A probabilistic approach to soil layer and bedrock-level modelling for risk assessment of groundwater drawdown induced land subsidence

A probabilistic approach to soil layer and bedrock-level modelling for risk assessment of groundwater drawdown induced land subsidence

E-mail to corresponding author:

jonas.sundell@chalmers.se

This is the accepted manuscript of the following publication:

Sundell, J., Rosén, L., Norberg, T., & Haaf, E. (2016). A probabilistic approach to soil layer and bedrock-level modelling for risk assessment of groundwater drawdown induced land subsidence. *Engineering Geology, 203*(Special Issue on Probabilistic and Soft Computing Methods for Engineering Geology), 126-139. doi:http://dx.doi.org/10.1016/j.enggeo.2015.11.006

1. Introduction

Improvement of infrastructure in urban areas generally involves tunneling and deep excavations. When building in cities founded on compressible clay, it is important to consider the risk of land subsidence caused by groundwater drawdown. If groundwater leaks into a tunnel overlain by clay deposits or other sediments with high compressibility, it can cause a considerable reduction in pore-pressure and induce subsidence in the soil deposits. Subsidence due to groundwater is a severe problem in many regions around the world, including cities in China (Xue et al., 2005), Mexico City (Ortega-Guerrero et al., 1999), Bangkok (Phien-wej et al., 2006), Las Vegas (Burbey, 2002) and Stockholm (Broms, 1970; Tyrén, 1968).

For assessing the risk of subsidence damage from groundwater leakage into planned sub-surface constructions, a detailed understanding of the soil stratification is necessary (Larsson et al., 1997). To acquire such understanding, 3D soil-strata models based on results from borehole logs can be a useful tool. Several methods for constructing geological 3D models based on information from borehole logs have been used in the past. One such method requires manual connection of soil layers in different boreholes, see e.g. Velasco et al. (2013) and Bozzano et al. (2000). Other methods are based on Kriging, a geostatistical interpolation method (Deutsch & Journel, 1997; Isaaks et al., 1989; Kitanidis, 1997), e.g. Asa et al. (2012), Bourgine et al. (2006) or Thierry et al. (2009). The primary advantage of Kriging is its ability to interpolate values from known data points and provide estimates of the uncertainty at all locations in the model domain. Kriging is commonly used to obtain an estimate with minimum error variance.

When constructing a 3D soil stratigraphic model for an urban area, large numbers of borehole logs commonly exist from previous construction projects (de Rienzo et al., 2008; Marache et al., 2009; Velasco et al., 2013). One challenge in handling such large amounts of data from borehole logs in an urban environment is that the data originates from unrelated successive investigation programs lacking a global strategy. Marache et al. (2009) found that this disordered accumulation of data does not provide a homogeneous or transient representation in terms of the nature, quality or spatial distribution of a city's underground. This means that the interpreted information from different boreholes is often

inconsistent; some logs include the bedrock level whereas others only include some of the soil layers (Bourgine et al., 2006). Furthermore, data quality varies substantially over different periods of a city's history. Consistent nomenclature is not always used since the data has not been collected for a common purpose (Marache et al., 2009). This inconsistency calls for an efficient method for soil-strata modelling that can manage large amounts of data of different character.

An additional challenge when constructing an urban geostratigraphic 3D-model from borehole data is how to account for uncertainties. Since a borehole log is an interpretation of the soil strata, uncertainties exist at the sampling point due to sampling errors (epistemic uncertainty). In addition, uncertainties are also present in the interpretation of soil strata between sampling points because of the natural variability in the geology (aleatory uncertainty). These two types of uncertainties can lead to a false understanding of the location of layers of compressible sediments in relation to the groundwater-pressure level, and hence result in incorrect interpretation of the risk for subsidence damage.

A false understanding of the hydrostratigraphy will lead to two major project risks: (1) neglecting to take action when there is a risk of harmful groundwater drawdown and subsidence, and, (2) taking action when there is no risk of harmful groundwater drawdown and subsidence. In an economic context, the first risk is associated with costs due to damage, and the second with costs for unnecessary mitigation measures. To reduce the risk of incorrect interpretation of the location of compressible sediments and subsidence damage, a three-dimensional probabilistic soil modelling approach can facilitate decision-making concerning cost-efficient actions.

These two challenges in dealing with inconsistent data and uncertainties have been handled differently in previous modelling approaches. Manually connecting soil layers in different boreholes, as in Velasco et al. (2013) and Bozzano et al. (2000), can ensure interpretations that are geologically realistic, even with inconsistent data, but can be time consuming if the dataset is large or is frequently updated. Since the method is based on geological expertise and experience and does not include explicit uncertainty analysis, a quantitative probabilistic analysis for estimating interpolation errors is difficult. For a structured interpolation procedure of soil layers, e.g. Bourgine et al. (2006) and L. Zhu et al. (2012), the interface between individual soil layers is interpolated independently in a first step. In

a second step, overlaps between layers are taken into account by stipulating rules for how one layer should relate to another, e.g. one layer is assumed to have eroded the other. When modelling each layer independently of the others, dependencies between different layers are not taken into account. Not considering dependencies between the different layers in the modelling process itself but in a subsequent professional assumption is a problem in a probabilistic analysis since each simulation needs to be adjusted manually.

The objective of this paper is to present a novel method for a probabilistic analysis of soil strata based on a stratified geostatistical (Kriging) procedure that can manage large amounts of data of different character. The methodology is explained by application to a case study. In section 2, the site for the case study is presented together with assumptions for the conceptual model and available data. Section 3 provides a description of a method for bedrock surface modelling. The model can both be applied deterministically to obtain a point estimate and probabilistically to describe the uncertainties in the model. In the same section an evaluation of the model's validity in relation to a reference dataset is presented. Section 4 presents the developed novel approach for soil stratification modelling. As with the bedrock model, the soil model can both be applied deterministically and probabilistically. The validity of the soil-layer model is evaluated using cross-validation. Finally, in section 5, a method for describing a risk area for groundwater-drawdown induced land subsidence is presented.

This method for soil-strata modelling facilitates a detailed understanding of the soil stratification, drainage paths, and magnitude of spatial variations. When assessing the risk for groundwater drawdown induced subsidence, it is important to recognize the entire drawdown-subsidence-damage chain. Even though the purpose of this case study was to detect compressible sediments, the method is expected to be useful for other purposes of soil-strata modelling.

2. Study area and available data

2.1 Study area

The study area in Stockholm (59°19'N 18°4'E) is located on the East Coast of Sweden and covers approximately 4.5 x 9.5 km². The area is characterized by Pre-Cambrian crystalline bedrock belonging

to the Svecofennian Orogeny with radiometric ages varying between 1750 – 2000 million years (Stålhös, 1969). The most important fracture zones are oriented E-W, WNW and NW (Persson, 1998) . Soil-covered valleys are located in some of the fracture zones.

The soil deposits in the study area are characterized by glacial till and glaciofluvial sediments such as clay, silt, and sand as well as post-glacial sands, clays and organic deposits. Filling material is common throughout the area. Glacial till has been deposited directly on the bedrock, and can be found on the surface level at locations with high bedrock altitudes as well as on the sides of the sedimentary filled valleys. In the valleys, the glacial till has commonly been washed out and glaciofluvial sediments have been deposited directly on the bedrock. Since the entire study area is situated under the highest shoreline, HS, (150 m above current mean sea level, m.a.s.l.) after the last glaciation (Weichsel), it is reasonable to assume that clay has been deposited across the entire valley areas, and partially also in higher areas below HS. Unconsolidated deposits in the western part of the area are dominated by a glaciofluvial esker deposit, the Stockholm Ridge. The central part of the esker consists of coarse sandy to gravelly fractions. The sediments at distal locations in the esker are often layered with both coarse and fine grained sediments. The Stockholm Ridge was previously a topographic high-point, but significant excavations and backfilling have evened out the topography of the ridge. Filling material of different material compositions and thicknesses are present throughout the whole area.

2.2 Conceptual model

When conceptualizing the soil strata at the study site, three distinct layers can be identified: (a) postglacial sand, filling material or organic deposits; (b) glacial and postglacial clay; (c) coarse grained glacial material (sand, glacial till); rock surface (see Figure 2). Even if clay in some areas is found to be inter-layered with other materials, only one clay layer is assumed. All soil layers are modelled as continuous units. In areas with missing strata, where bedrock outcrops are present, the layers are modelled as "zero-thickness" units, see e.g. L. Zhu et al. (2012).

2.3 Available data

In total, data from 20,300 drillings have been collected from various geotechnical reports. After a first check and correction, a MS Access database was built to code all the data in a common referential. The information originates mainly from different infrastructure and building projects. A large quantity of information had been inventoried and digitalized by the Stockholm City Planning Administration during 2012-2013. The oldest drilling was drilled at the end of the 19th century and the newest during 2014. Locations of rock outcrops have been collected and charted on an engineering geological map (Stockholms stad, 2014) at a scale of 1:10,000. A topographic model with a 2x2 meters grid resolution was bought from Metria in 2012. Groundwater pressure-level time series collected from various projects have been stored in another MS Access database.

2.4 Validation dataset

Out of 20,300 drillings, 30 percent were randomly selected as a reference dataset; the other 70 percent were used for the modelling dataset. The result of the bedrock-level and soil-strata modelling was subsequently compared with the validation dataset using cross-validation to assess the model performance.

3. Bedrock-level model

To model the level and thickness of individual soil layers, a bedrock-level model must be created first to obtain the total soil thickness. Since boreholes that both reach the bedrock (BH1) and do not reach the bedrock (BH2), see Figure 2, as well as areas of bedrock outcrops (BO) are considered in the model, the procedure needs to be divided into several steps. The first reason for several interpolation steps is described in Figure 3: As a first step, only BH1 is to be used to interpolate a surface (orange line). As a second step, information from both BH1 and BH2 are used to interpolate a surface (blue line in Figure 3). In the last step, the lowest level of the two surfaces is chosen as the model for the bedrock level (hatched line in Figure 3).

The second reason for why several interpolation steps are needed is because BO data is denser compared to the borehole dataset since it is interpreted as continuous surfaces. Modelling attempts conducted without considering the difference in density have shown that in valleys between areas with

BO, bedrock levels are consistently over-predicted. This leads to a geologically unrealistic model and to an underestimation of areas with compressible sediments. To compensate for the higher density of BO data, support-lines in the middle of valleys between areas with BO are introduced. The support lines were first automatically generated to be located exactly in the center of areas with BO. Thereafter, their positions were adjusted to locations with deep bedrock levels.

3.1 Management of available data

To model the bedrock surface level, the data has been divided into three different groups according to the following procedure:

- BH1: Soil-rock probing that penetrates the bedrock surface are included in this group. For the modelling dataset, 6,510 boreholes are included in this group.
- BH1 and BH2: In this group all boreholes are included. For BH1, the bedrock level is used in the interpolation, and for BH2 the lowest level is used. The modelling dataset contains 14,279 boreholes.
- Bedrock outcrops from the engineering geological map of Stockholm have been given altitude levels from the ground-surface model on a dataset with 10 x 10 m resolution. Bedrock outcrops are not used at a distance less than 30 meters from a borehole since information from boreholes are assumed to be more reliable.

Kriging requires data used for interpolation to be normally distributed. Figure 4 shows that it is reasonable to assume that the bedrock levels can be described by a normal distribution.

3.2 Variogram of bedrock level from boreholes

When modelling with Kriging, a variogram is necessary to determine spatial correlations between the data and for weighting the values of neighboring data points at unsampled sites. Moreover, a variogram can reveal anisotropy of a variable within the model field. Figure 5 shows an experimental and a fitted, theoretical variogram in two directions based on data from BH1. When the theoretical variogram was fitted to the experimental one, a lag distance of 400 meters was selected as the range for selecting the best model with the LS-method (Cressie, 1985). This distance was selected because only a few data points were located at longer distances than 400 meters to the closest data point. The

fitted theoretical variogram is a combination of an exponential and a Gaussian model, see e.g. Lloyd Smith (1999); the parameters of the variogram are presented in Table 1. A combination of the two models is necessary since the experimental variogram can be divided into two parts: a first asymptotical increase for distances up to 100-150 meters depending on direction and; a second at greater distances where the variogram not is leveled out to a constant sill, implying a constantly greater variance at longer lag distances. The Gaussian model is necessary for the first part whereas the exponential model describes the second part.

Both the experimental and the modelled variogram, Figure 5, reveals a clear anisotropy. In direction 0, EW (168), the variogram values are lower compared with other distances. This direction is also the most dominating direction of the valleys in the study area. Interpolated point estimates are affected of this anisotropy. Data points in EW directions of point estimates (along the dominating direction of valleys) are given higher weight than points in NS directions (perpendicular to the dominating direction of valleys).

3.3 Interpolation steps of deterministic bedrock-level model

In Figure 6, the interpolation steps of the bedrock-level model are presented. For the case study, the model was conducted with a 10 x 10 m grid resolution. The asymptotic variogram model presented in Table 1 and seen in Figure 5 has been used for all steps. The steps in Figure 6 are:

A. The data: BH1, BH2, and BO is presented together with section marker for cross-section in Figure 7.
B. The first interpolation is conducted only with BH1.
C. The second interpolation is conducted with BH1 and BH2.
D. The lowest absolute level from step B or C is selected at each interpolation point.
E. Support lines between areas with BO have been given values from step D along the lines at every 10th meter. The values at the support lines are then interpolated with data from BH1.
F. The values at the support-lines are interpolated with data from BH1 and BH2.
G. The lowest level from step E and F is selected at each interpolation point.

H. As a final step, the result from step G is compared with the ground-surface model to ensure that the modelled bedrock level does not increase ground-surface level.
I. The soil thickness is calculated based on the difference between the ground surface and the bedrock-level model.

In Figure 7 the result of the interpolation steps are shown in a cross-section. The grey area represents the soil thickness between the ground level and the modelled bedrock level. At meter 180 the necessity to include BH2 in the model is revealed. If not BH2 would have been included, the bedrock level would have been underestimated at those locations.

The effect of the support lines can be seen between m 100 and 150 in Figure 7 and more clearly between m 800 and 1000 in Figure 15. The black hatched line shows the bedrock level modelled according to the steps above but without support lines. This model follows the bedrock level model (H) until m 800. At m 800-900 there is an area between BO where drillings are not present. Since BO are located as dense clusters giving high weight to the model, these clusters results in high bedrock levels for the model without support lines. Since the purpose of the model is to estimate the probability of compressible sediments it is important not to underestimate the thickness of the soil cover.

3.4 Probabilistic bedrock-level model

When interpolating with Kriging, not only the weighted average, but also the error variance is computed at every interpolation point. Since the interpolation procedure is divided into several steps, the errors from the individual steps have to be accounted for when evaluating errors in the bedrock-level model. Moreover, dependencies between the interpolation steps also have to be taken into account. In Figure 8 A the standard deviation from interpolation with BH1 and BO is shown. Similarly, in Figure 8 B, the standard deviation from interpolation with BH1, BH2 and BO is presented. Since the bedrock level can be assumed to vary with a Gaussian distribution, see Figure 4, the standard deviations presented in Figure 8 can be used for a Monte Carlo simulation to predict the uncertainties in the bedrock-level model. Even though the support lines are used when BO is modelled together with BH1 and, BH1 and BH2 respectively, the standard deviations generated from interpolations without the support lines have been used for predicting the uncertainties in the model.

The reason for this is that the values at the support lines are not observations in themselves but modelled values. The following steps are followed for the prediction:

1. The average values for BH1 and BO are represented by the interpolation with BH1, BO, support-lines (step E in Figure 6) and the standard deviations from the generated error of the kriging interpolation with only BH1 and BO (Figure 8 A).
2. The average values for BH1, BH2 and BO are represented by the interpolation with BH1, BH2. BO (step F in Figure 6) and the standard deviations from the generated error of the kriging interpolation with only BH1 and BO(Figure 8 B).
3. In each realization a random quantile from a uniform distribution is generated. Subsequently, the random quantile is applied to two quantile functions of the normal distribution with parameters obtained from step 1 (QF1) and 2 (QF2). The resulting samples are compared, and the lowest value of the two is selected.
4. For the case study, an ensemble of 1,000 realizations are run to estimate the possible variations due to the uncertainties in the model. Subsequently, a distribution of possible outcomes for each location is obtained. From these, average values, percentiles (Figure 9) and standard deviations of the bedrock level are calculated.

When evaluating the uncertainties of the bedrock-level model using Monte-Carlo simulation, it is important to evaluate one interpolation point at a time. For example, the 95th percentile map shows, the bedrock level that will be exceeded by a probability of 0.05 at each location. Of course, the probability that the bedrock levels at all locations will exceed the 95th percentile value is much lower.

3.5 Validation with reference dataset

A comparison between the reference dataset and the simulated bedrock levels in the deterministic and the probabilistic models is presented in Figure 10 and Table 2. Figure 10 shows residuals of the bedrock level between the validation dataset and the simulated values for different percentiles. Negative residuals means that the modelled bedrock level is overestimated and vice versa. Figure 10 shows that the 50th percentile is homogeneous distributed around zero for all levels indicating unbiased estimates of the bedrock level.

Table 2 shows the portion of the modelled bedrock levels with a value lower than the validation dataset. In comparison to the validation dataset, the modelled 5th and 10th percentile underestimate the bedrock level since the portions below the validation dataset are 0.09 and 0.13 respectively. The deterministic model and the average, 50th, 90th and 95th percentile of the probabilistic model overestimate the bedrock level in comparison to the validation dataset. Since the low and high percentiles generally underestimate respectively overestimate the bedrock level, the range for the possible outcomes of the model is less than the variability in the validation dataset. This can be a result of the smoothing process that generally occurs in Kriging (Chung & Rogers, 2012).

The general overestimation of the bedrock levels in the model can be explained by the clustered BO data that leads to an increase of the levels. Even though support lines are introduced to keep the bedrock levels low where they are expected to be so, they cannot compensate fully for the dense BO data.

4. Soil-strata model

From the soil thickness (ground-surface to bedrock-surface), thickness and the level of individual soil layers are modeled. The purpose of the soil-layer model is to model the thickness and levels of clay which are crucial for estimating the risk of subsidence. In the layers above and below the clay layer, all the types of coarse grained material have been classified as one soil-class. Even though a more complex soil stratification than only three layers can be expected, the simplification is justified for the purposes of the model. With the assumption of only one clay layer, the clay thickness might be over-estimated at some locations.

The individual soil layers are not only dependent of the soil-thickness model, but also of each other. To take these dependencies into account, the borehole observations have been converted to proportions of the total soil thickness for the clay layer and proportions of the remaining soil thickness for the layer with friction material closest to the bedrock. Uncertainties in thickness and position of the clay layer is of primary importance for estimating the risk for subsidence, therefore the need for a probabilistic soil-strata model.

4.1 Management of available data

From all boreholes containing information about clay layers, the clay thickness has been calculated as the difference between the highest and the lowest level interpreted as clay. In the modelling dataset, 3960 boreholes contain information about the clay layer. Out of these 1680 contain incomplete information of the soil stratigraphy since the bedrock level is not reached. Moreover, 380 boreholes with information on clay layers do not reach stiff soil. The calculation of the clay thickness implies that if a borehole contains layers interpreted as clay, the whole clay thickness is assumed to have been pierced even if the borehole does not reach the bedrock. Since a common goal of a geotechnical sounding is to determine the depth to stable soil, this assumption can be justified for many situations (90 percent of the boreholes in the case study). To be able to observe the entire clay thickness with greater certainty, the whole soil layer down to the bedrock needs to be penetrated by the drilling.

If the model would only have used boreholes with complete information of the soil stratigraphy, incomplete boreholes with clay observations would have been left out of the modelling dataset. Leaving these boreholes out would imply a smaller number of data points for modelling. By leaving out incomplete drillings (see Figure 2), the reliability of the model can be increased in areas where drillings penetrate the entire clay layer. On the other hand, the model is less reliable in areas without drillings that are observed to pierce the entire clay layer since a part of the available information on clay thickness is notused. Consequently it is more advantageous to include incomplete drillings.

In urban areas it is common that the ground level changes throughout time due to shafting and filling. This means that the indicated ground level of an old borehole might not be representative for the present situation. To address this issue filling material is added to boreholes with a reference level below the ground-surface model. For borehole logs above the present surface level it is assumed that the upper part of the log has been peeled down to the present ground-surface level. To address the issue with incomplete drillings not reaching the bedrock, the bedrock-level is assumed from the bedrock-level model. This assumption means that additional coarse grained material is added between the lowest soil layer in the borehole and the bedrock-level (since the whole clay thickness is assumed to have been penetrated).

Apart from using borehole data containing information about clay layers, boreholes reaching the bedrock where the soil profile contains no clay have been included for modeling. For the latter, the whole profile is in the model assumed to belong to the lowest soil layer. In the modelling dataset, 1930 number of boreholes are classified to only contain friction material.

After the data has been converted, the proportion of clay in relation to the total soil thickness can be calculated by Eq. 7:

$$p_b = \frac{b}{a+b+c}, \quad (7)$$

where a, b, and c are soil layer thickness described by Figure 2.

The dependencies among the three soil-layers are taken into account by calculating the proportion of the lowest layer (c), in relation to the difference between the total soil thickness and the clay layer (a + c) (Eq. 8),

$$p_c = \frac{c}{a+c}, \quad (8)$$

where a and c are described by Figure 2.

As mentioned above, Kriging requires the data to be represented by a continuous normal distribution. The requirement for data to be normally distributed is not fulfilled when the soil layers are converted to proportions, since the proportions can take values only between 0 (a certain soil layer is not present) and 1 (a certain soil layer is the only layer in the profile). Due to this, the proportions p_b and p_c are transformed so they can be represented by a normal distribution, where the proportions represent the probability integral of the standardized normal distribution, *N(0,1)*. Consequently, the probabilities are transformed to the variables *z* of the standardized normal distribution as seen in Figure 11.

Figure 11 shows that a normal distribution is a good approximation of the transformed values p_b. The same conclusion can be drawn for p_c. The figure, however, does not include the values 0 and 1 for p_b. Since the probabilities 0 and 1 represent minus infinity and infinity on the standardized normal distribution, these values are instead assumed to 0.01 and 0.99 respectively. With this assumption, the proportions 0.01 and 0.99 are converted to the variables *z* -2.33 and 2.33 respectively. A consequence

of this assumption is that all three layers are always present at all locations even though some layers might be very thin.

4.2 Variogram for soil layers

In a similar manner as for the bedrock-level model, variograms have been calculated for the converted variables z of the soil-layer proportions. In Figure 12, experimental and theoretical variograms are presented for the clay-layer and the lowest layer. The experimental variograms are calculated from boreholes where: $0 < p_b < 1$ respective $0 < p_c < 1$. The theoretical variograms have been fitted to the experimental variogram using the LS-method over a range of 300 meters for the lag distance. The parameters of the theoretical variograms are presented inTable 3.

In contrast to the variogram for the bedrock surface, a clear nugget effect can be observed at the variograms for the soil layers. A likely explanation for the nugget effect is that it is more difficult to make accurate observations of distinctions between different soil layers compared with the bedrock level. An additional contribution to the nugget effect is probably that only one clay layer is assumed to exist even for the cases when many layers are observed.

Similarly to the variograms for the bedrock surface, anisotropy can be observed for the soil layers. This anisotropy is, however, not as clear as for the bedrock level model. At a lag distance of 30 to 50 meters, depending on the direction, the variograms start to level out. When a variogram has leveled out, the correlation length is reached. At lag distances longer than the correlation length, the whole dataset is influencing the interpolation result, not individual data points.

4.3 Interpolation steps of deterministic soil-strata model

Since the soil-thickness data has been converted in several steps - first to proportions and then to the z variable of the standardized normal distribution - it is necessary to conduct the interpolation in a stepwise procedure. Kriging is used together with variogram models presented in Figure 12. The procedure is as follows (see Figure 13):

A. Data points with proportions of clay of the total soil thickness (p_b) converted to the variable z_{pb} of the standardized normal distribution are interpolated with Kriging and the variogram in Figure 12.

B. The interpolation result in A is calculated from the z_{pb} values of the standardized normal distribution to proportions p_b.

C. From the proportions of clay and the total soil thickness in Figure 6 I, the clay thickness is calculated. The result of this calculation (clay thickness in meters) is presented in Figure 13 C.

With a corresponding procedure, the thickness of the lowest level, c, is modelled (see Figure 14).

From the thicknesses of layers *b* and *c*, the thickness of layer *a* is calculated. After this step, the altitude levels of the interface between the different layers are calculated from the ground surface elevation and the thicknesses of the layers. The result of this calculation is presented in the soil profile in Figure 15.

The coarse-grained sediments in the area where the profile cuts through the Stockholm esker are thick. Between m 600 and 750 in the cross-section, a lake which is now drained and filled existed previously. As can be seen in the profile, a very thin clay layer is modelled even in the areas where clay is not expected.

4.4 Probabilistic soil-layer model

Analogous to the bedrock-level model, the weighted average and error are computed at every interpolation point for each of the interpolations. Figure 16 shows the standard deviation from the interpolations of the clay layer (z_{pb}) and the lowest layer (z_{pc}).

Monte-Carlo simulations of the soil-layer model are conducted at each interpolation point according to:

1. Values for the bedrock-surface level are simulated according to the procedure in section 3.4.
2. Values for the soil-thickness: $s_{tot} = a + b + c$ are calculated from the difference between the ground-surface model and the simulated value for the bedrock-level model in step 1.
3. Values for z_{pb} are simulated from a normal distribution, where the mean-values are represented by the previous interpolation of z_{pb} in Figure 13 A, and the standard deviations presented in Figure 16 A.

4. The simulated values for z_{pb} are converted to p_b.
5. The clay thickness, b, is calculated according to the following formula, $b = p_b * s_{tot}$.
6. Values for z_{pc} are simulated from a normal distribution, where the mean-values are represented by the previous interpolation of z_{pc} in Figure 14 A, and the standard deviations presented in Figure 16 B.
7. The simulated values for z_{pc} are converted to p_c.
8. The thickness for the lowest layer, c, is calculated according to the following formula: $c = p_c * (s_{tot} - b)$.
9. The altitudes of the interface between the different layers are calculated.
10. The steps, 1-9 continued for each of the 1,000 simulations, the same number as for the bedrock-level model. When the simulations are complete, distributions for each of the interpolation points are obtained. From those, average values, percentiles and standard deviations of the soil-layer's thicknesses and levels are obtained.

The result of the simulations of the clay thickness is presented in Figure 17. As seen in the figure, the range of possible outcomes between the 5th and the 95th percentile can be very large, from about 1 meter up to 10 meters at the same location.

4.5 Validation of soil-strata model

A comparison between the validation dataset and the simulated clay thickness in the deterministic and the probabilistic models is presented in Figure 18 and Table 4. Figure 18 presents boxplots of residuals between the clay thickness of the validation dataset and the simulated values. The boxplots for the 50th percentile show that the thickness is generally over-estimated for thicknesses below two meters. Above two meters, the thickness is generally under-estimated. The general trend of under-estimating greater thicknesses can be explained by the smoothing process that generally occurs in Kriging.

Table 4 presents the part of simulated values in the clay-thickness model that have greater thickness than the validation dataset. Two subsets of the validation data are compared to the modelling data: i) drillings without clay and ii) where these drillings are not included. In the modelling dataset, the drillings without clay have been included for both comparisons. For the lowest percentile, P5, the

model and the validation dataset, for i) are in good correspondence. The correspondence decreases for ii). This is a consequence of a thin clay layer being generated even though no clay is observed. The portions in the validation dataset shows a good agreement with the simulated percentiles. Since the model is intended for estimating the risk for compressible sediments, it is important to confirm that the 90^{th} and 95^{th} percentiles are in good correspondence with the validation dataset.

5. Risk area for ground subsidence

When calculating groundwater drawdown induced subsidence, it is important to know whether the groundwater-pressure level is above the clay's lower boundary or not. Assuming that there is no significant capillary action in the coarse grained material, a groundwater drawdown can only cause subsidence between meters 650 to 750 along the cross-section, see Figure 15. The probabilistic soil-layer model together with an interpolated groundwater level-surface(about 300 observation wells were used for the interpolation in the study-area) was used for simulating the area where the groundwater pressure level in the lowest layer of coarse grained material is above the lower boundary of the clay layer, which constitutes the risk area. For practical reasons the boundary of the risk area shown in the case study was set to where the groundwater pressure level of the confined aquifer is at least one meter into the overlying clay. The different percentiles of the simulation of the risk area are presented in Figure 19. Since the clay thickness varies largely between the different percentiles, as presented in section 4.4, the extent of the risk areas also shows a great variability for the different percentiles. This variability occurs due to uncertainties in the bedrock-level and in the soil-layer models.

6. Conclusion

This paper presents a novel method for bedrock-level and soil-strata modelling. The method can be used both as a deterministic model to obtain the expected bedrock level and soil strata, and as a probabilistic model to describe uncertainties of bedrock levels and soil stratification in these areas. Applied to a case study in central Stockholm, the method has shown to be effective and to provide reliable results validated by a reference dataset. In the case study, the model was used to identify risk-areas where it is possible for a groundwater drawdown to induce subsidence in compressible soil deposits. The risk areas provide useful decision-support in planning for future sub-surface

constructions in Stockholm that can cause groundwater-drawdown. With the risk areas as a decision support, mitigation measures to avoid subsidence can be prioritized.

In comparison to existing approaches for bedrock-level and soil-strata modelling, the presented method provides several novelties on how different sources of information can be used in a probabilistic model. One example is how dependencies among bedrock level and soil layers are taken into account in a probabilistic model by modelling vertical proportions of soil strata. An additional novelty is the introduction of support lines to compensate for a dense dataset at bedrock outcrops that otherwise would cause too high bedrock levels in the model.

In comparison to other structured interpolation approaches of soil-layers, such as Bourgine et al. (2006) or Zhu et al. (2012), where dependencies are taken into account by first interpolating the different soil layers independently and then managing overlapping layers by establishing rules on how one layer has eroded the other, the approach in this paper first converts the soil-layer data to independent proportions beforemodelling is carried out. The modelling approach presented here is particularly suited for areas where all soil layers to be modelled could be reasonably expected to exist at all locations. Since the whole study area in Stockholm is well under the highest shoreline after the last glaciation, it is a reasonable to assume that clay was deposited at all locations across the entire valley. Still, there are locations where presence of clay is not only dependent on sedimentary condition, but also on erosion and shafting. Even though all layers cannot be assumed to exist at all locations, the presented model still provides reasonable results since non-observed layers are modelled to be very thin.

The presented method is a useful tool for identifying areas with risk for groundwater-drawdown induced subsidence. Depending on what is considered an acceptable level of uncertainty, risk areas for different percentiles can be provided. The 95^{th} percentile covers a larger area than the 50^{th} percentile and therefore represents a more precautionary approach for evaluating where safety measures should be implemented. Risk aversion leads to a larger risk areas which imply higher costs for safety measures. Correspondingly, if the risk area is underestimated, costs for subsidence damages are likely

since insufficient safety measures were implemented. Further model development is planned to estimate the optimal area for safety measures by means of risk-cost-benefit analysis.

The magnitude of the subsidence and its consequences do not, however, solely depend on the soil stratification and the groundwater-pressure level. It is also important to evaluate the possible extent of a planned groundwater drawdown, the compressibility of the clay layer and the sensitivity of the constructions at risk for damage. To describe the whole groundwater drawdown-subsidence chain, further research is needed for combining the soil-strata model with groundwater and subsidence models. Although several case studies of subsidence due to groundwater drawdowns exists, see e.g. Modoni et al. (2013), Touch et al. (2014) and L.-f. Zhu et al. (2013), there is a need for evaluating uncertainties in all components of this multidisciplinary problem. The presented model can be a good first step of a decision-support tool when assessing risk for groundwater drawdown induced subsidence.

Acknowledgements

We gratefully acknowledge the funding of this work by Formas (contract 2012-1933), BeFo (contract BeFo 333). and the COWI-fund. We also thank the anonymous referees whose comments and insights helped to improve the paper.

References

Asa, E., Saafi, M., Membah, J., & Billa, A. (2012). Comparison of linear and nonlinear kriging methods for characterization and interpolation of soil data. *Journal of Computing in Civil Engineering, 26*(1), 11-18.

Bourgine, B., Dominique, S., Marache, A., & Thierry, P. (2006, 6-10/09/2006). *Tools and methods for constructing 3D geological models in the urban environment; the case of Bordeaux.* Paper presented at the Engineering Geology for Tomorrow's Cities,, Nottingham, UK.

Bozzano, F., Andreucci, A., Gaeta, M., & Salucci, R. (2000). A geological model of the buried Tiber River valley beneath the historical centre of Rome. *Bulletin of Engineering Geology and the Environment, 59*(1), 1-21. doi: 10.1007/s100640000051

Broms, B. (1970). *Sättningar och sättningsskador på grund av grundvattensänkning*. Paper presented at the Nordisk hydrologisk konferens, Stockholm.

Burbey, T. J. (2002). The influence of faults in basin-fill deposits on land subsidence, Las Vegas Valley, Nevada, USA. *Hydrogeology Journal, 10*(5), 525-538. doi: http://dx.doi.org/10.1007/s10040-002-0215-7

Chung, J.-w., & Rogers, J. D. (2012). Estimating the position and variability of buried bedrock surfaces in the St. Louis metro area. *Engineering Geology, 126*(0), 37-45. doi: http://dx.doi.org/10.1016/j.enggeo.2011.12.007

Cressie, N. (1985). Fitting variogram models by weighted least squares. *Journal of the International Association for Mathematical Geology, 17*(5), 563-586.

de Rienzo, F., Oreste, P., & Pelizza, S. (2008). Subsurface geological-geotechnical modelling to sustain underground civil planning. *Engineering Geology, 96*(3–4), 187-204. doi: http://dx.doi.org/10.1016/j.enggeo.2007.11.002

Deutsch, C. V., & Journel, A. G. (1997). *GSLIB : geostatistical software library and user's guide* (2nd ed.): Oxford University Press.

Isaaks, E. H., Srivastava, R. M., & Knovel. (1989). *Applied geostatistics*. New York: Oxford University Press.

Kitanidis, P. K. (1997). *Introduction to geostatistics: applications to hydrogeology*. Cambridge: Cambridge Univ. Press.

Larsson, R., Bengtsson, P.-E., & Eriksson, L. (1997). Prediction of settlements of embankments on soft, fine-grained soils. *Information 13E*. Linköping: Swedish Geotechnical Institute.

Lloyd Smith, M. (1999). *Geologic and Mine Modelling Using Techbase and Lynx*. Abingdon: Taylor & Francis Group.

Marache, A., Breysse, D., Piette, C., & Thierry, P. (2009). Geotechnical modeling at the city scale using statistical and geostatistical tools: The Pessac case (France). *Engineering Geology, 107*(3–4), 67-76. doi: http://dx.doi.org/10.1016/j.enggeo.2009.04.003

Modoni, G., Darini, G., Spacagna, R. L., Saroli, M., Russo, G., & Croce, P. (2013). Spatial analysis of land subsidence induced by groundwater withdrawal. *Engineering Geology, 167*(0), 59-71. doi: http://dx.doi.org/10.1016/j.enggeo.2013.10.014

Ortega-Guerrero, A., Rudolph, D. L., & Cherry, J. A. (1999). Analysis of long-term land subsidence near Mexico City: Field investigations and predictive modeling. *Water Resources Research, 35*(11), 3327-3341. doi: 10.1029/1999WR900148

Persson, L. (1998). Engineering geology of Stockholm, Sweden. *Bulletin of Engineering Geology and the Environment, 57*(1), 79-90. doi: 10.1007/s100640050024

Phien-wej, N., Giao, P. H., & Nutalaya, P. (2006). Land subsidence in Bangkok, Thailand. *Engineering Geology, 82*(4), 187-201. doi: http://dx.doi.org/10.1016/j.enggeo.2005.10.004

Stockholms stad. (2014, 2014-07-01). Building geological map. Retrieved 2014-11-21, 2014, from http://www.stockholm.se/ByggBo/Kartor-och-lantmateri/Bestall-kartor/Geoarkivet/

Stålhös, G. (1969). *Stockholms omgivningar: geologi*: Norstedt.

Thierry, P., Prunier-Leparmentier, A.-M., Lembezat, C., Vanoudheusden, E., & Vernoux, J.-F. (2009). 3D geological modelling at urban scale and mapping of ground movement susceptibility from gypsum dissolution: The Paris example (France). *Engineering Geology, 105*(1–2), 51-64. doi: http://dx.doi.org/10.1016/j.enggeo.2008.12.010

Touch, S., Likitlersuang, S., & Pipatpongsa, T. (2014). 3D geological modelling and geotechnical characteristics of Phnom Penh subsoils in Cambodia. *Engineering Geology, 178*, 58-69. doi: http://dx.doi.org/10.1016/j.enggeo.2014.06.010

Tyrén, S. (1968). Grundvattenproblem i tätorter. *Byggforskning 68*, 87-100.

Velasco, V., Gogu, R., Vázquez-Suñè, E., Garriga, A., Ramos, E., Riera, J., & Alcaraz, M. (2013). The use of GIS-based 3D geological tools to improve hydrogeological models of sedimentary media in an urban environment. *Environmental Earth Sciences, 68*(8), 2145-2162. doi: 10.1007/s12665-012-1898-2

Xue, Y.-Q., Zhang, Y., Ye, S.-J., Wu, J.-C., & Li, Q.-F. (2005). Land subsidence in China. *Environmental Geology, 48*(6), 713-720. doi: 10.1007/s00254-005-0010-6

Zhu, L.-f., Li, M.-j., Li, C.-l., Shang, J.-g., Chen, G.-l., Zhang, B., & Wang, X.-f. (2013). Coupled modeling between geological structure fields and property parameter fields in 3D engineering geological space. *Engineering Geology, 167*(0), 105-116. doi: http://dx.doi.org/10.1016/j.enggeo.2013.10.016

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. *Engineering Geology, 127*(0), 1-13. doi: http://dx.doi.org/10.1016/j.enggeo.2011.12.001

Table 1 **Parameters of the fitted theoretical variogram model for bedrock level**

Variogram	Theoretical model	Max distance	Sill	Range	Aniso ratio	Aniso angle
BH1	Exponential	400	483.6	4927	3.0	163
	Gaussian	400	35.6	102	1.8	168

Table 2 **Portion of simulated values in the bedrock-level model that have a level below the validation dataset.**

Model	**Portion below validation dataset**
Deterministic	0.45
Average	0.43
5th percentile	0.09
10th percentile	0.13
50th percentile	0.45
90th percentile	0.81
95th percentile	0.85

Table 3 **Parameters of the fitted theoretical variogram model for z_{pb} and z_{pc}.**

Variogram	Theoretical model	Error variance	Max distance	Sill	Range	Aniso ratio	Aniso angle
Zpb	Nugget	0.16					
	Exponential		300	0.24	386.2	2.0	162
	Spherical		300	0.16	59.6	1.38	103
Zpc	Nugget	0.12					
	Exponential		300	0.50	44.6	1.67	174

Table 4 **Portion of simulated values in the clay-thickness model that have a greater thickness than in the validation dataset.**

	Deterministic	Average	Residuals P5	Residuals P10	Residuals P50	Residuals P90	Residuals P95
Portion overestimated - drillings without clay not included in validation dataset	0.37	0.45	0.04	0.06	0.36	0.84	0.90
Portion overestimated - drillings without clay included in validation	0.58	0.63	0.11	0.14	0.46	0.89	0.93

dataset							

Figure captions

Figure 1 The study area in Stockholm (big image) and its location (small image). The following figures in this paper only show the marked area in green. Only boreholes in the modelling dataset are shown: boreholes that reaches the bedrock (BH1) and not reaches the bedrock (BH2). Areas with bedrock outcrops (BO) are also presented.

Figure 2 Soil profile with coarse grained/filling material (a), clay (b) and coarse grained material (c). The lens with coarse grained material in the clay layer is in the model assumed to belong to the clay layer. The bedrock-level model considers both boreholes that reach the bedrock (BH1) and do not reach the bedrock (BH2). The soil strata model use information from both complete and incomplete drillings of the clay layer.

Figure 3 Interpolation with BH1 (B, orange line); BH1 and BH2 (C, blue line); and the lowest level of the two interpolations (D, hatched line). The location of the cross-section is indicated in Figure 6.

Figure 4 Histogram for bedrock altitude levels according to drillings grouped in BH1.

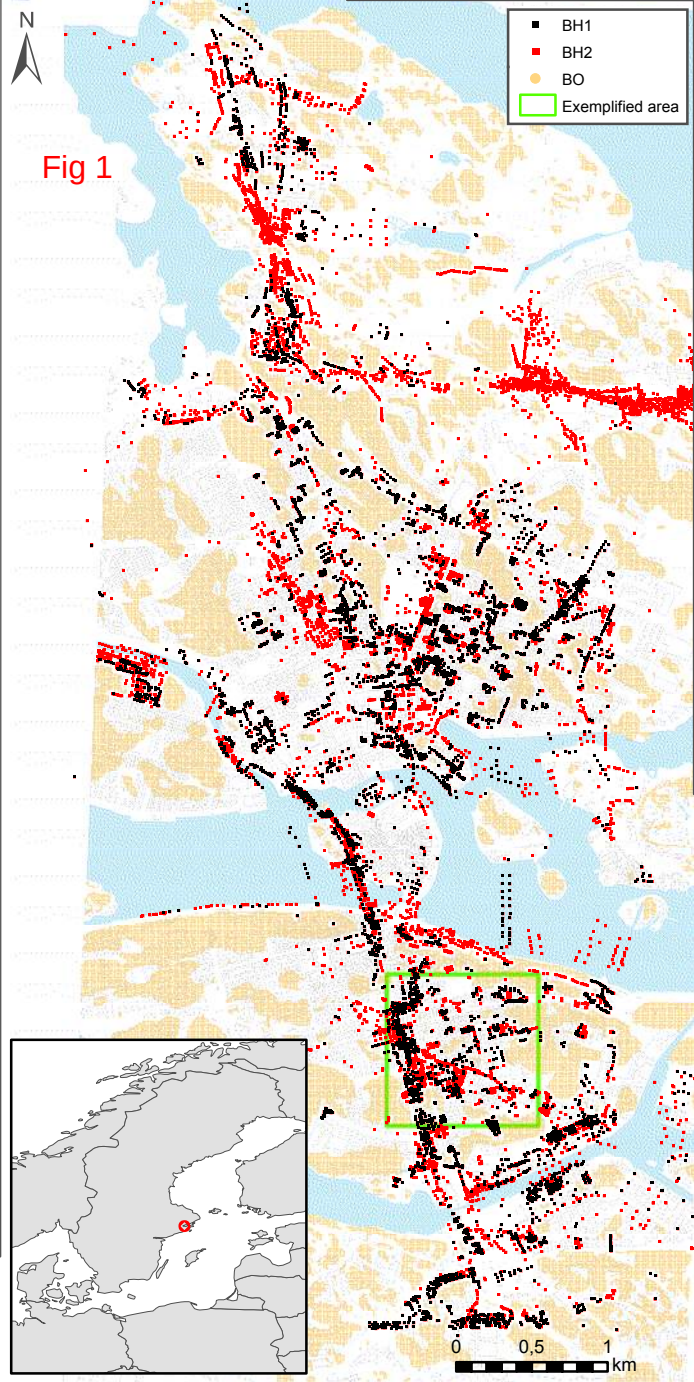
Figure 5 Experimental (points) and theoretical (blue and red lines) variogram from BH1 in two directions: 168 and 78.

Figure 6 Interpolation steps A-I for bedrock-level and soil-thickness model. The figure shows a section of the whole model. The distance between the ticks on the axis is 100 meters. Black squares shows BH1, white squares BH2 and red, clustered, points BO. Step E shows support lines between areas with BO as dotted lines. The line in the middle of each figure shows the location of the cross-section in Figure 7.

Figure 7 Interpolation steps B-H for bedrock-level and soil-thickness model (dotted area + grey area) together with a graph showing modelled bedrock level without use of support lines (soil thickness represented by the dotted area). The location of the cross-section is presented in Figure 6 A.

Figure 18 Boxplots of residuals between validation dataset and simulated percentiles (5th, 10th, 50th, 90th and 95th highest clay thickness) on Y-axis and clay thickness at validation points on X-axis.

Figure 19 Risk area for ground subsidence where the groundwater pressure level of the confined aquifer covers at least one meter of the clay thickness. The outmost area shows the 95th percentile, then follows the 90th, 50th 10th and 5th percentiles respectively.


N
BH1
BH2
BO
Exemplified area
Fig 1
0
0,5
1
km

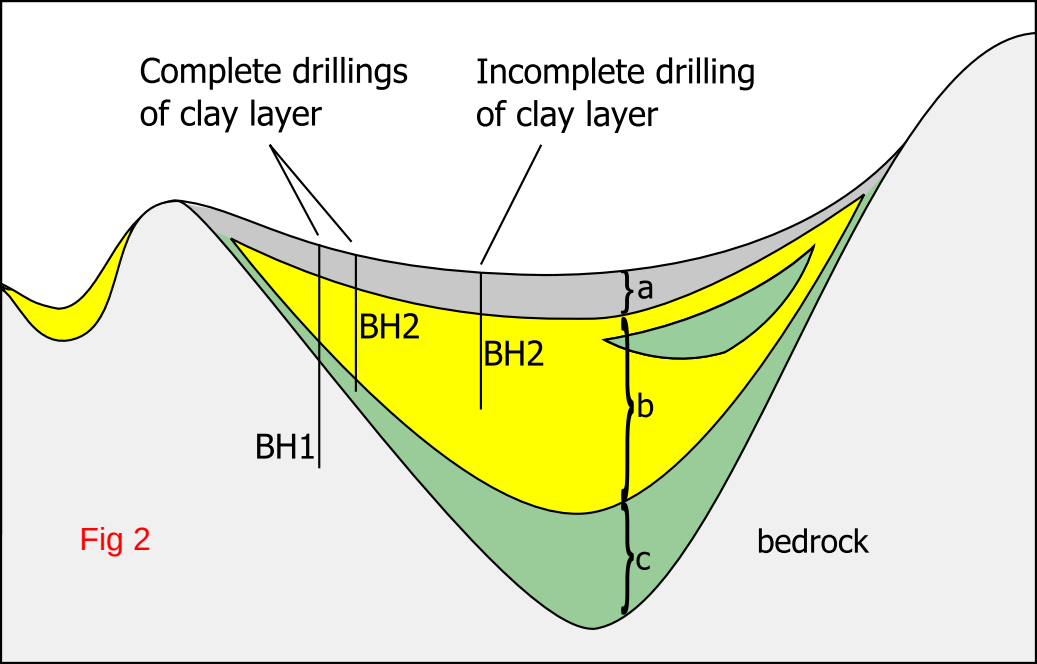
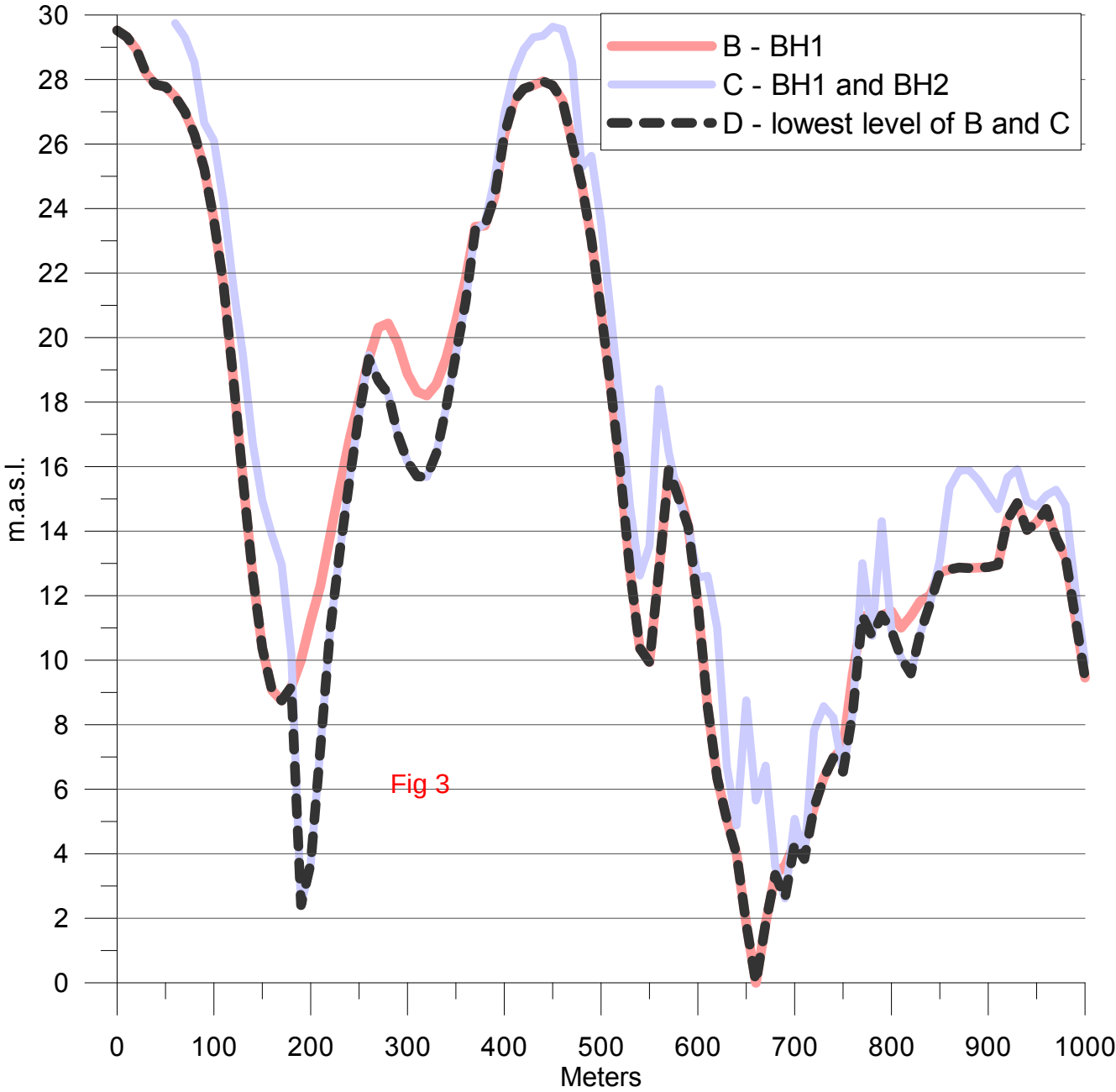

Complete drillings
of clay layer
Incomplete drilling
of clay layer
BH2
BH2
BH1
a
b
c
bedrock

Fig 2


B - BH1
C - BH1 and BH2
D - lowest level of B and C
m.a.s.l.
30
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0
0
100
200
300
400
500
600
700
800
900
1000
Meters
Fig 3

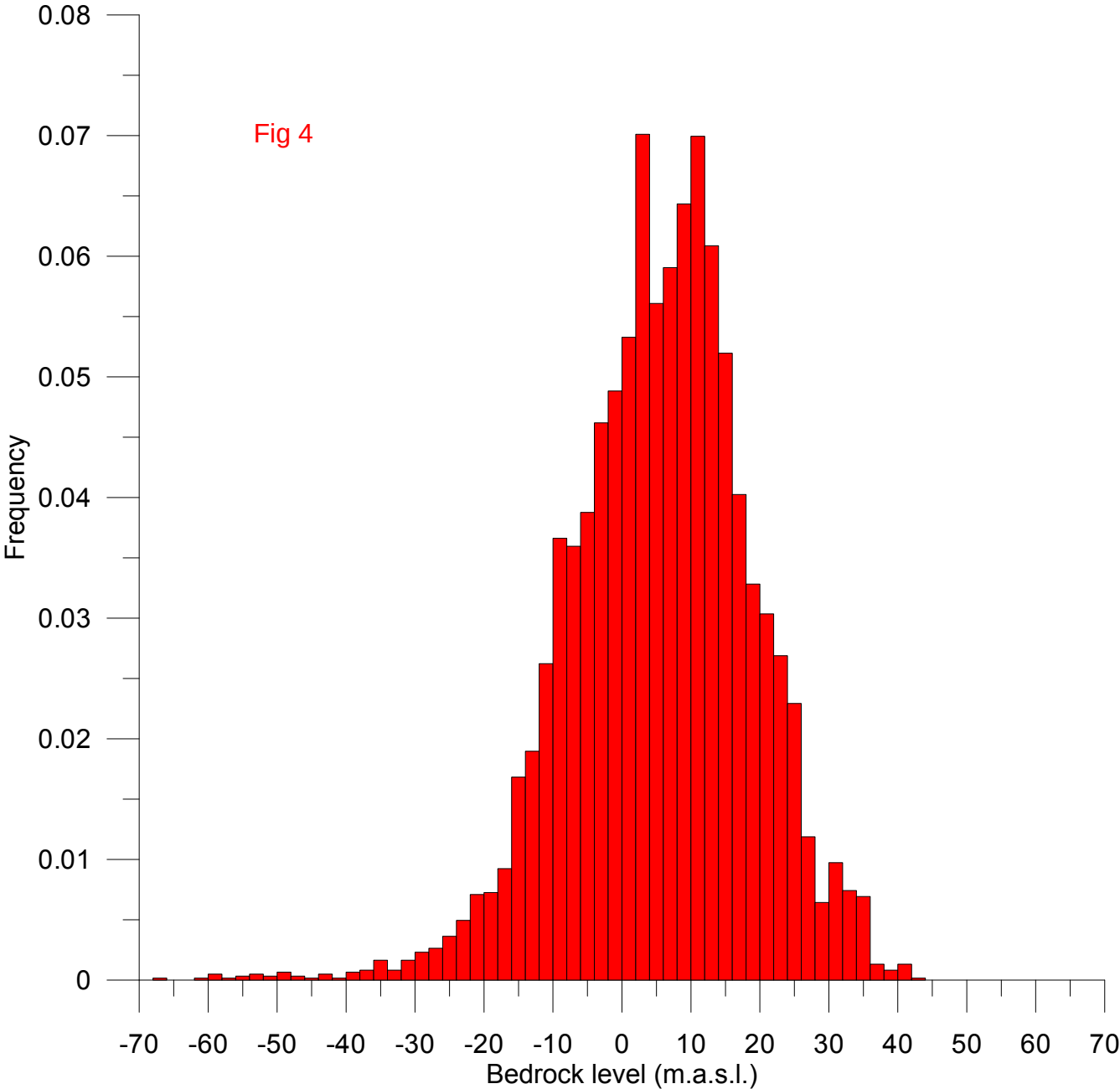


Fig 4
Frequency
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
-70
-60
-50
-40
-30
-20
-10
0
10
20
30
40
50
60
70
Bedrock level (m.a.s.l.)

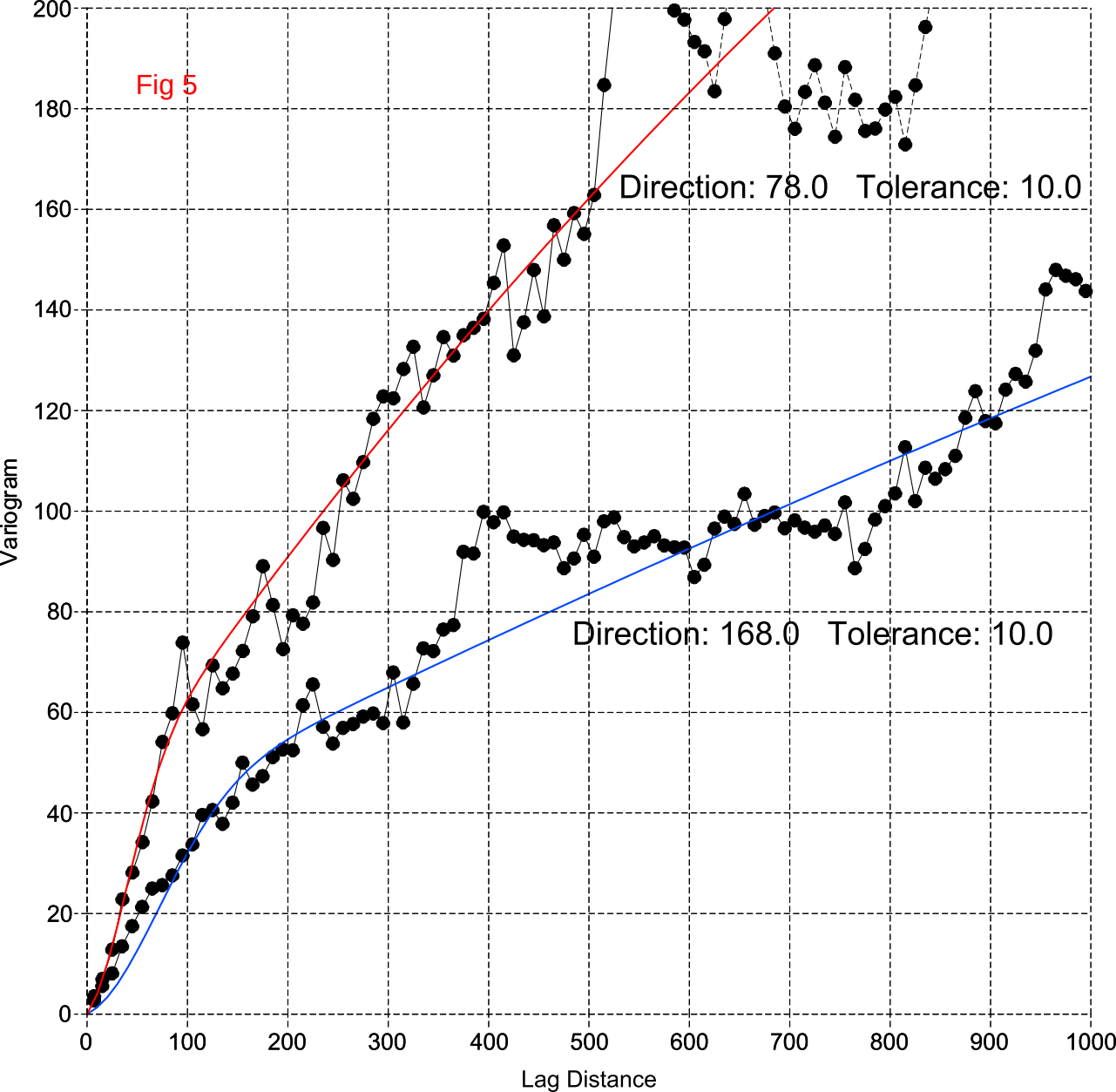
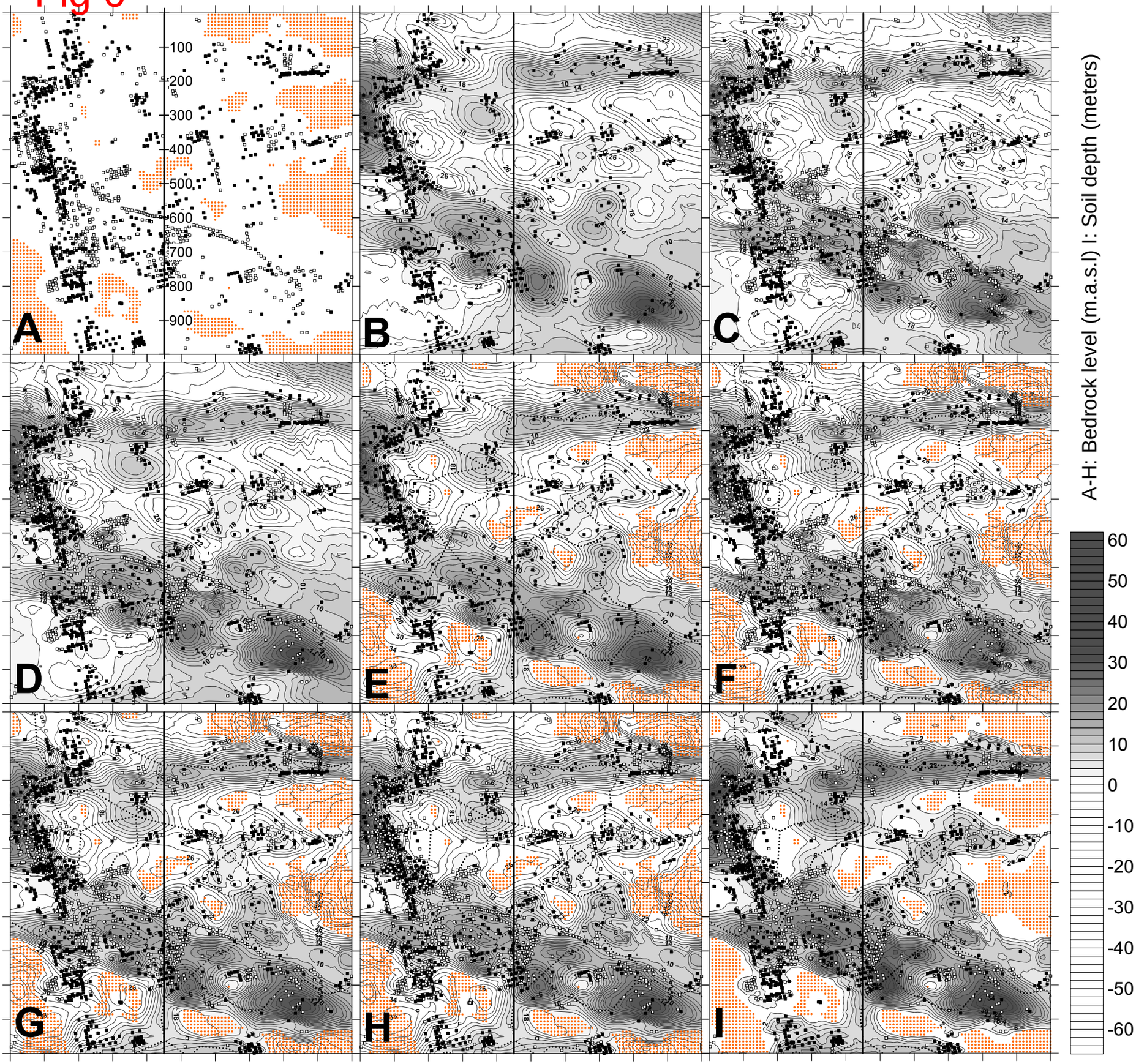


Fig 5
Direction: 78.0 Tolerance: 10.0
Direction: 168.0 Tolerance: 10.0
Variogram
Lag Distance
0
20
40
60
80
100
120
140
160
180
200
0
100
200
300
400
500
600
700
800
900
1000

Fig 6


A
B
C
D
E
F
G
H
I
A-H: Bedrock level (m.a.s.l) I: Soil depth (meters)
60
50
40
30
20
10
0
-10
-20
-30
-40
-50
-60

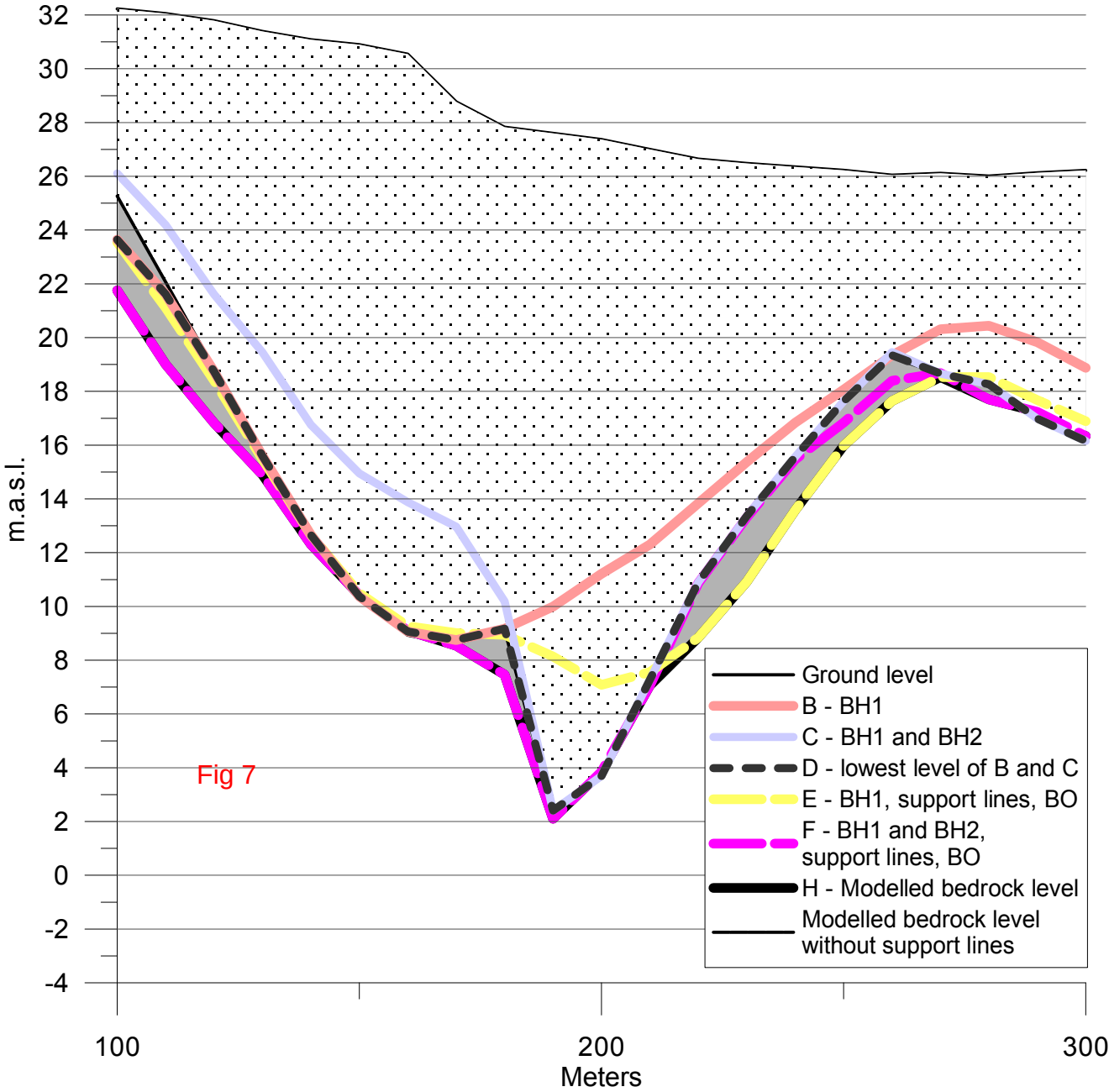
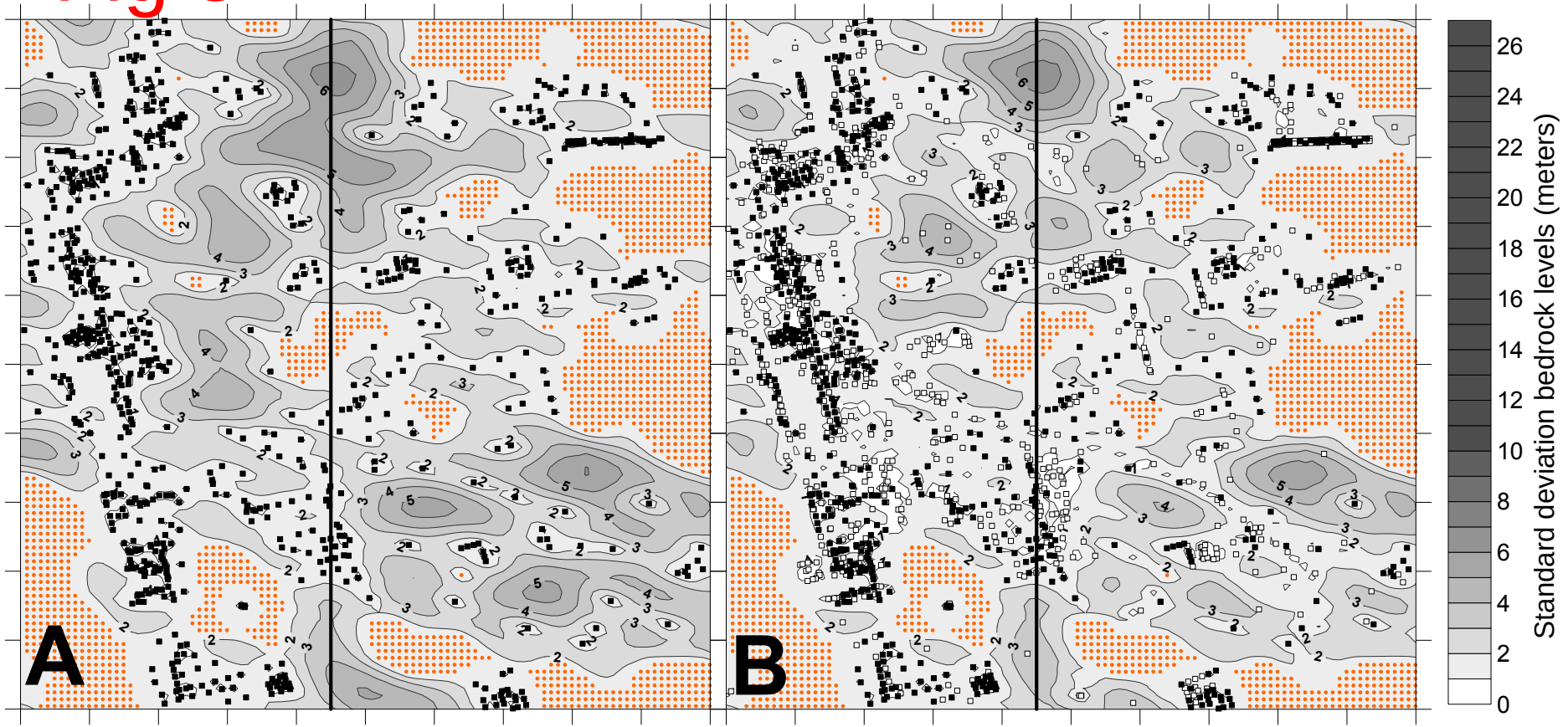
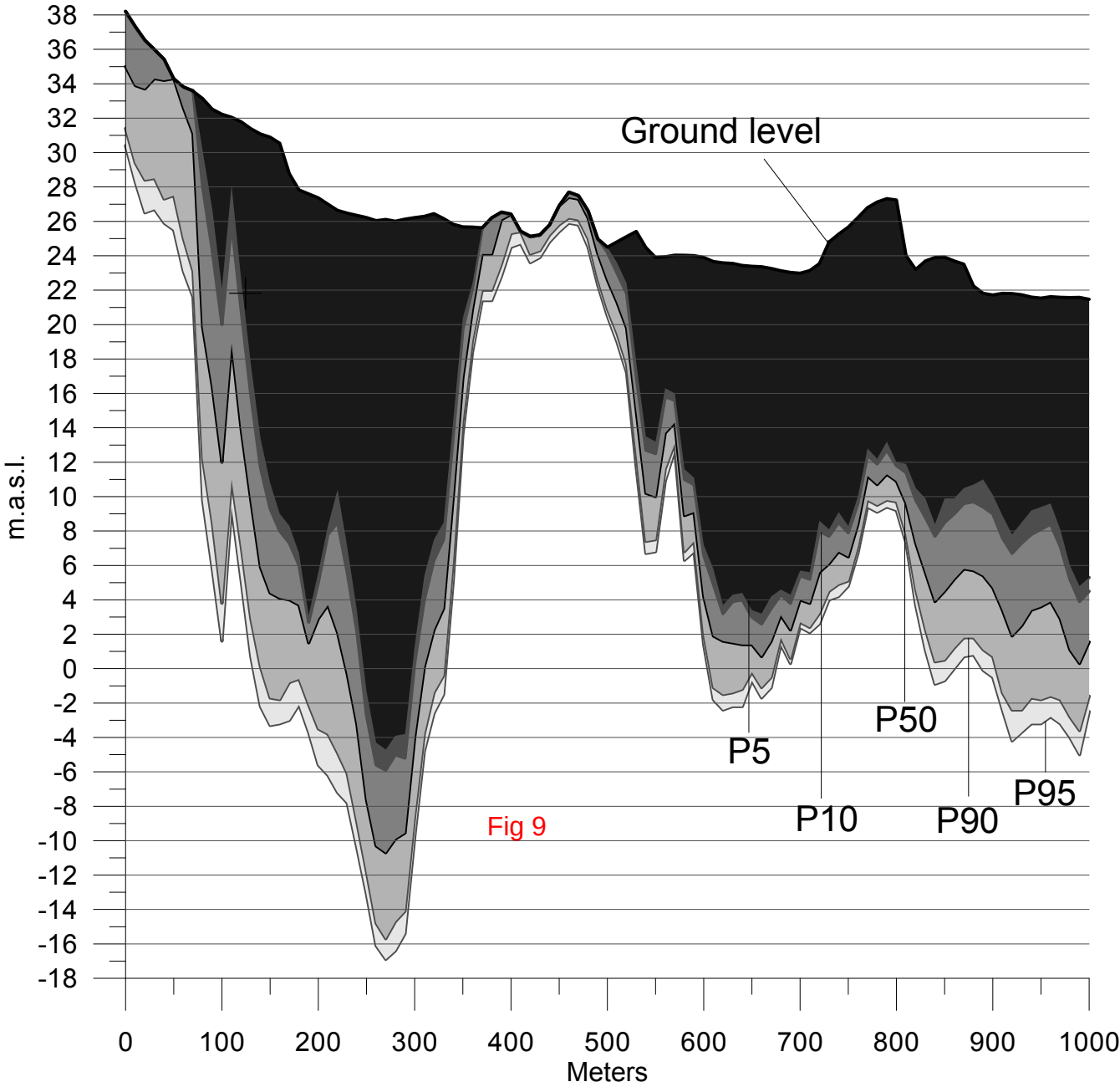

32
30
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0
-2
-4
m.a.s.l.
100
200
300
Meters
Fig 7
Ground level
B - BH1
C - BH1 and BH2
D - lowest level of B and C
E - BH1, support lines, BO
F - BH1 and BH2, support lines, BO
H - Modelled bedrock level
Modelled bedrock level without support lines

Fig 8


A
B
Standard deviation bedrock levels (meters)
0
2
4
6
8
10
12
14
16
18
20
22
24
26


Ground level
P5
P10
P50
P90
P95
Fig 9
m.a.s.l.
38
36
34
32
30
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0
-2
-4
-6
-8
-10
-12
-14
-16
-18
0
100
200
300
400
500
600
700
800
900
1000
Meters

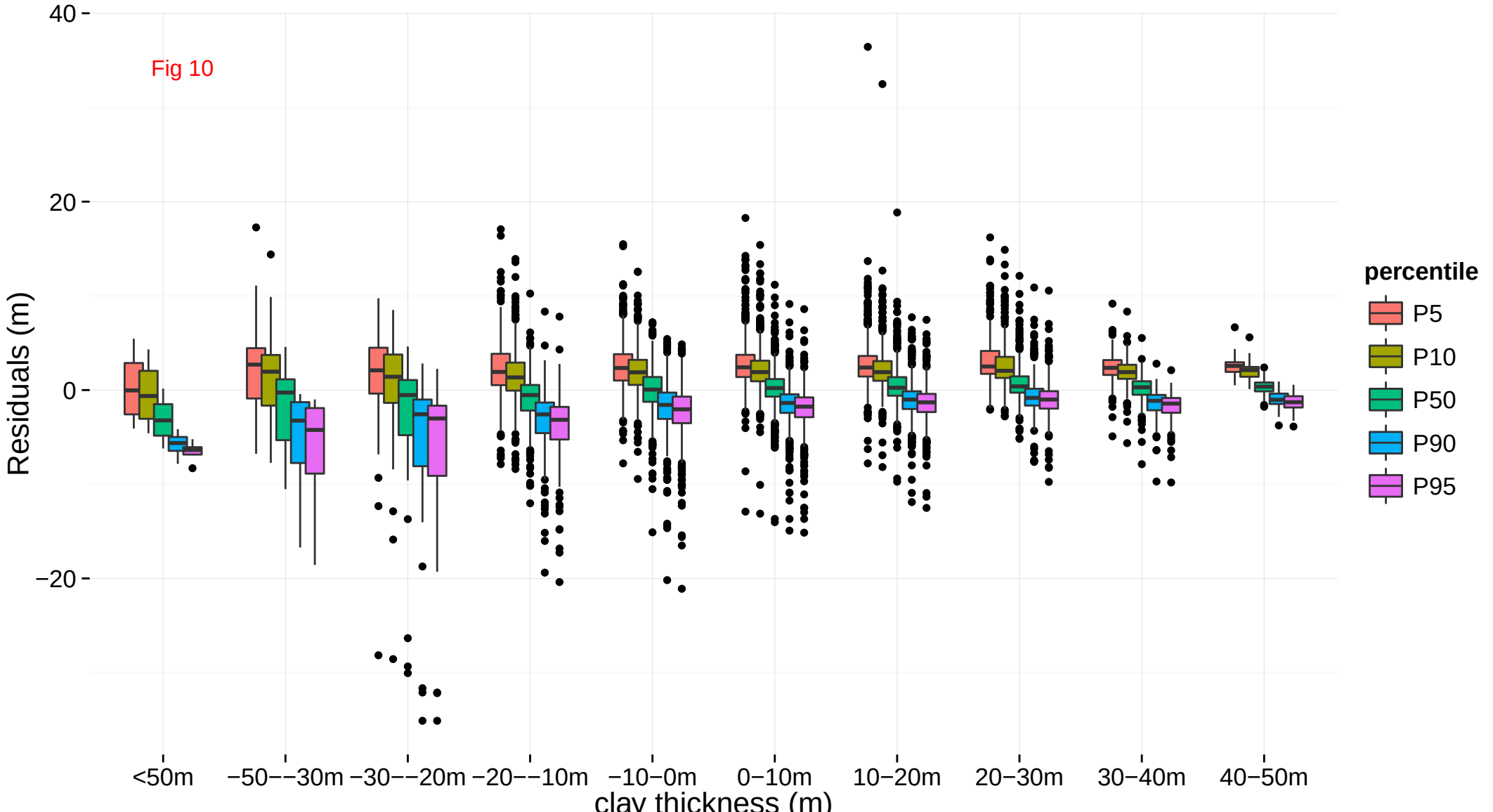
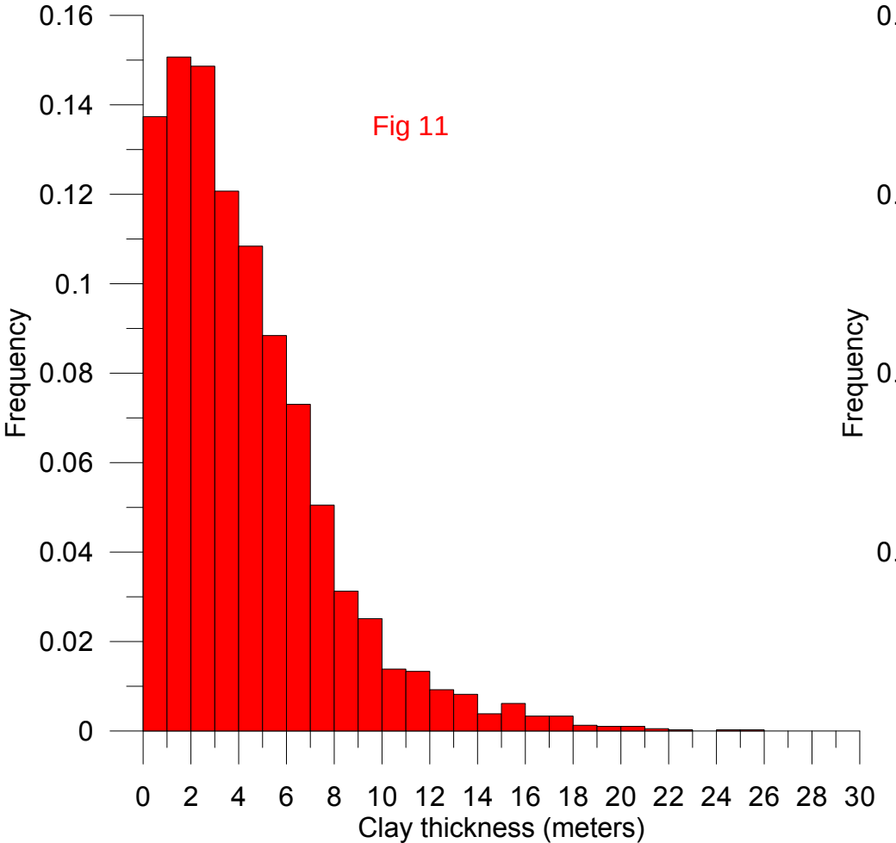
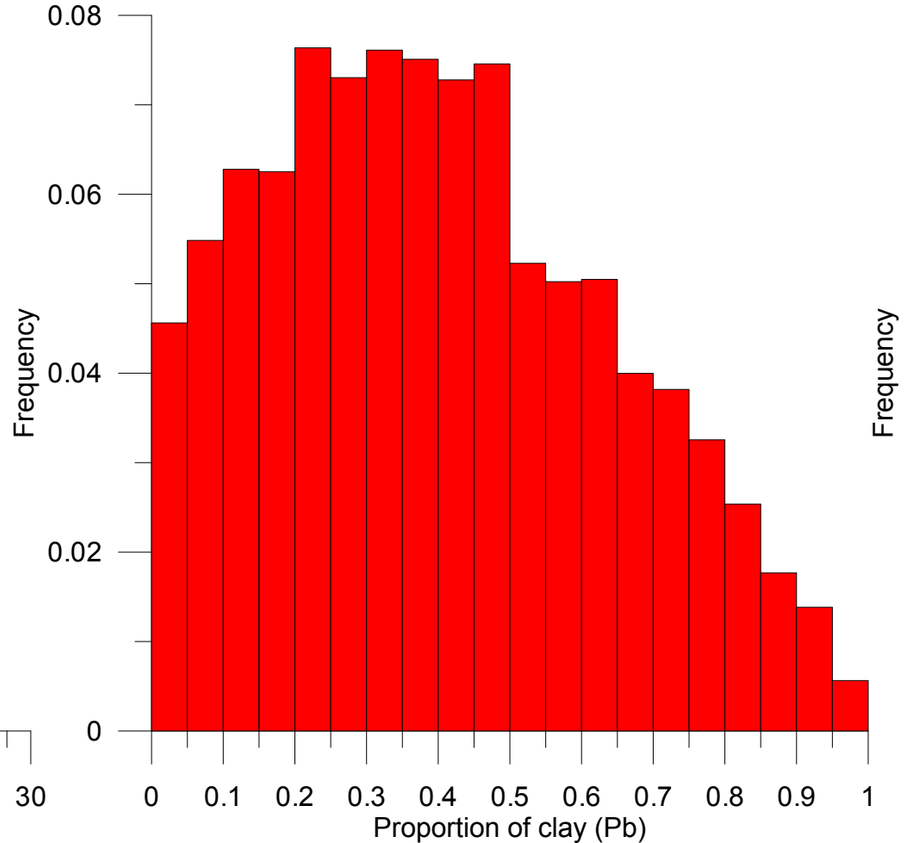
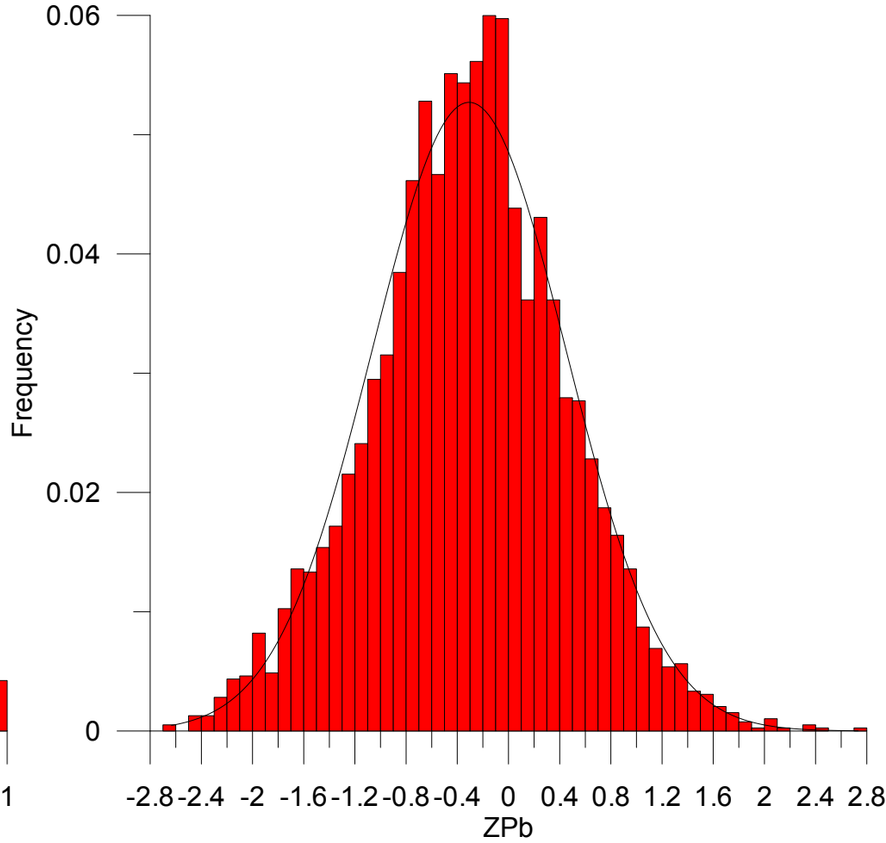


Fig 10
40
20
0
−20
Residuals (m)
<50m
−50−−30m
−30−−20m
−20−−10m
−10−0m
0−10m
10−20m
20−30m
30−40m
40−50m
clay thickness (m)
percentile
P5
P10
P50
P90
P95

Fig 11


Frequency
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
Clay thickness (meters)


Frequency
0.08
0.06
0.04
0.02
0
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Proportion of clay (Pb)


Frequency
0.06
0.04
0.02
0
-2.8 -2.4 -2 -1.6 -1.2 -0.8 -0.4 0 0.4 0.8 1.2 1.6 2 2.4 2.8
ZPb

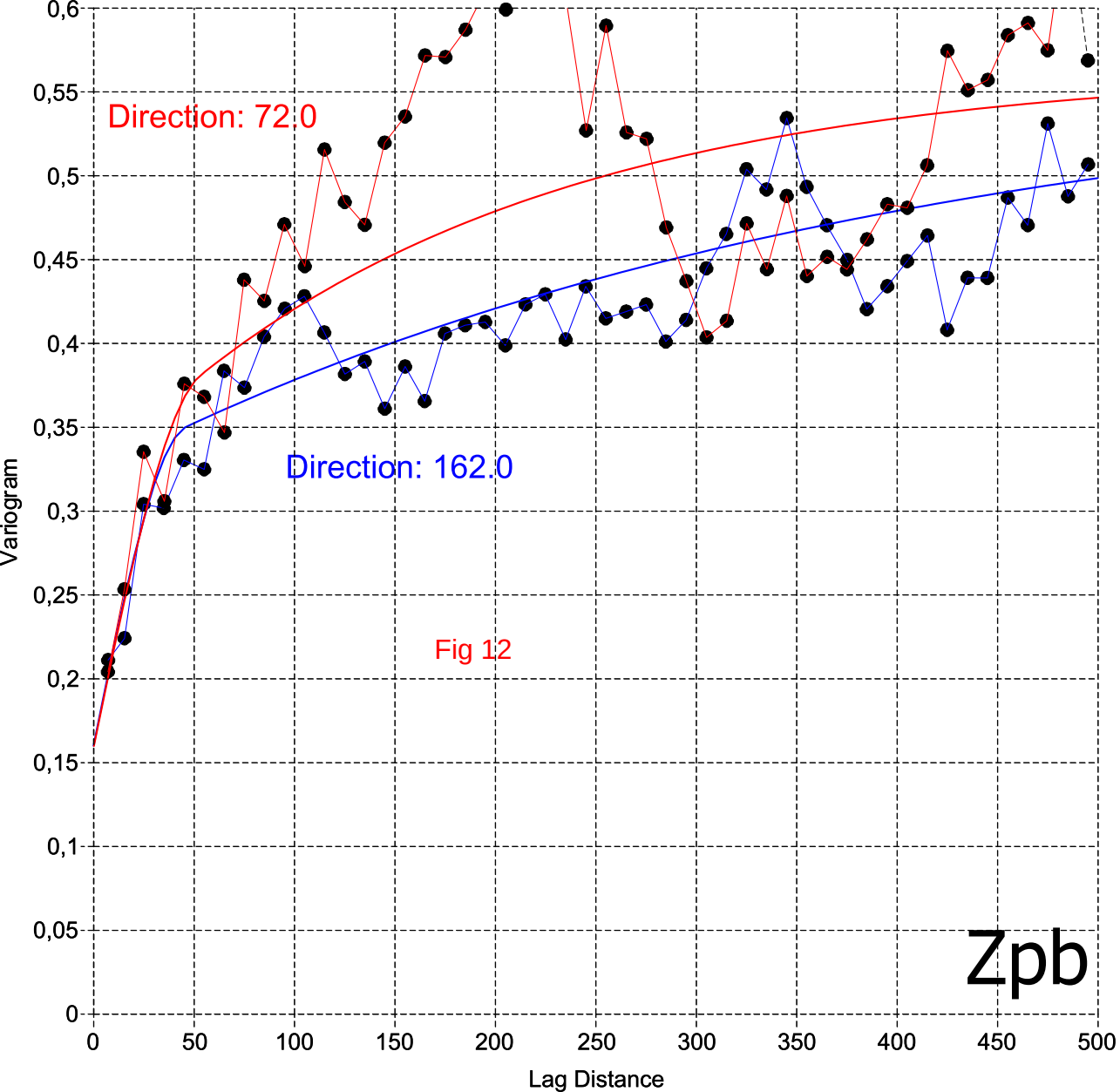

Direction: 72.0
Direction: 162.0
Variogram
Lag Distance
0
0,05
0,1
0,15
0,2
0,25
0,3
0,35
0,4
0,45
0,5
0,55
0,6
0
50
100
150
200
250
300
350
400
450
500
Zpb

Fig 12

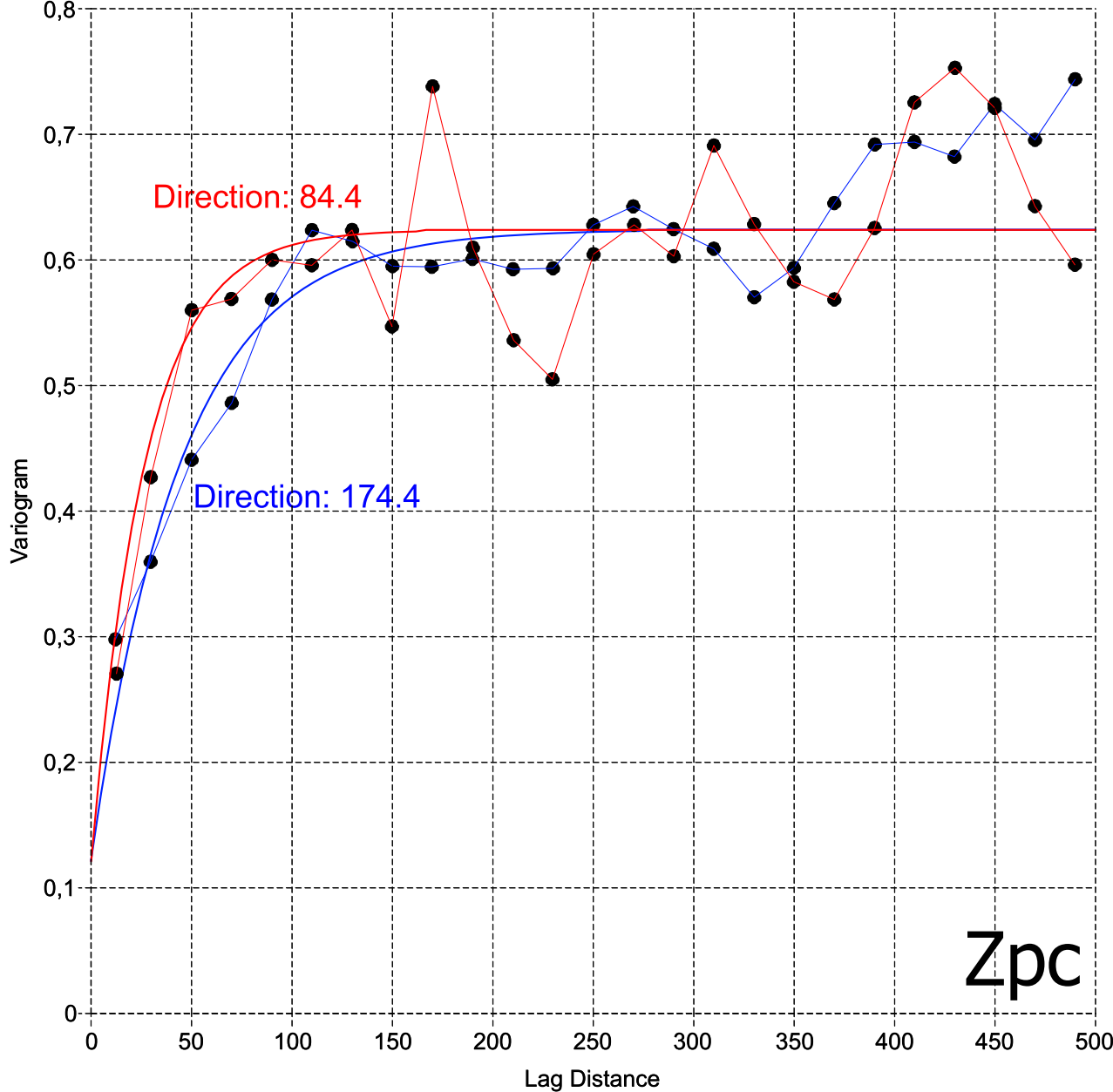

Direction: 84.4
Direction: 174.4
Variogram
Lag Distance
0
0,1
0,2
0,3
0,4
0,5
0,6
0,7
0,8
0
50
100
150
200
250
300
350
400
450
500
Zpc

Fig 13

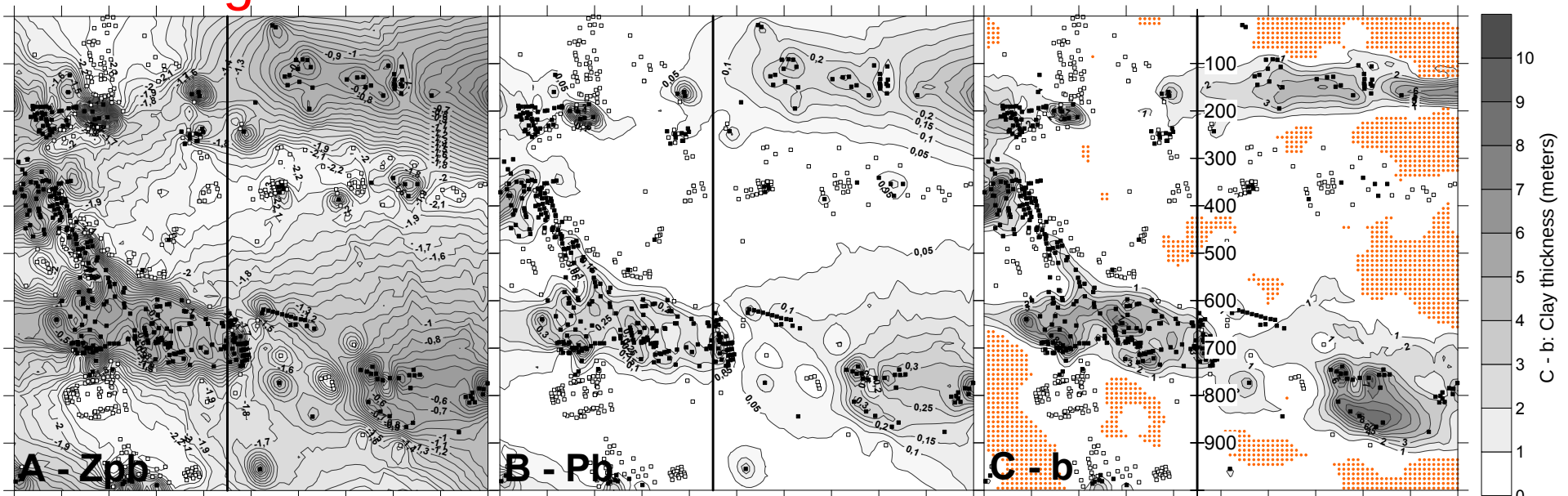
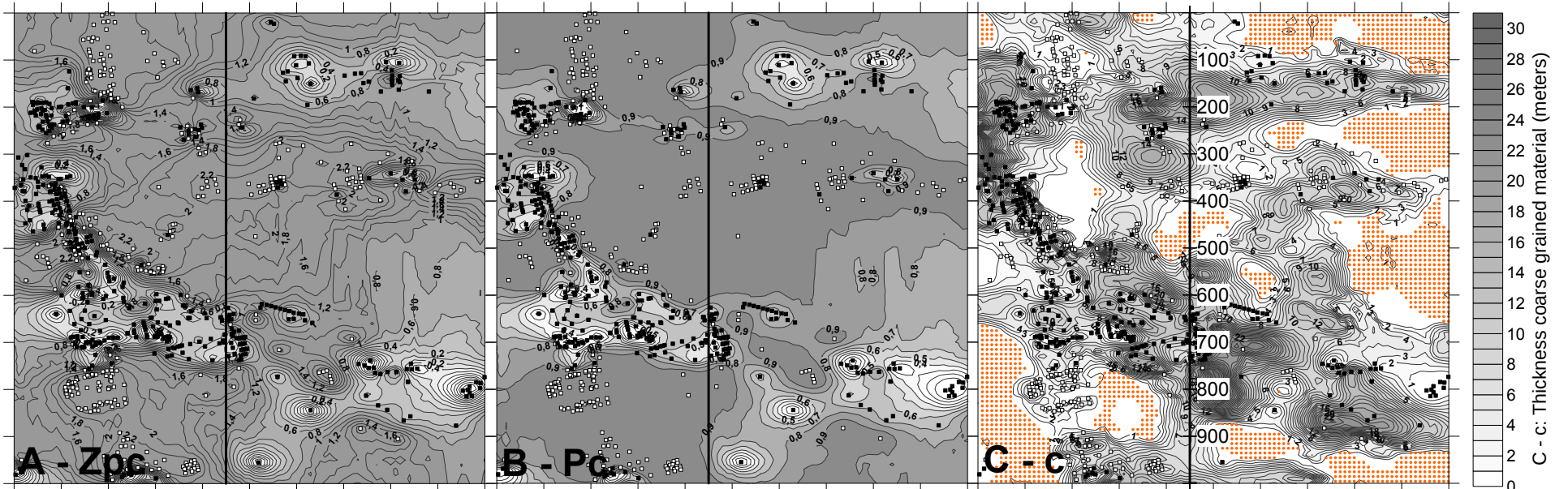


Fig 14


A - Zpc
B - Pc
C - c
100
200
300
400
500
600
700
800
900
C - c: Thickness coarse grained material (meters)
30
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0

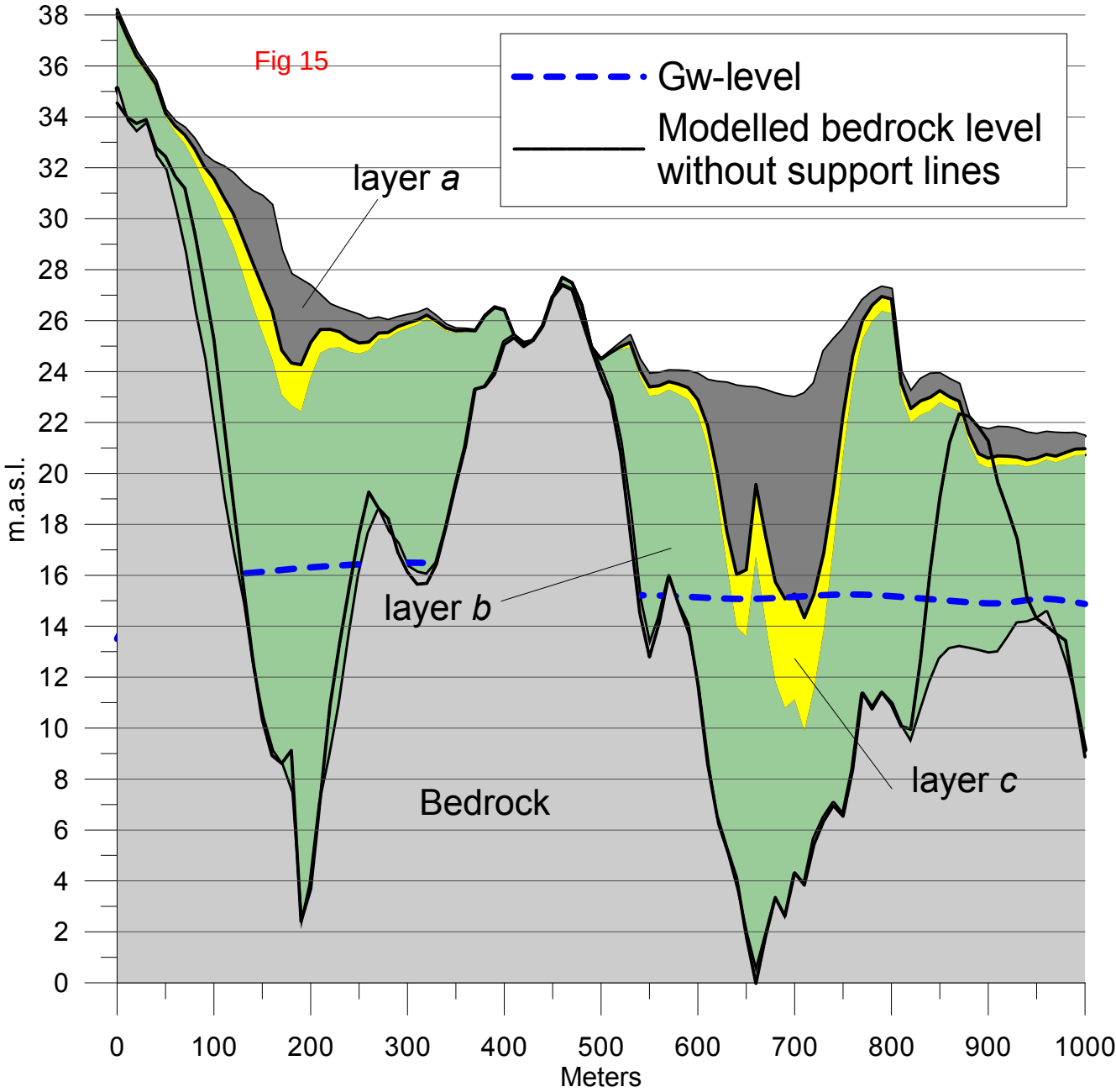
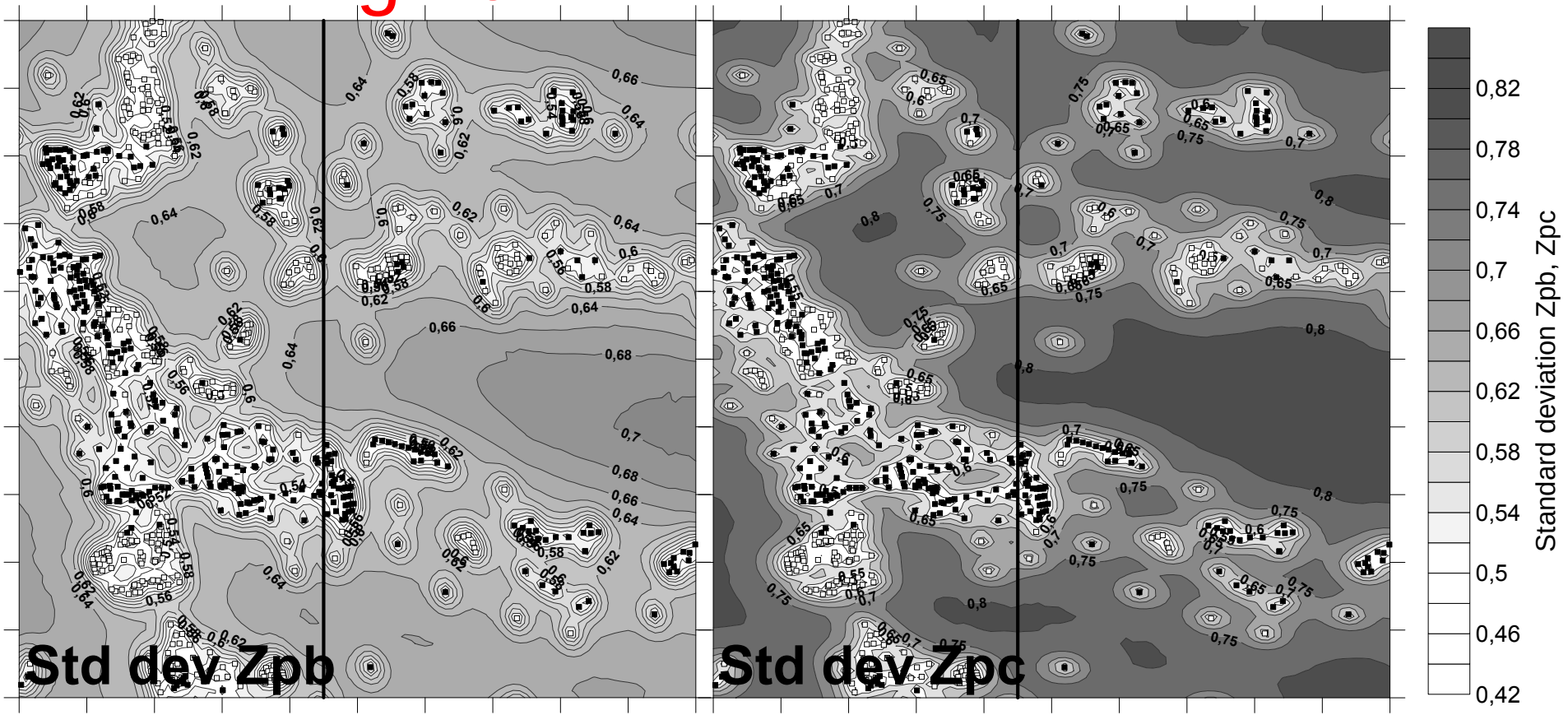


Fig 15
Gw-level
Modelled bedrock level
without support lines
layer a
layer b
layer c
Bedrock
m.a.s.l.
Meters
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
0
100
200
300
400
500
600
700
800
900
1000

Fig 16


Std dev Zpb
Std dev Zpc
Standard deviation Zpb, Zpc
0,82
0,78
0,74
0,7
0,66
0,62
0,58
0,54
0,5
0,46
0,42

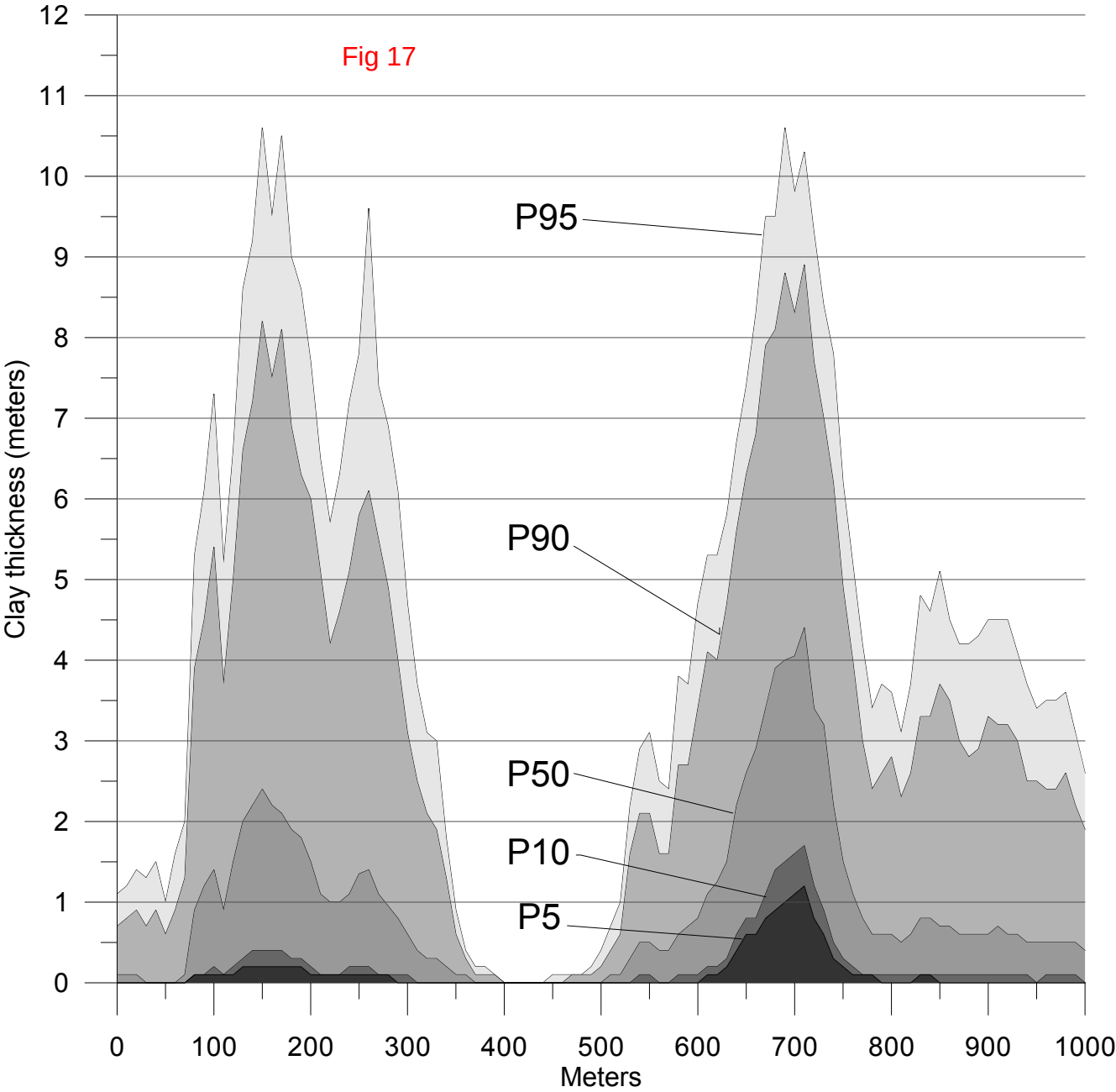


Fig 17
P95
P90
P50
P10
P5
Clay thickness (meters)
0
1
2
3
4
5
6
7
8
9
10
11
12
0
100
200
300
400
500
600
700
800
900
1000
Meters

Fig 18

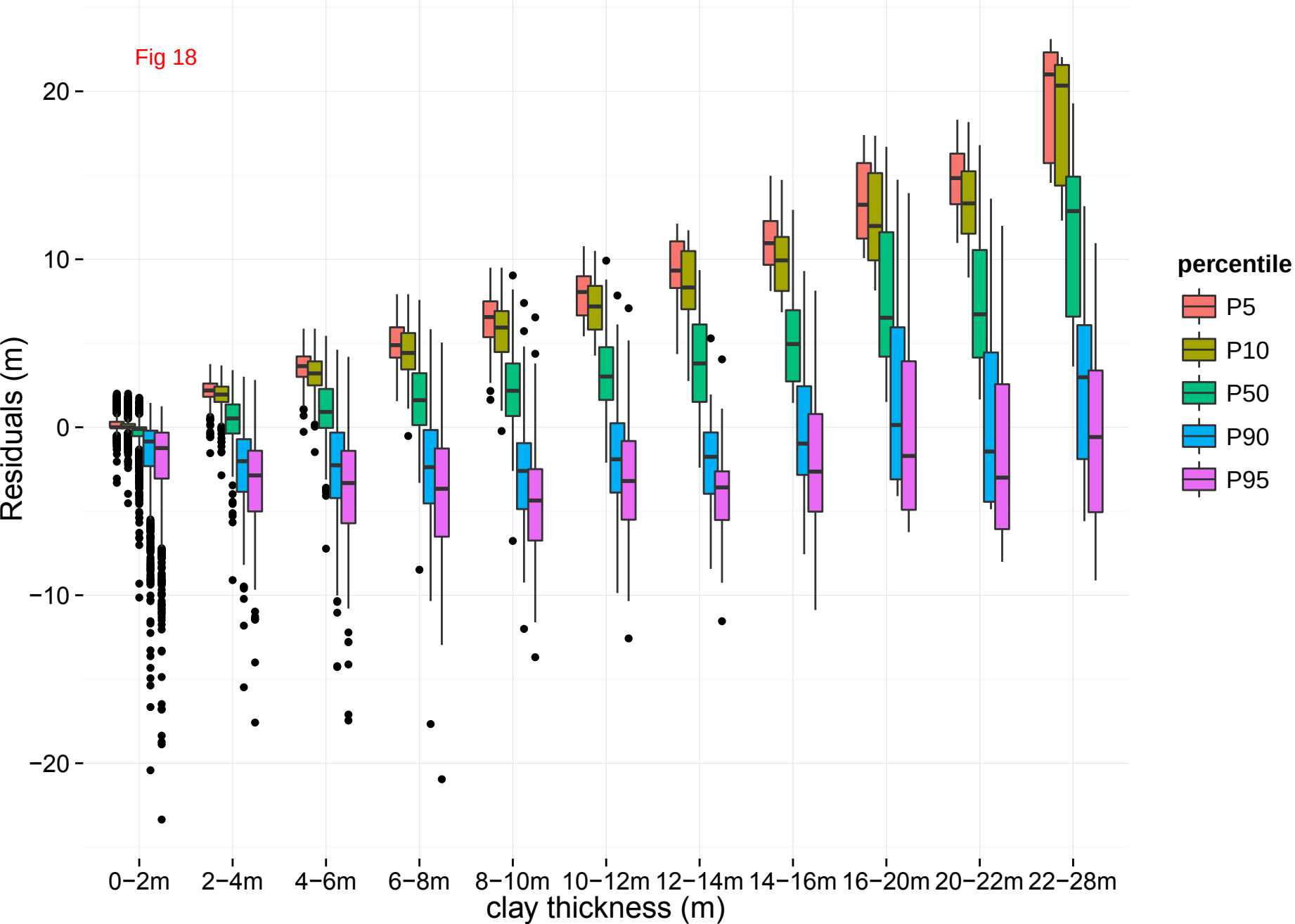

Residuals (m)
20
10
0
−10
−20
0−2m
2−4m
4−6m
6−8m
8−10m
10−12m
12−14m
14−16m
16−20m
20−22m
22−28m
clay thickness (m)
percentile
P5
P10
P50
P90
P95

Fig 19