally, different batch sizes (32, 64, and 128) are explored, with a batch size of 64 yielding the best stability and efficiency in training. The final hyperparameters are chosen based on the best performance observed during validation, ensuring the model achieves both accuracy and generalization.

Furthermore, training data proportions ranging from 60 % to 85 %, with a 5 % interval, are selected based on empirical trials and previous studies that explored the trade-off between model performance and generalization. These splits are used to investigate how varying amounts of training data impact on the model's ability to generalize. A higher proportion of training data (e.g., 85 %) can improve efficiency but may reduce the available validation data, increasing the risk of overfitting. Conversely, using a smaller proportion of training data may not allow the model to capture all relevant patterns. Thus, striking a balance between training and validation data is essential to optimize model performance while ensuring robustness and generalization.

2.3. Downscaling of GRACE/GRACE-FO TWS data

20-year TWS data from GRACE and GRACE-FO have been used for monthly downscaling. In the first stage (Fig. 1b), TWS data enters LSTM in a time series (with 1-month temporal resolution), where learning is

done in the temporal and spatial domain. After the flatten layer, which is to change the output dimension, they enter the fully connected layer that combine LSTM TWS results, and in the last stage, downscaled LSTM-TWS products are obtained. The first step shows the TWS calculation. The second step is the scenarios used to learn LSTM and the third step is validation with river flow discharge and precipitation. First, the C_{20} coefficient is replaced by that of satellite laser ranging (SLR; Xie et al., 2022a,b), followed by TWS calculation to the degree and order of 60. In the next steps, isostatic correction, and 300 km Gaussian filter (Tapley et al., 2019; to reduce stripe errors) are applied. Three training scenarios are selected based on empirical trials and previous studies approach, i.e., 60 %, 70 % and 85 % training data, and the remaining data used for validation (Fig. 2). The input of this model is the monthly TWS products of three center means: the Center for Space Research (CSR), the Jet Propulsion Laboratory (JPL), and the German Research Centre for Geosciences (GFZ) with a spatial resolution of $1^\circ \times 1^\circ$ degree and the output of this model is the 10-day TWS products with a resolution of $0.5^\circ \times 0.5^\circ$. Fig. 2 summarizes the downscaling workflow.

3. The Lorestan province flood (Iran)

3.1. Flood area

Lorestan Province, located in western Iran, features diverse elevations ranging from 100 to 3,000 m above mean sea level. Poldokhtar, a city in the southern part of the province, is one of its notable settlements. The areas of the case studies are in the west of Iran, specifically in Poldokhtar (Fig. 3 a and b with the associated rivers and topography), where the flood intensity was high during the flood period. The average annual precipitation is approximately 370 mm, with average temperatures reaching around 10°C . The city's elevation above sea level is approximately 713 m.

3.2. Precipitation event

The beginning of the solar new year in Lorestan province (Iran) was accompanied by the activities of two strong and extensive precipitation systems that affected almost the entire province and led to large and

devastating floods. From March 24 to April 2, 2019, Lorestan Kashkan Basin received 290 mm of precipitation. In Lorestan, the continuation of precipitation intensified the destruction caused by the floods. Due to its extent and pervasiveness, the damage to economic infrastructure, natural resources, and agriculture was enormous.

In March 2019, two waves of precipitation were experienced in Iran, leading to the formation of floods in the northern and western provinces. These precipitation waves occurred on (i) March 2nd to March 3rd, 2019, and in just two days, Iran's average precipitation increased by 72 % from its long-term precipitation average, and (ii), on March 17th to 26th, 2019. For the 10 days duration, the average precipitation of Iran increased by 50 mm.

These statistics mean that one-third of Iran's annual precipitation occurred within just two weeks. The average precipitation from September 23rd, 2018 to April 8th, 2019 was 310 mm, which is unprecedented in the last 51 years, and shows an increase of 49 % compared to the previous years. The average precipitation increases in 2019 compared to the average of 2018 in the Caspian Sea catchment area was 51 %, in the catchment area of the Oman Sea and the Persian Gulf was 56 %, and in the catchment area of Lake Urmia was 65 %. In the last 11 years, due to both natural factors and anthropogenic intervention (e.g., changing the natural balance by building dams and altering land use) in Iran, the average precipitation has never reached long-term average (approximately 230 mm/year), but the 2019 rains were much more than the long-term average.

The other precipitation wave in the Lorestan province was experienced from March 31st to April 2nd, 2019. Statistical analysis of the recorded precipitation results of this system in Khorramabad, Aleshtar, Noorabad, Kuhdasht and Poldokhtar meteorological stations shows 153 mm of precipitation in the whole Kashkan basin, with the highest amount of 172 mm occurring in Noorabad station, while the lowest amount of 136 mm occurring in Khorramabad station. On April 1st, 2019, the average precipitation reached 1058 mm. The annual amount of precipitation in the year 2018 was 275 mm and the average long-term precipitation in Lorestan was 228 mm. The flow of the Kashkan River reached $4,600\text{ m}^3/\text{s}$ on April 1st, causing the river to overflow in the cities of Poldokhtar and Mamulan. On the day of the flood (April 1st, 2019), the Khorramabad River flow increased to 700 m^3 , which was about 2 times the capacity of the river. With the flood discharge, the Karganeh River reached $280\text{ m}^3/\text{s}$, which has not been experienced for this small river during the last 20 years. The precipitations caused the Maruk Doroud Dam with a volume of 105 million m^3 to overflow for the first time.

The continuation of the rains was much more important than their intensity, an issue that played the biggest role in the destructiveness caused by the Lorestan flood. Fig. 3c shows the Kashkan Riverbed and the flood zone of Poldokhtar from the Sentinel-1 satellite image on March 25, 2019. The eastern and southern parts of Poldokhtar, which share a border with the river were completely flooded. The flood zone expanded about five times more than the riverbed, and the main eastern road of Poldokhtar was completely blocked in the first few days of the flood.

In this study, six sources of data are used: GRACE, global and land data assimilation system (GLDAS), precipitation, Sentinel-1, Sentinel-2, and river flow, with details for each described in Table 1.

Based on the 10-day GLDAS temporal data, downscaling of TWS changes has been conducted to improve spatial and temporal resolution. Inputs to the LSTM include four dimensions: longitude, latitude, time, and TWS changes. The LSTM optimizes downscaling by incorporating information from both monthly and 10-day GLDAS data. The use of the GLDAS model alone often produces unrealistic results, and accuracy decreases over extended time periods due to limitations in its model assumptions. To address this, monthly GRACE and GRACE Follow-On observations have been utilized, ensuring that mass displacements and measurement accuracy are maintained. The LSTM model effectively integrates location, time, and TWS changes to achieve optimal

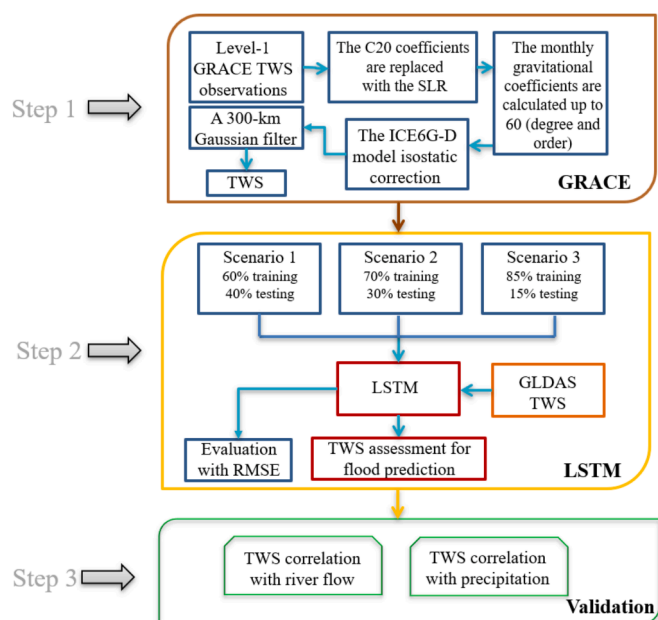


Fig. 2. TWS downscaling workflow comprising three components: GRACE satellite data processing, LSTM, and validation, which is conducted using two ground sensors measuring precipitation data and river flow.

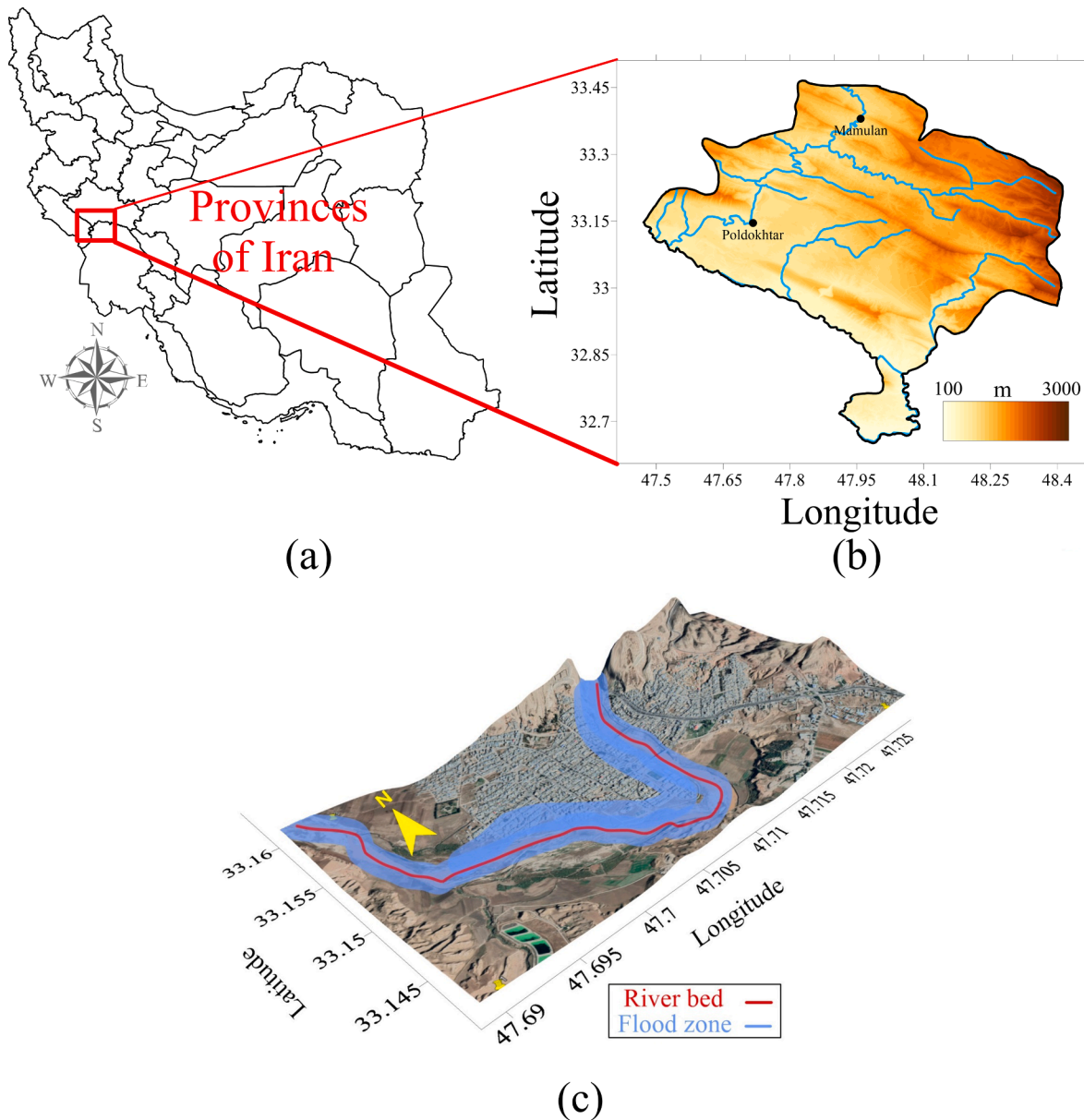


Fig. 3. a) Provinces of Iran, b) Poldokhtar and stream flow station, rivers, and topography in the southwest of Iran, and c), the Kashkan River bed and the flood zone of Poldokhtar from the Sentinel-1 satellite image of March 25th, 2019 showing the eastern and southern parts of Poldokhtar, which share a border with the river, which was completely flooded.

downscaling results, bridging the gap between coarse GRACE/GRACE-FO data. This approach enables a more reliable representation of TWS changes both temporally and spatially.

4. Results of LSTM based on the Lorestan (Iran) example

4.1. Application of LSTM to the 2019 Lorestan flood

As mentioned earlier, three datasets utilized in this research consist of GRACE (2002–2017) and GRACE-FO (2018–2021) satellite data, along with ground measurements of river flow (2002–2020) and precipitation data (3 months before and after the flood) from meteorological sensors, for verification purposes. Each data set contains information spanning periods before, during, and after the flood events indicated above.

Fig. 4 shows the annual TWS trends from the three centers (CSR, GFZ and JPL), and GLDAS trend. GLDAS TWS involves the hydrological

model to increase the spatial resolution so that it is not just a mathematical and unrealistic downscaling (see e.g., Rodell et al., 2004). GRACE observations are from three centers with a spatial resolution of 1 degree and GLDAS is 0.5 degree. TWS from the three centers are compatible with each other and TWS from GFZ in the northeast shows less decrease than CSR and JPL. Although GLDAS TWS has better spatial resolution than the three centers' solutions, it is nonetheless compatible with them. The 223-month GRACE and GRACE-FO data, along with the daily GLDAS (from which the monthly average is calculated), are employed in LSTM learning. The GLDAS TWS model is an integration of various hydrological components, including soil moisture, snow water equivalent, surface water, groundwater, and atmospheric data (Rodell et al., 2004). Before using GLDAS TWS, 100 random points are selected in different locations, and its compatibility with GRACE and GRACE-FO TWS evaluated, which shows 85 % compatibility. Also, TWS output is validated with river flow data in the flood time range. On the other hand, due to the validation capability of deep learning, which does not

Table 1

Data used in this study.

Data Source	Description	Provider	URL/ Source
GRACE (CSR, JPL, GFZ)	Gravity Recovery and Climate Experiment; gravity data	NASA – National Aeronautics and Space Administration	Jpl.nasa.gov
GLDAS	Global Land Data Assimilation System Noah Model (Version 2.1)	NASA	nasa.gov
Precipitation Data	Precipitation data with regional detail	IRMO – Iran Meteorological Organization	irmo.com
Sentinel-1	Synthetic Aperture Radar (SAR) imagery	ESA – European Space Agency, Copernicus Program	dataspace.cope.mnicus.eu
Sentinel-2	Multispectral imagery	ESA	dataspace.cope.mnicus.eu
River Flow	River flow data; regional hydrological records	WRM – Water Resources Management, Iran	wrm.ir

use part of the data in training, the reliability of the results increases.

In deep learning, the root mean square error (RMSE) is calculated using the validation data as a performance evaluation metric. A percentage of data, for example, 15 %, is set aside for validation and is not used during training. After training, the learned model is tested on TWS validation points. TWS is predicted, and RMSE is calculated using the following formula (Willmott and Matsuura, 2005):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2}, \quad (1)$$

where x_i represents the TWS values, $\hat{x}_i$ represents the predicted TWS values, and n is the number of validation data.

To mitigate overfitting and improve generalization, L2 regularization is applied to the loss function, which penalizes large weights, reducing model complexity and preventing it from fitting noise in the data. Additionally, several data augmentation techniques are tested, including random noise injection, time-series shifting, and scaling, to artificially expand the training dataset and help the model learn more robust features. These techniques are chosen because they can simulate variability in real-world data and improve the model's ability to make accurate predictions on new, previously unseen data.

After testing these methods, the L2 regularization, along with time-series shifting and noise injection, is found to significantly improve model generalization and reduced overfitting. These methods are selected as the optimal approach to enhance the performance of the LSTM model.

Three optimal scenarios are selected for downscaling TWS products employing LSTM: the first scenario uses 60 % of the data for training and 40 % for testing, the second scenario uses 70 % of the data for training and 30 % for testing, and the third scenario uses 85 % of the data for training and 15 % for testing.

Fig. 4b) shows the output of TWS scenarios based on LSTM. The annual TWS trend in Lorestan is negative, and the southwestern and northwestern regions have a greater decrease in TWS than the central and southeastern regions. As training data increases (in the third scenario), the range of TWS decreases. The highest RMSE is seen in Scenario 3, which has the most data used for training. The best-case scenario is given by the second scenario, which makes optimal use of the GRACE and GLDAS datasets.

Fig. 5 a) shows the JPL, CSR and GFZ time series for Poldokhtar after LSTM downscaling. Due to the high uncertainty, the gap between GRACE and GRACE-FO is not considered in the calculations. In general, the compatibility in the monthly solutions and seasonal signals from the

three centers are evident. A drop in the TWS changes signal is observed in mid-2007 and mid-2008. The annual range of TWS change from 2002 to mid-2007 is larger than from 2008 to 2014, and from then on, the annual range increases. After 2008, a drop in the TWS time series is evident, which is caused by water depletion (Fig. 5). It is also possible to calculate changes in groundwater storage with the help of GLDAS and GRACE/GRACE-FO (Sorkhabi et al., 2023). In the case of this study, the aim is to investigate the stepwise or gradual increase in TWS changes for specific locations due to abnormal precipitation during the time period, which is evident in the time series and spatial and temporal maps. At the end of 2018 and 2019, Iran's precipitation has been above normal. Although this TWS jump is temporary and the value of TWS decreased after this event, it cannot compensate for the water depletions of several years. For a better analysis of water depletions, GWS changes and indexes like the standard precipitation index (SPI) can be analyzed together, and linear trends can help.

Changes in the annual amplitude of TWS changes affect the amount of precipitation in the area. Torrential rains occurred in late March and early April 2019, with a TWS changes leap in all three centers' TWS products during this period, amounting to about 40 ± 2 cm. One of the reasons is the extent of precipitation during this period and the involvement of several provinces in Iran. The GRACE-FO satellite and LSTM has been able to clearly record this amount of leap (i.e., time series after 2018), which without employing LSTM is less i.e., 20 cm (see Fig. 5b). Furthermore, LSTM has been able to amplify the flood signal as shown in Fig. 5a.

One simple and statistical approach to detecting sudden increases in TWS changes is to use the mean and standard deviation as confidence intervals, where changes close to the confidence interval can be an indicator of flooding.

In the next step, the spatial TWS changes based on LSTM are further investigated. Fig. 6a shows the TWS changes for the period January to August 2019 after downscaling. In January, TWS changes are negative in all parts of Iran and the northern and western regions have further declines. In February, positive TWS changes are recorded in the western regions of Iran by GRACE-FO observations. This is an anomaly that is different from the rest of Iran. In March, the amount of TWS observed anomaly is tripled compared to February, and its extent covers almost the entire west of Iran. In April, the anomaly intensifies, and its magnitude is 1.5 times wider, with torrential rains measured in these areas on April 25th and 26th 2019. In the following months, it has gradually decreased, so that by August, the total TWS changes over Iran was negative. An increase in TWS changes is predictable when precipitation increases. The important point is the TWS positive anomaly in February before the flood in Fig. 6a, which can be a strong indicator of the impending floods. In the estimated TWS changes in Fig. 6a where LSTM is employed, the onset of floods from February, the period of progression in March, and in April, decrease in the following months and eventual cessation are observed. In Fig. 6b without employing LSTM (mean of the three centers products) months before the flood, an increase in TWS is not observed except in April (the month of the flood) and May (the month after the flood). The comparison of LSTM and original mode shows the superiority of LSTM-derived TWS changes, which with LSTM's long-term and short-term memory increases spatial resolution and reveals the pre-flood signals. Fig. 6a indicates a positive anomaly on the western side way before the April flood began. This TWS anomaly is an indicator/predictor of the April flood. In contrast, this is not seen in Fig. 6b where LSTM is not employed. In this study, multi-dimensional downscaling has been conducted using LSTM, where its advantage lies in its capacity for long- and short-term memory in deep learning, setting it apart from other ML methods. This is crucial for incorporating ML into downscaling over time.

Fig. 7a shows the time series of the TWS changes and river flow from 2002 to April 1st, 2019 (flood day). The behavior of the TWS changes is highly compatible with the river flow. Only the data of 3 months (mean monthly values) before and after the precipitation are available, and the

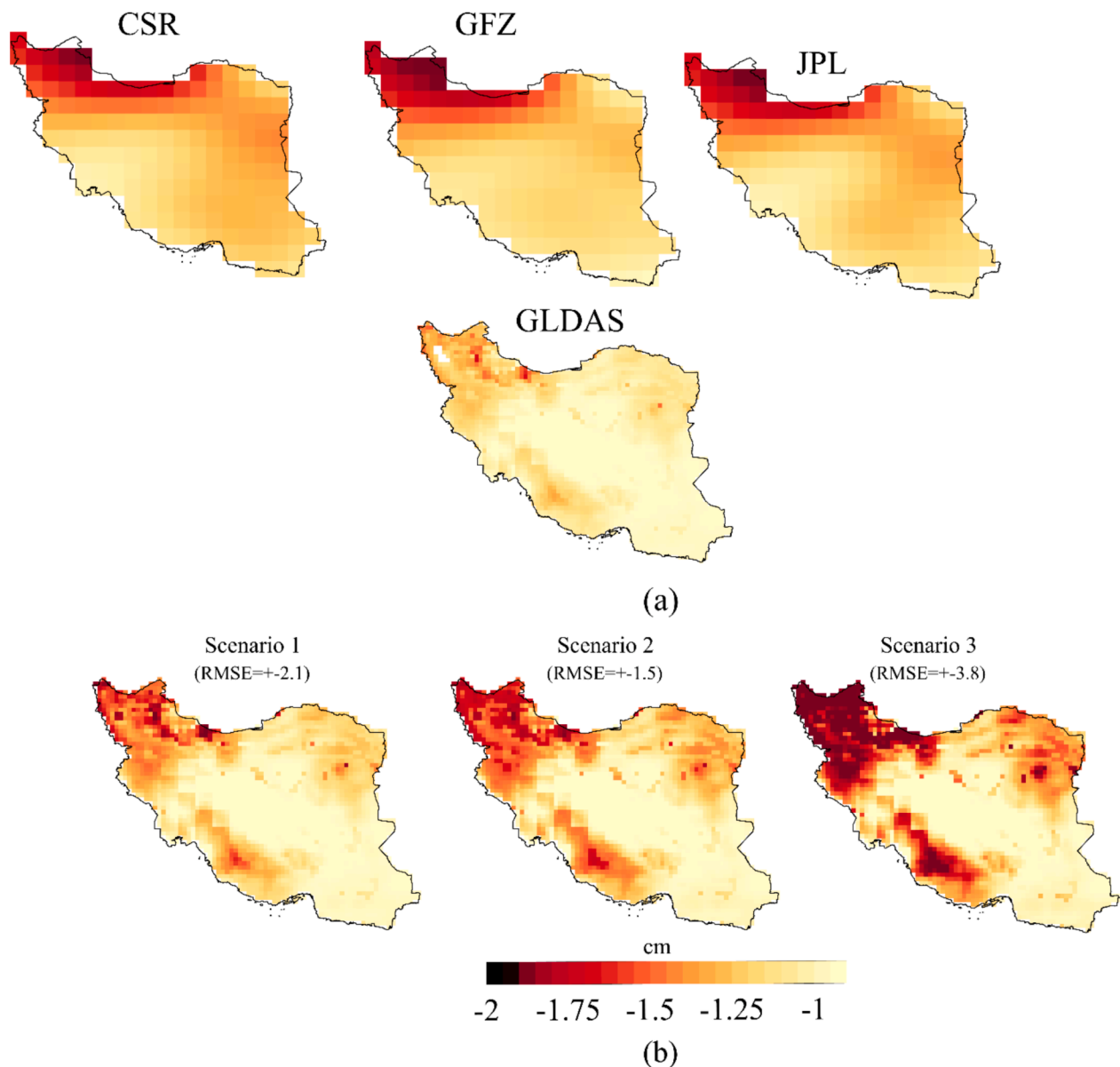


Fig. 4. A) TWS annual trends (cm/yr) of three centers (CSR, GFZ and JPL) and GLDAS, where darker areas show a further decrease, and b), output of TWS (cm) scenarios based on LSTM, which is evident in the improvement in spatial resolution and details.

correlation between TWS changes and the river discharge reaches its maximum on the day of the flood. This shows the LSTM-derived TWS product's capability for flood detection in the area within a month's time lag (Fig. 6). Fig. 7b shows regression between TWS changes and river discharge with 95 % and 99 % confidence intervals for statistical interpretations. TWS changes and Kashkan River flow have a correlation of 0.72 with 99 % confidence level. On flood days, the amount of correlation between the TWS changes and river flow is so high that it statistically becomes an outlier in the 99 % confidence interval.

For validation, the precipitation data of 7 meteorological stations in Lorestan province (Fig. 8a) have been used, and their precipitation correlation with TWS changes calculated. The names of the meteorological stations are Khorramabad (MS1), Rumeshkan (MS2), Nurabad (MS3), Dureh (MS4), Borujerd (MS5), Poldokhtar (MS6), and Kuhdasht (MS7). The TWS of these areas are also numbered, respectively.

Precipitation data from 3 months before and after the flood have been employed, considering GRACE's monthly temporal resolution. The average monthly precipitation has been utilized in the analysis. Fig. 7c and d show correlation results between precipitation and TWS changes

before and after flooding, respectively. The average correlations before and after flood flooding are 0.40 and 0.71, respectively. This approximate doubling of correlation between precipitation and TWS changes after the flood shows that the changes in both variables are synchronized. The reason for the negative correlation before the flood is due to the spatial difference between TWS and the precipitation of meteorological station. However, after the flood event, when the landscape is saturated and water storage mechanisms shift, the correlation becomes positive, indicating that these areas are directly influenced by increased surface water and replenished groundwater. Additionally, precipitation contributes to TWS through both direct accumulation and runoff. Before the flood, other factors such as soil moisture and groundwater storage play a more dominant role in TWS variability. In contrast, after the flood, the immediate impact of surface water dominates, leading to a stronger positive correlation between precipitation and TWS.

The GRACE data suffers from spectral leakage and low spatial resolution, which can result in inaccuracies when capturing fine-scale features. Additionally, the data is noisy, especially at higher degrees and orders (above 60), leading to signal attenuation at the maxima (peaks)

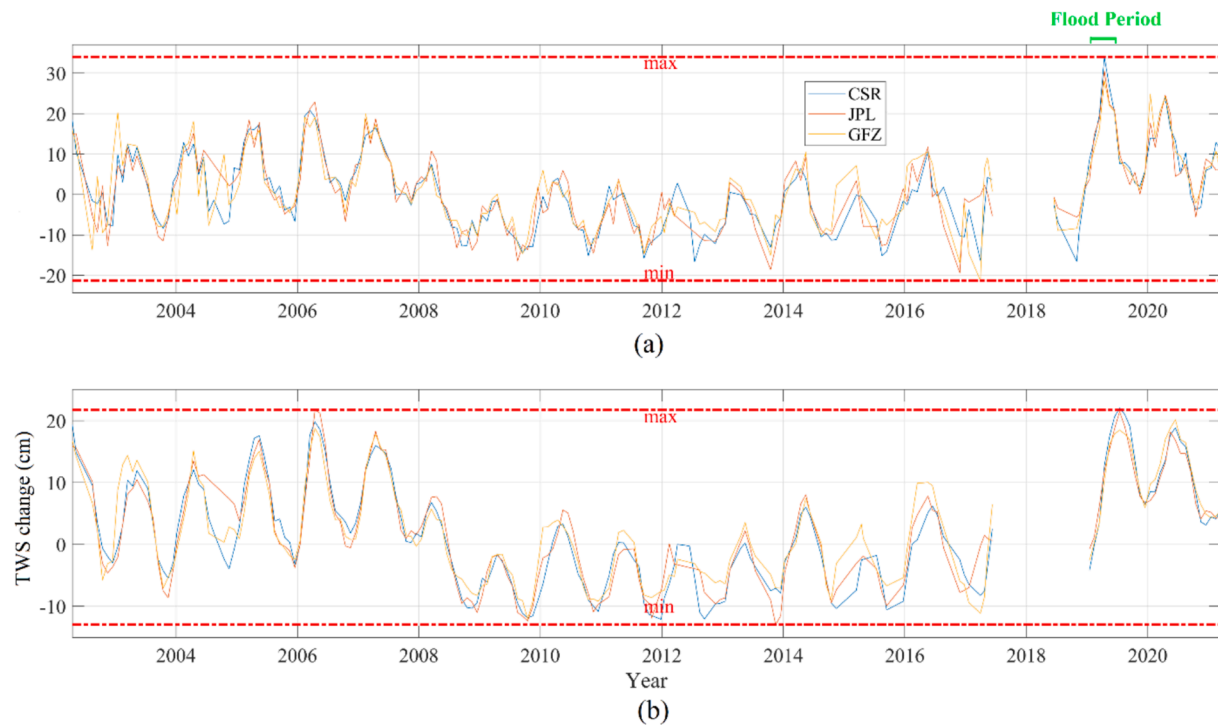


Fig. 5. JPL, CSR and GFZ time series a) after LSTM downscaling with maximum and minimum bands in red, and b), without employing LSTM. It is evident from figure that TWS change leap is more than 20 cm.

and minima. This issue requires localization methods and reconstruction to enhance the signal. LSTM has been utilized to address these challenges by improving both spatial and temporal resolution, as well as amplifying the signal to provide a clearer representation of the underlying data. To assess the accuracy of the LSTM output, independent data sources, such as river flow measurements (Fig. 7a) and weather station precipitation data, were used for comparison. The LSTM model showed an improvement in correlation with both the river flow and weather station data, with a 30 % increase compared to the non-LSTM model. This improvement suggests that the LSTM not only localized the signal more effectively but also enhanced the temporal resolution, which is evident in the more pronounced peak compared to the original GRACE data. The LSTM model's higher temporal resolution allows it to capture finer variations in the data, leading to sharper peaks that more accurately represent localized events. While the monthly averaged values of the LSTM align with GRACE, the enhanced peak in the LSTM output reflects its capability to better localize the signal both in time and space.

The Fig. 7c and d indicates that the correlation increased after the flood.

Fig. 8a shows the Sentinel-1 radar image of southwestern Iran on 2019-03-14 before the flood. Fig. 8b presents the Sentinel-1 image on 2019-03-26 after the flood, with black areas indicating precipitation covering an area of more than 100,000 km². The precipitation, occurring as snow in high altitudes and rain in low altitudes, led to floods. The expanded size increased the TWS changes in the region, discernible by the GRACE satellite. Fig. 8c shows a Sentinel-2 image the city of Poldakhtar before the flood on 2019-03-26, while Fig. 8d reveals the aftermath of the flood on 2019-04-16. Sentinel-2 true-color imagery is created using band 4 (red), band 3 (green), and band 2 (blue). After a flood, sediment and turbidity are often visible, with muddy water observed in affected areas. Fig. 8e shows the precipitation on 2019-04-01, with Poldakhtar city experiencing the highest precipitation at 142 mm, while the annual average precipitation in the stations of Lorestan province is 100 mm, indicating a significant volume of precipitation. Stations close to the flood location recorded higher precipitation. The results are consistent with Mehrabi (2021) who monitored the Lorestan

flood using Sentinel-1 products.

4.2. Discussion: Evaluation of the LSTM performance

To evaluate the performance of LSTM in three scenarios, both accuracy and number of iterations are investigated up to 250, as this threshold ensures a balance between computational efficiency and model convergence while preventing overfitting or underfitting. In this study, iteration refers to the number of sub-passes of the training data in each epoch (period), where the model parameters are updated. An epoch is defined as a full pass through the entire dataset. The number of iterations per epoch is determined by the batch size and the total number of training samples. The training process continues for a set number of iterations, or until the accuracy changes remain constant over the last 10 iterations, signaling that the model has reached the convergence stage. Fig. 9 illustrates the LSTM performance, showing that all three scenarios exhibit an increase in accuracy as the number of iterations increases. Scenario 2 achieves the highest accuracy, with a value of 97 %, while Scenario 3 exhibits the lowest accuracy at 71 %. Scenario 2 demonstrates the best fit, while Scenario 3 shows the worst fit. In Scenario 3, the accuracy decreases relative to the final iteration, indicating the presence of overfitting, after which the training process was terminated. Increased data and focus on extracting features from noisy data could be the cause (Song et al., 2019). The TWS products of LSTM-derived GRACE shows a mean significant correlation above 0.70 at the 95 % confidence level with both river discharge and precipitation in Lorestan flood while in original mode, after the flood, the correlation between precipitation and TWS changes reached 0.4. In time series, LSTM offers a better forecast by looking back and using sequences of data, and can learn TWS changes' behavior both long-term and short-term that can be useful for downscaling and forecasting, which is confirmed through the Lorestan flood case study. The disadvantages of deep learning include the need for large datasets to train the models effectively, which can be resource-intensive. The larger the training data set, the more accurate it can be. In small datasets, deep learning network optimization is required, and optimal prediction cannot be expected. Additionally, deep learning

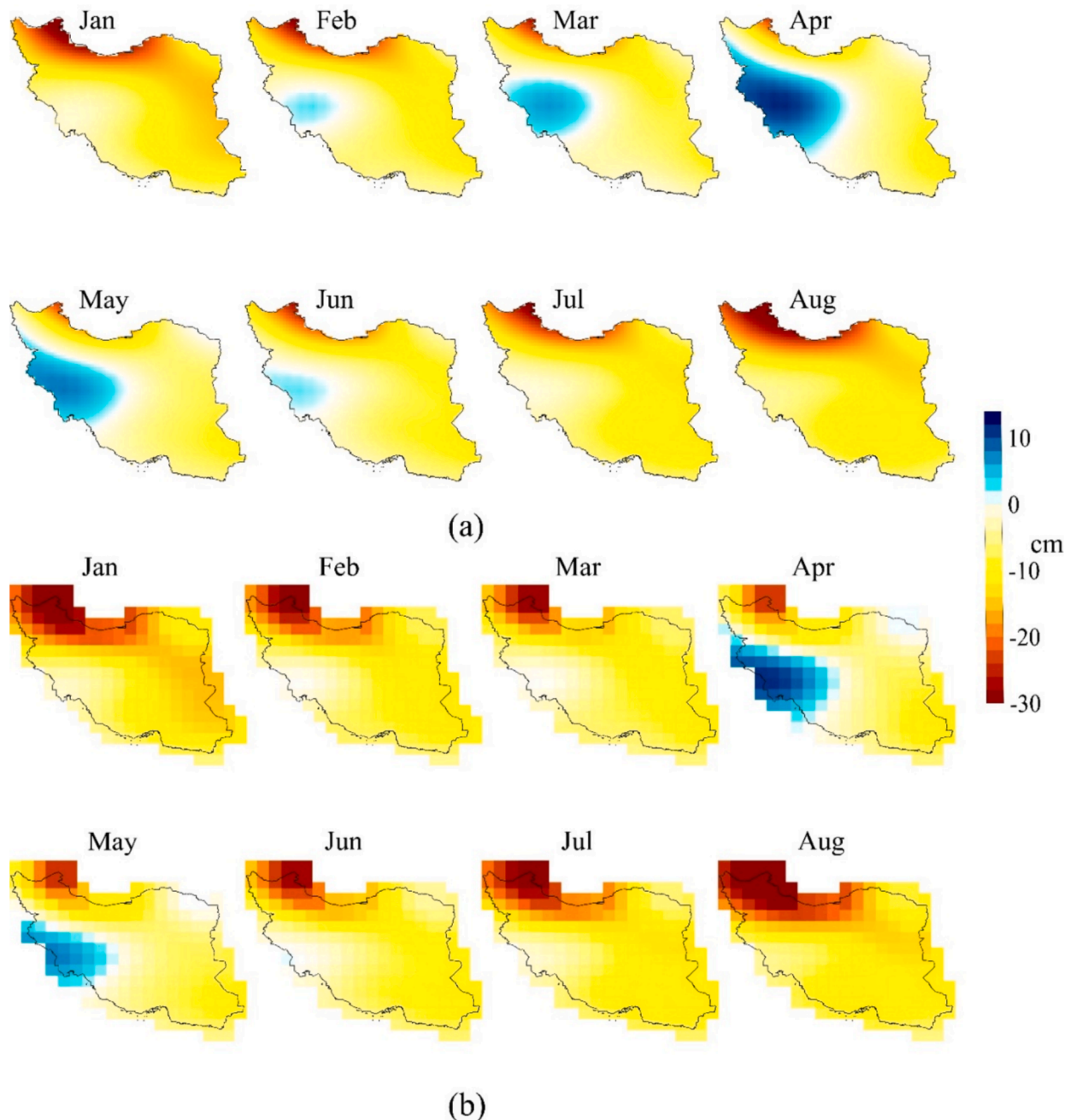


Fig. 6. A) TWS changes after LSTM downscaling of GRACE-FO's TWS from January to August 2019, and b), TWS changes without employing LSTM (mean of the three centers' products), where the brown areas indicate decrease while the blue areas show increase.

models often suffer from a lack of interpretability, making it difficult to understand how they reach their predictions or decisions.

GLDAS alone does not achieve high accuracy in capturing extreme hydrological events like floods, as demonstrated in this case study. While GLDAS provides reliable general trends, its reliance on modeled data often leads to the loss of localized signals, such as those associated with flooding (Rodell et al., 2004). This research aims to address this limitation by integrating GRACE/GRACE-FO observations with GLDAS to improve the accuracy, temporal resolution, and spatial resolution of TWS changes.

Several state-of-the-art downscaling approaches were reviewed, including statistical methods like partial least squares (Vishwakarma et al., 2021) and machine learning techniques such as Random Forest and Support Vector Machines (Seyoum et al., 2019; Zhong et al., 2021). This research demonstrates that the LSTM-based downscaling approach

achieves a significantly lower RMSE of ± 15 mm, outperforming other state-of-the-art methods. In comparison, the statistical approach by Vishwakarma et al. (2021) resulted in an RMSE of approximately ± 30 mm, while the machine learning method by Seyoum et al. (2019) had a much higher RMSE of ± 306 mm. These findings highlight the superior accuracy and efficiency of the LSTM model in reducing errors and improving correlation with independent sensor data. The results show that the LSTM approach consistently outperforms other methods in terms of reducing RMSE and improving the correlation with independent sensors.

One of the parameters that plays a crucial role in causing floods is soil moisture and ground water storage, which is a component of TWS. In this specific case study, there was no significant background of groundwater flooding, which is why groundwater storage was not thoroughly analyzed.

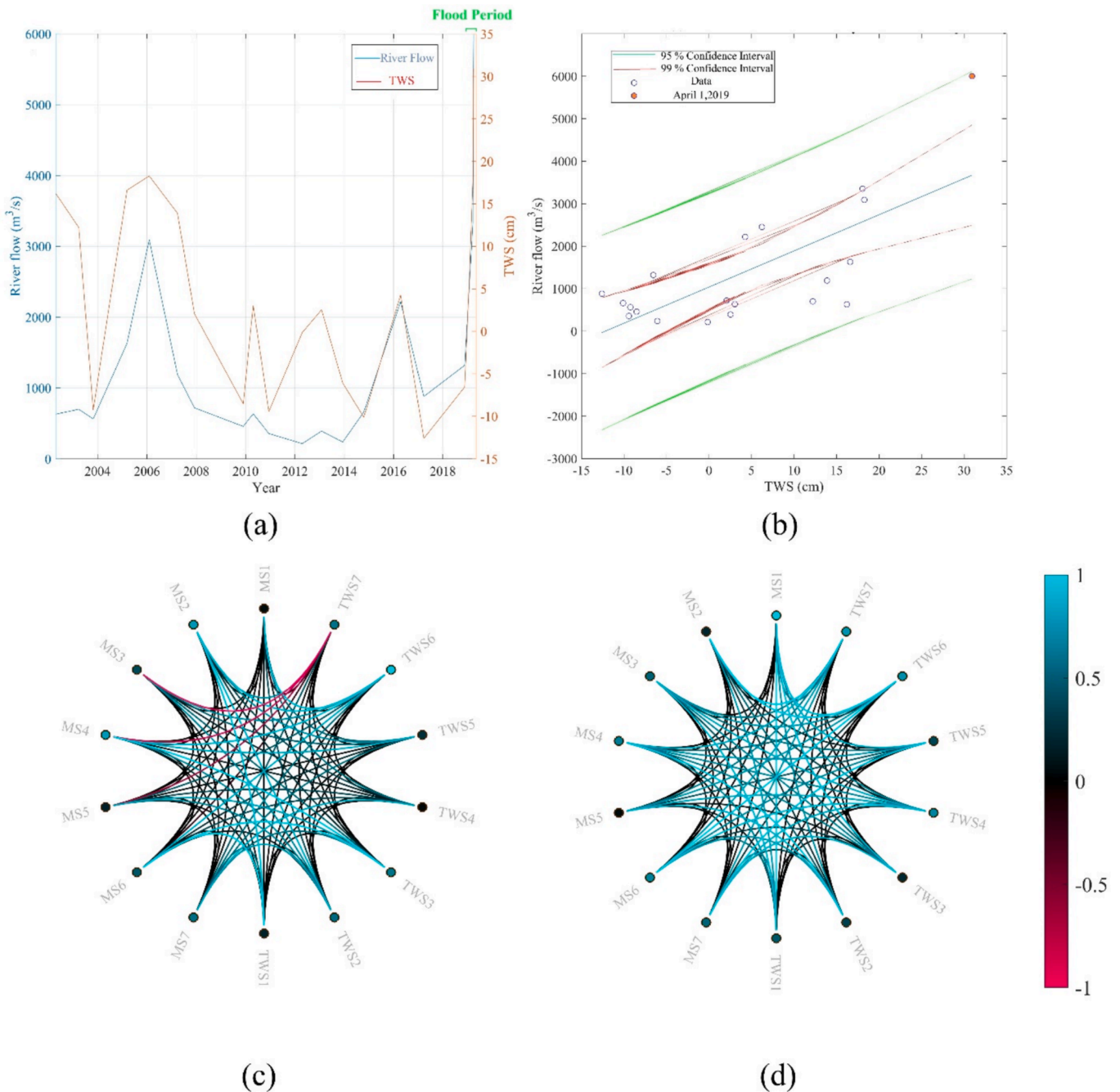


Fig. 7. A) TWS changes and Kashkan river flow time series, b) regression of TWS changes and river flow with confidence intervals (i.e., a correlation of 0.72 at 99% confidence level), c) precipitation correlation of 7 meteorological stations and TWS changes before the occurrence of the flood (i.e., an average correlation of 0.4 at 99% confidence level), and d), after the occurrence of the flood (i.e., an average correlation of 0.71 at 99% confidence level).

Future studies could focus on integrating additional datasets, such as soil moisture data from the soil moisture and ocean salinity (SMOS) satellite, alongside GRACE and GRACE-FO data, to further enhance flood prediction accuracy. By analyzing soil moisture and groundwater storage changes leading up to and during flood events, it will be possible to better understand the saturation levels of soil and groundwater, which can limit their ability to absorb further precipitation, potentially triggering floods. Additionally, hybrid modeling approaches that combine multiple datasets, such as TWS, soil moisture, and groundwater data, could be explored to reduce uncertainties in TWS estimates. These models could more accurately capture the complex interactions within the hydrological cycle, especially in the context of climate change, which is expected to intensify the frequency and severity of both

droughts and floods in the future (Yin et al., 2023a; Yin et al., 2023b; Yin et al., 2023c). Leveraging GRACE's long-term observational data, coupled with other relevant datasets, could further improve flood prediction models and reduce uncertainties in TWS estimates.

In future research, hybrid machine learning approaches, such as combining convolutional neural networks (CNNs) with LSTM, could be explored to enhance flood prediction capabilities. CNNs are effective at capturing spatial patterns, while LSTMs excel at modeling temporal dependencies. By integrating these two models, a hybrid approach could leverage both spatial and temporal features to more accurately capture flood events, providing a more comprehensive solution for flood monitoring and prediction. This approach holds the potential to improve predictive accuracy by utilizing the strengths of both methods.

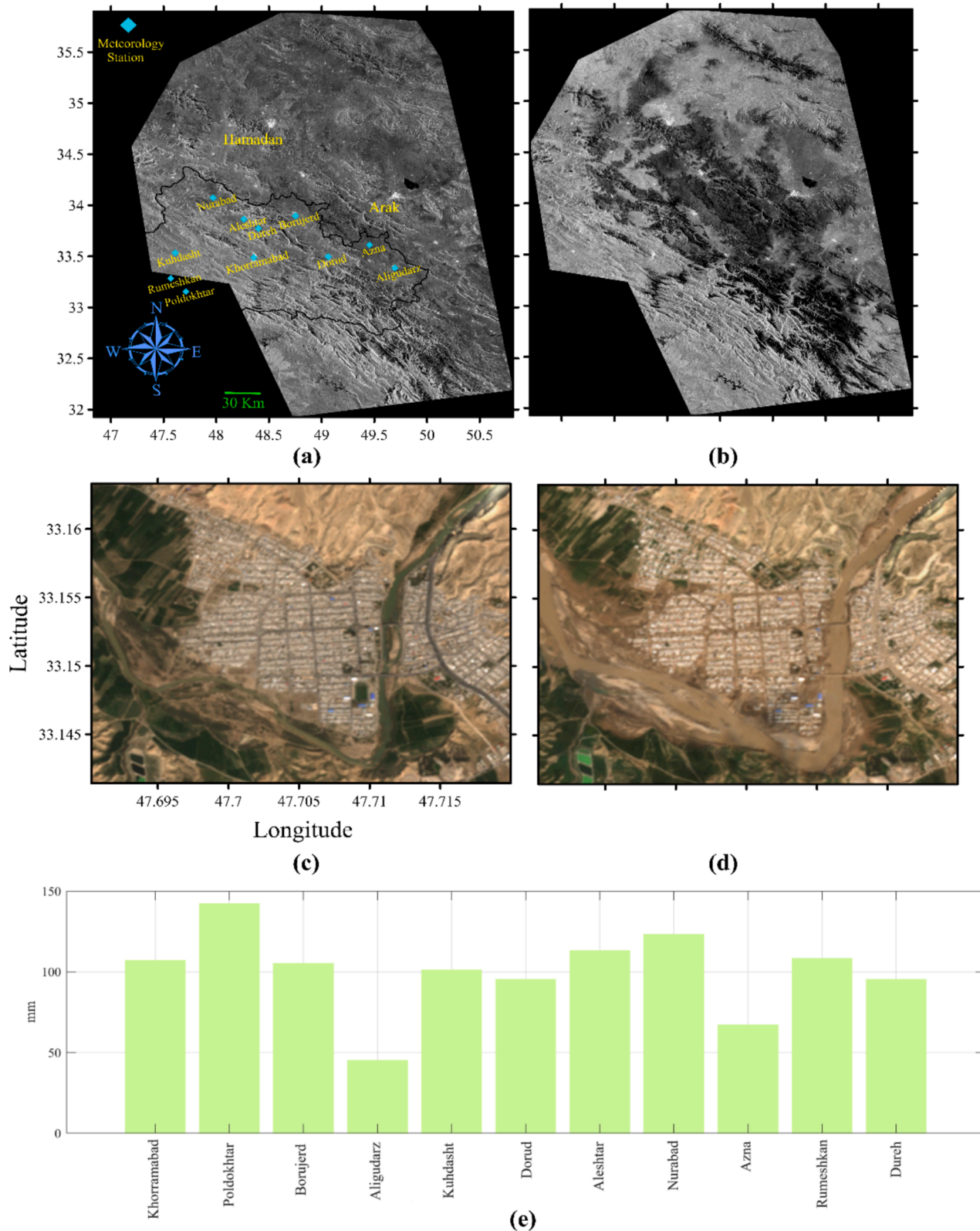


Fig. 8. (a) Sentinel-1 radar image of southwestern Iran on 2019-03-14 before the flood. (b) Sentinel-1 radar image on 2019-03-26 after the flood, with black areas indicating surface water and snow cover due to flooding and heavy precipitation. (c) Sentinel-2 true-color image of Poldokhtar on 2019-03-26 before the flood. (d) Sentinel-2 true-color image on 2019-04-16 after the flood, showing sediment and turbidity in flooded areas. (e) Precipitation on 2019-04-01, indicating Poldokhtar received 142 mm of rainfall, exceeding the annual average.

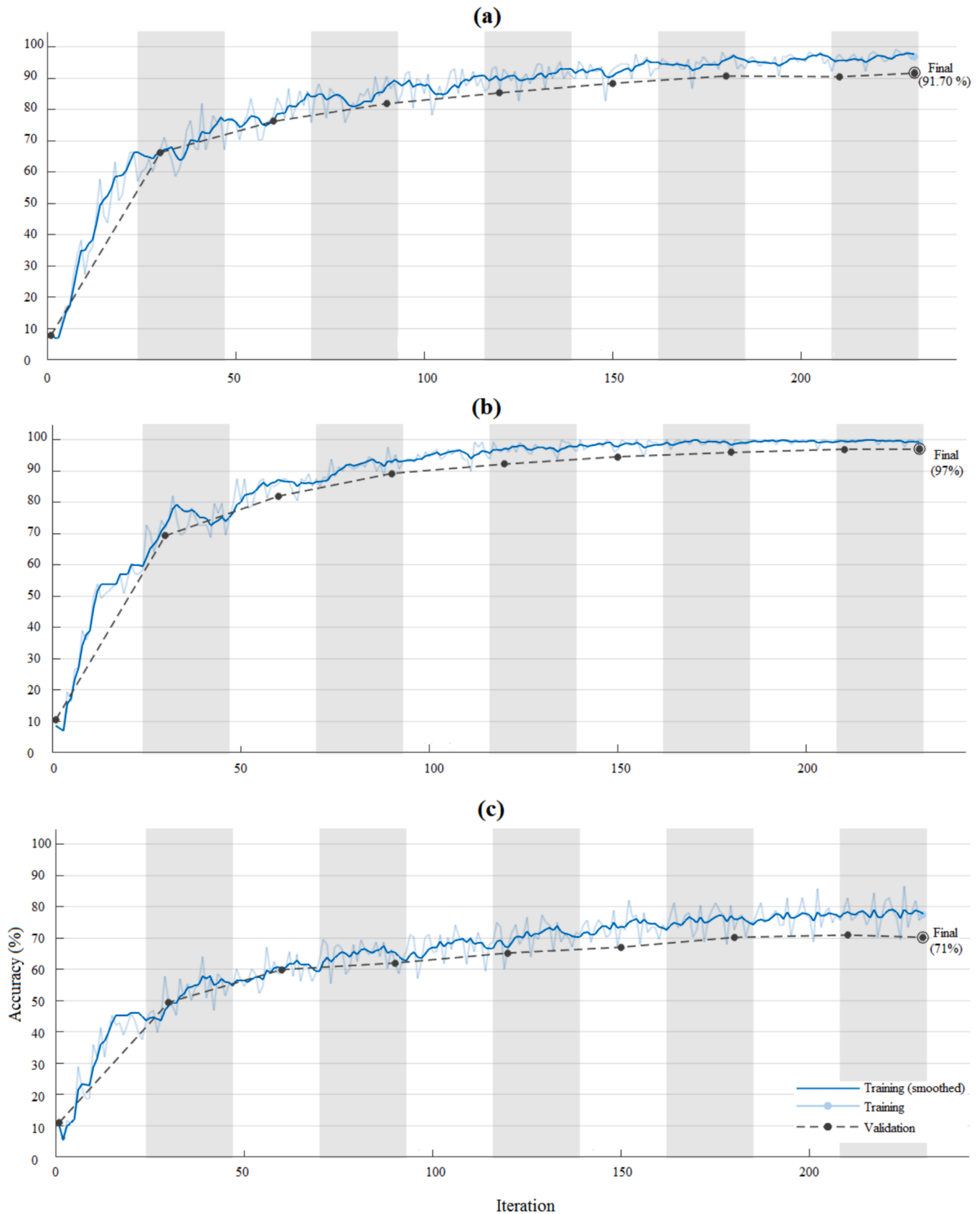


Fig. 9. LSTM performance a) scenario 1, b) scenario 2, and c), scenario 3. The accuracy of LSTM in each scenario increases with the increase in iterations, however the second scenario is seen to be the most accurate.

5. Conclusion

This study evaluated the capability of LSTM to downscale TWS changes. With its memory capability, LSTM can capture TWS changes based on its history. This allows for a better and more accurate understanding of the flood process. The proposed method can achieve enhanced spatial and temporal resolution by integrating observations with other sensors.

Based on the Lorestan flood case study, changes in the annual amplitude of TWS after downscaling are affected by the amount of precipitation. In February 2019, positive TWS is measured in the western regions of Iran. This is an anomaly that is significantly different from the rest of Iran. In April, this anomaly intensifies and its value increases 1.5 times. In March, the amount of TWS tripled compared to February, and its extent covers almost the entire west of Iran. Torrential rains occurred in late March and early April 2019, with a TWS leap of about 40 ± 2 cm. The TWS of GRACE-FO and LSTM has a mean significant correlation above 0.70 at the 95 % confidence level with river discharge and precipitation. The GRACE-FO and LSTM-derived GRACE have been able to clearly detect the amount of TWS leap in the month before the flood so that it was statistically identified as outlier. But one of its strongest signals was revealed in the western region of Iran in April and March 2019. A significant increase in TWS a few months before the flood, it indicated the occurrence of flood in the region. LSTM was able to model the spatial and temporal behavior of TWS with 97 % accuracy.

Subsequent studies can be useful in identifying TWS behavioral patterns to predict floods. The complex formation of the flood phenomenon makes it difficult to predict strongly. TWS downscaling can be used to predict floods in two aspects. In the first aspect, the incremental changes of TWS over several months, such as western Iran and the occurrence of floods, can be examined. In the second aspect, by examining the multi-year average of TWS and its sudden increase, the prognosis can be examined. What is important is the role of TWS in flood prediction and its relationship, showing the roadmap, and tools such as deep learning can use TWS LSTM.

CRediT authorship contribution statement

Omid Memarian Sorkhabi: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Joseph Awange:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration.

Declaration of competing interest

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.

Data availability

Data will be made available on request.

References

- Alhussein, M., Aurangzeb, K., Haider, S.I., 2020. Hybrid CNN-LSTM model for short-term individual household load forecasting. *IEEE Access* 8, 180544–180557.
- Arshad, A., Mirchi, A., Samimi, M., Ahmad, B., 2022. Combining downscaled-GRACE data with SWAT to improve the estimation of groundwater storage and depletion variations in the irrigated Indus basin. *Science of the Total Environment*.
- Awange, J., 2022. GHA's Greatest Freshwater Source: Victoria. In *Food Insecurity & Hydroclimate in Greater Horn of Africa*. Springer, Cham, pp. 85–105.
- Awange, J., Kiema, J., 2019. Water resources. In: *Environmental Geoinformatics*. Springer, Cham, pp. 431–468.

- Boergens, E., Kvas, A., Eicker, A., Dobslaw, H., Schawohl, L., Dahle, C., Flechtner, F., 2022. Uncertainties of GRACE-Based Terrestrial Water Storage Anomalies for Arbitrary Averaging Regions. *J. Geophys. Res. Solid Earth* 127 (2), e2021JB022081.
- Chen, L., He, Q., Liu, K., Li, J., Jing, C., 2019. Downscaling of GRACE-derived groundwater storage based on the random forest model. *Remote Sens. (Basel)* 11 (24), 2979.
- Chinnasamy, P., 2017. Inference of basin flood potential using nonlinear hysteresis effect of basin water storage: case study of the Koshi basin. *Hydrol. Res.* 48 (6), 1554–1565.
- Gohar, T., Hasan, L., Khan, G.M., Mubashir, M., 2022. In: March). Constraint Free Early Warning System for Flood Using Multivariate LSTM Network. *IEEE*, pp. 64–70.
- Gouweleeuw, B.T., Kvas, A., Gruber, C., Gain, A.K., Mayer-Gürr, T., Flechtner, F., Güntner, A., 2018. Daily GRACE gravity field solutions track major flood events in the Ganges-Brahmaputra Delta. *Hydrol. Earth Syst. Sci.* 22 (5), 2867–2880.
- Gupta, D., Dhanya, C.T., 2020. The potential of GRACE in assessing the flood potential of Peninsular Indian River basins. *Int. J. Remote Sens.* 41 (23), 9009–9038.
- He, H., Yang, K., Wang, S., Petrosians, H.A., Liu, M., Li, J., Li, J., 2021. Deep Learning Approaches to Spatial Downscaling of GRACE Terrestrial Water Storage Products Using EALCO Model Over Canada. *Can. J. Remote. Sens.* 47 (4), 657–675.
- Hu, J., Zhou, Z., Wang, J., Qin, F., Wang, J., Zhang, R., Huang, L., 2024. Enhancing the groundwater storage estimates by integrating MT-InSAR, GRACE/GRACE-FO, and hydraulic head measurements in Henan Plain (China). *Int. J. Appl. Earth Obs. Geoinf.* 131, 103993.
- Idowu, D., Zhou, W., 2019. Performance evaluation of a potential component of an early flood warning system—A case study of the 2012 flood, lower Niger River basin. *Nigeria. remote Sensing* 11 (17), 1970.
- Kao, I.F., Zhou, Y., Chang, L.C., Chang, F.J., 2020. Exploring a Long Short-Term Memory based Encoder-Decoder framework for multi-step-ahead flood forecasting. *J. Hydrol.* 583, 124631.
- Khain, P., Levi, Y., Shtivelman, A., Vadislavsky, E., Brainin, E., Stav, N., 2020. Improving the precipitation forecast over the Eastern Mediterranean using a smoothed time-lagged ensemble. *Meteorol. Appl.* 27 (1), e1840.
- Khorrami, B., Fistikoglu, O., Gunduz, O., 2022. A systematic assessment of flooding potential in a semi-arid watershed using GRACE gravity estimates and large-scale hydrological modeling. *Geocarto Int.* 37 (26), 11030–11051.
- Khorrami, B., Pirasteh, S., Ali, S., Sahin, O.G., Vaheddost, B., 2023. Statistical downscaling of GRACE TWSA estimates to a 1-km spatial resolution for a local-scale surveillance of flooding potential. *J. Hydrol.* 624, 129929.
- Landerer, F.W., Swenson, S.C., 2012. Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resour. Res.* 48 (4).
- Le, X.H., Ho, H.V., Lee, G., Jung, S., 2019. Application of long short-term memory (LSTM) neural network for flood forecasting. *Water* 11 (7), 1387.
- Li, W., Kiaghadi, A., Dawson, C., 2021. Exploring the best sequence LSTM modeling architecture for flood prediction. *Neural Comput. & Applic.* 33 (11), 5571–5580.
- Lin, J.C.W., Shao, Y., Djenouri, Y., Yun, U., 2021. ASRNN: a recurrent neural network with an attention model for sequence labeling. *Knowl.-Based Syst.* 212, 106548.
- Ma, Z., Fok, H.S., Tenzer, R., Chen, J., 2024. A novel Slepian approach for determining mass-term sea level from GRACE over the South China Sea. *Int. J. Appl. Earth Obs. Geoinf.* 132, 104065.
- Man, Y., Yang, Q., Shao, J., Wang, G., Bai, L., Xue, Y., 2022. Enhanced LSTM model for daily runoff prediction in the upper huai river basin, china. *Engineering*.
- Martinis, S., Plank, S., Ćwik, K., 2018. The use of Sentinel-1 time-series data to improve flood monitoring in arid areas. *Remote Sens. (Basel)* 10 (4), 583.
- Mehrabi, A., 2021. Monitoring the Iran Pol-e-Dokhtar flood extent and detecting its induced ground displacement using sentinel 1 imagery techniques. *Nat. Hazards* 105 (3), 2603–2617.
- Molodtsova, T., Molodtsov, S., Kirilenko, A., Zhang, X., VanLooy, J., 2016. Evaluating flood potential with GRACE in the United States. *Nat. Hazards Earth Syst. Sci.* 16 (4), 1011–1018.
- Pang, Z., Niu, F., O'Neill, Z., 2020. Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons. *Renew. Energy* 156, 279–289.
- Pascolini-Campbell, M., Fisher, J.B., Reager, J.T., 2021. GRACE-FO and ECOSTRESS Synergies Constrain Fine-Scale Impacts on the Water Balance. *Geophys. Res. Lett.* 48 (15), e2021GL093984.
- Reager, J.T., Famiglietti, J.S., 2009. Global terrestrial water storage capacity and flood potential using GRACE. *Geophys. Res. Lett.* 36 (23).
- Reager, J.T., Thomas, B.F., Famiglietti, J.S., 2014. River basin flood potential inferred using GRACE gravity observations at several months lead time. *Nat. Geosci.* 7 (8), 588–592.
- Rodell, M., Houser, P.R., Jambor, U.E.A., Gottschalk, J., Mitchell, K., Meng, C.J., Toll, D., 2004. The global land data assimilation system. *Bull. Am. Meteorol. Soc.* 85 (3), 381–394.
- Sakinah, N., Tahir, M., Badriyah, T., Syarif, I., 2019. In: September). LSTM with Adam Optimization-Powered High Accuracy Preeclampsia Classification. *IEEE*, pp. 314–319.
- Seyoum, W.M., Kwon, D., Milewski, A.M., 2019. Downscaling GRACE TWSA data into high-resolution groundwater level anomaly using machine learning-based models in a glacial aquifer system. *Remote Sens. (Basel)* 11 (7), 824.
- Shah, D., Mishra, V., 2021. Strong influence of changes in terrestrial water storage on flood potential in India. *J. Geophys. Res. Atmos.* 126 (1), e2020JD033566.
- Sherstinsky, A., 2020. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D* 404, 132306.
- Song, T., Ding, W., Wu, J., Liu, H., Zhou, H., Chu, J., 2019. Flash flood forecasting based on long short-term memory networks. *Water* 12 (1), 109.

- Sorkhabi, O., Asgari, J., Randhir, T.O., 2023. Monitoring groundwater storage based on satellite gravimetry and deep learning. *Nat. Resour. Res.* 32 (3), 1007–1020.
- Su, H., Hu, Y., Karimi, H.R., Knoll, A., Ferrigno, G., De Momi, E., 2020. Improved recurrent neural network-based manipulator control with remote center of motion constraints: Experimental results. *Neural Netw.* 131, 291–299.
- Tapley, B.D., Watkins, M.M., Flechtner, F., Reigber, C., Bettadpur, S., Rodell, M., Velicogna, I., 2019. Contributions of GRACE to understanding climate change. *Nat. Clim. Chang.* 9 (5), 358–369.
- Verma, K., Katpatal, Y.B., 2020. Groundwater monitoring using GRACE and GLDAS data after downscaling within basaltic aquifer system. *Groundwater* 58 (1), 143–151.
- Vishwakarma, B.D., Zhang, J., Sneeuw, N., 2021. Downscaling GRACE total water storage change using partial least squares regression. *Sci. Data* 8 (1), 1–13.
- Wang, F., Chen, Y., Li, Z., Fang, G., Li, Y., Wang, X., Kayumba, P.M., 2021. Developing a long short-term memory (Lstm)-based model for reconstructing terrestrial water storage variations from 1982 to 2016 in the tarim river basin, northwest china. *Remote Sens. (Basel)* 13 (5), 889.
- Wang, J., Shen, Y., Awange, J., Tabatabaeiasl, M., Song, Y., Liu, C., 2025. A novel generative adversarial network and downscaling scheme for GRACE/GRACE-FO products: Exemplified by the Yangtze and Nile River Basins. *Sci. Total Environ.* 969, 178874.
- Wang, J., Shen, Y., Awange, J.L., Yang, L., 2023. A deep learning model for reconstructing centenary water storage changes in the Yangtze River Basin. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2023.167030>.
- Widiasari, I.R., Nugoho, L.E., Efendi, R., 2018. In: September). *Context-Based Hydrology Time Series Data for a Flood Prediction Model Using LSTM*. IEEE, pp. 385–390.
- Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Res.* 30 (1), 79–82.
- Wu, Y., Ding, Y., Zhu, Y., Feng, J., & Wang, S. (2020). Complexity to forecast flood: Problem definition and spatiotemporal attention LSTM solution. *Complexity*, 2020.
- Wu, Y., Han, P., Zheng, Z., 2021. Instant water body variation detection via analysis on remote sensing imagery. *J. Real-Time Image Proc.* 1–14.
- Wu, Y., Liu, Z., Xu, W., Feng, J., Palaiahnakote, S., Lu, T., 2018. Context-aware attention LSTM network for flood prediction. In: 2018 24th International Conference on Pattern Recognition (ICPR). <https://doi.org/10.1109/icpr.2018.8545385>.
- Xie, J., Xu, Y.P., Yu, H., Huang, Y., Guo, Y., 2022a. Downscaling and monitoring the extreme flood events in the Yangtze River Basin based on GRACE/GRACE-FO satellites data. *Hydrol. Earth Syst. Sci. Discuss.* 1–39.
- Xie, J., Xu, Y.P., Yu, H., Huang, Y., Guo, Y., 2022b. Monitoring the extreme flood events in the Yangtze River basin based on GRACE and GRACE-FO satellite data. *Hydrol. Earth Syst. Sci.* 26 (22), 5933–5954.
- Xiong, J., Wang, Z., Guo, S., Wu, X., Yin, J., Wang, J., Gong, Q., 2022. High effectiveness of GRACE data in daily-scale flood modeling: case study in the Xijiang River Basin. *China. Natural Hazards* 1–20.
- Xu, Y., Gong, H., Chen, B., Zhang, Q., Li, Z., 2021. Long-term and seasonal variation in groundwater storage in the North China Plain based on GRACE. *Int. J. Appl. Earth Obs. Geoinf.* 104, 102560.
- Yang, T., Sun, F., Gentine, P., Liu, W., Wang, H., Yin, J., Liu, C., 2019. Evaluation and machine learning improvement of global hydrological model-based flood simulations. *Environ. Res. Lett.* 14 (11), 114027.
- Yin, J., Gentine, P., Slater, L., Gu, L., Pokhrel, Y., Hanasaki, N., Schlenker, W., 2023a. Future socio-ecosystem productivity threatened by compound drought–heatwave events. *Nat. Sustain.*, 6, 259–272.
- Yin, J., Guo, S., Wang, J., Chen, J., Zhang, Q., Gu, L., Zhang, Y., 2023b. Thermodynamic driving mechanisms for the formation of global precipitation extremes and ecohydrological effects. *Sci. China Earth Sci.* 66 (1), 92–110.
- Yin, J., Slater, L.J., Khouakhi, A., Yu, L., Liu, P., Li, F., Gentine, P., 2023c. GTWS-MLrec: global terrestrial water storage reconstruction by machine learning from 1940 to present. *Earth Syst. Sci. Data Discuss.* 2023, 1–29.
- Yu, Q., Wang, S., He, H., Yang, K., Ma, L., Li, J., 2021. Reconstructing GRACE-like TWS anomalies for the Canadian landmass using deep learning and land surface model. *Int. J. Appl. Earth Obs. Geoinf.* 102, 102404.
- Zhang, J., Liu, K., Wang, M., 2023. Flood detection using Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage and extreme precipitation data. *Earth Syst. Sci. Data* 15, 521–540.
- Zhong, D., Wang, S., Li, J., 2021. A Self-Calibration Variance-Component Model for Spatial Downscaling of GRACE Observations Using Land Surface Model Outputs. *Water Resour. Res.* 57 (1), e2020WR028944.