



# Leveraging GNSS tropospheric products for machine learning-based land subsidence prediction

Melika Tasan<sup>1</sup> · Zahrasadat Ghorbaninasab<sup>2</sup> · Saeid Haji-Aghajany<sup>3</sup> · Alireza Ghiasvand<sup>2</sup>

Received: 29 June 2023 / Accepted: 27 October 2023 / Published online: 7 November 2023  
© The Author(s) 2023

## Abstract

Land subsidence is a hazardous phenomenon that requires accurate prediction to mitigate losses and prevent casualties. This study explores the utilization of the Long Short-Term Memory (LSTM) method for time series prediction of land subsidence, considering various contributing factors such as groundwater levels, soil type and slope, aquifer characteristics, vegetation coverage, land use, depth to the water table, proximity to exploiting wells, distance from rivers, distance from faults, temperature, and wet tropospheric products. Due to the high spatial variability of wet tropospheric parameters, utilizing numerical weather models for extraction is impractical, especially in regions with a sparse network of synoptic stations. This hinders obtaining accurate prediction results because wet tropospheric products play a significant role in subsidence prediction and cannot be ignored in the subsidence prediction process. In this study, Global Navigation Satellite Systems (GNSS) tropospheric products, including Integrated Water Vapor (IWV) and EvapoTranspiration (ET), are employed as alternatives. Two scenarios were considered: one incorporating GNSS products alongside other parameters, and the other relying solely on the remaining parameters in the absence of GNSS tropospheric products. Ground truth data from Interferometric Synthetic Aperture Radar (InSAR) displacement measurements were used for evaluation and testing. The results demonstrated that the inclusion of GNSS tropospheric products significantly enhanced prediction accuracy, with a Root Mean Square Error (RMSE) value of 3.07 cm/year in the first scenario. In the second scenario, the absence of wet tropospheric information led to subpar predictions, highlighting the crucial role of wet tropospheric data in spatial distribution. However, by utilizing tropospheric products obtained from GNSS observations, reasonably accurate predictions of displacement changes were achieved. This study underscores the importance of tropospheric indices and showcases the potential of the LSTM method in conjunction with GNSS observations for effective land subsidence prediction, enabling improved preventive measures and mitigation strategies in regions lacking synoptic data coverage.

**Keywords** GNSS · InSAR · LSTM · Prediction · Subsidence

## Introduction

Land subsidence refers to the gradual settling or sudden sinking of the ground surface (Hu et al. 2022). It can occur due to natural factors such as earthquakes, volcanoes, and landslides, or as a result of human activities like fluid extraction (Motagh et al. 2017; Sorkhabi et al. 2022). One of the primary causes of land subsidence is excessive groundwater extraction, which has had significant impacts on both urban and agricultural areas worldwide (Shrestha et al., 2017; Ding et al. 2020; Sorkhabi et al. 2022; Haji-Aghajany et al. 2023). Iran heavily relies on groundwater as a crucial source for drinking water, industrial processes, and agricultural needs, particularly in arid and semi-arid regions (Motagh et al. 2008). However, unsustainable agricultural

---

Communicated by H. Babaie.

✉ Saeid Haji-Aghajany  
saeid.haji-aghajany@upwr.edu.pl

- <sup>1</sup> Department of Civil Engineering, Faculty of Environmental Engineering and Geodesy, Wrocław University of Environmental and Life Sciences, Wrocław 50-363, Poland
- <sup>2</sup> Department of Surveying Engineering, South Tehran Branch, Islamic Azad University, Tehran 1777613511, Iran
- <sup>3</sup> Institute of Geodesy and Geoinformatics, Wrocław University of Environmental and Life Sciences, Norwida 25, Wrocław 50-375, Poland

practices and the overexploitation of groundwater resources lead to increased stress on aquifers, resulting in a decline in the water table level, aquifer compaction, and ultimately land subsidence. Various regions in Iran, including Tehran (Mahmoudpour et al. 2016), Hamedan (Khanlari et al. 2012), Estahban (Sorkhabi et al. 2022), Zarand (Motagh et al. 2008), Mashhad (Akbari and Motagh 2012), Kurdistan (Haji-Aghajany and Amerian 2020), Yazd (Motagh et al. 2008), Tabriz (Haji-Aghajany et al. 2017), Mahyar (Davoodijam et al. 2015), Qazvin (Farshbaf et al. 2023), Rafsanjan (Motagh et al. 2017), Neyshabour (Dehghani et al. 2009), Zanjan (Farshbaf et al. 2023) and Kashmar (Anderssohn et al. 2008) have reported instances of land subsidence. Land subsidence can cause severe damage to various types of infrastructure, including residential areas, transportation networks, industrial zones, and agricultural lands. Additionally, it has indirect consequences such as increased flooding and alterations in soil-vegetation properties (Mohseni et al. 2017; Andreas et al. 2018). Unfortunately, land subsidence is often irreversible, making it challenging to fully mitigate the damage it inflicts. Therefore, it is crucial to conduct thorough studies, predict potential changes, and implement preventive measures to effectively control and manage land subsidence in susceptible regions.

In previous studies, researchers have focused on land subsidence mechanisms and utilized established constitutive models and numerical simulation models to predict displacements (Azarakhsh et al. 2022). However, accurately representing hydrological and climate information, which can be limited in space and time, is crucial for precise prediction. This limitation hinders the application of these models to large areas. To address this challenge, alternative approaches such as the Grey Model (GM) based on grey theory have been explored for short-term land subsidence prediction (Li et al. 2007). However, the GM model does not effectively capture the non-linear characteristics of land subsidence. To overcome this limitation, some researchers have proposed modified GM models combined with Artificial Neural Networks (ANNs) or other algorithms to account for non-linear features (Li et al. 2007; Wang et al. 2023). Although these methods show promise in short-term predictions and perform well with smaller datasets, they struggle to leverage deep information when dealing with large datasets or making long-term predictions. To tackle the challenges associated with land subsidence prediction, researchers have increasingly turned to machine learning techniques, which have demonstrated significant potential in various geospatial applications. Machine learning models excel at analyzing vast amounts of complex data and discovering patterns that traditional statistical approaches may miss. Among these techniques, Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM)

networks, have received considerable attention due to their ability to capture temporal dependencies in sequential data (Elman 1990; Hochreiter and Schmidhuber 1997; Lipton et al. 2015). LSTM networks are particularly well-suited for time series analysis, as they can effectively retain and exploit long-term dependencies. By constructing a multi-layer neural network, the LSTM model can extract temporal dynamic features from historical data, considering both non-linearity and temporal dependencies. The LSTM model has already shown successful applications in forecasting PM<sub>2.5</sub> concentration, which is a spatiotemporal phenomenon (Qi et al. 2019). The length of the prediction period depends on the time interval of the input data. In the context of land subsidence prediction, time series data plays a crucial role in capturing the dynamic behavior of the subsidence process.

Hydrological, geological, and environmental information reflects the behavior of subsidence. Additionally, Interferometric Synthetic Aperture Radar (InSAR) subsidence measurements offer high-resolution deformation maps, enabling the detection and monitoring of surface movements with millimeter-level accuracy (Berardino et al. 2002; Ferretti et al. 2007). By integrating these diverse data sources, a more comprehensive understanding of the subsidence phenomenon can be achieved. Furthermore, climate parameters, including temperature, precipitation, and humidity, have been recognized as important factors influencing land subsidence (Faunt et al. 2016; Haji-Aghajany et al. 2023). The necessary meteorological data can be divided into two different parts: the first part includes temperature, and the second part includes wet parameters such as precipitation and humidity. When it comes to temperature, if a dense network of synoptic stations is available, the interpolation method can be used to reconstruct a high-resolution temperature map with suitable accuracy. However, if a dense network is not available, a numerical weather model could be used instead. Generally, the spatial resolution of a numerical weather model grid is sufficient for interpolating temperature data due to its predictable behavior. Unlike temperature, wet parameters exhibit irregular behavior. Therefore, when a dense synoptic network is lacking, it is not advisable to rely on numerical weather models. This is because their resolution is not sufficient for accurately modeling the spatial variation of wet parameters through interpolation (Haji-Aghajany 2021). Global Navigation Satellite Systems (GNSS) technology can provide additional tropospheric products, which are highly accurate measures of atmospheric wet content (Dach et al. 2020; Haji-Aghajany 2021; Agnihotri et al. 2021; Haji-Aghajany et al. 2022). Incorporating these tropospheric parameters derived from a dense GNSS station network into predictive models can enhance the accuracy of subsidence predictions, particularly in regions with limited meteorological observations.

This study aims to check the effect of using GNSS wet tropospheric products on land subsidence prediction using machine learning, with a specific focus on LSTM networks. The proposed model will utilize GNSS tropospheric products, hydrological, geological and environmental information and InSAR measurements. By training the model on historical data and relevant environmental variables, it will be possible to predict future subsidence patterns and assess the impact of GNSS products on subsidence prediction. The results of this research can provide valuable insights for land subsidence mitigation strategies and support decision-making processes to safeguard vulnerable regions from the detrimental effects of subsidence. The rest of this paper is structured as follows: The next sections discuss the methodology and the process of obtaining training data. Subsequent sections delve into validation approaches and provide insights into the study area. Finally, the paper concludes by presenting the results and engaging in a discussion.

## Methodology

### Machine learning for land subsidence prediction

Machine learning has brought about a revolutionary transformation in the field of predictive modeling, enabling us to extract valuable insights and make accurate predictions based on complex data patterns. Among the many machine learning algorithms available, ANNs and RNNs have emerged as powerful tools for predictive analytics. ANNs consist of interconnected nodes, known as “neurons,” organized into layers. Each neuron performs a weighted sum of its inputs, applies an activation function to the result, and passes the output to the next layer. Through a process called backpropagation, the network’s parameters, including weights and biases, are learned. RNNs, on the other hand, are specialized neural networks designed to model sequential data, making them particularly well-suited for time-series forecasting and natural language processing tasks (Graves 2013). Unlike ANNs, RNNs incorporate a feedback loop that allows information to persist across time steps. This recurrent structure enables the network to capture dependencies and long-term patterns in sequential data. RNNs offer several advantages over ANNs in the context of predictive modeling. Firstly, RNNs excel at capturing temporal dependencies in sequential data, making them suitable for applications such as speech recognition, machine translation, and sentiment analysis (Lipton et al. 2015). Secondly, unlike ANNs, which require fixed-size inputs, RNNs can handle variable-length sequences, providing flexibility when dealing with data of varying lengths (Lipton et al. 2015). Thirdly, RNNs can maintain an internal

memory or hidden state that summarizes the context of the input sequence (Graves 2013). This contextual information enables the network to understand and remember relevant information across time steps, thereby enhancing the quality of predictions. Finally, RNNs do not make strong assumptions about the relationship between input and output, instead, they learn patterns directly from the data, reducing the need for manual feature engineering and making them adaptable to different problem types (Goodfellow et al. 2016). The most common variation of RNN is the LSTM network, which addresses the vanishing gradient problem and improves the modeling of long-range dependencies (Hochreiter and Schmidhuber 1997).

The LSTM is a powerful type of RNN that excels at capturing long-term dependencies in sequential data (Hochreiter and Schmidhuber 1997). In recent years, LSTM has gained significant attention in various fields due to its remarkable ability to perform accurate long-term predictions (Sak et al. 2014). This paper explores the structure and capabilities of LSTM networks, specifically focusing on their effectiveness in handling different sets and types of data for long-term prediction tasks. LSTM networks are composed of memory cells, input gates, forget gates, and output gates, all working together to retain and selectively update information over time (Hochreiter and Schmidhuber 1997). The architecture of an LSTM cell allows it to overcome the vanishing gradient problem, which is a common challenge faced by traditional RNNs. At the core of an LSTM cell is the memory cell, which maintains and propagates information throughout the network. The input gate determines how much new information should be stored in the memory cell, while the forget gate decides what information should be discarded. The output gate controls the amount of information extracted from the memory cell for generating predictions. The LSTM’s ability to mitigate the vanishing gradient problem is attributed to its gating mechanisms (Hochreiter and Schmidhuber 1997). By selectively updating and forgetting information, LSTM cells can retain crucial details over long sequences, enabling them to capture dependencies that traditional RNNs struggle with. LSTM networks have demonstrated impressive performance in a wide range of long-term prediction tasks. Whether applied to natural language processing, time series analysis, or financial forecasting, LSTM consistently exhibits superior predictive capabilities (Greff et al. 2017). One reason for LSTM’s success is its ability to effectively model temporal dependencies over extended periods. The memory cells’ capacity to remember past information and selectively update it allows the network to capture complex patterns and trends in sequential data. This enables LSTM to make accurate predictions even when the input data spans a large time horizon. Moreover, LSTM networks can handle diverse sets and types of data for long-term prediction

tasks (Gers et al. 2000). They can accommodate continuous or discrete-valued time series data, textual data, and even multimodal inputs (Greff et al. 2017). The flexibility of LSTM architecture makes it adaptable to various domains, ranging from weather forecasting and stock market prediction to speech recognition and machine translation. Recent advancements in LSTM training methods, such as deep LSTM architectures and attention mechanisms, have further improved its performance in long-term prediction tasks. These techniques enable the network to focus on relevant information, enhance feature extraction, and capture intricate relationships within the data (Greff et al. 2017). In conclusion, LSTM networks have emerged as a powerful tool for long-term prediction tasks, overcoming the limitations of traditional RNNs. With their unique structure and mechanisms, LSTM models effectively address the vanishing gradient problem and excel at capturing long-term dependencies (Hochreiter and Schmidhuber 1997). Their flexibility in handling different sets and types of data, combined with recent advancements in training methods, has made LSTM a preferred choice for various prediction applications. As the field continues to evolve, further research into LSTM networks promises even greater strides in long-term prediction capabilities. The LSTM network structure differs from the single-loop body structure as it incorporates three gates: the forgetting gate, input gate, and output gate. LSTM networks have demonstrated significant success in numerous problem domains, particularly for tasks involving sequence modeling. The majority of current recurrent neural networks utilize the LSTM structure as the foundation. Fig. 1 depicts the basic architecture of the LSTM neural network. In the diagram, the symbol  $\sigma$  represents the sigmoid function, producing outputs ranging from 0 to 1. The function  $\tanh$  represents the hyperbolic tangent, generating outputs between  $-1$  and  $1$ . The term  $h_{t-1}$  denotes the output of the previous cell, while  $X_t$  represents the input of the current cell. The initial phase of the LSTM neural network determines whether to retain or forget information within

the cell state. The calculation formula for the forgetting gate, denoted as  $f_t$ , is presented as follows:

$$f_t = \sigma(W_f[h_{t-1}, X_t] + bf) \tag{1}$$

Here,  $\sigma$  is the sigmoid activation function,  $h_{t-1}$  represents the output at time  $t - 1$ , and  $X_t$  represents the input vector at time  $t$ . The weight vector and bias vector of the forget gate are represented by  $W_f$  and  $bf$ , respectively. If the output value of  $f_t$  is close to 0, it implies that the previous data has been forgotten. However, a value close to 1 does not necessarily indicate that the previous data has been retained. The subsequent stage of the LSTM neural network determines which new data is to be stored in the cell state. This involves two steps: first, the sigmoid layer determines the information to be updated, while the  $\tanh$  layer generates an alternative candidate value, which is subsequently added to the cell state. Second, by combining these two pieces of information, the model creates new values to update the cell state. In the third stage of the LSTM neural network, the previous cell state  $c_{t-1}$  is updated using the forgetting gate,  $f_t$ , and the input gate,  $i_t$ . The values  $c_{t-1}$  and  $i_t$  are multiplied by  $f_t$  to eliminate redundant information. The updated cell state,  $c_t$ , is obtained by updating the previous state  $c_{t-1}$ . The calculation is represented by:

$$c_t = f_t \cdot c_{t-1} + i_t \cdot c'_t \tag{2}$$

Finally, the last stage of the LSTM neural network determines the output. Initially, a sigmoid layer is employed to determine which portion of the cell state will be output. Subsequently, the cell state is processed through the  $\tanh$  function (producing a value between  $-1$  and  $1$ ) and multiplied by the output of the sigmoid gate. As a result, only the determined portion of the output is generated. The calculations for the new output value,  $h_t$ , the output gate,  $O_t$ , and

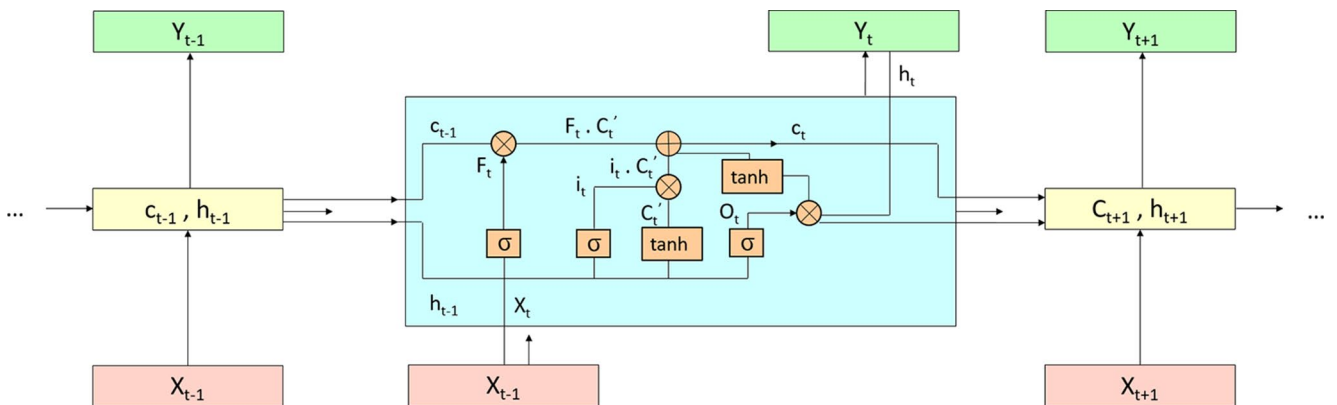


Fig. 1 The basic structure of LSTM method

the weight vector and bias vector of the output gate,  $W_o$  and  $b_o$ , respectively, are given by:

$$O_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o) \quad (3)$$

$$h_t = O_t \tanh(c_t) \quad (4)$$

### Different sources of data for machine learning

Like other machine learning approaches, training and validation play a crucial role in the effectiveness of LSTM networks. LSTM, a type of RNN, requires a robust training process with carefully curated training data to learn patterns and make accurate predictions. The training process of LSTM involves iteratively updating the network's parameters to minimize the discrepancy between its predictions and the ground truth. It requires a dataset consisting of input sequences and corresponding target outputs. LSTM networks leverage the concept of backpropagation through time (BPTT) to compute gradients and adjust the weights of the network (Gers et al. 2000). During training, the input sequences are fed to the LSTM network, and the resulting predictions are compared against the target outputs using a suitable loss function. The gradients are then computed, and the network's weights are updated using optimization algorithms such as stochastic gradient descent (SGD) or Adaptive Moment Estimation (Adam) (Kingma and Ba 2014). This iterative process continues until the network achieves satisfactory performance on the training data. The validation process is crucial for assessing the generalization ability of the LSTM network. After each training iteration or epoch, a separate validation set is employed to evaluate the network's performance on unseen data. The validation set helps in monitoring the network's progress and detecting overfitting, which occurs when the network memorizes the training data without effectively capturing underlying patterns (Goodfellow et al. 2016). During validation, the LSTM network takes the input sequences from the validation set and generates predictions. The predictions are compared against the corresponding target outputs, and performance metrics are computed. These metrics provide insights into how well the network generalizes to unseen data and helps determine if adjustments to the model architecture or training process are necessary. The choice of training data greatly influences the performance of an LSTM network. It is essential to curate a diverse and representative dataset that captures the patterns and variations present in the target problem domain. In this paper, various datasets are used for training, including GNSS tropospheric products, hydrological, geological, and environmental information. The InSAR displacement fields

are then utilized as validation and testing data to predict subsidence.

### InSAR displacement fields

InSAR time series analysis is a remote sensing technique based on radar acquisition and has emerged as a powerful method for monitoring ground deformations, such as subsidence, earthquakes, and volcanoes (Rosen et al. 2000; Haji-Aghajany et al. 2020). One commonly used method in InSAR time series analysis is the Small Baseline Subset (SBAS) technique. SBAS utilizes a set of interferograms acquired over a specific time period to estimate the ground deformation rates. The SBAS technique takes advantage of the temporal coherence of radar signals by selecting interferograms with small baselines, which minimizes the decorrelation effects caused by changes in the scene between acquisitions. By analyzing a series of these interferograms, it becomes possible to generate deformation maps and track subtle ground motion patterns over time. However, when dealing with interferograms, it is crucial to account for and remove additional effects apart from the desired displacement signal. InSAR interferograms are derived from the differential phase between two master and slave acquisitions. These observations contain different phase components at each point, which can be considered as follows (Hanssen 2001):

$$\begin{aligned} \Phi_{obs} &= \Phi_{Master} - \Phi_{Slave} \\ &= \Delta\Phi_{def} + \Delta\Phi_{top} + \Delta\Phi_{atm} + \Delta\Phi_{orb} \\ &\quad + \Delta\Phi_{flat} + \Delta\Phi_{noise} \end{aligned} \quad (5)$$

Where,  $\Delta\Phi_{def}^p$  is the phase change related to the ground displacement along the Line Of Sight (LOS) direction,  $\Delta\Phi_{top}^p$ ,  $\Delta\Phi_{atm}^p$ ,  $\Delta\Phi_{orb}^p$  and  $\Delta\Phi_{noise}^p$  are related to the topographic, atmospheric, orbital, and noise phase, respectively.  $\Delta\Phi_{flat}^p$  Corresponds to uncertainties in the Earth's ellipsoidal reference surface. To correct for orbital errors, precise satellite orbits are required. The orbital information is obtained from the satellite's navigation data or through the use of dedicated orbit determination techniques. Topographic variations can introduce phase differences in interferograms, leading to false deformation signals. To mitigate this effect, Digital Elevation Models (DEMs) are employed to remove the topographic component from the interferograms (Zebker and Villasenor 1992; Hanssen 2001). Atmospheric disturbances, mainly tropospheric and ionospheric delays, can significantly affect the interferometric phase. Generally, the effect of the ionospheric air refractivity on short wavelength InSAR is negligible (Gray et al. 2000). Tropospheric corrections can be estimated using the correlation between the delay and the topographic elevation (Remy et al. 2003)

stacking independent InSAR data (Peltzer et al. 2001), characterizing the statistical properties of phase delay patterns (Emardson et al. 2003) ray tracing techniques (Haji-Aghajany and Amerian 2018, 2020a; Haji-Aghajany et al. 2019; Khalili et al. 2023) or external data (Jolivet et al., 2014). By applying these correction methods, the interferograms can be cleaned from unwanted effects, enhancing the accuracy of the displacement measurements and improving the reliability of InSAR time series analysis.

**GNSS tropospheric products**

GNSS tropospheric products are data and models used to characterize and correct for the effects of the Earth’s atmosphere on GNSS signals. The use of GNSS products in climate and subsidence monitoring and prediction provides accurate and continuous data, enabling scientists and decision-makers to make informed assessments and take proactive measures.

Integrated Water vapor (IWV) is a tropospheric product that quantifies the total amount of water vapor in a vertical column of the atmosphere above a location (Bevis et al. 1992). IWV data obtained from GNSS can contribute to climate research by providing information on atmospheric water vapor trends and variability. Monitoring IWV variations over time helps researchers understand long-term changes in atmospheric moisture content and its implications for climate change (Bevis et al. 1992; Bock et al. 2003). Furthermore, IWV data from GNSS can aid in the study of subsidence, which refers to the gradual sinking or settling of the Earth’s surface. Changes in IWV can be indicative of subsidence, as the presence or absence of water vapor affects the moisture content and stability of underlying soil layers. GNSS-based IWV measurements can help monitor subsidence by detecting changes in the water vapor content of the atmosphere, providing valuable information for subsidence modeling and mitigation efforts (Davis et al. 2013; Tregoning and Ramillien 2014). The formula to calculate IWV from GNSS data using Zenith Tropospheric Delay (ZWD) is as follow:

$$IWV = 10^6 \frac{ZWD}{R_w} \left( k_2' + \frac{k_3}{T_M} \right)^{-1} \tag{6}$$

Where  $R_w$  is the specific gas constant for water vapor,  $k_2'$  and  $k_3$  are refraction constants and  $T_M$  is weighted mean water vapor temperature of the troposphere (Kleijer 2004). ZWD is the foundation for estimating of water vapor content in the troposphere. ZWD can be obtained by subtracting Zenith Hydrostatic Delay (ZHD) form Zenith Tropospheric Delay (ZTD):

$$ZWD = ZTD - ZHD \tag{7}$$

The estimation of the ZTD with high accuracy using GNSS processing techniques has been extensively studied (Bevis et al. 1992; Zhang et al. 2016). GNSS receivers measure the total delay experienced by satellite signals as they pass through the Earth’s atmosphere. Through analyzing phase and code measurements from multiple satellites, along with precise orbit and clock information, ZTD can be computed accurately (Bevis et al. 1992). To compute the ZHD, empirical models such as the Saastamoinen model are widely used (Saastamoinen 1972). These models take into account the surface pressure, to estimate the hydrostatic component of the ZTD (Saastamoinen 1972). These models have been extensively calibrated and provide high accuracy in estimating ZHD (Boehm et al., 2006).

EvapoTranspiration (ET) is a crucial parameter used to estimate the water requirements of crops and plays a significant role in irrigation planning, climate monitoring, and subsidence prediction. One widely used method for estimating ET is the Food and Agricultural Organization Penman-Monteith (PM) equation, proposed by Allen et al. (1998):

$$ET_{(PM)} = \frac{0.408 (R_n - G) + \gamma \frac{900}{T_a + 273} U_2 (e_s - e_a)}{\Delta + \gamma (1 + 0.34 U_2)} \tag{8}$$

Where  $G$  is soil heat flux density,  $R_n$  is net radiation at the crop surface,  $T_a$  is the mean daily air temperature at 2 m height,  $e_s$  is saturation vapor pressure,  $U_2$  is the wind speed at 2 m height,  $e_a$  is actual vapor pressure,  $\Delta$  is slope vapor pressure curve and  $\gamma$  is psychrometric constant. However, obtaining all the required meteorological data for the PM equation can be challenging as these data may not be readily available or may have different temporal and spatial resolutions.

Another method to estimate ET is the Thornthwaite (TH) method, introduced by Thornthwaite in 1948. The TH method estimates ET based solely on air temperature (Thornthwaite 1948):

$$ET_{(TH)} = \begin{cases} 0 & T \leq 0 \\ 16K \left( \frac{10T}{l} \right)^m & T > 0 \end{cases} \tag{9}$$

where  $T$  is temperature,  $K$  is the calibrated coefficient of latitude and month and  $l$  and  $m$  are the heat factor and coefficient, respectively. To improve the accuracy of ET estimation, a new method based on GNSS observations has been developed (Haji-Aghajany et al. 2023). This method utilizes the PM method as an accurate model and aims to model the difference between the PM and TH methods over time using water vapor data obtained from GNSS and temperature. By

applying a linear equation based on temperature and water vapor, the modeled difference is then added to the TH formula to correct it (Haji-Aghajany et al. 2023):

$$DET = ET_{PM} - ET_{TH} \tag{10}$$

$$DET = \begin{cases} a_0 \times PWV_1 + a_1 \times T + a_2 & T \geq 0 \\ b_0 \times PWV_2 + b_1 \times T + b_2 & T < 0 \end{cases} \tag{11}$$

where  $a_0$ - $a_2$  and  $b_0$ - $b_2$  are the model coefficients which can be determined based on the least squares method and  $PWV_1$  and  $PWV_2$  are the PWV at the moment of  $T \geq 0$  °C and  $T < 0$  °C, respectively. PWV can be computed using GNSS measurements based on the following formula (Bevis et al. 1992):

$$PWV = 10^6 \frac{ZWD}{R_v \rho_w} (k'_2 + \frac{k_3}{T_M})^{-1} \tag{12}$$

where  $\rho_w$  is the density of water. Finally, the more accurate value of ET index can be calculated using following equation:

$$ET_{Accurate} = DET + ET_{TH} \tag{13}$$

The accurate ET can be calculated at different times using the differential model without relying on multiple meteorological indices. Additionally, if there is a need to calculate this index as a two-dimensional map, the geographic location can be incorporated into the fitted model. It's important to note that the differential model is applied in the temporal dimension for calculating the index at the station location. For more details on using the differential model in both temporal and spatial dimensions, refer to the work of Zhao et al. (2021) and Haji-Aghajany et al. (2023).

**Hydrological, geological and environmental information**

Groundwater level data are crucial for predicting and mitigating subsidence, where the land sinks due to factors like groundwater extraction. Monitoring these levels helps understand hydrological processes and predict subsidence risks. Groundwater provides support to the soil, so monitoring levels helps prevent excessive extraction and reduce the risk of subsidence. Fluctuations in groundwater levels indicate areas prone to subsidence, whether due to heavy rainfall or droughts. Groundwater data also calibrate subsidence prediction models, aiding in land-use planning and infrastructure design (Ty et al. 2021). Access to accurate and up-to-date groundwater level data is essential for effective subsidence management. Moreover, temperature variations also contribute to subsidence phenomena and play an

important role in subsidence prediction (Chen et al. 2019; Ye et al. 2020). In addition, some geological information such as slope, aquifer media, depth to the water table, land use, distance from exploiting wells, distance from faults, and distance from rivers have been considered for the training step of subsidence prediction. These parameters have been selected based on the previous studies (Azarakhsh et al. 2022).

**K-Shape clustering algorithm**

Clustering plays a pivotal role in LSTM prediction by facilitating the identification and grouping of similar patterns in time series data. This process is essential for enhancing the performance and accuracy of LSTM models, especially when dealing with complex and diverse datasets. By applying clustering to the data, we can uncover underlying structures and relationships, enabling targeted model training, improved interpretability, and enhanced predictive capabilities. Clustering methods provide a means to segment the data into meaningful subsets, allowing LSTM models to capture the specific dynamics and patterns within each cluster. This approach not only enables a more focused and customized learning process but also enhances the model's ability to generalize and make accurate predictions on unseen data points. When it comes to subsidence prediction using LSTM, the choice of clustering method can significantly impact the performance and accuracy of predictive models. Clustering techniques offer a way to identify and group similar patterns within time series data, leading to improved model training and prediction. One notable clustering method employed in subsidence prediction is k-shape clustering. K-shape clustering is a shape-based algorithm that takes into account the shape and temporal dynamics of time series data (Paparrizos and Gravano 2015). The distinct advantage of k-shape clustering lies in its ability to capture and group time series based on their shape similarity, resulting in a more nuanced and accurate representation of the underlying patterns in the data. In comparison to traditional methods like k-means clustering, k-shape clustering offers several benefits for prediction using LSTM (Zhou et al. 2022). Firstly, k-shape clustering explicitly considers the temporal dynamics and shape characteristics of the time series, allowing for the identification of patterns that may be overlooked by other methods solely focused on magnitude or amplitude. This aspect is particularly relevant in subsidence prediction, where the shape and evolution of subsidence patterns over time are crucial factors. Secondly, k-shape clustering accommodates local shifts and warping in the time axis, making it suitable for capturing deformations and irregularities. This flexibility enables a more accurate alignment of similar shapes, resulting in improved

clustering performance. Furthermore, the incorporation of k-shape clustering in LSTM prediction enables the development of cluster-specific predictive models. By training separate LSTM models on each cluster, the models can specialize in learning the unique dynamics and patterns within each cluster (Zhou et al. 2022). This targeted approach can lead to improved prediction accuracy and a better understanding of the intricacies of subsidence behavior.

## Validation approaches

To assess and compare the efficacy of results, statistical indicators such as Root Mean Square Error (RMSE), Nash-Sutcliffe Efficiency (NSE), determination coefficient ( $R^2$ ), and Mean Absolute Error (MAE) are commonly employed. RMSE is employed to evaluate the predictive capability of various models, while  $R^2$  quantifies the relationship between the modeled and observed data. Ranging from zero to one, an  $R^2$  value of one indicates a strong association between the two sets of data. These indices are computed, respectively, as follows (Moriassi et al. 2007):

$$RMSE = \left[ \frac{1}{n} \sum_{i=1}^n (S_i - O_i)^2 \right]^{\frac{1}{2}} \quad (14)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (O_i - S_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (15)$$

$$R^2 = \left[ \frac{\frac{1}{n} \sum_{i=1}^n (S_i - \bar{S})(O_i - \bar{O})}{\sigma_s \times \sigma_o} \right]^2 \quad (16)$$

$$MAE = \frac{\sum_{i=1}^n |S_i - O_i|}{n} \quad (17)$$

where  $O_i$  and  $S_i$  are the observed and modeled data,  $\bar{O}$  and  $\bar{S}$  are the average of the observed and modeled data, and  $\sigma_o$  and  $\sigma_s$  are the standard deviation of the modeled and observed data, respectively.

## Study area and data sets

Karaj is one of the major cities in Iran and is also the capital of Alborz Province. It is located in the southern foothills of the Central Alborz Mountains and the northwest of Tehran. Geomorphologically, it is divided into mountainous, hilly, plain, and cone-shaped units. The mountainous

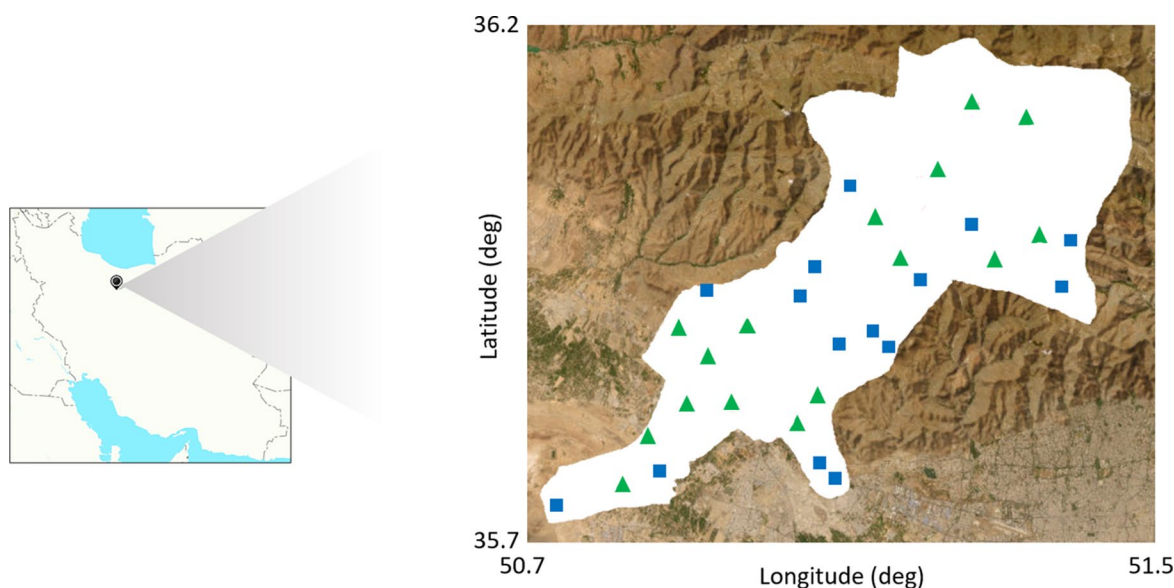
region is mostly situated in the northern, northeastern, and eastern parts, while smaller and lower areas are located in the western part. Karaj's climate is influenced by the Alborz Mountains, as well as the Chalus Valley and Karaj River. This results in a cooler and more humid climate compared to Tehran, and this distinction is observed throughout the year. According to long-term statistical analysis of the Karaj meteorological station, the city has a semi-arid climate with an annual rainfall of 247 millimeters, characterized by relatively cold winters and moderate summers. The population of Karaj was approximately 2 million people according to the National Statistical Center of Iran in 2016. After Tehran, Karaj is the most immigrant-friendly city in Iran. Among the major cities in Iran, it has the highest annual population growth rate of 14.3%. Therefore, studying the potential hazards facing this city is of great importance (Statistical Center of Iran). The geographical location of the study area can be seen in Fig. 2. Several Sentinel-1 A radar acquisitions taken from the study area have been used to implement the InSAR technique in this research. The specifications of radar acquisitions can be seen in Table 1. The spatiotemporal distribution of radar acquisitions is visible in Fig. 3. The measurements of GNSS stations have been used to compute IWV and ET index. The ERA5 reanalysis data from European Centre for Medium-Range Weather Forecasts (ECMWF) have been used to extract temperature and other necessary meteorological data for this study. ERA5 data presents values of meteorological information at 37 pressure levels, with a spatial resolution of about 31 km from 1950 to the present (Hersbach and Dee, 2016).

The study utilized groundwater level measurements obtained from piezometric wells monitored by Iran Water Resources Management and Regional Water Companies. Additionally, monthly differences in water table values were calculated at observation wells between 2015 and 2020, and these values were then used to create a continuous raster through kriging interpolation (Webster and Oliver 2007). The specifications of the other used data sets can be found in Table 2. The land use, slope, fault, geology, and vegetation coverage maps, and DEM of the study area can be seen in Fig. 4. The hydrological, geological and environmental parameters used for training were converted into raster format, and their values were extracted for the InSAR points identified through InSAR analysis.

## Results and analysis

To predict the land subsidence using GNSS, geological, hydrological, environmental information and InSAR data the following flowchart in Fig. 5 has been used.





**Fig. 2** The location of the study area and the distribution of GNSS (triangle) and groundwater wells (square)

**Table 1** Specifications of sentinel-1A radar acquisitions

Mission	Product	Track	Flight Direction	Beam Mode	Time period
Senti-nel-1 A	S1A-IW-SLC	28	ASCENDING	IW	Jan 2015 to Dec 21

The flowchart outlines the steps involved in processing and analyzing the data to predict land subsidence. In order to assess the impact of using GNSS tropospheric products on land subsidence prediction, the procedure was conducted both with and without utilizing these data (Fig. 5). Finally, the predicted results will be compared to the actual subsidence information obtained from InSAR.

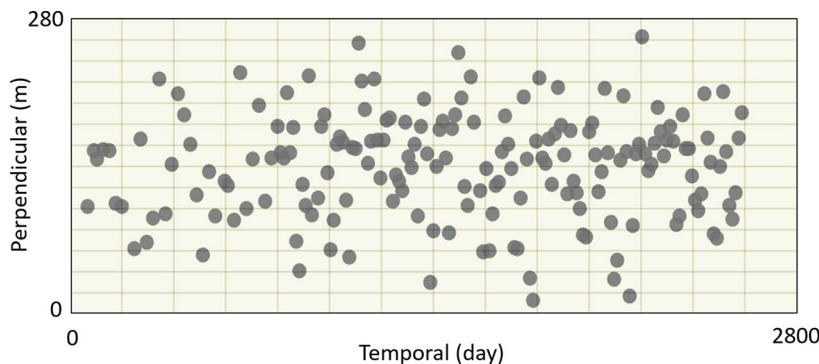
**Preparing data**

In the first step, the IWV and ET time series were computed from GNSS measurements. The Bernese 5.2 software was used for GNSS processing, and the Precise Point Positioning (PPP) method was employed to obtain the ZTD with a time

resolution of one hour for each of the radar acquisition days (Dach et al., 2015). After computing the total tropospheric delay, the Saastamoinen model was applied to subtract the hydrostatic part and obtain the ZWD. By utilizing ZWD, meteorological parameters, and the procedures mentioned in the previous sections, the IWV and ET time series were computed. The obtained time series for two sample points are visible in Fig. 6. The obtained IWV and ET were then interpolated on the InSAR points using the Kriging method.

Multiple radar acquisitions from Sentinel-1 A were utilized in the analysis for carrying out InSAR processing. A value of 0.42 was chosen for the Amplitude Dispersion Index based on the radiometric calibration. Initially, interferograms were selected based on their spatial coherence, and a reference point was determined by considering the smallest phase variance and the maximum number of coherent pixels. In order to calculate displacement velocity, any disruptive effects present in the interferograms were eliminated. The displacement time series and velocity map were then computed using the SBAS method. Fig. 7 illustrates sample displacement maps extracted from the obtained time

**Fig. 3** The spatiotemporal distribution of the radar acquisitions



**Table 2** Details of each data set

Data	Description
DEM	Obtained from Sentinel satellite data with a spatial resolution of 10 m
Slope map	Derived from the high-resolution DEM
Aquifer-media	The characteristics of Aquifer-media were determined based on data from well-driller's logs within the study areas, representing the soil texture and structure at various depths up to 100 m
Soil type	Categorized as clay and silt, silt and sand, and sand and gravel, considering the respective thicknesses of these soil types
Depths to water table values	Calculated by subtracting the water table levels from the bedrock depths, leading to the generation of interpolated maps
Land use map	Prepared by the Tehran and Shahriar municipalities in Iran.
Bedrock thickness	Map published by the Geological and Mineral Exploration Organization of Iran in 2007
Groundwater level and depth measurements from piezometric wells	Provided by Iran Water Resources Management and regional water companies
Depth to the bedrock	Determined by calculating the differences between the bedrock depth values and the surface levels in the piezometric wells
Euclidean distance from the exploitation wells	Computed using information provided by the Tehran Regional Water Company
Euclidean distance from faults	Derived from the Fault maps provided by the Geological Survey and Mineral Exploration of Iran
Euclidean distance from rivers	Calculated based on the river map within the study area

series. Fig. 8 displays the resulting displacement velocity in the study areas between 2014 and 2021. The analysis of the displacement velocity field reveals a concerning trend of subsidence occurring in the central part of the area at a rate of up to 30 cm/year. No significant displacement was observed in other parts of the area. This significant downward movement of the land poses a potential danger to the inhabitants of the region. It is crucial for local authorities and relevant stakeholders to take immediate action, conducting further investigations and implementing mitigation measures to safeguard the affected population and minimize the potential risks associated with this alarming subsidence issue.

In order to better visualize the variation in displacement in the area and observe the subsidence signal more clearly, two profiles have been considered on the displacement velocity field. The displacement variations along these two profiles can be seen in Fig. 9.

The temperature and other necessary meteorological data has been obtained from ERA5 model. Subsequently, the data mentioned in the previous section have been interpolated

at the location of InSAR points using kriging interpolation. A sample of the obtained temperature map can be seen in Fig. 10.

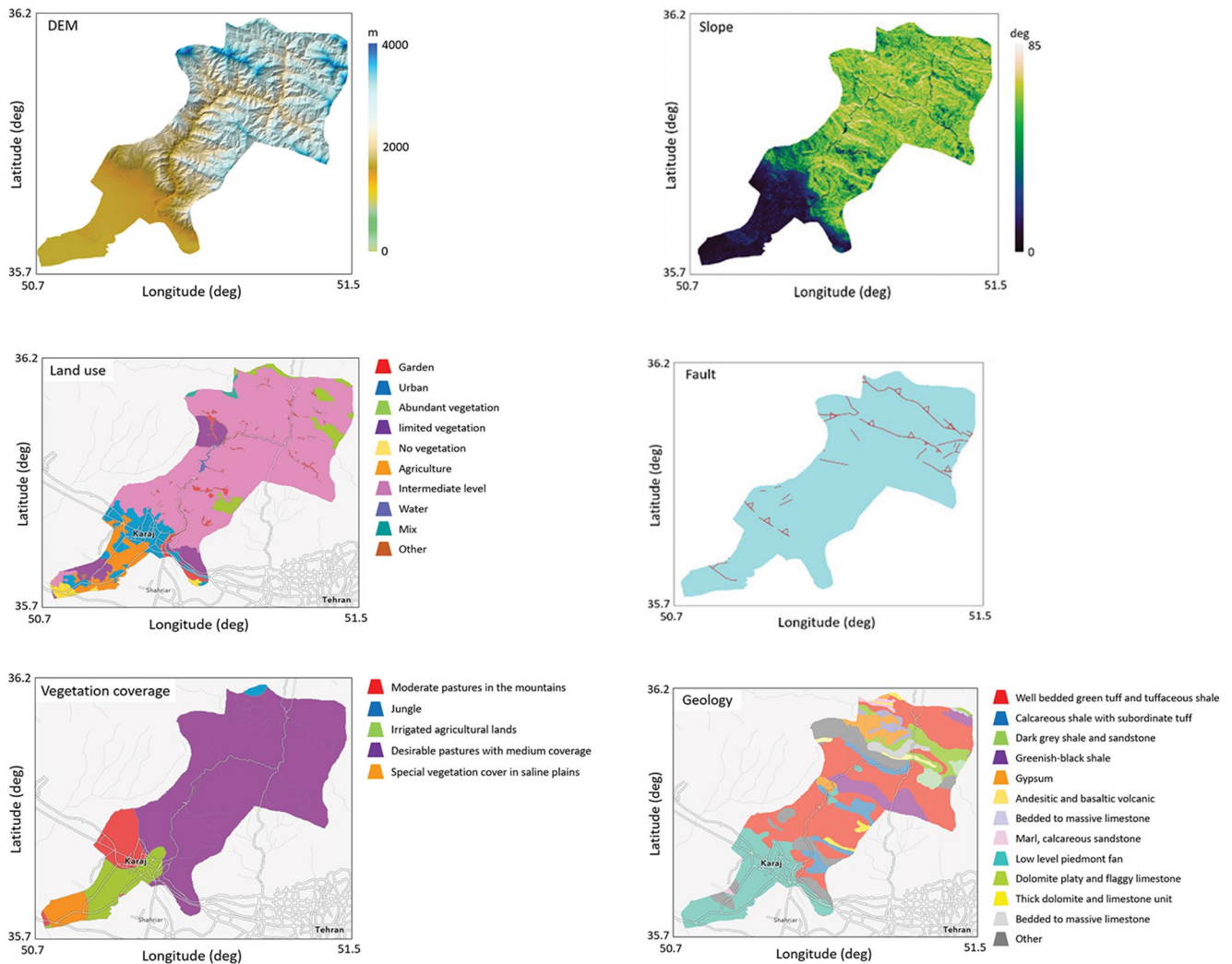
### Training and validation

Prior to the training step, the displacement signals within the study area were clustered using the k-shape method. The objective was to separate the subsidence points from the rest of the signals, as the focus of this study is specifically on subsidence. The prediction process has been performed on 1877 points that were extracted from k-shape clustering. A new dataset was then generated by employing a sliding window approach, taking into account the length of the historical sequence. This dataset was subsequently divided into three subsets: training, validation, and testing. It was determined that 70% of the data would be allocated for training, 15% validation and 15% for testing purposes. Fig. 11 depicts the step-by-step procedure employed in predicting the outcomes for the first case. In the second case, the prediction did not involve the use of GNSS products. In both cases, various training hyperparameters were established based on the genetic optimization algorithm. The used hyperparameters can be found in Table 3.

The diagram representing the loss function during the training process for each case is visible in Fig. 12. Both diagrams exhibit a decreasing trend of the loss function, gradually approaching a value very close to zero. It should be noted that the prediction process has been done on 1877 points that were extracted from k-shape clustering.

### Prediction and testing

After the training phase for both cases, in this stage, the prediction of displacement was performed for year 2021 on 1877 selected points. Fig. 13 displays the obtained displacement velocity from the prediction for both cases alongside the actual displacement velocity obtained from InSAR. As observed, the second case has predicted greater subsidence for this time period. Additionally, it is noteworthy that the spatial variations in the map obtained from the second case are less than the map obtained from the first case and the map obtained from InSAR. This can be attributed to the inappropriate resolution of the input data during the training phase. The spatial variations in the displacement of the second case can be clearly observed, which bears a significant resemblance to the spatial variations in the map obtained from InSAR. This issue is due to the use of GNSS tropospheric products as auxiliary data with appropriate resolution. In order to better observe the spatial variations of the two profiles on the map obtained from InSAR, Fig. 14 represents the predicted and actual displacement variations along

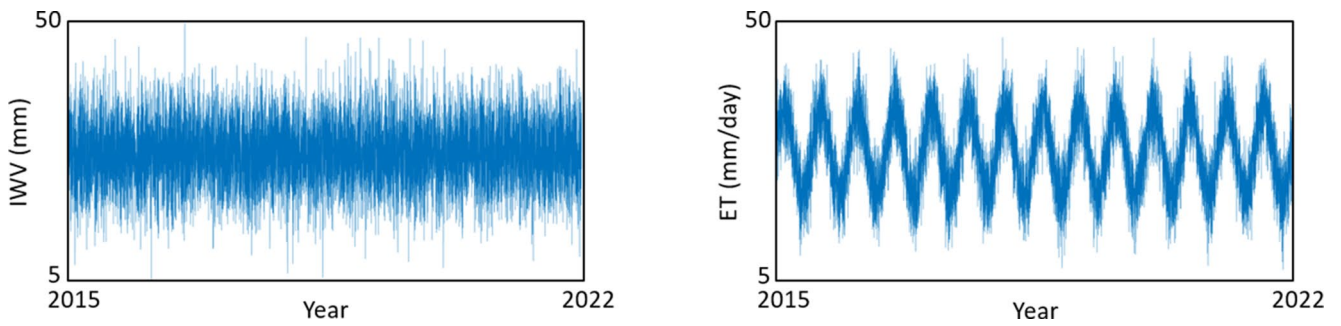
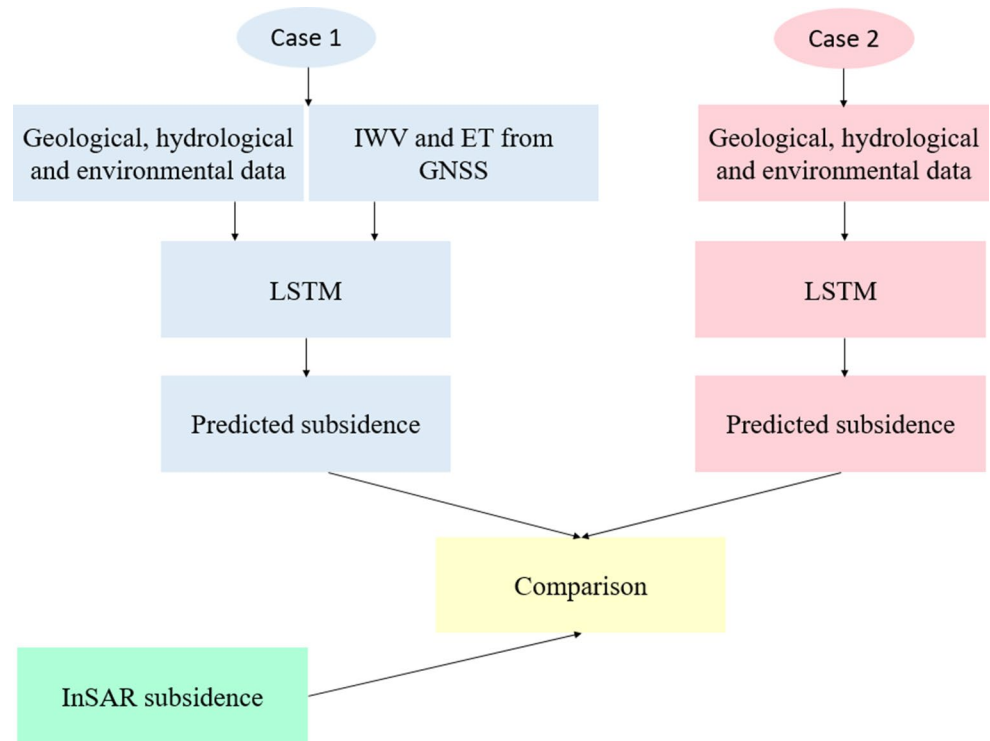


**Fig. 4** Study area specifications represented through maps

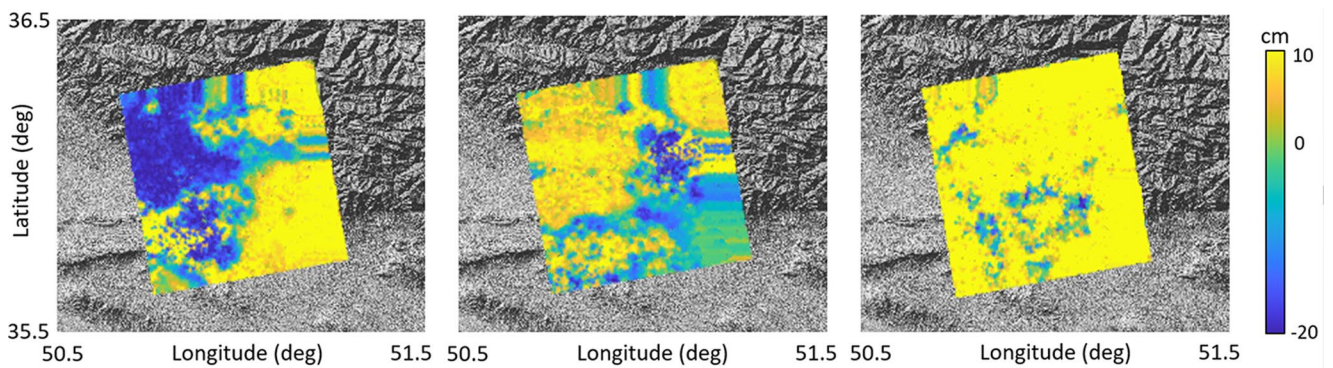
these two profiles. Based on this image, it can be said that the map obtained in the second case has a better alignment with the actual map. Furthermore, in some points, the map obtained from the second case shows different trends in displacement compared to the first case and the map obtained from InSAR. Fig. 15 illustrates the difference between the predictions obtained from both cases and the displacement obtained from InSAR. In this figure, the proximity of the results obtained from the second case to the actual data compared to the predictions obtained from the first case is clearly evident. Additionally, in order to better understand the statistical conditions of the problem, Table 4 has been used, which represents the statistical indicators of this comparison. Table 4 shows that the RMSE obtained from the first case is equal to 3.07 cm/year, and the RMSE obtained from the second case is equal to 6.23 cm/year. Furthermore, the standard deviation for the first case is equal to 2.32 cm/year, and for the second case, it is equal to 4.74 cm/year.

Both of these indicators indicate a significant difference between the predictions obtained from the first and second cases. Moreover, the NSE and  $R^2$  indicators also demonstrate a better alignment of the predictions from the first case with the actual data compared to the predictions from the second case. All the statistical results obtained indicate the role of using GNSS tropospheric products in the absence of meteorological observations, especially wet tropospheric indices. These products have been able to significantly increase the accuracy of predictions for subsidence in the region. It can be mentioned that their use, along with the utilization of groundwater levels, hydrological, geological, and environmental data, facilitates achieving accurate predictions of subsidence.

**Fig. 5** Flowchart related to the processes of this study

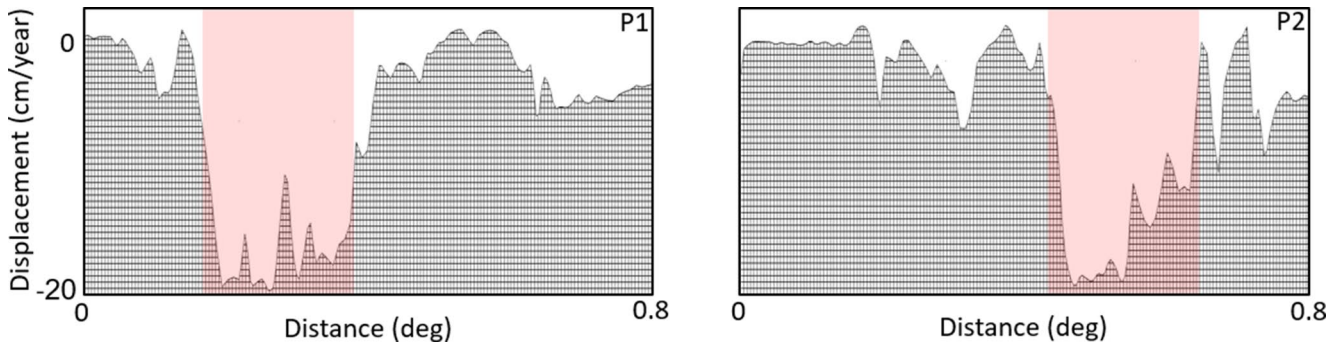
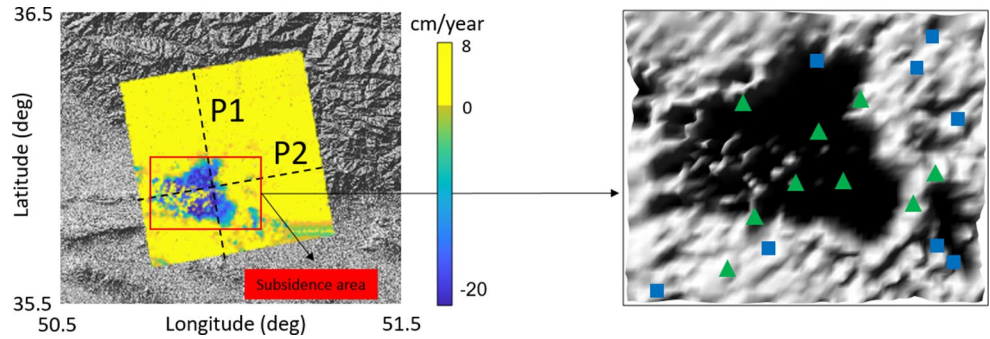


**Fig. 6** The time series of IWV and ET at a specific sample point



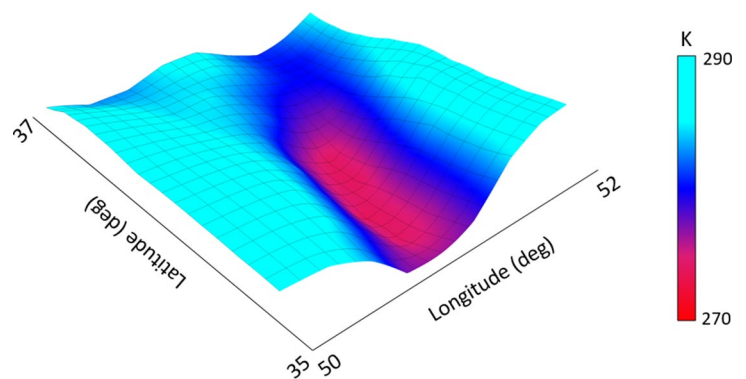
**Fig. 7** Three sample displacement maps extracted from the analysis of the InSAR time series. The time spans of these interferograms are 312, 228, and 156 days, from left to right, respectively

**Fig. 8** Displacement velocity map of the study area obtained from InSAR analysis showing the locations of GNSS (triangle) and groundwater wells (square) in the subsidence map



**Fig. 9** Profiles considered on the displacement velocity map

**Fig. 10** Temperature map on the first day of the year 2020



## Discussion

Land subsidence, influenced by numerous variables, necessitates precise prediction for effective risk mitigation. In this study, the application of GNSS tropospheric products, specifically IWV and ET, in machine learning-based land subsidence prediction was explored. Valuable insights were obtained from the numerical results: In the first scenario, where GNSS tropospheric products were integrated with other crucial parameters, an impressive RMSE of 3.07 cm/year was achieved. This low RMSE value signifies exceptional precision in capturing land subsidence patterns. In contrast, in the second scenario, where GNSS tropospheric products were omitted, a considerably higher RMSE of 6.23 cm/year was obtained. This substantial difference in RMSE values underscores the pivotal role played

by GNSS tropospheric data in enhancing predictive accuracy. Of particular note are the spatial distribution patterns of subsidence. The map generated in the first scenario closely mirrored the actual subsidence patterns observed through InSAR measurements. This high spatial correlation accentuates the ability of GNSS tropospheric products to accurately capture intricate spatial variations in subsidence. Conversely, the map produced in the second scenario, lacking GNSS tropospheric data, displayed less spatial variation and deviated significantly from actual subsidence patterns. This highlights the limitations of subsidence predictions in terms of spatial resolution when GNSS-derived information is absent. Furthermore, it's essential to emphasize the statistical parameters used for assessment. Beyond RMSE, metrics such as the MAE, NSE, and  $R^2$  consistently favored the first scenario. These metrics not only reiterate the significant improvement in predictive accuracy with

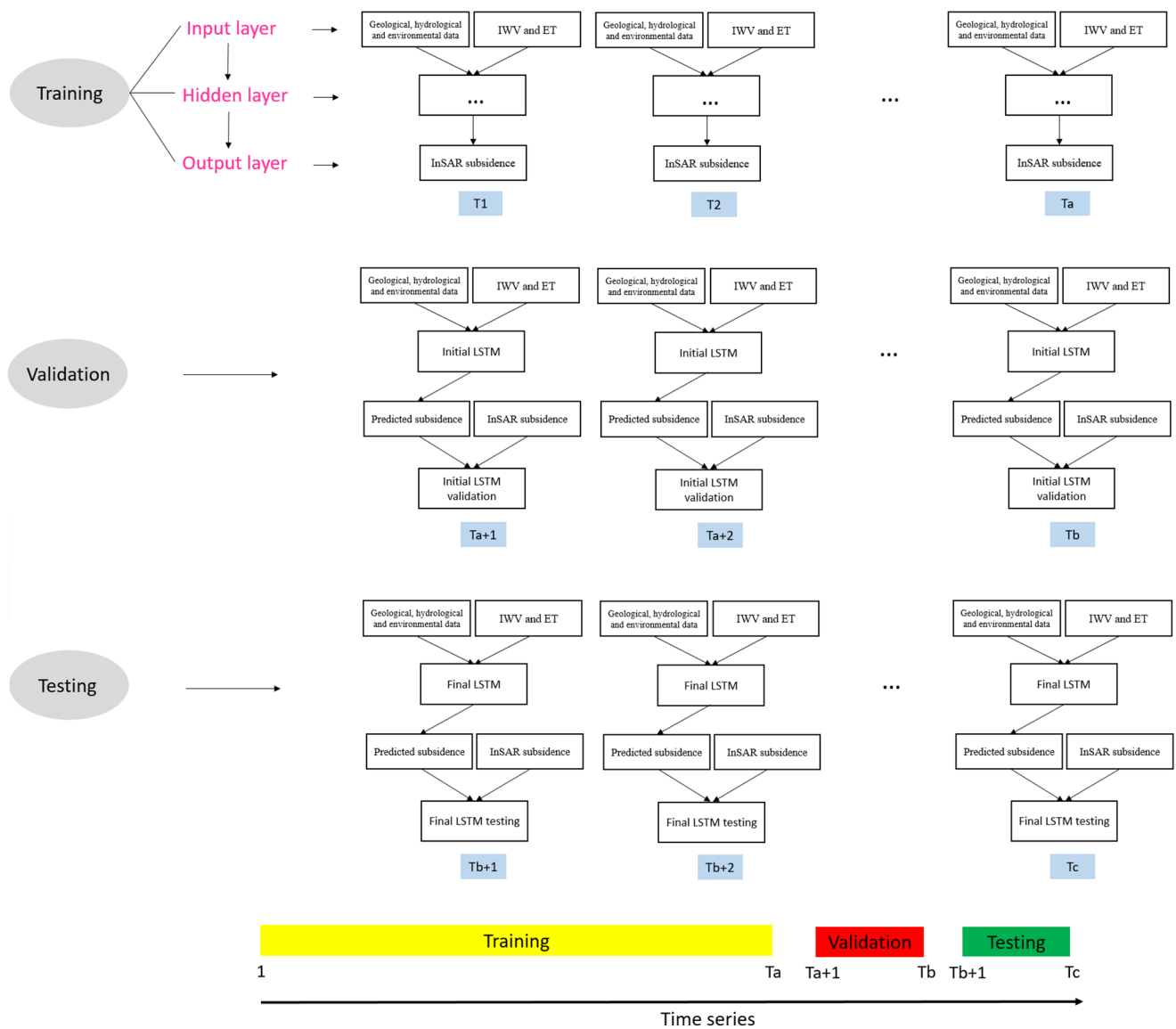


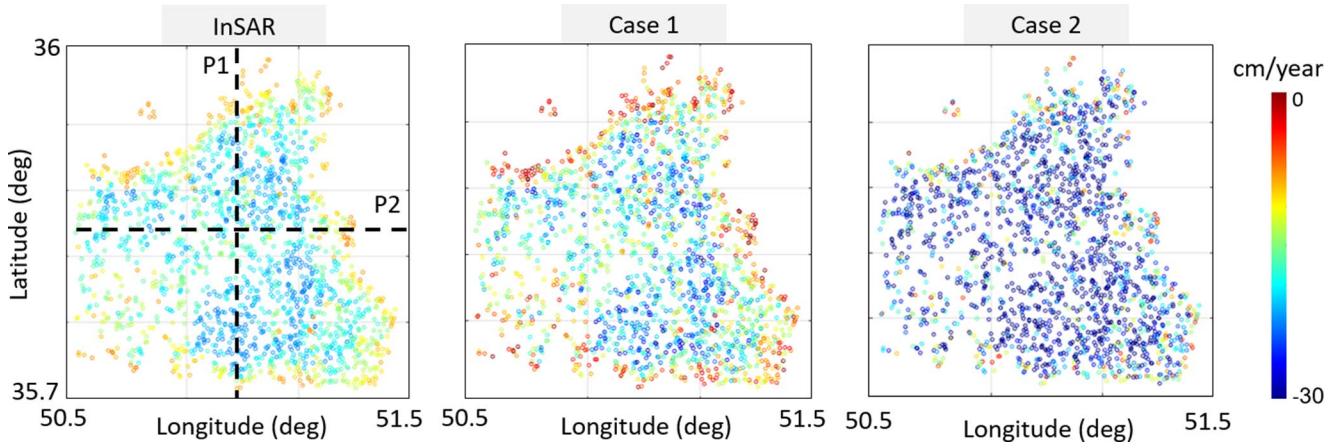
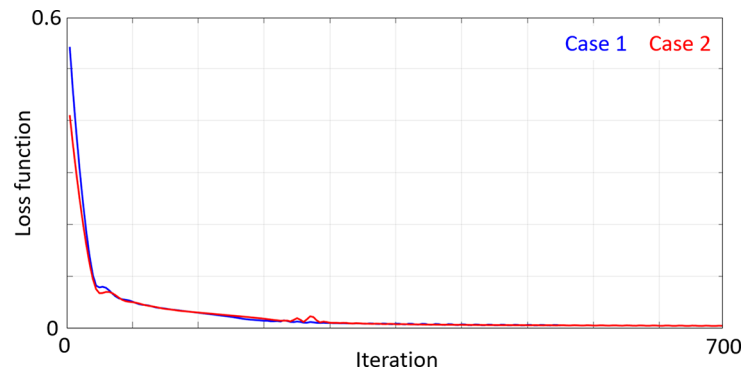
Fig. 11 Procedure for training, validation, and testing of the first case

Table 3 Optimal hyperparameters for LSTM

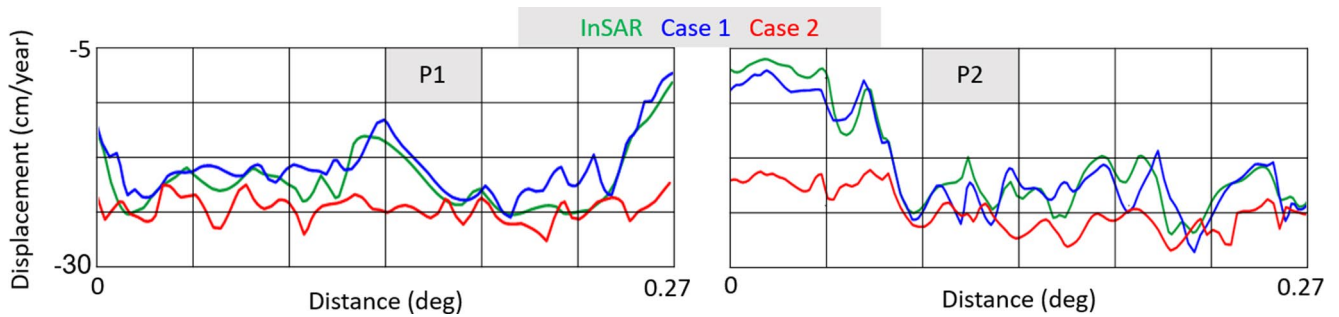
Hyperparameter	Optimal value for case 1	Optimal value for case 2
Number of LSTM layers	3	3
Number of neurons	100	100
Dropout rate	0.3	0.3
Activation functions	tanh	tanh
Batch size	32	32
Learning rate	0.01	0.01
Training iteration	500	700
Optimizer	Adam	Adam
Loss function	Mean Square Error (MSE)	MSE

GNSS tropospheric data but also provide a comprehensive view of the model’s performance. The results undeniably affirm that the integration of GNSS tropospheric products with other critical parameters leads to superior predictive performance. This integration empowers more precise subsidence predictions and facilitates the creation of accurate spatial maps of subsidence areas. The implications of these findings are substantial. Accurate land subsidence predictions enable proactive risk management and better-informed urban planning, which, in turn, safeguards public safety and infrastructure development. Looking ahead, future research should explore advanced modeling techniques and the incorporation of additional data sources to further enhance the accuracy and robustness of land subsidence predictions. The integration of GNSS observations and complementary

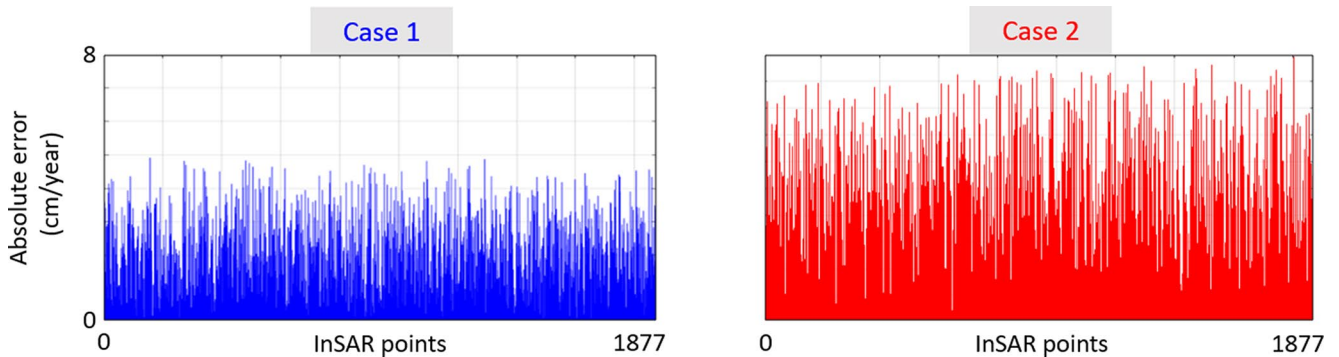
**Fig. 12** Loss function variation for each iteration of two cases



**Fig. 13** predicted and observed displacement velocity fields



**Fig. 14** Profiles considered on the predicted and observed displacement maps



**Fig. 15** Absolute error of different cases for each point

**Table 4** Statistical parameters of obtained results

Case	RMSE (cm/year)	MAE (cm/year)	NSE	R <sup>2</sup>
Case 1	3.07	2.32	0.56	0.79
Case 2	6.23	4.74	0.31	0.57

data presents a promising avenue for advancing the understanding of subsidence processes and contributing to more effective preventive measures and mitigation strategies. In summary, the numerical results obtained provide compelling evidence of the indispensable role played by GNSS tropospheric products in land subsidence prediction. These results emphasize their capacity to enhance predictive accuracy, particularly in regions with limited synoptic data coverage, and underscore their critical role in advancing the field of subsidence prediction.

## Conclusions

The numerical results of this study offer robust evidence of the instrumental role played by GNSS tropospheric products in machine learning-based land subsidence prediction. In the first scenario, where GNSS tropospheric products are incorporated, the remarkable RMSE of 3.07 cm/year, alongside lower MAE, higher NSE, and R<sup>2</sup> values, underscores the remarkable precision achieved when incorporating GNSS tropospheric data into the model, demonstrating the enhanced predictive accuracy and spatial resolution enabled by their inclusion. In the second scenario, the aim was to assess land subsidence prediction in the absence of GNSS tropospheric products. In this case, the predictive accuracy and spatial resolution were notably lower. The spatial concordance between the map generated in the first scenario and the actual subsidence patterns, contrasted with the map of the second scenario, further accentuates the value of GNSS-derived data in accurately capturing spatial variations. The statistics paint a clear picture: GNSS tropospheric products enhance predictive performance in the first scenario, making them indispensable for precise land subsidence predictions. In practice, these findings have practical implications for subsidence-prone regions. Accurate predictions enable better risk management and infrastructure safeguarding, benefiting both public safety and urban planning. Looking forward, further research should explore advanced modeling techniques and additional data sources to enhance the accuracy and robustness of land subsidence predictions. By harnessing GNSS observations and complementary data, an advanced understanding of subsidence processes can be achieved, and the ability to contribute to more effective preventive measures and mitigation strategies can be improved. In summary, the numerical results of this study underscore the pivotal role of GNSS tropospheric products in land

subsidence prediction, enhancing both predictive accuracy and spatial resolution, particularly in regions with limited synoptic data coverage.

**Acknowledgements** The authors would like to appreciate the European Space Agency (ESA) for providing the radar acquisitions. Detailed and constructive reviews and editorial remarks are greatly appreciated.

**Author contributions** All authors contributed to the study conception and design, material preparation, data collection, and analysis. Conceptualization was done by M.T. and S.H.-A. Formal analysis was carried out by M.T., Z.G., and A.G. The original draft was written by M.T. and Z.G., and S.H.-A provided supervision, review, and editing. All authors have read and agreed to the published version of the manuscript.

**Funding** Not applicable.

**Data Availability** Not applicable.

## Declarations

**Competing interests** The authors declare no competing interests.

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

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

## References

- Agnihotri G, Chaurasia V, Kumar S (2021) A comparative study of different machine learning approaches for Precise Tropospheric Parameter Estimation using GNSS Data. *IEEE J Sel Top Appl Earth Observations Remote Sens* 14:3732–3744
- Akbari V, Motagh M (2012) Improved ground subsidence monitoring using small baseline SAR interferograms and a weighted least squares inversion algorithm. *IEEE Geosci Remote Sens Lett* 9(3):437–444
- Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration. *FAO Irrigation and Drainage, Pa-per No. 56*. Food and Agriculture Organization of the United Nations, Rome
- Anderssohn J, Wetzel HL, Walter TR, Motagh M, Djamour Y, Kaufmann H (2008) Land subsidence pattern controlled by old alpine basement faults in the Kashmar Valley, northeast Iran: results from InSAR and levelling. *Geophys J Int* 74:287–294. <https://doi.org/10.1111/j.1365-246X.2008.03805.x>



- Andreas H, Abidin Z, Gumilar H, Sidiq I., P., Sarsito T., A., D., Pradipta D (2018) Insight into the correlation between Land Subsidence and the Floods in regions of Indonesia. *IntechOpen*. <https://doi.org/10.5772/intechopen.80263>
- Azarakhsh Z, Azadbakht M, Matkan A (2022) Estimation, modeling, and prediction of land subsidence using Sentinel-1 time series in Tehran-Shahriar plain: a machine learning-based investigation. *Remote Sens Appl Soc Environ* 25:100691 [Google Scholar] [CrossRef]
- Berardino P, Fornaro G, Lanari R, Sansosti E (2002) Small baseline subset (SBAS) interferometry: a Novel Method for Monitoring Elevation Changes Applied to volcanoes using ERS synthetic aperture Radar Data. *IEEE Trans Geosci Remote Sens* 40(11):2375–2383
- Bevis M, Businger S, Herring TA, Rocken C, Anthes RA, Ware RH (1992) GNSS meteorology: Remote sensing of atmospheric water vapor using the global positioning system. *J Geophys Research: Atmos* 97(D14):15787–15801
- Bock O, Bouin MN, Walpersdorf A, Doerflinger E (2003) Comparison of GNSS precipitable water vapor to Independent observations and numerical weather prediction models. *J Geophys Research: Atmos* 108:D12
- Boehm J, Heinkelmann R, Schuh H (2006) Short note: a global model of pressure and temperature for geodetic applications. *J Geodesy* 80(12):771–781
- Chen H, Luo Z, Zhou Z, Qi S (2019) Investigation on soil moisture and its influence on land subsidence based on remote sensing and field observations in Beijing plain, China. *Environ Earth Sci* 78(19):575. <https://doi.org/10.1007/s12665-019-8552-y>
- Dach R, Lutz S, Walser P, Fridez P (2015) Bernese GNSS Software Version 5.2. User manual. Astronomical Institute, University of Bern, Bern Open Publishing. <https://doi.org/10.7892/boris.72297>. ISBN: 978-3-906813-05-9
- Dach R, Arnold D, Grahsl A, Ge M (2020) Global tropospheric maps based on Global Navigation Satellite Systems: a review. *Remote Sens* 12(4):660
- Davis JL, Yu Z, Wdowinski S, Zhang P, Lee H (2013) Surface deformations caused by the 2010 Maule Earthquake, Chile: Ground-based and GNSS measurements. *J Geophys Research: Solid Earth* 118(2):823–8303
- Davoodijam M, Motagh M, Momeni M (2015) Land subsidence in Mahyar Plain, Central Iran, investigated using Envisat SAR Data. In: *Proceedings the 1st international workshop on the quality of geodetic observation and monitoring systems (QuGOMS'11)*. Springer, pp 127–130. <https://doi.org/10.1007/978-3-319-10828>
- Dehghani M, Valadan Zoj MJ, Entezam I, Mansourian A, Saatchi S (2009) InSAR monitoring of Progressive land subsidence in Neyshabour, northeast Iran. *Geophys J Int* 178:47–56 [CrossRef]
- Ding P, Jia C, Di S, Wang L, Bian C, Yang X (2020) Analysis and prediction of land subsidence along significant linear engineering. *Bull Eng Geol Environ* 79(10):5125–5139
- Elman JL (1990) Recurrent neural networks. *Cogn Sci* 14(2):179–211
- Emardson TR, Simons M, Webb HH (2003) Neutral atmospheric delay in interferometric synthetic aperture radar applications: statistical description and mitigation. *J Geophys Res* 108(B5):2231. <https://doi.org/10.1029/2002JB001781>
- European Space Agency (ESA). (n.d.). Sentinel-2. Retrieved from <https://sentinel.esa.int/web/sentinel/missions/sentinel-2>
- Farshbaf A, Mousavi MN, Shahnazi S (2023) Vulnerability assessment of power transmission towers affected by land subsidence via interferometric synthetic aperture radar technique and finite element method analysis: a case study of Zanjan and Qazvin provinces. <https://doi.org/10.1007/s10668-023-03127-x>. *Environ Dev Sustain*
- Faunt CC, Sneed M, Traum J, Brandt JT (2016) Impact of Climate Variability on Land Subsidence in the San Joaquin Valley, California, USA. *Hydrogeol J* 24(3):675–686
- Ferretti A, Prati C, Rocca F (2007) InSAR Time Series Analysis: methods and applications for Earth Surface Deformation Monitoring. *Rep Prog Phys* 70(7):1–74
- Gers FA, Schmidhuber J, Cummins F (2000) Learning to forget: continual prediction with LSTM. *Neural Comput* 12(10):2451–2471
- Goodfellow I, Bengio Y, Courville A (2016) *Deep learning*. MIT Press
- Graves A (2013) Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850
- Gray AL, Mattar KE, Sofko G (2000) Influence of ionospheric electron density fluctuations on satellite radar interferometry. *Geophys Res Lett* 27:1451–1454
- Greff K, Srivastava RK, Schmidhuber J (2017) LSTM: a search space odyssey. *IEEE Trans Neural Networks Learn Syst* 28(10):2222–2232
- Haji-Aghajany S (2021) *Function-Based Troposphere Water Vapor Tomography Using GNSS Observations*. PhD Thesis, Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology
- Haji-Aghajany S, Amerian Y (2018) An investigation of three dimensional ray tracing method efficiency in precise point positioning by tropospheric delay correction. *J Earth Space Phys* 44:39–52. <https://doi.org/10.22059/JESPHYS.2018.236885.1006913>
- Haji-Aghajany S, Amerian Y (2020) Atmospheric phase screen estimation for land subsidence evaluation by InSAR time series analysis in Kurdistan, Iran. *J Atmos Solar Terr Phys* 209:105314. <https://doi.org/10.1016/j.jastp.2020.105314>
- Haji-Aghajany S, Amerian Y (2020a) Assessment of InSAR tropospheric signal correction methods. *J Appl Remote Sens* 14(4):044503. <https://doi.org/10.1117/1.JRS.14.044503>
- Haji-Aghajany S, Voosoghi B, Yazdian A (2017) Estimation of North Tabriz Fault parameters using neural networks and 3D tropospherically corrected Surface Displacement Field. *Geomatics Nat Hazards Risk* 8(2):918–932. <https://doi.org/10.1080/19475705.2017.1289248>
- Haji-Aghajany S, Voosoghi B, Amerian Y (2019) Estimating the slip rate on the north Tabriz fault (Iran) from InSAR measurements with tropospheric correction using 3D ray tracing technique. *Adv Space Res* 64(11):2199–2208. <https://doi.org/10.1016/j.asr.2019.08.021>
- Haji-Aghajany S, Pirooznia M, Raoofian Naeeni M, Amerian Y (2020) Combination of Artificial neural network and genetic algorithm to Inverse Source parameters of Sefid-Sang Earthquake using InSAR technique and Analytical Model Conjunction. *J Earth Space Phys* 45(4):121–131. <https://doi.org/10.22059/JESPHYS.2019.269596.1007065>
- Haji-Aghajany S, Amerian Y, Amiri-Simkooei A (2022) Function-based Troposphere tomography technique for optimal down-scaling of precipitation. *Remote Sens* 14(11):2548. <https://doi.org/10.3390/rs14112548>
- Haji-Aghajany S, Amerian Y, Amiri-Simkooei A (2023) Impact of Climate Change parameters on Groundwater Level: implications for two subsidence regions in Iran using Geodetic observations and Artificial neural networks (ANN). *Remote Sens* 15:1555. <https://doi.org/10.3390/rs15061555>
- Hanssen RF (2001) *Radar interferometry: data interpretation and error analysis*. Springer Science & Business Media
- Hersbach H, Dee D ERA5 reanalysis is in production; ECMWF Newsl 147; ECMWF: Reading, UK, 2016.
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hu J, Motagh M, Guo J, Haghghi MH, Li T, Qin F, Wu W (2022) Inferring subsidence characteristics in Wuhan (China) through

- multitemporal InSAR and hydrogeological analysis. *Eng Geol* 297:106530
- Jolivet R, Agram PS, Lin NY, Simons M, Doin MP, Peltzer G, Li Z (2014) Improving InSAR geodesy using global atmospheric models. *J Geophys Res Solid Earth* 119:2324–2341. <https://doi.org/10.1002/2013JB010588>
- Khalili MA, Voosoghi B, Guerriero L, Haji-Aghajany S, Calcaterra D, Di Martire D (2023) Mapping of mean deformation rates based on APS-corrected InSAR data using unsupervised clustering algorithms. *Remote Sens* 15(2):529. <https://doi.org/10.3390/rs15020529>
- Khanlari G, Heidari M, Momeni AA, Ahmadi M, Beydokhti AT (2012) The effect of groundwater overexploitation on land subsidence and sinkhole occurrences, western Iran. *Q J Eng GeolHydrogeol* 45(4):447–456
- Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- Kleijer F (2004) Troposphere delay modelling and filtering for precise GNSS levelling, Ph.D. thesis, Mathematical Geodesy and Positioning, Delft University of Technology
- Li GD, Yamaguchi D, Nagai M (2007) A GM(1,1)–Markov chain combined model with an application to predict the number of Chinese international airlines. *Technological Forecast Social Change Technol Forecast Soc Change* 74:1465–1481
- Lipton ZC, Berkowitz J, Elkan C (2015) A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019.
- Mahmoudpour M, Khamehchiyan M, Nikudel MR, Ghassemi MR (2016) Numerical simulation and prediction of regional land subsidence caused by groundwater exploitation in the southwest plain of Tehran. *Iran Eng Geol* 201:6–28. <https://doi.org/10.1016/J.ENGGEOL.2015.12.004>
- Mohseni N, Sepehr A, Hosseinzadeh SR, Golzarian MR, Shabani F (2017) Variations in spatial patterns of soil–vegetation properties over subsidence-related ground fissures at an arid ecotone in northeastern Iran. *Environ Earth Sci* 76(6):234. <https://doi.org/10.1007/s12665-017-6559-z>
- Moriasi DN, Arnold JG, van Liew MW, Bingner RL, Harmel RD, Veith TL (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans ASABE* 50:885–900
- Motagh M et al (2008) Land subsidence in Iran caused by widespread water reservoir overexploitation. *J Geophys Res Lett* L0338:14. <https://doi.org/10.1029/2008G>
- Motagh M, Shamshiri R, Haghshenas Haghighi M, Wetzel H-U, Akbari B, Nahavandchi H, Roessner S, Arabi S (2017) Quantifying groundwater exploitation induced subsidence in the Rafsanjan plain, southeastern Iran, using InSAR time-series and in situ measurements. *Eng Geol* 1–18. <https://doi.org/10.1016/j.enggeo.2017.01.011>
- Paparrizos J, Gravano L (2015) k-shape: Efficient and accurate clustering of time series. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, Melbourne, VIC, Australia, 31 May–4 June; pp. 1855–1870
- Peltzer G, Crampé F, Hensley S, Rosen P (2001) Transient strain accumulation and fault interaction in the Eastern California shear zone. *Geology* 29(11):975–978
- Qi Y, Li Q, Karimian H, Liu D (2019) A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory. *Sci Total Environ* 664:1–10. <https://doi.org/10.1016/j.scitotenv.2019.01.333>
- Remy D, Bonvalot S, Briole P, Murakami M (2003) Accurate measurement of tropospheric effects in volcanic areas from SAR interferometry data: application to Sakurajima volcano (Japan). *Earth Planet Sci Lett* 213:299–310
- Rosen PA, Hensley S, Joughin IR (2000) Synthetic aperture radar interferometry. *Proc IEEE* 88(3):333–382
- Saastamoinen J (1972) Atmospheric correction for the Troposphere and stratosphere in radio ranging of satellites. *The Use of Artificial Satellites for Geodesy* 15:247–251
- Sak H, Senior A, Beaufays F (2014) Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In *Fifteenth Annual Conference of the International Speech Communication Association*
- Sorkhabi OM, Kurdpour I, Sarteshnizi RE (2022) Land subsidence and groundwater storage investigation with multi sensor and extended Kalman filter. *Groundw Sustain Dev* 1(19):100859
- Statistical Center of Iran (2018) Available online: <http://www.amar.org.ir> (accessed on 11
- Thorntwaite CW (1948) An approach toward a rational classification of climate. *Geogr Rev* 38:55–94 [CrossRef]
- Tregoning P, Ramillien G (2014) Water in the balance. *Nat Geosci* 7(9):613–614
- Ty TV, Minh HVT, Avtar R, Kumar P, Hiep HV, Kurasaki M (2021) Spatiotemporal variations in groundwater levels and the impact on land subsidence in CanTho, Vietnam. *Groundw Sustain Dev* 15:100680. <https://doi.org/10.1016/j.gsd.2021.100680>
- Wang H, Jia C, Ding P et al (2023) Analysis and prediction of Regional Land Subsidence with InSAR Technology and Machine Learning Algorithm. *KSCE J Civ Eng* 27:782–793. <https://doi.org/10.1007/s12205-022-1067-4>
- Webster R, Oliver MA (2007) *Geostatistics for environmental scientists*. John Wiley & Sons
- Ye S, Zhou G, Peng Y, Liu X, Sun C (2020) Impacts of soil moisture and temperature variations on land subsidence in Shanghai, China. *J Coastal Res* 95(sp1):57–63. <https://doi.org/10.2112/SI95-011.1>
- Zebker HA, Villasenor J (1992) Decorrelation in interferometric radar echoes. *IEEE Trans Geosci Remote Sens* 30(5):950–959
- Zhang K, Lu C, Chen W, Hu C, Zhang J, Jiao W (2016) Zenith wet delay estimation using ground-based GNSS and surface pressure observations: a comparative study. *J Geodesy* 90(10):951–964
- Zhao Q, Sun T, Zhang T, He L, Zhang Z, Shen Z, Xiong S (2021) High-Precision potential evapotranspiration model using GNSS Observation. *Remote Sens* 13:4848. <https://doi.org/10.3390/rs13234848>
- Zhou D, Zuo X, Zhao Z (2022) Constructing a large-scale Urban Land Subsidence Prediction Method based on neural Network Algorithm from the perspective of multiple factors. *Remote Sens* 14:1803. <https://doi.org/10.3390/rs14081803>

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