Master Thesis
TVVR 24/5012

Remote sensing-AI based prediction of subsidence in Ararat Valley, Armenia

Modeling Ground Subsidence Using InSAR, LSTM Networks, and Remote Sensing Data

Giuseppe Romano



Division of Water Resources Engineering
Department of Building and Environmental Technology
Lund University

Remote sensing-AI based prediction of subsidence in Ararat Valley, Armenia

Modeling Ground Subsidence Using InSAR, LSTM Networks, and Remote Sensing Data

By:
Giuseppe Romano

Master Thesis

Division of Water Resources Engineering
Department of Building & Environmental Technology
Lund University
Box 118
221 00 Lund, Sweden

Water Resources Engineering
TVVR-24/5012
ISSN 1101-9824

Lund 2024
www.tvrl.lth.se

Master Thesis
Division of Water Resources Engineering
Department of Building & Environmental Technology
Lund University

Swedish title:	
English title:	Remote Sensing-AI based prediction of subsidence in Ararat Valley, Armenia.
Author(s):	Giuseppe Romano
Supervisor:	Hossein Hashemi
Co-Supervisor:	Behshid Khodaei
Examiner:	Kenneth M.Persson
Language	English
Year:	2024
Keywords:	InSAR; Subsidence; Long-Short Term Memory; AI; Armenia; Remote Sensing

Acknowledgements

I want to express my deepest gratitude to my supervisor, Hossein Hashemi, at the Department of Water Resources Engineering and Center for Middle Eastern Studies, for allowing me to work on this research topic. I thank him for his expertise, feedback, time, encouragement, and insights in this field! I also extend my thanks to Behshid Khodaei in Hossein Hashemi's group for his valuable suggestions, motivation, and time.

Furthermore, I would like to express my sincere appreciation to the American University of Armenia, especially Garabet Kazanjian, for hosting me and facilitating my fieldwork during my stay in Yerevan. Their support was instrumental in the success of my research.

I would like to thank for the support, encouragement, and understanding given by all of you: my family and friends. Your belief in me has been an unfailing source of strength and motivation. Thank you for always being there for me.

Abstract

The Ararat Valley, being the cereal basket of Armenia and Turkey, is in danger with over-abstracted groundwater, which already induces colossal water loss and land subsidence. In the current work, advanced remote sensing techniques and machine learning models are applied to predict subsidence and give data-driven recommendations to effect sustainable water management. The tracking of terrain changes and analysis and prediction of the patterns of subsidence are carried out with the help of Synthetic Aperture Radar Interferometry and Long Short-Term Memory networks, respectively. The integration of InSAR data with LSTM models allows an in-depth analysis of subsidence dynamics, forming a basis for forecasting and management strategies. In that regard, early results confirmed the expediency of using the combination of remote sensing data and machine learning techniques, stating that it is possible to monitor subsidence accurately for all zones in the Ararat Valley. Further study investigates the socio-economic effects that subsidence will have and will give practical recommendations in relation to water management practice improvements, which include enforcement of water reuse regulations, enhancement of public awareness on the same, and adjustments in water pricing to its real value. By addressing both technical and socio-economic aspects of water management, the current research opens new horizons in providing a truly integrative system for coping with both land subsidence and securing the sustainable use of water resources in the Ararat Valley. It is considered a sign that new and very promising opportunities are related to

uniting innovative technologies with large, comprehensive, and far-reaching policy measures to effectively address environmental challenges.

Table of contents

1. Introduction 13
1.1 Objectives 14
2. Background 15
2.1 InSAR 15
2.2 Long Short-Term Memory 17
2.3 Study Area 18
2.4 Previous Studies 23
3. Data and Methodology 25
3.1 InSAR Data 25
3.2 Remote Sensing and Hydrological Data 28
3.3 Data Harmonization 32
3.3.1 Spatial Resampling 33
3.3.2 Temporal Resampling 34
3.4 Model Development 34
4. Results 39
5. Discussion 44
5.1 Data Utilization in the Research 44
5.2 Models' Performance 45
5.3 Future Model Improvement 47
5.4 Recommendations 50
6. Conclusions 52
References List 53
Appendix 57

List of Abbreviations

AI - Artificial Intelligence

AUA - American University of Armenia

CEOS - Committee on Earth Observation Satellites

CGLS - Copernicus Global Land Service

DEM - Digital Elevation Model

ET - Evapotranspiration

GEE - Google Earth Engine

GMTSAR - Generic Mapping Tools Synthetic Aperture Radar

GPS - Global Positioning System

ICA - Independent Component Analysis

InSAR - Interferometric Synthetic Aperture Radar

LSTM - Long Short-Term Memory

MAP - Maximum a Posteriori Probability

MODIS - Moderate Resolution Imaging Spectroradiometer

NASA - National Aeronautics and Space Administration

NDVI - Normalized Difference Vegetation Index

RNN - Recurrent Neural Network

RMSE - Root Mean Squared Error

SAR - Synthetic Aperture Radar

SBAS - Small Baseline Subset

SLC - Single Look Component

SNAPHU - Statistical-cost Network-flow Algorithm for Phase Unwrapping

SRTM - Shuttle Radar Topography Mission

SWI - Soil Water Index

TVVR- Division of Water Resources Engineering

USGS - United States Geological Survey

List of Figures

Figure 1: Picture captured from Google Earth.…….…… 6

Figure 2: Free spilling well monitored by Hydromet..…… 10

Figure 3: Temporal and perpendicular baselines of images …….…… 15

Figure 4: Long-term deformation map of the study area…… 18

Figure 5: Photo taken during the field trip……..… 19

Figure 6: Well and Deep well distribution in the Ararat Valley…… 20

Figure 7: Representation of what Spatial resample means …….…… 23

Figure 8: Division of the long-term deformation rate…..……..…… 25

Figure 9: Loss function plots for the different models. …….…… 29

Figure 10: Testing phase of zone 1 ……..…….…… 30

.

Figure 11: Feature Importance in the models…….…… 32

Figure 12: Graphs for Zones 1, 2, and 3. …… 33

Figure 13: Comparison of land subsidence measurements …… 34

Figure 14: Sample results from zone 1…… ……..…… 35

Figure 15: Comparison of water level and subsidence values…… 38

Figure 1A: Testing phase of zone 2 …… 46

Figure 2A: Testing phase of zone 3 …….……... 47

List of Tables

Table 1: Geological units from USGS …… 9

Table 2: Remote sensing data …… 17

Table 3: Hydromet data ……..…… 19

Table 4: RMSE and R^2 values of each model…….…… 42

1. Introduction

Water is one of the critical resources for humans, which ensures their survival, economic development, and environmental sustainability. Effective management of water resources in most parts of the world has become a matter of prime concern due to skyrocketing demand and change of climate. Ararat Valley is situated between Armenia and Turkey; it is a good example of how water management might directly affect the economic and social stability of some regions. In this valley, excessive abstraction of groundwater has given rise to very serious problems of water deficit and land subsidence, compromising agricultural activities and infrastructure security. This research starts from a collaboration between the Division of Water Resources Engineering of Lund University and the Acopian Center for the Environment of the American University of Armenia (AUA). The main goal is to carry out research and model subsidence phenomena in Ararat Valley for water resource improvement and sustainable use practices. This research will, thereby, set the framework to yield data for water resources planning and management in the region based on advanced remote-sensing methods and machine-learning approaches. In this case, SAR interferometry (InSAR) will be applied to the monitoring of relief changes and Long Short-Term Memory (LSTM) networks for analysis and prediction in this work. The said tools will give a detailed and dynamic view of the processes in place that is bound to provide a solid base for future studies and mitigation strategies.

1.1 Objectives

This research's primarily aim is to explore the potential of integrating Interferometric Synthetic Aperture Radar data with advanced machine learning techniques, particularly Long Short-Term Memory networks, to effectively model and predict land subsidence in the Ararat Valley. The approach fuses the benefits of remote sensing methods with those based on AI for availing a detailed and accurate understanding of subsidence patterns in the region. By the integration of these technologies, the study focuses on developing a good predictive model so as to give insights into subsidence issues closely related to over-extraction of groundwater. This study has a second objective: to evaluate the more general implications for sustainable water management in the Ararat Valley of such subsidence forecasts.

The intention is to provide and support future water management strategies in the region with the help of accurate predictions arising from the model. This study, through data-based recommendations, hopes to influence decisions on policy and planning related to resources so that the management practices adopted are effective not only for the current problem of subsidence but also for sustainability in the future. This dual emphasis on technical innovation and practical application places this research in a key position regarding the environmental and socio-economic impacts of land subsidence in the Ararat Valley.

2. Background

2.1 InSAR

To date, one of the most widely used remote sensing technologies for measuring millimeter changes in the Earth's surface is SAR interferometry. This advanced technique uses synthetic aperture radar (SAR). In particular, it exploits phase differences between SAR images acquired at different times to create detailed maps of topography and ground deformations. Phase differences are caused by variations in the distance between the radar and the Earth's surface. The images are combined to create an interferogram, a map representing the phase differences between two SAR images. The principle of interferometry is used to extract the phase difference between two SAR acquisitions. (USGS, 2018). By comparing a set of images that cover the same areas, we can determine if there has been a phase shift in the backscattered wave between two satellite passes. The images are paired as a master image (reference) and a slave image (subsequent). These images are aligned at the pixel level using bi-dimensional cross-correlation. After alignment, the two SLC images are multiplied to generate a complex interferogram, from which the phase change is derived, as described by equation 1 (Sandwell et al., 2016). The interferogram, therefore, can detect changes in topography and ground movements with high spatial resolution.

$$(\boldsymbol{\phi 2 - \phi 1}) = \boldsymbol{tan^{-1}}(\frac{\boldsymbol{R}}{\boldsymbol{I}}) \text{ (Eq. 1)}$$

There are three main factors on which the accuracy of InSAR measurements depends: the consistency of the SAR images, i.e., the similarity of the images to each other; the geometry of the radar, i.e., the angle at which the images are acquired; and finally, atmospheric effects. Differences in height and elevation variations can introduce errors due to parallax, especially in areas characterized by accentuated reliefs. These errors can compromise the accuracy of interferometric measurements, making it more complex to interpret the data and requiring advanced processing methods to correct these distortions and obtain reliable results.

However, the benefits of using InSAR far outweigh the limitations associated with it. First, its high spatial resolution allows it to detect detailed changes in elevation on a large scale. Another significant advantage is the continuity of data: InSAR allows the constant monitoring of large geographical areas, allowing their changes over time to be analyzed. In addition, InSAR is a highly versatile technology that can operate in different weather and light conditions. This feature makes it possible to obtain accurate data regardless of weather conditions or the availability of sunlight, ensuring reliable and continuous coverage in any geographical and climatic context with difficulty. (Committee on Earth Observation Satellites [CEOS], 2018).

Therefore, it is not surprising that the main applications of this technology are closely related to monitoring ground movements in a wide-ranging context. InSAR is used to monitor earthquakes, to detect deformations of the Earth's crust and thus provide valuable information on the distribution of tectonic movements. It then monitors soil subsidence, often caused by extracting groundwater resources such as water and oil.

InSAR is also used to monitor glacier movements, contributing significantly to understanding the dynamics associated with climate change. Due to its ability to accurately detect and measure small glacial surface displacements, InSAR provides crucial data for analyzing glaciers' behavior in global warming and its long-term consequences. Finally, InSAR can be used to monitor the stability of infrastructure and urban terrain. This application is crucial for identifying potential risks that could compromise the safety of buildings, bridges, dams, and other urban structures. (The European Space Agency).

2.2 Long Short-Term Memory

Recurrent Neural Networks (RNNs) are networks capable of maintaining a state that preserves information about previous data in the sequence. For this reason, they are used for processing sequential inputs such as time series, audio, and natural language text. However, it has been noted that the vanishing gradient problem and the consequent hindrance of learning long-term dependencies in data are rather common issues in RNNs. Nowadays, however, this problem seems to have been solved thanks to LSTM networks, introduced by Hochreiter & Schmidhuber in 1997, which utilize "memory cells" that allow the network to store information over time and decide which information to keep or forget.

The memory cells are controlled by three gates: input, output, and forget. These gates regulate the flow of information into the memory cell. The amount of information that can pass through these gates is controlled by sigmoid activation functions. The input gate regulates how much new information is allowed to enter the cell. This means the model can decide how much value to

assign to new information compared to previous information. The output gate regulates the amount of information that exits the cell to the rest of the network. This determines how much of the cell's memory will be used for output. The forget gate regulates how much of the cell's memory is forgotten. This allows the network to discard information no longer relevant or harmful to the current task. These gates, along with the memory cell, allow LSTMs to learn and remember long-term dependencies in sequential data. In other words, the network can capture and retain essential and relevant information over the long term, even across long time intervals, without being affected by issues such as vanishing gradients. (Aaqilah, 2023).

2.3 Study Area

The Ararat Valley is a portion of territory historically belonging to Armenia. However, following the war's conclusion in 1920, the valley was divided between Armenia, north of the Araks River, and Turkey, south of the river (Figure 1). The portion of the valley in the Armenian part represents a fundamental intermontane depression supporting thousands of inhabitants' livelihoods. This study delves into the valley's hydrogeological dynamics.



Figure 1: Picture captured from Google Earth. The red line represents the border of the Ararat Valley in Armenia, and the yellow line represents the border with Turkey.

The Ararat Valley stretches approximately 120 km in length in the Republic of Armenia. It covers an area of 1,177 km^2, characterized by different geological formations from volcanic mountains to alluvial and lacustrine sediments. The valley works as a natural groundwater reservoir, receiving water from surrounding mountain ranges and ancient riverbeds buried beneath volcanic deposits. Volcanic rocks dominate the drainage area, constituting over 90% of the geological substrate, facilitating significant infiltration of precipitation and surface runoff into aquifers. The geological environment of the Ararat Valley was formed following the eruption of basaltic lavas from the upper and lower Quaternary. The area is characterized by limn fluvial and effusive aquifer formations with a thickness of up to 500 meters. Geologically, the valley can be schematized as an alternation of permeable and impermeable zones that

have led to the formation of three aquifers: a superficial aquifer and two deeper artesian ones. (Valder J.F & al., 2018). In Table 1, the nine geological units described in the USGS report are presented.

Table 1: Geological units from USGS

Unit	Geological log
Unit 1	Loam/Clay
Unit 2	Gravel
Unit 3	Clay
Unit 4	Gravel
Unit 5	Dense Basalt/Clay
Unit 6	Gravel
Unit 7	Dense Basalt/Clay
Unit 8	Gravel/Fractured Basalt
Unit 9	Dense Clay

The Ararat Valley has a continental climate with cold winters and hot summers. The economic importance of water in the valley is fundamental, especially in the agricultural sector. The valley has fertile soils and favorable climatic conditions, making it a hotspot for high-value crop production, viticulture, and livestock farming. Groundwater resources play a crucial role in supporting agricultural activities, with extensive use for irrigation purposes. Additionally, the emergence of fish farming in the 2000s has further diversified water usage, with trout and sturgeon farming operations drawing from

underground reservoirs. (Ministry of Environment of the Republic of Armenia and USAID, 2021), It must be noted that fish farming is a tremendous water-consuming activity, with consumption that goes from 80 l/s for small fish farms and up to 3000 l/s for big ones. Fishes constantly need a fresh flow of water to survive, and it is in the interest of the fish farmers to keep the product of high quality.

The results of a study (Valder J.F & al., 2018) revealed a significant dependence on groundwater, with over 2,800 wells mapped in the valley. The groundwater extraction for different purposes, including drinking, irrigation, and industrial use, exceeds sustainable levels, leading to a water deficit. Moreover, a big problem are many artesian wells freely spilling water in the valley, see Figure 2. In 2016 alone, the estimated overexploitation of groundwater resources reached 681.34 million m^3, exacerbating the current water imbalance in the region.



Figure 2: Free spilling well monitored by Hydromet. Photo taken during the visit to the Ararat Valley.

Of particular importance are the clay layers presented in Table 1. These layers can be defined as aquitards, poorly permeable layers that significantly slow down the flow of water through the soil but are to be considered very subject to compression that can generate land subsidence (Hydraulic response of highly compressible aquitards during consolidation D. L. Rudolph, E. O. Frind). Subsidence is a soil phenomenon that appears as a lowering of the ground due to excessive extraction of water, oil, or gas from underground. This ground lowering could be attributed to the compaction of the clay layers, see Table 1, due to of groundwater decline and the weight of the upper layers, resulting in soil compaction and reduction of aquifer volume. The water deficit described in the previous paragraph has been pointed out as one of the principal

causes of the subsidence registered in the area. The overexploitation of the aquifers dries the clay aquitard, gets compressed, and reduces its volume. Subsequently, using the InSAR method, subsidence was recorded in the area.

2.4 Previous Studies

The use of the InSAR method to evaluate subsidence is already widely used in the research field. The correlation between soil lowering and water level decreasing is a valuable tool to model and perform predictions for future years and scenarios. The literature review underlines the importance of recent studies in land subsidence modelling and the application of advanced techniques for predicting subsidence phenomena. Bockstiegel et al. (2023) conducted a survey on simulating subsidence in the Rafsanjan Plain, Iran, using numerical models and analyzing InSAR data. Their research highlighted the impact of excessive groundwater exploitation on land subsidence. Naghibi et al. (2022) proposed an integrated InSAR-machine learning approach for modelling ground deformation rates in arid areas, demonstrating the effectiveness of combining satellite data with machine learning algorithms to monitor land changes. Peng et al. (2024) characterized and predicted ground movement derived from InSAR data using LSTM models assisted by ICA, emphasizing the importance of advanced data analysis techniques to enhance subsidence predictions. Finally, Liu et al. (2024) examined machine learning-based techniques for simulating land subsidence in urban areas, highlighting the potential of new methodologies in predicting and managing risks associated with ground subsidence. Together, these studies provide a comprehensive overview of current methods and applications in land subsidence modelling

and prediction, which is crucial for addressing environmental sustainability and water resource management challenges in high-water exploitation areas. In the Ararat Valley, studies like these have not been conducted, and this could be a starting point for further improvement of the area regarding water management and policymaking.

3. Data and Methodology

The creation of the model required many steps, from data collection and preparation to the working phase of the machine learning algorithm with calibration, validation, and testing. This section describes the development of these different phases.

3.1 InSAR Data

The following procedure has been carried out by Behshid Khodaei, Doctoral student at the Division of Water Resources Engineering of Lund University and is presented here just as informative discussion on how the InSAR were acquired for this research.

To conduct the InSAR analysis, a set of 173 Sentinel-1 A SAR images acquired from ascending orbit number 72 between July 2, 2017, and April 14, 2023, were used. By applying thresholds for the perpendicular and temporal baselines of 150 meters and 60 days, respectively, a continuous network of 635 small baseline interferograms was formed. The SBAS network for the study area is presented in Figure 3.

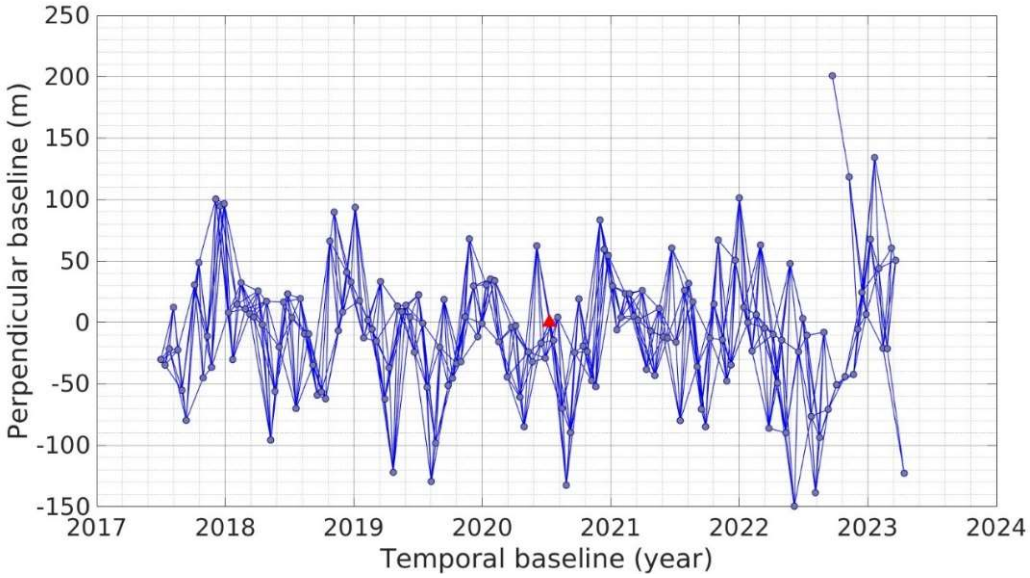


Figure 3: Temporal and perpendicular baselines of SAR images.

This network was then utilized to perform the SBAS-InSAR analysis using the Generic Mapping Tools Synthetic Aperture Radar (GMTSAR), which is an open-source InSAR processing package (Sandwell et al., 2011B).

The first step involved converting data formats to generic ones for GMTSAR, transforming raw SAR images into Single Look Component (SLC) images, and producing auxiliary files for alignment, including orbital position and velocity information. The second step entailed coarse alignment using precise orbit data, followed by fine alignment with sub-centimeter accuracy using a geometric-based registration method. GMTSAR utilized a down-sampled Digital Elevation Model (DEM) and precise orbital data to achieve 1–2-pixel accuracy, with adjustments made through de-ramping and optional Enhanced Spectral Analysis (ESD). In this work, the DEM used is the void-filled Shuttle

Radar Topography Mission (SRTM) DEM data with a 3-arc-second (90 m) spatial resolution acquired from the National Aeronautics and Space Administration (NASA). Images were co-registered relative to a middle-section reference image acquired on July 10, 2020, and DEM data were used to eliminate topographical phase contributions. The third step involved forming interferograms from reference and repeating SLCs at full resolution, removing the baseline-dependent topographical phase using the external DEM. Standard products, including amplitude, phase, and coherence, were generated through cross-multiplication and low-pass filtering. A 3D phase unwrapping algorithm, SNAPHU, which is a statistical-cost network-flow algorithm for phase unwrapping, was applied to compute the unwrapped phase of each interferogram, treating it as a series of Maximum a Posteriori Probability (MAP) estimations. Identifying a stable reference area was crucial for defining absolute phase values, achieved by identifying a stable region within the stack of unwrapped phase images using mean, standard deviation, and entropy measures. Here, we have considered a triangle area centered at a location with the coordinates of [44.240, 40.275] as the reference area. The final step converted results into a timeseries deformation by solving a weighted least-square problem, employing iterative spatiotemporal filtering to address time-dependent phases and atmospheric contributions, and averaging redundant interferograms of common scenes. (Sandwell et al., 2011A). Figure 4 presents the long-term deformation map.

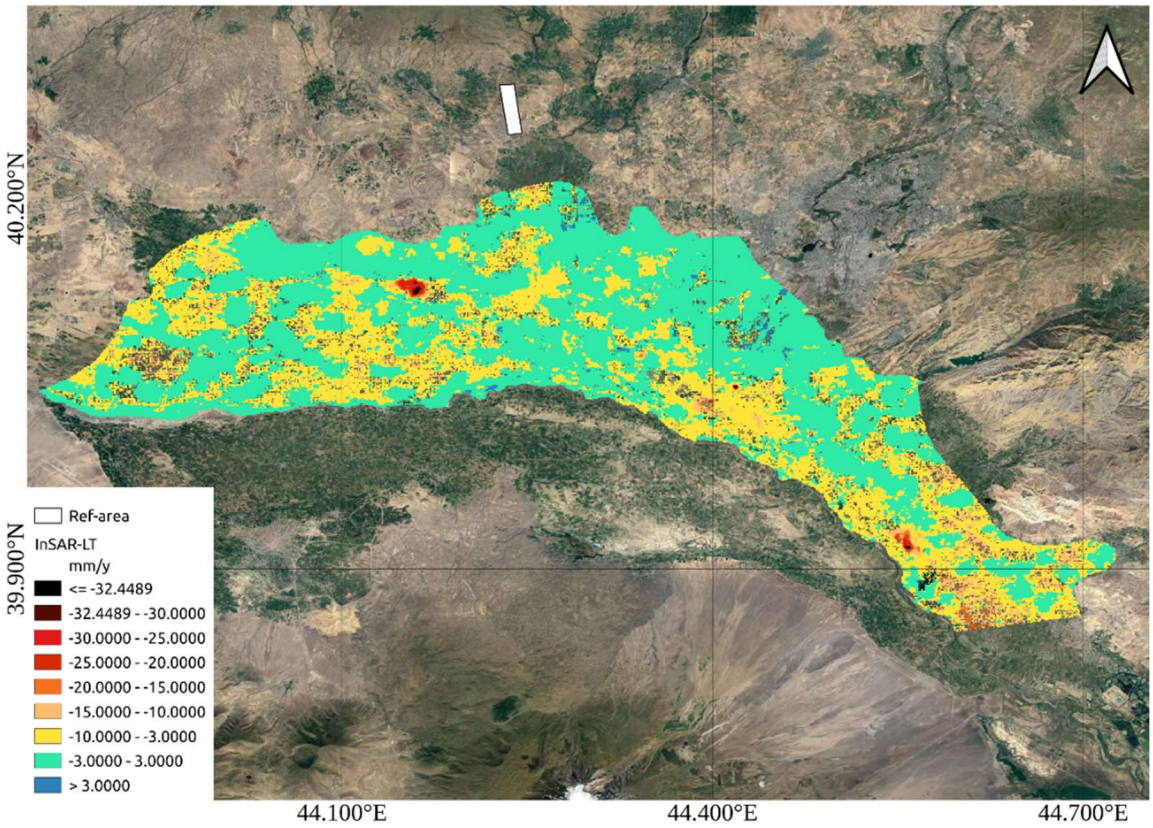


Figure 4: Long-term deformation map of the study area measured by SBAS-InSAR.

3.2 Remote Sensing and Hydrological Data

The data used in this project come from different sources. Most of the data were available online and downloaded using Google Earth Engine (GEE), a platform for geospatial analysis. GEE allows the introduction of a shapefile of the studied area and, through some coding, extracts geospatial data from the available satellites. The data were chosen considering their direct or indirect relation to ground subsidence.

The data were downloaded with a spatial resolution as close as possible to the InSAR data and, when available, a lifespan starting from 2014 to the end of 2023. It must be underlined that the time series were different for different datasets (ex., Evapotranspiration data are collected every 8 days). From the platform, the final products of Evapotranspiration (ET), Land surface temperature (LST), Precipitation, and Normalized Difference Vegetation Index (NDVI) were acquired. Moreover, from the European Copernicus Global Land Service (CGLS), Soil moisture and Soil water index (SWI) at different depths were downloaded, see Table 2.

Table 2: Remote sensing data used in this study.

Product	Spatial Resolution	Temporal Resolution	Unit	Source
ET	450m	8 days	kg/m^2/8 day	https://doi.org/10.5067/MODIS/MOD16A2GF.061
NDVI	250m	16 days	\	https://doi.org/10.5067/MODIS/MOD13Q1.061
Precipitation	5.5 Km	Daily	mm/day	https://www.nature.com/articles/sdata201566
LST	1 Km	Daily	Kelvin	https://doi.org/10.5067/MODIS/MOD11A1.061
Soil moisture	1 Km	Daily	%	10.1109/TGRS.2018.2858004
SWI (T=2)	1 Km	Daily	%	10.3390/rs10071030
SWI (T=3)	1 Km	Daily	%	10.3390/rs10071030
SWI (T=4)	1 Km	Daily	%	10.3390/rs10071030

Note: T means T-value, which represents different depths of soils.

Another important source was the field trip to Yerevan (Armenia). In January 2024, thanks to the help of the American University in Armenia, several meetings took place in the capital. During these meetings, it was possible to interview experts in the area, have a better understanding of the water management mechanisms, and request data in their possession. Moreover, two important meetings took place with two local companies, Hydromet and GeoRisk. GeoRisk is a private company in charge of all the GPS stations in Armenia that measure displacement to validate the InSAR data. Hydromet, a state company in charge of monitoring wells in the Ararat Valley, agreed to share data from 2015 to 2023 on 54 wells. The data are collected by farmers of the area every five days but the instruments they use are rudimentary (Figure 5).



Figure 5: Photo taken during the field trip while assisting a farmer taking the measurement of the monitored wells of Hydromet.

The data included different characteristics for different wells. Three types of wells were shared: well, deep well, and free fountain. For wells and deep wells, the water level (distance from the ground to the water) and temperature are monitored, while for free fountains, pressure, temperature, and exploitation are monitored. Table 3 presents a scheme of the data.

Table 3: Hydromet data

Type of well	Data monitored	Temporal Resolution	Unit
Free Fontain	Temperature	5 days	Celcius
	Exploitation		l/s
	Pressure		Bar
Well	Temperature	5 days	Celcius
	Water Level		m
Deep Well	Temperature	5 days	Celcius
	Water Level		m

However, the wells where the pressure (free fountains) is monitored were discarded because even Hydromet does not have sufficient knowledge about their characteristics to determine the related groundwater level or the aquifer they refer to. Moreover, the wells and deep wells were too few and not well-distributed to interpolate across the entire valley, see Figure 6. Additionally,

their reference aquifers were unknown, making it impossible to group and utilize these data. Therefore, a different approach was decided upon to model the InSAR data.



Figure 6: Monitoring well distribution in the Ararat Valley.

3.3 Data Harmonization

As evidenced in Table 2, the challenge in the model's data preparation process arose from integrating heterogeneous sources characterized by varying spatial resolutions and temporal frequencies. Sampling intervals ranged from every eight or sixteen days to daily. To ensure effective standardization and facilitate data interoperability, the spatial resolution of the InSAR data, set at 60 meters, was adopted.

InSAR data were plotted on QGIS to validate this resolution, measuring the distance between points in geographic space. This process confirmed the chosen resolution, solidifying the decision to standardize all data to 60 meters. A daily temporal resolution was also selected, considering that InSAR data were collected every 12 days. This choice was essential to facilitate data synchronization across different datasets during the time series creation necessary for the modelling phase.

3.3.1 Spatial Resampling

A resampling was performed to obtain the same spatial resolution for all the remote sensing data acquired on Google Earth Engine and Copernicus. On Python 3.11, an extension of the Geospatial Data Abstraction Library (GDAL) was used. This extension allows to work on raster and vector geospatial format. With a simple code, it was possible to open and resample the tiff files' pixels to the chosen resolution. Figure 7 presents a sketch to show how the code works, and here is the code line that performs it, "*dsRes = gdal.Warp("XXX.tif", ds, xRes = 60, yRes = 60, resampleAlg = "bilinear")* ". In Figure 7, an example of how the resampling works is presented.

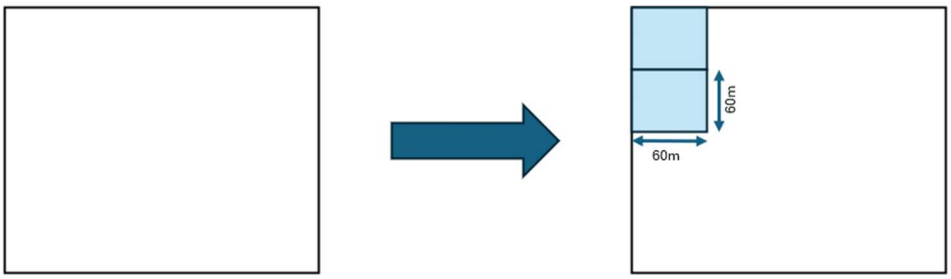


Figure 7: Representation of what Spatial resample means.

3.3.2 Temporal Resampling

The temporal resampling was trickier than the spatial one, but it had to be done only for two tiff files, ET and NDVI. To perform it, the format of the files downloaded from Google Earth Engine must be explained: ET has an image every eight days, but the pixels of every image contain a value that is the sum of the previous days, i.e., the image representing the 8th of January contains pixels with values that are the sum of the values of 1st, 2nd, 3rd, 4th, 5th, 6th, 7th, and 8th of January, while NDVI has an image every 16 days but the pixels only contain values calculated for that day. Regarding the ET resampling, the new daily images were created, splitting the images into eight new images, paying attention to the fact that the last image at the end of every year represented 5/6 days instead of 8. The NDVI resample was performed using a weighted average between existing images, creating the 15 new images between two existing images. In both cases, a Python code was used with the GDAL extension.

3.4 Model Development

After the acquisition and resampling of the data, we moved on to the creation of three datasets. This choice depended on the fact that it was chosen to develop a separate model for each zone based on the long-term deformation values of the InSAR data. Zone one (high subsidence) includes pixels with values less than -20mm, zone two (moderate) includes pixels with values between -20mm and -10mm, and zone three (low subsidence) contains pixels with values between 0mm and -10mm, see Figure 8. For the purpose of this research, the

uplifting values (> 0) have not been considered. Each dataset was created using the timeseries of the InSAR data, considering the dates to which they refer. Thanks to the resample, it was possible to collect the pixel values of the remote sensing data relating to the different zones and dates in which the InSAR timeseries data are present. By having the same spatial resolution, we were sure to have approximately the same number of pixels in the same position between georeferenced images and long-term deformation. Finally, by combining these new timeseries with those of the InSAR data, we were ready to feed them to the different models.

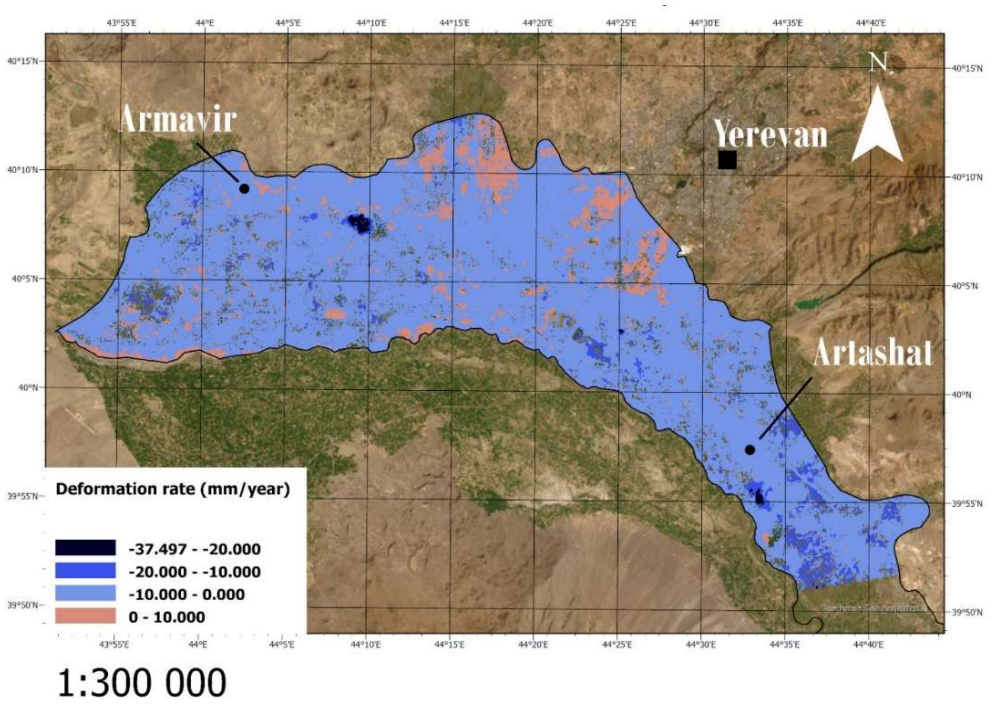


Figure 8: Division of the long-term deformation rate (presented previously) in the different zones modelled. Black zone 1, Dark blue zone 2, and Light blue zone 3. Moreover, the major cities in the area are shown on the map.

The LSTM algorithm with the TensorFlow package was utilized to create the model. Each model was calibrated differently due to the datasets being created with remote sensing and InSAR data by concatenating the values of each pixel for each zone. Thus, the number of data points per zone varies. However, each model followed the same schema.

The tanh activation function was used to introduce non-linearity into the model and to map input values into the range of [-1, 1]. This helps deal with the vanishing gradient problem often encountered in neural networks (LeCun et al., 2015). The Adam optimizer, which stands for Adaptive Moment Estimation, was employed to adjust the learning rate of the model. It combines the advantages of two other extensions of stochastic gradient descent: the ability to maintain a learning rate per parameter (AdaGrad) and the ability to improve the momentum method (RMSProp). This optimizer is particularly effective in handling sparse gradients on noisy problems (Kingma & Ba, 2014).

The learning rate determines the step size at each iteration while moving towards a minimum of the loss function. It's crucial for the convergence of the model, balancing between too fast, which can lead to overshooting the minimum, and too slow, which can result in a lengthy training process (Goodfellow et al., 2016). Dropout is a regularization technique that prevents overfitting by randomly dropping units (along with their connections) from the neural network during training. This ensures that the model does not rely too

heavily on nodes, thereby improving its generalization capabilities (Srivastava et al., 2014).

Each model was divided into batches based on the number of pixels per zone. Batching helps stabilize and speed up the training process by updating the model parameters after processing a subset of the training data rather than each data point (Ruder, 2016). The number of epochs refers to the number of complete passes through the training dataset. More epochs allow the model to learn, but too many can lead to overfitting. Finding an optimal number is critical for model performance (Bengio, 2012).

Each dataset was divided into training, validation, and testing. 60% of each dataset is for training, 20% is for validation, and 20% is for testing. Since each dataset comprises the concatenation of the pixels that form a zone, splitting the dataset means that some pixels were used to train the model, others to validate it, and the remaining part to test it. The testing phase was conducted subsequently to check the accuracy of the model's predictions. The Root Mean Squared Error (RMSE) and the coefficient of determination (R^2) were calculated for each model's testing phases. The RMSE metric measures the average magnitude of the errors between predicted and actual values. It helps understand how well the model performs in terms of prediction accuracy. The R^2 score, or the coefficient of determination, indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. It provides insight into the model's goodness of fit (Hastie et al.,

2009). These metrics were essential in evaluating the performance and accuracy of each LSTM model across different zones.

After the testing phase, we moved to a Feature Importance study. Using permutations, we studied the impact of each input variable on the final error of each model. The permutations allow us to modify each input value and observe how the Mean Square Error (MSE) changes, which measures the difference between the observed and predicted values. The higher the change in the RSE, the higher the importance of a feature will be.

Finally, a future projection of each model's trend was carried out. A linear Regression was used to estimate the course of every zone. It was important to give an idea of future changes to provide more solid help for future water management strategies.

4. Results

Training loss and validation loss in the different zones showed a downward trend. Both losses were similar and continued to decrease over time, see Figure 9. This shows that the model is not overfitting or underfitting. Also, the training loss and validation loss were reduced to near zero after iterating the entire training dataset with a different number of epochs for each model.

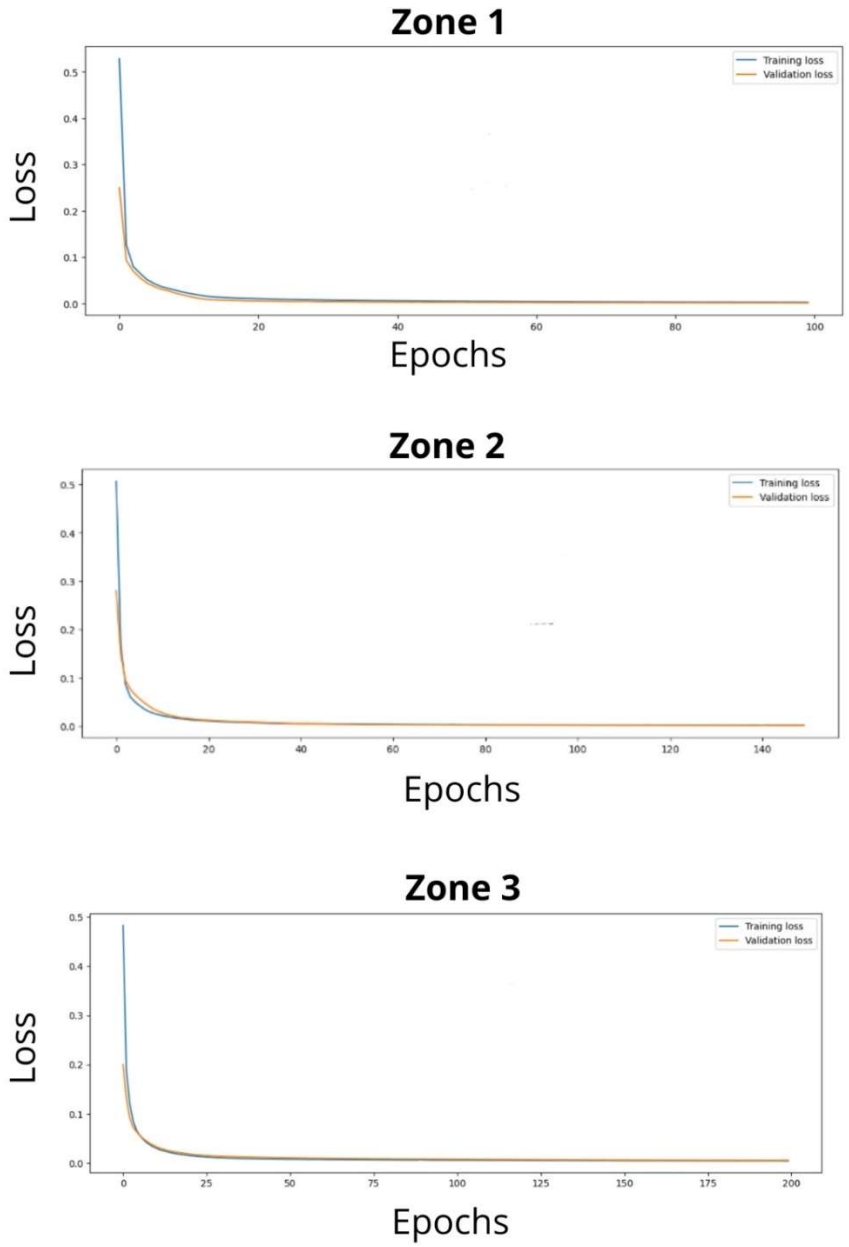


Figure 9: Loss function plots for the different models. 'Loss' indicates the summation of errors created for each sample in validation or training sets.

The testing phase of the different zones regarded evaluating the models' predictions over the actual data. The pixels used for the testing phase were not used during the previous phases, so the model did not 'see' them until this last phase. Figure 10 presents the testing phase of the High Subsidence zone. The testing phase of the other zones is presented in the Appendix.

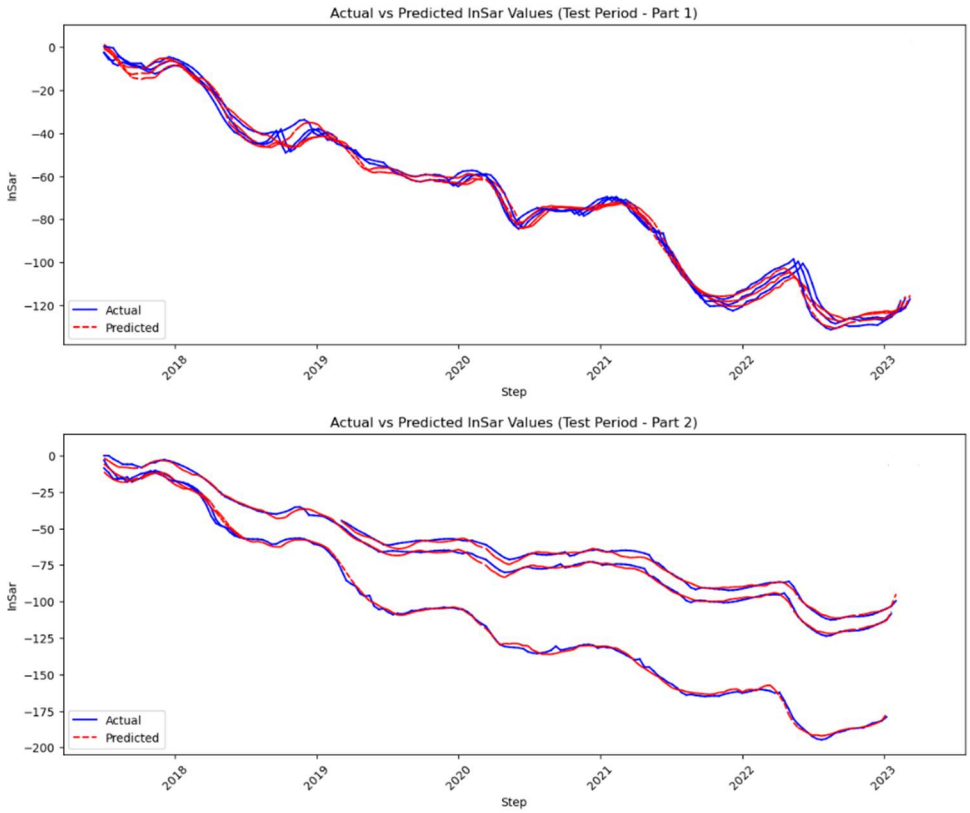


Figure 10: Testing phase of zone 1 (High Subsidence), every line represents a pixel. InSAR values are in mm. Step represent the timestep.

Table 5 presents the values of RMSE and R^2 used to evaluate the performance of the different models in the testing phase.

Table 4: RMSE and R^2 values of each model.

Zone	RMSE (mm)	R^2
1	2.43	0.91
2	2.92	0.89
3	3.21	0.87

Figure 11 presents the results of the importance calculated for every input feature.

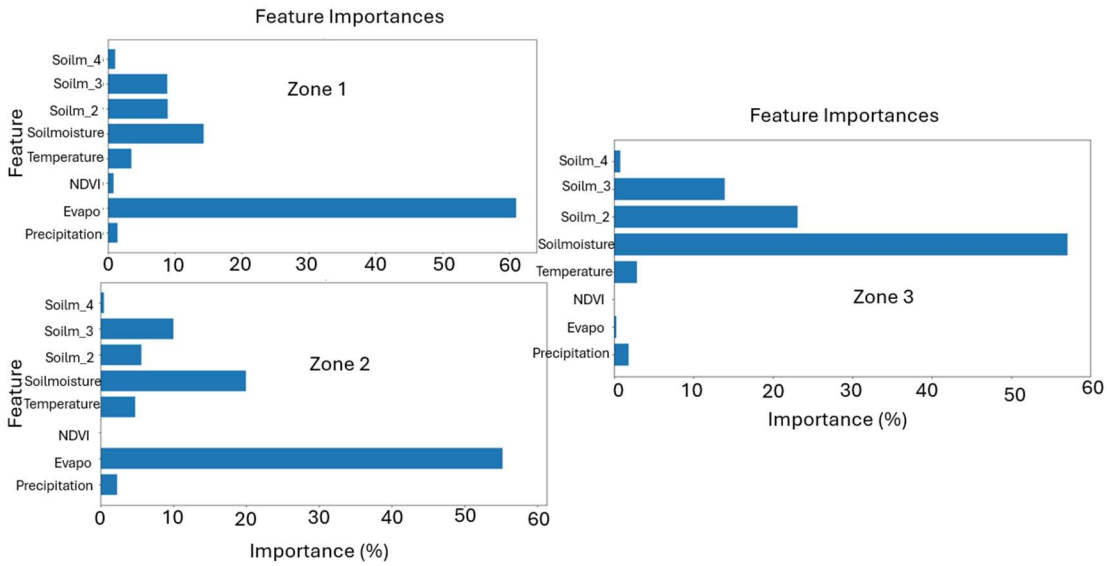


Figure 11: Feature importance in the models. The image shows bar charts depicting the permutation feature importance for three zones. Each chart highlights the

significance of various features, the values are represented in % to show how much they weight on the error. The abbreviations are used to easy the coding part.

Figure 12 shows the future projections of the trend for the different zones. The projections cover from the end of the dataset, 2023, to the end of 2025. Although the blue line indicates the average for visualization purposes, individual pixel data was used to study the future trend.

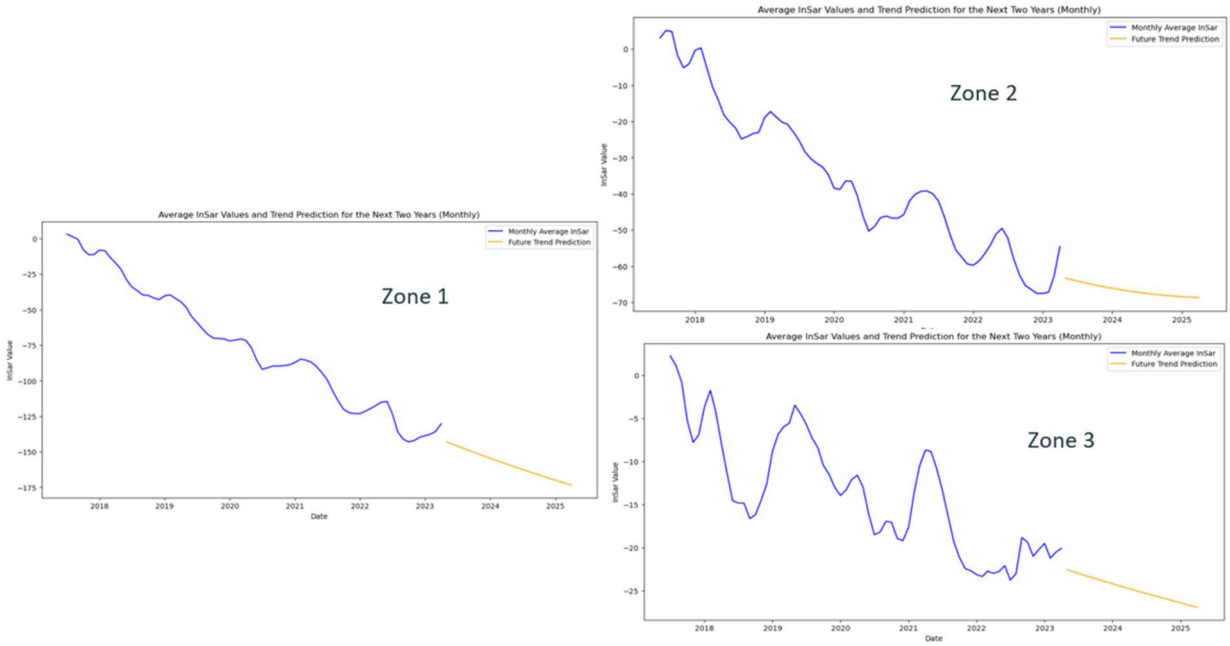


Figure 12: Graphs for Zones 1, 2, and 3. The blue line represents the average InSAR values for each zone, plotted monthly from 2018 to 2023. The orange line shows the predicted future trend.

5. Discussion

5.1 Data Utilization in the Research

The InSAR data used is from 173 Sentinel-1 A SAR images between July 2, 2017, and April 14, 2023. These data are to be considered reliable but not validated. There are several packages for extracting InSAR data. An example would be the LiCSBAS package, initially considered but discarded due to the more significant presence of noisy pixels in the images despite the more extended time coverage. The GMTSAR packet was less prone to noisy pixels and, therefore, was chosen for extracting land subsidence data. Figure 13 shows the visible difference in noisy pixels between the two packets. Thanks to the GPS data in our possession provided by GeoRisk, we can then validate both packages to understand which is the most reliable. For reasons of time and practicality, it has been chosen to avoid the validation part of this research.

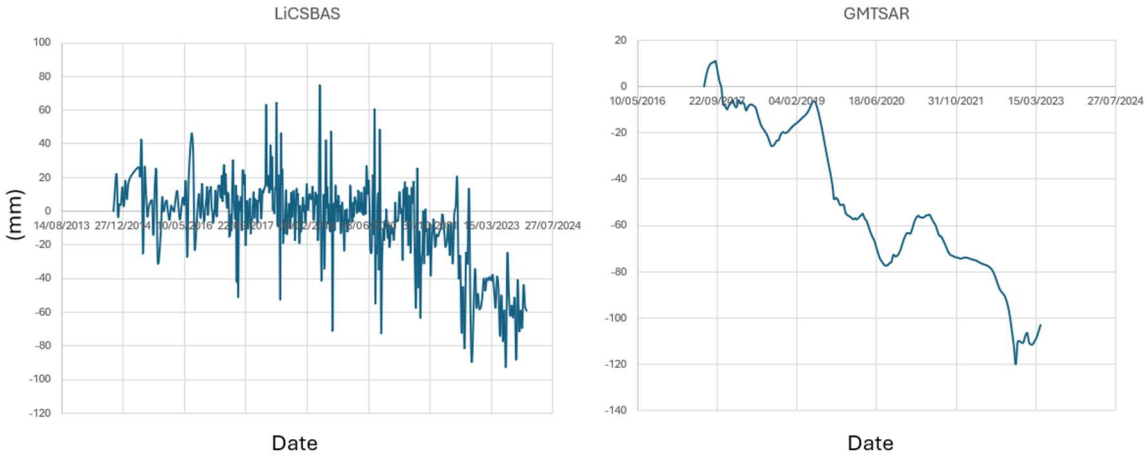


Figure 13: Comparison of land subsidence measurements in the High Subsidence zone using two different techniques: LiCSBAS (left) and GMTSAR (right). This comparison is only used to illustrate how LiCSBAS is

more susceptible to noisy pixels than GMTSAR, which shows a smoother and more consistent time series.

Remote sensing data was acquired from two different sources, GEE and CGLS. It was chosen to download the final product directly from each other's data, even though it is possible to calculate it using different satellite images to obtain the required products. This has led to different time series and relatively high spatial resolutions. Although the two resamples have led to images that are consistent with each other, it is necessary to consider some critical aspects. Switching from higher to lower resolution could lead to a loss of detail. On the other hand, interpolating data for the temporal resample can introduce values that do not reflect the real expected value. In both cases, some uncertainties can lead to incorrect data analysis.

5.2 Models' Performance

The quality of the input data is a fundamental aspect of modelling. As already discussed, the resampling may have increased uncertainty in the modelling phase and the results. However, the quality of InSAR data can also be compromised by various factors, such as environmental conditions. The structure of the model itself may not be considered adequate. The structure of the LSTM network, including the depth and number of neurons, must be appropriately calibrated to capture temporal dynamics without introducing excessive complexity. Training an LSTM model takes time and a large amount of data to adequately capture the long-term dependencies of sequential data. Despite the uncertainties and limitations, the results obtained by the LSTM models were satisfactory, demonstrating good accuracy in predicting land

subsidence. LSTM models are adaptable and can be used to model different datasets and scenarios. In addition, LSTM is an excellent algorithm for capturing long-term recurrences. This led to creating the three different models with more than satisfactory results.

A different approach was initially considered, instead of stacking the pixels, create an averaged timeseries for each zone. Each dataset would have been formed by the average of the values of the pixels of each zone. However, this choice has been discarded since the results were not satisfactory, see Figure 14. The reason is probably that the area is wide and averaging the data results in creating or losing long-term dependencies. This would result in a model with extremely high uncertainty and non-representative of reality.

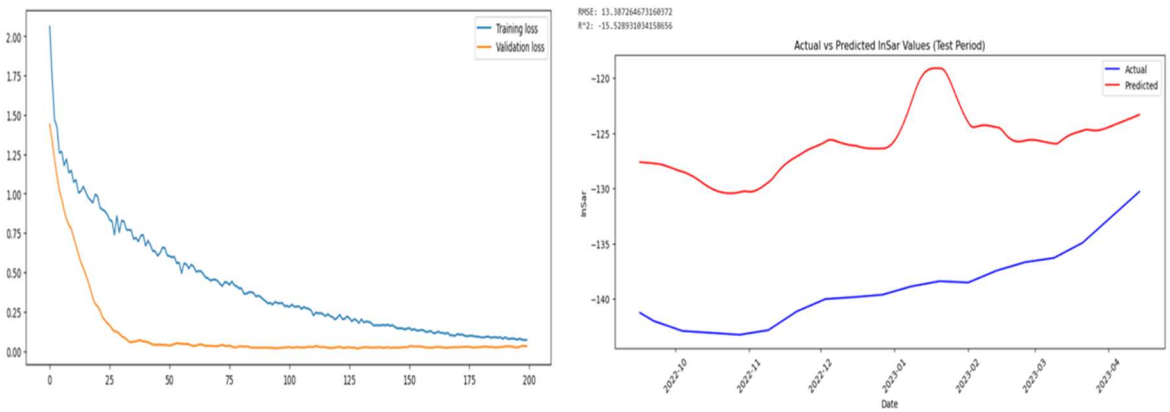


Figure 14: Sample results from zone 1. On the left is the trend of training and validation losses for 200 epochs. On the right is the comparison between the actual InSAR values and those expected during the test period, with an RMSE of 13.387 and an R^2 of -15.53.

Regarding the Feature's Importance, the graphs (Figure 11) show the different importance of the other zones. Evapotranspiration is an essential variable in zones 1 and 2. The reason behind this high dependence could be found in those areas' characteristics; Zone 1 is formed primarily by pits. The NDVI importance in all the zones is low or equal to zero. This result was unexpected because the increase and decrease of vegetation is usually directly correlated with exploitation and subsequent land subsidence. Soil moisture and WI were significant in Zone 2 and Zone 3. This indicates that water content significantly impacts subsidence in these areas, affecting soil cohesion and stability. Variability in soil content can cause soil contractions and expansions, contributing to the observed variations in subsidence values. Other factors, such as temperature and rainfall, show a variable influence, with a more significant impact in some areas than others.

The future projections show a negative trend in all three zones. As expected, Zone 1 has a steeper future trend; the most affected zone is likely to have a worse condition and a faster decline in soil level. Zones 2 and 3 also have a negative trend, but the difference in slope is notable. This shows that zones 2 and 3 are less affected by water exploitation for now, but in the future, this trend could increase in slope and make the lowering faster.

5.3 Future Model Improvement

The models' structures were kept as easy as possible. It was the right choice for this thesis, but a more advanced structure could improve the outcome. Moreover, this thesis presents a future projection of the trend of the next two

years, but LSTM allows to make predictions of a longer period, introducing a new input dataset after the creation of the model. For example, from the Copernicus website, it is possible to download different hydrogeological dataset predictions for the next 100 years. Introducing different datasets that represent different climate change scenarios could result in a better tool for policymaking.

The next step is to improve the well data (Hydromet). A suitable number of wells, in which all the characteristics are known, could result in interpolating the water level all over the valley or a part of it. Figure 15 shows how this relationship between the lowering of the water level and the lowering of the soil is correlated; a lowering of the water level corresponds to a subsequent lowering of the ground. A new law or regulation regarding water extraction or management would affect the water level in the valley and using it in the model and the forecasting would give a better understanding of their impact on the subsidence.

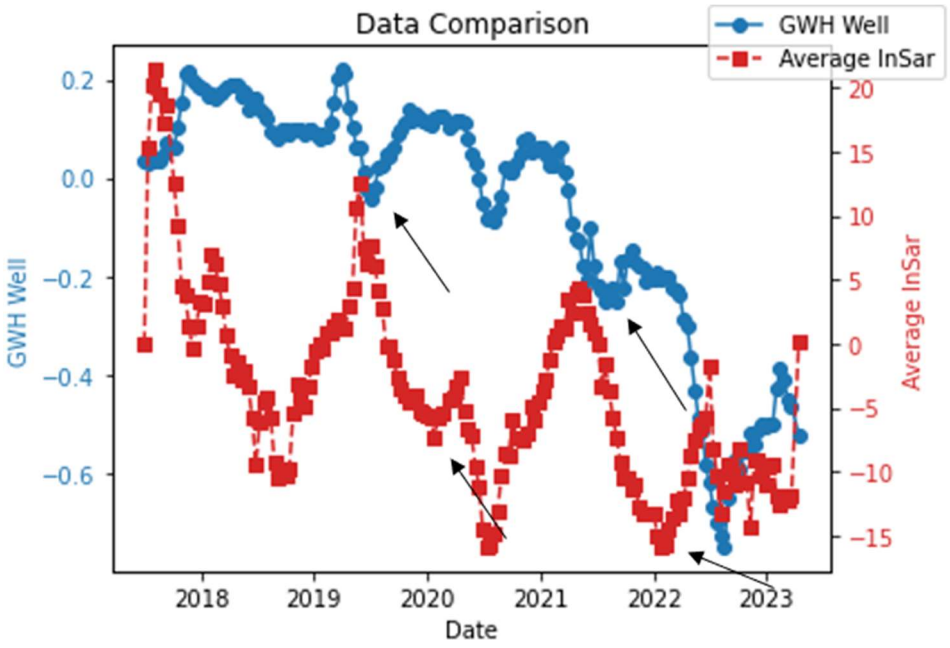


Figure 15: Comparison of water level and subsidence values. The blue line represents the GroundWater Head of a well monitored by Hydromet, while the red line represents the average InSAR data calculated within a 500 m radius around the well. The black arrows underline the agreement between the drop in groundwater levels and the corresponding land subsidence, which occurs after a certain time lag, as expected.

Regarding the improvement of the model, it can also be considered to incorporate additional types of data as inputs, such as spatial data of the valley. Using QGIS and downloading a Digital Elevation Model (DEM) file makes it possible to calculate various types of data, such as the slope of the valley or the topographic wetness index. Moreover, the most essential spatial data to include could be the geological conformation of the valley. Successfully reconstructing the clay layers in the valley or a part of it would make the model even more realistic. Additionally, spatial data can provide insights into the

hydrological connectivity and subsurface flow paths, which is critical for understanding groundwater-surface water interactions. However, integrating these spatial data into LSTM models is not straightforward but feasible. The challenge lies in the need to transform spatial information into a format that can be effectively used by the LSTM architecture, often requiring advanced preprocessing and feature extraction techniques.

5.4 Recommendations

The recent law enacted in January 2024, mandating the reuse of 40% of extracted water for extraction concessions, represents a potential step towards more sustainable water resource management. This legislation aims to address water scarcity issues by encouraging the recycling of water, thus reducing the overall demand for freshwater extraction. However, there are concerns regarding its effectiveness and practical implementation. During the field visit to Armenia, a fish farm owner mentioned that they already reuse 40% of their water. Despite this claim, the exact methods of how this water reuse is carried out were not made clear. This ambiguity raises questions about the actual enforcement and monitoring of the law. Without clear guidelines and transparency in practices, the true impact of the legislation remains uncertain.

Education plays a critical role in the success of sustainable water management practices. Currently, the population in the valley does not prioritize water conservation and often undervalues the importance of water. The focus tends to be more on the cost of electricity needed to pump the water rather than on the water itself. This perception can undermine efforts to conserve water.

The cost of water is presently very low, and increasing the price could potentially raise awareness about its value and encourage more conservative use. However, this approach must consider the economic conditions of the local population, as a sudden hike in water prices could face resistance and be economically burdensome for many residents.

6. Conclusions

InSAR data have been combined with LSTM networks to model and predict land subsidence in the Ararat Valley. Having to cope with difficulties of data reconciliation at different spatial and temporal resolutions, it was noted that the model gave very good accuracy in failure prediction. The present study showed that LSTM networks are particularly well suited for the extraction of long-term dependencies from sequential data and led to credible subsidence predictions. These predictions of land subsidence through the developed model can thus inform more sustainable water management strategies in Ararat Valley. These predictions might shape decisions on abstraction and policy implementations for groundwater that would help in addressing excessive consumption of water. The analyses underlined the importance of continuous monitoring and adaptive management of water resources for the mitigation of land subsidence risk. Advanced remote sensing and machine learning techniques are powerful tools in solving water management and land subsidence challenges. The results of the study will greatly provide a base for further research and implementation of sustainable strategies in the Ararat Valley, ensuring long-lasting economic and social stability within this region.

References List

Aaqilah (2023). Implementation of Long Short-Term Memory (LSTM). https://medium.com/@pennQuin/implementation-of-long-short-term-memory-lstm-81e35fa5ca54

Bauer-Marschallinger, B.; Freeman, V.; Cao, S.; Paulik, C.; Schaufler, S. ; Stachl, T. ; Modanesi, S. ; Massari, C. ; Ciabatta, L. ; Brocca, L. ; Wagner, W. Toward Global Soil Moisture Monitoring With Sentinel-1: Harnessing Assets and Overcoming Obstacles. IEEE Transactions on Geoscience and Remote Sensing 2019, 1 - 20. DOI 10.1109/TGRS.2018.2858004

Itger, B.; Paulik, C.; Hochstöger, S.; Mistelbauer, T.; Modanesi, S.; Ciabatta, L.; Massari, C.; Brocca, L.; Wagner, W. Soil Moisture from Fusion of Scatterometer and SAR: Closing the Scale Gap with Temporal Filtering. Remote Sensing 2018, 1019, 1030. DOI 10.3390/rs10071030

Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. In Neural Networks: Tricks of the Trade (pp. 437-478). Springer.

Bockstiegel, M., Richard-Cerda, J.C., Muñoz-Vega, E., Haghighi, M.H., Motagh, M., Lalehzari, R., Schulz, S., 2023. Simulation of present and future land subsidence in the Rafsanjan plain, Iran, due to groundwater overexploitation using numerical modeling and InSAR data analysis. Hydrogeol J. https://doi.org/10.1007/s10040-023-02657-y

Committee on Earth Observation Satellites (CEOS) (2018). The CEOS Earth Observation 46 (EO) Handbook-Catalogue of satellite instruments. The European Space Agency.

Didan, K. (2021). MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V061 [Data set]. NASA EOSDIS Land Processes Distributed Active Archive Center. Accessed 2024-05-07 from https://doi.org/10.5067/MODIS/MOD13Q1.061 NDVI

Funk, Chris, Pete Peterson, Martin Landsfeld, Diego Pedreros, James Verdin, Shraddhanand Shukla, Gregory Husak, James Rowland, Laura Harrison, Andrew Hoell & Joel Michaelsen. "The climate hazards infrared precipitation with stations-a new environmental record for monitoring extremes". Scientific Data 2, 150066. doi:10.1038/sdata.2015.66 2015. PRECIPITATION

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

Hochreiter, S., and J. Schmidhuber. 1997. Long Short-Term Memory. Neural Computation, 9: 1735-1780. https://www.bioinf.jku.at/publications/older/2604.pdf

Wan, Z., Hook, S., Hulley, G. (2021). MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V061 [Data set]. NASA EOSDIS Land Processes Distributed Active Archive Center. Accessed 2024-05-07 from LAND SURFACE TEMPERATURE https://doi.org/10.5067/MODIS/MOD11A1.061

Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

Liu, J., Liu, W., Allechy, F.B., Zheng, Z., Liu, R., Kouadio, K.L., 2024. Machine learning-based techniques for land subsidence simulation in an urban area. Journal of Environmental Management 352, 120078. https://doi.org/10.1016/j.jenvman.2024.120078

Liu, Z., Ng, A.H.-M., Wang, H., Chen, J., Du, Z., Ge, L., 2023. Land subsidence modeling and assessment in the West Pearl River Delta from combined InSAR time series, land use and geological data. International Journal of Applied Earth Observation and

Geoinformation 118, 103228. https://doi.org/10.1016/j.jag.2023.103228

Ministry of Environment of the Republic of Armenia and USAID (2021) The atlas of Ararat Valley. https://dec.usaid.gov

Naghibi, S.A., Khodaei, B., Hashemi, H., 2022. An integrated InSAR-machine learning approach for ground deformation rate modeling in arid areas. Journal of Hydrology 608, 127627. https://doi.org/10.1016/j.jhydrol.2022.127627

Peng, M., Motagh, M., Lu, Z., Xia, Z., Guo, Z., Zhao, C., Liu, Q., 2024. Characterization and prediction of InSAR-derived ground motion with ICA-assisted LSTM model. Remote Sensing of Environment 301, 113923. https://doi.org/10.1016/j.rse.2023.113923

Running, S., Mu, Q., Zhao, M., Moreno, A. (2021). MODIS/Terra Net Evapotranspiration Gap-Filled 8-Day L4 Global 500m SIN Grid V061 [Data set]. NASA EOSDIS Land Processes Distributed Active Archive Center. Accessed 2024-05-07 from https://doi.org/10.5067/MODIS/MOD16A2GF.061 EVAPOTRANSPIRATION

USGS (2018). Interferometric Synthetic Aperture Radar (InSAR). https://www.usgs.gov/centers/land-subsidence-in-california/science/interferometric-synthetic-aperture-radar-InSAR

Valder, J.F., Carter, J.M., Medler, C.J., Thompson, R.F., and Anderson, M.T., 2018, Hydrogeologic framework and groundwater conditions of the Ararat Basin in Armenia: U.S. Geological Survey ScientificInvestigations Report 2017–5163, 40 p., https://doi.org/10.3133/sir20175163

Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.

Sandwell, D., Mellors, R., Tong, X., Xu, X., Wei, M., Wessel. P, 2016.

GMTSAR: An InSAR Processing System Based on Generic Mapping Tools. UC San Diego: Scripps Institution of Oceanography.

Sandwell, D. ., R. . Mellors, X. Tong, M. Wei, and P. Wessel, 2011A. "Open radar interferometry software for mapping surface deformation." *Eos Trans. AGU* (2011): 92(28).

Sandwell, D., R.Mellors, X.Tong, & P.Wessel, 2011B. "GMTSAR: An InSAR processing system based on Generic Mapping Tools, technical report"

The European Space Agency. SAR (ERS) Interferometry https://earth.esa.int/eogateway/instruments/sar-ers/interferometry

Shi, X., Zhong, J., Yin, Y., Chen, Y., Zhou, H., Wang, M., Dai, K., 2023. Integrating SBAS-InSAR and LSTM for subsidence monitoring and prediction at Hong Kong international airport. Ore and Energy Resource Geology 15, 100032. https://doi.org/10.1016/j.oreoa.2023.100032

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research, 15, 1929-1958.

Appendix

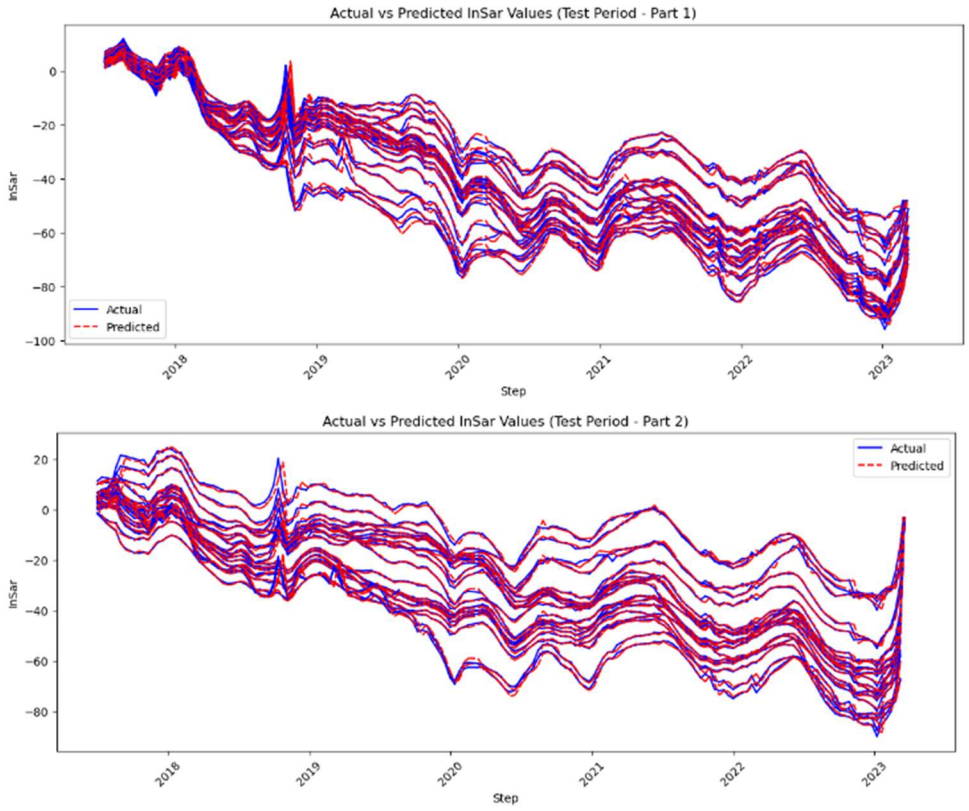


Figure 1A: Testing phase of zone 2 (Moderate Subsidence), every line represents a pixel. InSAR values are in mm. Step represents the timestep.

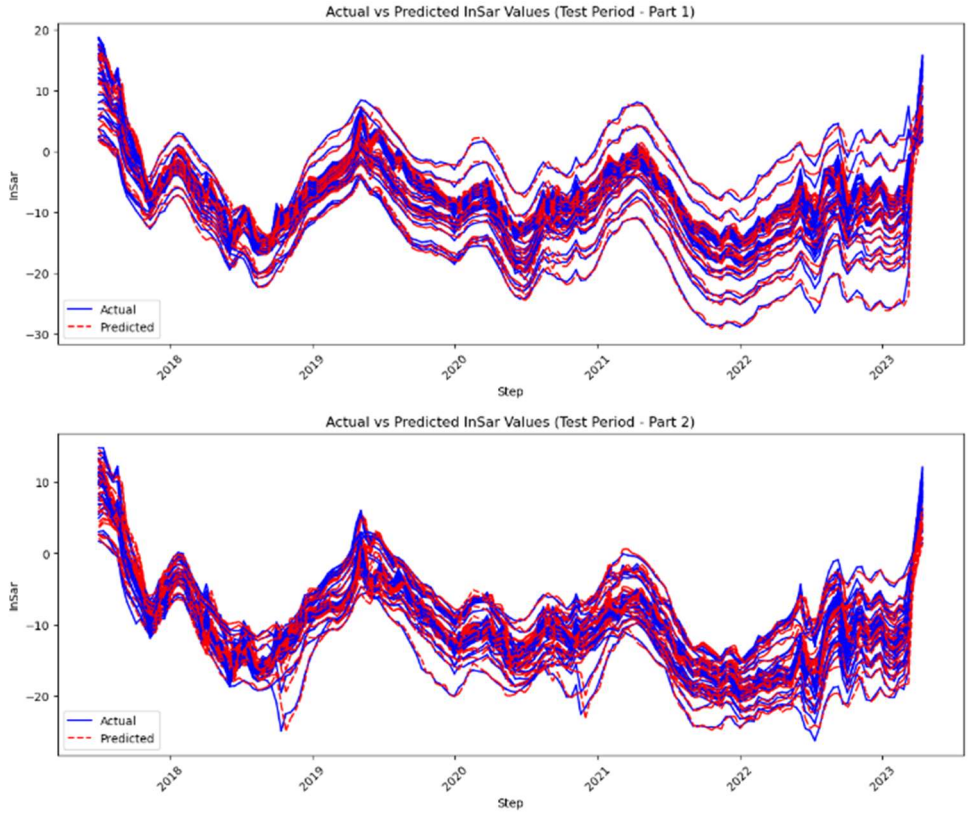


Figure 2A: Testing phase of zone 3 (Low Subsidence), every line represents a pixel. InSAR values are in mm. Step represents the timestep.