Developing a Spatio-Temporal Model to Predict InSAR-derived Hillslope Deformation

CHAKSHU GURURANI

June 2024

SUPERVISORS:
Dr. Hakan Tanyaş (First Supervisor)
Dr. Luigi Lombardo (Second Supervisor)
Prof.Dr.ir. Claudio Persello (Scientific Advisor)

Developing a Spatio-Temporal Model to Predict InSAR-derived Hillslope Deformation

CHAKSHU GURURANI
Enschede, The Netherlands, June 2024

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.

Specialization: Natural Hazards and Disaster Risk Reduction

SUPERVISORS:
Dr. Hakan Tanyaş (First Supervisor)
Dr. Luigi Lombardo (Second Supervisor)
Prof.Dr.ir. Claudio Persello (Scientific Advisor)

THESIS ASSESSMENT BOARD:
Prof.Dr. C.J. van Westen (Chair)
[Dr. Ivan Marchesini (External Examiner, Istituto di Ricerca per la Protezione Idrogeologica (IRPI), Italy)

ITC

ABSTRACT

Interferometric Synthetic Aperture Radar (InSAR) technology has revolutionized the monitoring of surface deformation, providing detailed and high-resolution data critical for assessing hillslope stability. Despite these advancements, integrating InSAR data with environmental parameters to predict deformation occurrences remains underexplored. This study aimed to bridge this gap by developing a novel spatio-temporal model to predict hillslope deformation by combining InSAR data from Sentinel-1, made available by the European Ground Motion Service (EGMS), with dynamic environmental variables obtained from ERA-5 dataset and static terrain properties from the British Geological Survey (BGS).

Utilizing InSAR's capability for precise surface deformation monitoring, a model using Graph Attention Networks (GAT) and Gated Recurrent Units (GRU) was developed, capturing both spatial and temporal dependencies in the dataset. The study was conducted for central England region, including the Peak District National Park, featuring significant surface deformation activity.

A correlation analysis was performed to explore relationships between dynamic variables—such as precipitation, temperature, Leaf Area Index (LAI), and terrestrial water storage (TWS)—and hillslope deformation, revealing varying correlations ranging from -0.4 to 0.6. The proposed GAT-GRU model achieved a final training loss of 0.00015. The predictions were evaluated using Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). The analysis indicated that the model captured the overall spatial and temporal dynamics. However, it struggled with extreme values and regions of high variability, with significant errors in regions of rapid surface movement beyond ±90 mm.

The findings demonstrate the model's ability to capture general trends and spatial patterns of deformation, including slow movements that are often overlooked in traditional landslide inventories and can lead to catastrophic failures. However, challenges remain in predicting extreme values and highly variable regions. This research advances the development of predictive tools for surface deformation monitoring and early warning systems. Future work should focus on improving model performance with higher resolution environmental data and exploring causal relationships between variables.

Keywords: Hillslope Deformation, InSAR, Graph Attention Networks, Gated Recurrent Units, Spatio-Temporal Model, Environmental Variables, Predictive Modelling

ACKNOWLEDGEMENTS

The path to get here was anything but straightforward, filled with many twists, turns, and bumps along the way. Without the support of those around me, it's hard to imagine having made it this far.

I would like to express my gratitude to my supervisors, Hakan and Luigi, for their guidance and kindness. Their academic acumen and expertise served as a guiding light, illuminating the path, and helping me move forward whenever I got stuck. I would like to thank Dr. Ing. Serkan Girgin for maintaining and providing the CRIB platform, which offered fast and efficient computation services to run the model. I am grateful to Ashok and Mei for their assistance in solving coding-related problems. Additionally, I would like to thank Bruno, for facilitating the entire process smoothly.

Thanks to my family: my parents, who are the motivation and reason behind everything I do and achieve, and Gayathri—thank you for being such a constant pillar of support.
My friends at ITC, thank you for being there for me, in sickness and in health: Salsa, Sachi, Shreya, Alma, Aparna, and everyone else (if I have ever cooked for you, you are included in this list, because food is my love language). You all believed in me when I lost faith in myself, and for that, I will always cherish your friendship. And Jelle- thanks for being the sunshine in my rainy days.

Immensely grateful to the Universe for my health and well-being throughout this nerve-wracking experience!

"There's always going to be another mountain,
I'm always going to make it move.
Always going to be another battle,
Sometimes I will have to lose.
It's not about how fast I get there, not about what's waiting on the other side,
It's all about the climb, it's all about the CLIMB..."

-M. Cyrus

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

1. INTRODUCTION

1.1. Background

Landslides cause significant harm to both lives and property. According to the Unified Landslide Database (UGLD), over 185,000 people have perished in 37,000 landslides from 1903 to 2020 (Gómez et al., 2023). Extensive research has been conducted to understand the various causal factors and their interactions that affect slope stability and contribute to landslides (Chen et al., 2020; Tanyaş et al., 2021). The literature predominantly focuses on creating susceptibility maps using either physically based (e.g., Van den Bout et al., 2021) or data-driven (e.g., Lombardo et al., 2020; Moreno et al., 2023) models, which can be heuristic (Corominas et al., 2014) or statistical (Pandey et al., 2023). Additionally, there has been an increase in the use of machine learning and deep learning algorithms for predictive landslide modelling (Dong et al., 2020).

In all these approaches, landslide inventories play a key role. In the realm of physically-based modelling, previously occurred landslides are essential to validate outputs of a given model. On the other hand, for data driven models, landslide inventories are required for both training and validating. This dependency on landslide inventories causes some limitation in the resultant hillslope stability assessments because of two main reasons. First, landslide inventories are mostly incomplete and thus non-landslide locations may not be fully represented in the dataset. Specifically, small landslides are mostly missed in many inventories (Van Den Eeckhaut et al., 2012; Tanyaş et al., 2019). Second, hillslope stability cannot be solely assessed by failed hillslopes. Slow hillslope deformation may always exist regardless of landslide occurrences. And yet, there are many reported cases that such slow moving landslides led catastrophic failures (e.g., Iverson, 2000; Nappo et al., 2019; Lacroix et al., 2020). This implies that monitoring and predicting hillslope deformations in addition to landslide occurrences could provide a better insight into hillslope stability assessments in general (Ji et al., 2020; Ganesh et al., 2023; Meng et al., 2023). In this regard, Interferometric Synthetic Aperture Radar (InSAR) technologies are invaluable. Spaceborne radar systems, particularly the Sentinel-1A and 1B satellites, provide SAR data with a 6–12-day temporal resolution globally, facilitating surface deformation monitoring on a regional and even global scale. For example, the Sentinel-1 mission's radar data has recently enabled comprehensive monitoring of surface deformation across Europe (Crosetto et al., 2021). Utilizing such deformation datasets could pave the way for developing advanced predictive tools and a new generation of early warning systems in the future.

1.2. Motivation and Research Gap

In the context of landslide monitoring, InSAR technologies have primarily been employed to observe slow-moving landslides (Y. Zhang et al., 2022; Dong et al., 2023). However, the development of predictive tools using InSAR-derived surface deformations that incorporate environmental parameters remains largely unexplored, with only a few exceptions. For example, He et al. (2023) utilized a multivariate statistical model, targeting mean slope velocity derived from InSAR data. They applied the Generalized Additive Mixed Model (GAMM) to analyse the contribution of various environmental factors to hillslope deformation. The study suggests that the accuracy of such predictive models could be enhanced by using high-resolution, precise input variables representing environmental factors like precipitation, snowfall, and groundwater table. Nava et al. (2023) evaluated various deep learning algorithms for forecasting landslide displacement, considering factors such as rainfall and reservoir water levels. They found that Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) outperformed 1D Convolutional Neural Network (CNNs). The study indicated that deep learning could

be effectively used in early warning systems for landslides and suggested that incorporating future weather forecast data could enhance predictive accuracy, despite potentially introducing uncertainty. Similar approaches have been also developed to predict deformation associated with land subsidence. For instance, Mirmazloumi et al. (2023) developed an early warning system to detect land subsidence in mining areas using the LSTM architecture. Radman et al. (2021) explored a weighted multivariate ensemble model to predict land subsidence, comparing it to individual models like LSTM, CNN, and multi-layer perceptron (MLP) networks. They concluded that the weighted ensemble model performed better in predicting subsidence. Nevertheless, none of these approaches are yet mature enough for operational use.

1.3. Deep Learning for Geohazards

Deep learning techniques in natural hazard analysis primarily focus on pixel-level image analysis (Ma & Mei, 2021). These techniques include CNN-based models for landslide detection (Sharma et al., 2017), Fully Convolutional Networks (FCN), and their variants such as U-Net and SegNet for object detection (Bianchi et al., 2021). These methods significantly reduce the time and resources traditionally needed for inventory generation. For example, Ji et al. (2020) implemented a CNN model with an attention mechanism, achieving over 95% accuracy in landslide detection.

One key advantage of deep learning models is their robustness to sampling methods. A study on landslide susceptibility evaluation (LSE) in China's Three Gorges Reservoir region by Hua et al. (2021) found that a Deep Neural Network (DNN) architecture outperformed traditional machine learning algorithms like Random Forest, Support Vector Machine (SVM), and Logistic Regression (LR) in verification accuracy. The study highlighted that deep learning surpasses machine learning algorithms by reducing overfitting and improving generalization.

For deformation prediction, Recurrent Neural Networks (RNN) are crucial because they can handle sequential data, such as time series. LSTM, a gated RNN variant that addresses the vanishing gradient problem, has been successfully used to predict landslide displacement caused by various factors (Xie et al., 2019; Meng et al., 2023). Efforts have also been made to predict InSAR time series using Independent Component Analysis (ICA) and LSTM (Peng et al., 2024). LSTMs can still lead to overfitting. GRUs offer a solution with fewer trainable parameters, faster training, and the ability to capture longer temporal dependencies (Chung et al., 2014). Predictive models for displacement using GRUs have demonstrated improved, or at least similar, predictive performance in a more computationally efficient manner (W. Zhang et al., 2022).

1.4. Novelty and Scientific Contribution

Although the above-mentioned algorithms have demonstrated impressive predictive capabilities, the mentioned recurrent networks only capture temporal sequences. Additionally, due to the irregular partitioning of the dataset in this research (into small catchments and slope units), CNNs are unsuitable since they work with regularly partitioned data, like pixels. To create a more representative spatio-temporal model, it is necessary to account for both irregular spatial relationships and temporal dependencies. Graph neural networks (GNNs) can capture this spatial dependence. Unlike CNNs, which operate in Euclidean spaces and use the distance between features, GNNs function in non-Euclidean spaces and consider neighbourhood relations (Jiang et al., 2022). GNNs have been successfully applied in various fields, such as forecasting virus spread (Tonks et al., 2022), modelling wildfire spread (Cisneros et al., 2023), and predicting traffic flow (Bhaumik et al., 2022).

In the field of land displacement, there have been a few attempts to incorporate spatial relationships using graph-based architectures. Ma et al. (2021) used a combination of graph convolution network and GRU to predict slope deformation. However, their approach was purely data-driven and external factors were not accounted for. Jiang et al. (2022) followed a similar approach. They applied graph convolution to GNSS displacement data to establish spatial dependence, then fed this output into a GRU network, thereby capturing both spatial dependence and temporal correlation (Figure 1). External triggering factors such as rainfall and reservoir water level were also incorporated but predisposing factors such as slope angle, were not included in the analysis. Their model outperformed ARIMA, SVM, and LSTM in prediction accuracy. Recently, Khalili et al. (2023) created a deformation prediction model based on Graph Convolution Networks (GCN) and LSTM. While they did utilize predisposing factors including slope and geology, dynamic environmental variables like rainfall and water storage capacity were not taken into account. Their model outperformed simple RNNs, GRU and simple LSTM, and the validation error was less than 4 mm for more than 90% of the data points. They suggested the use of more diverse variables in the future, along with attention-based mechanisms to capture more complex relationships in the dataset.

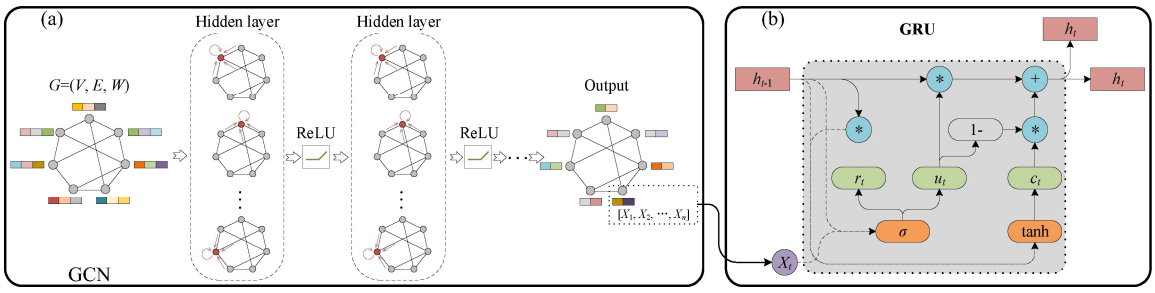


Figure 1 The architecture of (a) GCN and (b) GRU proposed by Jiang et al. (2022)

Given these research gaps and motivations, this thesis aims to build a novel spatio-temporal model for the prediction of hillslope deformation using GAT (Graph Attention) and GRU, while incorporating a number of predisposing factors such as slope and geology, as well as dynamic environmental factors such as precipitation and temperature.

1.5. Main Objectives and Research Questions

The main objective of this thesis is to develop a comprehensive spatio-temporal model that can predict hillslope deformation using InSAR-derived data. This involves not only understanding the complex relationships between environmental factors and surface deformation but also implementing advanced deep learning techniques to improve prediction accuracy. Specifically, the study aims to achieve the following objectives and address the related research questions (RQ):

A) To Examine the Correlation between Environmental Variables and Hillslope Deformation:

Objective A1: Investigate the relationships between dynamic environmental variables (e.g., precipitation, temperature, leaf area index, terrestrial water storage) and InSAR-derived surface deformations as an exploratory data analysis.

RQA1: How do precipitation, temperature, leaf area index, and terrestrial water storage relate to hillslope deformation?

Objective A2: Determine the optimal prediction and observation windows by analysing lagged correlations between these variables and hillslope deformation.

RQA2: What is the lag time for the effect of these environmental variables on hillslope deformation?

B. To Develop a Novel Spatio-Temporal Model for Predicting Hillslope Deformation:

Objective B1: Construct a spatio-temporal model incorporating both spatial and temporal dependencies using Graph Attention Networks (GAT) and Gated Recurrent Units (GRU).

RQB1: How accurately does the model predict hillslope deformation over different time periods and spatial extents?

Objective B2: Integrate predisposing factors such as slope and geology along with dynamic environmental factors into the model.

RQB2: What are the main limitations and challenges in the model's predictive capabilities?

2. STUDY AREA

This study will be conducted in an area covering 4,127 km² in central England (Figure 2), which includes the Peak District National Park and surrounding regions. The region experiences an average annual precipitation of 1,025 mm (compared to the England average of 985 mm) and an average temperature of 10.3°C (equal to the England average) (*Media Centre Facts and Figures: Peak District National Park*, n.d.).

From a regional point of view, the area is one of the five outcrops of British Carboniferous Limestone, featuring extensive karst topography. Overall, the area is characterized by Carboniferous rock formations deposited in a rift basin and affected by glacial scouring during the Quaternary period (Huggett, 2020). Spatial distribution of these carboniferous units is diverse over the study area, with limestone and blanket peat dominating the southern part, intermittent volcanic activity leading to basaltic layers, and fluvial deposits such as mudstones, sandstones, and shales in the northern and western areas, where active fluvial erosion occurs.

Landslides have been also reported in the area in relation with different geological formations/processes. For instance, periglacial and paraglacial weathering have weakened and eroded the strata, resulting in deep-seated landslides such as Alport Castles, one of the largest rotational landslides in the country (Donnelly et al., 2002). Additionally, shallow, rainfall-induced landslides are common in the region (Dixon & Brook, 2007; Huggett, 2020).

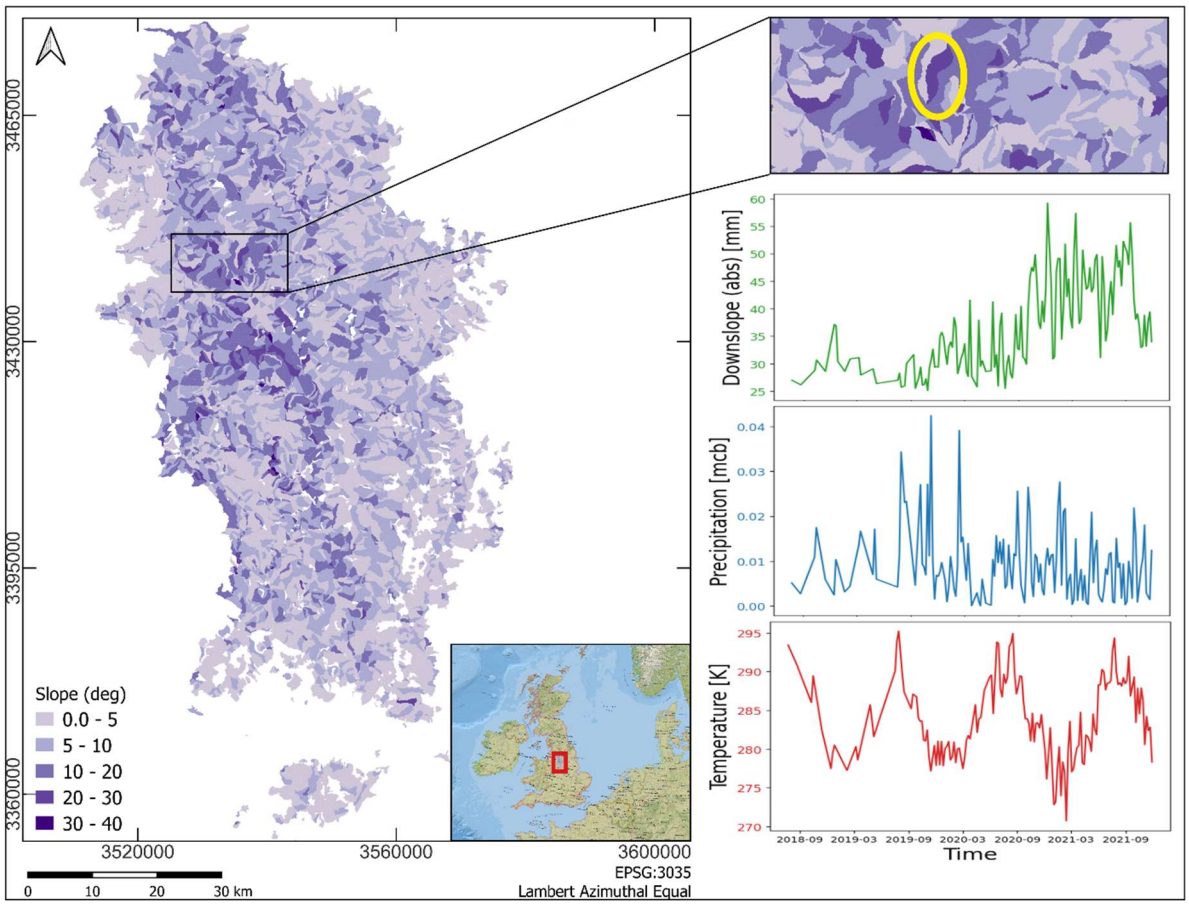


Figure 2 Map of the study area partitioned into slope units with associated time series sample data of downslope deformation, precipitation, and temperature.

3. DATASET: PREPARATION AND PREPROCESSING

This chapter details the process of data collection, the sources of the data, and the preprocessing steps taken to prepare the dataset for later analysis.

This study uses InSAR-derived surface deformations released by the European Ground Motion Service (EGMS) (Costantini et al., 2021), spanning a period of 6 years from 2016 to 2021, sampled every 7 days. To systematically examine this rich dataset, Slope Units (SUs) have been used as landscape partitions. SUs are geomorphologically meaningful mapping units showing similar aspect values and widely used to delineate hillslopes in the context of landslide hazard assessment (Reichenbach et al., 2018). For the study area, 6,352 SUs were generated (Figure 2) using the algorithm developed by (Alvioli et al., 2016). The scatterer giving the highest absolute velocity per SU was selected, and each SU was labelled with the corresponding vertical and east-west displacement time series. The target variable of the study is hillslope deformation, which has been provided as the net vector of east-west and vertical deformation. I converted these deformations to their downslope components (Cascini et al., 2010).

In addition to the surface deformation dataset, I also gathered several environmental factors to exploit them as the predictive variables of the model. Specifically, I used time series of temperature (weekly mean), precipitation (weekly cumulative), Leaf Area Index (LAI) and terrestrial water storage (TWS) provided by the ERA5 dataset (Hersbach et al., 2020). TWS is a measure of the total water content on the surface and the subsurface, including lakes, rivers, snow, ice, soil moisture (Girotto & Rodell, 2019). LAI is a dimensionless quantity that describes the canopy foliage. In barren land, the value is close to 0, and in the case of intensive timber cultivation, it can range from 16-45 (Bréda, 2008). In the given dataset, it ranges from 1.2 to 3, which is common in grasslands and shrublands. It indicates that the vegetation is not very dense. I sampled all these variables for SUs. These constitute the dynamic variables (Table 1).

I also generated terrain properties (i.e., static variables) such as slope, curvature, elevation, profile curvature, planar curvature and aspect using the Digital Elevation Model (DEM) provided by the British Geological Survey (BGS). In the study area, elevation ranges from 26 to 625 m, with a mean of 222 m and the mean slope angle is 6.8º. This shows that the area displays neither very high elevations nor very steep slopes. In addition to these variables, BGS also provided the upper soil thickness ang geology type, other static variables I used in the analyses.

Table 1 Summary of predictive variables used in the analyses.

	Variable Type	Variable	Temporal Resolution	Spatial Resolution	Preprocessing	Source
Target Variable	Time series	**Downslope displacement**	**Jan 2016-Dec 2021 Measured every 7 days.**	**Time series of absolute maximum deformation point per slope unit**	**Weekly observations**	Derived from EGMS vertical and EW deformation
Dynamic Variables		Rainfall	(Weekly cumulative)	Per Slope Unit	Scaled	ERA5 dataset resampled to ~11 km
		Temperature	(Weekly mean)			
		LAI				
		TWS				
Static	Continuous	Elevation	NA		Scaled (std=1; Mean=0)	BGS Dataset
		ruggedness				
		area_2				
		slop_avg				
		ss_thick				
		reg_thick				
		up_thick				
		profC_avg				
		planC_avg				
		northness				
		eastness				
	Categorical	geo			One-Hot Encoded	

4. METHODOLOGY

This chapter outlines the methodologies used in this study, detailing the steps taken to analyse the data and develop the predictive model. Section 4.1. provides an overview of the methodology related to the first objective and section 4.2. discusses the methods of the second objective including model construction. The preprocessing techniques applied are explained in section 4.3. Section 4.4. covers the model training process including data splitting, hyperparameter finetuning, loss function and model optimization.

4.1. Correlation Analysis

The first sub-objective is related to preliminary analysis of time series. Correlation provides an insight into the relationship between variables and have been used in the variable selection process for deep learning models (Ma et al., 2022). The literature shows that correlation between rainfall and deformation has shown complex interactions, that vary across regions. Studies have found significant correlation for lag time of 9 days in unstable regions with silt and clay (Lollino et al., 2006) to 9 months in the case of deep-seated landslides (Ardizzone et al., 2011). This study will attempt to capture the linear and non-linear relationships of the time-varying variables and deformation to investigate the lag period.

4.1.1. Pearson Correlation Coefficient

I used the Pearson Correlation Coefficient (PCC) to assess the variable relationships. PCC is a statistical measure that quantifies the strength and direction of the linear relationship between two continuous variables (Faizi & Alvi, 2023). This coefficient is calculated by dividing the sample covariance of the two variables by the product of their standard deviations.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

where x_i and y_i are the individual data points, and $\bar{x}$ and $\bar{y}$ are the means of the x and y variables, respectively.

Values of the PCC range from -1 to 1, where 1 indicates a perfect positive linear relationship, -1 indicates a perfect negative linear relationship, and 0 indicates no linear relationship (Rodgers & Nicewander, 1988).

The PCC is particularly useful for illustrating the strength of linear relationships. It may not recognize strong non-linear relationships and is sensitive to outliers, which can significantly distort the results. Despite its limitations, the PCC remains a widely used tool in statistical analysis for identifying and measuring linear relationships. It provides valuable insights into how variables are related and helps in understanding the nature of their association.

4.1.2. Lagged Correlation

While analysing the cause-and-effect relationship, there could be some lag time between two. For instance, hillslope deformation may occur sometime after a rainfall event because penetration of surface water to possible failure surface may take some time. This is particularly valid for deep seated landslides (Ardizzone et al., 2011). Therefore, assessing the lag time between dynamic and the target variables could provide some insights into deformation mechanism.

Lagged correlation is a statistical technique used to identify the relationship between two time series variables at different time lags. This method is useful when the impact of one variable on another may not be immediate but occurs with a delay.

The lagged correlation at lag k can be calculated based on the following equation:

$$r_k = \frac{\sum(x_{t+k}-\bar{x})(y_t-\bar{y})}{\sqrt{\sum(x_{t+k}-\bar{x})^2 \sum(y_t-\bar{y})^2}}$$

where x_t and y_t are the time series data points at time t, and $\bar{x}$ and $\bar{y}$ are the mean values of the respective series.
By shifting one time series relative to another and calculating the correlation for each lag, the time delay at which the relationship is strongest, can be identified.

Lagged correlation provides valuable insights into the temporal dynamics between variables, helping to reveal delayed effects and predict future behaviour based on past data. In this study case, a lagged correlation can be used to establish how past rainfall affects current downslope deformation, thus, providing an insight into the effect of current rainfall on future deformation trends.

4.2. Proposed Spatio-Temporal Model [GAT+GRU]

In order to establish a model that is truly spatio-temporal in nature, I used a combination of GAT and GRU. The model uses the PyTorch layer GAT from the 'PyTorch Geometric' library (Fey & Lenssen, 2019). It is based on the structure introduced by Veličković et al. in 2017. This model aims to capture both spatial dependencies among nodes in a graph and temporal dependencies across multiple time steps (Figure 3).

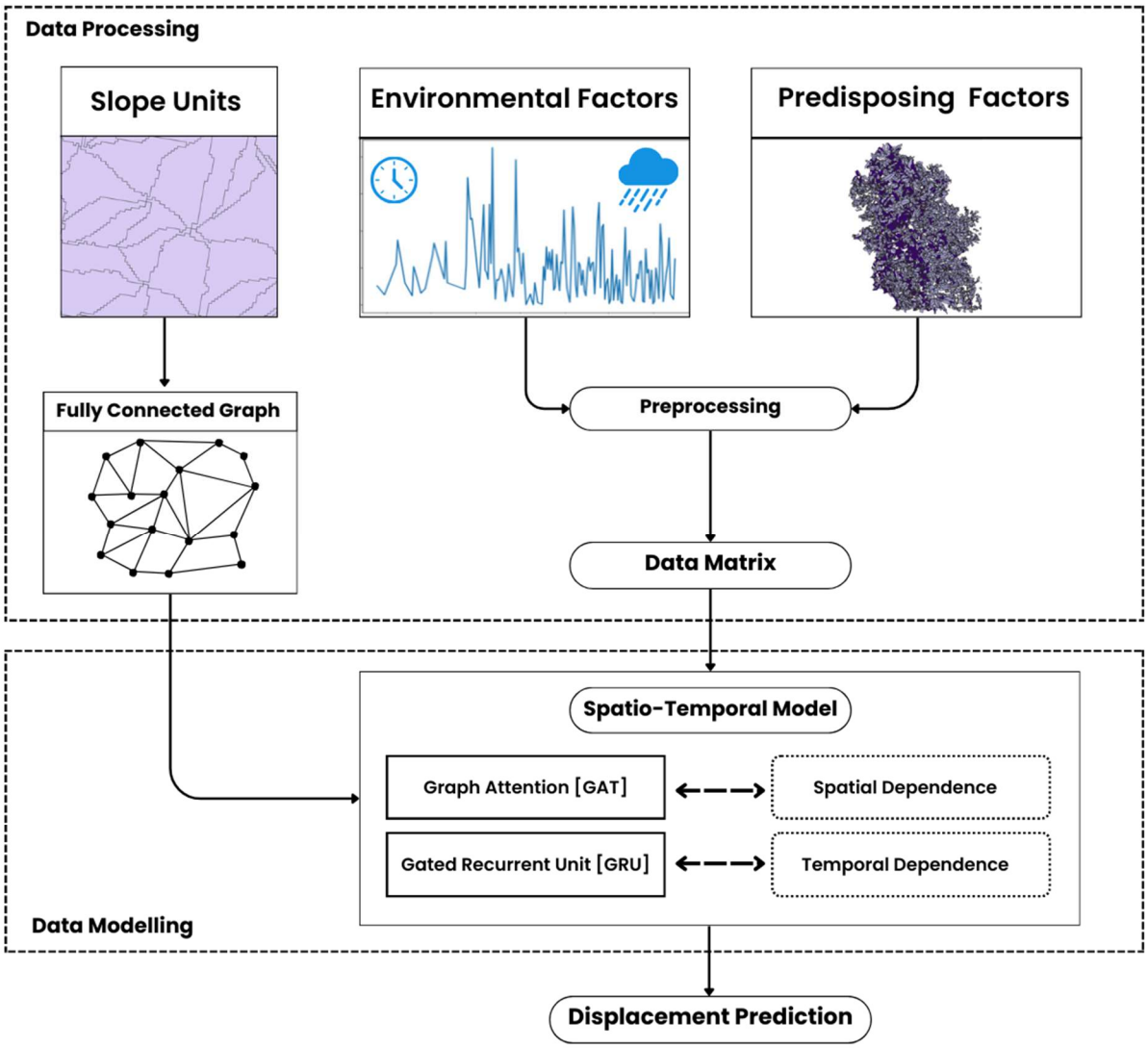
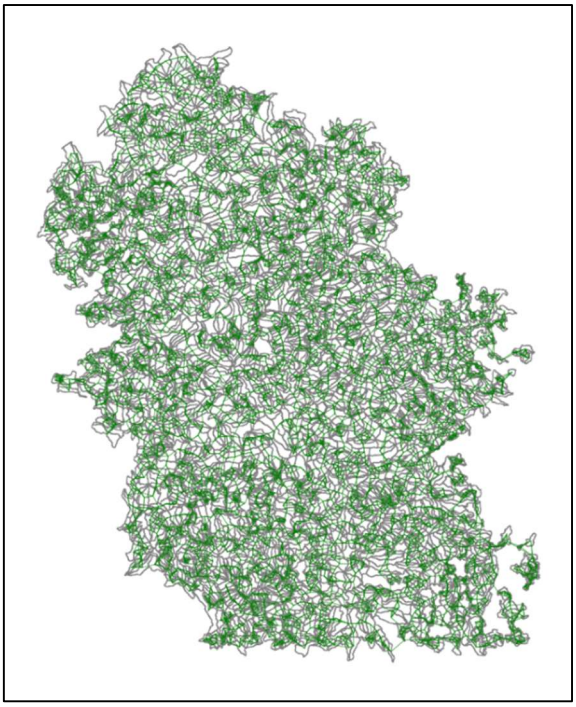


Figure 3 Conceptual Diagram of the proposed model

4.2.1. Adjacency Matrix

An adjacency matrix is crucial for the graph attention layer. It ensures that spatial dependencies are captured in the model in the form of a graph structure. I generated the adjacency matrix using the 'libpysal' library in python (Rey & Anselin, 2007). It uses the k-nearest neighbours method to create spatial weights with k=4, meaning each node (slope unit) in the graph will be connected to its four nearest neighbours (Figure 4).



(a)

Adjacency Matrix

```
[[0. 0. 0. ... 0. 0. 0.]
 [0. 0. 1. ... 0. 0. 0.]
 [0. 1. 0. ... 0. 0. 0.]
 ...
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]]
```

4688 * 4688

(b)

Figure 4 (a) Graph Structure defined by (b) Adjacency Matrix

4.2.2. Graph Attention Layer (GAT)

The core component of the Spatio-Temporal Model is the Graph Attentional Layer, which is inspired by the work of Velickovic et al. (2018). This layer performs self-attention on the nodes of the graph to compute attention coefficients that indicate the importance of neighbouring nodes' features to each node. The attention mechanism used in this layer allows for assigning different importances to different nodes within a neighbourhood while handling different sized neighbourhoods efficiently (Figure 5).

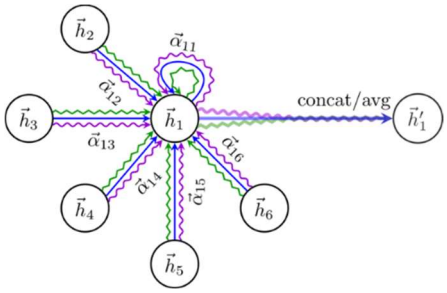


Figure 5 Multi-head attention with 3 heads for node 1. The features are then concatenated or averages to produce new weights $\vec{h}_1$' (Veličković et al., 2017)

4.2.2.1. Input and Linear Transformation

Given a set of node features $h = \{h_1, h_2, \ldots, h_N\}$, where $h_I \in R^F$ and N is the number of nodes, the layer applies a shared linear transformation to these features using a weight matrix $W \in R^{F_0 \times F}$. The transformed features are then used to compute attention coefficients.

4.2.2.2. Attention Mechanism

The attention coefficients a_{ij} are computed using a single-layer feedforward neural network parameterized by a weight vector $a \in R^{2F_0}$ and applying the LeakyReLU nonlinearity. The attention coefficients are then normalized across all neighbouring nodes using the softmax function:

$$\alpha_{ij} = \text{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in N_i} \exp(e_{ik})}$$

where N_i is the neighbourhood of node i in the graph and e_{ij} are the edge features. This formula normalizes the edge features using the softmax function, which ensures that the attention coefficients sum up to 1.

4.2.2.3. Multi-Head Attention

To stabilize the learning process and enhance the model's expressive power, multi-head attention is employed. In this setting, K independent attention mechanisms are used, each producing its own set of attention coefficients. The output features from each attention mechanism are then aggregated, resulting in the following output feature representation for each node:

$$\mathbf{h}_0^i = \sigma\left(\frac{1}{K} \sum_{k=1}^{K} \sum_{j \in N_i} \alpha_{kij} \mathbf{W}_k \mathbf{h}_j\right)$$

where σ is a nonlinearity function such as ReLU, $\mathbf{W_k}$ is the weight matrix for the k-th attention mechanism, and α_{kij} are the normalized attention coefficients computed by the k-th attention mechanism.

4.2.3. Gated Recurrent Unit (GRU)

Following the Graph Attentional Layer, the model utilizes a Gated Recurrent Unit (GRU) layer to capture temporal dependencies in the data. The output features from the Graph Attentional Layer are passed through the GRU layer, which produces hidden states representing the temporal information at each time step.

GRUs are gated recurrent networks with two gates: a reset gate and an update gate (Figure 6). The reset gate controls the amount of past information to discard, while the update gate regulates the amount of new information to preserve. This design enables GRUs to efficiently capture temporal dependencies within sequential data (Chung et al., 2014).

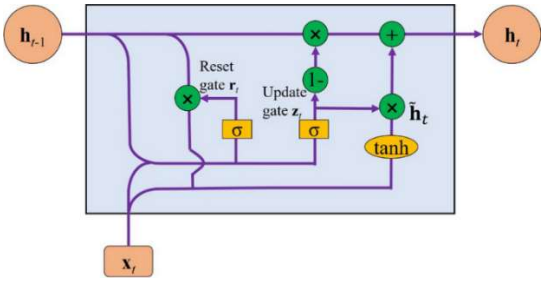


Figure 6 A GRU Unit with reset gate (r_t) update gate (z_t) and the current hidden layer vector ħ$_t$

(W. Zhang et al., 2022)

The equations governing the GRU operation are as follows:

$$
\begin{aligned}
z_t &= \sigma(\mathbf{W}_z \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_z) \\
r_t &= \sigma(\mathbf{W}_r \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_r) \\
\tilde{\mathbf{h}}_t &= \tanh(\mathbf{W} \cdot [r_t \odot \mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}) \\
\mathbf{h}_t &= (1 - z_t) \odot \mathbf{h}_{t-1} + z_t \odot \tilde{\mathbf{h}}_t
\end{aligned}
$$

where:

- x_t is the input at time t
- h_t is the hidden state at time t
- z_t is the update gate
- r_t is the reset gate
- ħt is the new memory content

Output

Finally, a linear layer is applied to the output of the GRU layer to obtain the final prediction for each node.

4.3. Preprocessing

To prepare the data for the model, it must first be preprocessed and transformed into a suitable format. The following sections describe these preprocessing steps in detail.

4.3.1. Normalizing Aspect into Eastness and Northness

The aspect, which defines the flow direction of a slope, is a cyclic property which ranges from 0º to 360º, where a north-facing slope will have an aspect of 0º, 90º is an east-facing slope, 180º means south-facing and a west-facing slope will have an aspect of 270º (Olaya, 2009). Since, aspect ‘a’ and a+360º refer to the same slope direction, in order to convert it into a more meaningful input variable, it is transformed to northness and eastness using trigononmetric functions (Wang et al., 2016).

$$Northness = cos\ (Aspect)$$
$$Eastness = sin\ (Aspect)$$

Northness of 1 represents a north-facing slope, while -1 represents a southwards slope. Similarly, for eastness, a value of 1 represents an east-facing slope and vice versa.

4.3.2. Z-score Normalization

Z-score normalization method was employed during data preprocessing to standardize features. It involves adjusting features by subtracting their mean and dividing by their standard deviation, resulting in a transformation where features have a mean of zero and a standard deviation of one (Figure 7).

$$z = \frac{x - \text{mean}}{\text{std}}$$

This procedure ensures that each variable contributes fairly to the model's learning process, preventing any single feature from exerting unequal influence due to its scale (Faraway et al., 2011). For instance, in our dataset, elevation ranges from 100 to 650 m, whereas another variable, 'upper soil thickness' ranges from 0 to 1.6 m. This difference in magnitude might cause the model to treat the two variables differently. Thus, to ensure equal contribution of all variables, this normalization was performed.

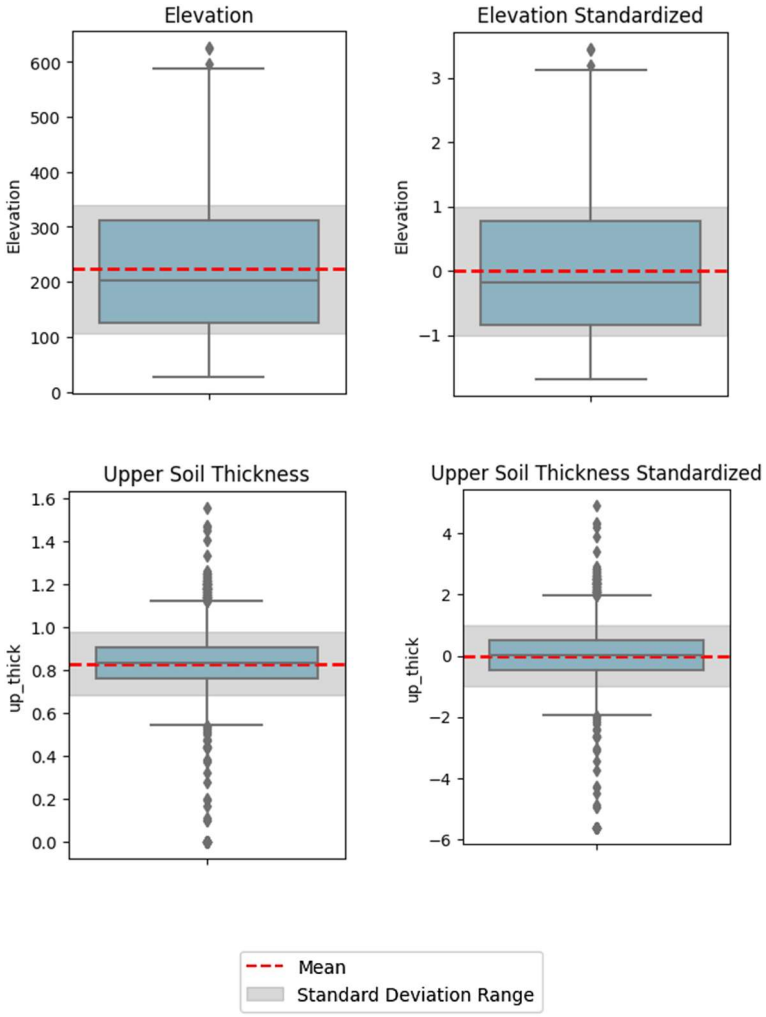
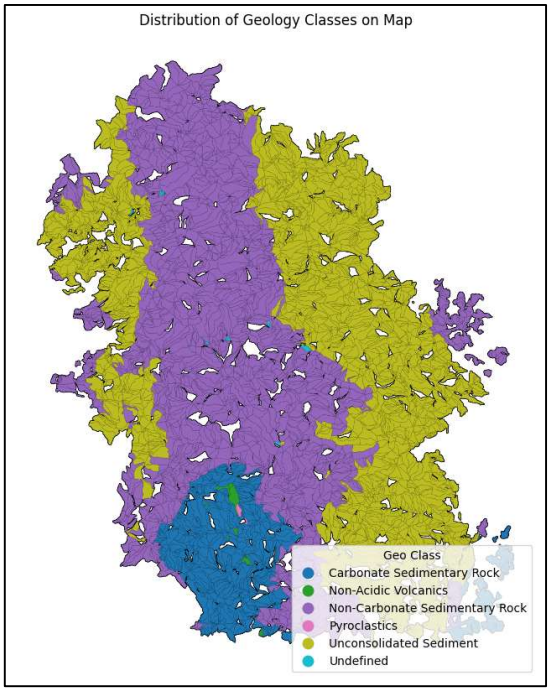


Figure 7 Example of two variables, Elevation and Upper soil thickness, before and after normalization

4.3.3. One-hot Encoding

One hot encoding is a technique used in data preprocessing to convert categorical variables into a binary format that is more suitable for machine learning algorithms (Brownlee, 2020). It involves creating binary columns for each category within the categorical variable and assigning a value of 1 to the column corresponding to the category present in each observation, while assigning 0 to all other columns. In this dataset, the geology column, 'geo' had 6 classes (Figure 8b). These were one-hot encoded to convert them to suitable input for the model (Figure 8a).



(a)

cat	original_geo	Carbonate Sedimentary Rock	Non-Acidic Volcanics	Non-Carbonate Sedimentary Rock	Pyroclastics	Unconsolidated Sediment	Undefined
200719	Non-Carbonate Sedimentary Rock	0	0	1	0	0	0
201594	Unconsolidated Sediment	0	0	0	0	1	0
204425	Undefined	0	0	0	0	0	1
212351	Carbonate Sedimentary Rock	1	0	0	0	0	0
213294	Non-Acidic Volcanics	0	1	0	0	0	0
214361	Pyroclastics	0	0	0	1	0	0

(b)

Figure 8 (a.) Map depicting the distribution of the various geological classes. (b.) One-hot encoded geology variable for model input

4.4. Model Training

With the model constructed and the data preprocessed, the next step was to train the model. The model training process is described in the following sections.

4.4.1. Data Preparation and Splitting

The dataset spans a period of 6 years, from January 2016 to November 2021. After preprocessing, the weekly dataset consists of a total of 360 timesteps. To ensure effective training and unbiased evaluation, the dataset was divided into three parts: training (2/3), validation (1/6), and test sets (1/6) (Figure 9). This division, based on time, ensured that all nodes of the graph were present in each of the three datasets, maintaining a consistent graph structure while having different time steps for each set (Table 2). This consistency is crucial for the integrity of the model, as it allows the model to learn and evaluate on the same set of nodes, differing only in the time steps used. The structured splitting was done to prevent data leakage and to ensure unbiased results.

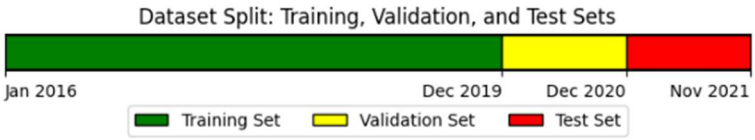


Figure 9 Dataset splitting into training, validation, and test sets

Table 2 Training, validation, and test set

Dataset	Time Period	Time Steps	Purpose	Nodes
Training set	Jan 2016- Dec 2019	240	Model Training	All graph nodes
Test Set	Dec 2019- Dec 2020	60	Model Evaluation (prediction)	All graph nodes
Validation Set	Dec 2020- Nov 2021	60	Hyperparameter finetuning	All graph nodes

4.4.2. Hyperparameter Finetuning

Hyperparameters are set before the model's learning process begins, unlike model parameters, which are learned during the training process from the data itself. They affect the learning process (Yang & Shami, 2020). The three hyperparameters in this model were- learning rate, GRU units and Batch Nodes (Table 3).

Table 3 Model Hyperparameters and their importance

Hyperparameter	Description	Importance
Learning Rate (LR)	Controls the step size at each iteration while moving toward a minimum of the loss function.	Affects the speed and quality of the convergence. A very high learning rate can cause the model to converge too quickly to a suboptimal solution, while a rate too low can result in a long training process.

GRU Units	Number of units (neurons) in each GRU (Gated Recurrent Unit) layer.	Determines the model's capacity to learn from sequential data. More units can capture more complex patterns but also increase the risk of overfitting and computational cost.
Number of Batch Nodes	Equivalent to the batch size, this is the number of nodes in each batch during training.	Impacts the model's ability to generalize. A larger batch size can provide a more accurate estimate of the gradient but requires more memory and can lead to longer training times. The adjacency matrix is trained in batches of fully connected nodes, to prevent model complexity. A method 'get_connected_nodes' (elaborated in Appendix 9.1) is defined to ensure that the nodes in each batch are fully connected.

Hyperparameter finetuning refers to adjusting the values of hyperparameters, to find an optimum solution- the best model performance. This was done by means of permutations. Different combinations of hyperparameter values were run on the validation set in nested loops, and the best combination was determined based on the overall loss for each epoch (detailed results in Appendix 9.2). The best performance was achieved with a learning rater of 0.01, 64 GRU units and 256 batch nodes. With this combination of hyperparameters, the model achieved a final loss of 0.00015, and displayed stable convergence over the epochs (Figure 10).

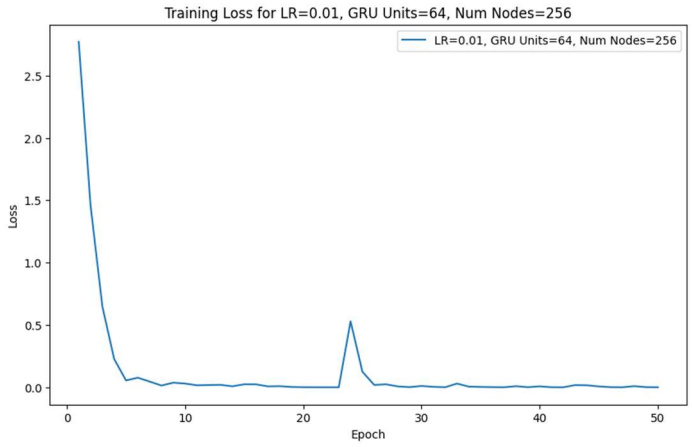


Figure 10 Training loss for the best hyperparameters over 50 epochs.

Best hyperparameters: {'lr': 0.01, 'gru_units': 64, 'num_nodes': 256}
Final loss for LR=0.01, GRU Units=64, Num Nodes=256: 0.000156

4.4.3. Loss Function and Optimizer

During training, the loss function quantifies the accuracy of the model by finding the similarity between the actual and predicted values. Since this is a predictive model for sequential data, Mean Squared Error (MSE) was used as the loss function. It tries to minimize prediction inaccuracies across multiple time steps, by penalizing larger errors more significantly (Brownlee, 2020). It computes the average squared difference between the predicted and actual values during training.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

where, n is the number of samples (or data points), y_i represents the actual value of the target variable for the *i*-th sample, $\hat{y}_i$ represents the predicted value.

The optimizer updates the model's weights to minimize loss function. The Adaptive Moment Estimation optimizer (Adam) was used (Kingma & Ba, 2014). The Adam optimizer is beneficial because it adapts the learning rate for each parameter, leading to faster and more efficient training. The adaptive learning rates for different parameters are calculated based on the first and second moments

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$$
$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

of gradients (m_t and v_t respectively). It incorporates momentum by using the moving averages of gradients.

Where, β_1 and β_2 are the decay rates for these moment estimates, g_t is the gradient of the loss function w.r.t. the parameters at time step *t*.

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t}$$
$$\hat{v}_t = \frac{v_t}{1-\beta_2^t}$$
$$\theta_t = \theta_{t-1} - \frac{\eta}{\sqrt{\hat{v}_t}+\epsilon}\hat{m}_t$$

η is the learning rate, θ_{t-1} represents the parameters from the previous time step, and θ_t is the updated parameter after applying the Adam optimization rule. These parameters of the model include the weights and biases, which need to be adjusted during training to minimize the loss function.

Once the model was trained on the training set, it was applied to the test set. The predicted values were then used to evaluate the model's performance, as discussed in Section 5.2.

5. RESULTS

This chapter presents the results of the analyses and the performance of the developed predictive model. Section 5.1 provides the findings from the correlation analysis, detailing the relationships between dynamic environmental variables and hillslope deformation. Section 5.2 discusses the performance of the spatio-temporal model.

5.1. Correlation Analysis

The objective was to assess the relationship between the dynamic variables and the target variable. For instance, to find the relationship between the precipitation at time *t-1* and downslope displacement at time *t*, the precipitation column was shifted by one step. Since each data point in the dataset represents weekly observations, this shift corresponds to a time lag of 7 days. Similarly, lagged correlation was also calculated between target variable and LAI, TWS and temperature (temp), respectively.

For lag=1, positive correlation was obtained with precipitation for most slope units. While the other three variables resulted in negative correlation values for the majority of slope units (Figure 11).

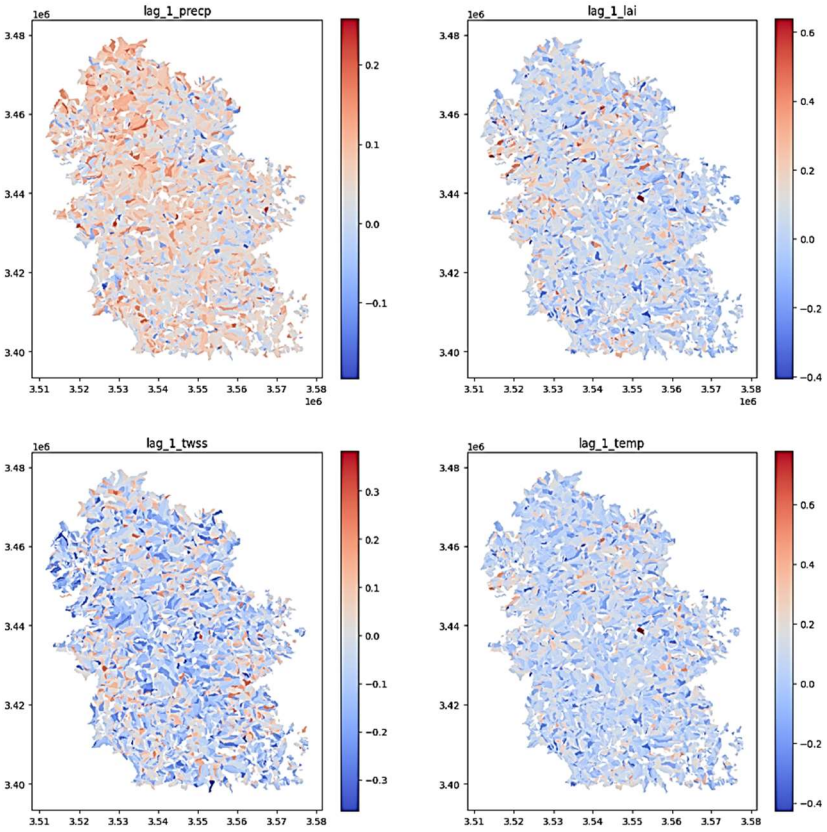


Figure 11 Spatial distribution of correlation values for the entire dataset, for a time lag= 1 (1 week)

Similar analysis was done was a lag of 2, 4, 8 and 12 (corresponding to a lag time of two weeks, one month, two months and 3 months respectively). The correlation values either did not change, or reduced further, as the lag time was increased (Figure 12). Overall, the correlation values with downslope displacement ranged from -0.2 to 0.3 for precipitation, -0.4 to 0.6 for temperature, -0.4 to 0.5 for TWS and between -0.4 to 0.6 for LAI.

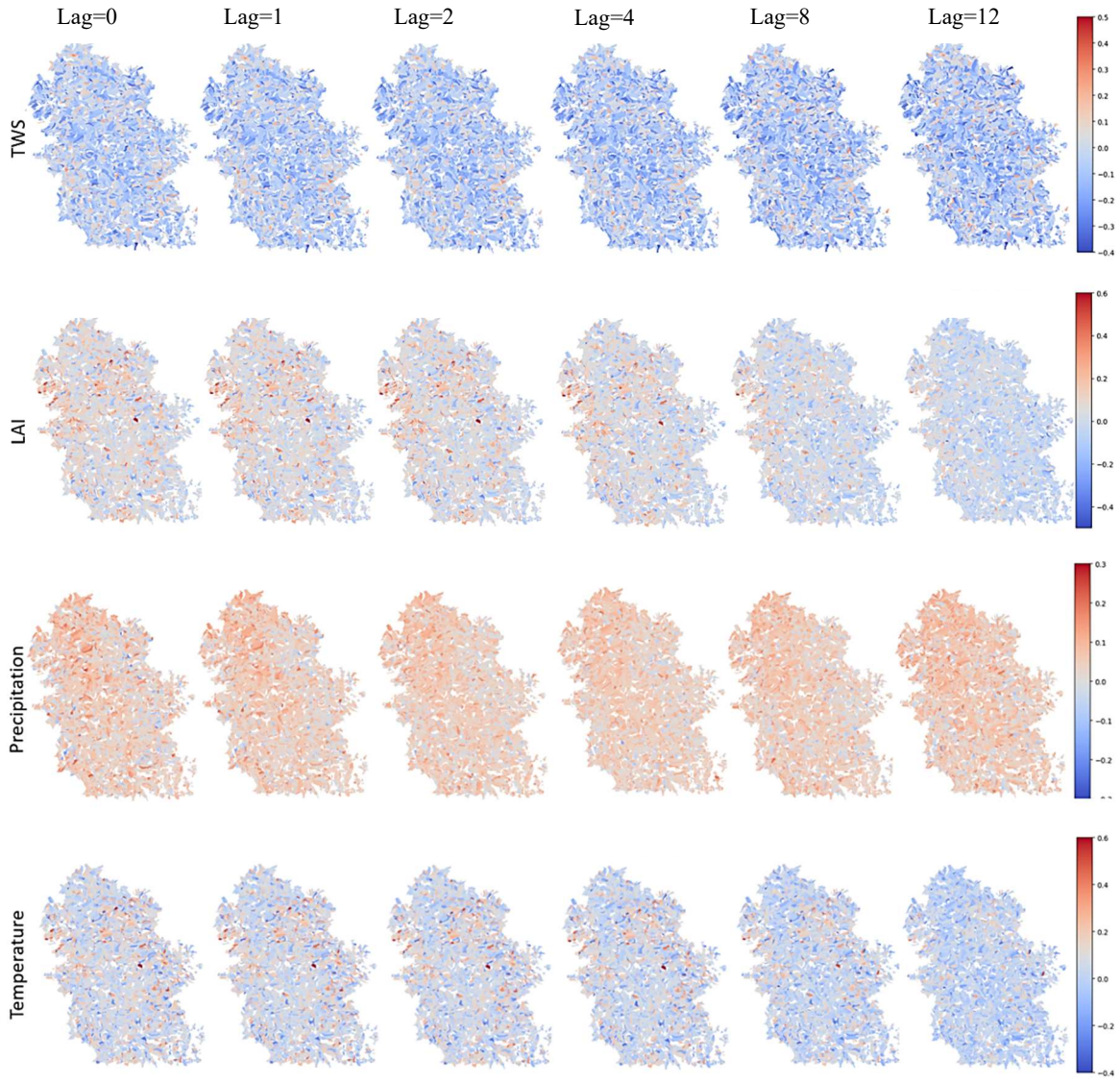


Figure 12 Correlation values between target variable (downslope displacement) and 4 other dynamic variables (precipitation, temperature, TWS and LAI), with a shift of lag= 0, 1, 2, 4, 8 and 12 (left to right).

To analyse the relationship between the correlation values and other variables, bivariate plots were created. However, no apparent patterns emerged (Figure 13). Similar plots for the various time lag correlations revealed no apparent trends (details in Appendix 9.3).

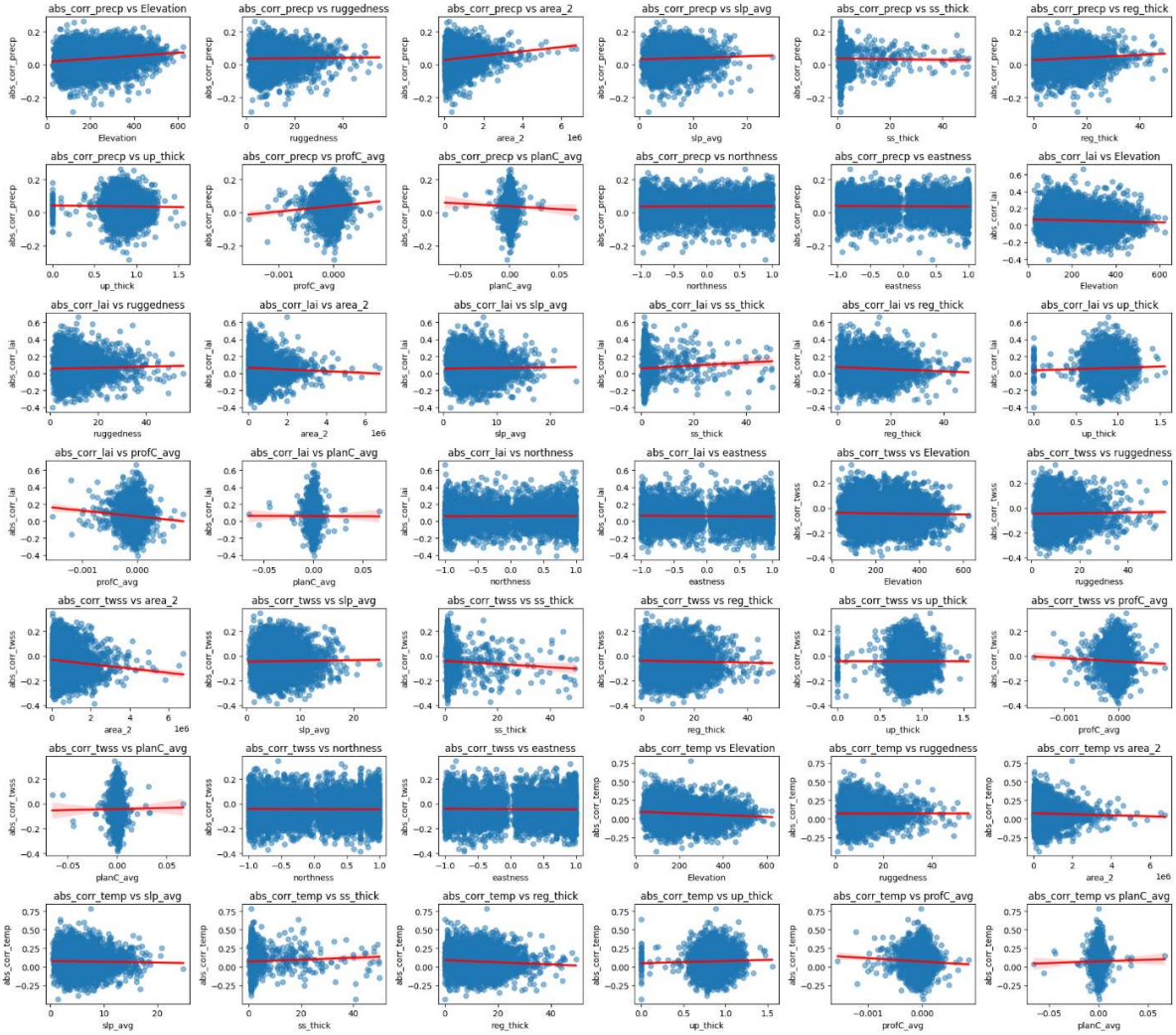


Figure 13 Bivariate plots between correlation values and other variables, without lag.

In order to account for the seasonality and cumulative effect of variables on the target variable, the data was aggregated monthly and seasonally (3 months), and the resulting correlation was calculated. For instance, the correlation between the maximum absolute displacement and monthly total rainfall was calculated to be 0.12. Even though on visual inspection, higher displacement appears to have occurred in months with higher rainfall in some parts of the dataset, the correlation value is extremely low (Figure 14).

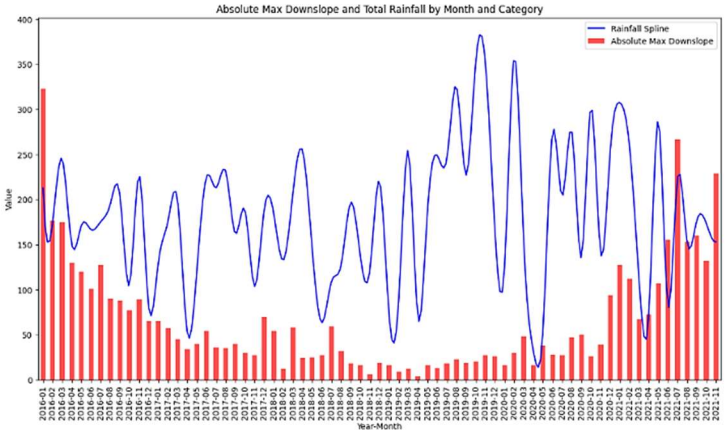


Figure 14 Downslope displacement and cumulative monthly rainfall

5.2. Prediction Model

After the hyperparameters were decided based on the validation set, the model was trained on the training set and evaluated on the test set to assess its predictive performance. The training process involved optimizing the model parameters to minimize the MSE on the training data. After training, the predictions were compared against the actual values to evaluate its generalization ability. This section presents a detailed analysis of the model's performance using various metrics and visualizations.

5.2.1. Temporal Dynamics

To understand the model's performance over time, I compared the actual and predicted values for specific nodes. This helps in understanding the temporal dynamics and accuracy of the model over time steps, for individual nodes (Figure 15). While the model captures the general trend and fluctuations of the actual values (Figure 15 (a) and (b)), there are nodes with sharp changes in signal where the predictions significantly deviate from the actual values (Figure 15 (c) and (d)).

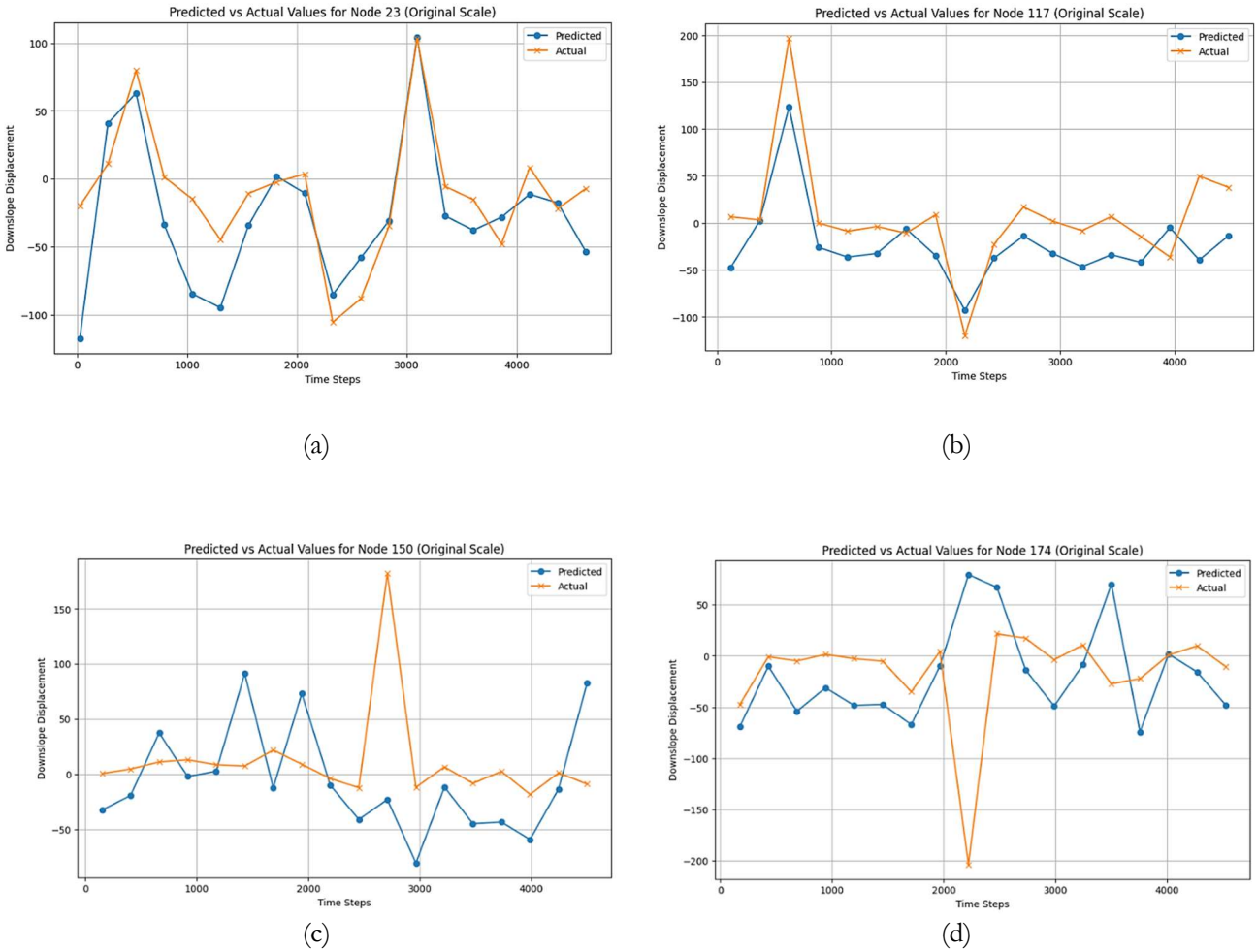


Figure 15 Examples presented for the time series of predicted and actual values for individual nodes, where model performed well (a and b) and where it could not capture the abruptly varying signal (c and d)

5.2.2. QQ Plot Analysis

A QQ (Quantile-Quantile) plot is used to compare the distribution of actual and predicted values. In Figure 16, the quantiles of the actual values were plotted against the quantiles of the predicted values. The closer the points are to the 45º reference line (y=x), the better the predicted values match the actual values. Deviations from this line indicate discrepancies between the distributions. It can be observed that most points in the central region of the data lie close to the y=x line, indicating that the predicted values follow the distribution of the actual values. The model's predictions for values near the mean are relatively accurate (Figure 16). However, at the extremes there are major deviations. Specifically, the model over and underestimates for low and high values, respectively. This pattern suggests the model tends to smooth out extreme values, predicting fewer extreme values than the actual observations.

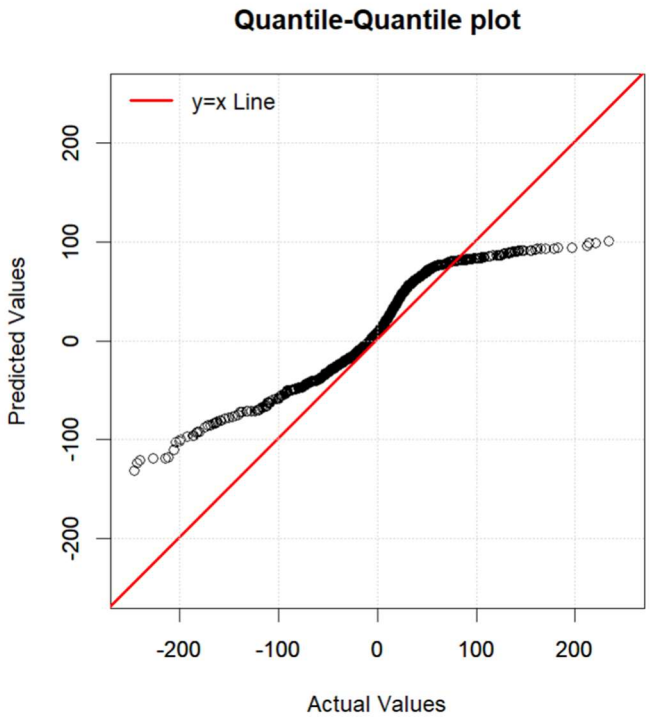
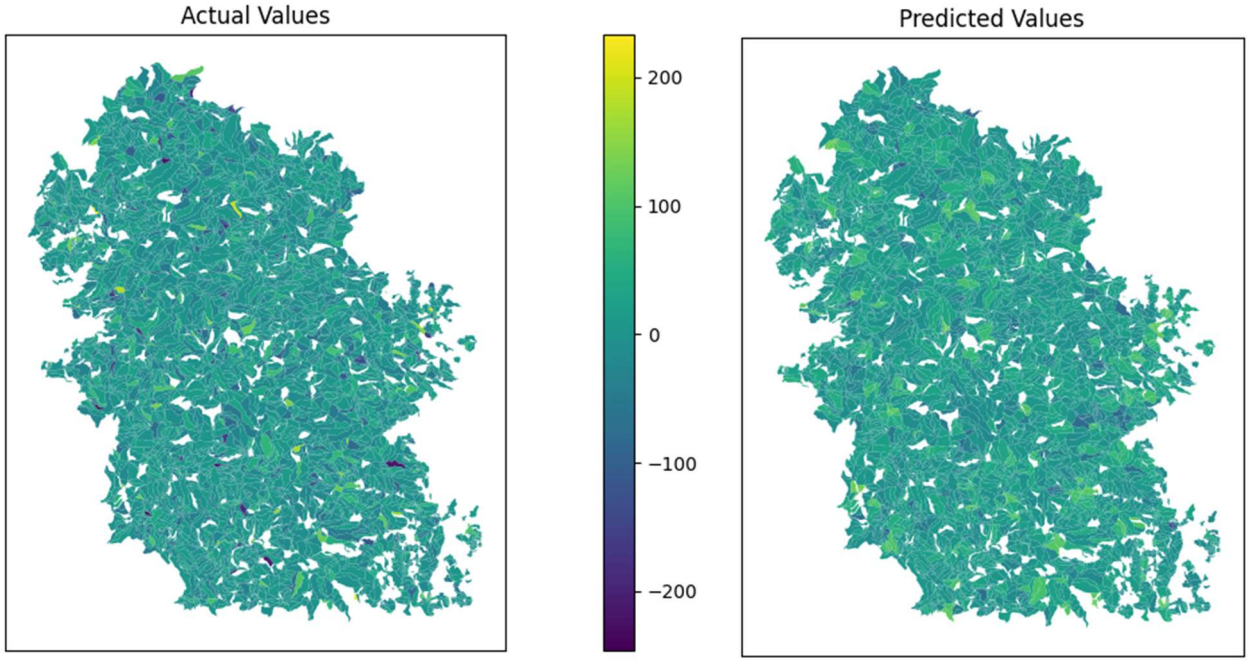


Figure 16 QQ Plot shows the quantiles of the actual values on the x-axis and the quantiles of the predicted values on the y-axis.

5.2.3. Spatial Distribution

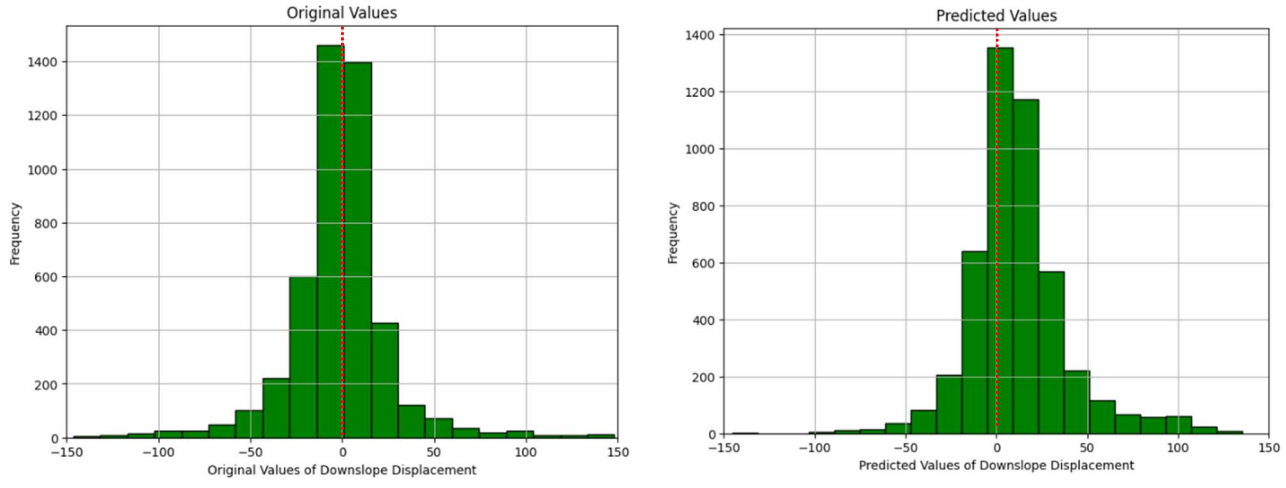
The visualization of the average actual and predicted values for the entire dataset provides an overview of the model's predictions (Figure 17). The model captures the general spatial patterns, but there are regions where the predicted values significantly deviate from the actual values.



(a) (b)

Figure 17 The spatial distribution of average (a) actual and (b) predicted values across the dataset.

The distribution of these values in the histogram shows the spread and central tendency of the actual and predicted values. Figure 18 shows a symmetric distribution centered around zero with most values falling within -50 to 50. The distribution of the predicted values is similar to the actual values, indicating that the model's predictions follow the overall distribution of the actual data. The histograms reveal that while the central tendency and spread are similar, there are deviations in the tails of the distribution. As also indicated by the QQ plot, the model overestimates low values (predicts them to be less negative than they are).

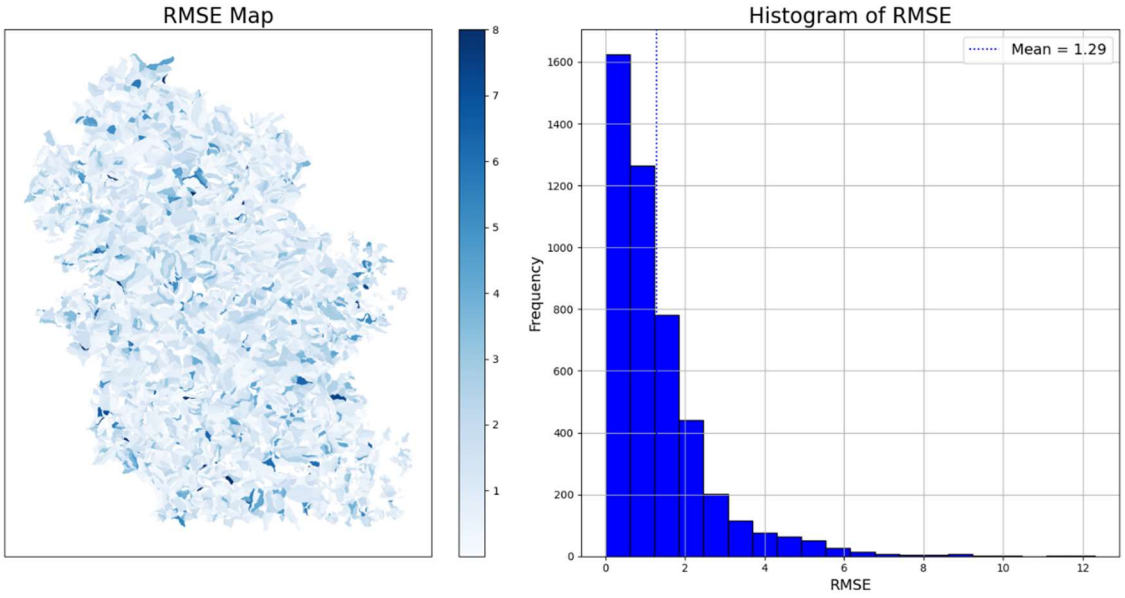


(a) (b)

Figure 18 Histogram showing the distribution of (a) actual values and (b) predicted values of target variable

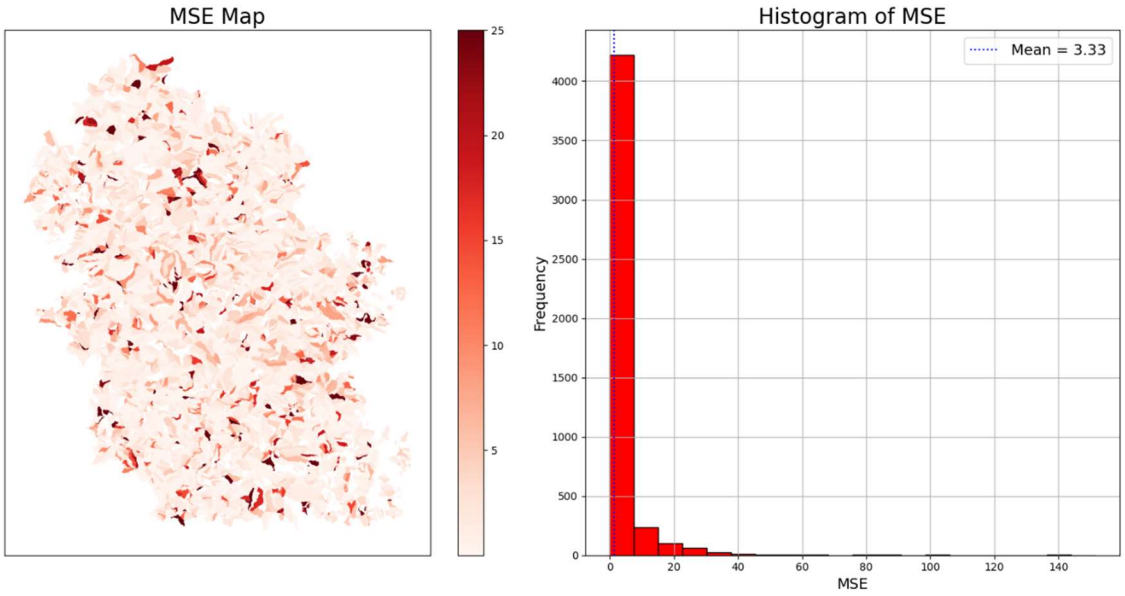
5.2.4. Prediction Error

Figures 19 and 20 show the Root Mean Square Error (RMSE) and MSE distribution in the study area, respectively. The map on the left showcases the spatial variability of prediction errors across the study area. The histogram on the right provides a frequency distribution of these errors. The RMSE ranges from 0 to 12 mm. It can be observed that the majority of the error values are concentrated at lower ranges, with a sharp decline as the error value increases. This indicates that the model generally performs well in most areas, with a few regions exhibiting higher prediction errors.



(a) (b)

Figure 19 Spatial and Frequency Distribution of RMSE (a) RMSE Map (b) Histogram of RMSE



(a) (b)

Figure 20 Spatial and Frequency Distribution of MSE (a) MSE Map (b) Histogram of MSE

6. DISCUSSION

This chapter discusses the implications of the results in the context of the study's objectives. Section 6.1 discusses the correlation analysis results, the challenges posed by the coarse resolution of dynamic variables from the ERA-5 dataset compared to the finer-resolution InSAR data, and its impact on capturing complex relationships between variables. Section 6.2 evaluates the performance of the spatio-temporal model, highlighting its strengths and limitations, especially in predicting extreme deformation values.

6.1. Spatial Resolution Discrepancy

Low values of correlation between the various dynamic variables and downslope displacement were obtained and there is an overall lack of trend of correlation values with increasing lag time. This could indicate that, since correlation measures linear relationships, it was unable to capture the complex relationships in the dataset. Mutual Information, a metric which captures non-linear relationships between variables, gave quite high values with precipitation (in Appendix 9.4). However, individually taken, it is not a conclusive metric.

Additionally, poor performance of the correlation analysis could also be attributed to the discrepancy in the data resolution. The dynamic variables from the ERA-5 dataset have a resolution of about 11 km. This means each pixel represents a large area, averaging the values of variables (e.g., precipitation, temperature, LAI) over this extensive region (Figure 21a). Such aggregation can smooth out local variations and extreme values that might be critical for understanding downslope displacement. Downslope displacement is often influenced by localized factors such as micro-topography, soil properties, vegetation cover, and small-scale weather events. These factors can vary significantly within an 11 km pixel, leading to a loss of critical information when using coarse-resolution data. For example, intense rainfall in a small part of the 11 km pixel might trigger a landslide, but the average precipitation value for the whole pixel might not reflect this intensity. The displacement data from EGMS is spatially rich, capturing fine-scale variations in slope movement. This high level of detail can include localized events and variations that are not visible in the coarser ERA-5 data.

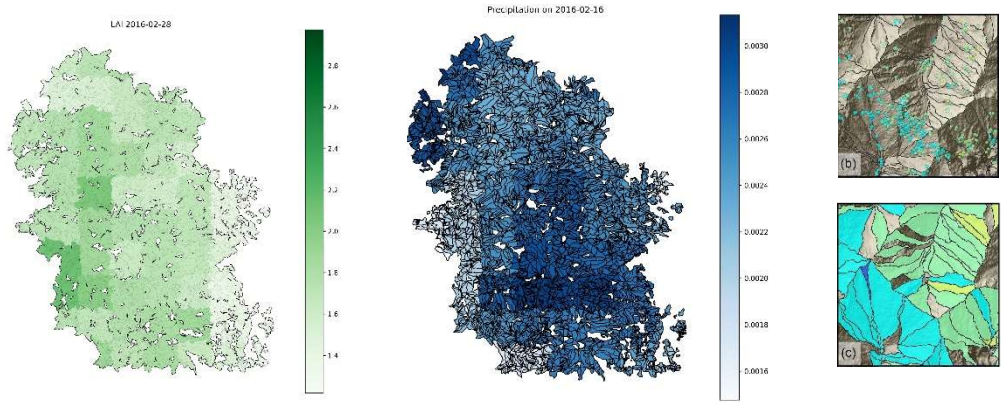


Figure 21 Discrepancy in the data resolution between the ERA5 (dynamic variables) and EGMS dataset (target variable)

(a) Low resolution of LAI and Precipitation (b) InSAR Scatterers (c) Slope Units

Efforts were made to smoothen the high spatial resolution EGMS displacement dataset by taking the signal of the scatterer with the highest mean deformation per slope unit (Figure 21b and c). Yet, the dynamic variables derived from the ERA-5 dataset are still coarser at a resolution of about 11 km. The significant difference in spatial resolution means that one pixel in the ERA-5 data covers a much larger area compared to multiple finer-resolution scatterer signals in the EGMS data. This spatial misalignment can lead to inaccuracies when correlating the two datasets because the precise location and extent of displacement events might not correspond directly to the averaged dynamic variables.

6.2. Outliers and Error Distribution

The analysis of the results indicates that the model performs reasonably well around the middle ranges in capturing both spatial and temporal dynamics of the dataset. The QQ plot reveals that the predicted values generally follow the distribution of the actual values, however, the model struggles with extreme values and certain regions with high variability, especially beyond the range of ±90 mm. This is reiterated by the distribution of the average prediction errors (Figure 22) which show higher values for slope units where the average downslope is beyond the range of ±100 mm.

The significant differences between MSE and RMSE across various bins of downslope displacement suggest the presence of outliers in the data. MSE, which involves squaring each error before averaging, is especially sensitive to these outliers, resulting in disproportionately larger values compared to RMSE (Chai & Draxler, 2014). This pattern might be also associated with data quality issues. The EGMS dataset was generated for the entire European region and in such a large scale processing of SAR data, some uncertainties that we cannot assess might exist in the repository.

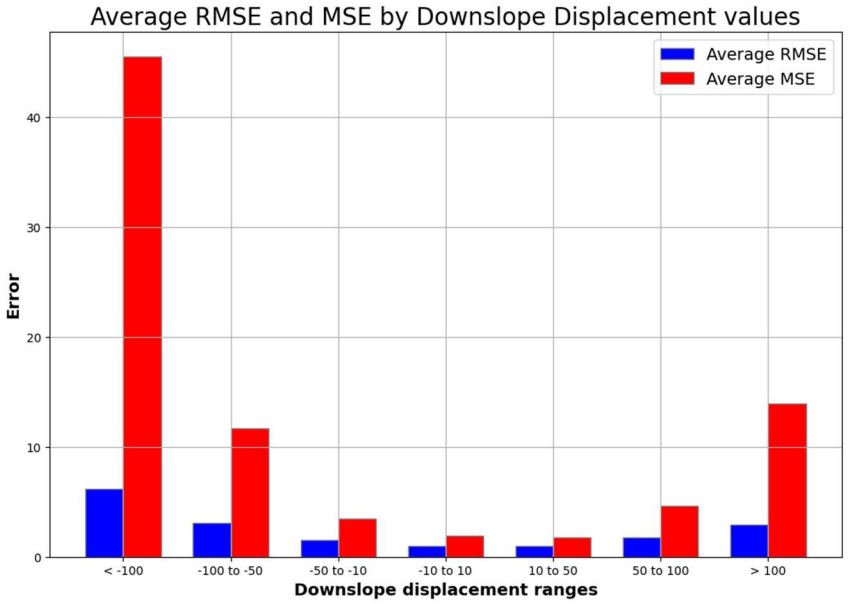


Figure 22 Distribution of RMSE and MSE across the downslope displacement ranges

6.3. Landslides and Error Distribution

The analysis revealed that the error is notably larger in slopes affected by flows and falls. This can be attributed to the fact that fast-moving processes are challenging for InSAR to capture accurately, as the signal is not well-suited for such rapid changes. Sudden changes in surface displacement lead to signal decorrelation and phase noise, making it difficult for InSAR to maintain coherence and provide reliable measurements (Moretto et al., 2021).

A comparison of the errors (Figure 20 a) with the landslide inventory provided by the BGS (Figure 23) indicates that most SUs with high error are associated with landslides. However, the reverse is not necessarily true - not all landslide locations correspond to high errors. This discrepancy can be attributed to the varying nature of landslides and the specific characteristics of different regions.

The bar chart (Figure 24) shows the distribution of average prediction error by landslide type. It indicates that the average error is highest for fast movements, including falls and flows, underlining the difficulty in modelling these rapid movements accurately.

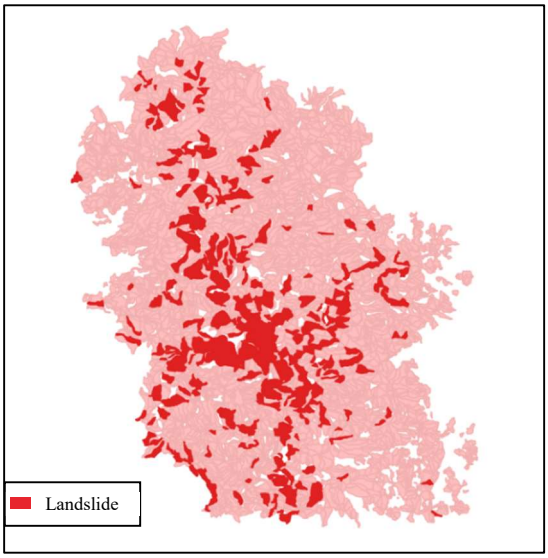


Figure 23 Distribution of Landslides in the region (2014 -2021) (Source: BGS)

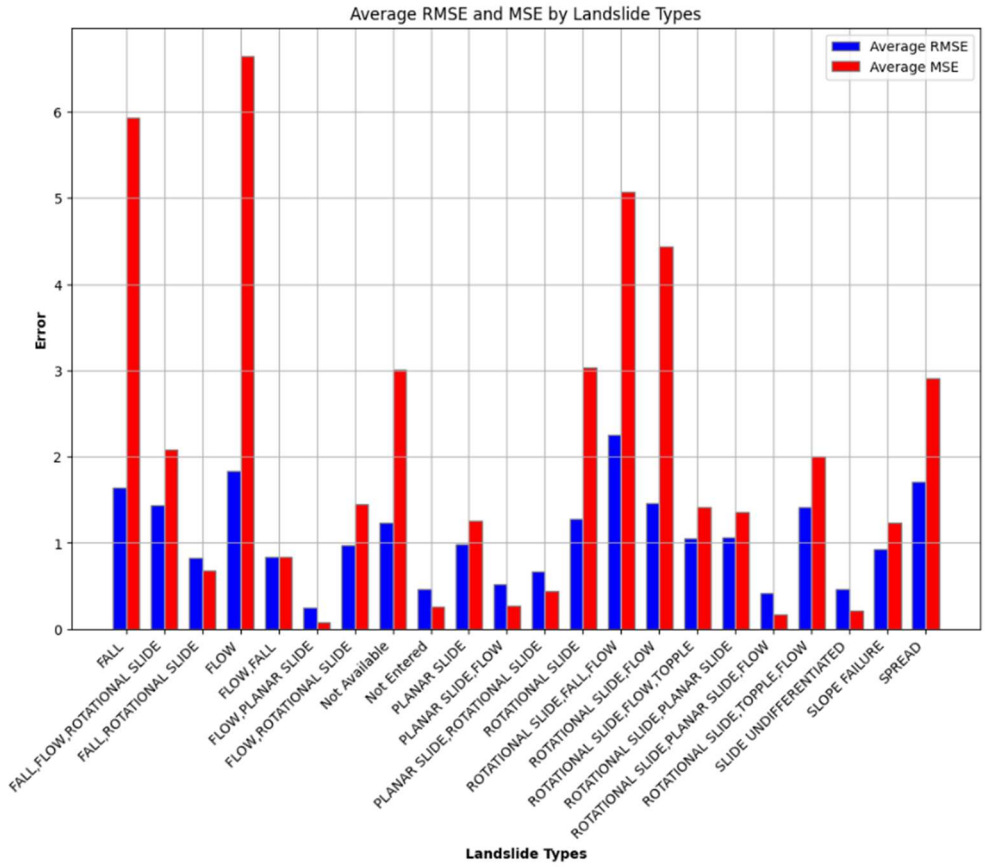


Figure 24 Distribution of average prediction error by landslide type, showing largest average error for flow and fall.

The effectiveness of InSAR technology diminishes when it comes to rapid movements such as those seen in debris flows and falls. The high velocity and sudden nature of these events result in phase unwrapping errors and loss of signal coherence, which in turn lead to inaccurate measurements and high prediction errors (Zhong et al., 2024). However, this technology is quite effective for detecting the onset of movement and monitoring slow, gradual displacements over time. For example, it can be used to identify areas where slope movements are beginning and to monitor landslides that exhibit slow, continuous deformation (Sun et al., 2016). This capability makes InSAR an invaluable tool for early warning systems and long-term monitoring of landslide-prone areas.

7. CONCLUSION

This chapter provides the concluding remarks. Section 7.1. summarizes the main findings in relation to the research objectives. Section 7.2. offers recommendations for future research directions in this field, while highlighting the importance and scope of further research.

7.1. Main Findings in relation to the Research Objectives

Objective 1: Relationship Between Environmental Variables and Deformation

- The study found varying correlations between dynamic variables and hillslope deformation, with precipitation showing positive correlations at certain lags, and temperature, leaf area index, and terrestrial water storage showing negative correlations. However, these correlations were generally low, indicating complex, non-linear interactions that may not be fully captured by linear correlation or given the low spatial resolution of the dynamic variables.

- The analysis of lag time did not reveal any apparent trends, suggesting that the impact of these environmental variables on hillslope deformation is not straightforward and may vary across different conditions and regions.

Objective 2: Development of the Spatio-Temporal Model

- The proposed GAT+GRU model successfully integrates both spatial and temporal dependencies. It utilizes graph structures to capture spatial relationships among slope units and recurrent units to model temporal sequences. The incorporation of predisposing and dynamic environmental factors in a spatio-temporal model is a novel approach.

- The model generally captures the trends and fluctuations in deformation data but struggles with extreme values. Spatially, the model's predictions align with actual deformations in many areas but deviate in regions with high variability. The performance analysis through QQ plots and spatial distribution maps indicates areas for potential improvement, particularly in handling extreme deformation values.

7.2. Future Recommendations

Slope stability and hillslope deformations are governed by a complex interaction of various factors (Dong et al., 2020; Scaringi & Loche, 2022). For example, Kang et al. (2021) identified seasonal changes in InSAR-derived deformation rates, attributing pore pressure and clay-rich body swelling as key factors. Other studies have linked monthly precipitation (Xu et al., 2019) and earthquakes (Fang et al., 2022) to positive correlations with InSAR-derived deformations. Given that the intensity and frequency of these influencing factors are rapidly changing due to climate change (Handwerger et al., 2019), understanding their contributions is crucial for gaining deeper insights into deformation patterns. Future research focusing on establishing feature importance and SHAP values of variables in the model could provide crucial insights into the influence of individual factors. These insights can significantly improve our understanding of slope evolution under varying conditions and enhance the predictive capabilities of the model (Lundberg & Lee, 2017).

To build on the current findings, employing different modelling approaches on the same dataset could also help in comparing results and identifying the most effective methods for prediction. Moreover, running the same model on different datasets could assess the model's robustness and generalizability. This comparative approach could highlight the model's strengths and areas needing improvement, ensuring it performs well across various conditions and datasets.

Future research could also consider incorporating ground-based measurements, which provide dynamic variables at a finer spatial resolution. This could lead to more significant correlation results, as seen in studies by Lollino et. al (2016) and Ardizzone et al. (2011). Utilizing high-resolution data, from sources such as the weather observations from the UK Meteorological Service (*Met Office WOW - Home Page*, n.d.) can capture localized variations and extreme values that are often critical for understanding downslope displacement.

In addition to correlation analysis, future studies could benefit from examining causal relationships between variables. Employing causal inference techniques can help identify and understand the direct effects of environmental factors on hillslope deformation, providing deeper insights into the mechanisms driving these processes (Runge et al., 2019) .

With more people living in hazardous areas and extreme weather events becoming more common due to climate change (Wallemacq et al., 2015), this direction of research could lay the foundation for a global hillslope displacement prediction model. Enhanced by AI-based accurate weather forecast models like the recent GraphCast (Lam et al., 2023), such a model would provide critical, timely warnings, potentially saving lives and reducing property damage. This global approach to hillslope deformation prediction is a vital step forward in disaster risk reduction and management.

8. REFERENCES

Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., & Guzzetti, F. (2016). Automatic delineation of geomorphological slope units with r.slopeunits v1.0 and their optimization for landslide susceptibility modeling. *Geoscientific Model Development*, *9*(11), 3975–3991. https://doi.org/10.5194/gmd-9-3975-2016

Ardizzone, F., Rossi, M., Calò, F., Paglia, L., Manunta, M., Mondini, A. C., Zeni, G., Reichenbach, P., Lanari, R., & Guzzetti, F. (2011). *Preliminary analysis of a correlation between ground deformations and rainfall: the Ivancich landslide, central Italy*. 81790L. https://doi.org/10.1117/12.899453

Bhaumik, K., Niloy, F., Mahmud, S., & Woo, S. (2022). *STLGRU: Spatio-Temporal Lightweight Graph GRU for Traffic Flow Prediction*.

Bianchi, F. M., Grahn, J., Eckerstorfer, M., Malnes, E., & Vickers, H. (2021). Snow Avalanche Segmentation in SAR Images With Fully Convolutional Neural Networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 75–82. https://doi.org/10.1109/JSTARS.2020.3036914

Bréda, N. J. J. (2008). Leaf Area Index. *Encyclopedia of Ecology, Five-Volume Set*, *1–5*, 2148–2154. https://doi.org/10.1016/B978-008045405-4.00849-1

Brownlee, J. (2020). *Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python (ebook)* (1st ed., Vol. 1). Machine Learning Mastery.

Cascini, L., Fornaro, G., & Peduto, D. (2010). Advanced low- and full-resolution DInSAR map generation for slow-moving landslide analysis at different scales. *Engineering Geology*, *112*(1–4), 29–42. https://doi.org/10.1016/J.ENGGEO.2010.01.003

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247–1250. https://doi.org/10.5194/GMD-7-1247-2014

Chen, L., Mei, L., Zeng, B., Yin, K., Shrestha, D. P., & Du, J. (2020). Failure probability assessment of landslides triggered by earthquakes and rainfall: a case study in Yadong County, Tibet, China. *Scientific Reports*, *10*(1), 16531. https://doi.org/10.1038/s41598-020-73727-4

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*. https://arxiv.org/abs/1412.3555v1

Cisneros, D., Richards, J., Dahal, A., Lombardo, L., & Huser, R. (2023). *Deep graphical regression for jointly moderate and extreme Australian wildfires*. https://arxiv.org/abs/2308.14547v1

Corominas, J., van Westen, C., Frattini, P., Cascini, L., Malet, J.-P., Fotopoulou, S., Catani, F., Van Den Eeckhaut, M., Mavrouli, O., Agliardi, F., Pitilakis, K., Winter, M. G., Pastor, M., Ferlisi, S., Tofani, V., Hervás, J., & Smith, J. T. (2014). Recommendations for the quantitative analysis of landslide risk. *Bulletin of Engineering Geology and the Environment*, *73*(2), 209–263. https://doi.org/10.1007/s10064-013-0538-8

Costantini, M., Minati, F., Trillo, F., Ferretti, A., Novali, F., Passera, E., Dehls, J., Larsen, Y., Marinkovic, P., Eineder, M., Brcic, R., Siegmund, R., Kotzerke, P., Probeck, M., Kenyeres, A., Proietti, S., Solari, L., & Andersen, H. S. (2021). EUROPEAN GROUND MOTION SERVICE (EGMS). *International Geoscience and Remote Sensing Symposium (IGARSS)*, *2021-July*, 3293–3296. https://doi.org/10.1109/IGARSS47720.2021.9553562

Crosetto, M., Solari, L., Balasis-Levinsen, J., Bateson, L., Casagli, N., Frei, M., Oyen, A., Moldestad, D. A., & Mróz, M. (2021). DEFORMATION MONITORING AT EUROPEAN SCALE: THE COPERNICUS GROUND MOTION SERVICE. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLIII-B3-2021*(B3-2021), 141–146. https://doi.org/10.5194/ISPRS-ARCHIVES-XLIII-B3-2021-141-2021

Dixon, N., & Brook, E. (2007). Impact of predicted climate change on landslide reactivation: case study of Mam Tor, UK. *Landslides*, *4*(2), 137–147. https://doi.org/10.1007/s10346-006-0071-y

Dong, J., Niu, R., Li, B., Xu, H., & Wang, S. (2023). Potential landslides identification based on temporal and spatial filtering of SBAS-InSAR results. *Geomatics, Natural Hazards and Risk*, *14*(1), 52–75. https://doi.org/10.1080/19475705.2022.2154574

Dong, M., Wu, H., Hu, H., Azzam, R., Zhang, L., Zheng, Z., & Gong, X. (2020). Deformation Prediction of Unstable Slopes Based on Real-Time Monitoring and DeepAR Model. *Sensors*, *21*, 14. https://doi.org/10.3390/s21010014

Donnelly, L. J., Northmore, K. J., & Siddle, H. J. (2002). Block movements in the Pennines and South Wales and their association with landslides. *Quarterly Journal of Engineering Geology and Hydrogeology*, *35*(1), 33–39. https://doi.org/10.1144/qjegh.35.1.33

Faizi, N., & Alvi, Y. (2023). Correlation. *Biostatistics Manual for Health Research*, 109–126. https://doi.org/10.1016/B978-0-443-18550-2.00002-5

Fang, H., Shao, Y., Xie, C., Tian, B., Zhu, Y., Guo, Y., Yang, Q., & Yang, Y. (2022). Using Persistent Scatterer Interferometry for Post-Earthquake Landslide Susceptibility Mapping in Jiuzhaigou. *Applied Sciences*, *12*(18). https://doi.org/10.3390/app12189228

Faraway, J. J., Tanner, M., Zidek, J., & Smith, P. J. (2011). Practical Multivariate Analysis. *Practical Multivariate Analysis*. https://doi.org/10.1201/9781466503243

Fey, M., & Lenssen, J. E. (2019). *Fast Graph Representation Learning with PyTorch Geometric*. https://arxiv.org/abs/1903.02428v3

Ganesh, B., Vincent, S., Pathan, S., & Garcia Benitez, S. R. (2023). Machine learning based landslide susceptibility mapping models and GB-SAR based landslide deformation monitoring systems: Growth and evolution. *Remote Sensing Applications: Society and Environment*, *29*, 100905. https://doi.org/https://doi.org/10.1016/j.rsase.2022.100905

Girotto, M., & Rodell, M. (2019). Terrestrial water storage. *Extreme Hydroclimatic Events and Multivariate Hazards in a Changing Environment: A Remote Sensing Approach*, 41–64. https://doi.org/10.1016/B978-0-12-814899-0.00002-X

Gómez, D., García, E. F., & Aristizábal, E. (2023). Spatial and temporal landslide distributions using global and open landslide databases. *Natural Hazards*, *117*(1), 25–55. https://doi.org/10.1007/s11069-023-05848-8

Handwerger, A. L., Huang, M.-H., Fielding, E. J., Booth, A. M., & Bürgmann, R. (2019). A shift from drought to extreme rainfall drives a stable landslide to catastrophic failure. *Scientific Reports*, *9*(1), 1569. https://doi.org/10.1038/s41598-018-38300-0

He, K., Tanyas, H., Chang, L., Hu, X., Luo, G., & Lombardo, L. (2023). Modelling InSAR-derived hillslope velocities with multivariate statistics: A first attempt to generate interpretable predictions. *Remote Sensing of Environment*, *289*, 113518. https://doi.org/https://doi.org/10.1016/j.rse.2023.113518

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hua, Y., Wang, X., Li, Y., Xu, P., & Xia, W. (2021). Dynamic development of landslide susceptibility based on slope unit and deep neural networks. *Landslides*, *18*(1), 281–302. https://doi.org/10.1007/s10346-020-01444-0

Huggett, R. (2020). The Peak District. In A. Goudie & P. Migoń (Eds.), *Landscapes and Landforms of England and Wales* (pp. 393–416). Springer International Publishing. https://doi.org/10.1007/978-3-030-38957-4_23

Iverson, R. M. (2000). Landslide triggering by rain infiltration. *Water Resources Research*, *36*(7), 1897–1910. https://doi.org/10.1029/2000WR900090

Ji, S., Yu, D., Shen, C., Li, W., & Xu, Q. (2020). Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks. *Landslides*, *17*(6), 1337–1352. https://doi.org/10.1007/s10346-020-01353-2

Jiang, Y., Luo, H., Xu, Q., Lu, Z., Liao, L., Li, H., & Hao, L. (2022). A Graph Convolutional Incorporating GRU Network for Landslide Displacement Forecasting Based on Spatiotemporal Analysis of GNSS Observations. *Remote Sensing 2022, Vol. 14, Page 1016*, *14*(4), 1016. https://doi.org/10.3390/RS14041016

Kang, Y., Lu, Z., Zhao, C., Xu, Y., Kim, J., & Gallegos, A. J. (2021). InSAR monitoring of creeping landslides in mountainous regions: A case study in Eldorado National Forest, California. *Remote Sensing of Environment*, *258*, 112400. https://doi.org/https://doi.org/10.1016/j.rse.2021.112400

Khalili, M. A., Guerriero, L., Pouralizadeh, M., Calcaterra, D., & Di Martire, D. (2023). Monitoring and prediction of landslide-related deformation based on the GCN-LSTM algorithm and SAR imagery. *Natural Hazards*, *119*(1), 39–68. https://doi.org/10.1007/S11069-023-06121-8/TABLES/3

Kingma, D. P., & Ba, J. L. (2014). Adam: A Method for Stochastic Optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. https://arxiv.org/abs/1412.6980v9

Lacroix, P., Handwerger, A. L., & Bièvre, G. (2020). Life and death of slow-moving landslides. *Nature Reviews Earth & Environment 2020 1:8*, *1*(8), 404–419. https://doi.org/10.1038/s43017-020-0072-8

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., & Battaglia, P. (2023). Learning skillful medium-range global weather forecasting. *Science*. https://doi.org/10.1126/science.adi2336

Lollino, G., Arattano, M., Allasia, P., & Giordan, D. (2006). Time response of a landslide to meteorological events. *Natural Hazards and Earth System Sciences*, *6*(2), 179–184. https://doi.org/10.5194/NHESS-6-179-2006

Lombardo, L., Opitz, T., Ardizzone, F., Guzzetti, F., & Huser, R. (2020). Space-time landslide predictive modelling. *Earth-Science Reviews*, *209*, 103318. https://doi.org/https://doi.org/10.1016/j.earscirev.2020.103318

Lundberg, S. M., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, *2017-December*, 4766–4775. https://arxiv.org/abs/1705.07874v2

Ma, J., Wang, Y., Niu, X., Jiang, S., & Liu, Z. (2022). A comparative study of mutual information-based input variable selection strategies for the displacement prediction of seepage-driven landslides using optimized support vector regression. *Stochastic Environmental Research and Risk Assessment*, *36*(10), 3109–3129. https://doi.org/10.1007/s00477-022-02183-5

Ma, Z., & Mei, G. (2021). Deep learning for geological hazards analysis: Data, models, applications, and opportunities. *Earth-Science Reviews*, *223*, 103858. https://doi.org/10.1016/J.EARSCIREV.2021.103858

Ma, Z., Mei, G., Prezioso, E., Zhang, Z., & Xu, N. (2021). A deep learning approach using graph convolutional networks for slope deformation prediction based on time-series displacement data. *Neural Computing and Applications*, *33*(21), 14441–14457. https://doi.org/10.1007/S00521-021-06084-6/FIGURES/11

Media Centre Facts and Figures: Peak District National Park. (n.d.). Retrieved November 20, 2023, from https://www.peakdistrict.gov.uk/learning-about/news/mediacentrefacts

Meng, Q., Wang, H., He, M., Gu, J., Qi, J., & Yang, L. (2023). Displacement prediction of water-induced landslides using a recurrent deep learning model. *European Journal of Environmental and Civil Engineering*, *27*(7), 2460–2474. https://doi.org/10.1080/19648189.2020.1763847

Met Office WOW - Home Page. (n.d.). Retrieved June 17, 2024, from https://wow.metoffice.gov.uk/

Mirmazloumi, S. M., Wassie, Y., Nava, L., Cuevas-González, M., Crosetto, M., & Monserrat, O. (2023). InSAR time series and LSTM model to support early warning detection tools of ground instabilities: mining site case studies. *Bulletin of Engineering Geology and the Environment*, *82*(10), 374. https://doi.org/10.1007/s10064-023-03388-w

Moreno, M., Steger, S., Tanyas, H., & Lombardo, L. (2023). Modeling the area of co-seismic landslides via data-driven models: The Kaikōura example. *Engineering Geology*, *320*, 107121. https://doi.org/10.1016/J.ENGGEO.2023.107121

Moretto, S., Bozzano, F., & Mazzanti, P. (2021). The role of satellite insar for landslide forecasting: Limitations and openings. *Remote Sensing*, *13*(18), 3735. https://doi.org/10.3390/RS13183735/S1

Nappo, N., Peduto, D., Mavrouli, O., van Westen, C. J., & Gullà, G. (2019). Slow-moving landslides interacting with the road network: Analysis of damage using ancillary data, in situ surveys and multi-source monitoring data. *Engineering Geology*, *260*, 1–14. https://doi.org/10.1016/J.ENGGEO.2019.105244

Nava, L., Carraro, E., Reyes-Carmona, C., Puliero, S., Bhuyan, K., Rosi, A., Monserrat, O., Floris, M., Meena, S. R., Galve, J. P., & Catani, F. (2023). Landslide displacement forecasting using deep learning and monitoring data across selected sites. *Landslides*, *20*(10), 2111–2129. https://doi.org/10.1007/s10346-023-02104-9

Olaya, V. (2009). Basic land-surface parameters. *Developments in Soil Science*, *33*(C), 141–169. https://doi.org/10.1016/S0166-2481(08)00006-8

Pandey, A., Sarkar, M. S., Palni, S., Parashar, D., Singh, G., Kaushik, S., Chandra, N., Costache, R., Singh, P., Mishra, A. P., Almohamad, H., Al-Mutiry, M., Hazem, &, & Abdo, G. (2023). Multivariate statistical algorithms for landslide susceptibility assessment in Kailash Sacred landscape. *Natural Hazards and Risk*, *14*(1). https://doi.org/10.1080/19475705.2023.2227324

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

Radman, A., Akhoondzadeh, M., & Hosseiny, B. (2021). Integrating InSAR and deep-learning for modeling and predicting subsidence over the adjacent area of Lake Urmia, Iran. *GIScience & Remote Sensing*, *58*(8), 1413–1433. https://doi.org/10.1080/15481603.2021.1991689

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., & Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, *180*, 60–91. https://doi.org/10.1016/J.EARSCIREV.2018.03.001

Rey, S. J., & Anselin, L. (2007). PySAL: A Python library of spatial analytical methods. *Review of Regional Studies*, *37*(1), 5–27. https://doi.org/10.52324/001C.8285

Rodgers, J. L., & Nicewander, W. A. (1988). Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician*, *42*(1), 59. https://doi.org/10.2307/2685263

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Muñoz-Marí, J., van Nes, E. H., Peters, J., Quax, R., Reichstein, M., Scheffer, M., Schölkopf, B., Spirtes, P., Sugihara, G., Sun, J., … Zscheischler, J. (2019). Inferring causation from time series in Earth system sciences. *Nature Communications 2019 10:1*, *10*(1), 1–13. https://doi.org/10.1038/s41467-019-10105-3

Scaringi, G., & Loche, M. (2022). A thermo-hydro-mechanical approach to soil slope stability under climate change. *Geomorphology*, *401*, 108108. https://doi.org/https://doi.org/10.1016/j.geomorph.2022.108108

Sharma, A., Liu, X., Yang, X., & Shi, D. (2017). A patch-based convolutional neural network for remote sensing image classification. *Neural Networks*, *95*, 19–28. https://doi.org/10.1016/J.NEUNET.2017.07.017

Sun, Q., Hu, J., Zhang, L., Ding, X., Lu, Z., Li, X., & Thenkabail, P. S. (2016). Towards Slow-Moving Landslide Monitoring by Integrating Multi-Sensor InSAR Time Series Datasets: The Zhouqu Case Study, China. *Remote Sensing 2016, Vol. 8, Page 908*, *8*(11), 908. https://doi.org/10.3390/RS8110908

Tanyaş, H., Kirschbaum, D., & Lombardo, L. (2021). Capturing the footprints of ground motion in the spatial distribution of rainfall-induced landslides. *Bulletin of Engineering Geology and the Environment*, *80*(6), 4323–4345. https://doi.org/10.1007/S10064-021-02238-X/FIGURES/13

Tanyaş, H., van Westen, C. J., Allstadt, K. E., & Jibson, R. W. (2019). Factors controlling landslide frequency-area distributions. *Earth Surface Processes and Landforms*, *44*(4), 900–917. https://doi.org/10.1002/ESP.4543

Tonks, A., Harris, T., Li, B., Brown, W., & Smith, R. (2022). Forecasting West Nile Virus with Graph Neural Networks: Harnessing Spatial Dependence in Irregularly Sampled Geospatial Data. *ArXiv E-Prints*, arXiv:2212.11367. https://doi.org/10.48550/arXiv.2212.11367

Van den Bout, B., Lombardo, L., Chiyang, M., van Westen, C., & Jetten, V. (2021). Physically-based catchment-scale prediction of slope failure volume and geometry. *Engineering Geology*, *284*, 105942. https://doi.org/10.1016/J.ENGGEO.2020.105942

Van Den Eeckhaut, M., Hervás, J., Jaedicke, C., Malet, J. P., Montanarella, L., & Nadim, F. (2012). Statistical modelling of Europe-wide landslide susceptibility using limited landslide inventory data. *Landslides*, *9*(3), 357–369. https://doi.org/10.1007/S10346-011-0299-Z

Veličković, P., Casanova, A., Liò, P., Cucurull, G., Romero, A., & Bengio, Y. (2017). Graph Attention Networks. *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*. https://doi.org/10.1007/978-3-031-01587-8_7

Wallemacq, P., Guha-Sapir, D., McClean, D., CRED, & UNISDR. (2015). *The Human Cost of Natural Disasters - A global perspective*.

Wang, H.-C., Lin, K.-C., & Huang, C. (2016). Temporal and spatial patterns of remotely sensed litterfall in tropical and subtropical forests of Taiwan. *Journal of Geophysical Research: Biogeosciences*, *121*, n/a-n/a. https://doi.org/10.1002/2015JG003113

Xie, P., Zhou, A., & Chai, B. (2019). The Application of Long Short-Term Memory(LSTM) Method on Displacement Prediction of Multifactor-Induced Landslides. *IEEE Access*, *7*, 54305–54311. https://doi.org/10.1109/ACCESS.2019.2912419

Xu, Y., Kim, J., George, D. L., & Lu, Z. (2019). Characterizing Seasonally Rainfall-Driven Movement of a Translational Landslide using SAR Imagery and SMAP Soil Moisture. *Remote Sensing 2019, Vol. 11, Page 2347*, *11*(20), 2347. https://doi.org/10.3390/RS11202347

Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, *415*, 295–316. https://doi.org/10.1016/J.NEUCOM.2020.07.061

Zhang, W., Li, H., Tang, L., Gu, X., Wang, L., & Wang, L. (2022). Displacement prediction of Jiuxianping landslide using gated recurrent unit (GRU) networks. *Acta Geotechnica*, *17*(4), 1367–1382. https://doi.org/10.1007/s11440-022-01495-8

Zhang, Y., Chen, Y., Ming, D., Zhu, Y., Ling, X., Zhang, X., & Lian, X. (2022). Landslide hazard analysis based on SBAS-InSAR and MCE-CNN model: a case study of Kongtong, Pingliang. *Geocarto International*, 1–22. https://doi.org/10.1080/10106049.2022.2136268

Zhong, J., Li, Q., Zhang, J., Luo, P., & Zhu, W. (2024). Risk Assessment of Geological Landslide Hazards Using D-InSAR and Remote Sensing. *Remote Sensing 2024, Vol. 16, Page 345*, *16*(2), 345. https://doi.org/10.3390/RS16020345

9. APPENDIX

9.1 Batched Adjacency Matrix

A function 'get_connected_nodes' is used to form a fully connected subgraph by identifying connected nodes. It ensures that every batch is fully connected, which is crucial for capturing the graph structure during training. The batch size (determining the number of nodes in each batch) is determined during hyperparameter-tuning and can be given as an input.

```python
def get_connected_nodes(self, start_node, adj_matrix):
    # Initializing a set to keep track of connected nodes, starting with the start_node
    connected_nodes = set([start_node])

    # Initializing a list to explore nodes starting with the start_node
    to_explore = [start_node]

    # Exploring until there are no more nodes to explore
    while to_explore:
        # Pop a node from the list of nodes to explore
        node = to_explore.pop()

        # Finding the neighbors of the current node by checking the adjacency matrix
        neighbors = adj_matrix[node].nonzero().squeeze(1).tolist()

        # For each neighbor of the current node
        for neighbor in neighbors:
            # If the neighbor is not already in the set of connected nodes
            if neighbor not in connected_nodes:
                # Add the neighbor to the set of connected nodes
                connected_nodes.add(neighbor)
                # Add the neighbor to the list of nodes to explore
                to_explore.append(neighbor)

    # Return the list of connected nodes
    return list(connected_nodes)
```

Figure 25 Function to define batched adjacency matrix

9.2 Hyperparameter finetuning

Training Loss for different hyperparameter combinations:

```
# Hyperparameters to search over
learning_rates = [0.0001, 0.001, 0.01, 0.05]
gru_units = [64, 128, 256]
num_batch_nodes = [64, 128, 256, 512]  # Equivalent to batch size in our case

losses = {}  # Dictionary to store loss for each combination of hyperparameters

# Training loop
for lr in learning_rates:
    print(f"Learning rate [{lr}]")
    for gru_unit in gru_units:
        print(f"GRU Units [{gru_unit}]")
        for num_nodes in num_batch_nodes:
            print(f"Nodes: [{num_nodes}]")

            # Instantiating the model
            model = SpatioTemporalModel(time_steps=30, features=22, nodes=num_nodes, gru_units=gru_unit)
```

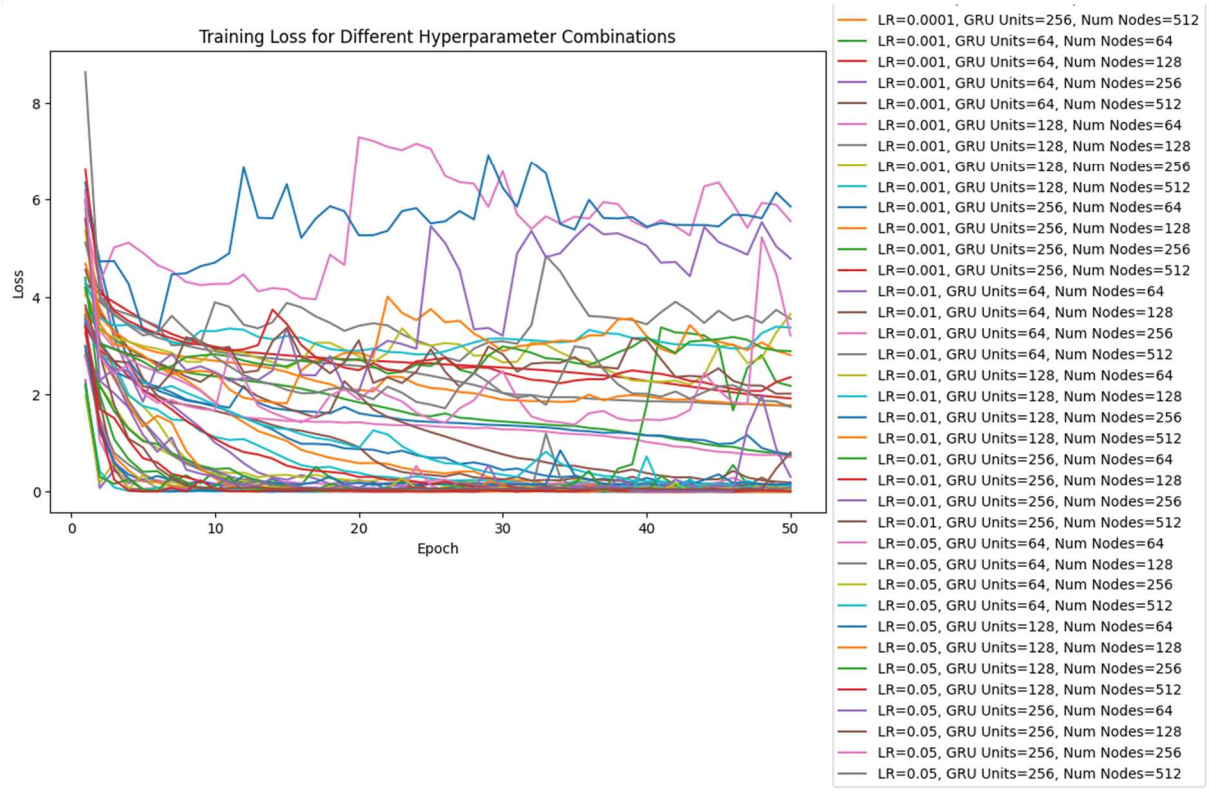


Figure 26 Results from hyperparameter finetuning for different combinations

9.3 Additional Results: Correlation Analysis

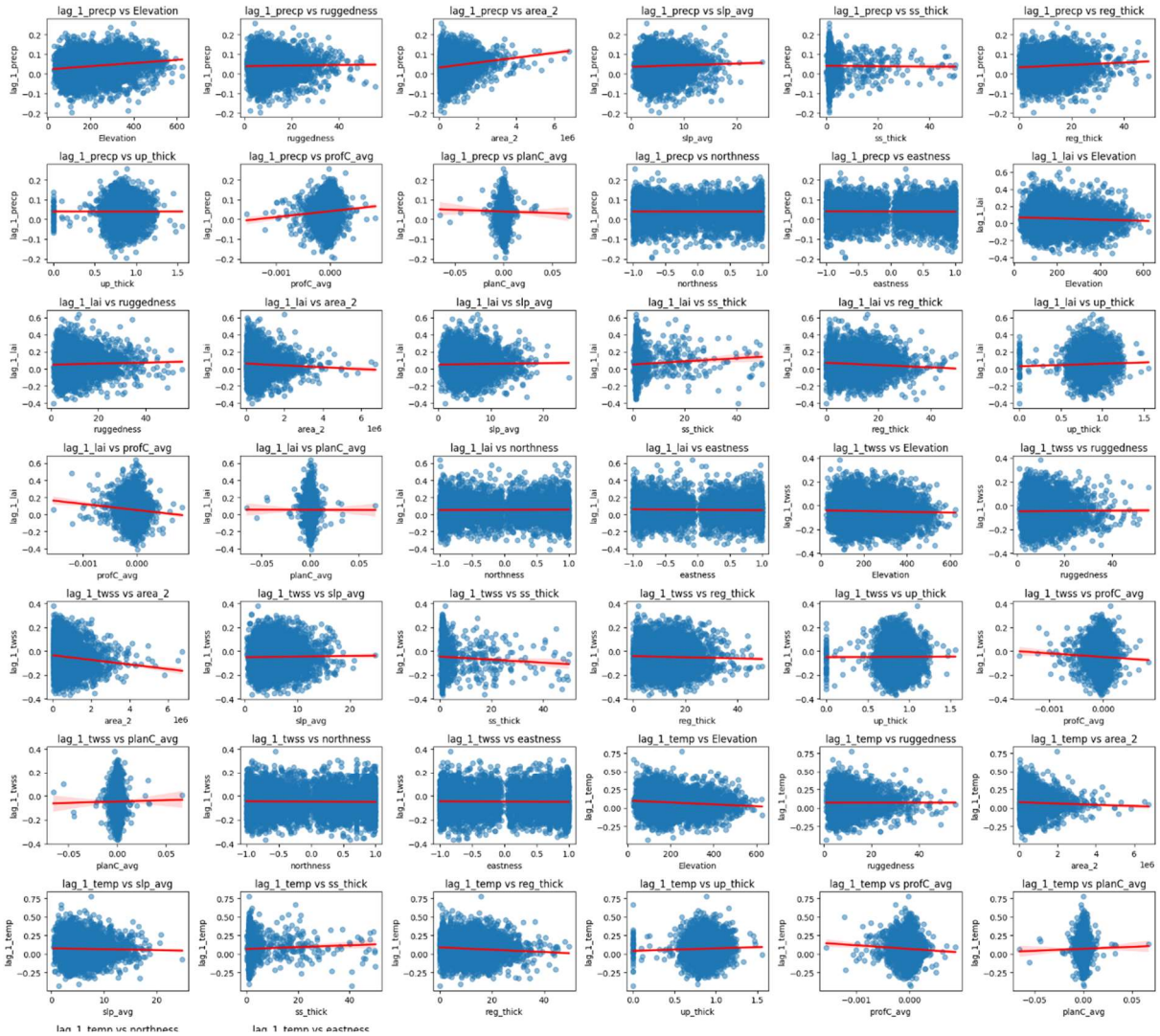


Figure 27 Bivariate plots between correlation values and other variables, with lag=1.

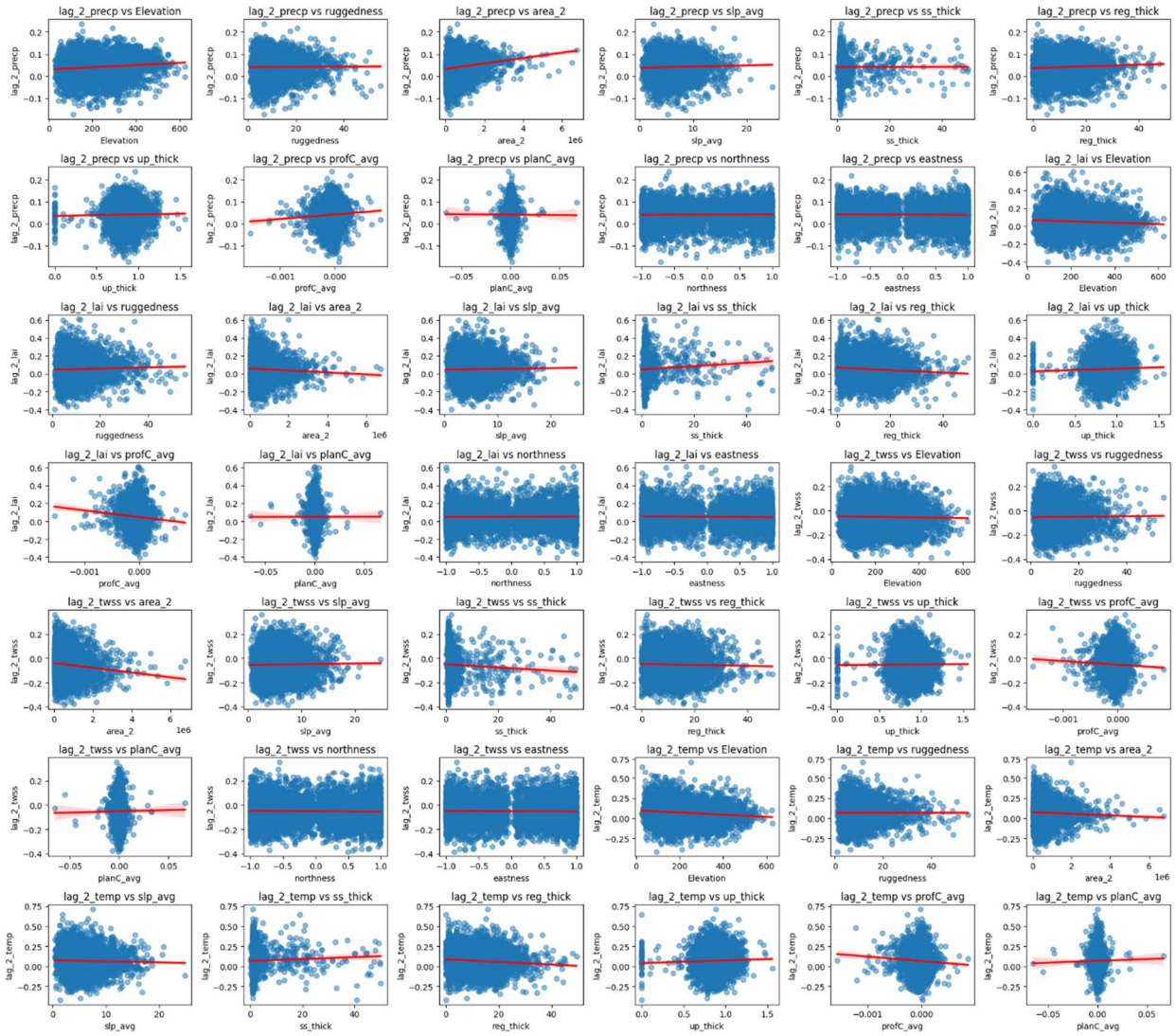


Figure 28 Bivariate plots between correlation values and other variables, with lag=2.

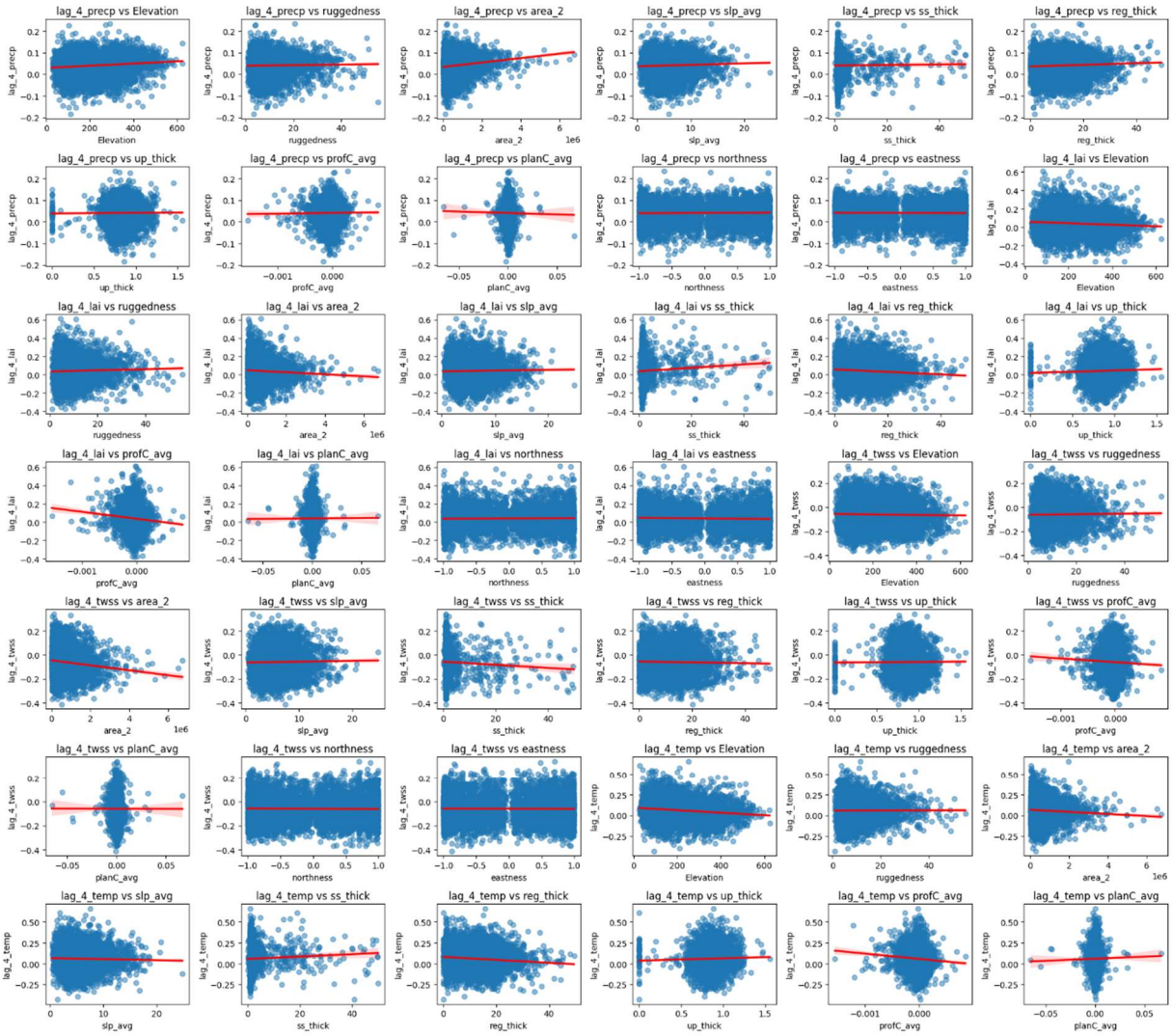


Figure 29 Bivariate plots between correlation values and other variables, with lag=4.

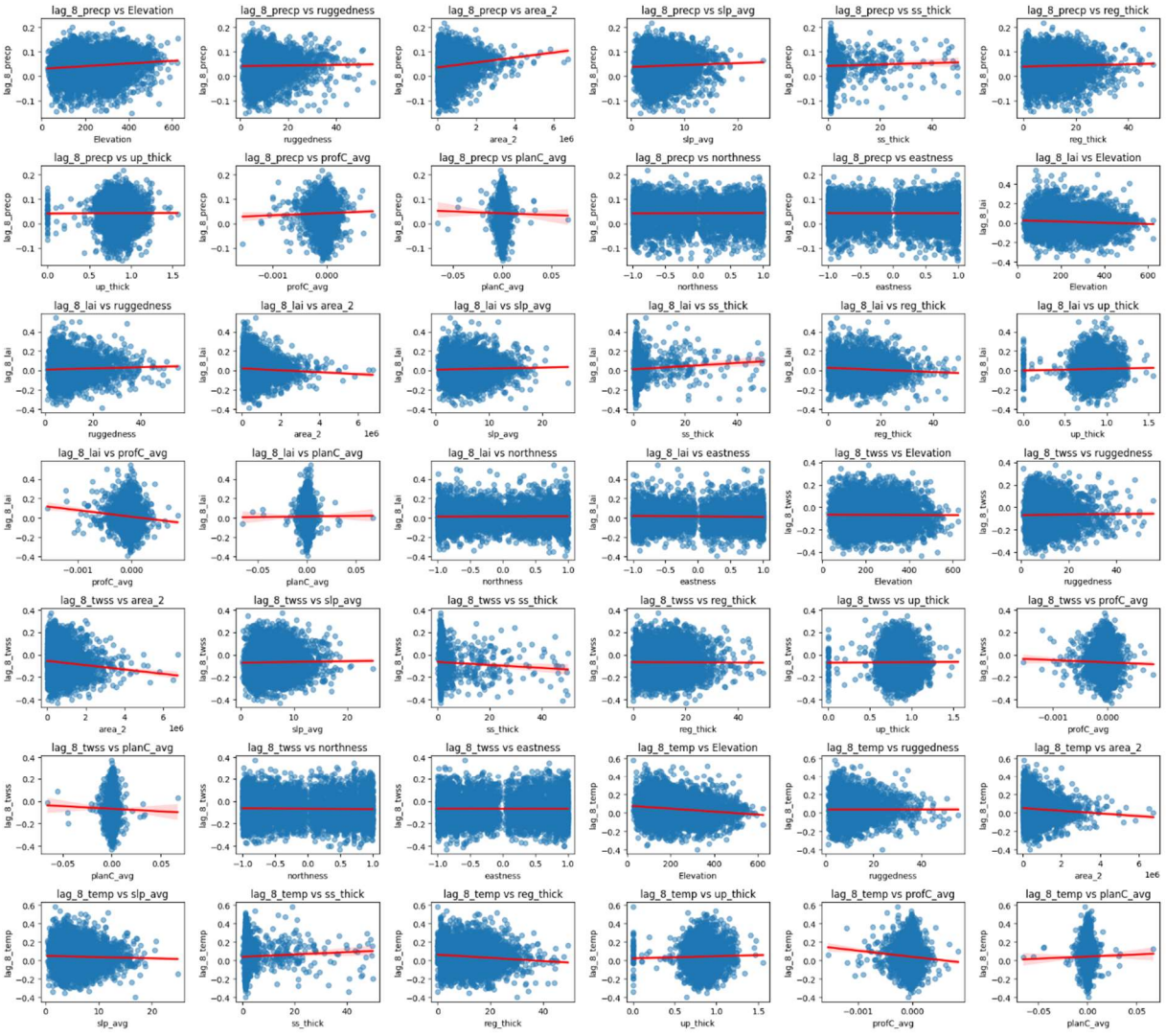


Figure 30 Bivariate plots between correlation values and other variables, with lag=8

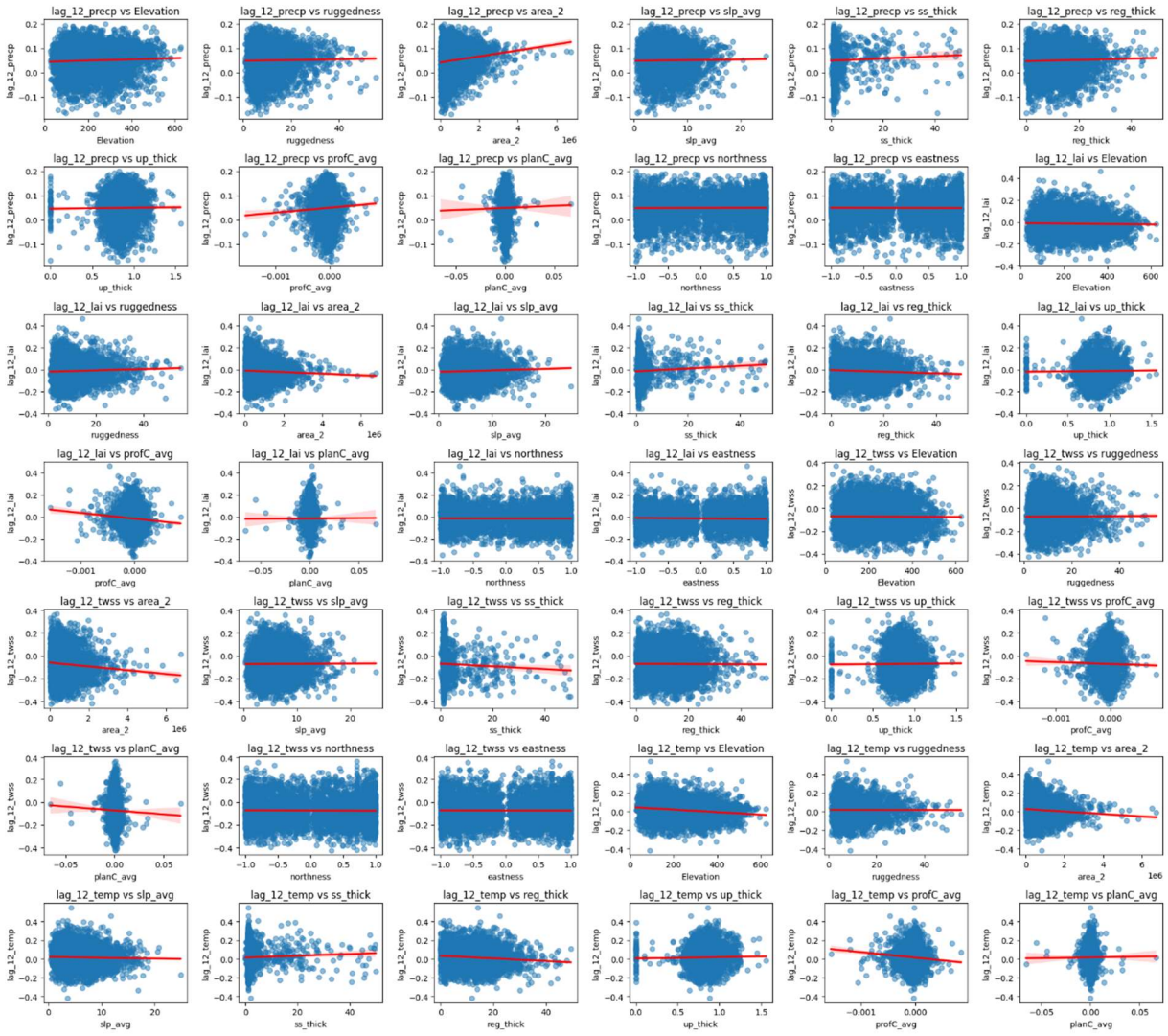
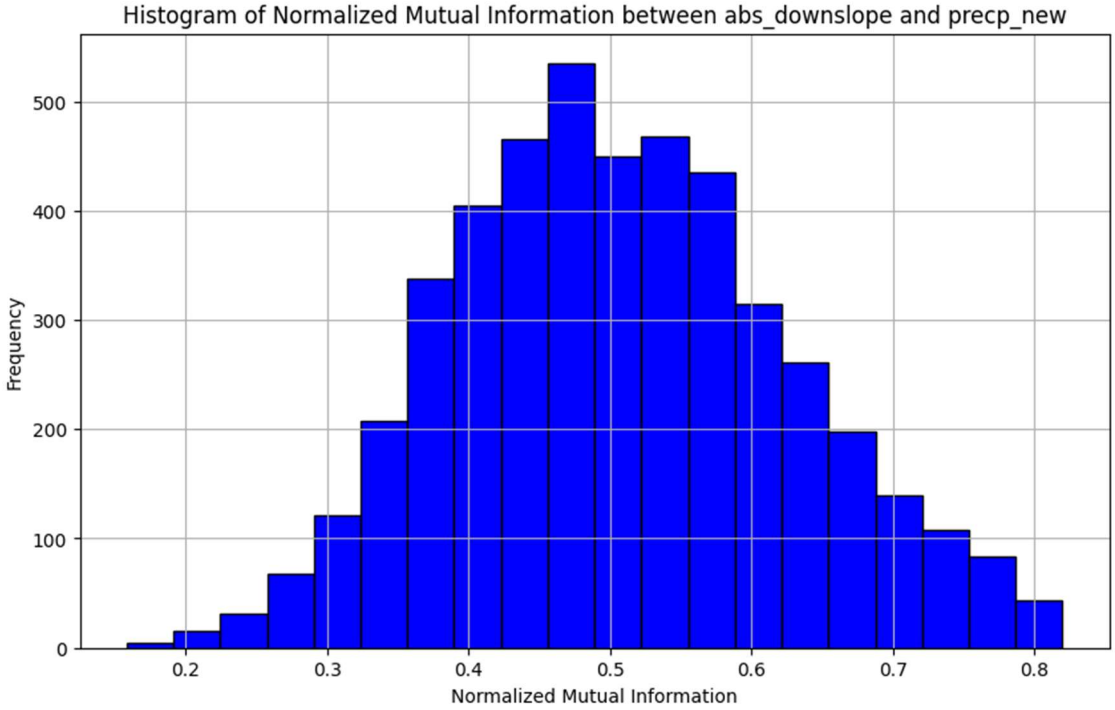
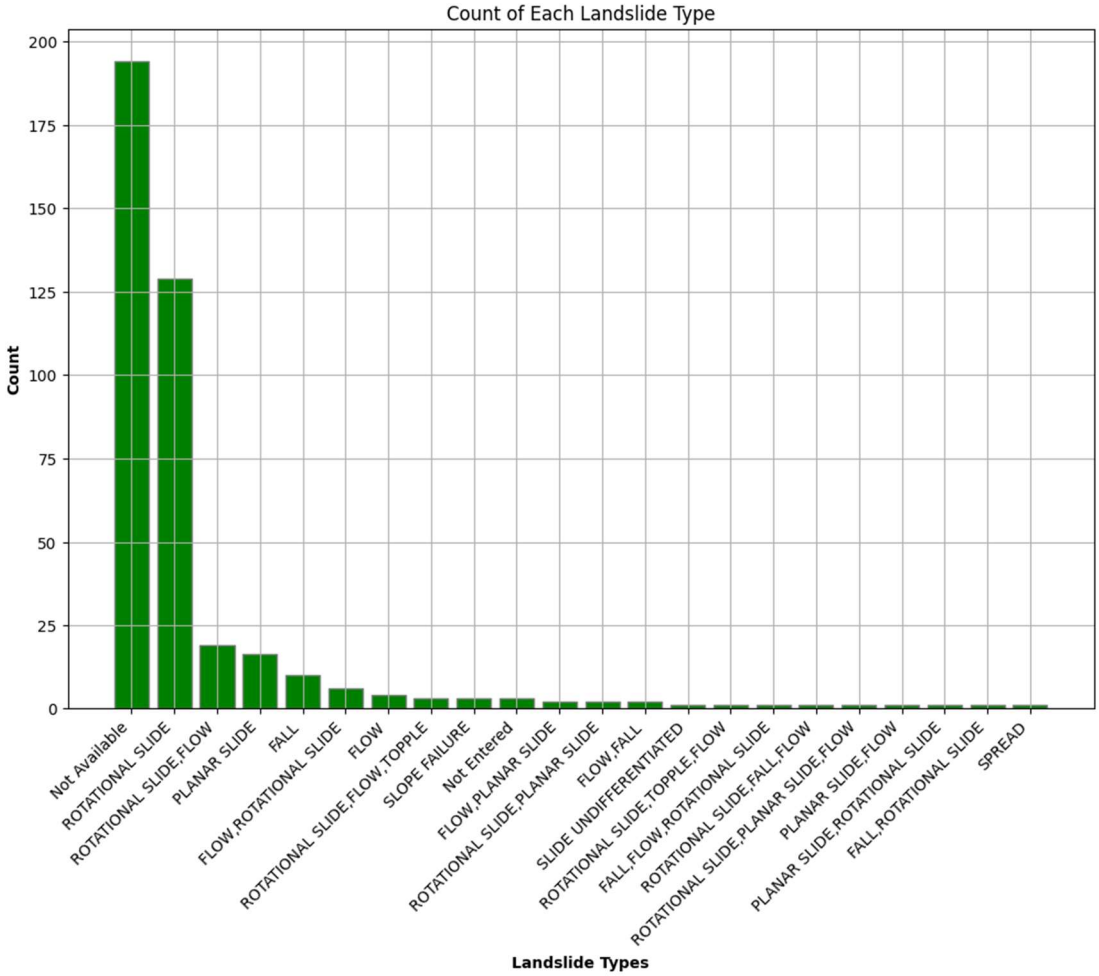


Figure 31 Bivariate plots between correlation values and other variables, with lag=12

9.4 Measuring non-linear relationship: Mutual Information



9.5 Landslide distribution by type in the region



9.6 Model Architecture composed of GAT + GRU layers

```
Model: "Spatio-Temporal model"
_________________________________________________________________________________
 Layer (type)                Output Shape              Param #    Connected to
=================================================================================
input (InputLayer)           (128, 22)                   0      []

att_src                       1 x 1 x 256                256 GATConv
att_dst                       1 x 1 x 256                256 GATConv
bias                          256                        256 GATConv
lin_src.weight                256 x 22                  5632 GATConv
weight_ih_l0                  384 x 256                98304 GRU
weight_hh_l0                  384 x 128                49152 GRU
bias_ih_l0                    384                        384 GRU
bias_hh_l0                    384                        384 GRU
weight                        1 x 128                    128 linear
bias                          1                            1 linear
=================================================================================
Total params: 154753
Trainable params: 154753
Non-trainable params: 0
```