Università degli Studi di Padova

Prevention of Potential Catastrophes Depending on Interferometric Radar Technique and Artificial Intelligence

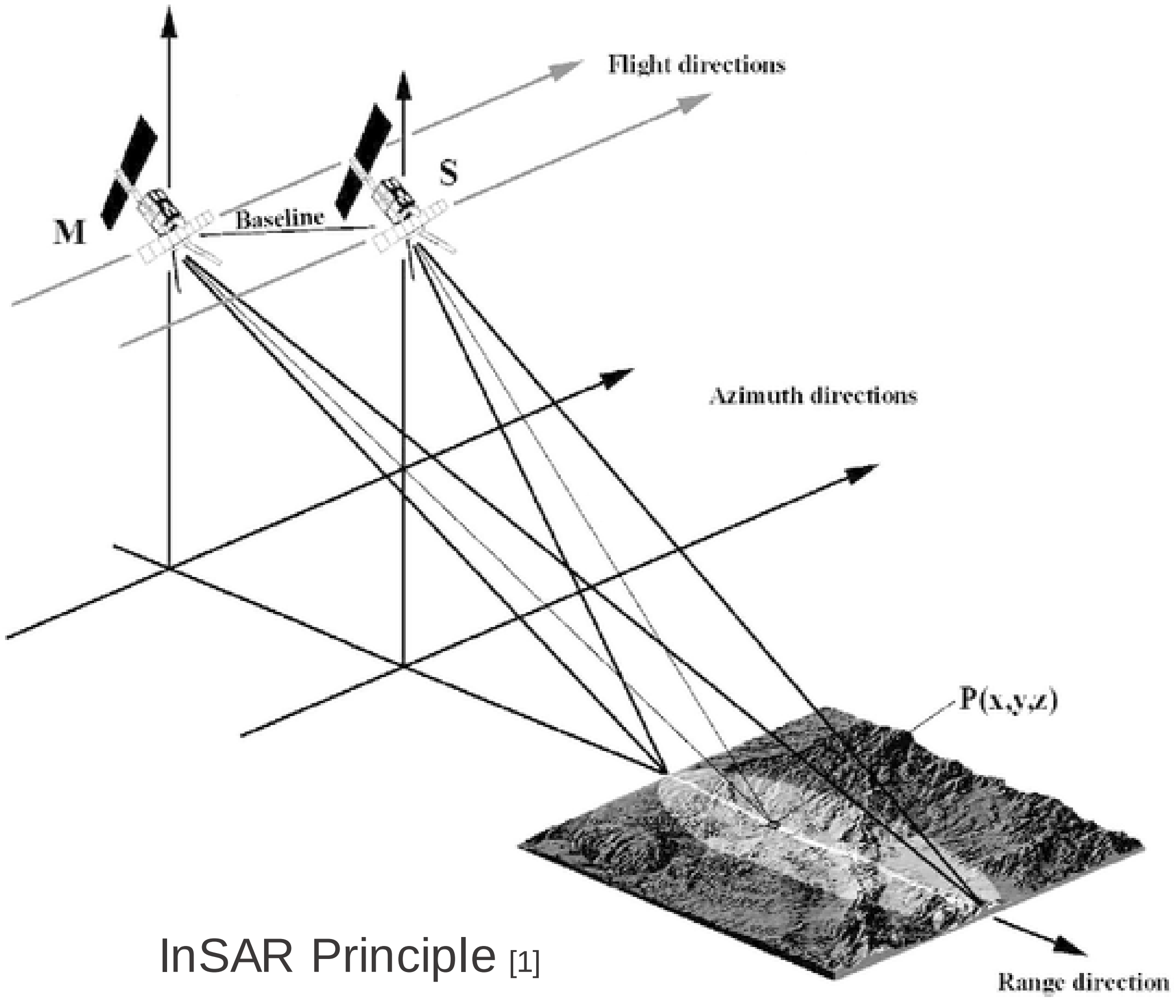
PhD Student: Lama Moualla - 36th Cycle

Supervisors: Dr. Alessio Rucci & Prof. Giampiero Naletto

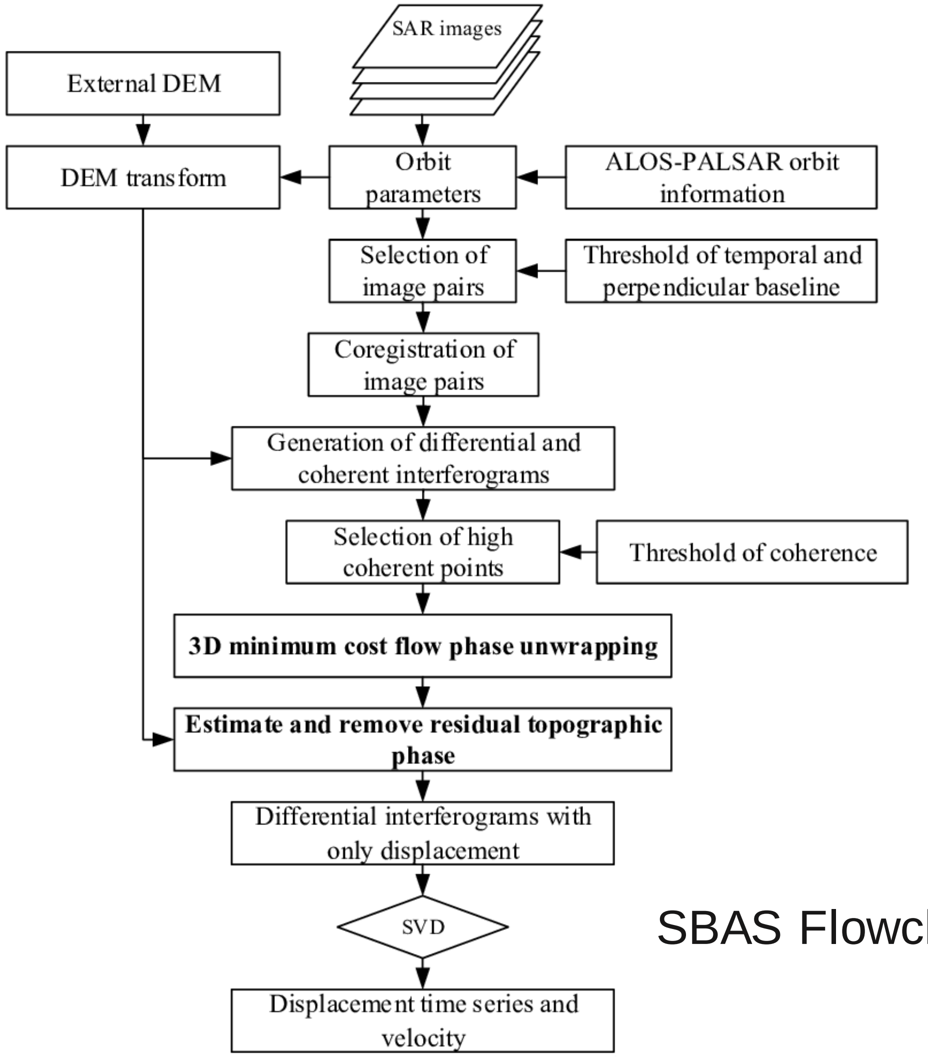
Meeting - 14/12/2023

Table of Contents

Challenges in Deploying InSAR Technology



InSAR Principle [1]



SBAS Flowchart [2]

UNIVERSITÀ DEGLI STUDI DI PADOVA

The First Objective

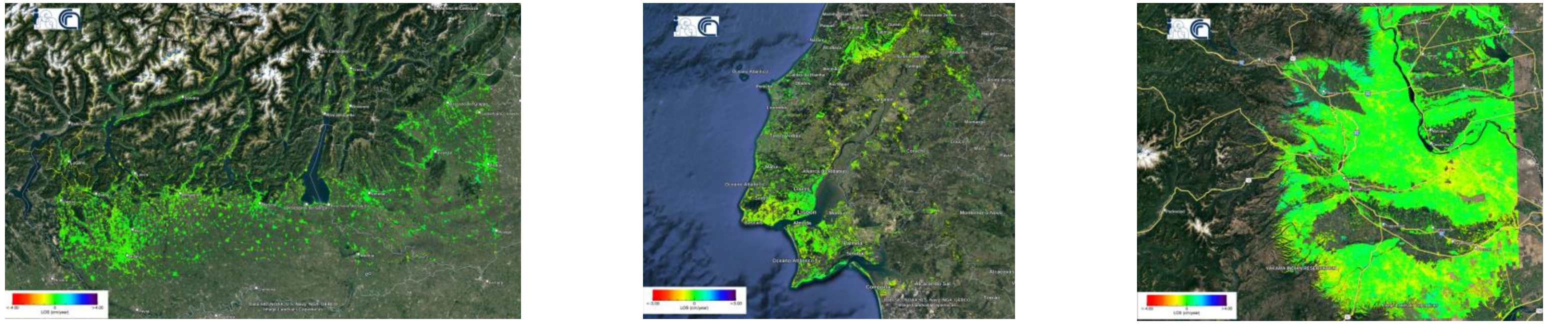
CISAS G. COLOMBO

Using Intelligent algorithms in developing a methodology that can automatically analyze large InSar data packets and identify areas where infrastructures are at risk of displacement due to ground movement

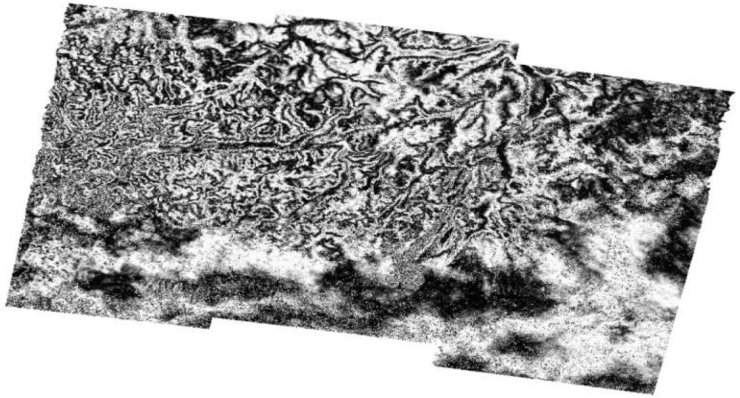
Università degli Studi di Padova

Recap of The First Objective Results

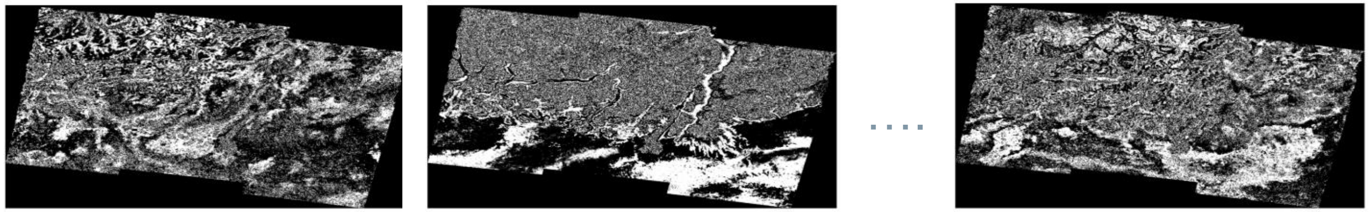
CISAS G. COLOMBO



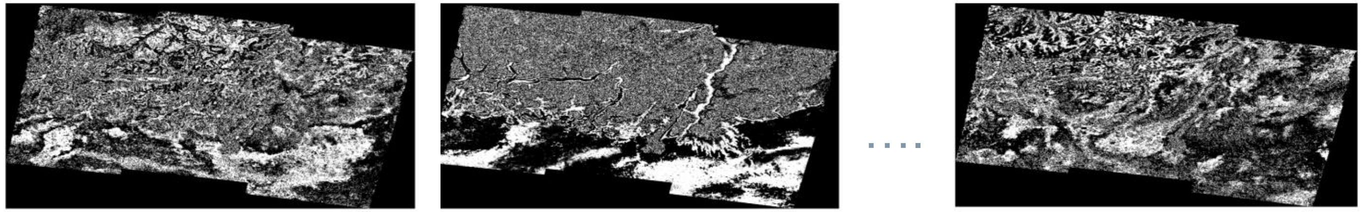
a)- Parallel - Small BAseline Subset Displacements Maps (Lombardy – Lisbon - Washington)



Wrapped Interferogram (Before Applying The High-Pass Filter)

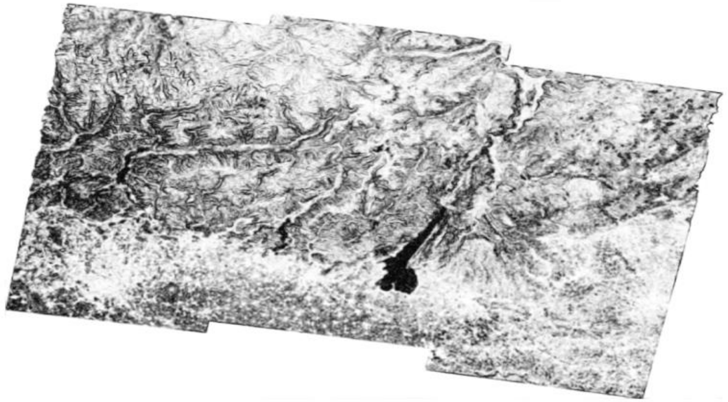


20191205 – 20191223 (3) 20200110 – 20200304 (2) …. 202000409 – 20200427 (1)



202000409 - 20200427 20200110 - 20200304 …. 20191205 - 20191223

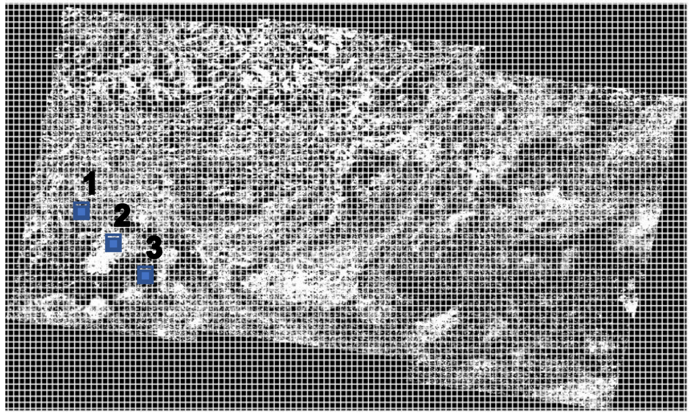
b)- The chronological sorting of the interferograms before creating the dataset and inputting the training samples into the model

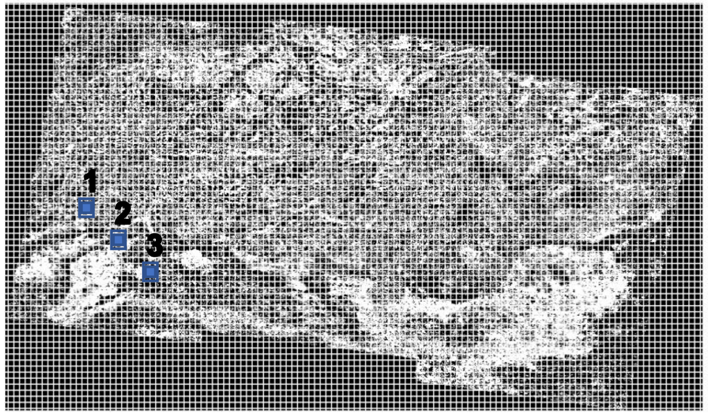


Wrapped Interferogram (After Applying The High-Pass Filter)

c)- Considering that the main power of the low frequency signal comes from the atmospheric artifacts , it was necessary to apply a high pass filter

Recap of The First Objective Results





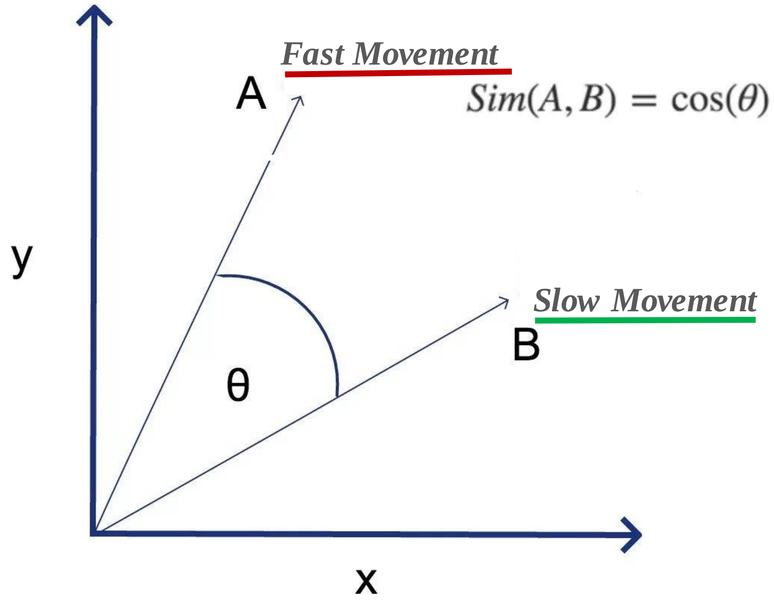
	1st Interferogram	2nd Interferogram	Last Interferogram	VEL	Label **(-0.7 cm/year)**
1	1.15	3.10	2.05	-1.5	Fast Negative
2	-2.10	2.55	-3.09	0.9	Undefined
3	-1.75	-1.15	-1.68	-0.1	Undefined

Training Examples
Phase Values (Radian)

Labels ??
The Velocity Values of The Measurement Point Inside The Pixel (cm/year)

The Best Trained Classifier: Cosine-KNN

- Handling of Angular Data
- Dimensionality and Feature Relationships
- Noise and Small Variations
- Suitability for Complex Patterns
- Generalization Capability

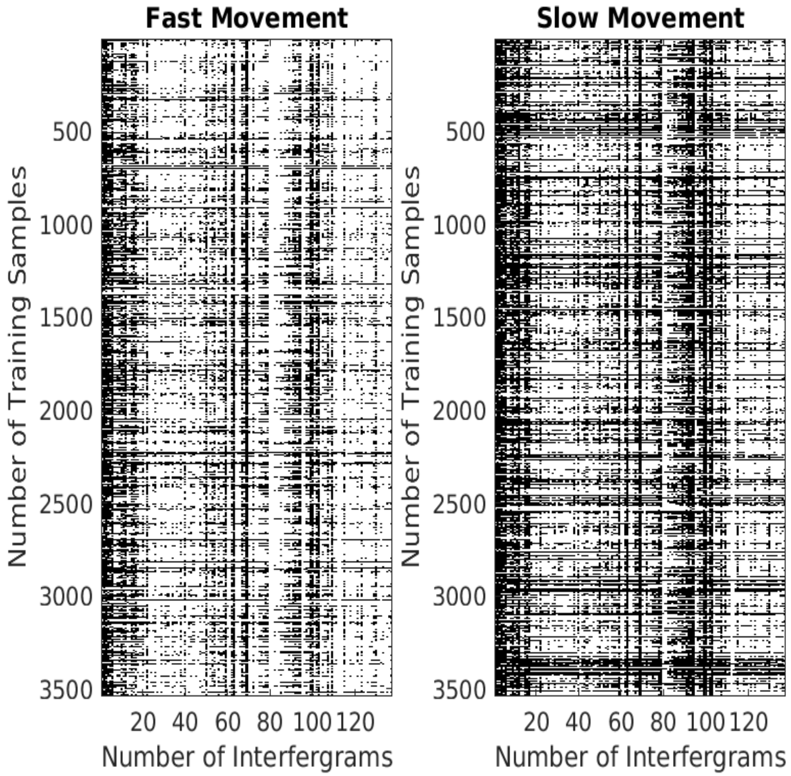


$$Sim(A,B) = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|}$$

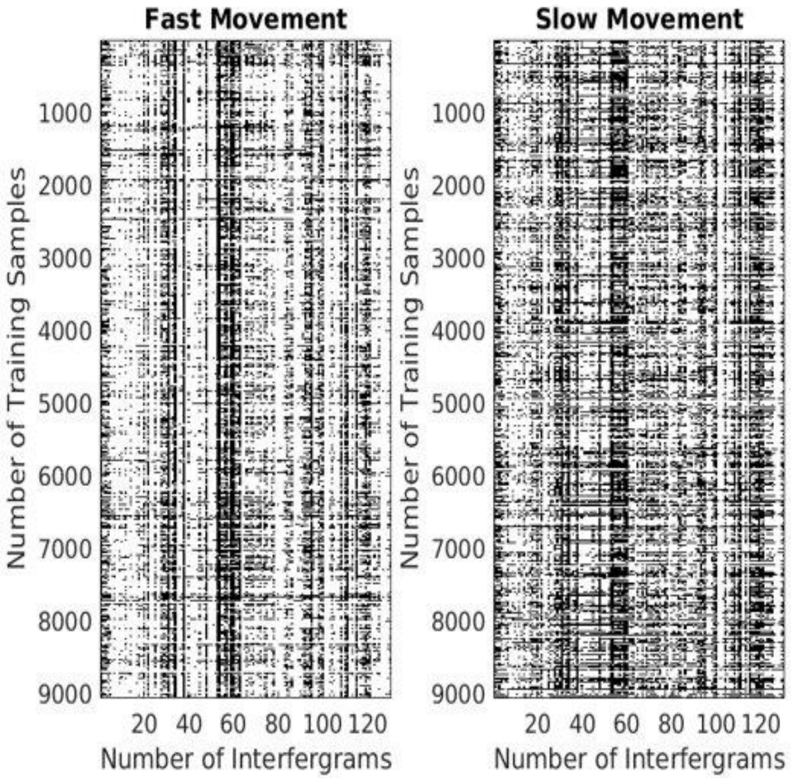
Trained Classifiers:
Cosine KNN, Subspace KNN (Ensemble), Medium Neural Network, Logistic Regression, Cubic SVM, Medium Tree, Fine Tree, Bagged Tree (Ensemble), Quadratic Discriminant, 2D Convolution Neural Network and Long short-term memory (LSTM)

UNIVERSITÀ DEGLI STUDI DI PADOVA

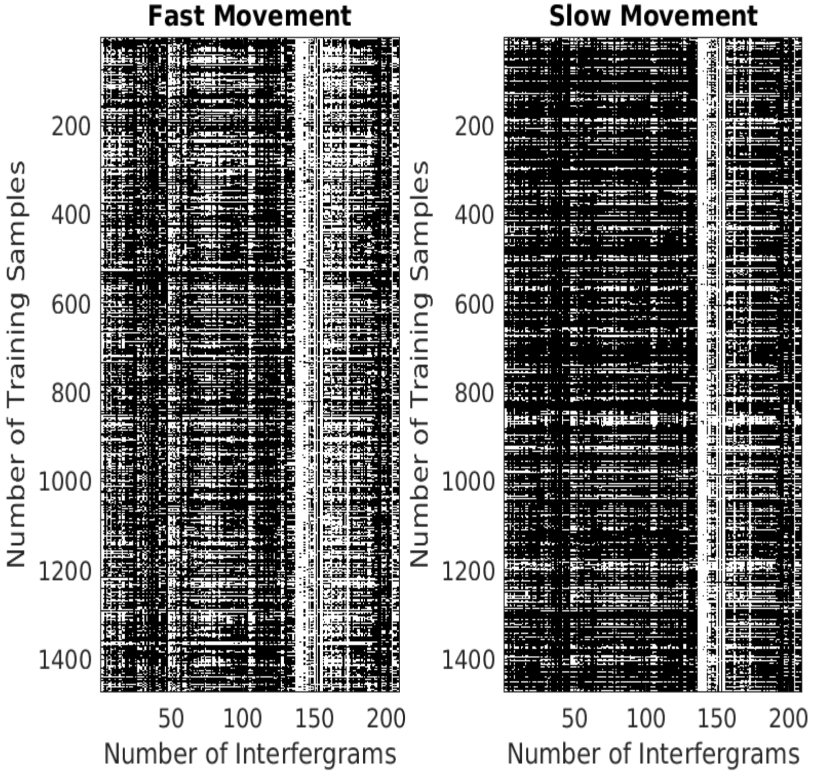
The Matrices Representing Slow and Fast Motion



Lombardy Dataset



Lisbon Dataset



Washington Dataset

The black color represents the magnitude values larger than 0.9 radian

The white color represents the magnitude values smaller than 0.9 radian

Psuedo-Labelling

A model is trained and tested on the labeled data

↓

The trained model is used to make predictions on the psuedo-labelled data

↓

A new training set is created by combining the testing samples with a prediction accuracy of 0.9 or higher with the original labeled data

↓

The pseudo-labeled training samples are removed from the original test set of unlabeled training samples

↓

The model is retrained using the combined labeled and pseudo-labeled data

Università degli Studi di Padova

Psuedo-Labelling Results

	Cosine K-NN	1st PS	2nd PS
Number of training samples	952	12334	29474
Number of testing samples	238	3084	7368
Accuracy of validation	72.8%	97.8%	98.7%
Accuracy of the test set	70.6%	97.9%	98.8%

Results of Pseudo-Labelling PS on Lombardy Dataset (Fast Positive/Undefined)

	Cosine K-NN	1st PS	2nd PS
Number of training samples	4884	6102	6684
Number of testing samples	1220	1526	1672
Accuracy of validation	81.2%	85.1%	86.1%
Accuracy of the test set	82.3%	84.3%	86.8%

Results of Pseudo-Labelling PS on Washington Dataset (Fast Negative/Undefined)

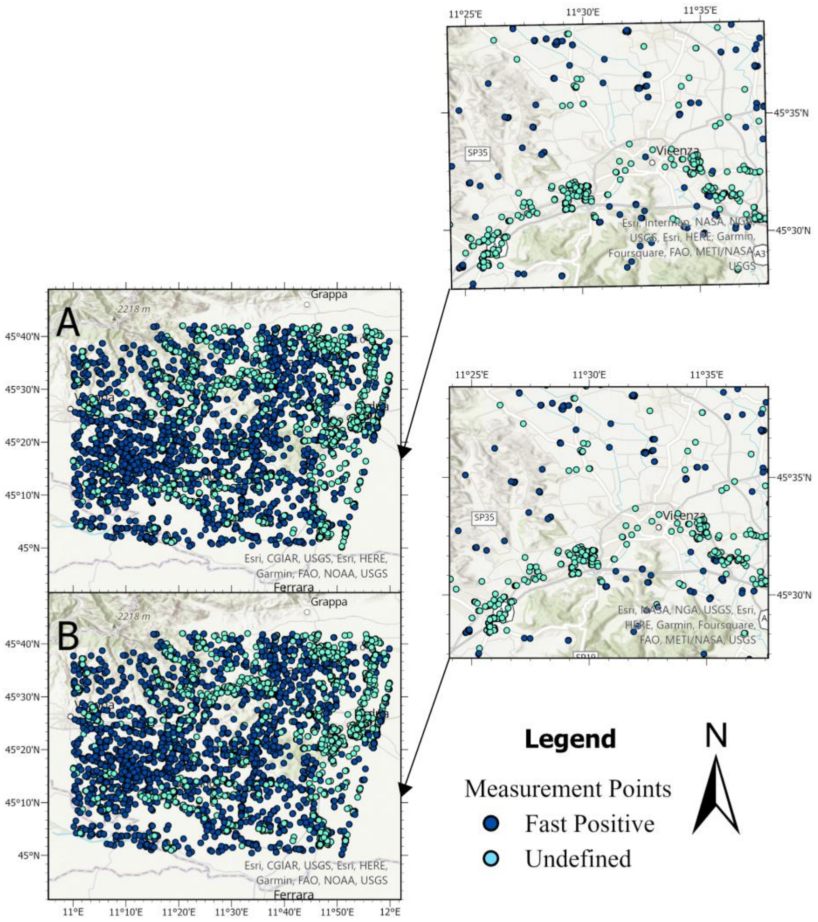
	Cosine K-NN	1st PS	2nd PS	3rd PS	4th PS
Number of training samples	14596	17812	20130	21576	22736
Number of testing samples	3650	4454	5032	5394	5684
Accuracy of validation	79.4%	83.4%	85.5%	86.6%	87.2%
Accuracy of the test set	79.4%	82.8%	85.7%	87.1%	87.1%

Results of Pseudo-Labelling PS on Lisbon Dataset (Positive/Negative)

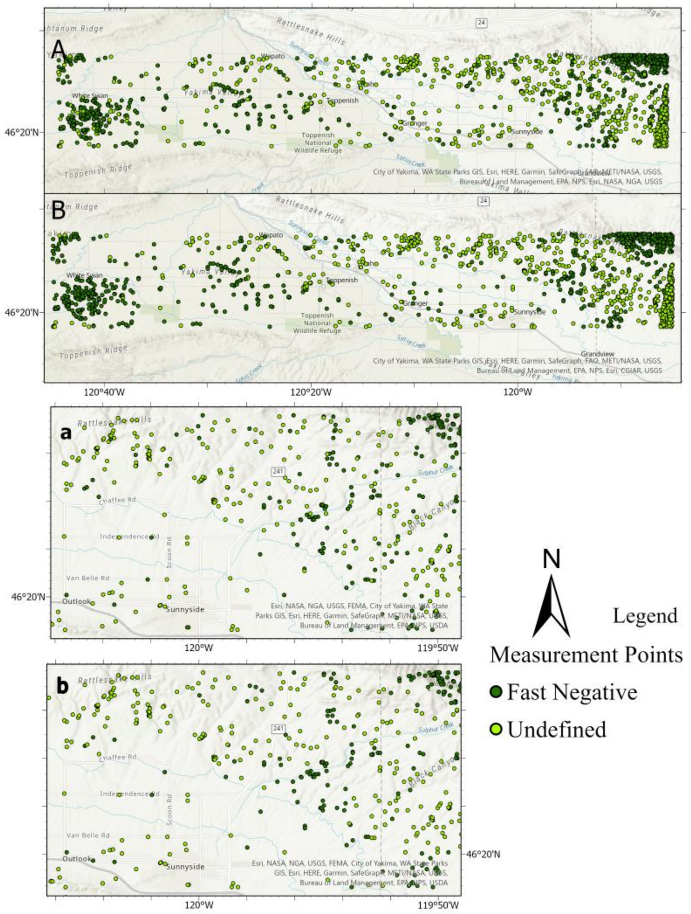
Università degli Studi di Padova

The Ground-truth Test Set vs Predictions

Lombardy Dataset (Fast Positive/ Undefined Movement Model)



Washington Dataset (Fast Negative/ Undefined Movement Model)



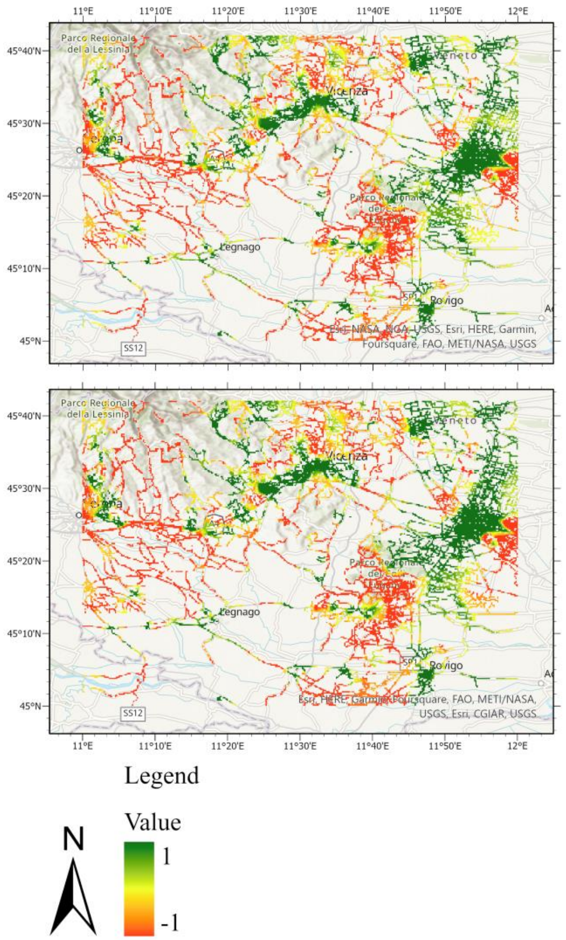
Lisbon Dataset (Positive / Negative Movement Model)

Figures a display the ground truth of the test dataset. Figures b display the predictions of the test dataset.

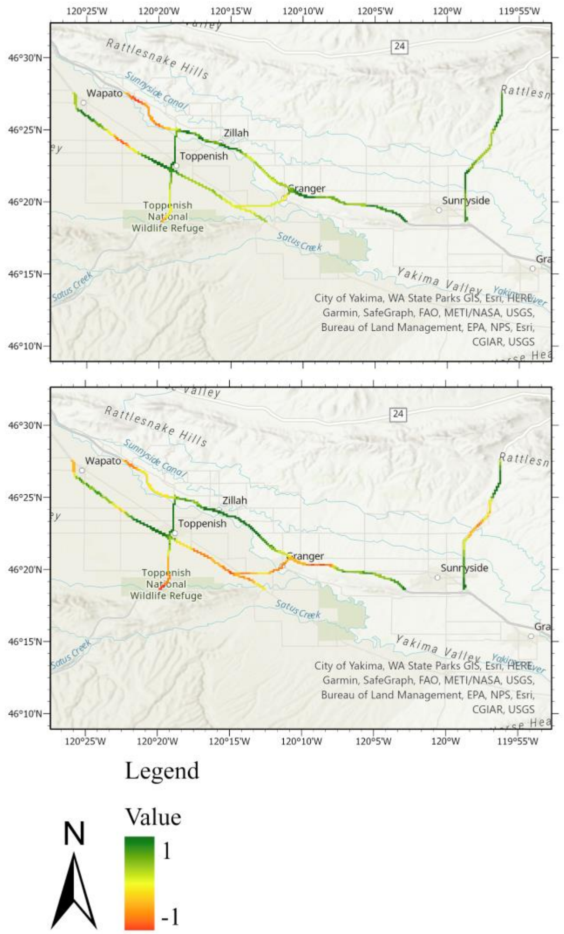
Università degli Studi di Padova

Masked Roads Network

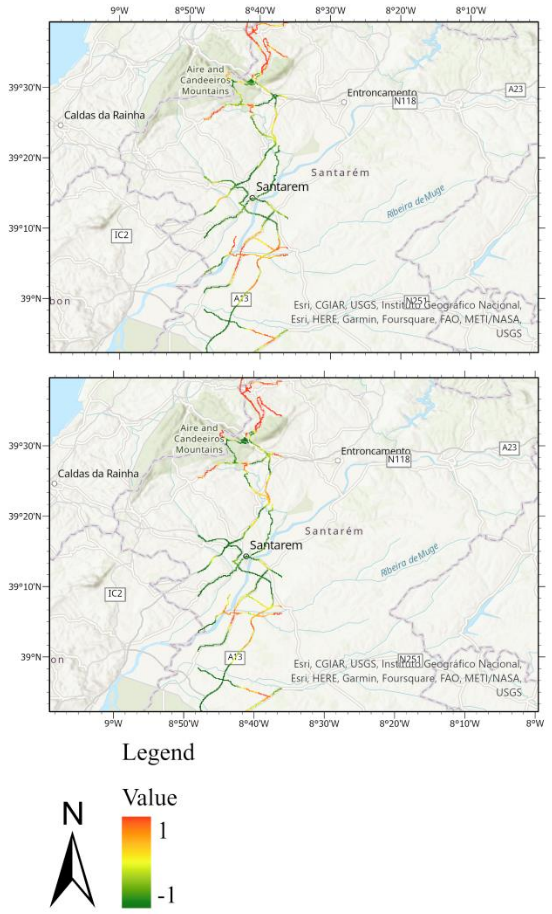
Fast Positive Movement Roads Indication (Milan Dataset)



Fast Negative Movement Roads Indication (Washington Dataset)



Positive/Negative Movement Roads Indication (Lisbon Dataset)



The top figures represent the masked roads of the ground truth test set; while the bottom figures represent the masked roads of the predicted test sets
The value of 1 expresses the positive movement while the value -1 expresses the negative movement

The Second Objective

Establishing a predictive model for the displacements of the infrastructure studied in the research, based on the methodology that we will try to develop

Data Preprocessing

Before Resampling

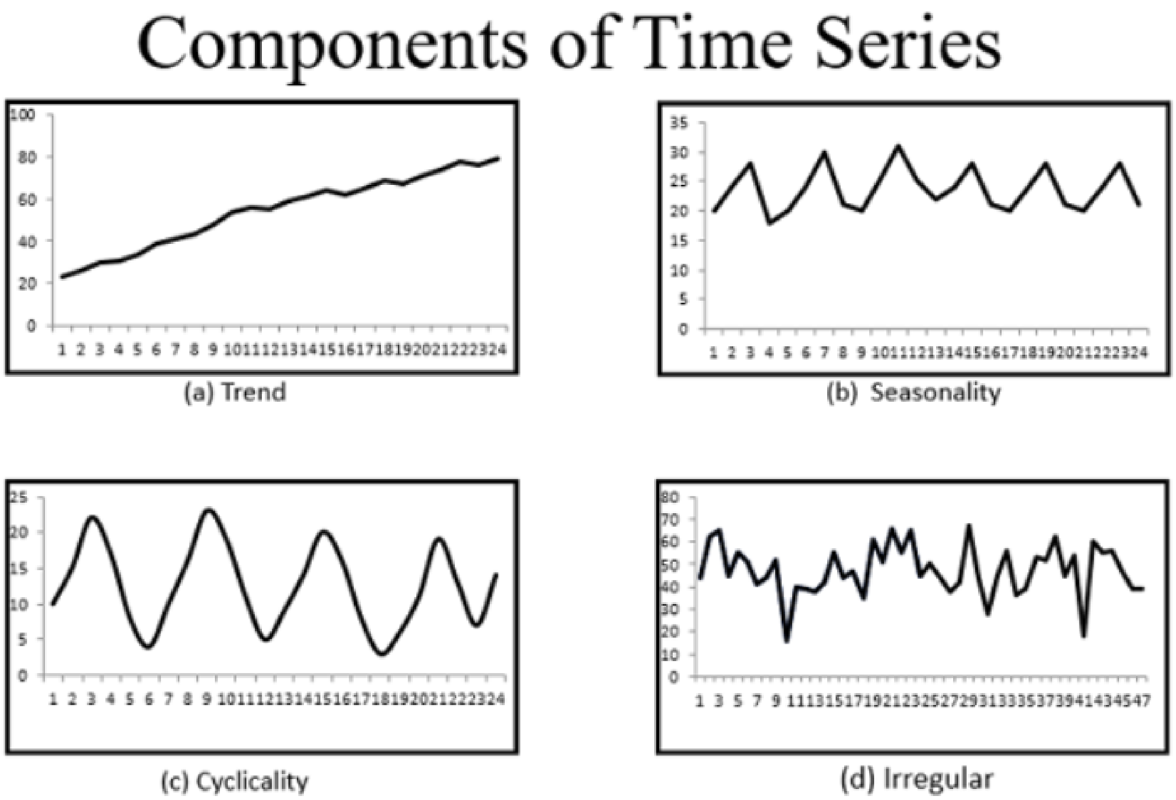
Name of the case study	Number of the time steps	Temporal difference between the time steps (days)
Milan	50	12
Lisbon	48	6, 12, 18, 24
Washington	75	12, 24, 48, 222

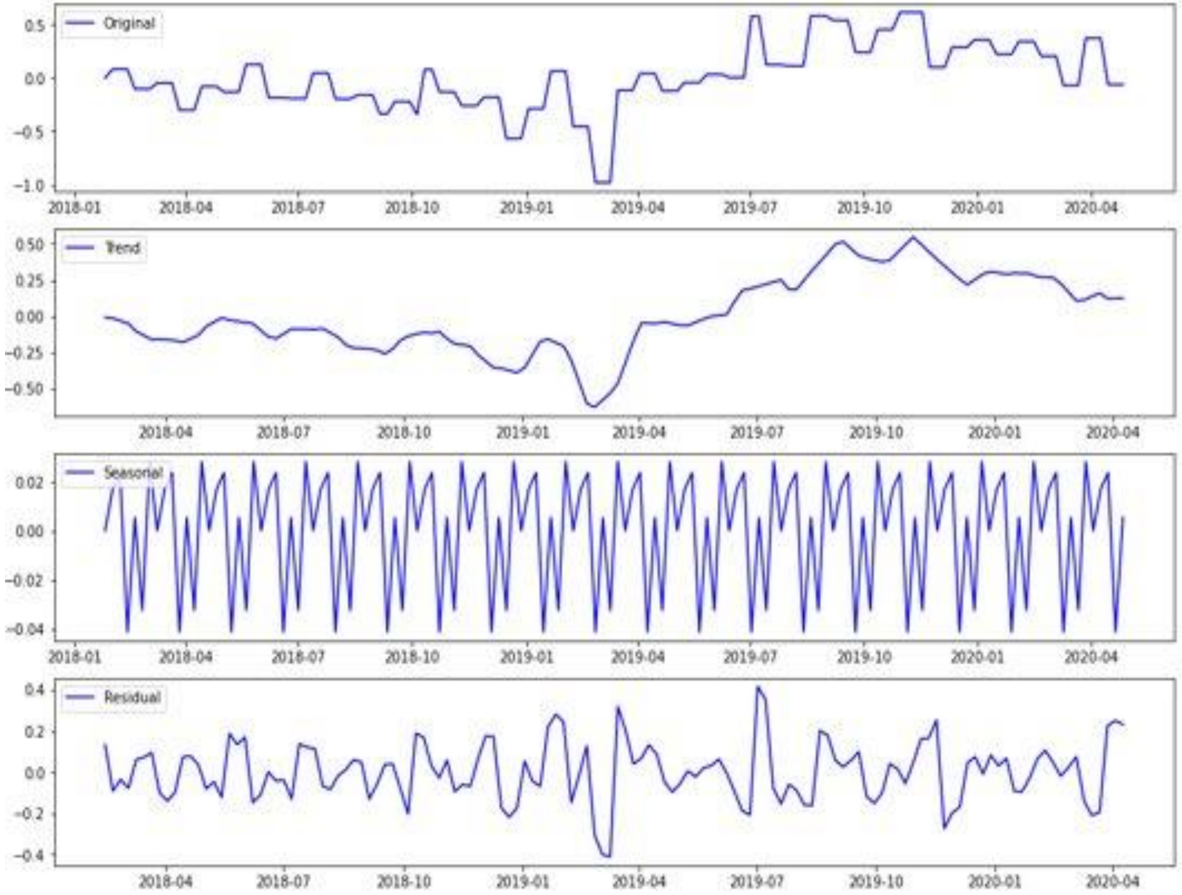
Data Preprocessing

- Missing Values Imputation (Backward Filling)

- Feature Engineering (Embedding Time as a Second Feature in Time Series Data)

Data Preprocessing





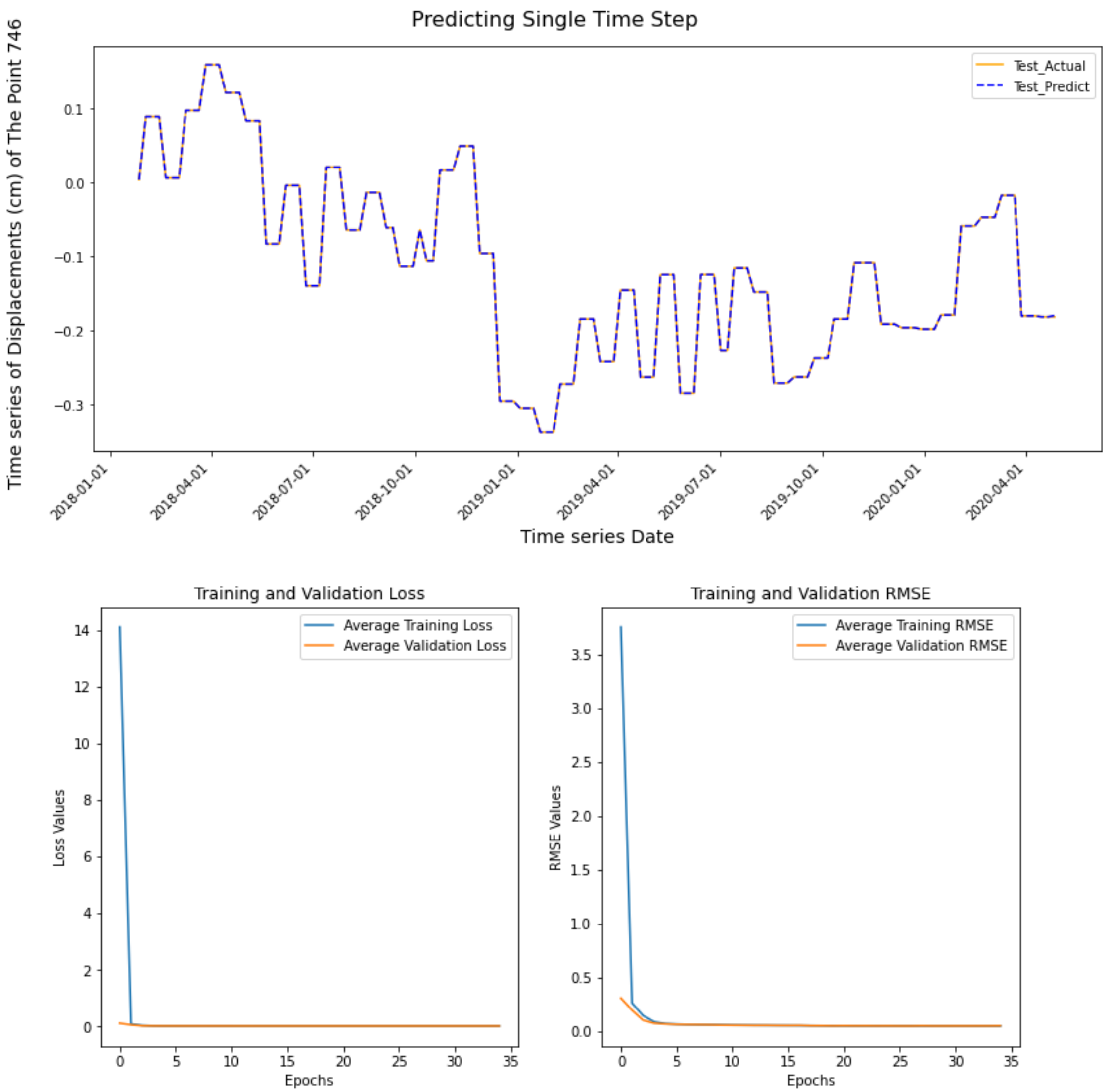
Lisbon Dataset

Why LSTM ?

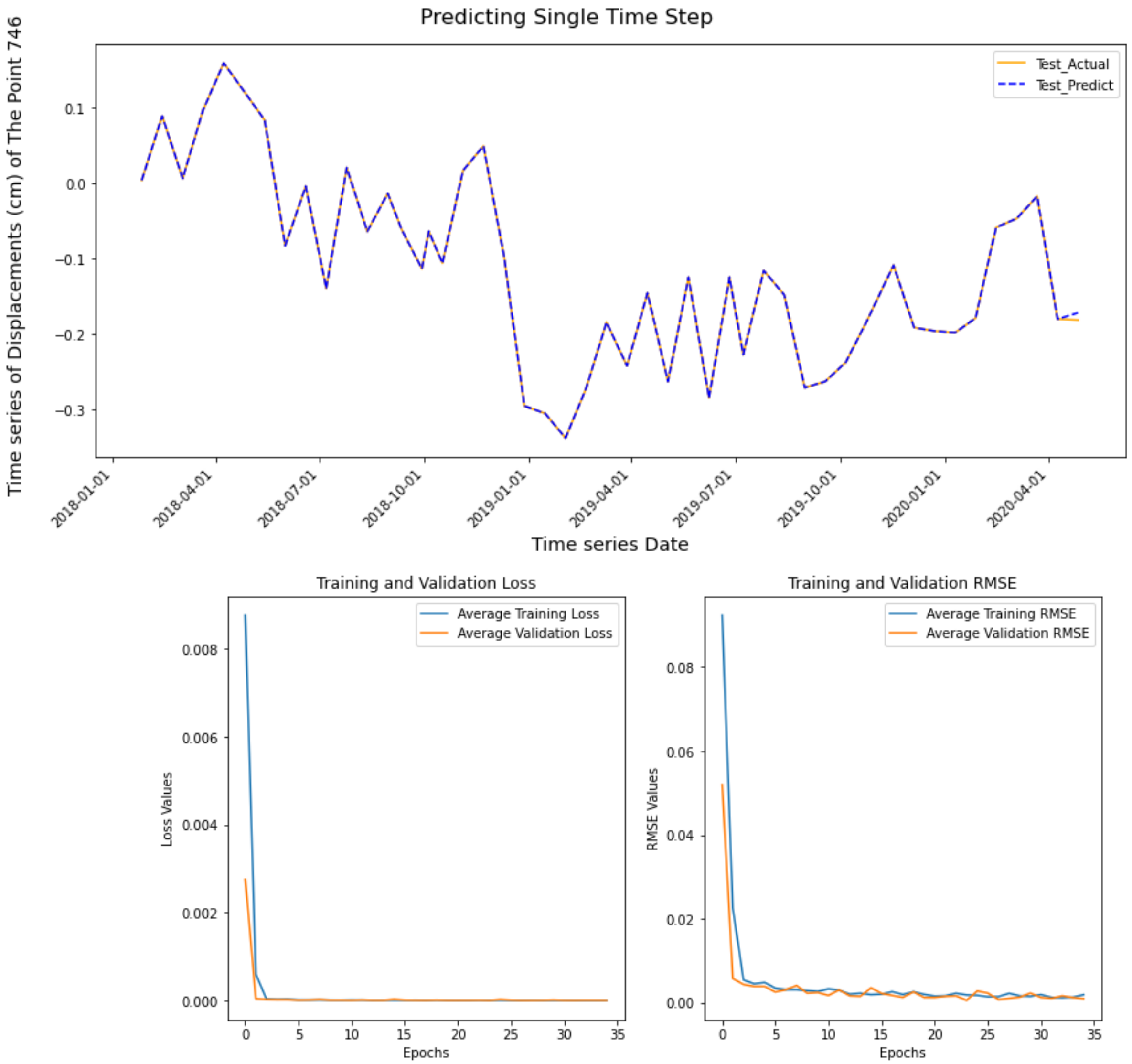
- Handling Temporal Dependencies
- Spatial Correlations
- Noise and Anomaly Tolerance
- Multivariate Time Series Capability
- Adaptability
- Real-Time Prediction

UNIVERSITÀ DEGLI STUDI DI PADOVA

Results of One Time Step Predictions (Lisbon Dataset)



Learning Curves Influenced by the Imputation of Missing Values

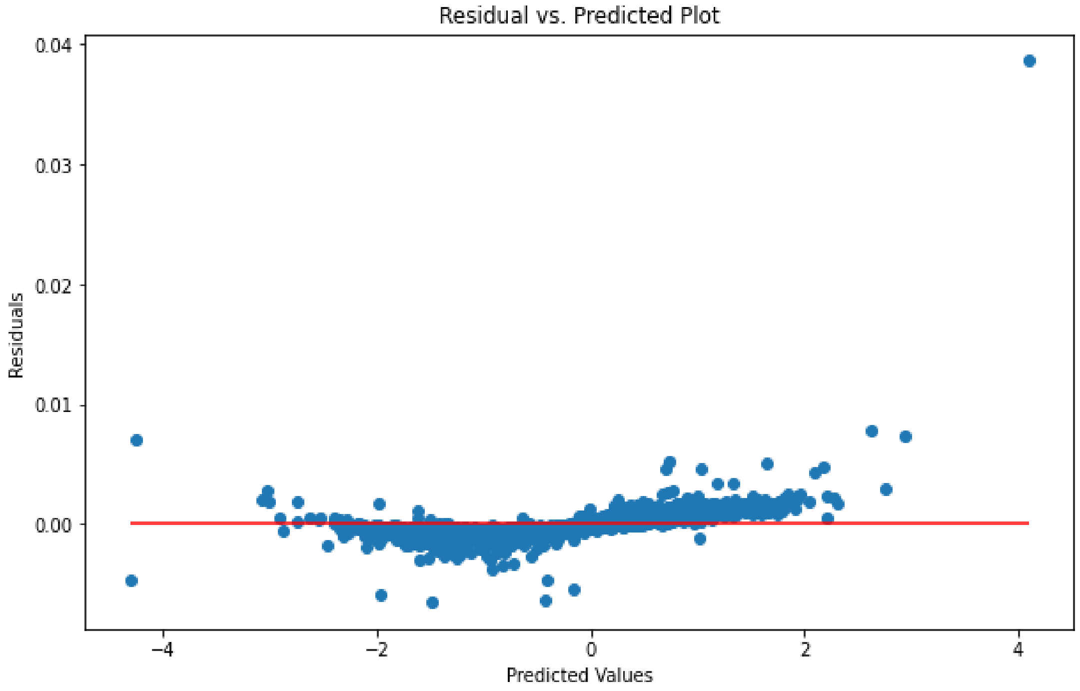


Learning Curves Influenced by the Time Embedding

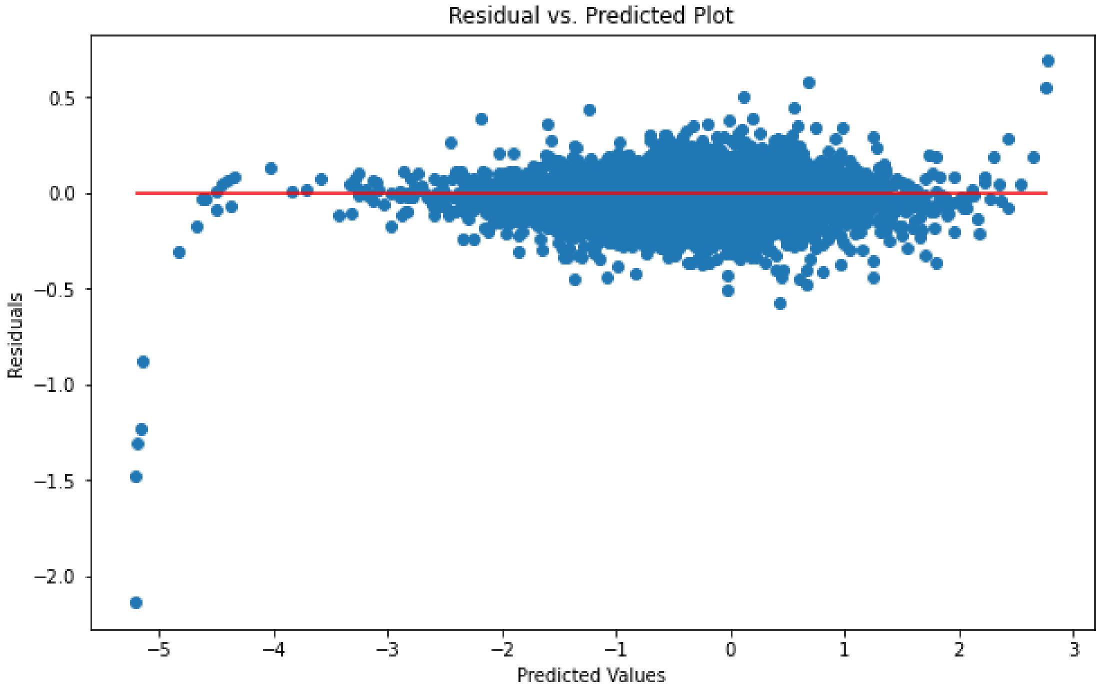
Results of One Time Step Predictions (Lisbon Dataset)

Are the residuals exhibiting constant variance across different levels of predicted values?

- **Homoscedasticity** is a desirable property in a regression model, as it suggests that the model appropriately captures the variance in the data across all levels of the independent variable(s).
- **Heteroscedasticity** often suggests that the model is missing important features, a misspecification in the model, or that a transformation of variables might be necessary.



Homoscedasticity as Affected by the Imputation of Missing Values

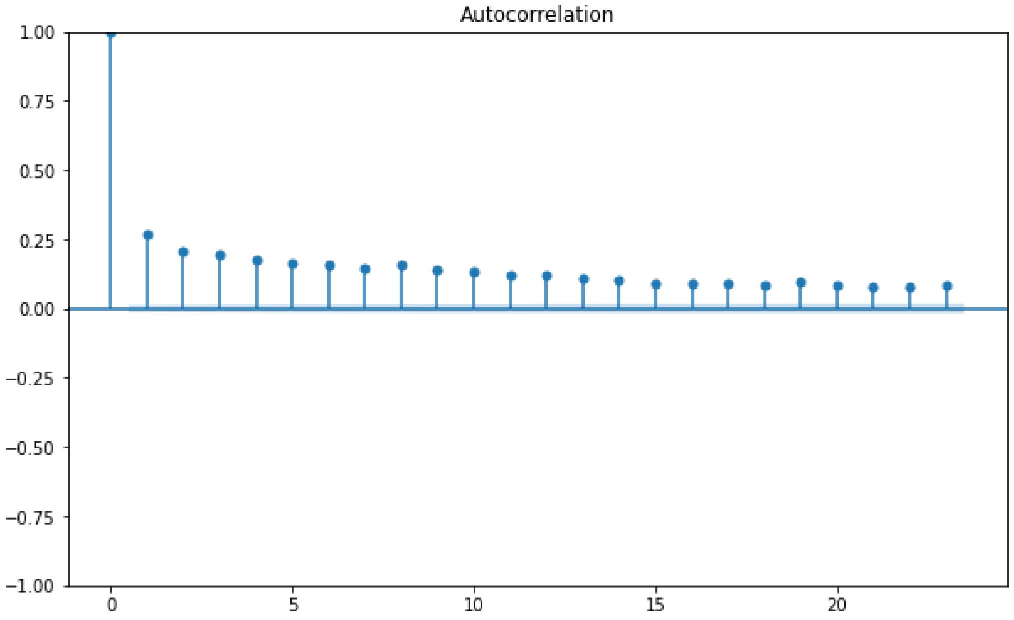


Homoscedasticity as Affected by the Time Embedding

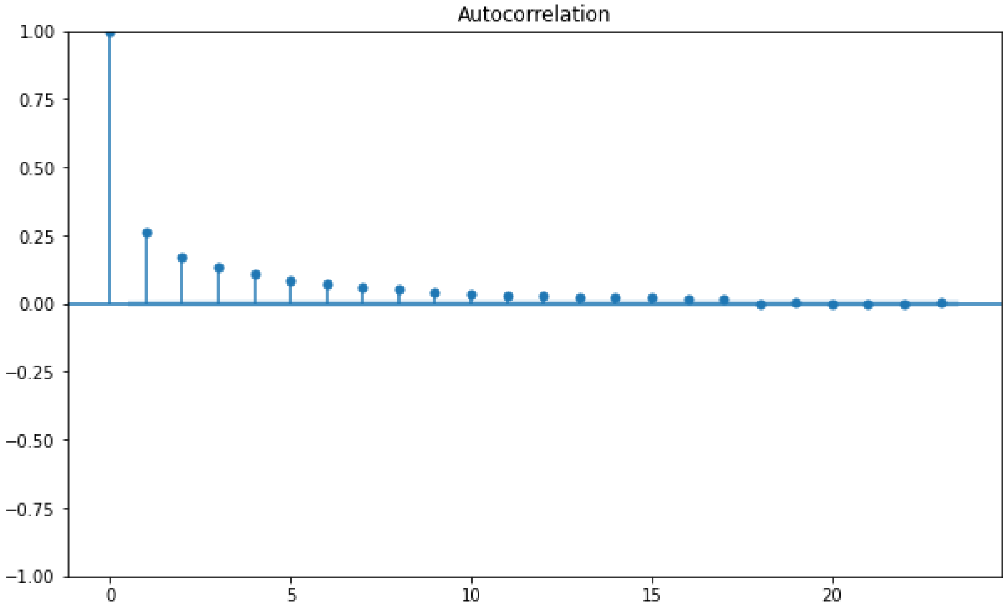
Results of One Time Step Predictions (Lisbon Dataset)

The Autocorrelation Function (ACF) is indispensable in the analysis of time series data as it reveals any remaining correlation in the residuals of the model's predictions.

- Significant autocorrelation at any lag could suggest that the model has not fully captured the predictive structure within the data, indicating room for improvement.
- The absence of such correlation, on the other hand, would affirm that the model's predictions are not systematically biased by overlooked temporal dependencies.



ACF Charts Depending on the Imputation of Missing Values

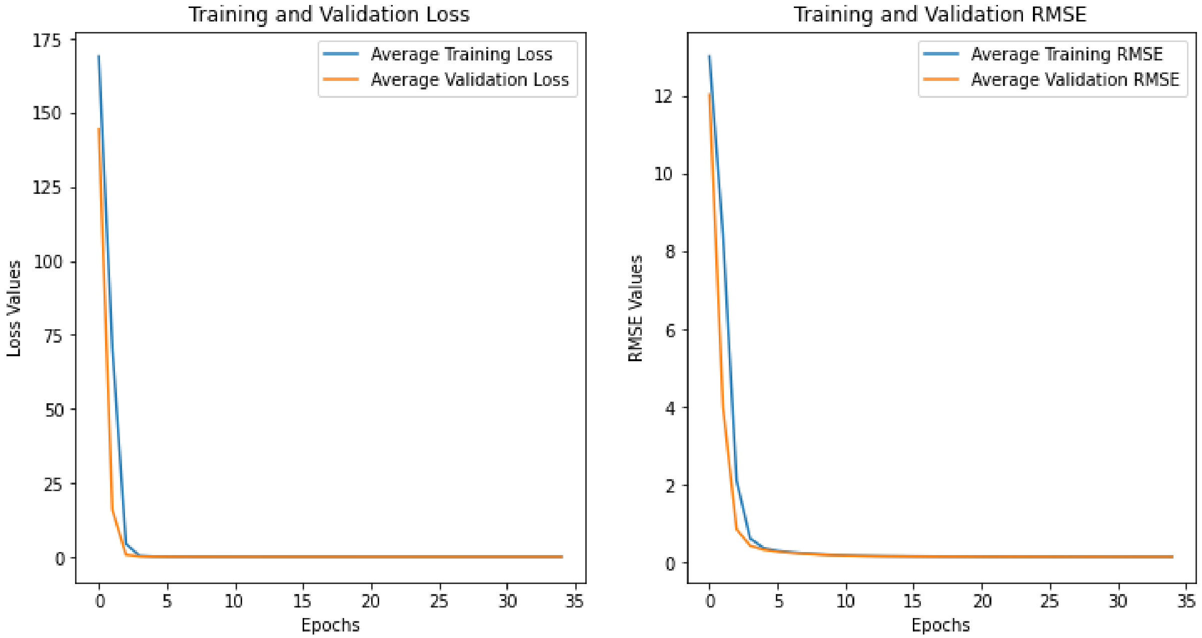


ACF Charts Depending on the Time Embedding

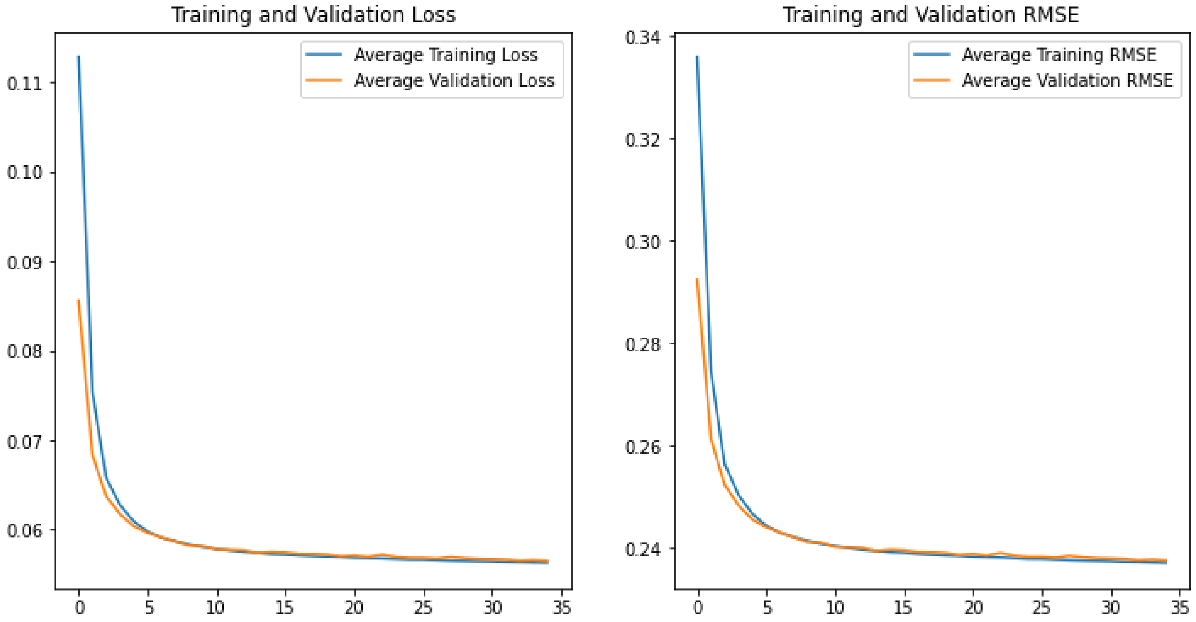
Results of Multiple Time Steps Predictions

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
rnn (RNN)                    (None, 38, 225)           206100

rnn_1 (RNN)                  (None, 75)                90600

dense (Dense)                (None, 20)                1520

reshape (Reshape)            (None, 10, 2)             0

=================================================================
Total params: 298,220
Trainable params: 298,220
Non-trainable params: 0
_________________________________________________________________
```

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm (LSTM)                  (None, 110, 100)          40800

lstm_1 (LSTM)                (None, 50)                30200

dense (Dense)                (None, 28)                1428

=================================================================
Total params: 72,428
Trainable params: 72,428
Non-trainable params: 0
_________________________________________________________________
```



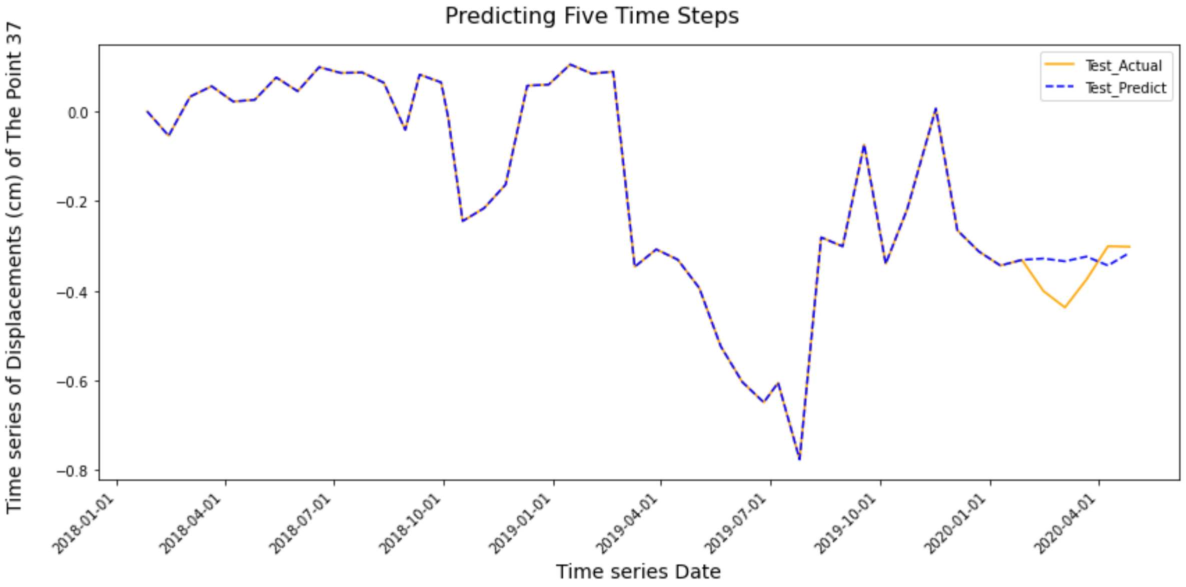
Learning Curves of training Time Gated LSTM model (10 Steps)

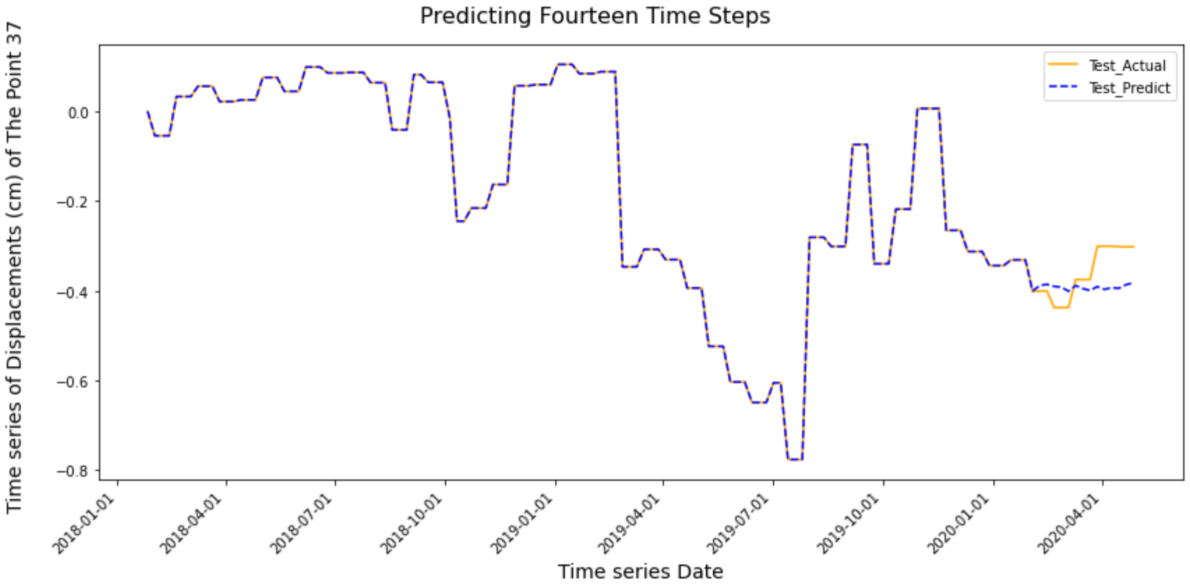


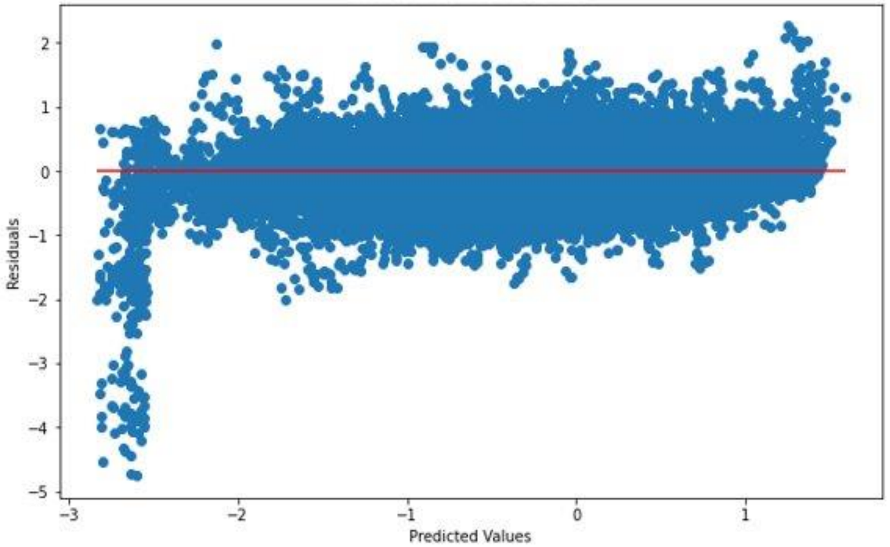
Learning Curves of training standard LSTM model (28 Steps)

Università degli Studi di Padova

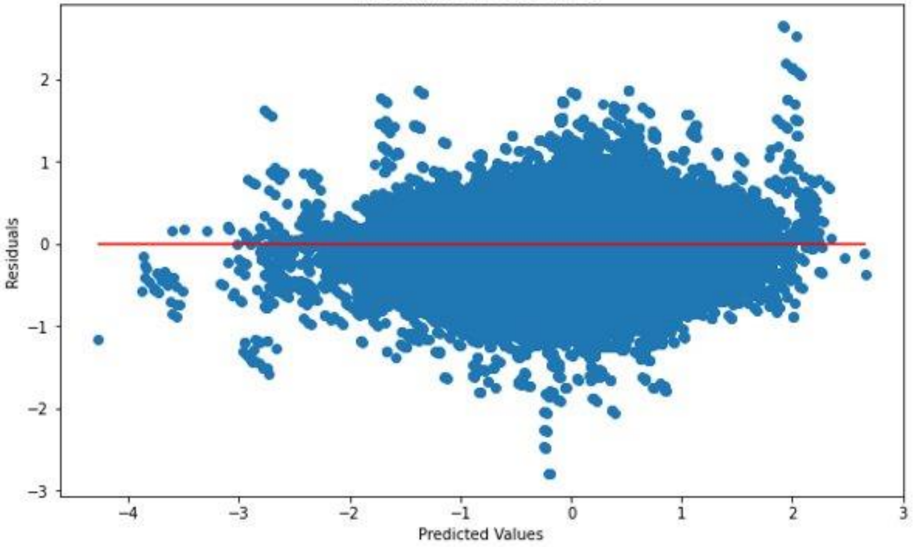
Results of Multiple Time Steps Predictions







Homoscedasticity of training Time Gated LSTM model

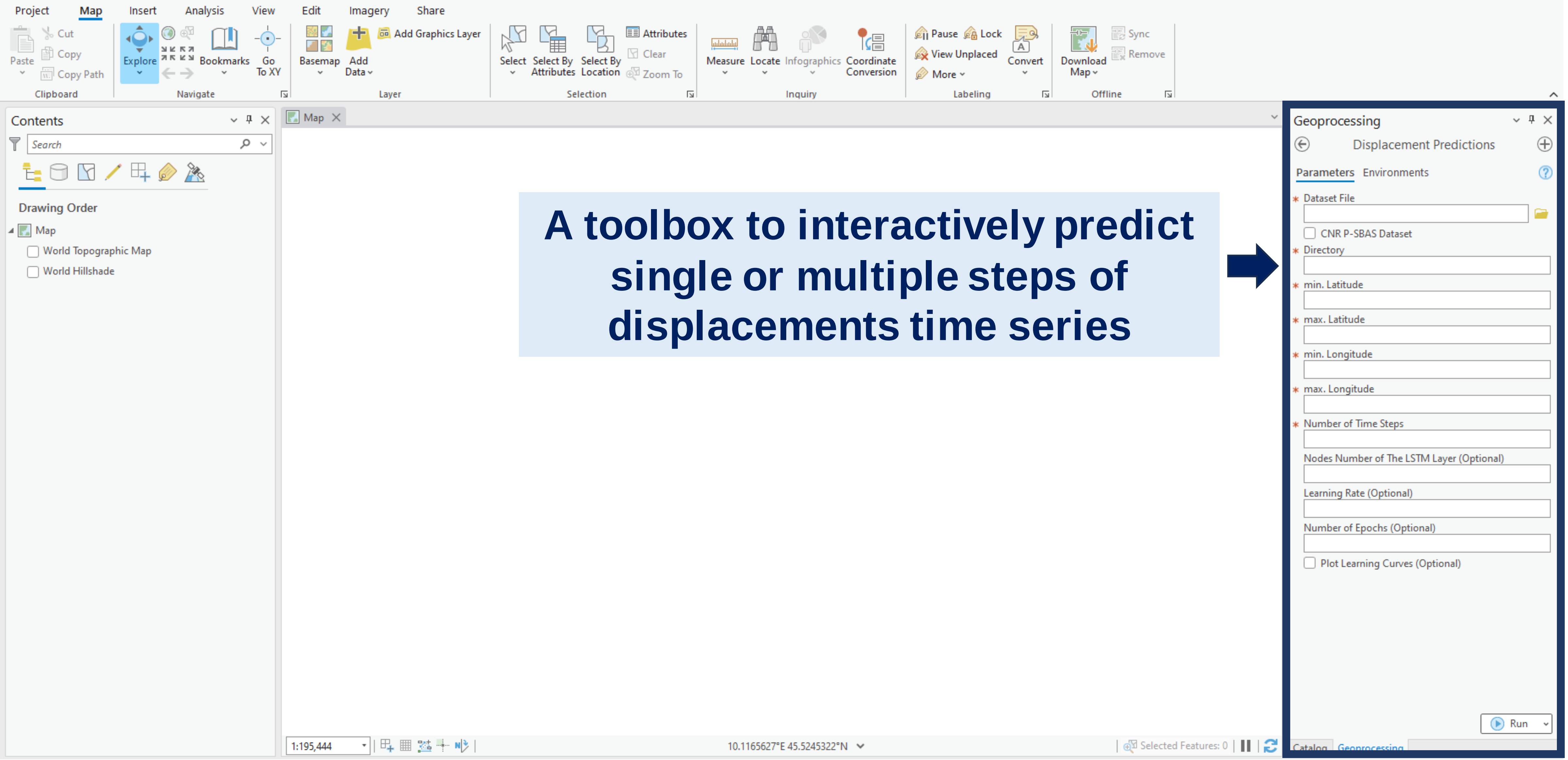


Homoscedasticity of training Standard LSTM model

The Third Objective

Determine the functions of the Geographic Information Systems (GIS) toolbox that will be developed to integrate the final work results within the ArcGIS Pro environment

Displacement Predictions Toolbox



A toolbox to interactively predict single or multiple steps of displacements time series

Università degli Studi di Padova

Displacement Predictions Toolbox

Insert the path for the dataset csv file and check the box if it generated by CNR P-SBAS service

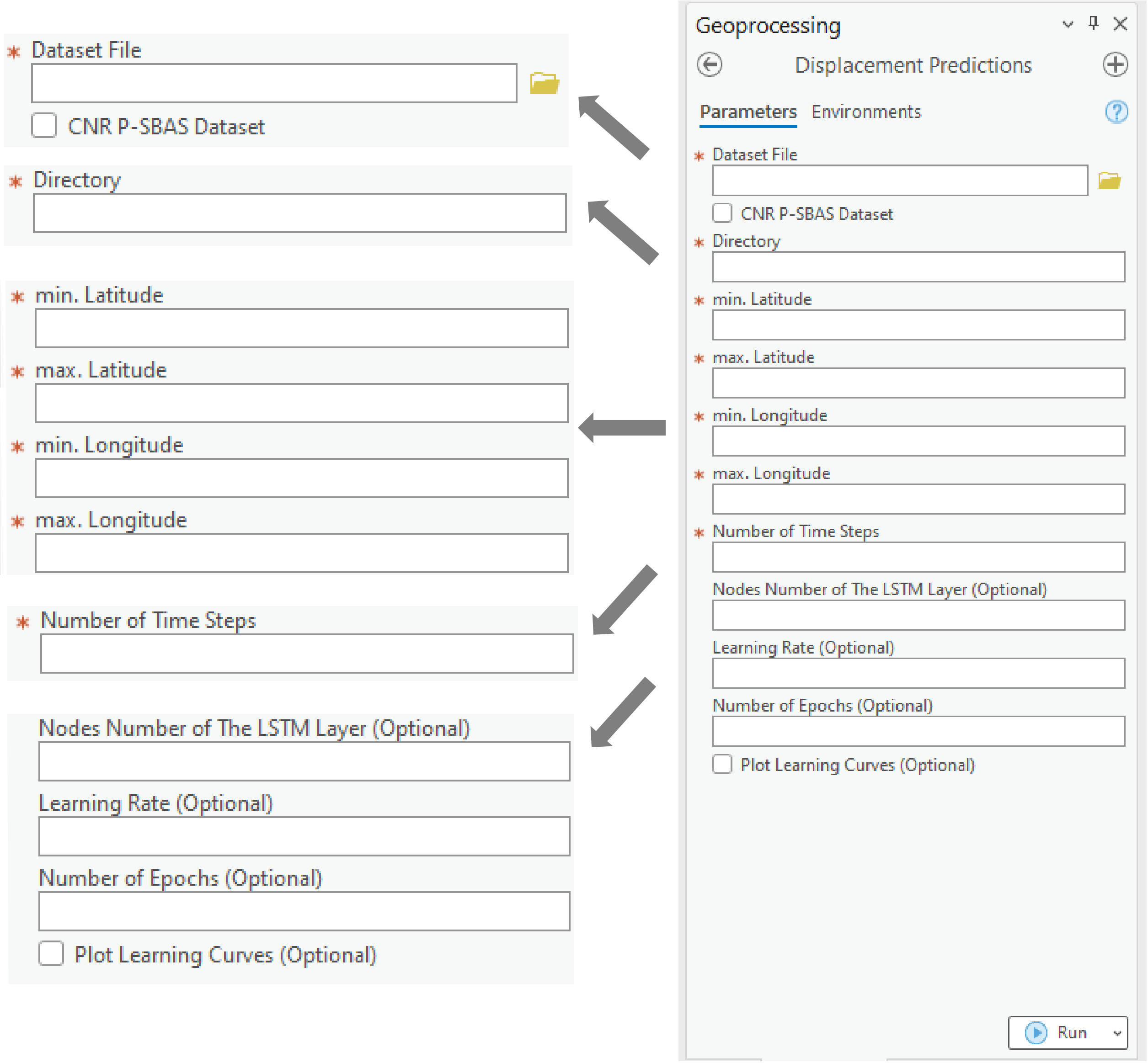
Insert the path for the output file where the results will be saved

Inter the geographic coordinates of the boundary for the studied area

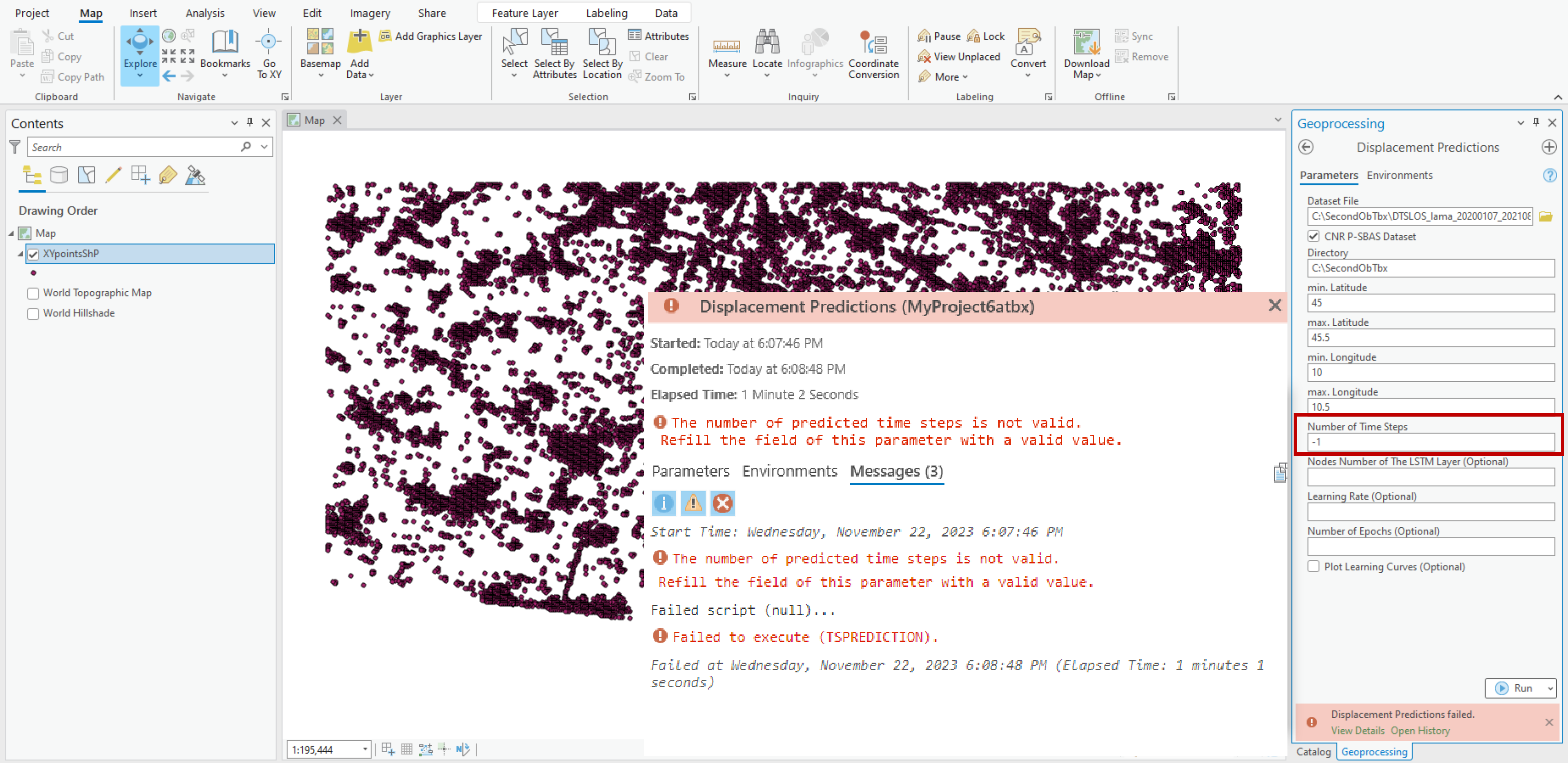
Insert the number of time steps you want to predict

Change the default values of the following hyperparameters in case you are not satisfied with the RMSE Result

Check the final box if it is important to be sure of the learning process efficacy

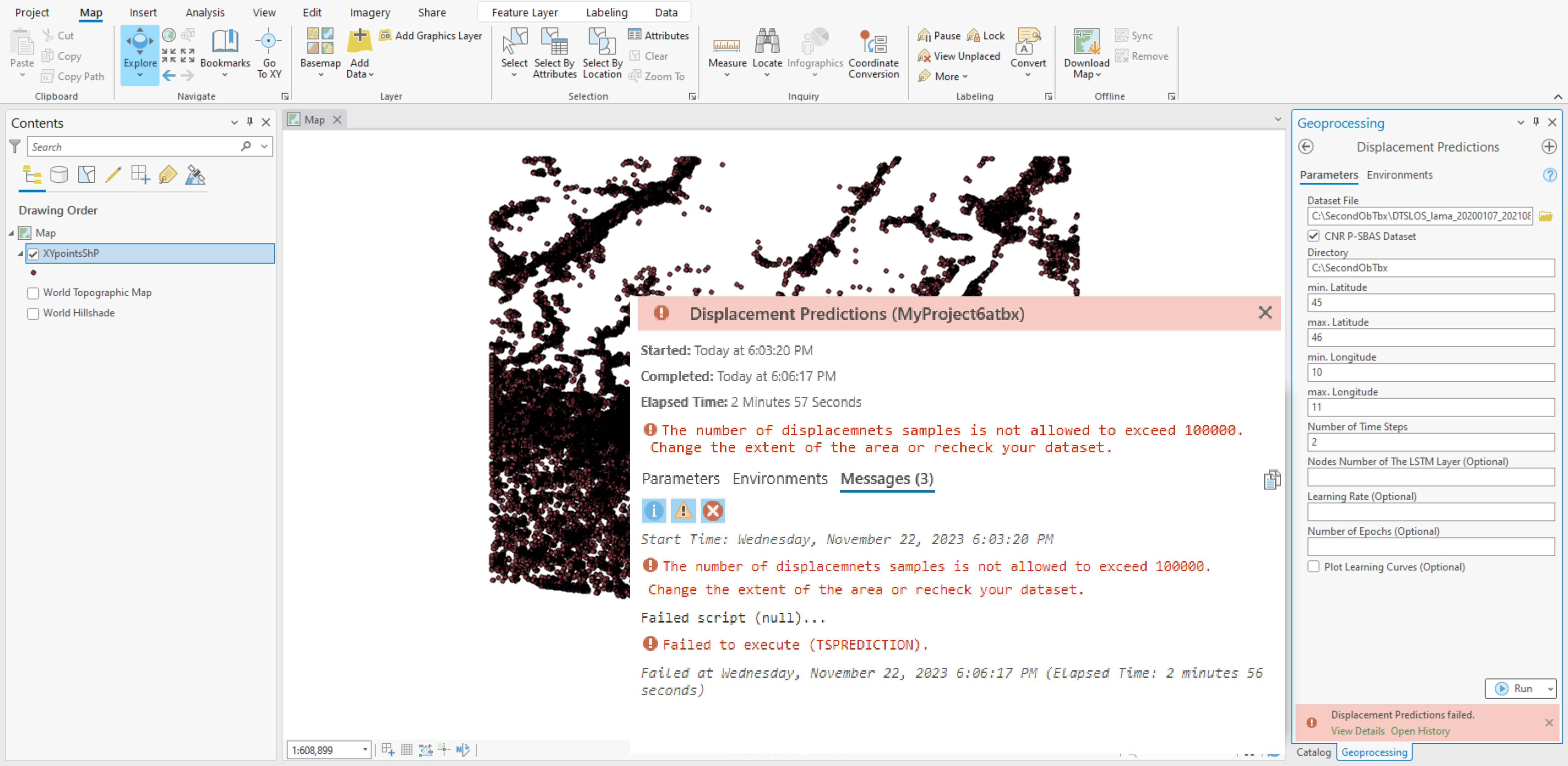


Toolbox Improper Execution Error Messages



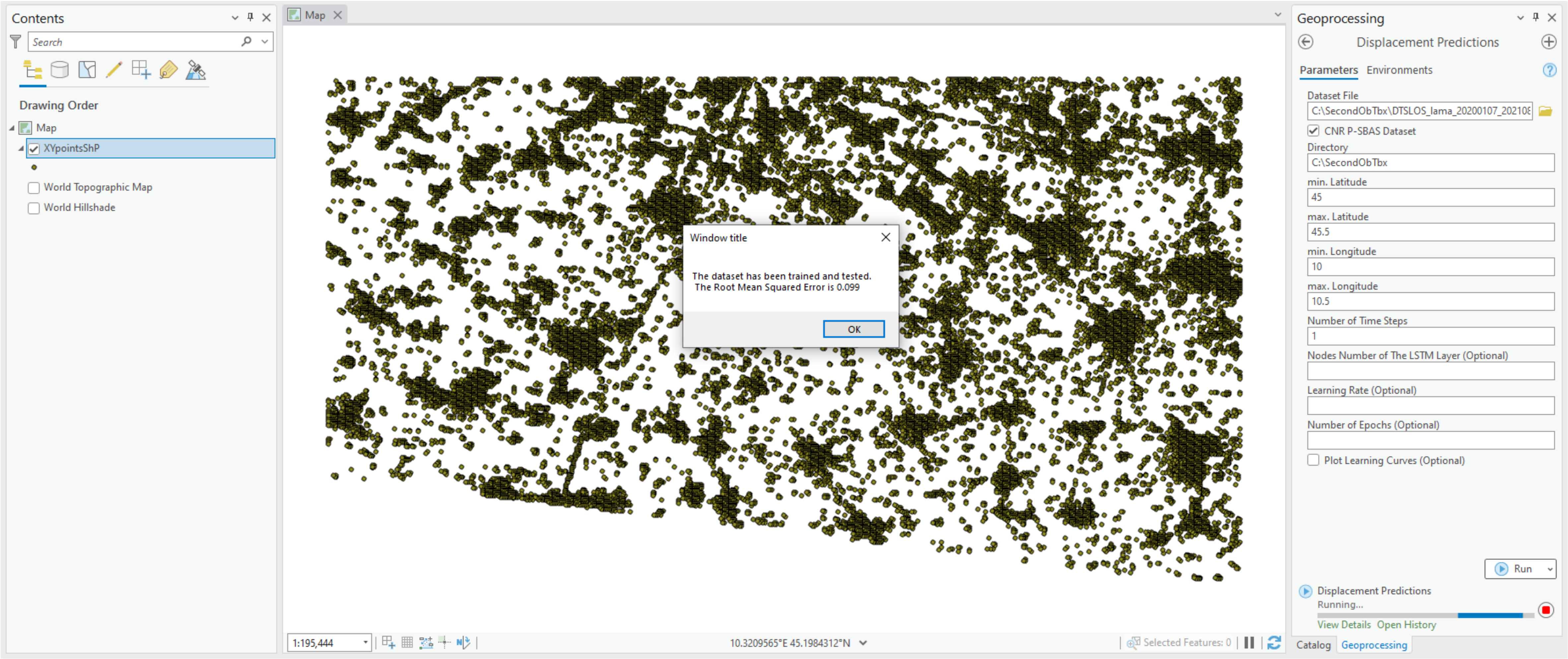
Università degli Studi di Padova

Toolbox Improper Execution Error Messages

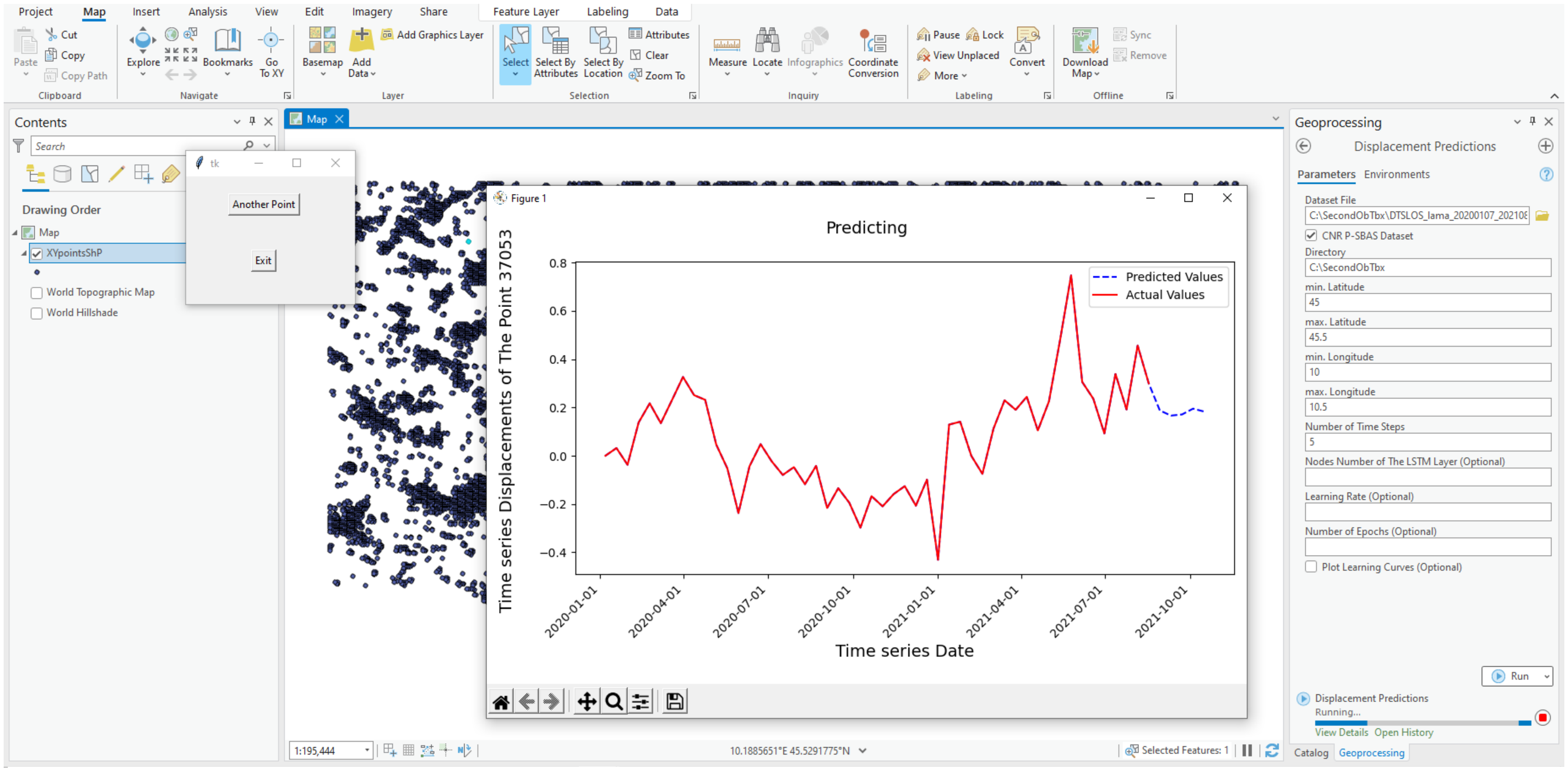


Università degli Studi di Padova

Results of Running The Toolbox (First Part)



Results of Running The Toolbox (Second Part)



Università degli Studi di Padova

Conclusions

1st Obj.

- The methodology developed in this research produced meaningful datasets to identify ground displacements using machine learning, with the Cosine K-Nearest Neighbor algorithm demonstrating superior performance, especially in adjacent areas
- The application of pseudo-labeling notably improved validation accuracy, marking a significant advancement in geospatial analysis techniques
- The trained models worked consistently across three different geographical datasets, although further validation is needed
- Testing the Persistent Scatterer Interferometry technique using this workflow for further evaluation is recommended

2nd Obj.

- LSTM models are proficient in forecasting single or multiple steps in InSAR displacement time series, particularly for regular sequences.
- For irregular time series, employing advanced LSTM models that are sensitive to time, such as the Time Gated LSTM, is advisable despite the increased computational costs involved.

3rd Obj.

- The fusion of AI and InSAR within the GIS framework remains in its initial stages, and the developed toolbox has underscored the significance of such an interdisciplinary amalgamation

References

[1] Crosetto, M., Castillo, M., & Arbiol, R. (2003). Urban subsidence monitoring using radar interferometry. Photogrammetric Engineering & remote sensing, 69(7), 775-783.

[2] He, L., Wu, L., Liu, S., Wang, Z., Su, C., & Liu, S. N. (2015). Mapping two-dimensional deformation field time-series of large slope by coupling DInSAR-SBAS with MAI-SBAS. Remote Sensing, 7(9), 12440-12458.

Thanks for the attention

UNIVERSITÀ DEGLI STUDI DI PADOVA