



# Predicting land deformation by integrating InSAR data and cone penetration testing through machine learning techniques

Melika Sajadian<sup>1</sup>, Ana Teixeira<sup>2</sup>, Faraz S. Tehrani<sup>2,3</sup>, and Mathias Lemmens<sup>1</sup>

<sup>1</sup>Faculty of Architecture and the Built Environment, TU Delft, Delft, the Netherlands

<sup>2</sup>Deltares, Delft, the Netherlands

<sup>3</sup>Faculty of Civil Engineering and Geosciences, TU Delft, Delft, the Netherlands

**Correspondence:** Ana Teixeira (ana.martinsteixeira@deltares.nl)

Published: 22 April 2020

**Abstract.** Built environments developed on compressible soils are susceptible to land deformation. The spatio-temporal monitoring and analysis of these deformations are necessary for sustainable development of cities. Techniques such as Interferometric Synthetic Aperture Radar (InSAR) or predictions based on soil mechanics using in situ characterization, such as Cone Penetration Testing (CPT) can be used for assessing such land deformations. Despite the combined advantages of these two methods, the relationship between them has not yet been investigated. Therefore, the major objective of this study is to reconcile InSAR measurements and CPT measurements using machine learning techniques in an attempt to better predict land deformation.

## 1 Introduction

Built environments developed on unconsolidated and/or organic sediments are susceptible to land deformation due to the weight of buildings, roads and fluctuation of ground water level (Kempfert and Gebreselassie, 2006; Peduto et al., 2016). Hence, the spatial and temporal monitoring and analysis of ground deformation is necessary for the sustainable development of cities.

More specifically, in case of roads, the deformation causes failure in serviceability and performance of the infrastructure and induces high maintenance and repair costs (Peduto et al., 2016; Du et al., 2018). Unevenly deformed roads are dangerous, damaging and inconvenient for both the vehicles and passengers (Wijeyesekera et al., 2016). Furthermore, the partial closures of the transportation networks during maintenance periods have adverse socioeconomic impacts. For these reasons, predicting and continuous monitoring of the ground deformation along infrastructure networks is of significant importance for improving the network resilience (Peduto et al., 2016; North et al., 2017).

For monitoring the rate of land deformation, advanced Interferometric Synthetic Aperture Radar techniques such as

Differential InSAR (DInSAR) can be used. SAR data currently has sufficient temporal resolution and by applying DInSAR techniques, land deformation can be monitored on the order of millimeters (SkyGeo, 2018). However, there are always gaps in the final deformation results due to occlusion and coherence loss in SAR imagery.

The potential of Cone Penetration Testing (CPT) for estimating land deformation has been extensively studied in the Geotechnical Engineering community (Koster et al., 2018a, b; Verruijt and Van Baars, 2007; Kempfert and Gebreselassie, 2006). The CPT measurements provide quantitative information about the characteristics of the soil layers including the compressibility. However, CPT-based methods in estimating land deformation suffer primarily from empiricism and spatial-temporal discontinuity.

Despite the application of these two methods in estimating ground deformation, the direct relationship between the CPT measurements and the rate of deformation acquired from DInSAR has not yet been investigated. Therefore, the major objective of this study is to reconcile DInSAR measurements and CT measurements using Machine Learning (ML) techniques to better predict land deformation. In Sect. 2, the proposed methodology for solving this problem is explained.

In Sect. 3, the proposed methodology is applied on an example case study of a road in the Netherlands. Finally, in Sect. 4, the final conclusions are stated and followed by the acknowledgements.

## 2 Methodology

The overall methodology consists of four main steps – see Fig. 1 (Sajadian, 2019). The first two steps are mainly concerned with gathering and pre-processing of the datasets, in which the relevant parameters for the next steps are extracted. In the third step, the correlations and similarities are investigated. In the fourth step, we use Machine Learning to define the relationship between soil properties, loading/unloading history and the linear rate of deformation.

### 2.1 Steps 1 and 2: Data Gathering and Extracting Parameters

The relevant parameters from CPT measurements such as depth, cone resistance  $q_c$ , sleeve friction  $f_s$ , friction ratio  $R_f$ , etc. and soil types are extracted. These parameters are considered as soil properties in this research. The Z coordinate of the CPT indicates the elevation of the terrain before construction of the road. The current elevation of the road is derived from the Digital Elevation Model (DEM), which is extracted from the LiDAR point cloud of the highway. Assuming a uniform thickness of 90 cm for the surface, base, sub-base and sub-grade of the road (based on the road construction standards), the difference between the current (DEM) and old elevation (CPT) indicates the amount of removed or added stress due to excavation or backfilling. The SAR images are processed by combining a sequence of radar images (Terrasar-X with the spatial resolution of  $3.00\text{ m} \times 2.80\text{ m}$  and revisit period of 11 d) from 2016 till 2019 to measure the ground deformation using D-InSAR techniques. The main product is the time series representing the amount of deformation with respect to the first acquisition. Each of the time series is decomposed to a linear trend over the 3-year period and a seasonal pattern using a least squares linear regression model. For each CPT measurement, the nearest InSAR measurement (within a distance of less than 5 m) is extracted as the deformation time series corresponding to that CPT point.

### 2.2 Step 3: Correlations and Similarities between Soil Properties, Loading/Unloading and Deformation

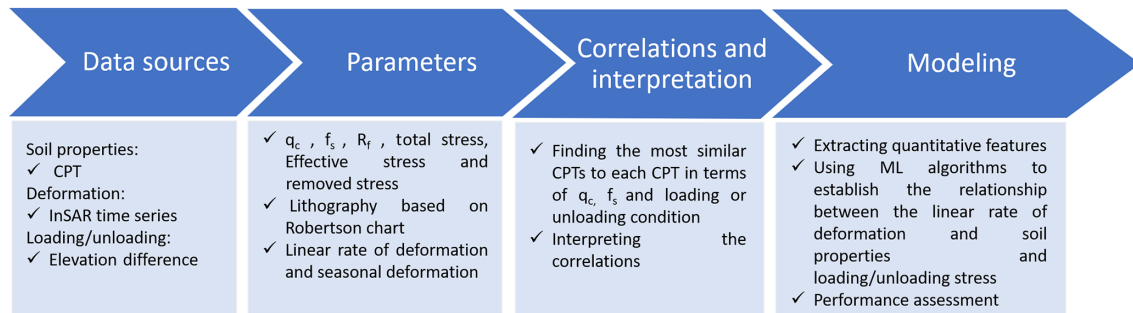
In this step, the similarities and correlations between soil properties, loading/unloading history and the resulting deformation are being studied. Serra and Arcos (2014) presents number of similarity measures for clustering and classification of time series. The  $q_c$  and  $f_s$  profiles are series of measurements in depth and can be treated as time series. Hence, we can use the similarity measures discussed in Serra and Arcos (2014) to measure the similarities of  $q_c$  and  $f_s$  profiles

on the road. In this research, the hypothesis is that if two CPT measurements are similar in terms of both  $q_c$  and  $f_s$  profiles and the loading history is the same, the deformation behavior should be the same. Here, we used the simplest similarity measure, i.e. the Euclidean (Serra and Arcos, 2014) distance between the time series, which is computationally efficient and suitable for comparison of samples that are at exactly the same depth location. By our definition, two CPT are considered similar if the sum of normalized distances of their  $q_c$  and  $f_s$  is less than 0.2 and the difference between their loading/unloading stress is less than 10 kPa (these thresholds are based on expert's knowledge and trial and error). If the aforementioned hypothesis is correct, the deformation rate of a reference point in the dataset should be more or less the same as the mean of the linear rates of deformation of the similar points (with similar CPT profiles). The coefficient of determination between the deformation rate of the reference point and the similar points is regarded as a measure that describes the degree that the deformation rate can be taken as a function of soil properties (CPT measurements) and loading/unloading stress.

### 2.3 Step 4: Feature Extraction and Modeling Using Machine Learning

In this step, first we extract quantitative descriptors from CPT profiles. (Coerts, 1996) lists the possible quantitative features and their interpretation and shortcomings for CPT segments. Ultimately, he introduces a set of the most suitable and interpretable descriptors for CPT measurements, which we use in our case study: the Interquantile range (IQR), Indicator of simple trend (T), Indicator of convexity or concavity (C), Normalized number of fluctuations around the median (R) and Sharpness of upper boundary (B). The quantitative features are extracted from the CPT profiles to the depth of 15 m under the ground surface. The choice of 15 m is due to the good trade-off between having the maximum possible depth and not losing too many CPT measurements shallower than that depth, as well as the fact that peat and clay layers are mostly present above this depth. The loading/unloading stress is another feature. The goal is to establish the relationship between the these features and the linear rate of deformation, which is predicted. In the research work of Sajadian (2019) a qualitative as well as a qualitative Machine Learning (ML) are shown and compared.

There are multiple ML algorithms one could use (Breiman, 2001). However, there are no previous studies for our case study, so the choice of the proper set of features is unknown. The ML algorithm should provide information about the significance of each of the features and the established model through the ML algorithm should be interpretable. Taking this into account we selected tree-based algorithms, which satisfy all these conditions, which are Gradient-Boosting and Random Forest (Hastie et al., 2005).



**Figure 1.** Overall methodology applied in this paper.

### 3 Case Study and Results

The required data for modeling land deformation significantly depends on the case study. In this research, the newly constructed part of A4 highway connecting Delft to Schiedam (the Netherlands) is studied ([https://www.wegenwiki.nl/A4\\_\(Nederland\)](https://www.wegenwiki.nl/A4_(Nederland)), last access: October 2019). For the sake of brevity, we will only present the results of the last part of step 4 of the methodology. For the complete study readers are referred to Sajadian (2019).

The soil properties and loading/unloading history influence the rate of deformation in different directions. The complicated interactions between the driving mechanisms suggests that the relationship is definitely not linear to be recognized by the simple correlations. This led us to using ML to model the relationship between these datasets. The location under study is about 5 km long and has 368 CPT's for which deformation measurement points are available.

Figure 2 shows the quantitative features extracted, at different depths, for one CPT. These features are the descriptors of the segments of every 5 m. In this research, rather than being interested in importance of each of these descriptors in estimating the target value, we are more interested in investigating which of the profiles and which depth of measurement is more significant in estimating the linear rate of deformation.

For the A4 highway case study, and using the features mentioned above, the two ML algorithms are tested on the dataset: Gradient-Boosting and Random Forest. Eighty percent of the measurements are taken for training each of the algorithms and 20 % of the measurements are used to validate the results. Table 1 summarizes the performance metrics for each method.

As shown in Table 1, the outcomes of the two ML algorithms are very similar and the histograms show a very similar distribution. The estimated deformation rates are between  $-1$  and  $4 \text{ mm yr}^{-1}$ , meanwhile the measured values are between  $-7$  and  $8 \text{ mm yr}^{-1}$  – see Fig. 3. This means that both algorithms fail to detect subsiding patterns and extreme heaving patterns. The errors (mostly between  $-3$  and  $3 \text{ mm yr}^{-1}$ ) could be explained by the fact that the data set is imbalanced

(15 % subsiding and 85 % heaving), the total number of data points is not enough and/or the features selected are not representative of the observed deformation.

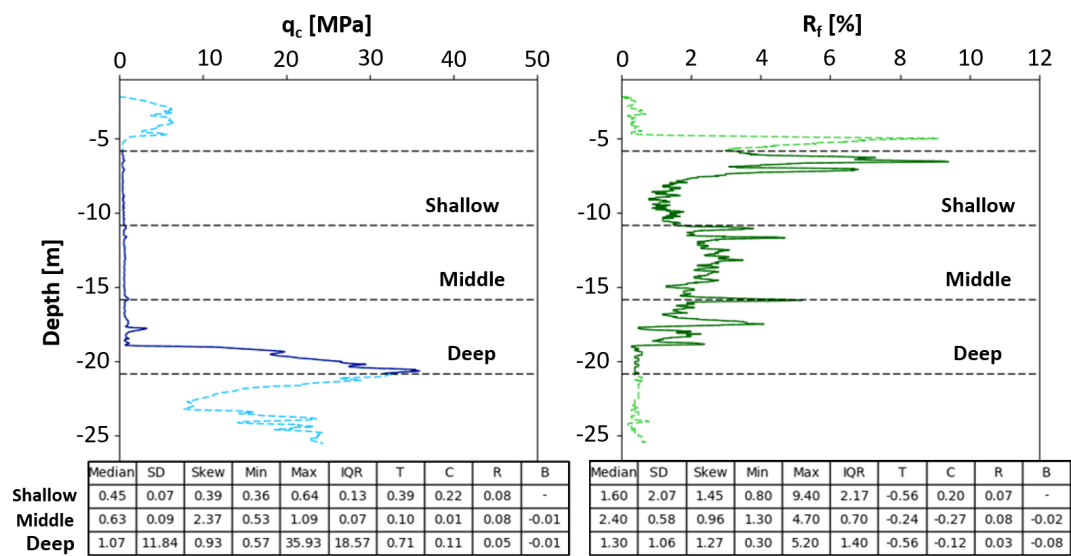
Nevertheless, both ML algorithms give consistent results in terms of generalization performance and feature importance with negligible differences. When we look at the significance of each feature on predicting the target deformation rate, both algorithms showed that features extracted from CPT's  $q_c$  are the most dominant ones. After this, the loading/unloading stresses are also important.

### 4 Conclusions

In this research, the main focus was studying and modeling the deformation on roads due to loading/unloading and based on soil characterization using ML algorithms. The desired output of the research was an ML model trained by standard data that enabled the prediction of surface movements of roads susceptible to soil deformation. The case study was the newly constructed part of A4 highway (Delft-Schiedam) in the Netherlands.

It was concluded, for this case study, that:

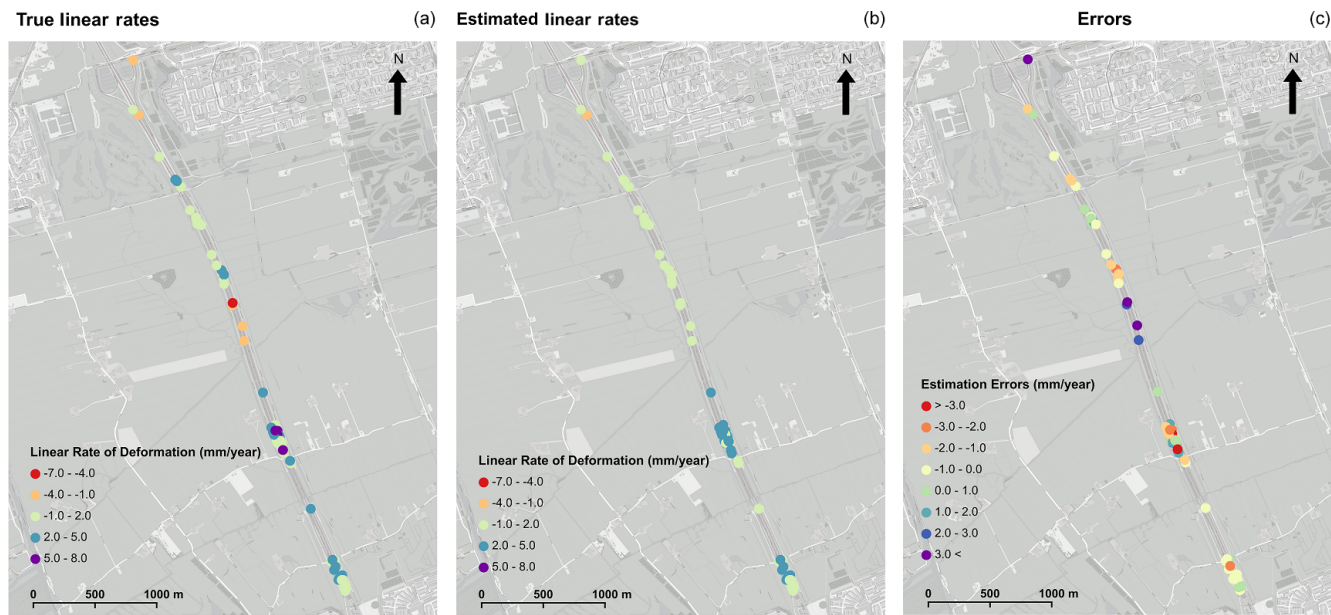
- the available data sources on soil data do not provide all the necessary information, e.g. information on presence of certain expansive minerals or information about ground water conditions in the soil are missing. Here, only the latest (simplified) loading/unloading step was estimated while information on the previous stress conditions (which were discontinuous in time) were missing;
- the InSAR measurements only provided information about the first three years after the construction of the road and, therefore, the information on the amount of deformation was limited to this time span;
- and finally, although this case study showed diverse deformation behavior (both heave and subsidence), which made it interesting for investigating more influential parameters of road deformation, it presented another limitation. i.e. the diversity of the behavior in this case study



**Figure 2.** Example of the quantitative features extracted from one CPT, per depth level: shallow, middle and deep.

**Table 1.** Performance metrics “Average over 10-folds” and “Best Performing Model” of Gradient Boosting and Random Forest ML algorithms with quantitative features.

Performance metrics:	Gradient Boosting		Random Forest	
	Averaged	Best	Averaged	Best
Mean Absolute Error ( $\text{mm yr}^{-1}$ )	1.1	1.1	1.2	1.2
Mean Squared Error ( $\text{mm yr}^{-1}$ )	2.4	2.2	2.6	2.5
Root Mean Squared Error ( $\text{mm}^2 \text{yr}^{-2}$ )	1.6	1.5	1.6	1.6
$R^2$	0.5	0.5	0.4	0.5



**Figure 3.** Deformation rates in  $\text{mm yr}^{-1}$  for the case study. (a) The true rates of deformation. (b) The estimated rates of deformation of Gradient Boosting model with quantitative features. (c) The error of estimated rates.



was due to the special construction history and therefore it could not be easily generalized to the other roads.

In this study, the trained ML algorithm, rather than presenting a general relationship between InSAR data and CPT data, helped to investigate the effectiveness of the gathered data in explaining the studied phenomena. The results this research study could be improved by adding more data points and features as well as more accurate boundary conditions. This could help to explain the diverse deformation pattern of A4 highway.

**Data availability.** The data that support the findings of this study are available on request from the corresponding author.

**Author contributions.** MS designed the computational framework and analysed the data. AT, FST and ML contributed to the design, implementation and supervision of the research. All authors contributed to the writing of the manuscript.

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

**Special issue statement.** This article is part of the special issue “TISOLS: the Tenth International Symposium On Land Subsidence – living with subsidence”. It is a result of the Tenth International Symposium on Land Subsidence, Delft, the Netherlands, 17–21 May 2021.

**Acknowledgements.** The work presented is part of the MSc joint-research study between TU Delft and Deltares. The discussions and contributions of Ramon Hanssen and SkyGeo are gratefully acknowledged.

## References

- Breiman, L.: Random forests, *Mach. Learn.*, 45, 5–32, 2001.
- Coerts, A.: Analysis of static cone penetration test data for subsurface modelling: a methodology, Koninklijk Nederlands Aardrijkskundig Genootschap, the Netherlands, 1996.
- Du, Z., Ge, L., Ng, A. H.-M., Zhu, Q., Yang, X., and Li, L.: Correlating the subsidence pattern and land use in Bandung, Indonesia with both Sentinel-1/2 and ALOS-2 satellite images, *Int. J. Appl. Earth Obs.*, 67, 54–68, <https://doi.org/10.1016/j.jag.2018.01.001>, 2018.
- Hastie, T., Tibshirani, R., Friedman, J., and Franklin, J.: The elements of statistical learning: data mining, inference and prediction, *Math. Intell.*, 27, 83–85, 2005.
- Kempfert, D. H.-G. and Gebreselassie, D. B.: Excavations and Foundations in Soft Soils, Springer Science & Business Media, Springer, Berlin, Heidelberg, <https://doi.org/10.1007/3-540-32895-5>, 2006.
- Koster, K., De Lange, G., Harting, R., de Heer, E., and Middelkoop, H.: Characterizing void ratio and compressibility of Holocene peat with CPT for assessing coastal-deltaic subsidence, *Q. J. Eng. Geol. Hydrogeol.*, 51, 210–218, <https://doi.org/10.1144/qjegh2017-120>, 2018a.
- Koster, K., Stafleu, J., and Stouthamer, E.: Differential subsidence in the urbanised coastal-deltaic plain of the Netherlands, *Neth. J. Geosci.*, 1–13, 2018b.
- North, M., Farewell, T., Hallett, S., and Bertelle, A.: Monitoring the response of roads and railways to seasonal soil movement with persistent scatterers interferometry over six UK sites, *Remote Sensing*, 9, 922, <https://doi.org/10.3390/rs9090922>, 2017.
- Peduto, D., Huber, M., Speranza, G., van Ruijven, J., and Cascini, L.: DInSAR data assimilation for settlement prediction: case study of a railway embankment in the Netherlands, *Can. Geotech. J.*, 54, 502–517, 2016.
- Sajadian, M.: Spatial and Temporal Analysis of Road Deformation based on Remote Sensing and Subsurface Exploration, MSc thesis, TU Delft, available at: <https://repository.tudelft.nl/islandora/object/uuid%3A9e836468-7291-4279-aa08-7d723ef11327> (last access: 11 March 2020) 2019.
- Serra, J. and Arcos, J. L.: An empirical evaluation of similarity measures for time series classification, *Knowl.-Based Syst.*, 67, 305–314, 2014.
- SkyGeo: Technical background SkyGeo InSAR, Tech. rep., available at: <https://www.skygeo.com/insar-technical-background/> (last access: 11 March 2020), 2018.
- Verruijt, A. and Van Baars, S.: Soil mechanics, VSSD, Delft, the Netherlands, 19–25, 2007.
- Wijeyesekera, D. C., Numbikannu, L., Ismail, T., and Bakar, I.: Mitigating Settlement of Structures founded on Peat, in: IOP Conference Series: Materials Science and Engineering, Vol. 136, Soil Engineering International Conference 2015 (SEIC2015) 27–29 October 2015, Langkawi, Malaysia, 12042, IOP Publishing, <https://doi.org/10.1088/1757-899X/136/1/012042>, 2016.