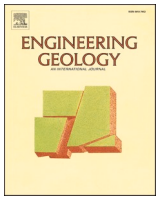
A metamodel for estimating time-dependent groundwater-induced subsidence at large scales

Engineering Geology 341 (2024) 107705

Contents lists available at ScienceDirect

Engineering Geology

journal homepage: www.elsevier.com/locate/enggeo



A metamodel for estimating time-dependent groundwater-induced subsidence at large scales

Ezra Haaf [a,*], Pierre Wikby [a], Ayman Abed [a], Jonas Sundell [a,b], Eric McGivney [c], Lars Rosén [a], Minna Karstunen [a]

[a] Department of Architecture and Civil Engineering, Chalmers University of Technology, Sweden
[b] Swedish Geotechnical Institute, Sweden
[c] Qasa AB, Sweden

ARTICLE INFO

Keywords:
Regional subsidence
Metamodeling
Machine learning

ABSTRACT

Construction of large underground infrastructure facilities routinely leads to leakage of groundwater and reduction of pore water pressures, causing time-dependent deformation of overburden soft soil. Coupled hydro-geomechanical numerical models can provide estimates of subsidence, caused by the complex time-dependent processes of creep and consolidation, thereby increasing our understanding of when and where deformations will arise and at what magnitude. However, such hydro-mechanical models are computationally expensive and generally not feasible at larger scales, where decisions are made on design and mitigation. Therefore, a computationally efficient Machine Learning-based metamodel is implemented, which emulates 2D finite element scenario-based simulations of ground deformations with the advanced Creep-SCLAY-1S-model. The metamodel employs decision tree-based ensemble learners random forest (RF) and extreme gradient boosting (XGB), with spatially explicit hydrostratigraphic data as features. In a case study in Central Gothenburg, Sweden, the metamodel shows high predictive skill (Pearson's r of 0.9–0.98) on 25 % of unseen data and good agreement with the numerical model on unseen cross-sections. Through interpretable Machine Learning, Shapley analysis provides insights into the workings of the metamodel, which alignes with process understanding. The approach provides a novel tool for efficient, scenario-based decision support on large scales based on an advanced soil model emulated by a physically plausible metamodel.

1. Introduction

During construction and operation of underground infrastructure, leakage of groundwater into these structures can cause a reduction of groundwater heads and result in subsequent reduction of pore water pressure. In areas with deformation-sensitive clays, this pore pressure reduction is delayed, leading to time-dependent ground deformations. Such deformations have the potential to cause significant damage to buildings and other infrastructure as groundwater head decline may occur at larger scales (district-, city-, or regional scale), where the impact area is dependent on the hydraulic and stratigraphic features of the soil (e.g., Burbey, 2002; Chaussard et al., 2014; Guzy and Malinowska, 2020; Huang et al., 2012; Larsson et al., 1997; Zhu et al., 2012).

When planners of underground infrastructure assess subsidence risks, physical process-based numerical models are generally applied that combine groundwater flow with deformation analysis. With such models the complex problems of subsurface variations (e.g., hydro-stratigraphic, and geotechnical properties) across spatial scales, as well as the time-dependent nature of the coupled hydro-geomechanical processes can be addressed. For large-scale subsidence predictions, where hydro-stratigraphic variations are represented, Finite Difference (FD) methods are commonly used, such as MODFLOW coupled with the MODFLOW SUB package (Harbaugh, 2005; Hoffmann et al., 2003) or Finite Element (FE) methods such as FEFLOW (Trefry and Muffels, 2007). Other models (both FD and FE methods) that have been applied on large scales combine groundwater flow and linear elastic or elasto-plastic stress-strain relationship (Calderhead et al., 2011; Mahmoudpour et al., 2016; Ochoa-González et al., 2018; Sundell et al., 2019a; Teatini et al., 2006; Ye et al., 2016). Such applications have been demonstrated for many subsidence problems with acceptable accuracy in relation to historic piezometric and subsidence measurements (Galloway and Burbey, 2011).

* Corresponding author.
E-mail address: ezra@chalmers.se (E. Haaf).

https://doi.org/10.1016/j.enggeo.2024.107705
Received 21 February 2024; Received in revised form 19 August 2024; Accepted 2 September 2024
Available online 10 September 2024

However, in cases where highly deformable sensitive clays with large thicknesses dominate, an advanced constitutive (soil stress-strain-strength) relationship is needed due to the highly non-linear material response. Creep, anisotropy of stress-strain response as well as degradation of bonding are features which affect the compressibility of natural sensitive clays. Therefore, a rate-dependent model, which accounts for these different features, such as Creep-SCLAY1S (Gras et al., 2017b; Sivasithamparam et al., 2015), is required to accurately represent the response soft sensitive clays. To the authors' knowledge, no applications of advanced constitutive models on large scale subsidence problems currently exist. Presumably, the main reason for this is the associated computational complexity, which is compounded by the need for the high number of elements in the numerical models at large scales, short time steps due to the high non-linearity of the in constitutive models, the many iterations required for convergence, as well as small enough element sizes for acceptable precision, mesh quality and the reduction of smearing effects. In addition to that, applying coupled formulations in advanced constitutive models is numerically complex.

When a physical process cannot be simulated with feasible computing time, surrogate models or metamodels are increasingly used (e.g., Fienen et al., 2018; Furtney et al., 2022; Kang et al., 2016). A metamodel is an approximate mathematical model of the outcome that is more computationally efficient than a more detailed numerical model (e.g., Asher et al., 2015). Metamodels generally only emulate outcomes and not processes, which means that models can at best only be as good as the underlying numerical model and will generally only be valid for the range of inputs over which the metamodel was trained on (Furtney et al., 2022). However, metamodels can capture complex, nonlinear behavior and be efficiently applied in larger scale assessments, given that statistical stationarity holds between the training model and the metamodel (e.g., Starn et al., 2021). To be able to emulate outcomes that have complex, nonlinear relations to input variables, different underlying algorithms of metamodels have been used in geotechnics and hydrogeology, such as traditional statistical methods, e.g., polynomial approaches using response surfaces (Obel et al., 2020; Zhang, 2020) or polynomial chaos expansion (Zoccarato et al., 2021) and Machine Learning and Artificial Neural Networks (Zhang, 2020). More commonly, varying types of regression models are employed, including Machine Learning models (Fienen et al., 2018; Furtney et al., 2022; Kang et al., 2016). Furtney et al. (2022) show that metamodels can be more than 99 % accurate and argue that they are generally underused in soil and rock mechanics.

Here, we propose an approach for constructing a Machine Learning-based metamodel for subsidence due to a pore-water pressure drop under a soft sensitive clay layer at a large scale. The metamodel emulates the outcome, i.e., the magnitude of subsidence de at a given time, of the Creep-SCLAY1S constitutive model, which is run as a coupled hydro-geomechanical FE model in Plaxis 2D. The modeling strategy is applied to a case study in Central Gothenburg, Sweden. Here, the construction of a commuter train tunnel and station necessitates investigating multiple scenarios of pore pressure drops in the clay layer. This is crucial to determine acceptable levels of groundwater drawdowns over time and space to minimize damages with timely mitigation measures.

2. Methodology

2.1. Approach

A three-stage strategy is proposed to predict pore pressure drawdown induced ground deformation on large scales (Fig. 1). In the preliminary stage, the required input data is collected over the entire domain of interest, including hydro-mechanical properties, borehole logs, and groundwater levels. The borehole logs and groundwater levels are processed into a hydrostratigraphic model, with thicknesses and depths of soil layers as well as a piezometric map (see Section 3.2). In the process modeling phase I, a scenario of a pore-water pressure drop (groundwater drawdown) is chosen, and a section of the domain (here, a 2D cross section is proposed) is extracted from the hydrostratigraphy. Further, representative soil data is prepared for process modeling (see Section 3.3). These cross-sections are then implemented in a numerical model, and ground deformations (here, subsidence) resulting from the drawdown are computed along the section. In phase II, a statistical learning framework is used to train a machine learning-based metamodel on the subsidence results of the simulated cross sections using features such as hydrostratigraphy that are available for the entire domain (Section 2.2). Finally, the learned relationships from the cross-section are used to spatially predict subsidence on the entire model domain.

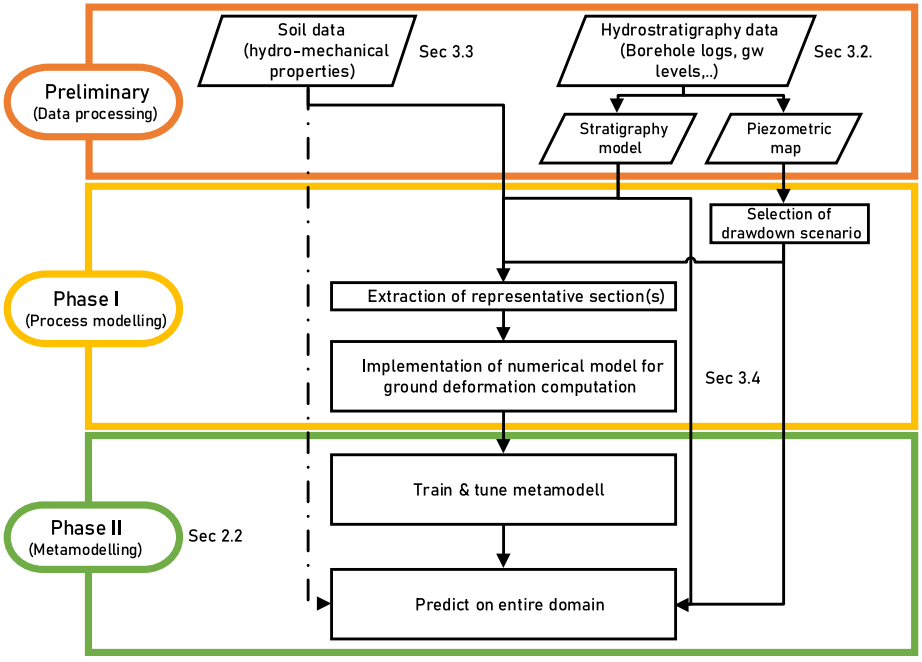


Fig. 1. Flowchart of the proposed metamodeling strategy.

2.2. Machine learning-based metamodel

The relationship between soil properties, hydrostratigraphy, and ground deformation under external stresses is complex, which requires a metamodel that supports description of nonlinearity. While in principle independent of any physical process modeling, for this task we examined the use of regression by comparing two commonly used decision tree-based ensemble machine learning for metamodel development, namely extreme gradient boosting (XGB) (Friedman, 2001) and random forests (RF) (Breiman, 2001). Both techniques have previously been applied for complex metamodeling (Bjerre et al., 2022; Starn et al., 2021) to model various environmental prediction problems involving nonlinearity and interaction, such as in both studies of groundwater quantity and quality (e.g., Haaf et al., 2023; Hou and Lu, 2018; Meray et al., 2024) and geotechnics (Wang et al., 2021; Bharti et al., 2021; Wang et al., 2020; Zhou et al., 2017; Zhu et al., 2021). Several authors have shown that both XGB and RF outperform other machine learning algorithms and neural networks in empirical modeling of subsidence, with a slight advantage of XGB due to higher prediction accuracy (e.g., Kim et al., 2022; Zhang et al., 2021). When building regression trees, decision trees are trained additively to a certain depth (i.e., the number of nodes of the tree). Splitting of decision trees at nodes into new branches is determined through an objective function. During the training phase, all potential training features are tested at each split to improve the prediction of the target variable. In XGB regression trees are further "boosted" by training each tree on the residuals of the previous tree sequentially. The RF algorithm on the other hand, builds multiple decision trees in parallel, after which the mean prediction of the decision trees is returned.

Here, the RF and XGB algorithms are used independently to predict subsidence (target variable: the vertical displacement u_y (cm)) based on a data set containing training features, which in the proposed approach are proxy variables derived from the hydrostratigraphic model (example features in Table 1) along the cross-section. To improve model performance, RF and XGB require model tuning, which is the process of selecting the hyperparameters of the model that control the learning of the model and overfitting. To limit underfitting (model too simple, not using all information available) or overfitting (model too complex, fitting noise), a cross-validation and tuning strategy for training is carried out.

Two cross-validation techniques were employed in this study:

(i) Random Hold-Out (RHO): In RHO, the entire dataset, which includes all available cross-sections, is randomly divided into two sets – a training set (75 % of the data) for model training and an evaluation set (25 % of the data) for assessing the performance of the model.

(ii) Leave-One-Cross-Section-Out Cross-Validation (LOCSO): For LOCSO, one of the available cross-sections is held out entirely for evaluation purposes, while model training is performed using the remaining cross-sections.

Table 1
Proxy variables for 2D cross sections, derivable from hydrostratigraphy and used as training features for the metamodel. Compare Fig. 2b.

Training variables (short name)	Description
t_soil	Thickness of sediment (Depth to bedrock)
t_lyr0	Thickness of layer of filling material (upper aquifer)
t_lyr1	Thickness of layer of soft soil
t_lyr2	Thickness of layer of coarse material (lower aquifer)
gw_upp	Depth of groundwater head in upper aquifer
gw_low	Depth of groundwater head in lower aquifer
OCR_i	Thickness of zone with uniform over-consolidation ratio (OCR), i zone identifier

While RHO is a conventional approach, LOCSO is applied in this study to further demonstrate the quality of prediction of an entire cross section not trained by the metamodel. The exact setup of the cross-validation in the case study is shown in Section 3.5.

Before model fitting, hyperparameter tuning is performed using a latin hypercube grid search (to reduce computational cost) based on Root Mean Squared Error (RMSE) as an objective function. Here, we used the *R* implementation of *XGBoost* and *ranger* for random forest generation (Wright and Ziegler, 2017), with *dials* for hyperparameter tuning (Kuhn and Frick, 2022) and *tidymodels* for setting up the model workflow (Kuhn and Wickham, 2020). In the proposed metamodel, prediction on cross sections (and the entire domain) is carried out pointwise based on information at each grid point.

To understand the plausibility of the metamodel, both the importance of features and their contributions over the value range of the features to the predictions of subsidence were estimated from Shapley Additive Explanations (SHAP) values (Lundberg et al., 2020). SHAP values can be used to understand the relationship between the model features and the model output in statistical-learning models by estimating the contributions of each feature to each prediction (for subsidence u_y, this is in units of cm). The method is based on the Shapley value, which assigns a value to each model feature based on its contribution to the overall output. Here, we show the relationship between each model feature and subsidence by evaluating i), the mean absolute SHAP value across all observations to investigate feature importance for the entire model, ii), feature importance along the selected cross-section, and iii), locally visualizing the value ranges of individual features versus their SHAP values using the implementation *kernelshap* (Mayer and Watson, 2023).

3. Case study, data sets and models

3.1. Problem and site description

The method is applied to a case in Gothenburg, a city of about 600,000 inhabitants located in Southwestern Sweden, where a commuter train tunnel with multiple underground stations is being built. The construction is expected to cause leakage of groundwater into the subsurface construction causing pore pressure reduction in the overburden soft soils over widespread areas, typically at district scale. This reduction, in turn, may result in time-dependent ground deformations, potentially damaging buildings, and utilities. Mitigation of such damages necessitates the formulation of damage scenarios covering extensive areas, enabling planners to implement timely safety measures, such as the installation of sealing in the tunnel or artificial infiltration wells to sustain pore pressures. To be able to deliver such damage scenarios, a metamodel is constructed for an area in the city center of about $1.5 \times 1.5\ km^2$ to evaluate different drawdown scenarios at different time scales. More specifically, subsidence is estimated for scenarios, representing a major or a minor leakage of groundwater into the tunnel. From experience, such leakages may result in pore pressure reduction below the clay of between 10 kPa and 40 kPa. Here persistent drawdown responses for 1 year and 30 years were chosen to allow investigation of short-term and long-term responses to minor and major groundwater drawdown, resulting in four different scenarios (1y,10 kPa; 1y,40 kPa; 30y,10 kPa, 30y,40 kPa).

The topography of the site varies substantially in the study area with a steep decline from a higher area in the southeast around 70 m.s.l. to the canal (Rosenlund canal) at sea level. Then, elevation rises north of the canal before reaching the low-lying areas along the Göta River (Fig. 2a). The crystalline, predominantly metamorphic, bedrock on which Gothenburg rests is part of the Southwestern Swedish gneiss province and contains water-bearing fractures with mainly west-northwest-southeast strike. In the high areas in the south, the bedrock crops out, while it is covered with soils of varying thickness in the lower areas around the watercourses and in the valleys. In general, above the

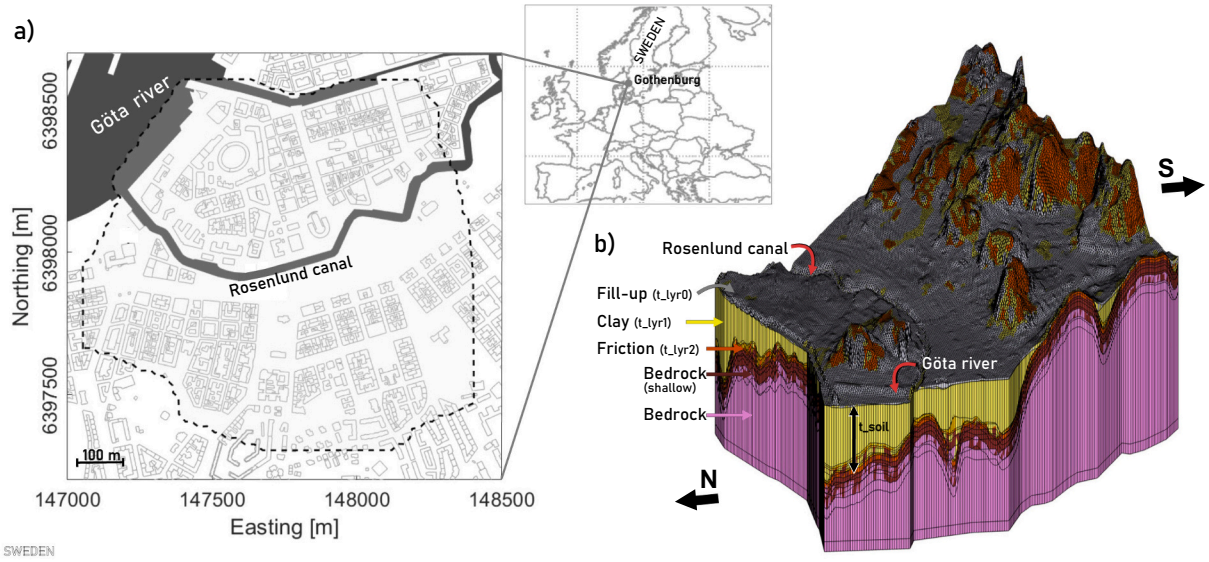


Fig. 2. a) City map of downtown Gothenburg, Sweden. Study area indicated by dashed line. b) Stratigraphic model of study area (vertical scale ten times exaggerated).

bedrock, is a layer of coarse-grained soils, usually glacial till, and glacial outwash, overlain by glacial and post-glacial marine clay, topped by a mix of post-glacial beach deposits (mainly sands) and heterogenous anthropogenic fill. The fill has a thickness of about 1 to 3 m, but locally at the Rosenlund canal, may reach a thickness of 5 to 7 m. The thickness of the clay varies considerably, with the largest depths of clay, ranging between 40 and up to 100 m at Rosenlund/Haga. Three types of aquifers can thus be found at the study site, a fractured rock aquifer, which is partially connected to a confined aquifer in the coarse grained soil located beneath the glacial clay layer, as well as an unconfined aquifer in the post-glacial sand and fill materials (Agrell, 1979).

3.2. Geostatistical hydro-stratigraphic model

The stratigraphy is modeled using a comprehensive borehole data set (ca 28.000 logs) with a geostatistical modeling procedure previously presented in Sundell et al. (2016) and Sundell et al. (2017). The procedure simplifies the soil stratification into three continuous layers as described above: coarse-grained fill (top-most), soft clay, and coarse-grained material (glacial till and glacio-fluvial deposits) above crystalline bedrock (Fig. 2b). The model follows a stepwise procedure to take advantage of all available information in the borehole data set. The procedure results in grids of bedrock level, thickness of the non-cohesive material, soft clay, and fill-up at 5-m resolution. The piezometric map of groundwater heads is modeled with a steady-state MODFLOW-NWT model (Niswonger et al., 2011) following the methodology described by Sundell et al. (2019b). The model geometry is based on modeled stratigraphy and calibrated on long-term averages of measured groundwater levels in the confined aquifer.

3.3. Soil properties

The soft clays in the study area were deposited relatively recently, during (glacial) and after (post-glacial) the last glacial period (Weichselian glaciation). Therefore, these natural clays are typically only slightly over-consolidated with relatively high creep rates. This effect has been further increased by urbanization since the foundation of Gothenburg in 1621, including fill materials piled over the soft clay, as well as the weight of the built environment, resulting in heterogeneous soil conditions and background creep. Consequently, the stress history across the study area is also heterogeneous, making it challenging to retrieve a representative stiffness. Herein, a stress-dependent stiffness is assumed, with the stiffness and rate of creep, controlled by the vertical over-consolidation ratio (*OCR*).

The OCR values were initially retrieved from oedometer tests performed on samples extracted via ST-II piston tubes, see Fig. 3b. However, due to poor sample quality at depths larger than 25 m at the Haga site, the values had to be adjusted (Fig. 3c; Fig. S1, supplementary information). Therefore, for larger depths, incremental/depth-discrete green-field displacements measured from 2011 and 2018 retrieved from a bellow hose, see Fig. 3a, were used to tune the background creep predicted by the model. A bellow hose is an instrument which measures the incremental absolute settlements with depth (Andersson et al., 2015) similarly to en extensiometer. By assuming that the measurements were purely vertical, and that the ongoing settlements were natural (creep), as well as representative of the model domain, back-calculations with a Finite Element model (see Section 3.4) were made to retrieve a more representative $OCR = 1.5$ at those depths. The large, measured displacements at the surface layers can be attributed to excessive retraction (during installation, the hose is stretched and eventually attached to the soil while in stretched state) due to the lack of horizontal soil stresses acting on the bellow hose. However, the ground displacement at this point fits well with InSAR data at roughly the same location (Wikby et al., 2023). Hence, the chosen final trend of OCR (Fig. 3c) is based on the bellow hose in the deeper clay layers (notice the large difference between the measured *OCR* and the trend), the oedometer test results in the shallower layers, and InSAR measurements at surface level. A table of the layering and its respective *OCR* values can be found in the supplementary information (Table S2). The *OCR* profile yields a relatively low pre-overburden pressure (POP) at the shallower layers, however, increasing drastically after 30 m depth.

Fig. 3c shows the index properties over depth. Bulk density, ρ, is shown to increase from around 1.6 t/m^3 at ground level to around 1.8 t/m^3 at 50 m depth, whereas the natural water content, w, decreases from 80 % at the ground level to 40 % at 50 m depth. The sensitivity, S_t, like w, decreases from around 20 to 10 with increased depth. The hydraulic conductivity (vertical), k, was evaluated from Constant rate of strain (CRS) oedometer tests, decreasing from 1e-9 m/s to 0.5e-9 m/s with increased depth. Finally, derived *OCR* values are shown to decrease from around 1.5 at the top to 1.1 at 20–23 m, following an increase to 1.5 at depths above 40 m. Apart from soil properties at the Haga station, parameter values from the nearby, more well-characterized Göta tunnel

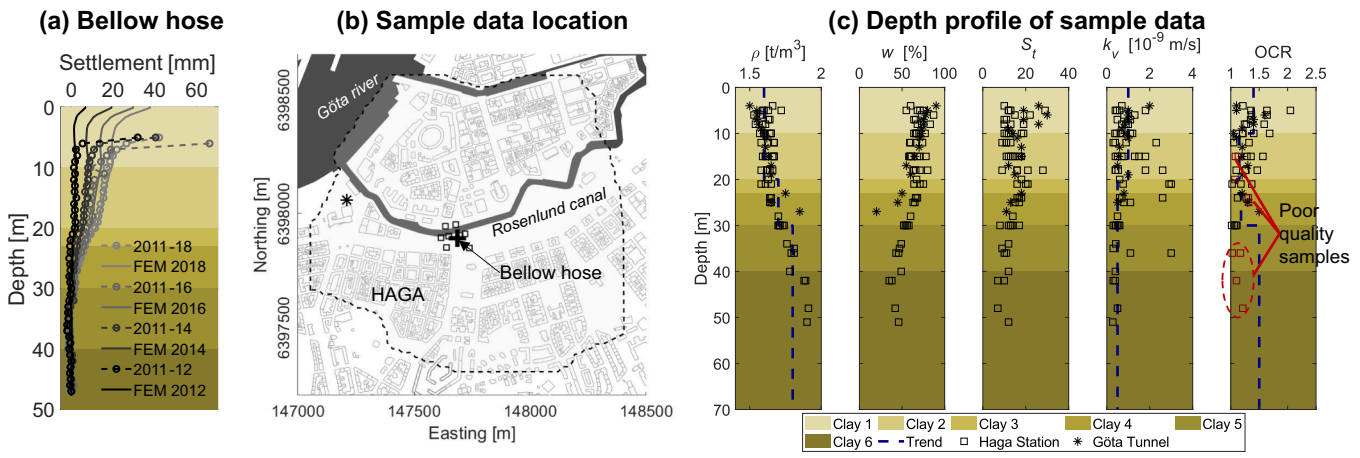


Fig. 3. a) Accumulated settlements between 2011 and 2018: Bellow hose measurements vs. back-calculated FEM results. b) Location of model domain, Haga station and Göta tunnel samples, as well as the bellow hose. c) Index properties comparison between this study's (Haga station) and Göta tunnel's (Tornborg et al., 2021) data. The trend line shows the selected value for each clay sublayer (shaded areas).

project (Tornborg et al., 2021) were used. Due to the presence of till at shallower depths at the Göta Tunnel site, deviations can be seen between the sites at 20-30 m depths due to higher density and OCR, as well as lower natural water content. Apart from that, the index properties are similar enough, such that model parameters from Tornborg et al. (2021) could be used. No hydraulic anisotropy was assumed, since it is not an influential factor of the specific problem, as shown by Wikby et al. (2023).

3.4. Hydro-mechanical modeling

The clay response was simulated using an in-house implementation of Creep-SCLAY1S constitutive model (Gras et al., 2017a, 2017b; Karstunen et al., 2005; Sivasithamparam et al., 2015; Wheeler et al., 2003). In addition to creep, the rate-dependent model accounts for initial and evolutionary anisotropy and bonding. Creep-SCLAY1S has been successfully benchmarked against oedometer and triaxial tests on Scandinavian clays (Sivasithamparam et al., 2015), as well as embankments (Amavasai et al., 2017; Amavasai et al., 2018), and deep excavations at field scale (Tornborg et al., 2021; Bozkurt et al., 2023) thus proving its representativeness for Gothenburg clay. Three model reference surfaces define the constitutive relationship, the Intrinsic Compression Surface (ICS), the Current Stress Surface (CSS) and the Normal Compression Surface (NCS) (Fig. S2, supplementary information). Creep-SCLAY1S calculates strain rates through an elastic term and a visco-plastic (creep) term. The volumetric (Eq. (1)) and deviatoric (Eq. (2)) creep strain rates are calculated as shown below:

$$\dot{\varepsilon}_p^c = \dot{\Lambda}\left(\frac{\partial p'_{eq}}{\partial p'}\right) \tag{1}$$

$$\dot{\varepsilon}_q^c = \dot{\Lambda}\left(\frac{\partial p'_{eq}}{\partial q}\right) \tag{2}$$

$$\dot{\Lambda} = \frac{\mu_i^*}{\tau}\left(\frac{(1+\chi)p'_{m,i}}{p'_{eq}}\right)^{-\left(\frac{\lambda_i^*-\kappa^*}{\mu_i^*}\right)}\frac{M_c^2-\alpha_0^2}{M_c^2-\eta_0^2} \tag{3}$$

where p'_{eq} is the equivalent mean effective stress defining the size of CSS, μ_i^* is the intrinsic modified creep index, τ is the reference time, χ is the initial amount of bonding (linking to the sensitivity), $p'_{m,i}$ is the intrinsic preconsolidation pressure defining the size of ICS, λ_i^* is the intrinsic modified compression index, κ^* is the modified swelling index. M_c is the stress ratio at critical state in triaxial compression, α_0 and η_0 are the initial rotation of the reference surfaces (representing initial anisotropy) and the initial stress ratio, respectively, at normally consolidated state. The second term in Eq. (3), is an inverse of OCR, thus when *OCR* (which is the vertical component) is fed into the model, it calculates the initial size of the NCS which differentiates large creep strains from small creep strains.

The fill and the till, above and below the clay layer, respectively, were both modeled using the Mohr-Coulomb material model, whilst the bedrock was simulated using a linear elastic model. The model parameters for all layers can be found in Table S2 (supplementary information).

Simulations were run using the finite element software PLAXIS 2D (version 21) with full hydro-mechanical coupling. Three cross-sections (AA, BB and CC) were chosen for the coupled hydromechanical analyses, with location and meshes shown in Fig. 4. The hydro-stratigraphic levels are based on the results of the geostatistical hydro-stratigraphic model (Section 3.2). The mesh fineness was chosen after carrying out a sensitivity analysis to ensure mesh-independent results, by rerunning the simulation and adjusting the mesh density from lower to higher. Mesh-independence was found when the average element size was 4 m^2 in the clay. Cross-sections AA, BB, and CC contain 22,384, 19,181, and 21,506 6-noded elements, respectively, with three stress integration points per element. The displacements were fixed in the horizontal direction at the boundary of the bedrock.

A global groundwater level (GW_{upper}) was retrieved from the groundwater flow model for the top clay/fill set as the hydraulic head of the upper aquifer. For the friction material/till and bedrock, a local groundwater level was set as the hydraulic head of the lower aquifer. For initial conditions in the clay layer modeled with six sub-layers, a hydraulic interpolation function was used to interpolate the head in the clay layer between the upper and lower aquifer heads. Thereafter, a relatively instantaneous drawdown ($GW_{lower,0}$ - > $GW_{lower,1}$) and pore water pressure (PWP) drop was simulated in the lower aquifer (underdrainage), followed by delayed PWP drop in the clay starting at the bottom, continuing upwards until full consolidation (PWP_{final}) is reached, as predicted by the coupled hydromechanical analyses. A scenario of 10 kPa pressure drop (1 m drawdown) is shown in Fig. 5 for times t1 and t2 where t2 > t1 > 0.

The underdrainage was created by using phases in PLAXIS. Firstly, an initial phase was run using a linear elastic material to calculate the in situ effective stresses. This was followed by a NIL step where the linear elastic material was replaced with Creep-SCLAY1S for the clay (to initialize the state parameters of the model correctly) and with Mohr-Coulomb for the till and fill layers, and the strains were reset to zero (model parameters can be found in Table S1 in the supplementary

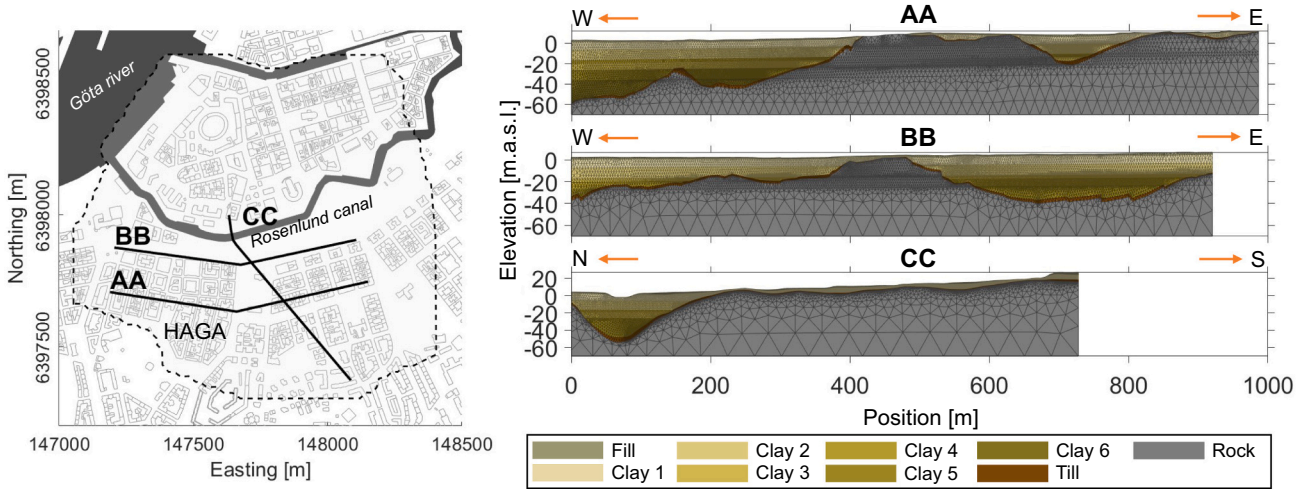


Fig. 4. Cross-section locations in the metamodel domain (left) and cross-sections with respective layering and mesh. The depression in the soil at x ~ 70 m in profile CC is the bottom of the Rosenlund canal.

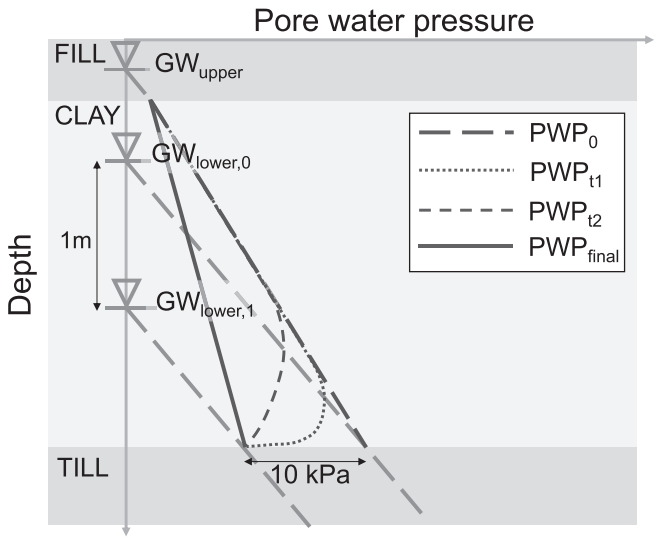


Fig. 5. Excess pore pressure dissipation process of a typical underdrainage problem in clays bounded by an upper and a lower aquifer (modified from Wikby et al. (2023)).

information). The first two steps are thus only needed to simulate a computationally sound initial stress state and the initialization of the state variables for Creep-SCLAY1S. Following the NIL step, the actual drawdown could be simulated in a plastic calculation phase. Finally, two consolidation phases with full hydro-mechanical coupling were simulated representing periods since the drawdown occurred (1 year and 30 years), emulating extreme cases in greenfield conditions if no measures are taken.

Two drawdown scenarios were chosen, a small pressure drop of 10 kPa, and a more severe scenario of 40 kPa, resulting in a total of four combinations of scenarios: 1y,10 kPa, 1y,40 kPa, 30y,10 kPa and 30y,40 kPa. Additionally, a zero-drawdown scenario (0 kPa) was set up with equally many consolidation phases to calibrate the OCR values (see Section 3.3). This scenario also enables to determin the contribution from pure background creep for each drawdown scenario.

3.5. Metamodel evaluation setup

First, the training data set was created by extracting the subsidence values computed with the coupled hydro-geomechanical model at positions along the cross-section at corresponding grid points of the hydrostratigraphic model. At the same locations, features were then extracted from the hydrostratigraphic model and combined with the computed subsidence values for each scenario and cross-section. To comprehensively test the performance of the metamodel, twelve different combinations of training and evaluation data were created. The setups shown in Table 2 are designed to evaluate the model on all combinations of cross sections with both RHO and LOCSO. For example, evaluation number 1 in Table 2 is trained on 75 % of data available in cross sections AA and BB. For RHO, evaluation of the relationship between observed and simulated subsidence with Pearson's correlation coefficient (R) and Root Mean Square Error (RMSE) against the numerical model is carried out on the remaining 25 % of the data in cross-sections AA and BB. For LOCSO, the model trained on the 75 % training data is used to predict the entire cross section CC, based on the available features of CC. All further evaluation numbers are the remaining combinations of model scenarios and cross sections. Each evaluation setup is modeled with both Machine Learning techniques, resulting in a total of 12 trained metamodels for each RF and XGB. As described above, a trained metamodel is evaluated twice (on RHO and LOCSO), leading to 48 evaluations (12 RF/RHO, 12 RF/LOCSO, 12 XGB/RHO, 12 XGB/LOCSO). In addition, a comparison of average InSAR displacement data for the period 2022–2024 with a metamodel trained on creep on the entire domain was carried out, to compare metamodel performance for areas of the study domain that were not represented by cross sections (Supplementary material S6).

4. Results & discussion

4.1. Hydro-geomechanical model

Fig. 6 shows settlement results from all cross-sections. For all scenario combinations, the largest settlements occur in the shallower layers in the transitional zones between soft clay and frictional layer/ rock. This is most likely due to the relatively low pre-overburden pressures at the bottom boundary of the clay layer, resulting in low stiffness, as well as faster rates of excess pore water pressure dissipation, due to smaller drainage lengths (in 2D).

Creep is more significant in the thicker clay layers for the 30 years consolidation cases (Fig. S3, supplementary information), which reflects the bellow hose results in terms of which layers settle the most (0-30 m), ca. 15 cm after 30 years (5 mm/year). This is very similar to values found in InSAR measurements in the area.

For positions 200–730 m in cross-section CC, it is especially

Table 2
Evaluation setups for different drawdown scenarios, training, and evaluation strategies.

Evaluation number	Scenario name	Time [y]	Drawdown [kPA]	Training sections	Total data	Split [%]	Evaluation			
							RHO		LOCSO	
							Section	#Data	Section	#Data
1	1y,10 kPa	1	10	AA+BB	567	75/25	AA+BB	190	CC	291
2	1y,10 kPa	1	10	AA+CC	510	75/25	AA+CC	171	BB	368
3	1y,10 kPa	1	10	BB+CC	492	75/25	BB+CC	164	AA	393
4	1y,40 kPa	1	40	AA+BB	567	75/25	AA+BB	190	CC	291
5	1y,40 kPa	1	40	AA+CC	510	75/25	AA+CC	171	BB	368
6	1y,40 kPa	1	40	BB+CC	492	75/25	BB+CC	164	AA	393
7	30y,10 kPa	30	10	AA+BB	567	75/25	AA+BB	190	CC	291
8	30y,10 kPa	30	10	AA+CC	510	75/25	AA+CC	171	BB	368
9	30y,10 kPa	30	10	BB+CC	492	75/25	BB+CC	164	AA	393
10	30y,40 kPa	30	40	AA+BB	567	75/25	AA+BB	190	CC	291
11	30y,40 kPa	30	40	AA+CC	510	75/25	AA+CC	171	BB	368
12	30y,40 kPa	30	40	BB+CC	492	75/25	BB+CC	164	AA	393

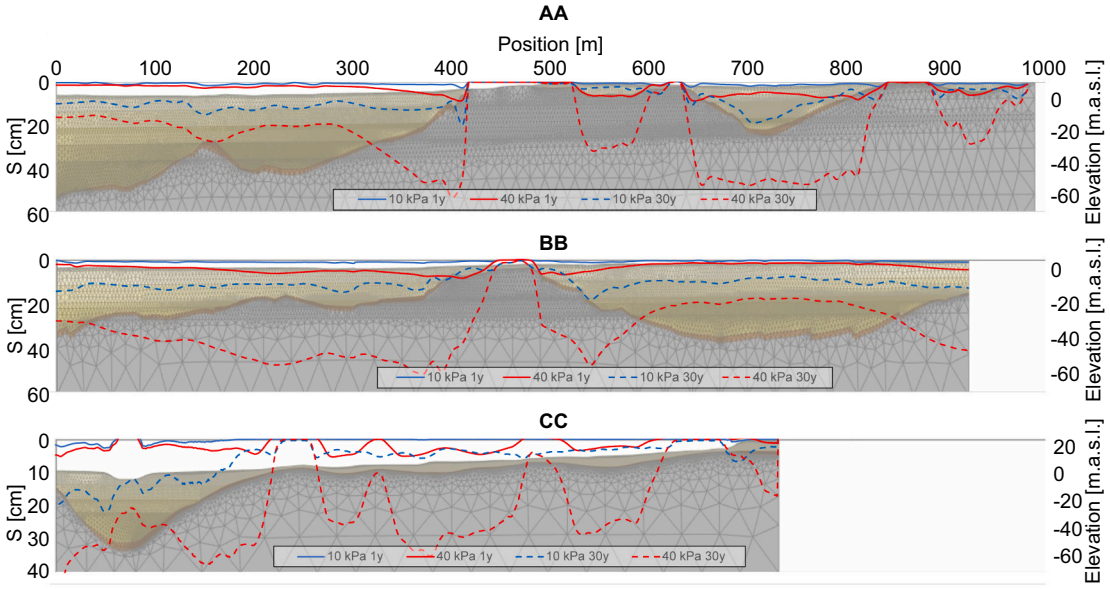


Fig. 6. Calculated vertical settlement (i.e., excluding background creep) results using Plaxis.

noticeable that the settlements are highly sensitive to the clay thickness, where extremely thin layers do not yield any significant settlements and slightly thicker layers yield significantly higher settlements. Thereafter, with increasing clay layer thickness, settlement magnitudes decrease. Another factor that influences settlement magnitude is the initial level of the groundwater head, which affects the pre-overburden pressure (Fig. S4, supplementary information). Generally, at locations with high groundwater heads, settlements are large, particularly when groundwater heads are higher in the confined than in the unconfined aquifer.

4.2. Metamodel evaluation

Pearson correlation of the subsidence prediction on five-fold cross-validated random hold-out data of the metamodel compared to the numerical model was consistently high (0.87–0.98) across the three different test sets (containing 25 % of data from sections AA+BB, AA+CC or BB+CC), see non-hatched bars in Fig. 7. The bottom panel of Fig. 7 also shows that prediction errors, as measured in RMSE, are small compared to estimated prediction ranges (Fig. 8), with averages of 0.24 cm, 0.87 cm, 1.3 cm, and 4 cm for scenarios 1y,10 kPa; 1y,40 kPa; 30y,10 kPa; and 30y,40 kPa, respectively and normally distributed errors (Fig. S4, supplementary information). Results do not vary substantially between the algorithms XGB and RF across the scenarios, achieving generally slightly higher correlation and lower error with XGB.

Compared to results from random hold-out, RF often outperforms XGB substantially regarding Pearson correlation and prediction error of entirely unseen cross sections. However, on unseen sections, the metamodel shows lower skill than for random hold-out (hatched bars, Fig. 7). This is particularly true for scenarios of small, short-term drawdowns (1 year, 10 kPa). The performance also differs across the test sections, where the test section AA and BB yield results that generally are only slightly below those of random hold-out test data (except for 1 year, 10 kPa). The cross-section CC on the other hand yields satisfactory results for only one scenario (30 years, 10 kPa). A reason for this is likely the different stratigraphy of CC compared to AA and BB. While AA and BB are relatively similar with larger sections of thick clays and small sections of thin clay layers, section CC is dominated by a large area with thin clay layers and shallow soil depth (compare Fig. 4).

The lower fit for the LOCSO evaluation is seen in Fig. 8, where for the CC test section, values are consistently overestimated in the section with shallow soils (180-700 m) and underestimated in deep clay deposits (50–150 m). However, the general subsidence pattern across all cross-sections is captured well by the metamodel. In scenario 1y,10 kPa for example, the effect of significantly larger subsidence in transition zones described in Section 4.1 is mainly exhibited in cross-section AA (e.g.,

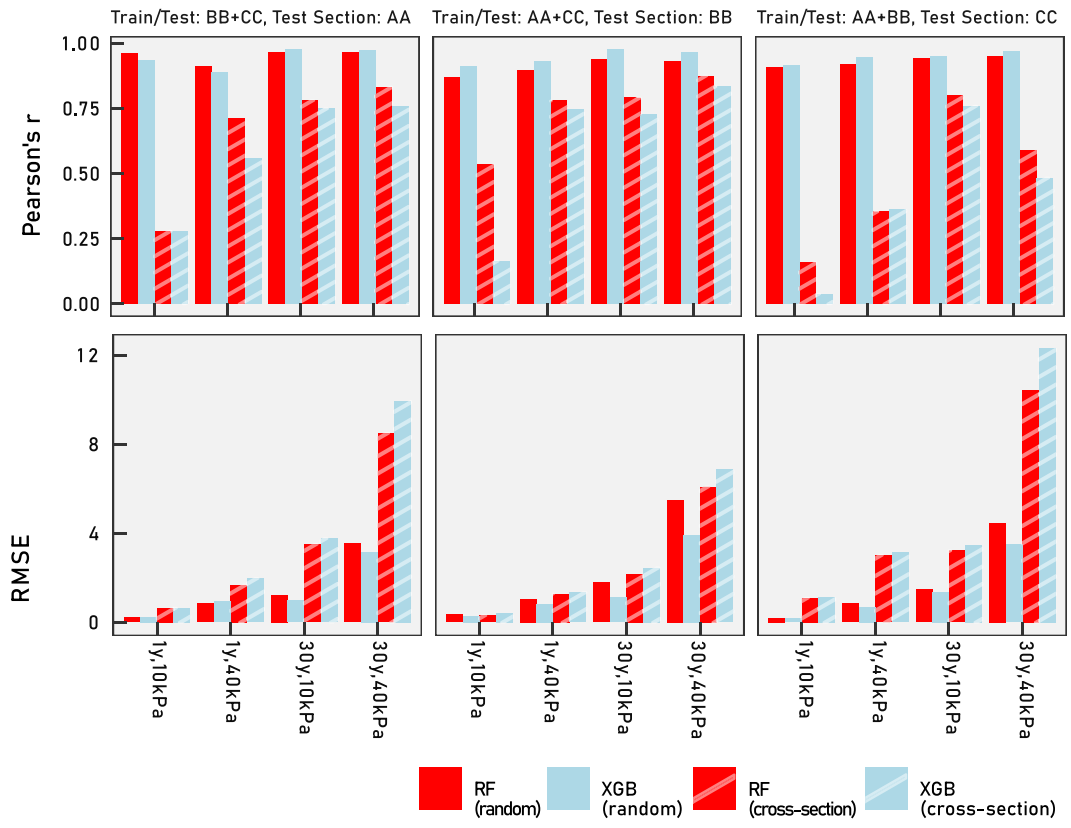


Fig. 7. Pearson correlation and root mean square error (RMSE) for different time and drawdown scenarios, machine learning models and test (evaluation) data sets.

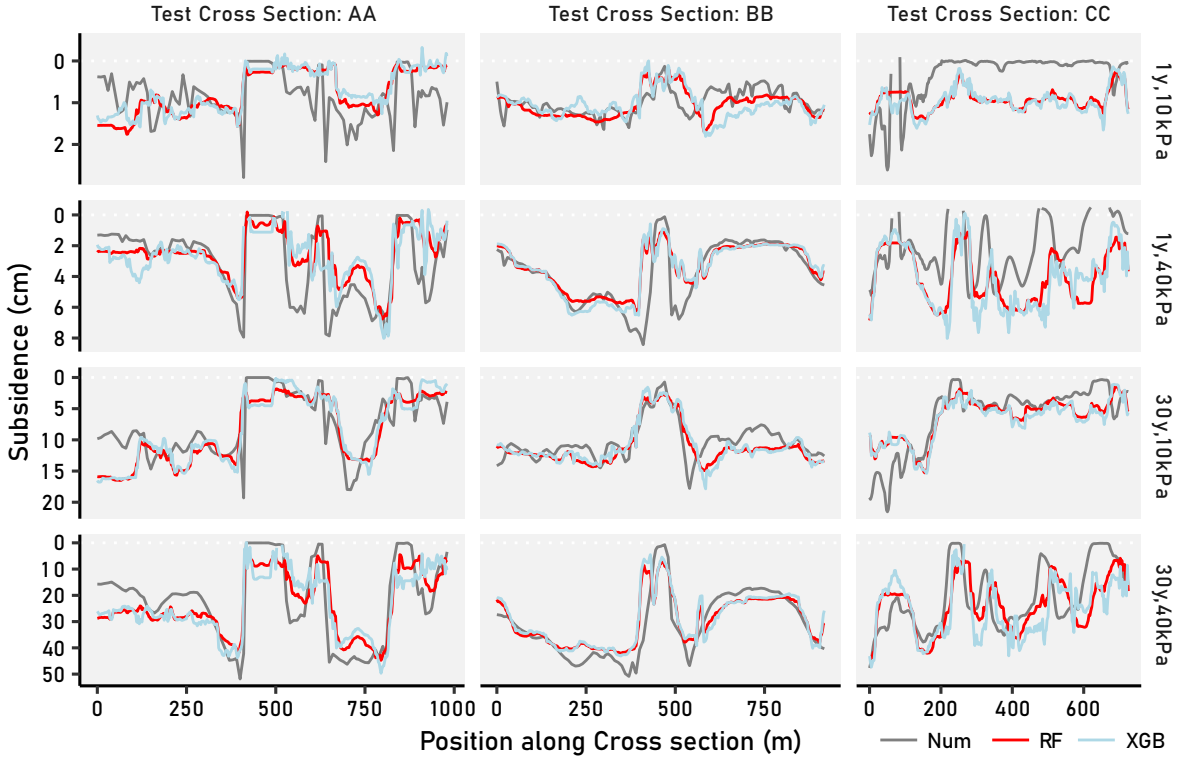


Fig. 8. Subsidence magnitude estimated by RF and XGB-based metamodels compared with numerical model (Num) along unseen test (evaluation) cross sections.

160 m, 400 m, 520 m, 600 m, 630 m, …), but is generally underestimated by the metamodel. This is likely due to the relatively few transition zone data points for the metamodel to train on and the lack of information being passed into the metamodel from neighboring grid points, which could indicate the presence of a transition zone (e.g., strong incline of bedrock surface from one grid point to the next). Cross-section BB contains fewer such transition zones, and the clay layer is thicker, resulting in lower variability of subsidence, and therefore higher overall performance in the 1y,10 kPa scenario. Similar effects can be seen for the remaining scenarios, but with much lower relative prediction errors, resulting in better overall correlation metrics.

In summary, the metamodel performs better when scenarios with larger spatial contrasts in subsidence are simulated, while low impact (e.g., 1y, 10 kPa) gives larger errors. For this scenario, the errors are mostly due to small shifts along the alignment and thus are still useful for decision support. Overall, very good correspondence can be achieved between the metamodel and the physical process-based numerical model, particularly when the data stratigraphy used to train the metamodel is representative of the area to be modeled.

4.3. Feature contribution to predictions

Across all metamodels, the most important features for prediction of subsidence magnitude are the clay thickness and soil thickness as quantified by mean absolute SHAP value, seen in Fig. 9. Soil thickness and clay layer in this environment are correlated since the soil profile in most areas is dominated by clay. However, the importance of clay thickness is generally significantly higher than the overall soil thickness. This is plausible since clay thickness is not only a proxy for sensitivity to subsidence but has a role in controlling the drainage length for dissipation of excess pore water pressure, and the soil stiffness at which consolidation occurs. This unique relationship between effective stress increase at different depths and the layer-dependent Over-Consolidation Ratio (OCR) for a specific consolidation time is affecting the soil stiffness and the resulting displacements, as described in detail by Wikby et al. (2023). The other soil layer thickness features, of the glacial till between bedrock and clay and fill are expected to have lower importance since the predicted soil displacements in these layers are negligible.

Depth to the groundwater (GW depth confined), particularly in the confined aquifer, is consistently important, controlling the initial pore water pressures, and thus the initial effective stresses which will have direct impact on the pre-consolidation pressures through OCR. While the pattern of the importance of the features is relatively consistent across the different scenarios, some differences can be seen. For instance, the relative importance of the soil depth (and remaining soil layers) and confined groundwater level depth is much higher. Furthermore, the importance increases with the thickness of the zone of OCR3 over OCR2, i.e., the OCR at higher clay depths, which is more often the zone of excess pore water dissipation, when the drawdowns are larger. Finally, while features have a very similar order of importance for both Machine Learning metamodels, RF and XGB, feature importance is more broadly distributed in the RF model. The XGB model on the other hand lays more weight on the most important features overall (Soil layer thicknesses), while RF gives more weight to the less important OCR features.

The contribution of features to the differences in predictions between RF and XGB can be nuanced. Fig. 10 shows the relationship between the contribution for a particular feature (here, clay and depth to groundwater) and each individual point in the data set. RF metamodels have very similar shapes and magnitudes of feature contribution across trained models, i.e., when the scenario is the same but different data sets

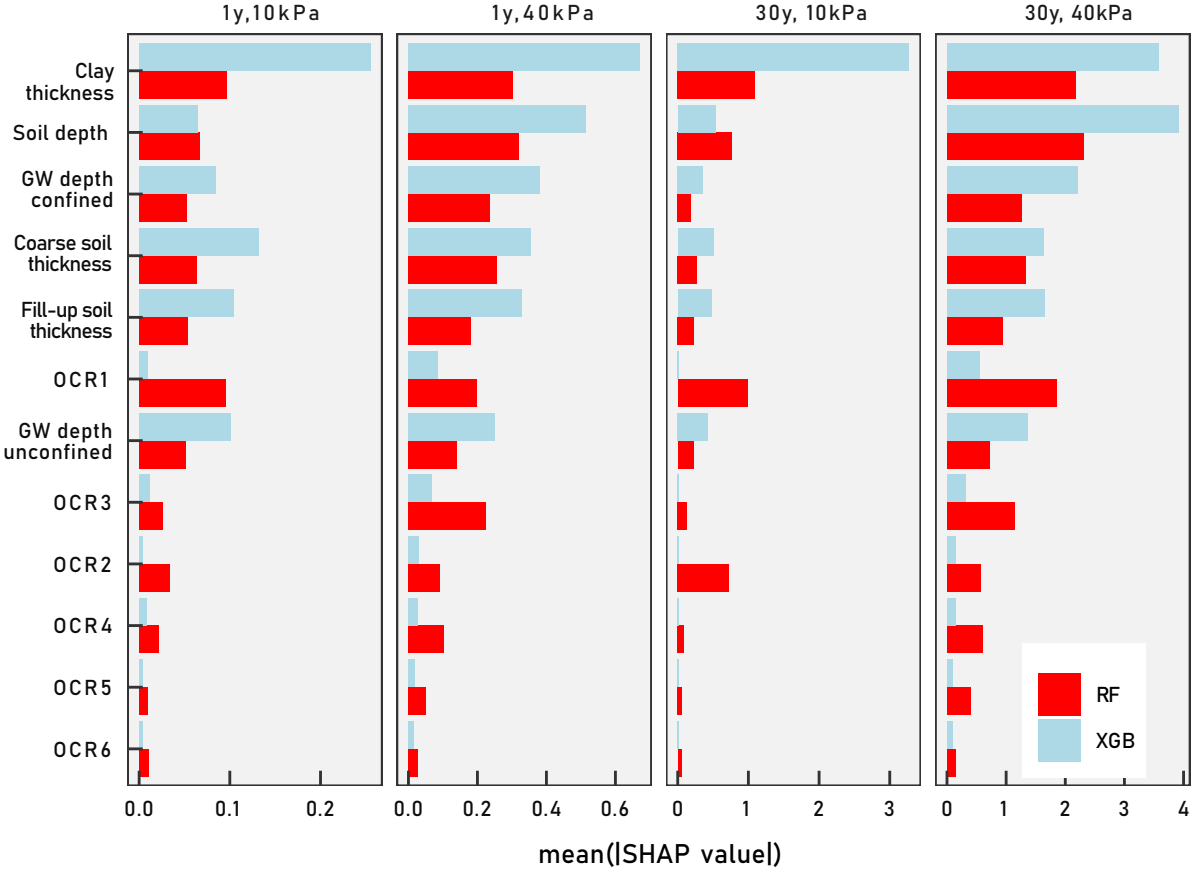


Fig. 9. Feature importance for prediction of subsidence (sorted by overall importance across models) based on mean absolute Shapley values.

(combinations of cross sections are used). The change of contribution of a feature across its value range also means that the effect of the features is not constant but exhibits varying contributions to the prediction with their physical quantity. Very low clay thickness results in a reduction of subsidence (negative SHAP values), before rising sharply with thickness, leading to an increase in the predicted subsidence with maxima between 10 and 20 m (Fig. 10A). After a peak, contribution falls again to zero or leads to reduced subsidence with larger clay thickness, coinciding with thicknesses with larger soil stiffness at 30–35 m (compare OCR in Fig. 3). The inflection points of the relationship between clay thickness and its contribution to the prediction depends also on the scenario, i.e., the actual drawdown scenario and the consolidation (and creep) period. Also, the range of absolute SHAP values varies, since the SHAP value is in units of the prediction value (cm subsidence), and consequently greater for scenarios with larger drawdowns and consolidation periods.

In the example of clay thickness shown in Fig. 10A, the scenarios with 40 kPa drawdown show higher absolute SHAP values compared to 10 kPa with respective consolidation time, which is related to a higher subsidence magnitude in these scenarios. The contribution of the depth to groundwater head in the confined aquifer varies as well, where high values relative to the surface level (negative values are equal to artesian conditions, i.e., above ground surface level) results in higher subsidence and vice versa in a mostly linear manner (Fig. 10B). This is expected since groundwater head governs the initial pore pressure distributions, which in turn govern the effective stresses (compare Fig. S5, supplementary information). (See Fig. 10.)

In summary, both the relative importance of features to subsidence prediction and the variation across the value range of the features is coherent with process understanding, as demonstrated for clay thickness and initial pore pressures. The two Machine Learning algorithms show

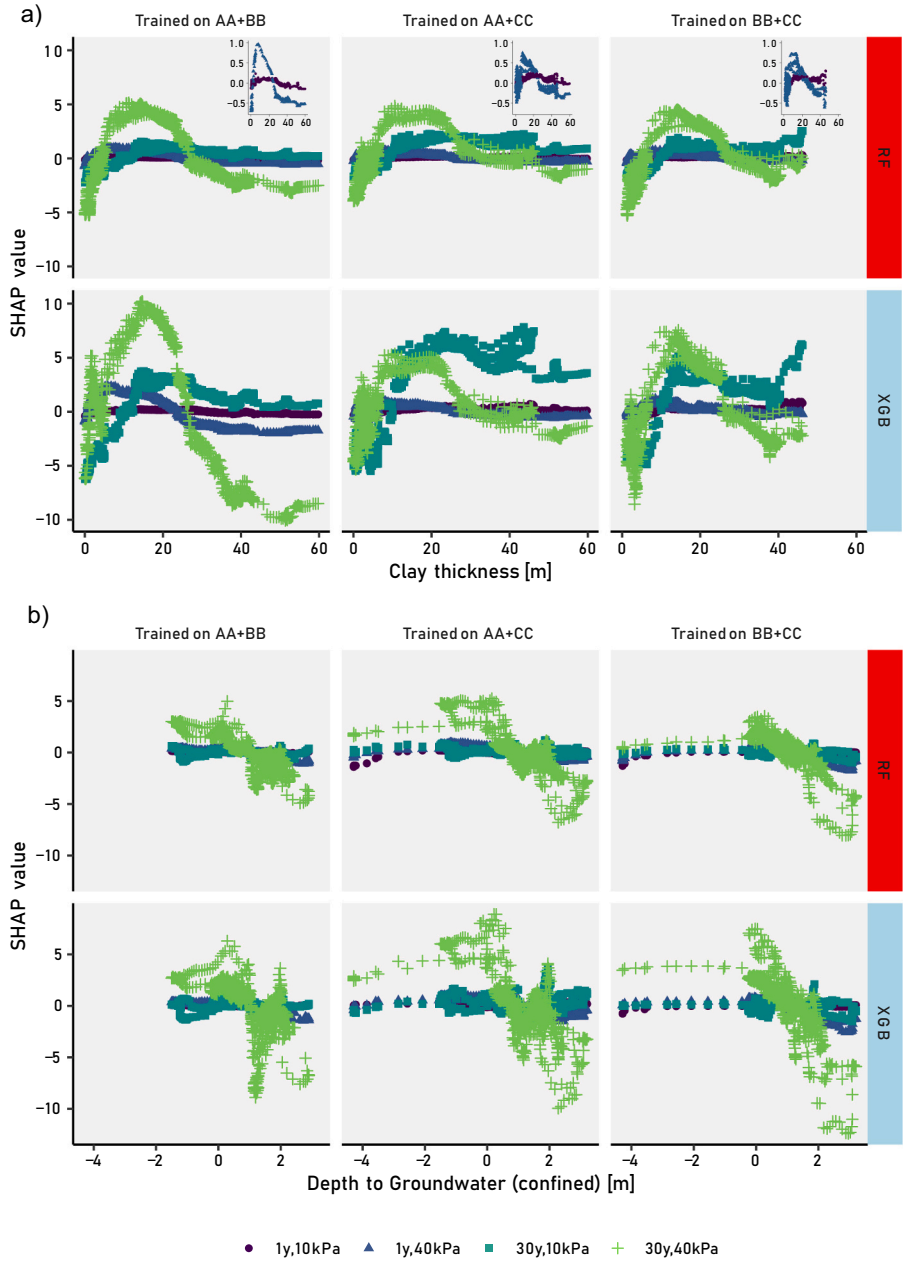


Fig. 10. a) Contribution of clay thickness range to metamodel prediction separated by the algorithm. Insets show enlarged distributions of scenarios 1y10kPa and 1y,40 kPa to better display their shapes, which are less visible in the main plots due to smaller magnitudes. The clay thickness value range is shorter when no training data from section AA is used (column "Trained on BB + CC) since only cross section AA has areas with clay thicker than 50 m. b) Contribution of depth to groundwater head in the confined aquifer. The value range of depth to groundwater is different since CC contains more points with artesian conditions.

similar results; however, RF learns from the importance of features more equally and shows stronger consistence across different data sets, indicating higher reliability, when predicting on data the metamodel has not been trained on.

4.4. Metamodel for large scale prediction

When running the trained models on features of the entire study area, mostly similar spatial patterns of subsidence emerge, both across training data sets and scenarios, as seen in Fig. 11a. In both scenarios, 1y, 40 kPa and 30y, 40 kPa shown in, RF-based predictions show higher predicted subsidence in areas with larger soil depth (Fig. 11b) and vice versa. However, in transition zones, i.e., areas where the clay tapers off from larger soil depths, the highest subsidence values are exhibited analogous the numerical model (Section 4.1). The higher subsidence in transition zones is more pronounced relative to other areas in the 30y than in the 1y scenario, which has also been observed in the numerical model (Fig. 6). However, this effect is also larger in cross sections AA and BB than CC, due to a higher presence of thicker clay layers tapering off. Therefore, metamodels trained on data from CC show a weaker transition zone effect, which can be seen when comparing the spatial patterns of the models trained on different cross sections in Fig. 11a. The standard deviation plots also show this clearly, highlighting the differences between the models that were trained on different data sets, where transition zones are the spatial features that exhibit the highest standard deviations.

Apart from closely approximating the results of an advanced physical process-based 2D numerical model, the proposed metamodeling approach is computationally efficient, where prediction on a 2.3 km^2 study site can be achieved in tens of seconds on a current mid-level PC. In contrast, calculation, and setup for physical process-based subsidence estimation at such large scales using a 3D numerical model would require significant computational time that is not feasible for most applications. An alternative approach to subsidence estimation at this scale is described by e.g., Sundell et al. (2019a) using a 1D semi-analytical model, solved at each grid cell. However, semi-analytical solutions that consider e.g., underdrainage, rate-dependency, degradation of bonding as well as anisotropy or 2D pore water pressure dissipation in clay are not available. Therefore, such semi-analytical solutions that do not model the advanced constitutive relationship for the soft clay and 2D processes will deliver potentially misleading results.

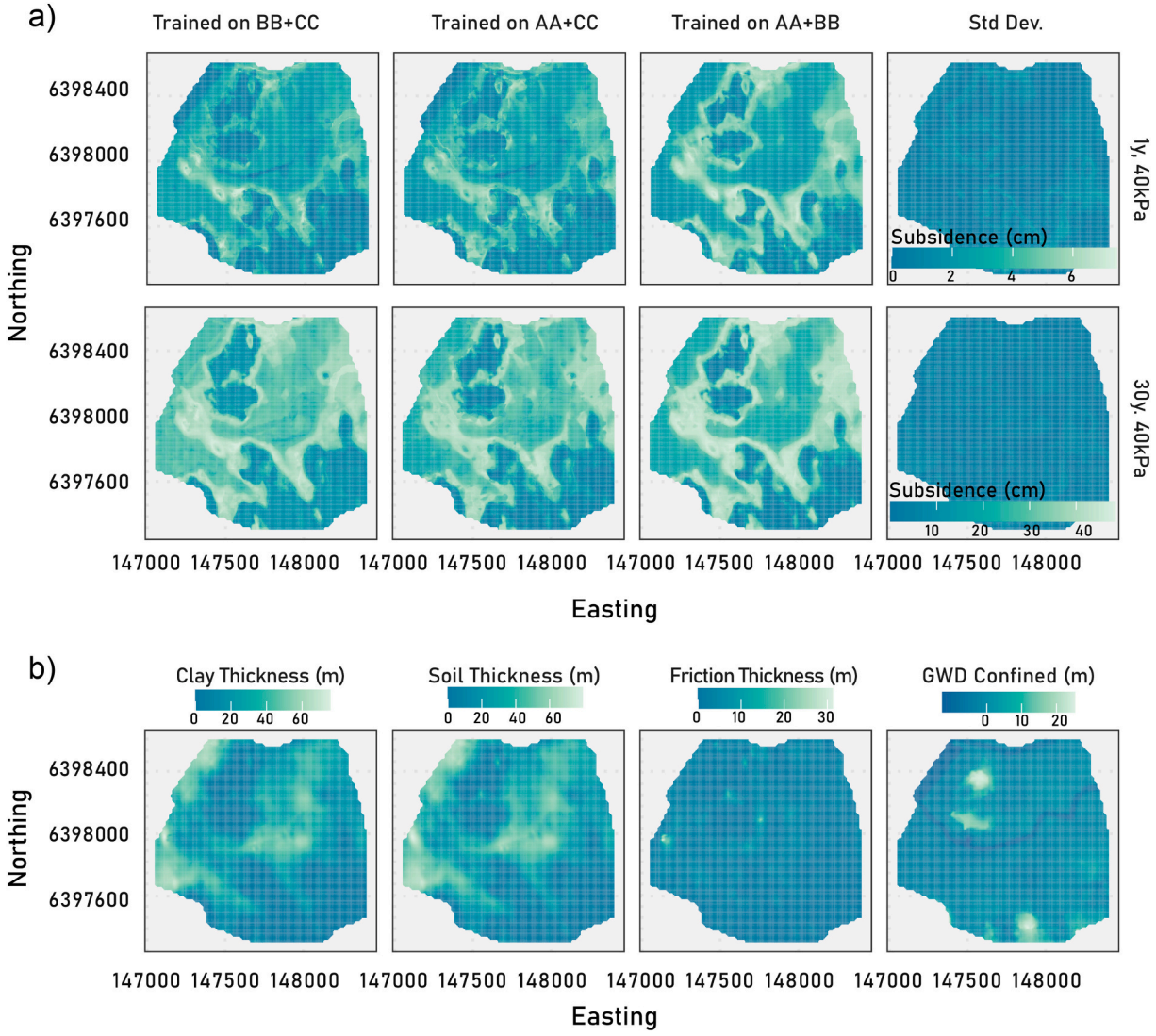


Fig. 11. a) Predictions with RF-metamodel on the entire modeling domain, based on different training data for scenarios 1y, 10 kPa (upper row) and 30y, 40 kPa (lower row). b) Spatial distribution of the four most important features (clay, soil, friction layer thickness, and depth to groundwater head from the surface of the confined aquifer) for prediction of subsidence.

4.5. Uncertainties and outlook

Despite the metamodel demonstrating promise in emulating the physical process-based model, model setup can be improved, and uncertainties further considered. An important factor is the dependence of the predictions on the hydrostratigraphic data and model. Because of this, the training data for the metamodels must be chosen with diligence, to ensure covering the value range of the features and their possible combinations over the entire study area for representative modeling as shown by the standard deviation plots and results in Section 4.2. Diligence is further necessary to satisfy the assumption of spatial stationarity of tree-based metamodels (Fienen et al., 2018; Furtney et al., 2022). In this study, cross sections were chosen according to expert knowledge covering the major morphological features of the subsurface. However, a more systematic spatial analysis and soil property sampling strategy across the entire study domain would possibly improve the robustness of the predictions (reducing standard deviations). Furthermore, using larger training data sets will also improve robustness and predictive skill. Nevertheless, this study shows that despite using relatively small training data sets (~500 data points), differences in the magnitude between areas are generalized well by the metamodel with small standard deviations relative to the FE predictions. Subsidence magnitudes in the transition zones are, however, sometimes underestimated. As shown in Section 4.1, sharp changes in subsidence occur at these transition zones, an effect seen covering a distance with multiple prediction points. Therefore, this effect may be improved by including relevant information to the set of features, which could be engineered based on neighboring prediction points. An example of such an engineered feature could be the slope of the bedrock between neighboring cells to be able to more explicitly pick up on the presence of transition zones. Such a step would turn the metamodel from an implicit 2D model into an explicit 2D model. In summary, further increase in predictive skill could be achieved by improving training data curation and selection and engineering of features.

Another uncertainty linked to the input data is the uncertainty of the parameters of the hydro-geomechanical and stratigraphical models. Regarding the hydro-geomechanical model, both the overall number of samples and the quality of the CRS tests used to estimate soil stiffness in this study was low, and representative prediction of the background creep rate could only be carried out thanks to the measurements from a bellow hose. While this proved effective in determining the representative stiffness and creep rate, only one bellow hose was available in the study area, resulting in a model with deterministic soil properties. To consider the heterogeneity and entailing uncertainty in soil parameters across the study area, more measurement points are required, which could be included in a probabilistic model setup, see e.g., Sundell et al. (2019a); Wikby et al. (2023). Such a probabilistic model setup could also include uncertainties regarding the hydrostratigraphic model, which is here based on tens of thousands of boreholes, but still has uncertainties (c.f. Sundell et al., 2016) that may play a role in e.g., the uncertainty of the transition zone effect.

The metamodel is intended as a part of a model chain for risk-based decision support in planning subsurface infrastructure in urban areas (Merisalu et al., 2021). To inform decision makers on costs arising from subsidence risk must be quantified. The next step in such a chain is to feed the subsidence computed with the metamodel into a building damage model (Wikby et al., 2024). Since the proposed model assumes green-field conditions, it ignores any soil-structure interaction and may overestimate settlements at building locations leading to overpredicted damage. However, emulating coupled soil-structure interaction for entire districts with a metamodel is currently out of reach. At this stage, when using this model for decision support, the overestimation must be considered, as well as the compounded errors of the geostatistical and numerical model. In future work, the errors may be reduced by improving the features and the metamodel setup (described above) but also can be quantified by rerunning the computationally efficient metamodel for many stochastic realizations of the geostatistical model. Further, at a stage where more knowledge of the dimensions and construction times of an underground structure is known, a hydrogeological model could be used to integrate and train the metamodel on non-uniform drawdown scenarios.

5. Conclusions

In this study, a metamodel is presented using a tree-based ensemble Machine Learning solution, which shows high skill in emulating a coupled hydro-geomechanical model, quantifying land subsidence due to pore-pressure reduction. With the presented approach, subsidence can be efficiently computed over large areas (2.3 km^2) for different pore-pressure reduction scenarios based on a spatially explicit hydrostratigraphic model accounting for the complex features of soft senstive clays. The predictions of the metamodel are coherent with the understanding of the physical processes as seen using the SHAP approach, which allows attribution of subsidence magnitude to proxy variables, such as over-consolidation Ratio and thus interpretation and plausibility checks. The prediction of the magnitude of subsidence is strongly dependent on the soil stratigraphy data, particularly clay depth and the drainage length for consolidation. The largest magnitudes of subsidence arise in the so-called transition zones, where the clay layer tapers off. Due to the lower stiffness and shorter drainage lengths, the pore pressure changes have rather immediate effect. The analysis has shown that this calls for more comprehensive uncertainty analysis, as well as features engineered to include information from neighboring cells, where the clay layer gradually thins out, capturing the spatial dependence of the processes. Together with uncertainty analysis and a building damage model, the metamodel can inform planners in a time- and cost-efficient way to carrying out risk-based decision and cost-benefit analysis in subsurface infrastructure at large scales.

CRediT authorship contribution statement

Ezra Haaf: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Pierre Wikby:** Writing – review & editing, Writing – original draft, Visualization, Validation, Formal analysis. **Ayman Abed:** Writing – review & editing, Validation, Supervision, Methodology. **Jonas Sundell:** Writing – review & editing, Validation, Supervision, Funding acquisition. **Eric McGivney:** Writing – review & editing, Software. **Lars Rosén:** Writing – review & editing, Validation, Supervision, Funding acquisition. **Minna Karstunen:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The research presented in this manuscript was financed by the Swedish Transport Administration (Trafikverket), Grant number TRV 2020/54637. The work is carried out as part of Digital Twin Cities Centre that is supported by Sweden's Innovation Agency VINNOVA under Grant No. 2019-00041.

Appendix. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.enggeo.2024.107705.

References

Agrell, H., 1979. The Quaternary of Sweden. Sveriges Geologiska Undersökning, Uppsala.

Amavasai, A., Gras, J.-P., Sivasithamparam, N., Karstunen, M., Dijkstra, J., 2017. Towards consistent numerical analyses of embankments on soft soils. Eur. J. Environ. Civ. Eng. 26, 2616–2634.

Amavasai, A., Sivasithamparam, N., Dijkstra, J., Karstunen, M., 2018. Consistent Class A & C predictions of the Ballina test embankment. Comput. Geotech. 93, 75–86.

Andersson, M., Andersson, J.V., Hallingberg, A., Lind, B., Löfroth, H., 2015. Förbättrade kontrollsystem för uppföljning av sättningar, Trafikverkets forskningsportföljer. Statens geotekniska institut, Linköping, p. 48.

Asher, M.J., Croke, B.F.W., Jakeman, A.J., Peeters, L.J.M., 2015. A review of surrogate models and their application to groundwater modeling. Water Resour. Res. 51, 5957–5973.

Bharti, J.P., Mishra, P., Moorthy, U., Sathishkumar, V.E., Cho, Y., Samui, P., 2021. Slope Stability Analysis using Rf, Gbm, Cart, Bt and Xgboost. Geotech. Geol. Eng. 39, 3741–3752.

Bjerre, E., Fienen, M.N., Schneider, R., Koch, J., Højberg, A.L., 2022. Assessing spatial transferability of a Random Forest metamodel for predicting drainage fraction. J. Hydrol. 612 (Part B), 128177.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Burbey, T.J., 2002. The influence of faults in basin-fill deposits on land subsidence, Las Vegas Valley, Nevada, USA. Hydrogeol. J. 10, 525–538.

Calderhead, A.I., Therrien, R., Rivera, A., Martel, R., Garfias, J., 2011. Simulating pumping-induced regional land subsidence with the use of InSAR and field data in the Toluca Valley, Mexico. Adv. Water Resour. 34, 83–97.

Chaussard, E., Wdowinski, S., Cabral-Cano, E., Amelung, F., 2014. Land subsidence in Central Mexico detected by ALOS InSAR time-series. Remote Sens. Environ. 140, 94–106.

Fienen, M.N., Nolan, B.T., Kauffman, L.J., Feinstein, D.T., 2018. Metamodeling for Groundwater Age forecasting in the Lake Michigan Basin. Water Resour. Res. 54, 4750–4766.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29 (1189–1232), 1144.

Furtney, J., Thielsen, C., Fu, W., Le Goc, R., 2022. Surrogate models in rock and soil mechanics: integrating numerical modeling and machine learning. Rock Mech. Rock. Eng. 55, 1–15.

Galloway, D.L., Burbey, T.J., 2011. Review: Regional land subsidence accompanying groundwater extraction. Hydrogeol. J. 19, 1459–1486.

Gras, J.P., Sivasithamparam, N., Karstunen, M., Dijkstra, J., 2017a. Permissible range of model parameters for natural fine-grained materials. Acta Geotech. 13, 387–398.

Gras, J.P., Sivasithamparam, N., Karstunen, M., Dijkstra, J., 2017b. Strategy for consistent model parameter calibration for soft soils using multi-objective optimisation. Comput. Geotech. 90, 164–175.

Guzy, A., Malinowska, A., 2020. State of the art and recent advancements in the modelling of land subsidence induced by groundwater withdrawal. Water 12.

Haaf, E., Giese, M., Reimann, T., Barthel, R., 2023. Data-driven estimation of groundwater level time-series at unmonitored sites using comparative regional analysis. Water Resour. Res. 59 https://doi.org/10.1029/2022wr033470.

Harbaugh, A.W., 2005. MODFLOW-2005 : the U.S. Geological Survey modular ground-water model–the ground-water flow process. In: Techniques and Methods.

Hoffmann, J., Leake, S.A., Galloway, D.L., Wilson, A.M., 2003. MODFLOW-2000 Ground-Water Model-User Guide to the Subsidence and Aquifer-System Compaction (SUB) Package. Open-File Report, Reston, VA.

Hou, Z., Lu, W., 2018. Comparative study of surrogate models for groundwater contamination source identification at DNAPL-contaminated sites. Hydrogeol. J. 26, 923–932. https://doi.org/10.1007/s10040-017-1690-1.

Huang, B., Shu, L., Yang, Y.S., 2012. Groundwater overexploitation causing land subsidence: hazard risk assessment using field observation and spatial modelling. Water Resour. Manag. 26, 4225–4239.

Kang, F., Li, J.-S., Wang, Y., Li, J., 2016. Extreme learning machine-based surrogate model for analyzing system reliability of soil slopes. Eur. J. Environ. Civ. Eng. 21, 1–22.

Karstunen, M., Krenn, H., Wheeler, S.J., Koskinen, M., Zentar, R., 2005. Effect of anisotropy and destructuration on the behavior of Murro Test Embankment. Int. J. Geomech. 5, 87–97.

Kim, D., Kwon, K., Pham, K., Oh, J.-Y., Choi, H., 2022. Surface settlement prediction for urban tunneling using machine learning algorithms with Bayesian optimization. Autom. Constr. 140.

Kuhn, M., Frick, H., 2022. Dials: Tools for Creating Tuning Parameter Values.

Kuhn, M., Wickham, H., 2020. Tidymodels: A Collection of Packages for Modeling and Machine Learning Using Tidyverse Principles.

Larsson, R., Bengtsson, P.-E., Eriksson, L., 1997. Prediction of Settlements of Embankments on Soft, Fine-Grained Soils : Calculation of Settlements and their Course with Time. SGI, Linköping.

Lundberg, S.M., Erion, G., Chen, H., Degrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2, 56–67.

Mahmoudpour, M., Khamehchiyan, M., Nikudel, M.R., Ghassemi, M.R., 2016. Numerical simulation and prediction of regional land subsidence caused by groundwater exploitation in the southwest plain of Tehran, Iran. Eng. Geol. 201, 6–28.

Mayer, M., Watson, D., 2023. kernelshap: Kernel SHAP. R package version 0.3.5.

Meray, A., Wang, L., Kurihana, T., Mastilovic, I., Praveen, S., Xu, Z., Memarzadeh, M., Lavin, A., Wainwright, H., 2024. Physics-informed surrogate modeling for supporting climate resilience at groundwater contamination sites. Comput. Geosci. 183, 105508 https://doi.org/10.1016/j.cageo.2023.105508.

Merisalu, J., Sundell, J., Rosén, L., 2021. A framework for risk-based cost-benefit analysis for decision support on hydrogeological risks in underground construction. Geosciences (Switzerland) 11, 82.

Niswonger, R.G., Panday, S., Ibaraki, M., 2011. MODFLOW-NWT, A Newton formulation for MODFLOW-2005. In: Techniques and Methods.

Obel, M., Ahrens, M.A., Mark, P., 2020. Metamodel-based prediction of structural damages due to tunneling-induced settlements. ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A Civ. Eng. 6.

Ochoa-González, G.H., Carreón-Freyre, D., Franceschini, A., Cerca, M., Teatini, P., 2018. Overexploitation of groundwater resources in the faulted basin of Querétaro, Mexico: a 3D deformation and stress analysis. Eng. Geol. 245, 192–206.

Sivasithamparam, N., Karstunen, M., Bonnier, P., 2015. Modelling creep behaviour of anisotropic soft soils. Comput. Geotech. 69, 46–57.

Starn, J.J., Kauffman, L.J., Carlson, C.S., Reddy, J.E., Fienen, M.N., 2021. Three-dimensional distribution of groundwater residence time metrics in the glaciated United States using metamodels trained on general numerical simulation models. Water Resour. Res. 57.

Sundell, J., Rosén, L., Norberg, T., Haaf, E., 2016. A probabilistic approach to soil layer and bedrock-level modeling for risk assessment of groundwater drawdown induced land subsidence. Eng. Geol. 203, 126–139.

Sundell, J., Haaf, E., Norberg, T., Alén, C., Karlsson, M., Rosén, L., 2017. Risk mapping of groundwater-drawdown-induced land subsidence in heterogeneous soils on large areas. Risk Anal. 39, 105–124.

Sundell, J., Haaf, E., Tornborg, J., Rosén, L., 2019a. Comprehensive risk assessment of groundwater drawdown induced subsidence. Stoch. Env. Res. Risk A. 33, 427–449.

Sundell, J., Norberg, T., Haaf, E., Rosén, L., 2019b. Economic valuation of hydrogeological information when managing groundwater drawdown. Hydrogeol. J. 27, 1111–1130.

Teatini, P., Ferronato, M., Gambolati, G., Gonella, M., 2006. Groundwater pumping and land subsidence in the Emilia-Romagna coastland, Italy: Modeling the past occurrence and the future trend. Water Resour. Res. 42.

Tornborg, J., Karlsson, M., Kullingsjö, A., Karstunen, M., 2021. Modelling the construction and long-term response of Göta Tunnel. Comput. Geotech. 134.

Trefry, M.G., Muffels, C., 2007. FEFLOW: a finite-element ground water flow and transport modeling tool. Groundwater 45, 525–528.

Wang, M.-X., Huang, D., Wang, G., Li, D.-Q., 2020. SS-XGBoost: a machine learning framework for predicting newmark sliding displacements of slopes. J. Geotech. Geoenviron. Eng. 146, 04020074.

Wheeler, S.J., Näätänen, A., Karstunen, M., Lojander, M., 2003. An anisotropic elastoplastic model for soft clays. Can. Geotech. J. 40, 403–418.

Wikby, P., Abed, A., Karlsson, M., Sundell, J., Karstunen, M., 2023. The Influence of Parameter Variability on Subsidence.

Wikby, P., Haaf, E., Abed, A., Rosén, L., Sundell, J., Karstunen, M., 2024. A grid-based methodology for the assessment of time-dependent building damage at large scale. Tunn. Undergr. Space Technol. 149, 105788 https://doi.org/10.1016/j.tust.2024.105788.

Wright, M.N., Ziegler, A., 2017. Ranger: a fast implementation of random forests for high dimensional data in C++ and R. J. Stat. Softw. 77, 1–17.

Ye, S., Luo, Y., Wu, J., Yan, X., Wang, H., Jiao, X., Teatini, P., 2016. Three-dimensional numerical modeling of land subsidence in Shanghai, China. Hydrogeol. J. 24, 695–709.

Zhang, W., 2020. A Review of Surrogate Models, MARS Applications in Geotechnical Engineering Systems: Multi-Dimension with Big Data. Springer Singapore, Singapore, pp. 7–17.

Zhang, W.G., Li, H.R., Wu, C.Z., Li, Y.Q., Liu, Z.Q., Liu, H.L., 2021. Soft computing approach for prediction of surface settlement induced by earth pressure balance shield tunneling. Undergr. Space 6, 353–363.

Zhou, J., Shi, X., Du, K., Qiu, X., Li, X., Mitri, H.S., 2017. Feasibility of random-forest approach for prediction of ground settlements induced by the construction of a shield-driven tunnel. Int. J. Geomech. 17, 04016129.

Zhu, L., Zhang, C., Li, M., Pan, X., Sun, J., 2012. Building 3D solid models of sedimentary stratigraphic systems from borehole data: an automatic method and case studies. Eng. Geol. 127, 1–13. https://doi.org/10.1016/j.enggeo.2011.12.001.

Zhu, X., Chu, J., Wang, K., Wu, S., Yan, W., Chiam, K., 2021. Prediction of rockhead using a hybrid N-XGBoost machine learning framework. J. Rock Mech. Geotech. Eng. 13, 1231–1245.

Zoccarato, C., Ferronato, M., Teatini, P., 2021. A Surrogate Model for Fast Land Subsidence Prediction and Uncertainty Quantification. Springer International Publishing, pp. 943–950.