

This is the preprint version of the contribution published as:

Paasche, H. (2018):

About probabilistic integration of ill-posed geophysical tomography and logging data: A knowledge discovery approach versus petrophysical transfer function concepts illustrated using cross-borehole radar-, P- and S-wave traveltimes tomography in combination with cone penetration and dielectric logging data

J. Appl. Geophys. **148**, 175 – 188

The publisher's version is available at:

<http://dx.doi.org/10.1016/j.jappgeo.2017.11.017>

About probabilistic integration of ill-posed geophysical tomography and logging data: A knowledge discovery approach versus petrophysical transfer function concepts illustrated using cross-borehole Radar-, P- and S-wave traveltimes tomography in combination with cone penetration and dielectric logging data

Hendrik Paasche

UFZ - Helmholtz Centre for Environmental Research, Department of Monitoring and Exploration Technologies, Permoserstr. 15, 04318 Leipzig, Germany

phone: +49 341 2351414; fax: +49 341 2351939

Email: hendrik.paasche@ufz.de

Manuscript submitted for publication in *Journal of Applied Geophysics*, February, 2017

Revised version, August, 2017

Revised version, November, 2017

Abstract

Site characterization requires detailed and ideally spatially continuous information about the subsurface. Geophysical tomographic experiments allow for spatially continuous imaging of physical parameter variations, e.g., seismic wave propagation velocities. Such physical parameters are often related to typical geotechnical or hydrological target parameters, e.g. as achieved from 1D direct push or borehole logging. Here, the probabilistic inference of 2D tip resistance, sleeve friction, and relative dielectric permittivity distributions in near-surface sediments is constrained by ill-posed cross-borehole seismic P- and S-wave and radar wave traveltime tomography. In doing so, we follow a discovery science strategy employing a fully data-driven approach capable of accounting for tomographic ambiguity and differences in spatial resolution between the geophysical tomograms and the geotechnical logging data used for calibration. We compare the outcome to results achieved employing classical hypothesis-driven approaches, i.e., deterministic transfer functions derived empirically for the inference of 2D sleeve friction from S-wave velocity tomograms and theoretically for the inference of 2D dielectric permittivity from radar wave velocity tomograms. The data-driven approach offers maximal flexibility in combination with very relaxed considerations about the character of the expected links. This makes it a versatile tool applicable to almost any combination of data sets. However, error propagation may be critical and justify thinking about a hypothesis-driven pre-selection of an optimal database going along with the risk of excluding relevant information from the analyses. Results achieved by transfer function rely on information about the nature of the link and optimal calibration settings drawn as retrospective hypothesis by other authors. Applying such transfer functions at other sites turns them into a priori valid hypothesis, which can, particularly for empirically derived transfer functions, result in poor

predictions. However, a mindful utilization and critical evaluation of the consequences of turning a retrospectively drawn hypothesis into an a priori valid hypothesis can also result in good results for inference and prediction problems when using classical transfer function concepts.

Keywords: parameter estimation, petrophysics, borehole geophysics, data-driven, hypothesis-driven, discovery science

1 Introduction

For taking educated decisions, engineers require detailed and ideally spatially continuous information about the subsurface. Geophysical tomographic imaging offers unique potential to image ground variability in 2D or 3D with a spatial resolution of a few meters. However, geophysical tomograms image physical parameters, e.g., propagation velocities of seismic or electromagnetic waves. Based on theoretic models or empirical testing it is known that some physical parameters are particularly linked to classical geotechnical or hydrological target parameters, e.g., porosity, tip resistance or sleeve friction, which are usually more familiar and easily interpretable for engineers. While direct-push- or borehole-based logging techniques provide detailed information about the vertical variability of the acquired target parameters, logging techniques inherently lack lateral resolution. Geophysical imaging has been used for a long time striving to mitigate this gap and support geotechnical and hydrological site characterization for engineering purposes (e.g. Ward 1994; Friedel et al. 1997; Butler 2005).

However, in most cases geophysical tomograms have only been qualitatively integrated in the interpretation of logging data (Niederleithinger 2015), e.g., laterally linking different logs by taking the spatial variability of geophysical data or tomograms as soft (visual) constraint (e.g., Shahri et al. 2015). A more rigorous integration is challenging, since the inter-relations between typical hydrological or geotechnical (e.g., sleeve friction) and geophysical (e.g., seismic shear wave velocity) target parameters are often not exactly known, site specific, spatially variable, non-linear and even non-unique (e.g., Schön 2004). Furthermore, experimental data carry errors despite high efforts to guarantee optimal data quality when carrying out field experiments (e.g., Taylor 1996). In geophysical tomographic experiments, limited number of observations and limited observational accuracies result in ill-posed tomographic reconstruction problems inherently suffering ambiguity when striving to find a

physical tomographic model explaining the observed geophysical tomographic data set (e.g., Friedel 2003). Classical geophysical tomographic imaging relies on regularized optimization techniques searching the solution space locally in the vicinity of a given initial model (e.g., Aster et al. 2005). Such deterministic imaging techniques do not allow for realistic assessment of geophysical imaging ambiguity and are thus generally limited in their potential to support probabilistic geotechnical or hydrological site characterization.

Many deterministic transfer functions have been developed formulating the link between a distinct geophysical imaged parameter, e.g., shear (S-) wave velocity, and a geotechnical or hydrological target parameter, e.g., sleeve friction (e.g., Sykora and Stokoe 1983; Hegazy and Mayne 1995; Piratheepan 2002; Robertson 2009; Tonni and Simonini 2013; Mola-Abasi et al. 2015). Such transfer functions “may be derived either theoretically on the basis of models or by empirical observations” (Schön, 2004). They are typical examples for hypothesis-driven scientific approaches (e.g., nmeth0409-237, 2009; Mazzocchi, 2015) and particularly the latter must be regarded as retrospectively drawn hypotheses about existing parameter interrelations based on a limited set of observations. Frequently, classical transfer functions are limited to the description of the link between two distinct parameters. The applicability of the thus developed transfer functions is usually limited to specific geological settings, for example Holocene sands. The sheer existence of multiple transfer functions for the same geological setting and geophysical-geotechnical or geophysical-hydrological parameter combination indicates that approaches striving to explicitly formulate the inter-relation between geophysical and the target parameters cannot cope well with non-unique, spatially variable and site-specific parameter inter-relations. Nevertheless, transfer function concepts have been successfully employed for geotechnical and hydrological site characterization (e.g., Dannowski and Yaramanci 1999; Kaufmann et al. 2012; Lorenzo et al. 2014).

In the last two decades various statistical and geostatistical approaches have been developed to link geophysical imaged parameters to laterally sparse measured geotechnical or hydrological target parameters accounting for complex (multi-facies) and a priori unknown inter-parameter relations in the survey area. For example, co-kriging (Cassiani et al. 1998; Gloaguen et al. 2001), Bayesian inference (Ezzedine et al. 1999; Hubbard et al. 2001; Chen et al. 2001; Bosch et al. 2010; Dubreuil-Boisclair et al. 2011; Ruggeri et al. 2013), fuzzy cluster analysis (Paasche et al. 2006; Hachmöller and Paasche 2013) or conditional stochastic simulation (Tronicke and Holliger 2005; Dafflon et al. 2009) based approaches have been developed. Some of them can cope with unknown, spatially variable, non-linear, and even non-unique inter-relations thus following data-driven or discovery science ideas (e.g., nmeth0409-237, 2009; Mazzocchi, 2015) relaxing hypothesizing about the particular usefulness of a distinct data set for the envisaged task or the existence or the character of an inter-data link. Applications have been limited to deterministic geophysical tomography, e.g., ignoring the geophysical ambiguity resulting in potential artifacts of the finally produced geophysical tomograms.

Recently, Rumpf and Tronicke (2014) employed alternating conditional expectation (ACE; Breiman and Friedman 1985) to link ensembles of 125 radar, P-wave, and S-wave velocity tomograms with sparsely measured different geotechnical target parameters, including effective grain size and sleeve friction, to predict 2D models of the target parameters. All considered radar and seismic tomograms fit the underlying data set equally well and are thus considered as a discrete description of the tomographic reconstruction ambiguity. The Rumpf and Tronicke (2014) approach is probabilistic and prediction accuracy is illustrated by mean or median values in combination with percentile ranges. Asadi et al. (2016) demonstrate with a synthetic database how feed-forward Artificial Neural Networks (ANNs) can be used to link ensembles of 30 radar and seismic tomograms with sparse porosity data in order to transduce tomographic ambiguity into 2D probabilistic porosity prediction. The same is achieved by

Paasche (2017) using fuzzy concepts instead of ANNs to link ensembles of physically disparate geophysical tomograms with sparse logging data. Their approach is found more robust with regard to measurement errors of the logging data used for calibration or learning of the inter-relation between the target parameter and the logging data than those of Asadi et al. (2016).

Here, we apply the technique of Paasche (2017) to a field database recorded by Linder et al. (2010) comprising cross-borehole tomographic radar, P-wave and S-wave traveltimes data sets as well as two CPT measurements and direct-push based dielectric permittivity logs. With this approach we follow a pure data-driven or discovery science approach simply including all available data sets regardless of their theoretically or empirically proven suitability for the desired prediction task, which is in our example the probabilistic prediction of 2D tip resistance, sleeve friction, and dielectric permittivity models. Note, that here no petrophysical model derived prior to our analyses, e.g., based on theoretical insights or empirical analyses of other data, is considered in the prediction. The outcome of this approach is opposed to prediction results achieved when employing (i) empirically found transfer functions linking tomographic images of S-wave velocity with sleeve friction logs and (ii) a theoretically derived transfer function linking dielectric permittivity logs and radar wave velocity tomograms. We discuss the strengths and weaknesses of data-driven or discovery science approaches and classical transfer function concepts, which can both result in good predictions if mindfully applied.

2 Field site and database

The database has been acquired by Linder et al. (2010) and used for deterministic tomographic imaging and geotechnical prediction. The field site is located 30 km south of Berlin close to the village Horstwalde and is part of a larger test area of the German Federal

Institute for Materials Research and Testing (BAM; Niederleithinger 2009). The near-surface geological settings have been shaped during the last glaciations. Sandy and gravelly glacial and fluvioglacial sediments with inter-bedded thin layers of organic material below 8 m depth are deposited (Linder et al. 2010; Schmelzbach et al. 2011). During the time of measurements the ground water table was approximately 3 m below surface.

Concrete piles and PVC cased boreholes have been installed on the site, which was also used as test site for non-destructive testing of foundations (see Figure 1). Between boreholes BH2, BH3, and BH5 (Figure 1) cross-borehole geophysical experiments have been carried out. Additionally, at two locations within the tomographic planes cone penetration tests (CPT) have been done. At these locations a direct-push based soil-moisture probe (Shinn et al. 1998) has been employed to record depth profiles of dielectric permittivity.

2.1 Logging data

At two locations in the tomographic planes between BH2 and BH3 as well as BH3 and BH5 direct-push experiments have been performed. Cone penetration tests were carried out using a piezocone probe of 4.4 cm diameter that was pushed into the ground with controlled force and constant speed of 2 cm/s. Tip resistance and sleeve friction are the measured quantities (Lunne et al. 1997). Additionally, a direct push soil moisture probe has been employed to measure the dielectric permittivity at a frequency of 30 MHz. The resultant logs with a vertical sample spacing of 1 cm are shown in Figure 2. These data provide high-resolution information essentially required for geotechnical site characterization. However, despite the laterally small distance of approximately 5 m between the two logging positions, the logs exhibit some differences, particularly at depth below 8 m. The dielectric permittivity logs show some prominent peaks of increased permittivity indicating presence of some organic

layers with high water content extending only a few centimeters or decimeters in their thicknesses.

2.2 Crosshole travelttime tomography

Between adjacent boreholes BH2, BH3, and BH5 seismic P-wave, seismic S-wave, and radar-wave crosshole tomographic experiments have been carried out (Figure 1). A detailed discussion of the tomographic experiments and an overview about geophysical acquisition parameters and signal characteristics is given by Linder et al. (2010). First cycle onsets have been picked in the recorded shotgathers to achieve tomographic travelttime data sets suitable for 2D P-wave, S-wave and radar-wave velocity reconstruction by inversion. The inversion strategy chosen here differs from those used by Linder et al. (2010).

We invert each travelttime data set separately using particle swarm optimization (Kennedy and Eberhart 1995) to search the solution space of a bi-objective inverse problem concurrently addressing a data misfit measure and a model roughness constraint. This approach is similar to those of Schwarzbach et al. (2005), but used the model roughness constraint of Pham (2001) to avoid the model parameter amplitude damping inherent to the Tikhonov regularization used by Schwarzbach et al. (2005). A game theoretic decision strategy is used to balance the interests of both objectives when solving the inverse problem (Paasche and Tronicke 2014). We repeatedly run the inversion to achieve ensembles of 30 velocity models explaining the underlying data set equally well. The reconstructed spatial velocity variability varies for each model so that a set of equivalent models illustrates the ambiguity inherent to the tomographic velocity reconstruction procedure in discrete form. Note, other global search inversion approaches resulting in ensembles of 2D tomographic velocity models matching the underlying data set equally well could have been used alternatively. For practical considerations, we stopped our efforts after finding 30 equivalent velocity tomograms

explaining the underlying data set equally well. Other authors invested more attention in evaluating the representativeness of the found ensemble for ambiguity assessment (e.g., Fernández Martínez et al. 2012; Tronicke et al. 2012; Fernández-Martínez et al. 2013; Sen and Stoffa 2013).

We parameterized the 2D model reconstruction area along the inter-borehole planes between BH2, BH3, and BH5 using rectangular grid cells with horizontal and vertical side lengths of 1 m and 0.5 m, respectively (Figure 3a). Given the vertical grid cell extension of 0.5 m and the vertical sample distance of 1 cm used for logging of tip resistance, sleeve friction and dielectric permittivity, 50 log readings are available for the corresponding tomographic grid cell (Figure 3b). The lateral direction in the mesh chosen for model area parameterization follows the direction of the inter-borehole planes. Thus, boreholes BH2, BH3, and BH5 are projected at $x = 0$ m, $x = 5.01$ m, and $x = 10.96$ m. Direct push injection points for cone penetration and dielectric permittivity measurements are then located at $x = 2.75$ m and $x = 8.0$ m, respectively.

The 2D velocity models achieved by inversion are presented as 11 laterally continuous panels of 1D velocity variation over depth (Figure 3a). Figure 4 shows the seismic P-wave, seismic S-wave and radar-wave velocity models in the form of 1D panels. Each panel contains the information of 30 equivalent P-wave, S-wave, and radar-wave velocity models. For example, all 30 2D P-wave velocity models in Figure 4a show increased velocities around 6 m depth and below 14 m depth. The P-wave velocities found by the 30 equivalent models in the top left corner (panel $x = 0.5$ m in Figure 4a, depth above 7 m) of the 2D model reconstruction area scatter strongly as indicated by a rather broad velocity range. Thus, Figure 4 provides information about 2D P-wave, S-wave and radar-wave velocity as well as the velocity range and distribution for each tomographic grid cell, i.e. model parameter. This tomographic reconstruction ambiguity shall be taken into account when linking geophysical

tomographic models with target parameters when striving to realize a more quantitative integration of geophysical tomographic models and sparse logging data for site characterization.

3 Probabilistic inference of 2D target parameter distributions

3.1 Methodology based on fuzzy sets

The processing flow for our probabilistic inference of 2D geotechnical parameter fields follows Paasche (2017, option 1 in their discussion). As summarized in Figure 5 it infers spatially continuous distributions of a target parameter from physically disparate tomographic images and sparse calibration data, e.g., measured in boreholes or by direct push technology. The inference approach is generally probabilistic and can account for tomographic ambiguity as well as scale differences between tomographic grid cell size and logging sample interval. Fundamental assumption is that geophysical tomographic data and geotechnical logging data image the same subsurface reality and must therefore be compliant. From each available tomographic data set multiple tomograms are reconstructed, which fit the underlying data set equally and sufficiently well. In our example, we have three tomographic data sets (crosshole seismic P-wave, S-wave and radar-wave traveltime data) and for each data set we computed 30 2D velocity tomograms, which fit the underlying data set equally well (see Figure 4). This allows for the definition of $30^3 = 27000$ different triplets comprising a P-wave, S-wave and radar wave velocity model each.

These triplets shall be transformed into a fuzzy domain. This requires the definition of nodes in the 3D parameter space spanned by the P-wave, S-wave and Radar-wave velocity tomograms. The number of nodes to be used equals the number of tomographic grid cells for which calibration data are available. In our example, calibration logs are available at $x = 2.75$ m and $x = 8.0$ m over the full vertical extension of the model area (see Figures 1 and 2).

Therefore, each log provides calibration information for 23 grid cells, which defines the number of nodes used in the parameter space in our example. Figure 6 shows the parameter space spanned by the tomograms and the arrangement of the nodes. Details about suitable node arrangement are given by Paasche (2017). For the given nodes, the spatial variability of the velocity tomograms forming a triplet can be integrated and described by fuzzy sets. These are dimensionless numbers summing to unity for each grid cell. They describe the spatial variability of all three tomograms in a triplet independently from their actual physical velocity values.

Following option 1 in Paasche (2017) all fuzzy sets are then commonly calibrated with a target parameter log by solving an over-determined system of linear equations in a least-squares sense. The found prediction model consisting of a target parameter value for each node is applied to the fuzzy set description of each triplet to achieve different 2D prediction scenarios of the target parameter distribution. Note, to avoid inference of target parameter properties for one material from other materials, we recommend to produce a (pseudo-)lithological plot, prior to parameter inference. For example, Linder et al. (2010) have clustered Radar, P-wave and S-wave velocity tomograms to achieve a zonal model outlining major subsurface units at our test site. For all units calibration data are available in our study, since all found zones are penetrated by the logs. In cases where no calibration data are available for some zones, these zones can be excluded from the parameter inferences analogue to Hachmöller and Paasche (2013).

Since 50 logging samples per grid cell are available for calibration, we repeatedly calibrate the fuzzy sets. In doing so, we want to avoid prior upscaling the logging data to a half-meter discretization corresponding to the vertical side lengths of the tomographic grid cells in our example. Instead, we consider one log reading per grid cell. In doing so, we first calibrate the fuzzy sets using the uppermost log readings falling into a grid cell (Figure 3b). After this, we

repeat calibration considering the second-uppermost samples per grid cell. We continue with this procedure until we perform calibration employing all lowermost readings falling into a grid cell. This ensures that all log readings contribute equally to the calibration procedure and scale differences between tomographic parameterization and logging data sampling can be considered in the probabilistic parameter inference procedure.

3.2 Discovery science approach - results achieved when considering all available information

First, we apply the methodology sketched in Figure 5 to achieve scenarios of 2D tip resistance distribution. We consider all available tomographic data sets and all 30 tomograms per data set. This results in 27000 triplets of tomograms to be converted into 27000 fuzzy membership matrices for the node arrangement shown in Figure 6. The tip resistance log acquired at $x = 2.75$ m is considered for calibration. We repeatedly calibrate 50 times and achieve 1350000 2D scenarios of tip resistance or 1350000 tip resistance values per tomographic grid cell. Figure 7a shows these tip resistance models in terms of relative frequency information for each grid cell. A perceptually balanced colormap has been used to ensure optimal contemplation free of visual artifacts (Smith and van der Walt 2015). The 2D model area is shown as series of laterally neighbored 1D panels (Figure 3a). Available tip resistance values per grid cell have been binned into 100 classes equally dividing the range between 0.1 and 50 MPa. Values exceeding the upper and lower limits of the considered range are binned in the highest and lowest bins, respectively. The inferred tip resistance models are in good coincidence with the log used for calibration. The log at $x = 2.75$ m coincides largely with high relative frequencies and is fully inside the range of predicted tip resistance. However, the second available tip resistance log at $x = 8.0$ m appears to be systematically higher than the most likely predictions indicated by high relative frequencies. The log is close to the upper limit of the predicted range and partly exceeds the prediction range. It appears like a

systematic mismatch either founded in a systematically different calibration of the CPT tool or by the presence of a systematic lateral change in ground conditions not captured by the available three tomographic data sets.

When repeating the prediction employing the tip resistance log measured at $x = 8.0$ m for calibration, we achieve again 1350000 2D scenarios of tip resistance illustrated in Figure 7b. Now, the log measured at $x = 2.75$ m is systematically lower than the most likely tip resistance values whereas the log measured at $x = 8.0$ m now matches much better the most likely predicted values and falls within the predicted range. This confirms again that for each log an acceptable match can be found between the tomograms and the individual log used for calibration. However, some systematic mismatch in lateral direction exists. The available database does not allow to further specify the source for this mismatch, which is either founded in missing information (lateral changes not present in the available tomograms) or systematic errors in the logging data. Since the employed prediction technique is purely data-driven and not limited to the consideration of geophysical and geotechnical data combinations of particular and theoretically founded interrelations, incorporation of additional tomograms emanating from a fourth physically different tomographic data sets could potentially result in improved prediction performance, if such tomogram would be able to capture potentially present lateral variations not captured by the three available tomographic data sets. Note, Asadi et al. (2017) applied an Artificial Neural Networks (ANNs) based probabilistic data-driven parameter inference approach to this database and similarly observed a mismatch between the logs at $x = 2.75$ m and $x = 8.0$ m and the available tomograms. Hence, we do not consider this mismatch as methodological flaw of the employed data-driven inference approach relying on fuzzy sets.

Given the current problem of lateral mismatch, the best and at the same time most conservative prediction possible with the given database is to aggregate the prediction

scenarios shown in Figures 7a and b. In Figure 7c we show the relative frequency information based on all 2700000 computed tip resistance scenarios calibrated either by the log measured at $x = 2.75$ m or $x = 8.0$ m. Relative frequencies show clearly bi-modal distributions at depths below 8 m illustrating the systematic difference in predictions when calibrating with different logs. Instead of equally considering predictions over the full 2D area like done here, one could inversely weight the prediction contributions to the relative frequency according to the lateral distance of a grid cell from the calibration positions at $x = 2.75$ m and $x = 8.0$ m. Thus, it would be possible to achieve relative frequency distributions that are uni-modal at the calibration positions and bi-modal elsewhere. However, the weighting would impose some subjective assumptions about the behavior of the lateral mismatch which can currently only be guessed.

Next, we repeat the procedure for inference of 2D probabilistic sleeve friction distributions. The results are shown in Figure 8a based on 2700000 2D sleeve friction scenarios achieved by individual calibration at $x = 2.75$ m and $x = 8.0$ m. Again, at depths below 8 m bi-modal distributions are found indicating a similar discrepancy in lateral direction between logging data and tomograms as achieved when inferring tip resistance. The logging data are fully in the prediction range and even prominent anomalies of slightly more than 1 m vertical direction in the logging data (e.g., at $x = 8.0$ m and depth around 11 m) are reflected by changes in the tomogram ensembles, here indicated by increased prediction ranges at these depth for grid cells laterally centered from 7.5 m to 9.5 m. Such anomalies are close to the tomographic resolution limit and not present in all tomograms of an ensemble. This could for instance indicate the presence of a lens structure of approximately 1 m thickness and a lateral extension of approximately 2-3 m. Here, tomograms and log mutually confirm each other in their information content.

Figure 8b shows the prediction results for dielectric permittivity based on 2700000 2D scenarios. Here, relative frequency distributions are uni-modal indicating good fit of tomograms and logs also in vertical direction. This is different from the findings resulting from CPT data prediction. However, a different tool has been used and an electrical parameter has been measured. Error components may be different from the CPT data thus resulting in the missing of a systematic shift between both dielectric logs. We interpret this finding as weak indicator that the source for the lateral misfit between tomograms and CPT logging data may be rather on the side of the logging data than in missing information not captured by the tomographic data sets. A number of prominent peaks in the logging data caused by layers of a few centimeters or decimeters thickness exceed the upper limits of the predicted range. However, such thin layers or lenses are clearly beyond the resolution limits of the geophysical tomograms and cannot be brought into compliance with any tomographically imaged contrasts and hence not predicted satisfactorily. This is not a weakness of the parameter inference technique but rather a physical limit of the spatial resolution of the tomographic experiments.

4 Hypothesis-driven science - the classical transfer function concept

Many explicitly formulated deterministic transfer functions have been developed to convert geotechnical parameters, e.g., sleeve friction achieved by cone penetration measurements, into parameters achieved from geophysical experiments, e.g., seismic S-wave velocity. The development of such functions builds on physical insight as well as empirical testing. For example, transfer functions are available for linking S-wave velocity with sleeve friction and/or tip resistance, parameters that are inherently mechanic, which reflects physical insight, but no linkage function is known for explicitly relating sleeve friction or tip resistance with radar-wave velocity propagation, which is based on electromagnetic subsurface properties.

The explicitly formulated link is often based on empirical analyses of laboratory measurements employing a limited number of material samples. The derived equations build on simple and unique regression functions, e.g., linear or exponential, between the linked quantities, and contain a number of empirical regression or tuning parameters. To keep the number of parameters low, the applicability of the derived transfer functions is usually limited to materials of similar geologic evolution and deposition history, e.g., Holocene sands or Quaternary soils. At our test site, sediments consist mainly of sand and gravel deposited after the last glaciation dating approximately 22 000 yBP (Schmelzbach et al. 2011).

We employ the deterministic transfer functions listed in Table 1 to convert the available sleeve friction and tip resistance logs into 1D S-wave velocity logs and superimpose the results on the corresponding 1D S-wave velocity panels extracted from Figure 4b (Figure 9). Note, in doing so we convert the retrospectively drawn hypotheses, that these functions and corresponding tuning parameters were found useful by their authors when working at other sites than we do, into an a priori valid hypothesis at our site. Since the direct push log at $x = 8$ m is positioned at the right and left boundary of the tomographic grid cells centered at $x = 7.5$ m and $x = 8.5$ m, we overlay the converted logging information on both velocity panels (Figures 9b and c). The S-wave velocity information achieved by sleeve friction and tip resistance conversion resembles the major trends in the tomographically imaged S-wave velocity models. Velocity values differ slightly and define a range somewhat larger than, and not fully superimposing, those resulting from tomographic reconstruction ambiguity. It empirically reflects the inherent uncertainty when striving to determine a transfer function describing the relation between different parameters realistically and independent from specific site conditions and measurement scales. When employing the transfer functions in the forward direction, i.e., converting sleeve friction and tip resistance into S-wave velocity, most of them give satisfactory results.

However, 1D S-wave velocity profiles can be rather directly measured by geophysical experiments in boreholes or by direct push technology (Hunter et al. 2010). S-wave velocity can be even tomographically imaged by geophysical experiments in 2D or 3D using cross-borehole or direct-push based tomography (Paasche et al. 2013) as well as surface-based geophysical experiments (Hunter et al. 2010). Geotechnical experiments or tomography resulting for example in 2D or 3D images of sleeve friction are not known. Hence, it would be appealing to employ the available transfer functions linking sleeve friction and S-wave velocity inversely, i.e., to convert available S-wave velocity information into sleeve friction. So we rearrange the transfer functions in Table 1 except for those of Sykora and Stokoe (1983) so that they can be used to convert the tomographically reconstructed 2D S-wave velocity into 2D sleeve friction distributions. We apply the transfer functions to all 30 S-wave velocity models available and thus transduce tomographic ambiguity of S-wave velocity reconstruction into uncertainty of sleeve friction determination. Figure 10 shows the 2D sleeve friction models as series of laterally neighbored 1D panels. Measured sleeve friction at $x = 2.75$ m and $x = 8.0$ m is superimposed on the panels laterally centered at $x = 2.5$ m and $x = 7.5$ m and $x = 8.5$ m, respectively. The quality of 2D sleeve friction models is poor for all considered equations except for those of Piratheepan (2002) for Holocene soils (Figure 10a). 2D sleeve friction prediction results in Figures 10b and e suffer very broad ranges making a trustworthy and useful specification of most likely sleeve friction values almost impossible. The large ranges are a result of utilizing very low exponents in the transfer functions to keep the importance of sleeve friction low when calculating S-wave velocity. When employing this equation in inverse direction this results in high sensitivity of the calculated sleeve friction on the provided S-wave velocities. Hence, these equations consider different sensitivities between sleeve friction f_s and S-wave velocity v_s when either imaging $f: f_s \rightarrow v_s$ or $f: v_s \rightarrow f_s$. Physically it is hard to argue why such assumption should be considered realistic, since in

general the link between both parameters is described by a deterministic and unique transfer equation.

Results in Figures 10c and d show very narrow ranges but are systematically resulting in sleeve frictions lower than those measured. This is in correspondence with the results achieved when using this transfer functions in the forward direction (Figure 9). It indicates that these transfer functions are generally not suitable to describe the realistic link between S-wave velocity and sleeve friction at our site sufficiently well. The larger exponents employed to control sleeve friction importance result in relatively high sensitivity of S-wave velocities on sleeve frictions when employing them in the forward direction. This is also reflected by large amplitudes of the cyan and magenta lines in Figure 9. In the inverse direction, the larger exponents result in reduced sensitivity of the calculated sleeve friction on S-wave velocity, which is reflected by narrow ranges in Figures 10c and d. Note, the variability in sleeve friction prediction is only induced by the tomographic ambiguity as described by the available 30 S-wave velocity tomograms. Uncertainty of regression parameter determination, e.g., scalar exponents or factors in the equations in Table 1, is not considered (Tronicke and Paasche, 2017).

Based on the results in Figure 10, one may come to the conclusion, that a simple consideration of a retrospectively drawn hypothesis about existing inter-relations between sleeve friction and S-wave velocity should not simply be turned into an a priori valid hypothesis at our site. So, we have to give up at least on some information provided by the authors of the transfer functions found in the literature and replace it by site-specific (data-driven) information. For example, this can be done by discarding the published values of the tuning parameters (which was analogue to acceptance of a remote calibration off the current field site) and recalibrating the linkage functions with the data available at our site. When doing this, it is advisable to make sure that recalibration can be done for all relevant

subsurface units in the survey area, e.g., by producing a (pseudo-)lithology plot as suggested in section 3.1 for the discovery science approach. If unrecognized material variations occur in the survey area away from the locations where calibration is possible, then, analogue to the discovery science approach, involuntarily poor predictions based on wrong calibration may be achieved for these ground regions. Naturally, this should improve prediction results compared to those shown in Figure 10. Nevertheless, when doing such recalibration, we still assume that the general nature of the linkage function found retrospectively by analyzing data from other locations may be valid at our site. Appendix A illustrates, that this may sometimes hold and sometimes not. If the function itself is unsuitable, even a re-calibration may not provide good results, despite some minor improvements. In such cases, even the general nature of the provided transfer function has to be discarded, when turning a retrospectively drawn hypothesis into an a priori valid hypothesis at our site. In such cases one would be left with the task to find a new linkage function and corresponding tuning parameters, and this can be generally done at every site using a discovery science approach, too, albeit prediction models may not be represented by simple deterministic transfer equations.

5 Considerations about the optimal database for target parameter inference constrained by geophysical tomograms

When utilizing the transfer functions in section 4, we ignored the presence of the P-wave and radar wave tomograms when striving to achieve 2D sleeve friction distributions. This has been done based on the hypothesis that this information is only weakly or not linked to the target parameter of interest or of no relevance since it does hypothetically not provide additional information beyond those included in the favored S-wave tomograms. This decision is based on observations made by others in the past and their conclusions or hypothesis drawn retrospectively from their observations, e.g., as summarized by Schön

(1994). Building on their observations and retrospective hypothesis we a priori hypothesized that a known deterministic equation derived from previous observations and experiences made by others may be transferable to our site. In our experiment, such an a priori hypothesis severely limits the scope of our data analysis, potential findings and quality of results. In many experiments even the experimental effort in geophysical or geotechnical site characterization, e.g., by solely collecting S-wave data when striving for sleeve friction or vice versa is limited based on the a priori consideration of retrospectively drawn hypotheses, for example with regard to spatial sample density or selection of data types expected to be particularly useful. However, already Francis Bacon argued 1620 in his *Novum Organum*, that scientific knowledge should rely on data rather than preconceived notions. Deductive reasoning is potentially limited because setting a premise in advance of an experiment would constrain the reasoning so as to match that premise (Mazzocchi, 2015). This would be clear argumentation against turning a retrospectively drawn hypothesis about the nature of parameter inter-linkage (e.g., linear, exponential, etc.) and chosen tuning parameters into an a priori valid hypothesis at the field site under investigation. It should be noted, that also ensemble modeling approaches (e.g., Seni and Elder, 2010), e.g., realized by combining different petrophysical transfer functions, even when accounting for their tuning parameter uncertainty (e.g., Tronicke and Paasche, 2017), and their results in a simple or more sophisticated stacking approach (e.g., Wolpert, 1992), do not ensure to fully overcome preconceived notions potentially present in the considered transfer functions to predict the target parameters.

Alternatively, in section 3 we followed a slightly different strategy. We still use a priori hypotheses, e.g., we believe that data at a point can be represented by a combination of other types of data at the same point, or more generally, that data providing information about the same ground must be compliant within their observational accuracy. But we strive to collect and mine a broad database considering even data, which have never indicated their particular

usefulness for the prediction of our target parameter, e.g., for the combination of radar velocity tomograms and sleeve friction no transfer function has been developed (hypothesized) in the past. Here, we follow a more discovery science (Aebersold et al. 2000) data exploration approach striving to keep the limiting effect of a priori hypotheses on the potential results small. This stands in the tradition of data mining and knowledge discovery strategies striving to find solutions that are solely driven by the available set of observations and not constrained by a priori consideration of hypotheses retrospectively drawn from previous observations made at other locations and times (e.g., Gaber 2010). Ideally, such approaches could reveal new insights into inter-parameter relations so far overseen in the hypothesis-based approaches, albeit they might not automatically generate new (theoretic) petrophysical models.

In the following we repeat the inference of 2D sleeve friction using the data-driven approach sketched in Figure 5. There is no proven reason why we should expect to find radar tomograms useful when predicting tip resistance or sleeve friction, which are both inherently relying on mechanic ground properties like seismic wave propagation. We repeat our prediction of sleeve friction considering only the P-wave and S-wave tomograms and exclude the radar tomograms hypothesizing that they may be of no importance for achieving high-quality predictions of sleeve friction. Note, this hypothesis can be seen like a filter on our database. P- and S-wave velocity tomograms allow for 900 pairwise combinations of tomograms leading us to 90000 2D sleeve friction scenarios when individually calibrating the fuzzy patterns with the logs measured at $x = 2.75$ m and $x = 8.0$ m, respectively. The results are shown in Figure 11a. The prediction ranges are now reduced, which is in full compliance to expectations resulting from error propagation. If we exclude the set of radar tomograms, then no imaging ambiguity of this data set is transduced into the probabilistic inference of the target parameter. This would be good argumentation to follow the classical approaches of considering only the data sets hypothetically expected to be of particular usefulness.

In most parts of the 2D area the prediction ranges are systematically narrowed compared to those in Figure 8a but preserved the relative differences in narrow and broad ranges.

However, the prominent anomaly present in the log at $x = 8.0$ m and depth = 11 m is now clearly exceeding the prediction range. There is not even a hint in the prediction that this anomaly may be reality, since the seismic tomograms do not show any indication for the presence of a lens-like anomaly in this region. The significantly increased prediction range in Figure 8a at this location is now fully missing. Here, the radar tomograms contributed unique information not captured by any of the seismic tomograms but captured by the sleeve friction log. This may indicate the presence of a link between electromagnetic radar wave velocity propagation and sleeve friction which is an inherently mechanic parameter. Physically we have no argumentation for the presence of such a link and it is beyond the scope of this work to speculate on this in terms of physical or petrophysical theory, but the data-driven discovery science approach revealed such a link at least here in our case study at this field site.

This has some consequences when working with data and mining them for compliancy of data sets imaging the same ground. When a priori limiting our database to data sets expected to be of particular importance we can expect to achieve good results in many cases, albeit not perfect results. Care must be taken when turning retrospectively drawn hypothesis into a priori valid hypothesis which is to some degree always the case when using deterministic transfer functions, recalibrated or not. However, when just regarding all possible or available data sets and considering them just as information about the ground free from hypothesis about their particular suitability for the desired prediction task, we can hope to achieve good predictions and achieve hints on inter-relations between data so far not recognized. Whether these unexpected links are results of superposition of measurement errors or true unrecognized hints that can be physically validated by further experiments and theoretical advancements cannot be finally answered with the experiment discussed here.

In Figure 11b we show dielectric permittivity prediction based on the radar tomograms alone. We employed a relation theoretically derived from Maxwell's equations and constitutive equations between electromagnetic wave propagation and dielectric permittivity. For simplicity, we employ here a simple low-loss approximation which can be proven to be of sufficient accuracy to be regarded as almost error free with regard to experimental errors resulting from nowadays physical experimental accuracy limits (Butler 2005). Here, we converted the 30 radar velocity models into dielectric permittivity and achieve a prediction quality that is excellent. The ranges are even reduced compared to those in Figure 8b (benefits of reduced error propagation by considering less data sets in the prediction procedure) but the most likely predictions are of similar quality like those achieved when employing all available tomograms. This proves that our inference approach learning the link between all considered tomograms integrated in the fuzzy set delivers good results regardless of the considered database. However, employing such theoretical and physically proven transfer functions can likely result in good predictions and may justify keeping other data sets out of the prediction procedure to keep error propagation low. However, empirical transfer functions, or theoretical models striving to describe inter-parameter relations in the sense of strongly idealized models, e.g. building on rounded grains or materials disturbed when carried from the field to the laboratory, may never reach this accuracy, and data-driven inference approaches may be considered as flexible and powerful alternatives.

6 Conclusions

We employed classical deterministic transfer functions and a recently developed data-driven inference approach suitable to predict spatially continuous distributions of sparsely measured target parameters constrained by geophysical tomography. Since we employ a fully non-linear tomographic imaging strategy resulting in ensembles of equivalent geophysical tomograms

we can transduce tomographic imaging ambiguity into the probabilistic inference of the target parameter distributions. Furthermore, the employed data-driven parameter inference approach allows even for consideration of differences in spatial resolution or probed volume between geophysical tomographic imaging and borehole or direct-push logging data and is applicable to any combination of tomograms and logging data.

While empirically derived transfer functions may result in acceptable conversion of a given parameter into another, they may suffer reduced applicability when being used for parameter conversion in the inverse direction. Physically, this behavior of direction-dependent inter-parameter sensitivity is hard to explain. It may be even necessary to give up on some information of empirically derived transfer functions when turning them into a priori valid hypothesis. Decisions on that require careful analyses. Results achieved by theoretical transfer functions of general applicability are considered more robust and thus a better choice. Data-driven approaches provide generally good inference results, and are flexible with regard to the database to be considered for parameter inference.

However, error propagation results in increasing prediction ranges when using the data-driven inference approach, which may be a disadvantage compared to theoretically founded transfer functions. Thus, before deciding to employ theoretical or data-driven approaches for probabilistic inference of spatially continuous target parameter distributions constrained by ill-posed geophysical tomographic imaging critical analysis of eventually present hypothetical assumptions may impact the result of the predicted parameter distributions. In many practical situations the data-driven approach may be applicable where theoretical transfer functions are lacking.

A Appendix

Based on the results shown in Figure 10 we decided to test the performance of transfer functions when recalibrating them. Recalibration can be done formulating the transfer function in the forward or in the inverse direction. Since we are interested in sleeve friction prediction in our examples, it would be natural to do the calibration in the inverse direction, e.g., as done when using the discovery science approach discussed in section 3. Analogue to the calibration of fuzzy sets, we repeatedly calibrate each transfer function using fifty calibration sets drawn from each log. Tuning parameters are determined in a least squares sense averaging over all available seismic velocity tomograms.

Figure A1 shows the calibration results for the transfer function suggested by Mayne (2006). Two tuning parameters c_1 and c_2 (see Table 1) had to be recalibrated. In Figures A1a and b the found values are shown for each log and forward and inverse formulations of the transfer function. The new calibration parameters vary for each calibration set and are off the values initially suggested by Mayne (2006). However, prediction performance slightly improved. In the case of forward calibration (Figure A1c) the prediction of sleeve friction scatters and problems exist to fit the general shape of the given logs. At depths below 14 m predictions indicate values larger than those measured by the logs, which leaves doubts on the general suitability of the (retrospectively hypothesized) transfer function for our data set. In the inverse direction (Figure A1d), the most likely predictions approximately match the mean value of the measured logs, which is an improvement compared to the results in Figure 10c. However, minor variations in the logging data cannot be accommodated by the transfer function, still illustrating a relative insensitivity to minor S-wave variations.

In Figure A2 we show the prediction results for Hegazy and Mayne's (1995) transfer function for Quaternary sands when using it in inverse direction for calibration. This equation has three tuning parameters and the recalibration settings achieved are shown in Figures A2a-c. Again, the found parameters vary for each calibration set and are clearly different from the

values initially suggested by Hegazy and Mayne (1995; see Table 1). Prediction performance (Figure A2d) is significantly improved compared to the results shown in Figure 10e.

Prediction ranges are clearly reduced and the expression is highly flexible and allows for accommodation of most minor variations observed in the logging data. However, there is no ability to accommodate or predict the prominent anomaly around the depth of 11 m for the log acquired at $x = 8$ m,

7 References

- Aebbersold, R., Hood, L.E. and Watts, J.D. 2000. Equipping scientists for the new biology. *Nature Biotechnology* **18**, 359.
- Asadi, A., Dietrich, P. and Paasche, H. 2017. Spatially continuous probabilistic prediction of sparsely measured ground properties constrained by ill-posed tomographic imaging considering data uncertainty and resolution. *Geophysics* **82**, V149-V162.
- Asadi, A., Dietrich, P. and Paasche, H. 2016. 2D probabilistic prediction of sparsely measured earth properties constrained by ill-posed geophysical tomographic imaging. *Environmental Earth Sciences* **75**, 1487.
- Aster, R.C., Borchers, B. and Thurber, C.H. 2005. *Parameter Estimation and Inverse Problems*. Academic Press.
- Bosch, M., Mukerji, T. and Gonzalez, E. 2010. Seismic inversion of reservoir properties combining statistical rock physics and geostatistics: a review. *Geophysics* **75**, A165-A176.
- Breiman, L. and Friedman, J.H. 1985. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association* **80**, 580-598.
- Butler, D.K., 2005. *Near-Surface Geophysics*. Society of Exploration Geophysicists.
- Cassiani, G., Böhm, G., Vesnaver, A. and Nicolich, R. 1998. A geostatistical framework for incorporating seismic tomography auxiliary data into hydraulic conductivity estimation. *Journal of Hydrology* **206**, 58-74.
- Chen, J., Hubbard, S. and Rubin, Y. 2001. Estimating the hydraulic conductivity at the south oyster site from geophysical tomographic data using Bayesian techniques based on the normal linear regression model. *Water Resources Research* **37**, 1603-1613.

- Dafflon, B., Irving, J. and Holliger, K. 2009. Simulated-annealing based conditional simulation for the local-scale characterization of heterogeneous aquifers. *Journal of Applied Geophysics* **68**, 16-24.
- Dannowski, G. and Yaramanci, U. 1999. Estimation of water content and porosity using combined radar and geoelectrical measurements. *European Journal of Environmental and Engineering Geophysics* **4**, 71-85.
- Dubreuil-Boisclair, C., Gloaguen, E., Marcotte, D. and Giroux, B. 2011. Heterogeneous aquifer characterization from ground-penetrating radar tomography and borehole hydrogeophysical data using nonlinear Bayesian simulations. *Geophysics* **76**, J13-J25.
- Ezzedine, S., Rubin, Y. and Chen, J. 1999. Bayesian method for hydrogeological site characterization using borehole and geophysical survey data: theory and application to the Lawrence Livermore National Laboratory Superfund Site. *Water Resources Research* **35**, 2671-2683.
- Fernández Martínez, J. L., Fernández Muñoz, M. Z. and Tomkins, M. J. 2012. On the topography of the cost functional in linear and nonlinear inverse problems. *Geophysics* **77**, W1-W15.
- Fernández-Martínez, J. L., Fernández-Muñoz, M. Z., Pallero, J. L. G. and Pedruelo-González, L.M. 2013. From Bayes to Tarantola: New insights to understand uncertainty in inverse problems. *Journal of Applied Geophysics* **98**, 62-72.
- Friedel, M.J., Scott, D.F. and Williams, T.J. 1997. Temporal imaging of mine-induced stress change using seismic tomography. *Engineering Geology* **46**, 131-141.
- Friedel, S. 2003. Resolution, stability and efficiency of resistivity tomography estimated from a generalized inverse approach. *Geophysical Journal International* **153**, 305-316.
- Gaber, M.M. 2010. *Scientific Data Mining and Knowledge Discovery*. Springer.

- Gloaguen, E., Chauteau, M., Marcotte, D. and Chaouis, R. 2001. Estimation of hydraulic conductivity of an unconfined aquifer using cokriging of GPR and hydrostratigraphic data. *Journal of Applied Geophysics* **47**, 135-152.
- Hachmöller, B. and Paasche, H. 2013. Integration of surface-based tomographic models for zonation and multimodel guided extrapolation of sparsely known petrophysical parameters. *Geophysics* **78**, EN43-EN53.
- Hegazy, Y.A. and Mayne, P.W. 1995. Statistical correlations between Vs and cone penetration data for different soil types. *Proceedings of the International Symposium on Cone Penetration Testing*, Volume II, Linköping, 173-178.
- Hubbard, S., Chen, J., Peterson, J., Majer, E., Williams, K., Swift, D., Mailliox, B. and Rubin, Y. 2001. Hydrogeological characterization of the D.O.E. bacterial transport site in Oyster, Virginia using geophysical data. *Water Resources Research* **37**, 2431-2456.
- Hunter, J.A., Motazedian, D., Crow, H.L., Brooks, G.R., Miller, R.D., Pugin, A. Pullan, S.E. and Xia, J. 2010. Near-surface shear-wave velocity measurements for soft-soil earthquake-hazard assessment: some Canadian mapping examples, in: Miller, R.D., Bradford, J.H., Holliger, K. (Eds.), *Advances in Near-Surface Seismology and Ground-Penetrating Radar*. Society of Exploration Geophysicists, pp. 339-359.
- Kaufmann, O., Deceuster, J. and Quinif, Y. 2012. An electrical resistivity imaging-based strategy to enable site-scale planning over covered palaeokarst features in the Tournaisis area (Belgium). *Engineering Geology* **133**, 49-65.
- Kennedy, J. and Eberhart, R. 1995. Particle swarm optimization. *Proceedings of the IEEE International Conference on Neural Networks*, IEEE, 1942-1948.

- Linder, S., Paasche, H., Tronicke, J., Niederleithinger, E. and Vienken, T. 2010. Zonal cooperative inversion of crosshole p-wave, s-wave, and georadar traveltime data sets. *Journal of Applied Geophysics* **72**, 254-262.
- Lorenzo, J.M., Hicks, J. and Vera, E.E. 2014. Integrated seismic and cone penetration test observations at a distressed earthen levee: Marrero, Louisiana, U.S.A.. *Engineering Geology* **168**, 59-68.
- Lunne, T., Robertson, P.K. and Powell, J.J.M. 1997. *Cone Penetration Testing in Geotechnical Practice*. Blackie Academic Professional.
- Mayne, P.W. 2006. In situ test calibrations for evaluating soil parameters. in: Tan, T.S., Phoon, K.K., Hight, D.W., Leroueil, S. (Eds.), *Characterisation and Engineering Properties of Natural Soils*, Volume III. Taylor and Francis, pp. 1601-1652.
- Mazzocchi, F. 2015. Could big data be the end of theory in science?. *EMBO reports* **16**, 1250-1255.
- Mola-Abasi, H., Dikmen, U. and Shooshpasha, I. 2015. Prediction of shear-wave velocity from CPT data at Eskisehir (Turkey), using a polynomial model. *Near Surface Geophysics* **13**, 155-167.
- Niederleithinger, E. 2009. The BAM site for non-destructive testing methods (NDT) in civil engineering. *International Foundation Congress & Equipment Expo 2009*, Expanded abstracts. 8 pp.
- Niederleithinger, E. 2015. 3G – Geophysical methods delivering input to geostatistical methods for geotechnical site characterization, in: Schweckendiek, T., van Tol, A.F., Pereboom, D., van Staveren, M.Th., Cools, P.M.C.B.M. (Eds.), *Geotechnical Safety and Risk V*. IOS Press, pp. 871-879.

- nmeth0409-237, 2009. Defining the scientific method. Editorial, *Nature Methods* **6**, 237. doi: 10.1038/nmeth0409-237.
- Paasche, H., Tronicke, J., Holliger, K., Green, A.G. and Maurer, H. 2006. Integration of diverse physical-property models: subsurface zonation and petrophysical parameter estimation based on fuzzy c-means cluster analyses. *Geophysics* **71**, H33-H44.
- Paasche, H., Rumpf, M., Hausmann, J., Fechner, T., Werban, U., Tronicke, J. and Dietrich, P. 2013. Advances in acquisition and processing of near-surface seismic tomographic data for geotechnical site assessment. *First Break* **31**, 59-65.
- Paasche, H. and Tronicke, J. 2014. Non-linear joint inversion of tomographic data using swarm intelligence. *Geophysics* **79**, R133-R149.
- Paasche, H. 2017. Translating tomographic ambiguity into the probabilistic inference of hydrologic and engineering target parameters. *Geophysics* **82**, EN67-EN79.
- Pham, D. L. 2001, Spatial models for fuzzy clustering. *Computer Vision and Image Understanding* **84**, 285-297.
- Piratheepan, P. 2002. *Estimating Shear-Wave Velocity from SPT and CPT Data*. M.Sc. Thesis, Clemson University.
- Robertson, P.K. 2009. Interpretation of cone penetration tests – a unified approach. *Canadian Geotechnical Journal* **46**, 1337-1355.
- Ruggeri, P., Irving, J., Gloaguen, E. and Holliger, K. 2013. Regional-scale integration of multi-resolution hydrological and geophysical data using a two-step Bayesian sequential simulation approach. *Geophysical Journal International* **194**, 289-303.
- Rumpf, M. and Tronicke, J. 2014. Predicting 2D geotechnical parameter fields in near-surface sedimentary environments. *Journal of Applied Geophysics* **101**, 95-107.

- Schmelzbach, C., Tronicke, J. and Dietrich, P. 2011. Three-dimensional hydrostratigraphic models from ground-penetrating radar and direct-push data. *Journal of Hydrology* **398**, 235-245.
- Schön, J. 2004. *Physical Properties of Rocks*. Pergamon.
- Schwarzbach, C., Börner, R. U. and Spitzer, K. 2005. Two-dimensional inversion of direct-current resistivity data using a parallel, multi-objective genetic algorithm. *Geophysical Journal International* **162**, 685-695.
- Sen, M. K., and Stoffa, P.L. 2013. *Global optimization methods in geophysical inversion*. Elsevier.
- Seni, G., and Elder, J. 2010. *Ensemble methods in data mining: Improving accuracy through combining predictions*. Morgan and Claypool.
- Shahri, A.A., Malehmir, A. and Juhlin, C. 2015. Soil classification analysis based on piezocone penetration test data – A case study from a quick-clay landslide site in southwestern Sweden. *Engineering Geology* **189**, 32-47.
- Shinn, J.D., Timian, D.A., Morey, R.M., Mitchell, G., Antle, C.L. and Hull, R. 1998. Development of a CPT deployed probe for in situ measurement of volumetric soil moisture content and electrical resistivity. *Field Analytical Chemistry and Technology* **2**, 103-109.
- Smith, N. J. and van der Walt, S.. 2015. How we designed a new default colormap for Matplotlib (and you can too). Talk presented at Scipy 2015 conference, Austin, TX. Available at: <https://bids.github.io/colormap/> (accessed 22.12. 2016).
- Sykora, D.E. and Stokoe, K. H. 1983. Correlations of in-situ measurements in sands of shear wave velocity. *Soil Dynamics and Earthquake Engineering* **20**, 125-136.
- Taylor, J.R. 1996. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books.

- Tonni, L. and Simonini, P. 2013. Shear wave velocity as function of cone penetration test measurements in sand and silt mixtures. *Engineering Geology* **163**, 55-67.
- Tronicke, J. and Holliger, K. 2005. Quantitative integration of hydrogeophysical data: conditional geostatistical simulation for characterizing heterogeneous alluvial aquifers. *Geophysics* **70**, H1-H10.
- Tronicke, J., Paasche, H. and Böniger, U. 2012. Crosshole travelttime tomography using particle swarm optimization: A near-surface field example. *Geophysics* **77**, R19-R32.
- Tronicke, J. and Paasche, H. 2017. Integrated interpretation of 2D ground-penetrating radar, P-, and S-wave velocity models in terms of petrophysical properties: Assessing uncertainties related to data inversion and petrophysical relations. *Interpretation* **5**, T121-T130.
- Ward, S.H. 1994. *Geotechnical and Environmental Geophysics*, Volumes I, II, and III. Society of Exploration Geophysicists.
- Wolpert, D.H. 1992. Stacked generalization. *Neural Networks* **5**, 241-259.

Figure captions

Figure 1: Plan view of the Horstwalde field site. Tomographic planes are marked by black arrows with borehole 2 (BH2) equal to $x = 0$ m and BH5 equal to $x = 10.96$ m. Direct push (DP) injection points are at $x = 2.75$ m and $x = 8.0$ m, respectively (Linder et al. 2010).

Figure 2: (a) tip resistance q_c and (b) sleeve friction f_s logging data acquired by direct push cone penetration tests (CPT). (c) Dielectric permittivity ϵ_r logging data acquired using a direct push soil moisture probe (SMP).

Figure 3: (a) Sketch illustrating the visualization of a 2D geophysical velocity tomogram by means of 1D velocity-depth profiles. Dimensionality and grid cell arrangement match those used to invert all tomographic data sets discussed in this work. A single velocity panel corresponds to the tomographic velocity information of a column of grid cells. Lateral grid cell center positions are shown on top of each panel. (b) Sketch of logging data and tomographic grid cells for the logs measured at $x = 8$ m. Log readings are spaced 1 cm, whereas tomographic grid cells have 0.5 m vertical side lengths. Therefore 50 log readings are available per grid cell for calibration (Paasche, 2017, modified).

Figure 4: Results of different co-located geophysical 2D cross-borehole tomographic imaging experiments. The data have been inverted using a global-search optimization algorithm resulting in 30 different 2D tomographic models fitting the underlying data set equally well. Tomographic reconstruction ambiguity is discretely illustrated by the 30 equivalent models. (a) The black lines illustrate 30 P-wave velocity models fitting the underlying data equivalently well. (b) Illustration of 30 S-wave velocity models and (c) 30 Radar wave velocity models fitting the underlying data equivalently well.

Figure 5: Conceptual sketch of the processing flow used to transduce tomographic ambiguity into the probabilistic inference of geotechnical target parameter distributions

constrained by sparse logging data (modified, Asadi et al., 2017). For a detailed methodological description see Paasche (2017).

Figure 6: Orthogonal projections of the 3D parameter space spanned by the geophysical models after normalization to the interval [0 1]. Black symbols mark a sample in the parameter space defined by a P-, S-, and Radar wave velocity information in a collocated grid cell. All 27 000 possible triplets of the 30 P-wave, 30 S-wave, and 30 Radar velocity models are shown. Gray circles mark 23 nodes. 8 nodes have been manually positioned at the corner points of the 3D parameter space. The remaining ones have been randomly positioned.

Figure 7: Relative frequency information of (a) 1350000 realizations of 2D tip resistance distributions relying on all available geophysical models and the sparse tip resistance information acquired at DP1. (b) the same as in (a), but using the sparse tip resistance information acquired at DP2. (c) aggregation of the distributions shown in (a) and (b). Green lines show the available logging data (see Figure 2).

Figure 8: The same as in Figure 7c but for (a) sleeve friction and (b) dielectric permittivity.

Figure 9: Black lines are reproductions of the S-wave velocity panels in Figure 4b with grid cell center positions at (a) $x = 2.5$ m, (b) $x = 7.5$ m and (c) $x = 8.5$ m. Colored lines outline S-wave velocity information deduced from the CPT measurements (Figures 2a and b) using various deterministic transfer functions (Table 1). Data acquired at location DP1 (see Figure 1) are shown in (a). Data acquired at location DP2 are repeatedly shown in (b) and (c), since the DP2 location at $x = 8.0$ m is exactly between the tomographic grid cells laterally centered at $x = 7.5$ m and $x = 8.5$ m.

Figure 10: 2D sleeve friction information deduced from all 30 S-wave tomographic models (Figure 4b) after applying deterministic transfer functions (Table 1). The employed transfer functions have originally been developed for the determination of S-wave

velocities from CPT measurements and have been rearranged for our purpose. (a) Piratheepan (2002), all soils, Holocene. (b) Piratheepan (2002), sand, Holocene. (c) Mayne (2006), all soils, Quaternary. (d) Hegazy and Mayne (1995), all soils, Quaternary. (e) Hegazy and Mayne (1995), sand, Quaternary. Gray lines indicate sleeve friction data acquired by CPT measurements.

Figure 11: (a) The same as in Figure 8a but now derived considering only P- and S-wave tomograms and excluding the radar tomograms from the sleeve friction inference. Number of 2D sleeve friction scenarios underlying the relative frequencies is now limited to 90 000. (b) Dielectric permittivity achieved by converting the radar velocity tomograms using a theoretically founded conversion equation. 30 dielectric permittivity scenarios form the basis for the shown relative frequencies. The number of bins has been reduced by 50 % compared to Figure 8b given the low number of 2D dielectric permittivity scenarios. Green lines show the available logging data (see Figure 2).

Figure A1: Results achieved using the transfer function of Mayne (2006) when re-calibrating the tuning parameters (a) c_1 and (b) c_2 . Calibration has been done for 50 different sets of logging samples (for details see text) in forward (x symbols) and reverse (+ symbols) direction. Black symbols represent new tuning parameter when calibrating at $x = 2.75$ m. For tuning parameters originally suggested by Mayne (2006) see Table 1. Gray symbols represent new tuning parameter when calibrating at $x = 8$ m. (c) and (d) show the prediction results when calibrating the transfer equation in forward and reverse direction, respectively.

Figure A2: Results achieved using the transfer function of Hegazy and Mayne (1995) for Quaternary sands when re-calibrating the tuning parameters (a) c_1 , (b) c_2 , and (c) c_3 in reverse direction. Black and gray symbols depict results for calibration at $x = 2.75$ m

and $x = 8$ m, respectively. Solid lines correspond to the values originally suggested by Hegazy and Mayne (1995; see also Table 1).

Table caption

Table 1: Different deterministic transfer functions for converting sleeve friction f_s into seismic S-wave velocity v_s . y is the depth below ground surface. q_c is tip resistance. f_s and q_c are provided in kPa and y in m. The given values for tuning parameters c_i have been originally suggested by the authors of the transfer functions and are based on empirical analyses.

Figure 1

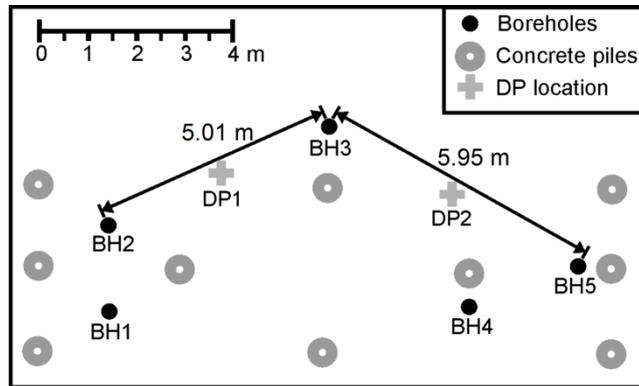


Figure 2

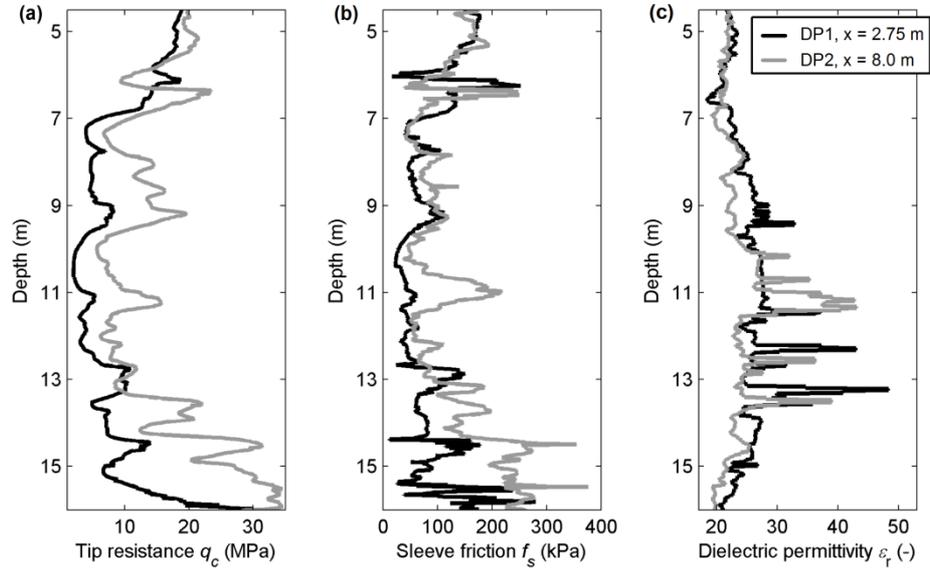


Figure 3

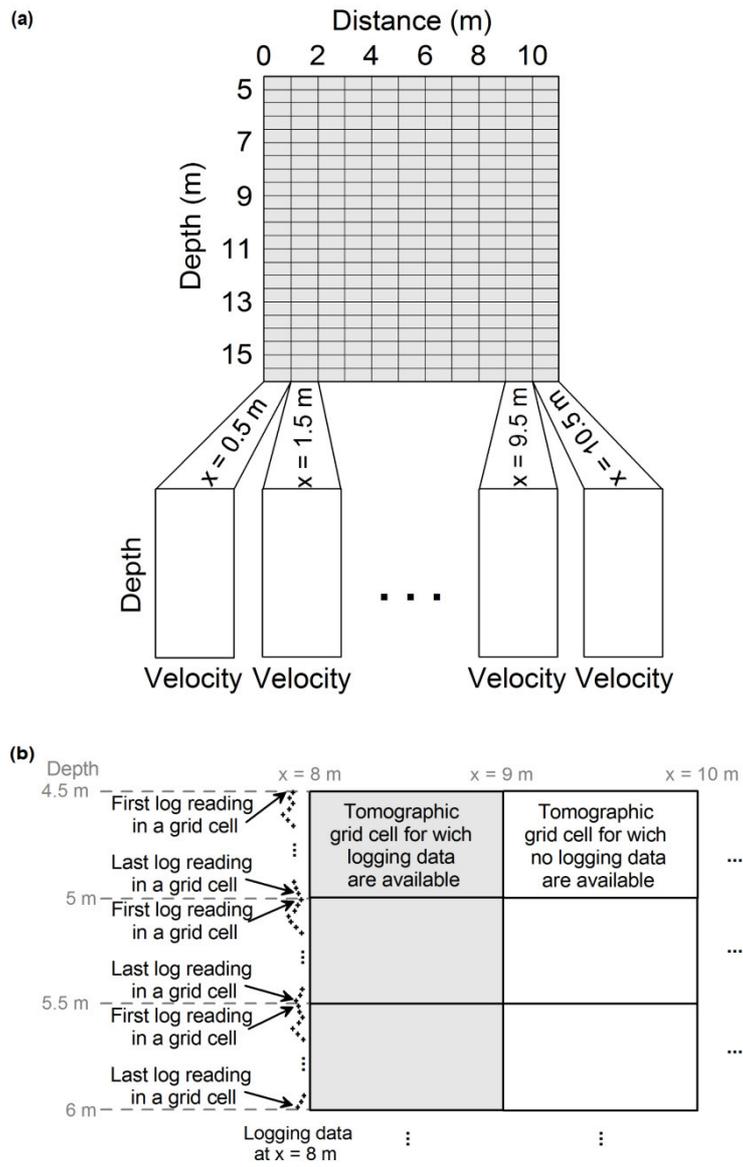


Figure 4

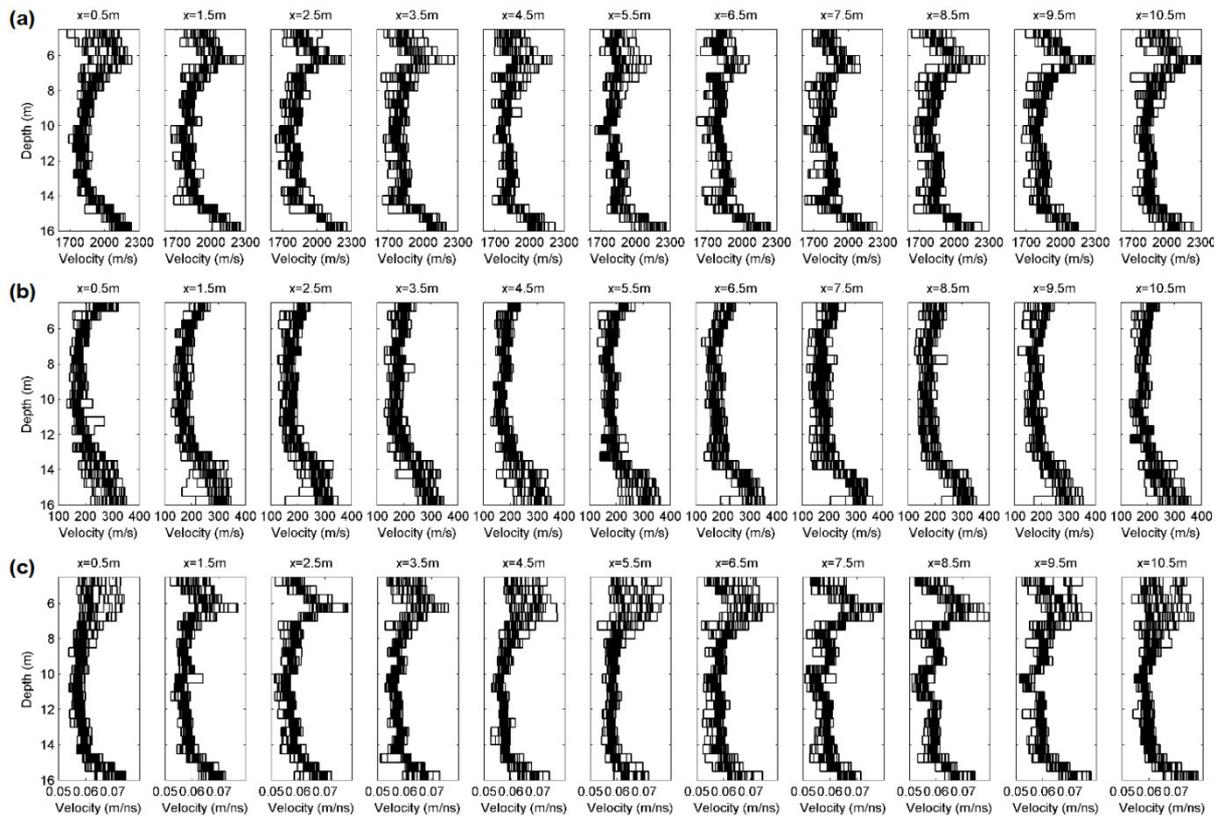


Figure 5

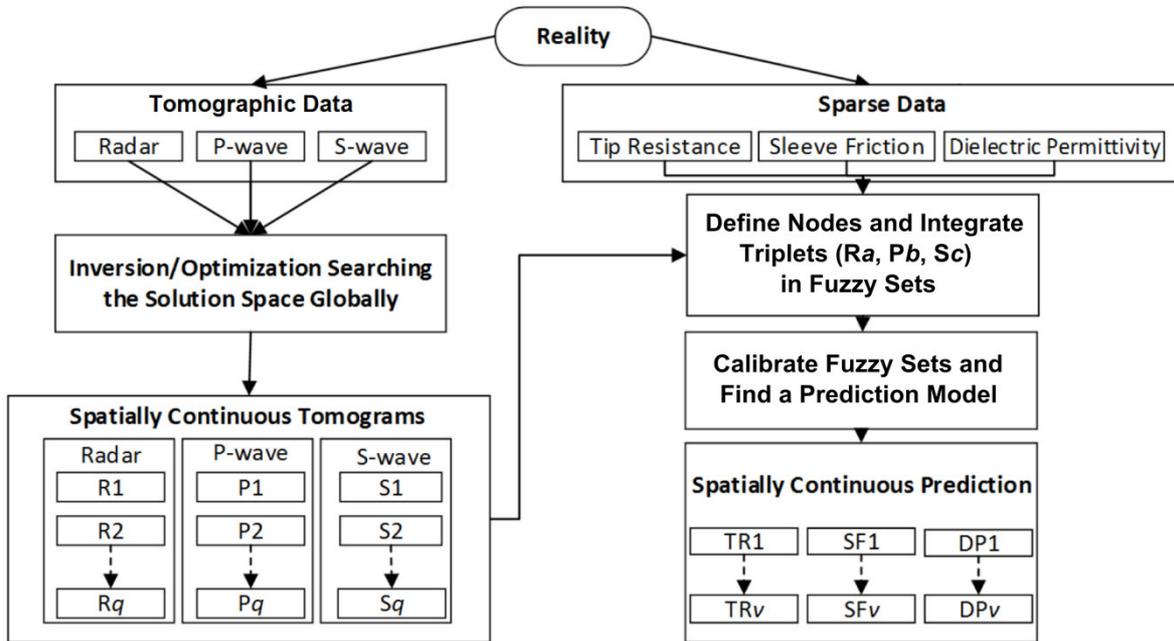


Figure 6

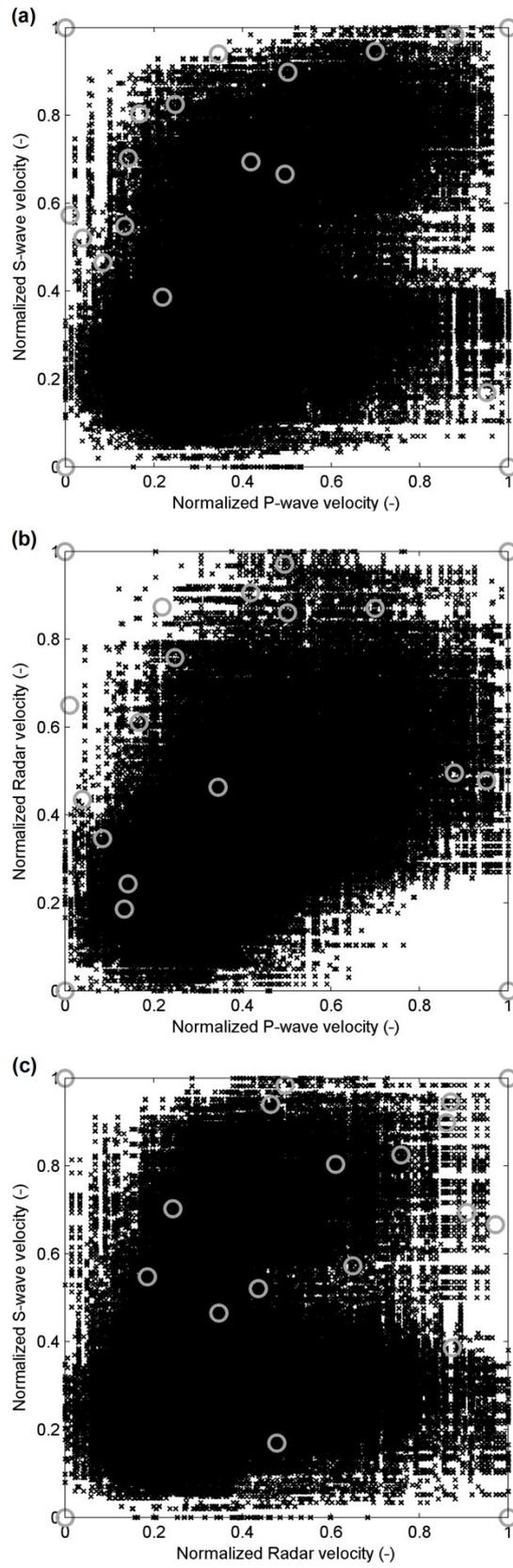


Figure 7

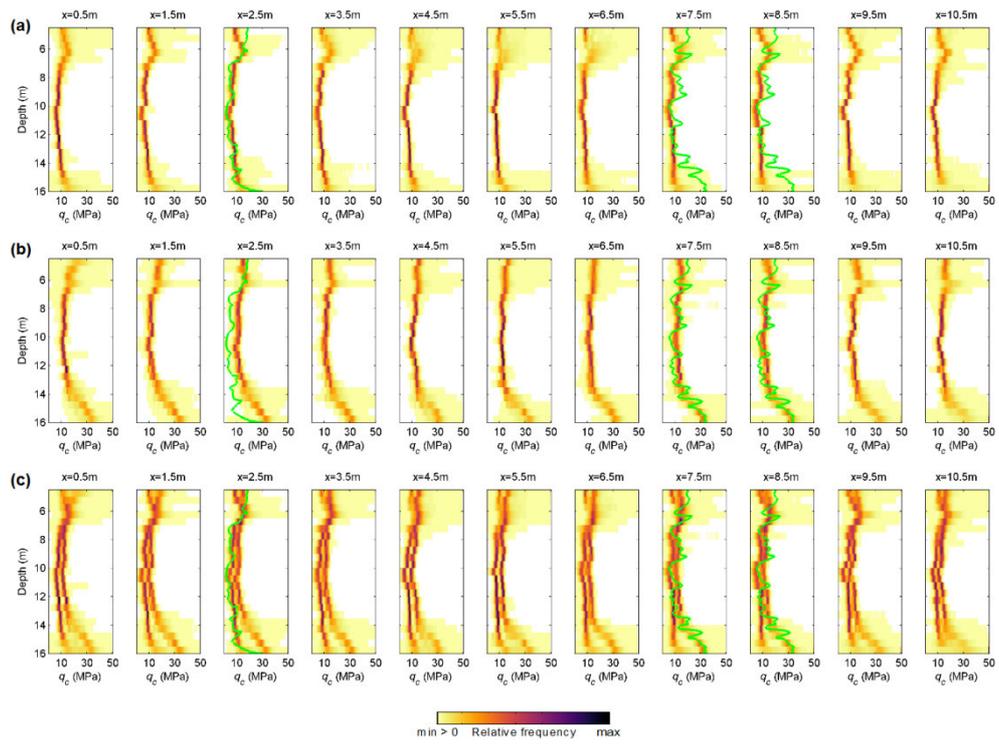


Figure 8

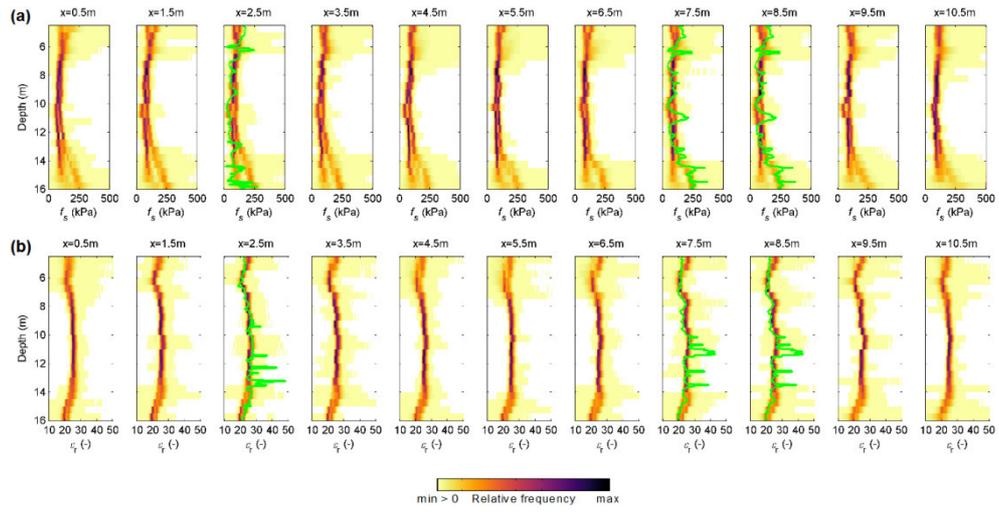


Figure 9

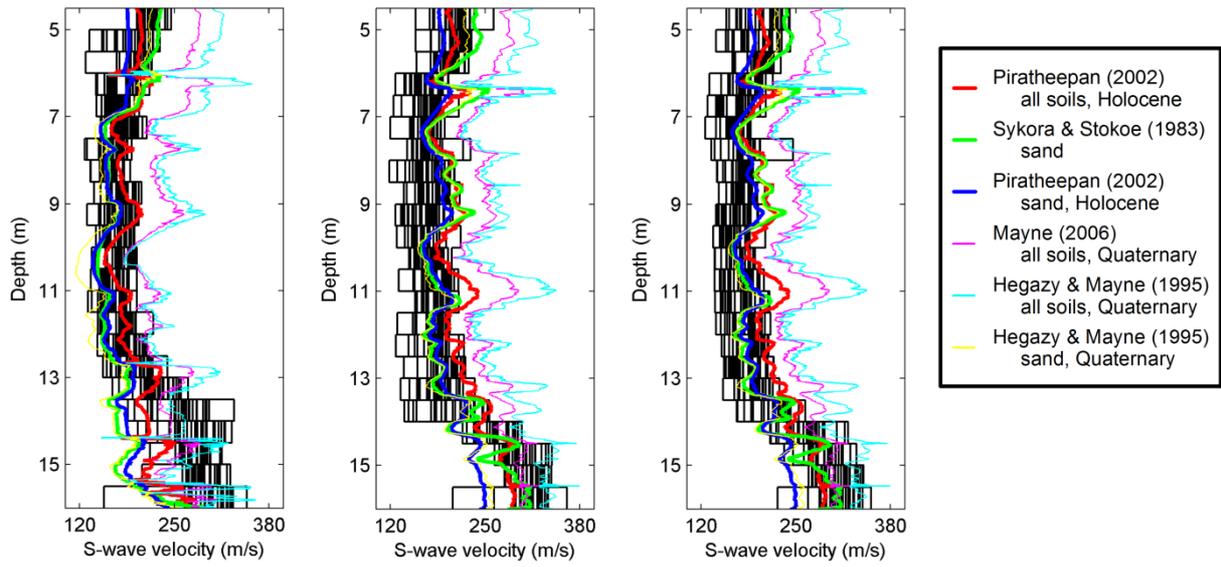


Figure 10

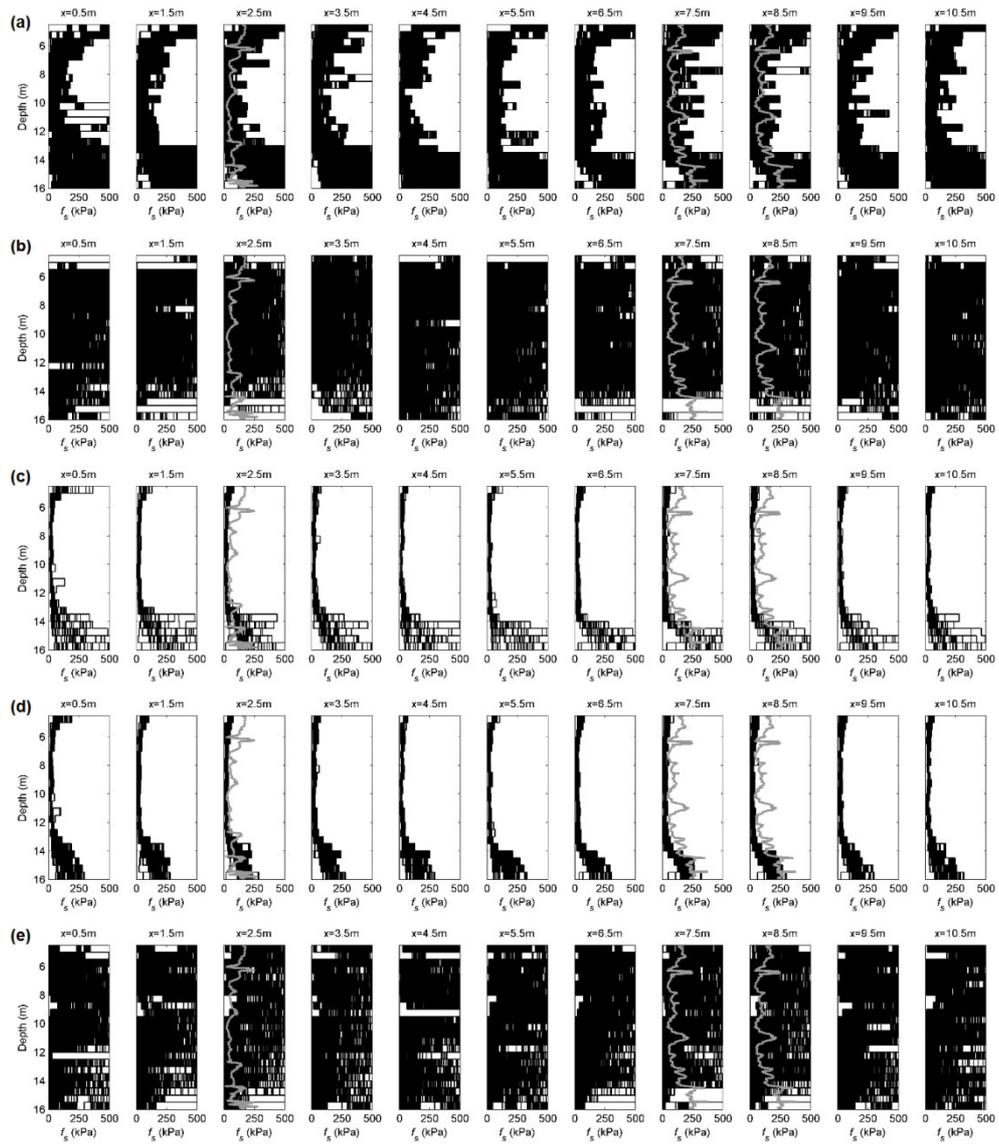


Figure 11

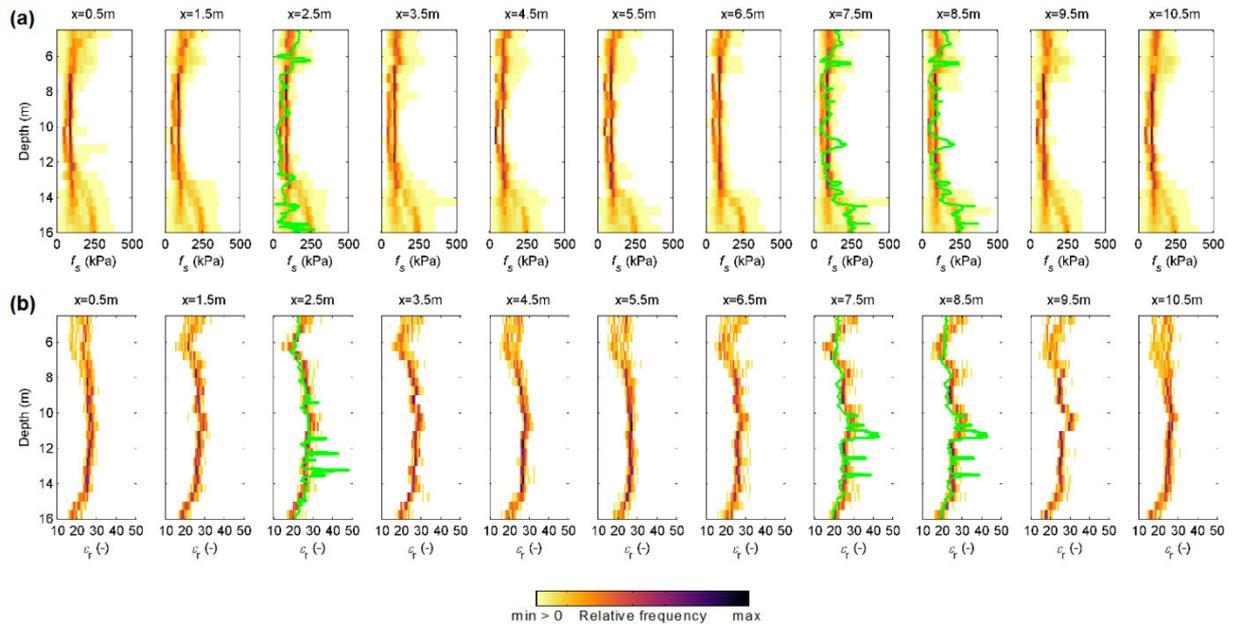


Figure A1

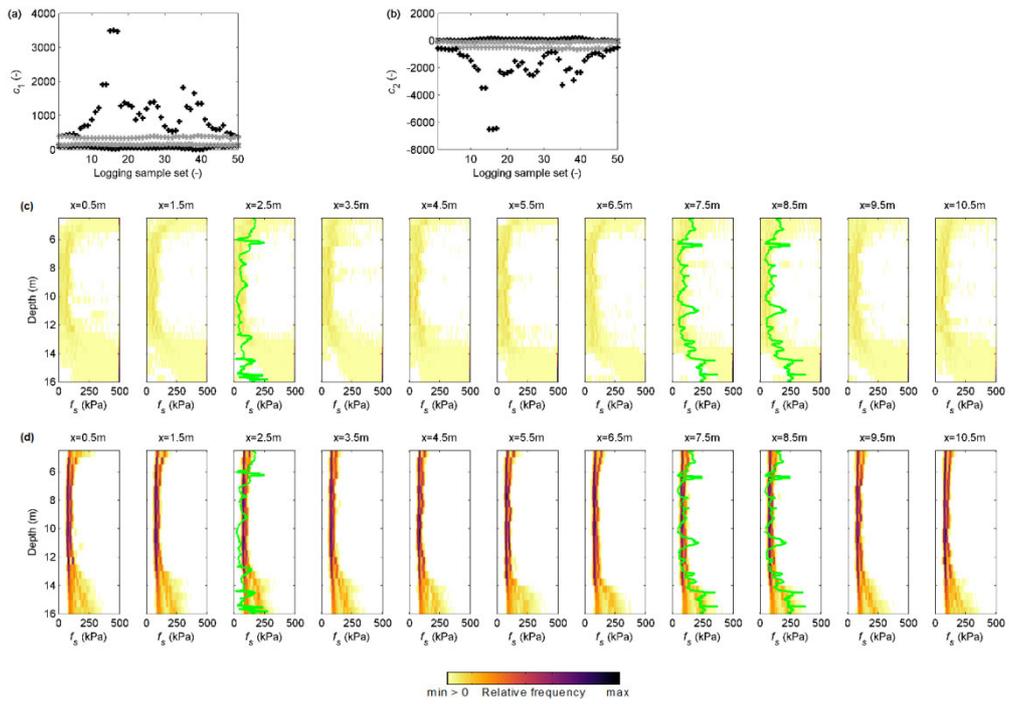


Figure A2

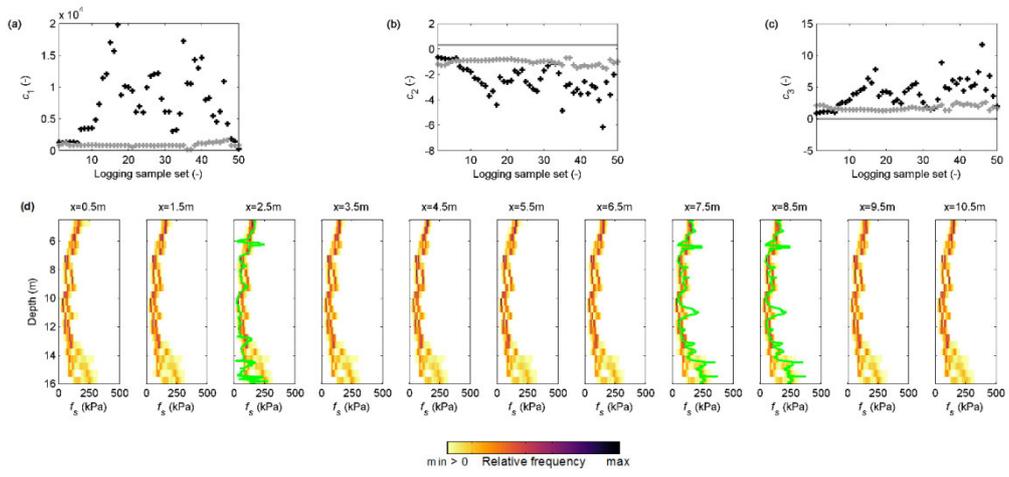


Table 1

Transfer function	Material	Suggested by	Suggested optimal settings
	Holocene soils	Pirathepan (2002)	$c_1 = 32.3$ $c_2 = 0.089$ $c_3 = 0.121$ $c_4 = 0.215$
	Sand	Sykora and Stokoe (1983)	$c_1 = 134.1$ $c_2 = 0.0052$
	Holocene sands	Pirathepan (2002)	$c_1 = 25.3$ $c_2 = 0.163$ $c_3 = 0.029$ $c_4 = 0.155$
	Quaternary soils	Mayne (2006)	$c_1 = 118.8$ $c_2 = 18.5$
	Quaternary soils	Hegazy and Mayne (1995)	$c_1 = 10.1$ $c_2 = 11.4$ $c_3 = 1.67$ $c_4 = 100$ $c_5 = 0.3$
	Quaternary sands	Hegazy and Mayne (1995)	$c_1 = 12.02$ $c_2 = 0.319$ $c_3 = -0.0466$