Hybrid Artificial Intelligence to model land subsidence

Koster, K.[1], Molhoek, M.[2], de Bakker, M.P.[2], Candela, T.[1], Esteves Martins, J.[1], Fokker, P.A.[1], Verberne, M.[1,3], van Laanen, J.[2,4], Trantas, T.[2], and Soustelle, V.[1]

1 TNO, Energy Transition
2 TNO, Information and Communication Technology
3 Utrecht University, Department of Earth Sciences
4 TU Eindhoven, Department of Mathematics and Computer Science
Contact: kay.koster@tno.nl

Rationale

Land subsidence is hazardous as it inflicts damage and increases flood risks. In recent years in the Netherlands, it has gained substantial attention from governmental bodies, industry, and researchers. This momentum resulted in the process of maturation of subsidence as a research topic alongside with an increase in the amount of available data and physical modelling output. This process however, also revealed limitations in the state-of-the-art of subsidence modelling. For instance, often applied physics-based models to describe below-ground subsidence processes are not always able to explain above-ground observations. An example is the slow-paced shallow subsidence processes in soft soils by increased vertical effective stress as a result of groundwater level lowering (Verberne, 2021). In addition, disentangling of different superimposed land subsidence processes in areas subjected to various subsurface interventions from observations is presently challenging (Candela and Koster, 2022).

Using Artificial Intelligence (AI) to understand system behaviour is well established in many scientific disciplines, especially where physics-based models cannot describe processes based on available heterogeneous big datasets. The field of land subsidence is a prime example where AI techniques could aid physical predictive modelling, given the availability of big data sets of various quality and spatiotemporal resolution on different aspects of the below and above-ground domains.

In this contribution, we explore applications of different AI techniques within the field of land subsidence and present preliminary results. We focus on the coastal plain of Friesland in the Netherlands (Fig. 1), where shallow phreatic groundwater level lowering and deep gas extraction results in superimposed land subsidence processes. Furthermore, extensive, different and big datasets relevant to subsidence analysis have been made available for this area: GeoTOP, a 3D geological model of the upper 50 m of the subsurface (resolution 100x100x0.5m, cells attributed with lithology and stratigraphic unit) (TNO, 2022), a phreatic groundwater level change model (resolution 100x100m, averaged per month) (Dabekaussen et al., 2020), geodetic data consisting of levelling, GNSS and InSAR data (Esteves Martins et al., in prep), and gas reservoir pressure changes data (Candela et al., in prep).

Method

Our approach regards different Hybrid AI techniques, meaning here the connection between physics-based modelling and advanced deep learning. More specific, the presented techniques fall within the categories of Federated Learning (FL) and Explainable AI (XAI).

FL jointly trains a machine learning model while keeping the training data separate at different locations. Gas reservoir pressure change data from hydrocarbon operators is often kept confidential and, therefore, difficult to obtain by researchers. FL has as an advantage that it can train models while keeping confidential reservoir pressure data at the premises of hydrocarbon operators. This way, data leakage is prevented and thus predictive land subsidence can be run securely (cf. Van Laanen, 2022).

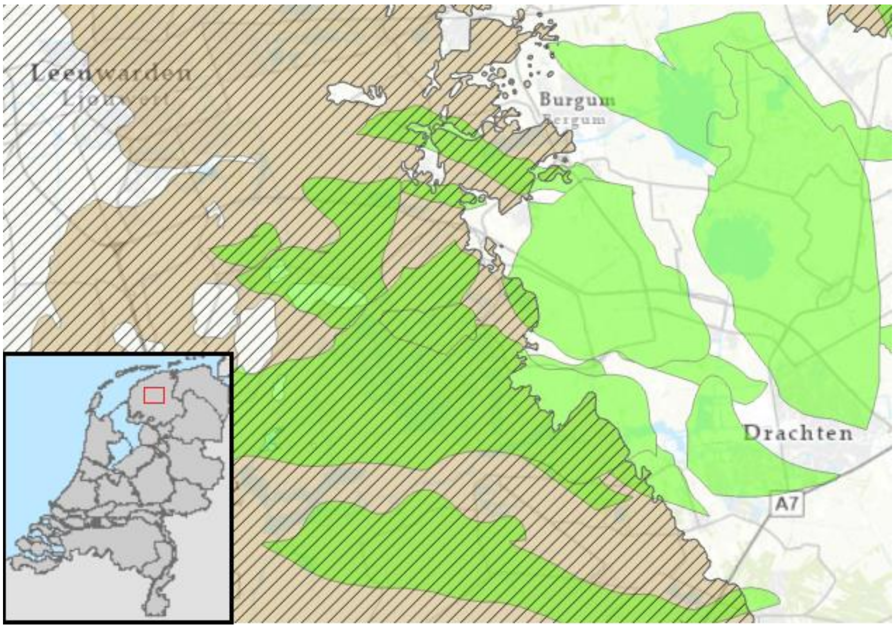

Figure 1 Topographical map of the Netherlands showing the area of interest (red square). The zoomed-in map shows for the area of interest the spatial distribution of gas fields (green) (NLOG, 2022), and the inland extent of the coastal plain accommodating soft soil deposits (hatched), including surficial peat beds (brown) (TNO, 2022).

XAI regards the use of for humans understandable machine learning approaches, without relying on non-transparent ‘black-box’ methods. Insight in the causes of land subsidence modelling are critical for governance and industry to implement mitigation strategies. This requires modelling outcomes to be interpretable by humans. Generally, AI leans on black-box concepts, with often difficult interpretable outcomes. XAI therefore is vital in communicating and implementing modeling results.

FL is applied on vertically partitioned data mimicking 2 nodes. The library PySyft is selected, combining the secure inner sect (Multi-Party Computation) technique with the Split Neural Networks (NN) method, averaging the outputs of the split trained models with max pooling. The machine learning technique Neural Network is selected for flexibility in algorithms including physics informed NN. Subsidence predictions are performed over the area of interest including different NN splits. Actual subsidence in the area is compared with the performed subsidence predictions.

Counterfactual explanations are used as a means to make the XAI approach transparent. They are calculated based on a trained machine learning model. Counterfactual explanations provide insights in the prediction of a machine learning by calculating which features need to change by how much in order to change the predicted outcome. As such, counterfactual explanations can be used to explain the causes of certain predictions generated by ‘black-box’ algorithms. Here, a model of Mothilal et al. (2019) with associated Python library is used to generate counterfactuals and feature importance on our data.

Synthetic data

The extensive different datasets are used to create synthetic data that describes total observed subsidence using different below-ground subsidence processes at different depths. By using synthetic data, complete control is obtained on the behaviour of the developed algorithms, which is critical to fine-tune modelling processes. The synthetic subsidence data is generated by different well established physical models describing shallow (Fokker et al., 2019) and deep (Candela et al., 2022) subsidence, with the addition of noise, and interpolated towards the locations of the real geodetic measurements.

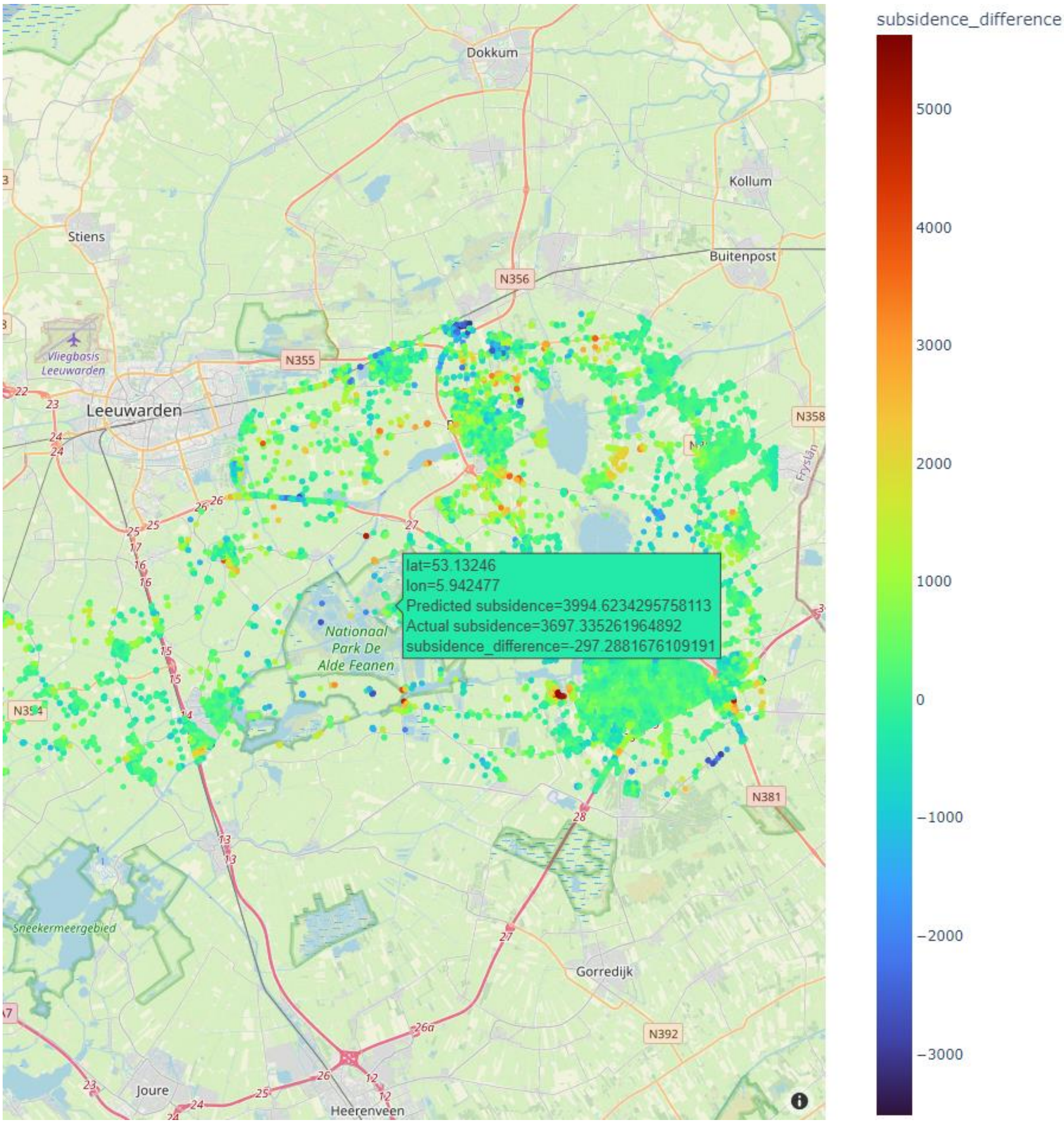

Figure 2 Results of the FL prediction of total subsidence (mm) by deep and shallow processed based on synthetic data. A datapoint describing actual (synthetic) and predicted subsidence in a federated setting is selected indicating minimal differences.

Preliminary results

The FL results shows an R2: 0.83 (MSE of 0.0046, MAE of 0.042) between predicted and actual (synthetic) subsidence (Fig. 2). The predictions in federated settings are equal to those in non-federated settings, indicating that training the data while being at different servers provides comparable results as training the data while being at the same server. Furthermore, the best results are obtained in areas with both shallow and deep subsidence processes, i.e. on top of the gas fields. The latter is most likely because there, the subsidence rates are the highest, resulting in a relatively favourable subsidence signal-to-noise ratio. These preliminary results are promising in view of training subsidence models using confidential (hydrocarbon) data in a secure way.

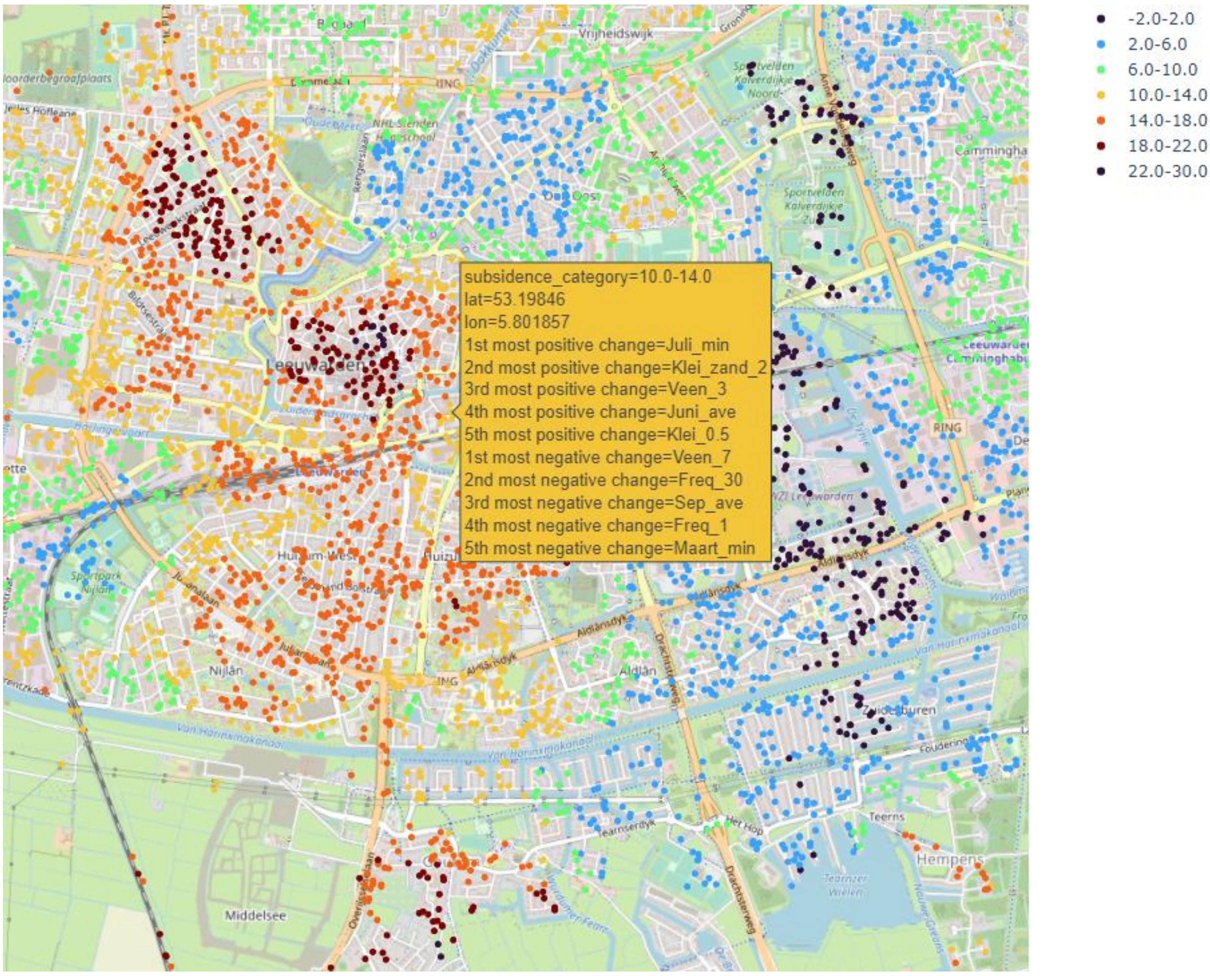

Figure 3   XAI results for the city centre of Leeuwarden. The legend indicates synthetic cumulative subsidence (mm) for a period of five years on the locations on geodetic data points. A datapoint with the class 10-14 cm is selected. The popped-up table shows 2x five features that practice most influence when changed (positively or negatively) on the 10-14 cm class. Note that the table is in Dutch (juli = July, juni = June, sep = September, maart = March, veen = peat, klei = clay, zand = sand).

The XAI results produced counterfactual explanations for 75 features, e.g. soil types at different depths and groundwater levels per month. For now, this accounts all features available in the dataset, including those we expect to barely affect subsidence (such as the features that comprise 'sand' at different depths). Nevertheless, we are able to identify features per point that have most influence on the synthetically generated total subsidence class. For example, in Figure 3, the desired total subsidence for the selected point is the class 10 – 14 cm. The counterfactual explanation revealed that this class is best reached for this point by obtaining higher values of the lowest water level in July ('juli_min' in the table), or lower values of peat at 7 m below the surface ('veen_7' in the table). In this example, especially the July groundwater levels result is potentially useful for groundwater management strategies to mitigate subsidence, as this can be physically adjusted by humans.

Look forward

In this contribution, FL combined with XAI is used to explain land subsidence in a single point in a secure way with the aid of (federated) synthetic data generation and Multi-Party Computation in a federated learning setting. Future steps would focus on i) generalizing the secure counterfactual explanation over a larger area to improve explaining the relative contributions to subsidence of shallow and deep processes and the reduce of the number of features, and ii) generating a diverse set of candidate synthetic counter factual explanations to create a robust explanation. Furthermore, synthetic data should be substituted by the real available datasets to finalize the workflow.

References

Candela, T. and Koster, K. (2022). The many faces of anthropogenic subsidence. Science 376, 1381-1382

Candela, T., Chitu, A., Peters, E., Pluymaekers, M., Hegen, D., Koster, K., and Fokker, P. (2022). Subsidence induced by gas extraction: a data assimilation framework to constrain the driving rock compaction process at depth. Frontiers in Earth Sciences 10, 1-18

Dabekaussen, W., Lourens, A., and W.-J. Zaadnoordijk (2020). Simulation of groundwater heads for the Think or Sink project. TNO report, unpublished, pp. 12

Fokker, P., Gunnink, J., Koster, K., and De Lange, G. (2019). Disentangling and parameterizing shallow sources of subsidence: application to a reclaimed coastal area, Flevoland, the Netherlands. Journal of Geophysical Research: Earth Surface 124, 1099-1117

Mothilal, R.K., Sharma, A., and Tan, C. (2019). Explaining machine learning classifiers through diverse counterfactual explanations. Proc. Con. of Fairness, Accountability, and Transparency, 607-617

NLOG (2022). Online portal for subsurface energy and resources information in the Netherlands: www.nlog.nl

TNO (2022). Online portal of the Geological Survey of the Netherlands for subsurface information: www.dinoloket.nl

Van Laanen, J.C.A. (2022). Secure Multi-Party Counterfactual Explanations in Vertical Federated Learning. MSc Thesis Technical University Eindhoven, Eindhoven, pp. 89

Verberne, M. (2021). Towards an improved methodology for data assimilation to disentangle the processes leading to subsidence. A case study on the methodology for an area in North-Friesland, the Netherlands. MSc Thesis Utrecht University, Utrecht, pp. 59