A statistical assessment of the uncertainty in a 3-D geological framework model

R.M. Lark[1*], S.J. Mathers[1], S. Thorpe[1], S.L.B. Arkley[2], D.J. Morgan[1], D.J.D. Lawrence[2].

[1]*British Geological Survey, Keyworth, Nottinghamshire NG12 5GG, U.K.,* [2]*British Geological Survey, Murchison House, West Mains Road, Edinburgh, EH9 3LA, U.K.*

Abstract

Three-dimensional framework models are the state of the art to present geologists' understanding of a region in a form that can be used to support planning and decision making. However, there is little information on the uncertainty of such framework models. This paper reports an experiment in which five geologists each produced a framework model of a single region in the east of England. Each modeller was provided with a unique set of borehole observations from which to make their model. Each set was made by withholding five unique validation boreholes from the set of all available boreholes. The models could then be compared with the validation observations. There was no significant between-modeller source of variation in framework model error. There was no evidence of systematic bias in the modelled depth for any unit, and a statistically significant but small tendency for the mean error to increase with depth below the surface. The confidence interval for the predicted height of a surface at a point ranged from ±5.6 m to ±6.4 m. There was some evidence that the variance of the model error increased with depth, but no evidence that it differed between modellers or varied with the number of close-neighbouring boreholes or distance to the outcrop. These results are specific to the area that has been modelled, with relatively simple geology, and reflect the relatively dense set of boreholes available for modelling. The method should be applied under a range of conditions to derive more general conclusions.

Keywords: 3-D Geological Modeling; 3-D Visualization; Uncertainty; Error; Linear Mixed Model.

*Corresponding author: *E-mail address*: mlark@bgs.ac.uk (R.M. Lark).

1. Introduction

Geological objects are three-dimensional (3-D), but for many years geological surveyors were constrained by available technology to present their understanding in the form of two-dimensional maps, based on their conceptual 3-D models. Geology can now be represented in 3-D by computer technology, and users of geological information now generally recognize that 3-D representations of the subsurface are necessary to support planning and decision-making (e.g. Mathers and Kessler, 2010; Royse et al., 2010). Furthermore, the users of 3-D geological information may also require measures of uncertainty. An example is provided by the construction of the Channel Tunnel (Blanchin and Chilès, 1993). The aim of the tunnel engineers was to stay within the Chalk Marl, avoiding both the underlying Gault Clay and the overlying, altered and fractured, Grey Chalk. The output of a geostatistical analysis of borehole observations of the depth to the top of the Gault Clay allowed engineers to quantify for proposed tunnel routes the risk of tunnelling into the Gault, and so to adjust their plans to reduce this risk to acceptable levels.

In this paper we are concerned with what we call *3-D geological framework models*. These represent the distribution of lithostratigraphic units in 3-D, so they are, effectively, 3-D geological maps. We use the term 'framework models' to distinguish them from the statistical models discussed below which are used in the analysis of the data on errors in the framework model. Each unit in a framework model is bounded by an upper and lower surface, and may be folded and faulted. A single unit is a 3-D object (also called a *volume* or a *shell*), and the model allows the user to predict which unit would be found at a particular position in 3-D and, by extension, the sequence of units that would be encountered by a borehole at any point, or a tunnel or section along a specified route.

There are various ways to generate geological information in 3-D. In purely geostatistical prediction from data, the user develops a statistical model of some continuous variable of interest (e.g. depth to a given stratigraphic horizon). This statistical model may comprise both systematic trends and a random component (Stein, 1999), and defines

a prediction distribution for the modelled spatially distributed (*regionalised*) variable at any unsampled location, conditional on the observed data. The expected value of this distribution (its mean value) is generally computed using the geostatistical method of *kriging* (Matheron, 1963), as a predicted value with an associated prediction error variance. However, more complex results (including uncertainty measures not just at points) can be obtained by *conditional simulation*: the computer-based simulation of a spatially distributed variable such that it honours both the spatial dependence of the data and its values at known data points (Journel and Huijbregts, 1978; Goovaerts, 1997). Once the user has made any necessary decisions about the structure of the statistical model, the procedure uses an algorithm to generate both the predictions and their associated measures of uncertainty.

The production of a geological framework model may, in principle, be performed by a geostatistical algorithm, and geostatistics has been used to predict the elevation of one or more surfaces as continuous variables (see, for example, Blanchin and Chilès, 1993; Lark and Webster, 2006). However, this approach is not always suitable. In particular, it requires both reasonably uniformly spatially distributed observations, and sufficient observations to support the statistical modelling. Furthermore, in a purely geostatistical algorithm, there is limited scope to constrain the possible form of particular geological units through interpretation of available observations in the light of geological expertise.

The *Geological Surveying and Investigation in 3D* (GSI3D) software (Kessler et al., 2009; described in detail by Mathers et al., 2011), is designed specifically to exploit the ability of the geologist to interpret data in the light of geological knowledge and a capacity to visualise 3-D structure which is consistent with his or her understanding of processes and landscape evolution. In summary, the geologist produces a set of interlocking cross-sections as the result of expert interpretation of available borehole data and surface geological map line-work, coupled to a digital terrain model (DTM; Kessler and Mathers, 2004). The areal distribution of each geological unit in the stack is then defined

and, finally, the 3-D model is calculated using the triangle-based tesselation method of Delaunay (1934). The final model thus comprises closed 3-D objects which correspond to the stratigraphic units represented in the cross-sections and geological succession. Automated steps in the procedure are limited to the final triangulations, and are not based on a statistical model. This means that, in contrast to the output of a purely geostatistical algorithm, there is no model-derived measure of uncertainty associated with the 3-D framework model. Hence, a different approach is required in the evaluation of the inherent uncertainty in 3-D framework models, such as those produced in GSI3D, which are not produced by statistical prediction.

Lelliott et al. (2009) proposed a procedure which synthesises information on different factors that, *a priori*, are expected to contribute to the uncertainty of a framework model. These include the local geological complexity, and the distances from a location under consideration to neighbouring boreholes. Their procedure indicates how the uncertainty of a model may be expected to vary spatially, and it was shown to produce results consistent with experts' expectations. However, it is not a direct quantification of uncertainty, such as would allow one to give, for example, *confidence intervals* for the predicted depth of a particular surface at some location. By a confidence interval is meant a range of values which, given the variability of the target variable, can be expected to contain the true value of the variable with a specified probability (commonly 95%). The magnitude of uncertainties in 3-D framework models remains unknown.

In this paper we present the results of an experiment to investigate directly the uncertainty in 3-D framework models of a region produced by a methodology, such as that of GSI3D, which depends primarily on expert interpretation rather than automated interpolation. The objective was to obtain quantitative measures of model uncertainty, and to investigate factors that might explain variation in the uncertainty of the model from one place to another. The basic idea was to compare a framework model with a set of boreholes which had not been used to produce it. The concept is simple, but it was necessary

to design the experiment in such a way that sufficient validation boreholes could be used without unduly reducing the density of boreholes available to establish the framework model. Furthermore, as the available boreholes were not distributed according to a statistical design yielding a known probability distribution for their inclusion, model-based statistical analysis was necessary to obtain estimates of appropriate statistical measures of uncertainty.

2. Materials and methods

2.1 The general approach

The geologist is asked to produce a framework model of a region, with GSI3D, using a dataset from which a subset of the available boreholes has been withheld. The framework model predictions can then be compared with the observations at these *validation boreholes*. If we aim for a validation subset of about 100 boreholes, then even in a reasonably densely investigated study area, the density of boreholes remaining for modelling will be reduced significantly. For this reason we decided to use a team of five modellers, and each geologist independently produced a framework model of the study region, according to a common set of instructions. The modellers, however, worked with unique, but partly overlapping, subsets of the available boreholes. Each of these subsets was formed by withholding a unique, non-overlapping validation subset from the full set of boreholes. Hence, any borehole in the validation subset for the ith modeller occurs in the modelling subset used by each of the other modellers.

The validation subsets of boreholes for each modeller together constitute an overall validation set. Each borehole in this set belongs to the validation subset for just one of the five independent models, and a measure of framework model error for that validation borehole is obtained by comparing it with that model, for which it is a validation, rather than a modelling, borehole. For example, one may calculate the difference between the modelled and observed height of a particular surface at the location of a given validation

borehole.

Hence a total of, say, 100 observations of framework model error can be obtained by withholding a validation subset of just 20 unique boreholes from each of the five modellers.

One consequence of this, of course, is that one source of variation in the observations of framework model error may be differences between the modellers, but this contribution can also be assessed in the analysis. From the perspective of the geological survey organization this is a potentially important component of variation in framework model because, if modeller differences are a significant source of variation then this would imply that it is necessary to improve the consistency of modeller performance. Details of the statistical modelling with these data are described in Section 2.5 below.

2.2 The study area

The area used in this study was the TM24 Ordnance Survey mapsheet around Ipswich and Woodbridge, Suffolk in Eastern England which covers an area of 10-km×10-km. The British National Grid Co-ordinates of the south-west corner of the area are 620 000, 240 000. The geology of the study area has been surveyed at 1:10 000 scale in recent times and published as 1:50 000 geological map sheets (British Geological Survey, 2001; 2006) together with accompanying explanatory accounts (Mathers and Smith, 2002; Mathers et al., 2007). For the purposes of this exercise the modellers were asked to model only a subset of the units. These were:

- Thames Group: Eocene clay and silty clay comprising both the London Clay and Harwich formations (undifferentiated).
- Red Crag Formation: Pliocene shelly, medium- to coarse-grained sand.
- Chillesford Sand: Pliocene sand. This did not, in fact, appear in the models or boreholes at any validation sites.
- Kesgrave Sand and Gravel: Quaternary (Cromerian to Beestonian).

- A sheet of Glacial Sand and Gravel overlying the Kesgrave Sand and Gravel but below and extending beyond the Lowestoft Till.
- Lowestoft Till: Quaternary Anglian diamicton.
- Worked Ground and Worked and Made Ground were included in the framework model where the modeller thought it necessary.

Post-Anglian Quaternary deposits confined to the valleys within the area were ignored for purposes of modelling (see Section 2.4 below). Figure 1 shows a 2-D map of the modelled units and an example cross-section within the area.

2.3 Borehole subsets

A total of 347 boreholes were available for the TM24 area. All were in the set of coded boreholes (i.e. interpreted in terms of the standard stratigraphy for the area), although it was discovered later that five has not actually been coded. All had been interpreted by the same geologist (SJM). This ensured a certain consistency and quality in the boreholes, a factor which will not hold in all study areas where 3-D framework models are made.

The borehole locations are not distributed at random across the area, but reflect purposive sampling decisions and the range of purposes for which the original investigations were made (see Figure 2). For this reason we cannot undertake design-based statistical analysis of data from these boreholes, or any subset of them, as though an independent random set of sites had been selected from across the area according to a fully-specified design. However, to ensure a reasonable distribution of validation sites across the area, in order to draw a sample it was first divided into nine equal square blocks (called *strata* in statistical terms) each of side length 3 333.3 m. Six of these nine strata contained a large number of boreholes, but three of them contained 8, 10 and 13 boreholes respectively. For each modeller, it was decided to draw one unique validation borehole at random from each of these last three strata, and three, again at random, from each of the remaining six strata. This gave 21 validation boreholes per modeller, a total of 105. Figure 2 shows

the overall distribution of boreholes across the area, with the validation boreholes for Modeller 1 highlighted.

2.4 Modelling

Ideally the modellers would have been selected at random from a population of individuals involved in 3-D framework modelling at the British Geological Survey, but in practice we were constrained by the availability of staff to undertake the work. The set of five modellers had varying degrees of experience in 3-D framework modelling, in use of the GSI3D software and familiarity with the geology of the study area. Each modeller was provided with a subset, X_i, of 326 boreholes to use for modelling. In addition they were provided with the same rasterized topographic map of TM24, the same set of vector files representing the surface and bedrock geology as 2-D shapefiles and a common DTM.

Although between-modeller variation is of interest, we set out to constrain it to some extent in two ways. First, all modellers were given a common set of instructions about how to use the data and what assumptions to make as follows.

1. To proceed as they would in any normal geological framework modelling project (Kessler et al., 2009).

2. To use a common version of the modelling software: GSI3D Version 2011.

3. To assume that the boreholes were correct with respect to their location (northings and eastings) and start height, but to ignore any obvious rogues.

4. To assume that the information on the surface outcrops (surface geological map) was correct.

5. When drawing sections, to display the outcrops as a colour band or ribbon along the trace of the DTM and to snap correlation lines to them.

6. To construct the distributions of each unit in plan view by combining the surface outcrop portion from the map with any subcrop distribution defined in the sections,

these distributions were snapped to the extent of the unit in section.

7. To draw 'docking' sections positioned along the bounding grid lines at the limits of the TM24 mapsheet to provide a boundary condition for the calculation of the model.

8. Not to model any valley-floor deposits, or isolated patches of Glacial Sand and Gravel within the valleys, either in the sections or as distributions. However their presence and likely thickness were to be taken into account when constructing the remaining geological units at depth.

9. To model only the units listed in Section 2.2 above.

10. To ignore any tiny patches of Glacial Sand and Gravel or Glacial Silt and Clay lying stratigraphically above the Lowestoft Till.

Second, once the models were complete all were inspected by an experienced modeller with knowledge of the area (also one of the modellers, SJM), and any obvious errors in the use of the software, rather than of interpretation, were discussed and corrected by the modeller before information was extracted to validate the framework model.

2.5 Data analysis

Once all the models had been completed, each framework model was interrogated at the locations of its unique subset of validation boreholes. The framework model elevations of the top and the base of each modelled unit were extracted and the corresponding boreholes were examined. At this stage, it was discovered that five of the boreholes had not been coded and the elevation data for a sixth were clearly in error. This left a full set of 99 validation data for the analysis.

2.5.1 Consistency

The first comparison between the models and the validation data was an assessment of their overall consistency, that is to say whether units included in the framework model at a site were present in the actual borehole. A unit was designated as present at a site

if it was recorded in the borehole record. It could be designated as absent only if there was evidence that the borehole was deep enough to record it should it be present. In this particular area, where the modelled units are not subject to marked folding, a unit could be unambiguously designated as absent if, and only if, it were not recorded in the borehole record but a stratigraphically-lower unit was recorded. The numbers of validation boreholes at which each unit could be identified either as present or unambiguously absent were: Lowestoft Till, 98; Glacial Sand and Gravel, 98; Kesgrave Sand and Gravel, 97; Red Crag, 89; and Thames Group, 66. The framework model was examined at each of these subsets of the validation sites to see whether the corresponding unit was present or absent, and the results were tabulated.

The primary interest of this study is in the agreement, or otherwise, between the modelled and observed positions of modelled objects in 3-D. We compared the observed and modelled height (with respect to Ordnance Survey datum) of the bases of the five modelled units. This comparison could be made, for any unit, in any validation borehole where the base of that unit is proven and the unit also appears in the framework model. Table 1 gives the numbers of validation boreholes, out of the full set of 99, for which the base of each unit was both proven and modelled. This shows that only in the cases of the Kesgrave Sand and Gravel and the Red Crag formation were there a sizeable number of validation sites where this criterion was met. For this reason, the initial analyses were done by grouping the observed and modelled heights for:

1. the base of the formation present at the surface;
2. the second base below the surface;
3. the Kesgrave Sand and Gravel; and
4. the Red Crag.

These rules for grouping mean that, at any validation borehole, there was at most one unit for which the observed and modelled height of the base was considered in any one of

these groups.

2.5.2 Error in the height of surfaces: simple linear models with stationary mean and variance

In order to establish a statistical model for our analysis, we denote by y the height with respect to Ordnance Survey datum of the base of a unit in one of the four sets listed in the previous paragraph. Let $y_{\mathrm{b}}(\mathbf{x}_{i,j})$ denote the observed height in the framework model at the jth member of the unique subset of validation boreholes for the ith framework model which is at location $\mathbf{x}$. Let $y_{\mathrm{m}}(\mathbf{x}_{i,j})$ denote the height of the same unit in the same validation borehole as represented in the ith framework model. On the assumption that $y_{\mathrm{b}}(\mathbf{x})$ is observed without error, the framework model error for that unit and location is defined as

$$z(\mathbf{x}_{i,j}) \;=\; y_{\mathrm{b}}(\mathbf{x}_{i,j}) - y_{\mathrm{m}}(\mathbf{x}_{i,j}).$$

For purposes of analysis we assume that $z(\mathbf{x}_{i,j})$ is a realization of the random variable, $Z(\mathbf{x}_{i,j})$, that it to say, it is one of a set of values to which the random variable could give rise. Furthermore, we assume that this random variable may be represented by a statistical model so that its values are given by a linear equation:

$$Z(\mathbf{x}_{i,j}) \;=\; \mu \;+\; \alpha_i \;+\; \eta(\mathbf{x}_{i,j}) \;+\; \varepsilon(\mathbf{x}_{i,j}). \tag{1}$$

The first term on the right hand side of Equation (1) is a fixed effect, the overall mean error in the framework model which is a constant. The remaining three terms are random effects (i.e. they are random variables each with a probability distribution); α_i is an effect for the ith model, the difference between the overall mean model error and the mean model error in the ith model. The α_i for different i have a mean of zero and an unknown variance, σ_α^2. The terms $\eta(\mathbf{x}_{i,j})$ and $\varepsilon(\mathbf{x}_{i,j})$ both represent random variables. Each is assumed to be normally distributed, with a mean of zero, and mutually independent (i.e. not correlated with each other). It is a common convention to denote the variance of ε by σ^2 and that of η by $\sigma^2\zeta$ where ζ is a scaling factor. It is assumed that $\eta(\mathbf{x}_{i,j})$ is a

second-order stationary autocorrelated random variable. This implies that the expected values (mean) and variance are constant and that the covariances Cov $\{\mathbf{x}_i, \mathbf{x}_i + \mathbf{h}\}$ depend only on the separation (called the lag) between two locations, and not on their absolute position. Hence, the correlation of the values of η at two locations, $\mathbf{x}_{i,j}$ and $\mathbf{x}_{k,l}$ is

$$\text{Corr}\left\{\eta(\mathbf{x}_{i,j}), \eta(\mathbf{x}_{k,l})\right\} \;=\; R\left(\mathbf{x}_{i,j} - \mathbf{x}_{k,l}|\boldsymbol{\theta}\right), \tag{2}$$

where $R\left(\cdot|\boldsymbol{\theta}\right)$ denotes some authorised spatial correlation function with parameters in $\boldsymbol{\theta}$. Thus, by including the term η in Equation (1), we allow for the possibility of spatial dependence in the variation of framework model error, that is to say, it is implicit in the model that the error at two neighbouring validation sites is, in general, more similar than the error at two sites which are further apart.

The random variable $\varepsilon(\mathbf{x}_{i,j})$ is a residual term which is assumed to be independently and identically distributed (i.i.d) at all possible locations.

Note that this model assumes that the expected framework model error is the same everywhere (the overall mean, μ) and that the variability of the model error is also the same everywhere. For this reason the model is said to be stationary in the mean and the variance. In the section 2.5.3 we consider models in which this assumption is relaxed.

If the validation boreholes had been located according to an independent and random design-based sampling scheme, then, conditional on the design, the framework model errors could be regarded as independent. In fact, the distribution of boreholes depends on past *ad hoc* decisions rather than a single sampling scheme, so it is necessary to use a model-based approach to analysis in which the random properties of the data are not based on a randomized sampling scheme but rather on a proposed statistical model for a random variable, such as Equation (1) above (de Gruijter et al., 2006).

A model of the form Equation (1) may be fitted to data by the method of residual maximum likelihood (REML) (Patterson and Thompson, 1971) to obtain estimates of the random effects parameters. In the case of the model in Equation (1) these parameters are σ^2_α, σ^2, ζ and the set of autocorrelation parameters in $\boldsymbol{\theta}$. To estimate these parameters

by maximum likelihood (ML) one finds parameter values such that the probability of observing the data values that we have is maximized. REML is a refinement of ML, and the REML estimates maximize a residual likelihood function which is not dependent on the unknown values of the fixed effects. Once the random effects parameters have been estimated, the fixed effects (just the mean μ in this case) are estimated by weighted least squares.

One may make inferences about data by comparison of alternative linear mixed models. Consider, for example, a simplified version of Equation (1), in which the modeller effects, α_i, are ignored

$$Z(\mathbf{x}_{i,j}) \;=\; \mu \;+\; \eta(\mathbf{x}_{i,j}) \;+\; \varepsilon(\mathbf{x}_{i,j}). \tag{3}$$

One can think of these two models as nested, because Equation (3) is a simplified version of Equation (1) in which, in effect, $\alpha_i = 0$ for all i. This model has one parameter fewer than does Equation (1) because σ_α^2 is not estimated. The fit of two nested models, with the same fixed effects, can be compared by computing the log-ratio of their maximized residual likelihoods, the usual statistic is

$$L \;=\; 2\left\{\ell_{\mathcal{R}}\left(\boldsymbol{\psi}_{\text{complete}}|\mathbf{Z}\right) - \ell_{\mathcal{R}}\left(\boldsymbol{\psi}_{\text{null}}|\mathbf{Z}\right)\right\}, \tag{4}$$

where $\ell_{\mathcal{R}}\left(\boldsymbol{\psi}_{\text{complete}}|\mathbf{Z}\right)$ and $\ell_{\mathcal{R}}\left(\boldsymbol{\psi}_{\text{null}}|\mathbf{Z}\right)$ denote, respectively, the maximum residual log-likelihoods (the natural logarithm of the maximum residual likelihood) obtained in fitting a model with the full set of parameters, $\boldsymbol{\psi}_{\text{complete}}$ and the so-called null model with a reduced parameter set $\boldsymbol{\psi}_{\text{null}}$.

With nested models L is always positive (i.e. the likelihood for the complete model is always larger). The inference as to whether the terms missing from the null model contain real information about the variable Z is based on the asymptotic distribution of L under the null model, which is a chi-squared distribution with degrees of freedom (shape parameter) equal to the number of extra parameters which are in the complete model but not in the null model (Verbeke and Mohlenbergs, 2000). This is the correct distribution

provided the model comparison meets certain conditions (Cox and Hinkley, 1990; Stram and Lee, 1994). These conditions hold for all the comparisons that we make in this study except for comparisons between a model which has the spatially autocorrelated random variable η and a null model with only an i.i.d. residual (Lark, 2012). In these cases the distribution of L under the null model was found by Monte Carlo simulation, as described by Lark (2012).

Analyses. The framework model errors for the four groups of bases: the base of the surface unit; the base of the first sub-surface unit; the base of the Kesgrave Sand and Gravel; and the base of the Red Crag, were analysed as follows:

Exploratory statistics were calculated, a histogram of the errors was plotted, and a scatter plot of the modelled height of the base and the observed height at each validation borehole was prepared.

Mixed models for framework model error were fitted, based on Equation (1), in which the only fixed effect was an overall mean model error. Likelihood ratio tests were conducted to decide whether the modeller effects (α_i) and spatially-dependent random effects (η) were required, and so to select an appropriate random effects statistical model for the framework model errors. Inferences about η were based on sample distributions for L under the null model obtained by simulation, as described by Lark (2012). In all cases two models were fitted. In the first the correlation function for η was assumed to be a spherical function (Webster and Oliver, 2007). In the second it was assumed to be exponential (Webster and Oliver, 2007). Details of these functions are given in the Appendix. Since the two models have the same number of parameters the best-fitting model can be identified as the one with the largest residual likelihood. Models were fitted using the NLME package (Pinheiro et al., 2012) for R, a widely used freeware statistical platform (R Development Core Team, 2012).

2.5.3 Error in the height of surfaces: models for possible sources of bias and non-stationarity in the variance of model error

The analyses described above assumed that the expected framework model error is the same everywhere, a constant mean. It is possible, however, that the expected model error might vary systematically. For example, if the modeller makes incorrect assumptions about the dip of a surface then the expected model error might increase with depth. The analyses also assume that the random component of framework model error has a uniform (stationary) variance. Again, this might not be true. For example there may be greater uncertainty in the framework model at locations with a sparse set of neighbouring boreholes, and the variance of the random effects in the model should be larger. These effects were accounted for by extensions of the linear mixed model presented in Equation (1), and this is now described.

Because these further models are more complex, it was decided to combine all available comparisons between a modelled and an observed unit base into a single data set to provide a larger set of observations. This means that more than one comparison between a modelled and observed unit height may be considered at any one validation borehole, and so spatial dependence in the model error cannot be accounted for in terms of covariance functions, as in Equation (2). The model errors must be treated as spatially independent. Since no evidence for spatial dependence in the model error was found in the analyses of the four initial sets of bases, this assumption was thought to be reasonable. However, it was necessary to include a borehole effect in the linear mixed model, to allow for the fact that framework model error for different units observed at the same validation borehole are likely to be correlated. The basic model was therefore

$$Z(\mathbf{x}_{i,j,k}) \;=\; f\{s(\mathbf{x}_{i,j,k})\} \;+\; \alpha_i \;+\; \xi_{i,j} \;+\; \varepsilon(\mathbf{x}_{i,j,k}), \tag{5}$$

where $Z(\mathbf{x}_{i,j,k})$ is the framework model error of the height of the kth unit observed at the borehole at location $\mathbf{x}$. The expression $f\{s(\mathbf{x}_{i,j,k})\}$ is a linear function of some covariate(s) known at the validation boreholes, and is a fixed effect, giving the expected model error at any location. In the simplest case it is a constant, the mean. As before, the term α_i is a random effect for the ith framework model, the difference between the overall mean

framework model error and the mean error of the ith framework model. The term $\xi_{i,j}$ is a random effect which represents the difference between the mean framework model error in the ith model and the mean framework model error for the jth validation borehole in the unique subset for the ith model. As in previous models ε is a residual random term.

2.5.3.1 Bias. A scatter plot of the framework model error against the depth of the framework modelled base below the DTM suggested that there might be some depth-dependent bias in the framework model with the modelled base tending to be too shallow at greater depths. One way to investigate this would be to include a linear function of the depth of the unit base in the framework model below the DTM as the term $f(s(\mathbf{x}_{i,j,k})$ in Equation (5) effectively as a regression predictor.

However, because the model error is the difference between the depth of the observed and modelled base, a relationship between this variable and the depth of the modelled base would arise through random variation as a result of the phenomenon of regression to the mean (Stigler, 1997), which means that for any random variables X_1 and X_2, the difference, $X_1 - X_2$ is necessarily correlated with both X_1 and X_2. One way to investigate whether there is a correlation between X_1 and X_2 and $X_1 - X_2$ in addition to the effect of regression to the mean is to examine the relationship between $X_1 - X_2$ and the average $(X_1 + X_2)/2$ (Oldham, 1962), which is robust provided that the observation error variances are the same for X_1 and X_2.

Let $t(\mathbf{x})$ denotes the height with respect to Ordnance Survey datum of the DTM at location $\mathbf{x}$. The depth of the kth unit in the framework model at location $\mathbf{x}$ and the depth of the same unit in the validation borehole at that location are, respectively, $t(\mathbf{x}) - y_{\mathrm{m}}(\mathbf{x}_{i,j,k})$ and $t(\mathbf{x}) - y_{\mathrm{b}}(\mathbf{x}_{i,j,k})$ and the average of these depths is

$$d(\mathbf{x}_{i,j,k}) \;=\; t(\mathbf{x}) - \frac{y_{\mathrm{m}}(\mathbf{x}_{i,j,k})}{2} - \frac{y_{\mathrm{b}}(\mathbf{x}_{i,j,k})}{2}.$$

To investigate the evidence for a depth-dependent bias in the framework model we there-

fore fitted a model of the form

$$Z(\mathbf{x}_{i,j}) \;=\; \beta_0 \;+\; \beta_1 d(\mathbf{x}_{i,j,k}) + \;\xi_{i,j} \;+\; \alpha_i \;+\; \varepsilon(\mathbf{x}_{i,j}), \tag{6}$$

As before likelihood ratio tests were conducted to decide whether the random effects for borehole and modeller were required.

2.5.3.2 Variance model The model in Equation (6) allows us to assess whether there is an effect of depth on framework model error, relaxing the assumption that the expected error is a constant mean. The assumption that the variance of the model error is constant is still implicit, however, because the residual term is i.i.d. and so we may write:

$$\mathrm{Var}\left\{\varepsilon(\mathbf{x}_{i,j})\right\} \;=\; \sigma^2,$$

which is a constant. This can be relaxed in the context of mixed models. Rather than simply estimating a fixed parameter, σ^2 we may estimate parameters of a function to calculate this variance for some particular $\mathbf{x}_{i,j}$, $\sigma^2(\mathbf{x}_{i,j})$ as a function of available covariates. This approach has been used generally in statistical modelling (Nelder and Lee, 1998) and in geostatistics (e.g. Lark, 2009). This function to compute σ^2, called a variance function, must be guaranteed to return a non-negative solution in all circumstances, we use a simple linear function which ensures this,

$$\sigma(\mathbf{x}_{i,j}) \;=\; \sigma_0 \;+\gamma s(\mathbf{x}_{i,j}), \tag{7}$$

where σ_0 and γ are parameters of the variance model and s is a continuous covariate, such as the depth of the unit base below the DTM in the framework model. The two parameters of the variance model are estimated together with the other random effects parameters of the linear mixed model by REML. Note that a mixed model in which a variance model such as Equation (7) was substituted for a fixed variance, σ^2, would have one more parameter than the simpler model. The two models can be compared by the log-likelihood ratio statistic, L, assumed to be asymptotically distributed as a chi-squared variable with 1 degree of freedom. In this study variance models were considered in which

the continuous covariate was (i) the average model and borehole depth of the unit, $d(\mathbf{x}_{i,j,k})$ and (ii) the framework model depth of the unit $y_{\mathrm{m}}(\mathbf{x}_{i,j,k})$ (which was included after the first model was shown to be significant, since in practice we could not compute a framework model error variance from $d(\mathbf{x}_{i,j,k})$ except at validation sites). In addition two further variance models were fitted in which the covariates were the number of boreholes, used by the modeller to form sections, (iii) within 200 m of location $\mathbf{x}$, $\nu_{200}(\mathbf{x})$ and (iv) the number of such boreholes within 1 500 m of location $\mathbf{x}$, $\nu_{1\,500}(\mathbf{x})$. Finally a model was fitted in which the covariate was (v) the shortest distance from the validation borehole to the outcrop of the base of the unit on the 2-D map, $c(\mathbf{x}_{i,j,k})$. However, since the modellers did not have information on units below the Thames Group, the base of this unit was excluded from this analysis.

We also considered the possibility that the variance of the framework model error differs between modellers. This is a categorical form of the variance model in which

$$\sigma(\mathbf{x}_{i,j}) \;=\; \sigma_i \;, \tag{8}$$

where σ_i denotes a constant value for all observations at validation boreholes in the unique subset for the ith modeller.

The code to fit these variance models was written in the FORTRAN programming language. In all cases the same code was used to fit models with stationary covariance structures to check that the results were the same as those obtained with the NLME package.

3 Results

3.1 Summary statistics.

Table 2 shows the cross-tabulation of presence and absence of the different units in those validation boreholes where a unit is either present or unambiguously absent. For all units the model and borehole are in agreement in most cases, the greatest disparity being for Kesgrave Sand and Gravel where there are a total of 11 cases where the model either

includes the unit where it is not present or includes it where it is absent.

Figure 3 shows scatter plots of framework-modelled height of unit bases against the observed heights, for the different subsets of units defined. In all cases the scatter plots are clustered around the bisector, where the models and observations are in agreement. Summary statistics for the model errors are presented in Table 3, and histograms are in Figure 4. In all cases the distribution of errors appears more or less symmetrical about a mean close to zero, so the assumption that they are drawn from a normal random variable is reasonable. Note that the mean error furthest from zero is for the surface unit (Figs. 3a, 4a). In this case the mean error is negative, indicating that the base of the surface unit tends, on average, to be modelled slightly higher than the borehole observations. Whether this is a significant effect or a random fluctuation is tested later.

3.2 Model results.

3.2.1. Separate groups of bases Table 4 presents the log-likelihoods for sets of models for the framework model error for bases of different groups of units. The models for any given unit differ with respect to their random effects. The first model, denoted A1 in the case of the surface unit, B1 for the second unit etc., has a random effect for modeller, and also a spatially correlated random component, η as in Equation (1). The second model (A2 etc.) has a modeller effect, but no spatially correlated random effect. These models can be compared by the log-likelihood ratio, L, which is compared to thresholds which correspond to a P-values of 0.05, 0.01 and 0.001 obtained by simulation using the method of Lark (2012). A P-value is the probability of observing, through random variation, a value of L as large or larger than the one actually observed if the true model for the data were the second model (A2 etc.) with no spatially correlated random effect (see, for example, Webster and Oliver, 1990). In the third model in each case the modeller effect is excluded, but a spatially correlated random effect is included. In the fourth model the only random effect is the i.i.d. term ε. The fourth and third models are compared by computing the log-likelihood ratio. In all cases it was found that there was no significant

evidence for a spatially correlated term, η (the threshold value of L for 0.05 is 3.51, and the largest value of L for a comparison between models with and without a correlated term was 1.90, most were rather smaller. We then compared the models with a random effect for modeller, but no spatially correlated component (the second model) and no random effect for modeller, or spatially correlated component (the fourth model). In most cases the P-value for this comparison was large (in the case of the first sub-surface unit the likelihoods for the two models were the same to 4 significant figures). This suggests that the evidence for a modeller effect is very weak, evidence of equal strength or stronger, would occur with rather large probability in cases where the model with no modeller effect is the correct one. The strongest evidence for a modeller effect was in the case of the Kesgrave Sand and Gravel (P=0.07), but this P-value is still rather larger than is conventionally thought necessary to judge an effect significant.

The variance components attributable to modeller effects, as well as being statistically insignificant, are also small. In the case of the surface unit the residual variance (the variance of the i.i.d component ε, in model A2, was 7.96, whereas the modeller variance component was 0.28. In the case of the first sub-surface unit, the residual variance was 10.17 and the modeller variance component was <0.001. In the case of the Kesgrave Sand and Gravel the residual variance was 9.12 and the modeller variance component was 1.93. This is larger than for the other units, but still not statistically significant. In the case of the Red Crag, the residual variance was 8.14 and the modeller variance component was 0.36.

In all cases the fourth model, with the i.i.d. component the only random effect, appears to be most appropriate. Table 5 shows the residual variances for this model applied to each unit. The t-ratio for a test of the null hypothesis that the mean model error is zero is also shown. Note that in no case was there evidence that the mean error was significantly different from zero.

To summarize the results for the separate groups of bases:

- In no case was the modeller effect statistically significant, and, with the exception of the Kesgrave Sand and Gravel, the estimated variance component for modeller was very small.

- In no case was there evidence of spatial correlation in framework model error.

- In no case was their evidence that the mean framework model error was significantly different from zero, i.e. the framework models appear to be unbiased.

- Out of 100 randomly selected test locations we would expect the framework model and the true base to be within a confidence interval ranging from ±5.6 m to ±6.4 m at 95 sites.

3.2.2. Combined set of bases In Table 6 we consider the models for the combined data set on framework model error for all bases that appear at the validation boreholes and in the corresponding framework model. Recall that the random effects for these models do not include a spatially-correlated component, but a component is included for borehole effects, to allow for the possibility that framework model errors for different units at the same validation borehole are correlated. In the first instance, as described above, we consider a model in which the average of the framework model depth and borehole depth of the base of the unit below the DTM is treated as a fixed effect in the simple linear model presented in Equation (6). This is to explore the possibility that there is a bias in the framework model associated with the depth of the unit. Figure 5 shows a plot of framework model error against average of model and borehole depth for all framework model and observed bases. The first part of Table 6 shows the different random effects models that were considered. The comparison of model D2 with model D1 shows that there is no evidence for a modeller effect, as with the models reported in Table 4. However, the comparison of model D3 and model D2 shows that there is evidence for a borehole effect. The residual variance for model D2 is 6.15 and the borehole variance component is 2.93. The overall variance about the expected framework model error, given the fixed effects, is 9.08. Table

5 shows the fixed effects parameter estimates for model D2. Note that the null hypothesis that the parameter β is zero can be rejected, with a very small P-value. This shows that there is a significant effect of depth on model error: for shallower bases the model error is generally smaller; at greater depth there is a tendency for the model error to become large (positive), suggesting that the framework model tends to underestimate the height of deeper bases. The fixed effect model is drawn on the plot in Figure 5.

Model D4 is reported in Table 6. In this model, the residual variance is modelled as a function of a covariate, using the expression in Equation (7) above, with the average of the framework model depth and the borehole depth below the DTM as the covariate. This model was compared to model D2, in which the residual variance is a constant, by the log-likelihood ratio test. The table shows that model D2 can be rejected in favour of D4 on this basis. The estimated coefficients of the variance model σ_0 and γ are 2.06 and 0.0953 respectively. This shows that the variance of the framework model error appears to increase with depth below the DTM.

A further model, model D5, was fitted with the same random effects as model D2 but with the mean framework model error the only fixed effect (see Table 7). In this case the borehole variance component was 1.85 and the residual variance was 7.65, giving an overall variance of 9.5. The estimates mean error was -0.55 m with a standard error of 0.281 m, this is not significantly different from zero (P >0.05). By comparing the variances for models D5 and D2, we may compute an approximate adjusted R^2 value for D2. The adjusted R^2 is a measure of the goodness of fit of the linear model (Webster and Oliver, 1990). Specifically it is the proportion of the variance in framework model error that the model (here in terms of average depth) explains. This is $(9.5 - 9.08)/9.5 = 0.04$, which is very small. So, although there is a significant effect of depth on framework model error, the effect is very small by comparison to the variation in framework model error.

For this reason we also considered models for the framework model error for the data combined for all units in which the mean was the only fixed effect. Model D5 is one

of these. Some others are reported in Table 7. These were fitted to allow us to test (i) the significance of modeller effect (D6), and then to consider a set of variance models.

Model D6, with a random effect for modeller, was not significantly better than model D5 with just an i.i.d. residual term. In models D7, D9, D10 and D12, the residual variance is modelled as a function of continuous covariates, respectively: the framework model depth of the unit base below the DTM, the number of boreholes used for modelling within 200 m of the validation borehole (ν_{200}), the number of such boreholes within 1 500 m of the validation borehole ($\nu_{1\,500}$ and the shortest distance from the validation borehole to the outcrop of the base of the unit (excluding Thames Group). In model D8 the variance model took the form of Equation (8) with a separate residual variance for each modeller. Table 7 shows that there was no evidence that variance models based on the modeller, on the number of neighbouring boreholes or on the distance to outcrop were preferable to model D5 with an i.i.d. residual. However, Model D5 could be rejected in favour of model D7, with an effect of framework model depth below the DTM with $P = 0.017$. The coefficients for this significant variance model were $\sigma_0 = 2.51$ and $\gamma = 0.0594$ so the variance of framework model error appears to increase with depth below the DTM.

To summarize the results for the combined set of bases:

- There was no significant modeller effect.
- There was evidence of a bias in the framework model with increasing depth below the DTM, but this effect is very small.
- The uncertainty of the framework model, as measured by the residual variance, appears to increase with depth below the DTM.
- The uncertainty of the framework model does not seem to vary with depth from the outcrop, the number of neighbouring boreholes or the modeller.

4 Discussion and conclusions

We discuss our results in terms of their relevance to the users of framework models and the producers of these models. We then consider the scope to develop the methodology used in this paper further and to apply it.

From the perspective of users of framework models the following findings are of interest:—

First, for none of the framework models, on groups of units or all units combined, was the mean error significantly different from zero, so there is no evidence of overall bias in the framework models. Although there was significant evidence for a depth-dependent bias, the effect was very small relative to other sources of uncertainty. This suggests that framework models, at least those produced in similar conditions to the one used in this study in terms of geology, the quality and density of data, the modelling methods used and the expertise of the modellers, can be regarded as reliable in terms of the average depth of units that they depict. There is uncertainty in predictions at individual locations, but no reason to believe that, in general, these predictions are subject to a systematic error.

Second, while the confidence intervals of $\pm$ around 5 m may seem large, it must be borne in mind that this uncertainty has various sources. There will be contributions from uncertainty in borehole heights (which contribute to both the framework model error and inflate our estimates of this error since our validation observations also have uncertainty from this same source).

Finally, there was no evidence for spatial dependence of framework model error, so at the scale of sampling the error looks like white noise (a sequence of i.i.d. serially uncorrelated random variables with zero mean and finite variance). If the framework model tended to represent a real surface with a model surface which is over-generalized, smoothing out variations in the height of the surface which could be resolved by observations collected at the density of our validation set, then we would expect to see spatial dependence in the framework model error. Our results therefore suggest that the generalization of the shape of the surface by the framework modellers was reasonable and smoothed out

only the short-range variations of the real surface.

The following findings are particularly relevant for framework modellers and organizations that undertake 3-D geological framework modelling:—

First, the overall consistency of framework models and observations (Table 2) is encouraging, as is the lack of overall bias and the agreement between the framework models and observations shown in Figure 3. The lack of spatial dependence in the framework model error suggests that this is unlikely to be reduced markedly by improved modelling methodology from borehole observations, but rather by improved information (such as increasing the density of boreholes or incorporating of higher-resolution geophysical information into different stages of the framework modelling process).

There was no strong evidence that differences between modellers was a significant source of variation in observed framework model error, nor that there were differences between modellers with respect to the residual variance term. The strongest evidence for a modeller effect was when we examined the errors in Kesgrave Sand and Gravel. This is a challenging unit to model (see Table 2), being the deepest Quaternary unit, mainly occurring as a subcrop, where modeller experience becomes critical. Overall, however, there is no suggestion of a pressing need to improve the consistency of modelling practice between different modellers. We note below, however, that different conclusions might be drawn in areas where the geology is more complex or data are sparser, or both.

It is acknowledged that the geological structure in the study area is very simple compared to many geological terrains. It is likely that model error would be larger where the geology is more complex, in particular the variation between framework modellers might be larger in more complex conditions, reflecting differences in experience and knowledge of the local geology. The study area is also rich in data with respect to surface geological linework and borehole observations. Uncertainty will be greater where data are sparser and, again, the effect of variation in the experience of framework modellers may also be larger. Further studies are therefore required to develop the methodology and to apply it

in a wider range of conditions. We summarize below some priorities for such work.

First, we have shown that this experimental approach allows us to quantify framework model error but, as noted above, our particular conclusions are unlikely to apply to all framework models. The method should be applied in contrasting landscapes, with more complex geology, and in more difficult conditions for modelling, for example, with sparser borehole data, to give indicative measures of uncertainty for these different conditions.

Second, there was a significant borehole effect in models for the combined data set, suggesting that errors in different units in a single borehole are correlated. This is not surprising since if one unit is too high in the framework model then associated units are likely to be too. It may also be that error in borehole height contributes both to error in the model and to errors in our assessment at validation boreholes. This must be borne in mind when considering the error variances reported. The magnitude and effect of borehole elevation error should be the subject of further study.

Third, there is evidence that the error variance, and hence the magnitude of model uncertainty increases with depth below the surface. We believe that this reflects increasing sparsity of borehole data. The fitted variance model would allow us to represent framework model uncertainty as a variable (increasing with depth) rather than a constant value.

Finally, there was no evidence that distance to boreholes accounted for any differences in the magnitude of variations in framework model error. This is counter to expert expectations (see Lelliott et al., 2009). It was also interesting that distance to the outcrop was not a significant predictor of the framework model error variance. At the limit it must be, because as the modelled base gets closer to the surface so the scope for variation in the error is curtailed. However, simply looking at the numbers of neighbouring boreholes, or shortest distance to the outcrop may be too crude. Borehole interpretation is done on sections, which are then interpolated to produce the final model. There is a need for further study to look at error along sections, and then how this error is propagated in the triangulation steps.

In conclusion, we have shown how a carefully designed framework modelling experiment, with held-back subsets of validation boreholes, can be used to study and quantify the errors in a framework model that are attributable to the modelling process. This methodology should be applied in a range of contrasting geological terrains to give an overall picture of the range of quality that can be expected in 3-D geological framework models and the dependence of this quality on geological conditions, modeller experience and the quality and quantity of data.

Acknowledgements

This paper is published with the permission of the Director of the British Geological Survey (NERC). This work was undertaken as an internal science project at BGS and the project team acknowledge the contribution of Jonathan Ford in developing the ideas on which it was based. We are grateful for invaluable discussions with our colleagues Holger Kessler, Don Aldiss, Mark Cave and Andy Kingdon. We also acknowledge helpful suggestions about the presentation of this paper made by a reviewer.

References

Blanchin, R., Chilès, J.P. 1993. The Channel Tunnel: geostatistical prediction of the geological conditions and its validation by the reality. Mathematical Geology, 25, 963–974.

British Geological Survey. 2001. Woodbridge and Felixtowe. England and Wales Sheets 208 and 225 Solid and Drift Geology. 1:50 000. British Geological Survey, Keyworth, Nottingham.

British Geological Survey. 2006. Ipswich. England and Wales Sheet 207 Solid and Drift Geology. 1:50 000. British Geological Survey, Keyworth, Nottingham.

Cox, D.R., Hinkley, D.V. 1990. Theoretical Statistics. Chapman & Hall, London.

de Gruijter, J., Brus, D., Bierkens, M.F.P., Knotters, M. 2006. Sampling for Natural Resource Monitoring. Springer, Heidelberg.

Delaunay, B. 1934. Sur la sphère vide. Izvestia Akademii Nauk SSSR, Otdelenie Matematicheskikh i Estestvennykh Nauk 7, 793–800.

Diggle, P.J., Ribeiro, P.J. 2007. Model-based geostatistics. Springer, New York.

Goovaerts, P. 1997. Geostatistics for natural resources evaluation. Oxford University Press, New York.

Journel, A.G., Huijbregts, C.J., 1978, Mining Geostatistics. Academic Press, London.

Kessler, H., Mathers, S.J. 2004. From geological maps to models - finally capturing the geologists' vision. Geoscientist, 14,(10), 4–6.

Kessler, H., Mathers, S.J., Sobisch, H.-G. 2009. The capture and dissemination of integrated 3D geospatial knowledge at the British Geological Survey using GSI3D software and methodology. Computers & Geosciences, 35, 1311–1321.

Lark, R.M. 2009. Kriging a soil variable with a simple non-stationary variance model. Journal of Agricultural Biological and Environmental Statistics, 14, 301–321.

Lark, R.M. 2012. Distinguishing spatially correlated random variation in soil from a 'pure nugget' process. *Geoderma* **185–186**, 102–109.

Lark, R.M., Webster, R. 2006. Geostatistical mapping of geomorphic variables in the presence of trend. Earth Surface Processes and Landforms 31, 862–874.

Lelliott, M.R., Cave, M.R., Wealthall, G.P. 2009. A structured approach to the measurement of uncertainty in 3D geological models. Quarterly Journal of Engineering Geology and Hydrogeology, 42, 95–105.

Matheron, G. 1963. Principles of geostatistics. Economic Geology 58, 1246–1266.

Mathers, S.J., Kessler, H. 2010. Shallow sub-surface 3D geological models for Earth & Environmental Science decision making. Environmental Earth Science, 60, 445–448.

Mathers, S.J., Smith, M.A. 2002. Geology of the Woodbridge and Felixtowe District — a brief explanation of the geological map. Sheet explanation of the British Geological Survey 1:50 000 Sheets 208 and 225 Woodbridge and Felixtowe (England and Wales). British Geological Survey, Keyworth, Nottingham.

Mathers, S.J., Wood, B., Kessler, H. 2011. GSI3D 2011 software manual and methodology. British Geological Survey Internal Report, OR/11/020. 152 pp.

Mathers, S.J., Woods, M.A., Smith, N.J.P. 2007. Geology of the Ipswich District — a brief explanation of the geological map. Sheet explanation of the British Geological Survey 1:50 000 Sheet 207 Ipswich (England and Wales). British Geological Survey, Keyworth, Nottingham.

Nelder, J.A., Lee, Y.G. 1998. Joint modelling of mean and dispersion. *Technometrics* **40**, 168–171.

Oldham, P.D. 1962. A note on the analysis of repeated measurements of the same subjects. Journal of Chronic Diseases 15, 969–977.

Patterson, H.D., Thompson, R. 1971. Recovery of inter-block information when block sizes are unequal. Biometrika, 58, 545–554.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & R Development Core Team. 2012. *nlme: Linear and Nonlinear Mixed Effects Models.* R package version 3.1-105.

R Development Core Team. 2012. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. `http://www.R-project.org/`.

Royse, K.R., Kessler, H., Robins, N.S., Hughes, A.G., Mathers, S.J. 2010. The use of 3D geological models in the development of the conceptual groundwater model. Zeitschrift der Deutschen Gesellschaft für Geowissenschaften, 161, 237–249.

Stein, M.L. 1999. Interpolation of spatial data: some theory for kriging. Springer, New York.

Stigler, S.M. 1997. Regression toward the mean, historically considered. Statistical Methods in Medical Research 6, 103–114.

Stram, D.O., Lee, J.W. 1994. Variance components testing in the longitudinal mixed effects setting. Biometrics 50, 1171–1177.

Verbeke, G., Molenberghs, G. 2000. Linear Mixed Models for Longitudinal Data. Springer, New York.

Webster, R. & Oliver, M.A. 1990. Statistical Methods in Soil and Land Resource Survey. Oxford University Press, Oxford.

Webster, R. & Oliver, M.A. 2007. Geostatistics for Environmental Scientists. 2nd Edition John Wiley & Sons, Chichester

Appendix: the spherical and exponential correlation functions

In Equation (2) we referred to the general spatial correlation function $R\left(\mathbf{x}_{i,j} - \mathbf{x}_{k,l}|\boldsymbol{\theta}\right)$ which represents the correlation between values of a random variable at two locations $\mathbf{x}_{i,j}$ and $\mathbf{x}_{k,l}$. In this study we considered functions which are isotropic, that is to say they are functions of the distance $h = |\mathbf{x}_{i,j} - \mathbf{x}_{k,l}|$ irrespective of the direction of the vector $\mathbf{x}_{i,j} - \mathbf{x}_{k,l}$. The spherical function has the following form:

$$\begin{aligned} R_{\text{sph}}(h|a) &= 1 - \frac{3h}{2a} + \frac{1}{2}\left(\frac{h}{a}\right)^3 \quad a > h \\ &= 0 \qquad a \leq h, \end{aligned}$$

where a is a distance parameter, the *range*, at which the correlation is equal to zero. By contrast, the correlation in the exponential model declines with increasing distance, but never goes exactly to zero. It is defined by the following function:

$$R_{\text{exp}}(h|r) = \exp\left(-h/r\right), \tag{9}$$

where r is a distance parameter. The correlation is small (< 0.05) for distances larger than $3 \times r$. Note that $R_{\text{sph}}(h|a)$ and $R_{\text{exp}}(h|r)$ each have a single parameter.

Table 1. Numbers of validation boreholes by formation with both a proven and a modelled base.

Formation	Lowestoft Till	Glacial Sand and Gravel	Kesgrave Sand and Gravel	Red Crag Formation	Thames Group
	7	17	50	57	11

Table 2. Cross-tabulation, by formation, showing the presence or absence of a formation in the model and validation borehole for each borehole where the formation is present or unambiguously absent. The agreement is the proportion of validation boreholes at which the model and the observations agree.

Unit	Borehole record	Model prediction		Agreement
		Absent	Present	
Lowestoft Till	Absent	90	1	0.99
	Present	0	7	
Glacial Sand and Gravel	Absent	76	2	0.95
	Present	3	17	
Kesgrave Sand and Gravel	Absent	28	6	0.89
	Present	5	58	
Red Crag Formation	Absent	6	2	0.96
	Present	2	79	
Thames Group	Absent	0	0	1.0
	Present	0	66	

Table 3. Summary statistics for model error for the height of the base of (i) the surface unit (ii) the first unit below the surface (iii) Kesgrave Sand and Gravel (iv) Red Crag (v) All modelled and observed bases.

Unit	Number of observations	Mean /m	Median /m	Standard deviation /m	Minimum /m	Maximum /m	Skewness
Surface	84	−1.13	−1.45	2.86	−11.72	6.82	−0.66
First below surface	43	0.33	0.89	3.19	−7.14	7.75	−0.17
Kesgrave Sand and Gravel	50	−0.50	−0.39	3.27	−9.71	6.95	−0.17
Red Crag	57	−0.41	−0.43	2.91	−11.72	7.75	−0.60
All combined	143	−0.55	−0.44	3.09	−11.72	8.61	−0.21

Table 4. Linear mixed models for framework model error for different groups of units. In all the mean is the only fixed effect with i.i.d. residual random variation. Random effects include combinations of modeller (ε) and correlated random variation (η).

Model	Unit	Random effects		Correlation	Residual	Comparison	$L^{\dagger}$	P
		Modeller	η	model for η^{*}	log-likelihood			
A1	Surface unit	✓	✓	Exp	−206.1799			
A2	Surface unit	✓	×	None	−207.0182	A1 vs A2	1.676	>0.05
A3	Surface unit	×	✓	Exp	−206.2752			
A4	Surface unit	×	×	None	−207.2346	A3 vs A4	1.919	>0.05
						A2 vs A4	0.433	0.51
B1	First sub-surface unit	✓	✓	Exp	−109.8175			
B2	First sub-surface unit	✓	×	None	−110.1993	B1 vs B2	0.764	>0.05
B3	First sub-surface unit	×	✓	Exp	−109.4862			
B4	First sub-surface unit	×	×	None	−110.1993	B3 vs B4	1.426	>0.05
						B2 vs B4	0.0	1.0
C1	Kesgrave Sand & Gravel	✓	✓	Exp	−127.3570			
C2	Kesgrave Sand & Gravel	✓	×	None	−127.9016	C1 vs C2	1.089	>0.05
C3	Kesgrave Sand & Gravel	×	✓	Exp	−129.2385			
C4	Kesgrave Sand & Gravel	×	×	None	−129.5374	C3 vs C4	0.598	>0.05
						C2 vs C4	3.27	0.07
D1	Red Crag	✓	✓	Exp	−140.9460			
D2	Red Crag	✓	×	None	−140.9887	D1 vs D2	0.085	>0.05
D3	Red Crag	×	✓	Sph	−141.2024			
D4	Red Crag	×	×	None	−141.2024	D3 vs D4	0.0142	>0.05
						D2 vs D4	0.427	0.51

*Selected model from exponential (Exp) or spherical (Sph). $^{\dagger}L$ has 1 degree of freedom for all comparisons in this table.

Table 5. Residuals variances, t statistic and P-value for the null hypothesis that the mean model error is zero, and width of the 95% confidence interval for the modelled depth of the base of the unit for each group of units, based on the fourth model (an i.i.d. random variable is the only random effect).

Unit	Residual variance	t-ratio*	P-value*	95% confidence interval
Surface unit	8.19	−1.27	0.21	±5.6 m
First sub-surface unit	10.17	0.237	0.81	±6.3 m
Kesgrave Sand & Gravel	10.7	−0.38	0.71	± 6.4 m
Red Crag	8.44	−0.37	0.71	± 5.7 m

* Null hypothesis is that the mean model error is zero.

Table 6. Linear mixed models for framework model error from all modelled and observed bases with average of framework model and borehole-derived depth of the base below the DTM, $d(\mathbf{x}_{i,j,k})$, as the fixed effect. No spatially correlated random effect is included. Each model has some combination of modeller and borehole in its random effects. Each model has a residual term, ε, which is either i.i.d. or has a variance given by a variance model which is a linear function of the average depth.

Model	Random effects†			Residual	Comparison	$L^\dagger$	P
	Modeller	Borehole	Variance model	log-likelihood			
D1	✓	✓	×	−229.05			
D2	×	✓	×	−229.07	D2 vs D1	0.04	0.84
D3	×	×	×	−233.74	D3 vs D2	9.34	0.002
D4	×	✓	$\sigma = \sigma_0 + \gamma d(\mathbf{x}_{i,j,k})$	−225.00	D4 vs D2	8.14	0.004

†L has 1 degree of freedom for all comparisons in this table.

Fixed effects parameter estimates for model D2: framework model error = $\beta_0 + \beta$ Average depth.

Parameter	Estimate	Standard error	t-ratio‡	P-value‡
β_0	−0.481	0.287		
β	0.135	0.036	3.75	0.0002

‡The null hypothesis is that the parameter equals zero.

Table 7. Linear mixed models for framework model error from all modelled and observed bases with mean framework model error as the only fixed effect. No spatially correlated random effect is included. Each model has borehole as a random effect and a residual term, ε, which is either i.i.d. or has a variance given by a variance model defined in the text. Note that model D11 is the equivalent of D5, but fitted to a subset of data from which the framework model errors in the base of the Thames group are excluded. This is for comparison with Model D12.

Model	Random effects†		Residual log-likelihood	Comparison	L	Degrees of freedom	P
	Modeller	Variance model					
D5	×	×	−231.91				
D6	✓	×	−231.63	D6 vs D5	0.56	1	0.45
D7	×	$\sigma = \sigma_0 + \gamma_1$framework model depth	−229.10	D7 vs D5	5.62	1	0.017
D8	×	$\sigma = \sigma_i$	−230.20	D8 vs D5	3.42	4	0.49
D9	×	$\sigma = \sigma_0 + \gamma_2 \nu_{200}$	−231.40	D9 vs D5	1.0	1	0.32
D10	×	$\sigma = \sigma_0 + \gamma_3 \nu_{1\,500}$	−230.72	D10 vs D5	2.4	1	0.12
D11	×	×	−210.19				
D12	×	$\sigma = \sigma_0 + \gamma_3$distance to outcrop	−209.84	D12 vs D11	0.7	1	0.17

Figure Captions.

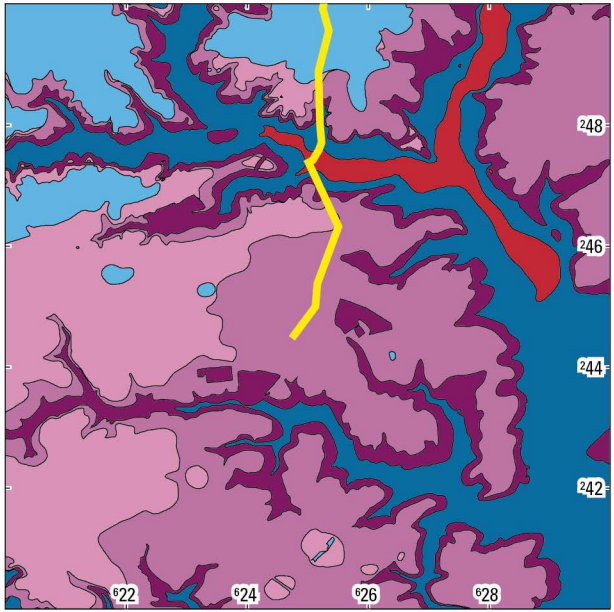
1. A 2-D map of the modelled units in the study area with an example cross-section.

2. Distribution of boreholes across the region, the encircled boreholes are the validation set for Modeller 1. The line of the cross-section in Figure 1 is also shown.

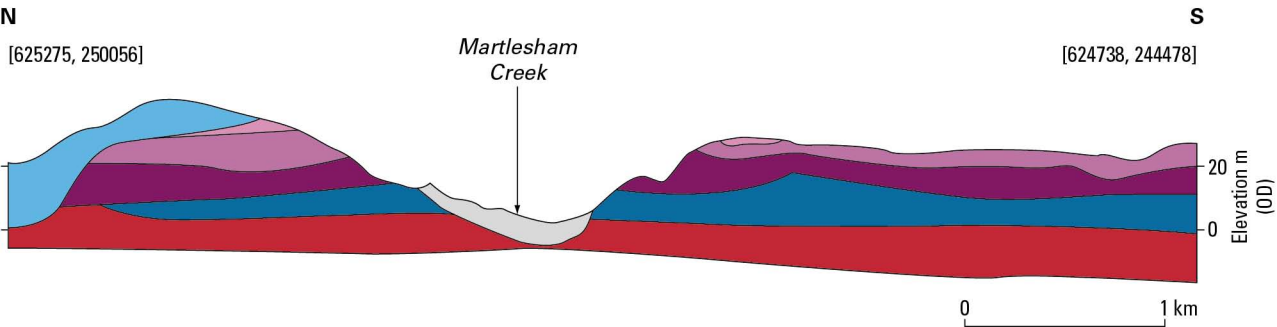
3. Scatter plot of observed height of unit base with respect to Ordnance datum against framework model height for validation boreholes. (a) surface unit (b) first unit below surface (c) Kesgrave Sand and Gravel (d) Red Crag (e) All framework model and observed bases. The line is the bisector where framework model height equals observed height.

4. Histogram of framework model error (observed height − model height) for validation boreholes. (a) surface unit (b) first unit below surface (c) Kesgrave Sand and Gravel (d) Red Crag (e) All framework model and observed bases.

5. Plot of framework model error against average of model and borehole depth below the DTM for all framework model and observed bases. The fitted line is from model D2.



Devensian-Holocene sediments
Lowestoft Till
Glacial Sand and Gravel
Kesgrave Sand and Gravel
Red Crag Formation
Thames Group
Lambeth Group



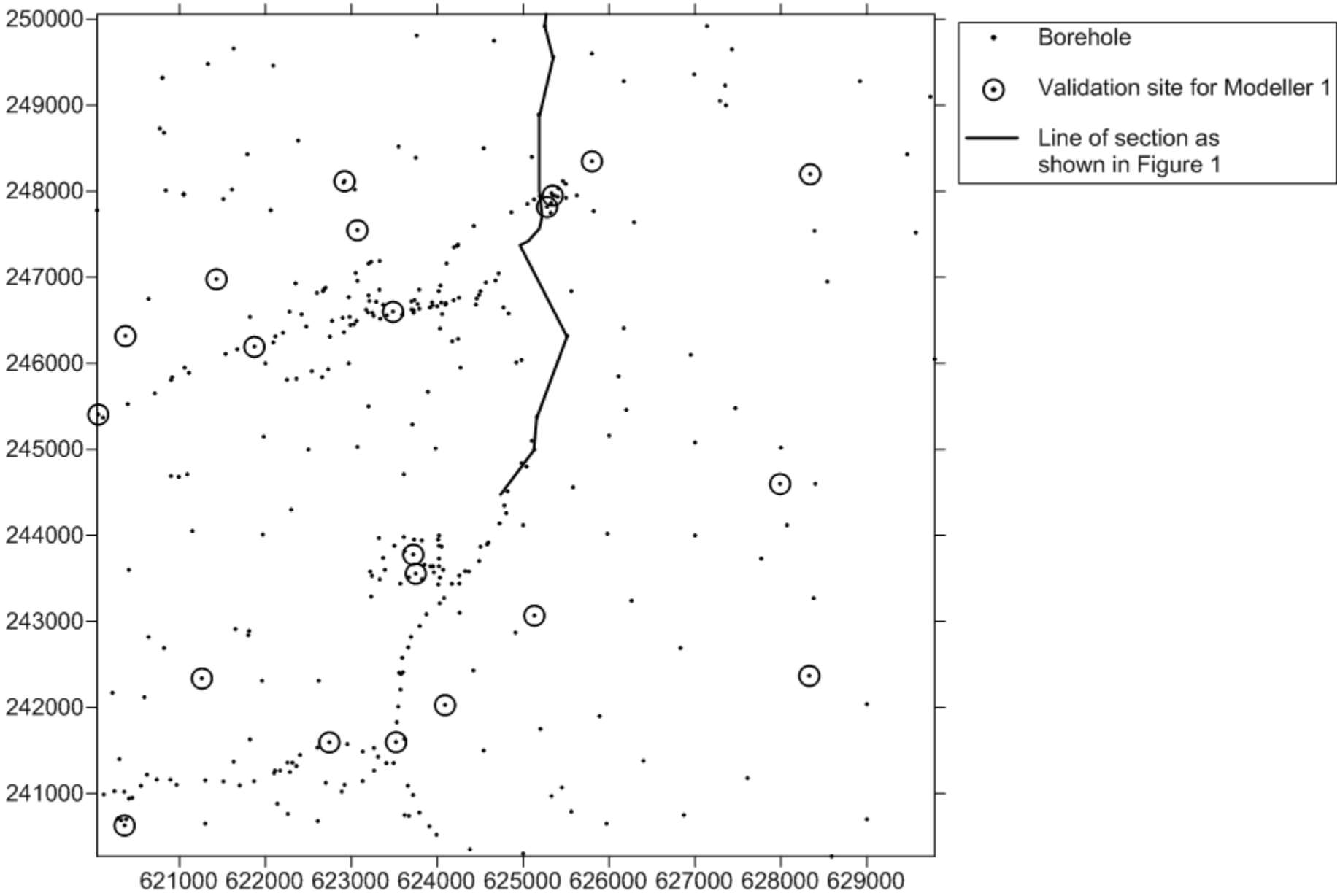
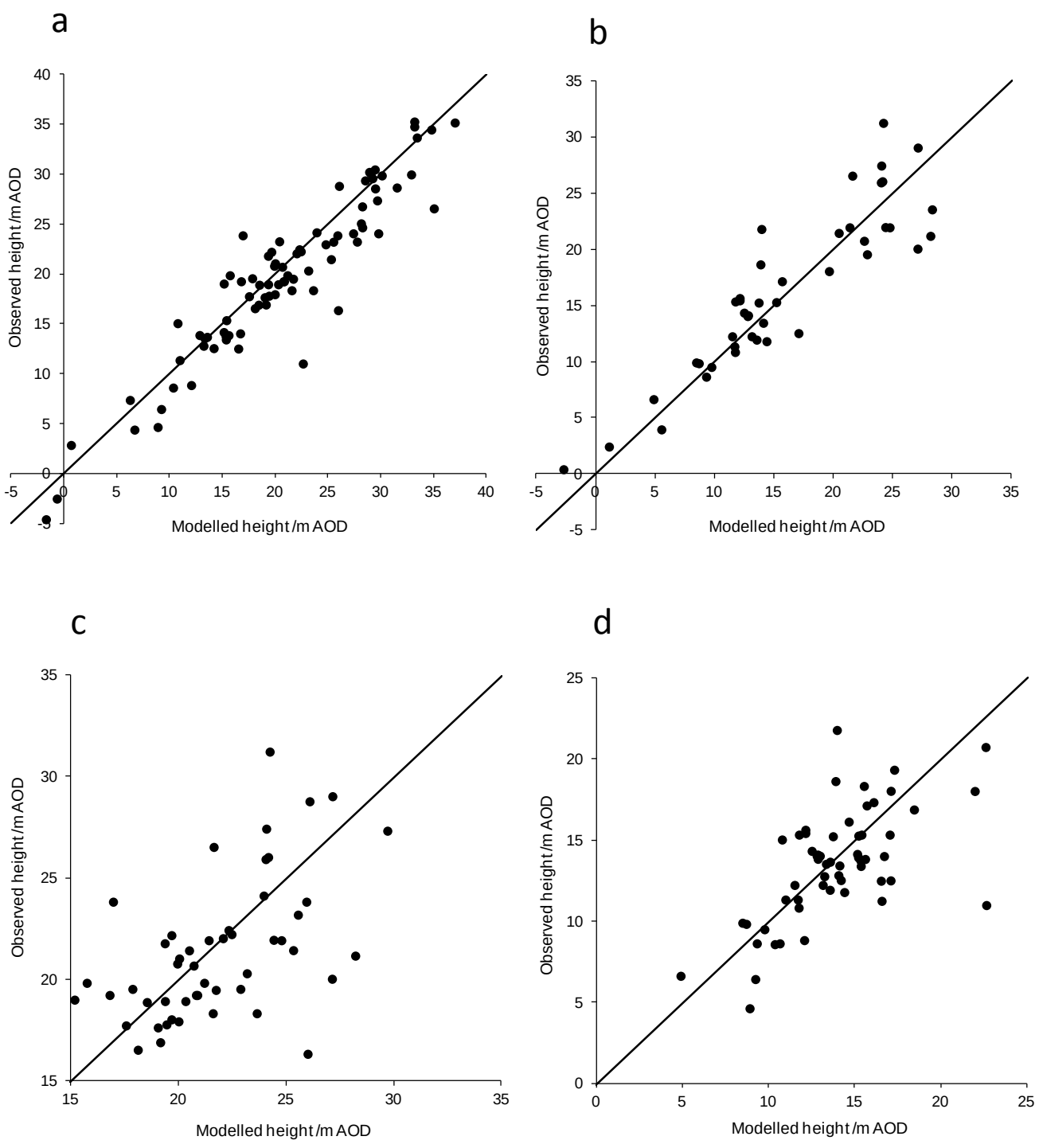

250000
249000
248000
247000
246000
245000
244000
243000
242000
241000
621000 622000 623000 624000 625000 626000 627000 628000 629000
Borehole
Validation site for Modeller 1
Line of section as shown in Figure 1

Figure 3


a
Observed height /m AOD
Modelled height /m AOD
b
Observed height /m AOD
Modelled height /m AOD
c
Observed height /m AOD
Modelled height /m AOD
d
Observed height /m AOD
Modelled height /m AOD

e

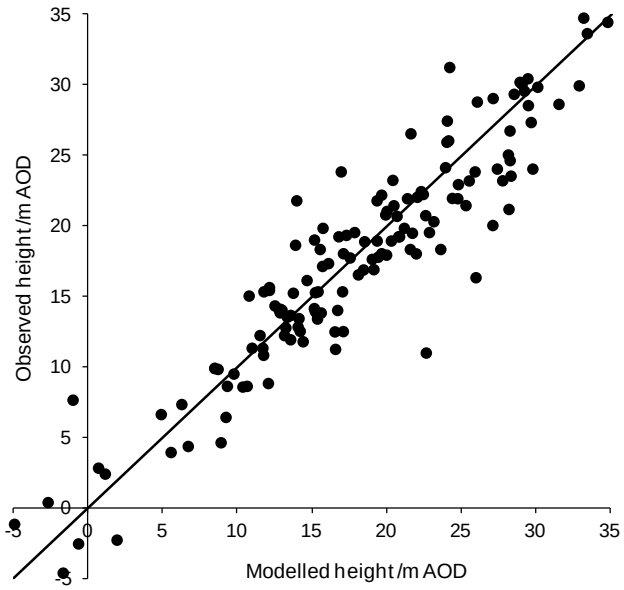
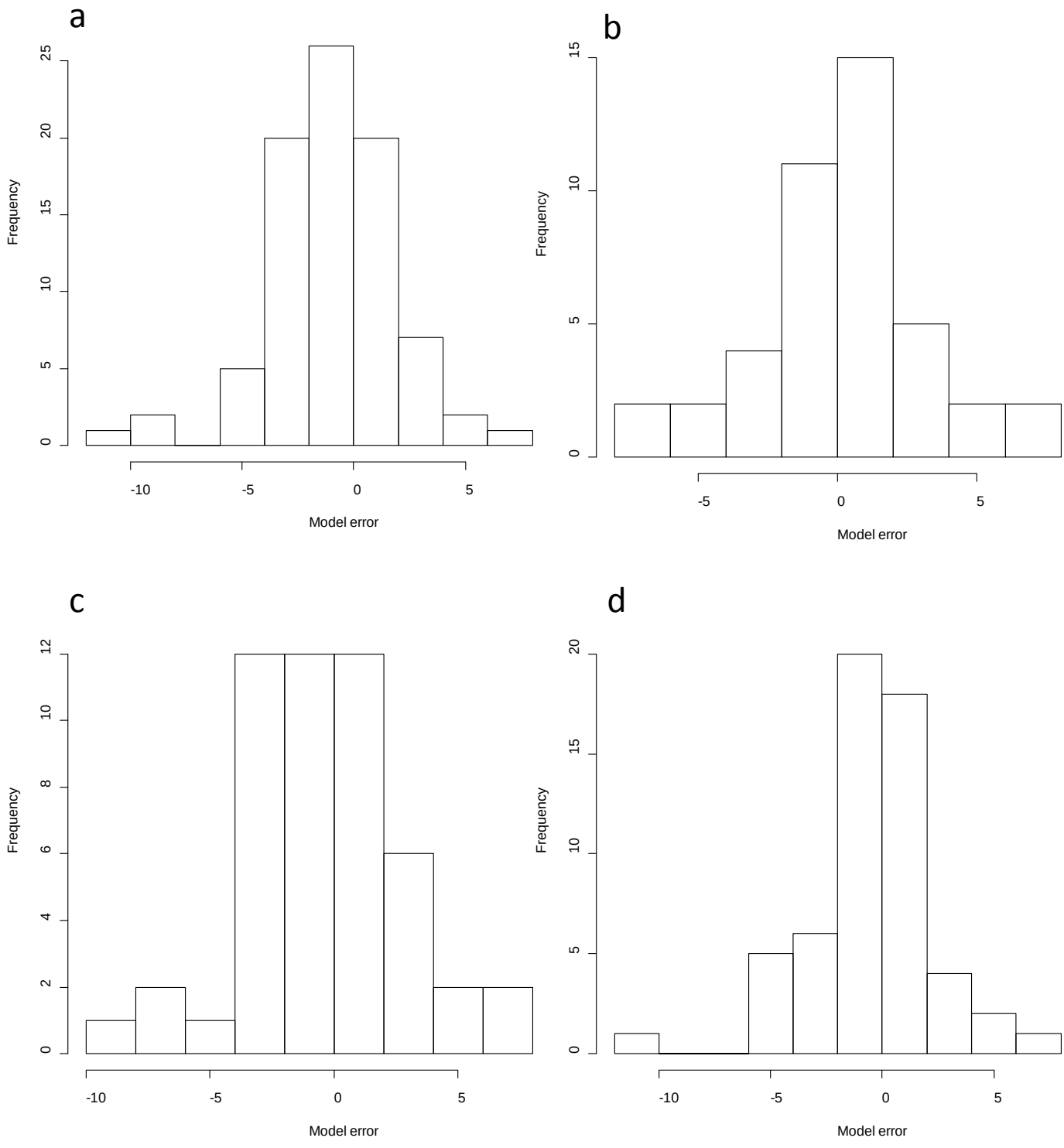

35
30
25
20
15
10
5
0
-5
-5
0
5
10
15
20
25
30
35
Observed height /m AOD
Modelled height /m AOD

Figure 4


a
Frequency
0
5
10
15
20
25
-10
-5
0
5
Model error
b
Frequency
0
5
10
15
-5
0
5
Model error
c
Frequency
0
2
4
6
8
10
12
-10
-5
0
5
Model error
d
Frequency
0
5
10
15
20
-10
-5
0
5
Model error

e

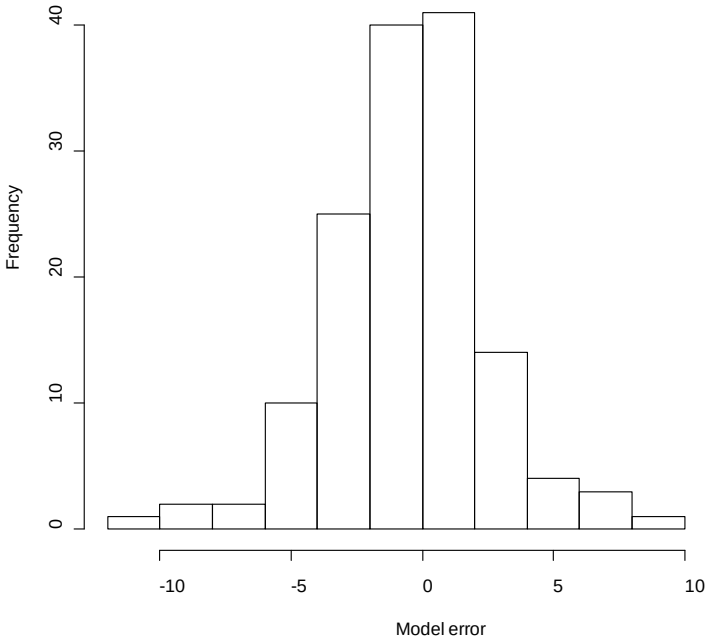
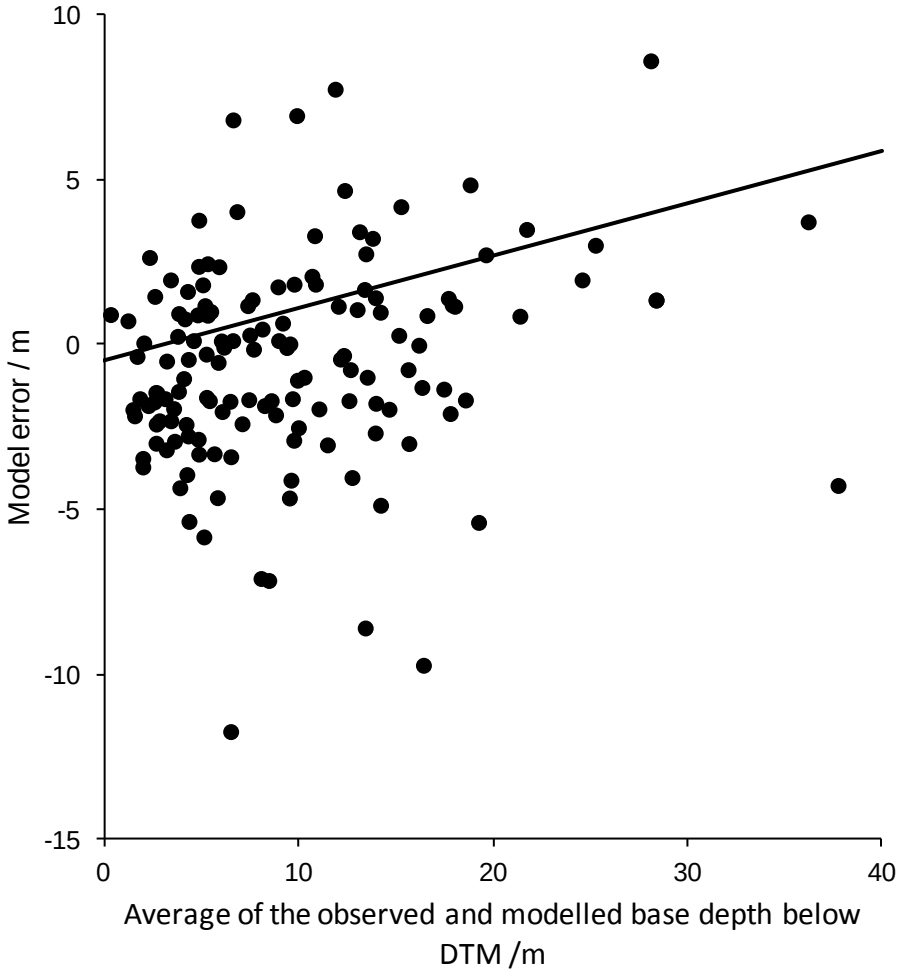


Figure 5


10
5
0
-5
-10
-15
0
10
20
30
40
Model error / m
Average of the observed and modelled base depth below DTM /m