



# Uncertainty in mapped geological boundaries held by a national geological survey: eliciting the geologists' tacit error model

R. M. Lark, R. S. Lawley, A. J. M. Barron, D. T. Aldiss, K. Ambrose, A. H. Cooper, J. R. Lee, and C. N. Waters

British Geological Survey, Keyworth, Nottingham, UK

*Correspondence to:* R. M. Lark (mlark@bgs.ac.uk)

Received: 22 December 2014 – Published in Solid Earth Discuss.: 22 January 2015

Revised: 26 May 2015 – Accepted: 31 May 2015 – Published: 17 June 2015

**Abstract.** It is generally accepted that geological line work, such as mapped boundaries, are uncertain for various reasons. It is difficult to quantify this uncertainty directly, because the investigation of error in a boundary at a single location may be costly and time consuming, and many such observations are needed to estimate an uncertainty model with confidence. However, it is recognized across many disciplines that experts generally have a tacit model of the uncertainty of information that they produce (interpretations, diagnoses, etc.) and formal methods exist to extract this model in usable form by elicitation. In this paper we report a trial in which uncertainty models for geological boundaries mapped by geologists of the British Geological Survey (BGS) in six geological scenarios were elicited from a group of five experienced BGS geologists. In five cases a consensus distribution was obtained, which reflected both the initial individually elicited distribution and a structured process of group discussion in which individuals revised their opinions. In a sixth case a consensus was not reached. This concerned a boundary between superficial deposits where the geometry of the contact is hard to visualize. The trial showed that the geologists' tacit model of uncertainty in mapped boundaries reflects factors in addition to the cartographic error usually treated by buffering line work or in written guidance on its application. It suggests that further application of elicitation, to scenarios at an appropriate level of generalization, could be useful to provide working error models for the application and interpretation of line work.

## 1 Introduction

### 1.1 What geological boundaries are, and why they are uncertain

The geological map, with boundaries delineating the surface expression of different stratigraphic or lithological units, is the classical form of spatial geological information. These boundaries are drawn by a geologist on the basis of field observations and interpretation of borehole records, remote sensor data and other information. The boundaries delineated by the geologist eventually are presented as boundaries on the published map, be this a paper or a digital product, and may also appear on the basis of subsequent interpretation, as boundaries in other derived maps: susceptibility maps for geohazards, for example, or maps of mineral resources or soil parent material. Recent developments in computer-based geological modelling make it easier for the geologist to represent their three-dimensional (3-D) understanding of geology, but mapped geological boundaries in two dimensions remain an important source of information in the era of 3-D modelling. Boundaries in two dimensions represent important information, e.g. on the position of outcrop lines, which assist and constrain the 3-D interpretation. Mapped geological boundaries, particularly those held in the records of large national geological surveys, remain an important source of geological information. For this reason it is important to understand and to quantify their inherent uncertainties.

The objective of this study was to investigate the feasibility of one particular approach to the quantification of uncertainty of geological boundaries – expert elicitation. The motivation for the study was specifically to see whether elicitation might be used to quantify the uncertainty in the geological boundaries mapped over the course of onshore surveys by staff of

the British Geological Survey (BGS), and so used to communicate this uncertainty to users of the maps and derived products.

Geological boundaries are uncertain for various reasons. The first is *conceptual* uncertainty. In some cases a geological boundary on a map can reasonably be expected to correspond (subject to other sources of uncertainty) to an unambiguous physical reality, a contact between two contrasting units. In such circumstances the boundary is conceptually certain, even if there is uncertainty about its position in space. In other cases a mapped boundary represents an interpretation of variation that is essentially spatially continuous, i.e. a gradational boundary. Representing such geological variation by a boundary on a map is somewhat artificial, even if it is practically useful, and the conceptual meaning of the boundary is therefore uncertain. By conceptual uncertainty we mean, therefore, the uncertainty attached to a mapped boundary which represents a division of spatially continuous variation rather than a contact between two unambiguously different geological units. Conceptual uncertainty arises because a boundary is an artificial way to represent geological variation in such circumstances, and there is likely to be variation between different surveyors or users of the final map in their understanding of the boundary and where it should be represented in space. This is also true of many boundaries on soil maps. Metamorphic boundaries, particularly those resulting from regional metamorphism, are often diffuse and defined by geochemical or mineralogical assemblage. In this case the identification of a specific boundary is rare, relying upon a balance of evidence that supports the transition from one assemblage to another. Similarly, facies boundaries representing different sedimentological environments can present a range of boundary types (gradual, interdigitating, complex) where a clear separation of the units is difficult to establish, but must occur within an implicit zone. In this paper we do not consider conceptual sources of uncertainty, but consider cases where the geological reality that the mapped boundary aims to represent could, in principle, be observed directly and unambiguously. This would require the removal of overlying material – all vegetation and material altered by pedogenesis and anthropogenic processes such as cultivation where the delineated units are superficial deposits, and all superficial material when the solid geology is mapped.

The second type of uncertainty is *scale-dependent* uncertainty. Even where a boundary is conceptually unambiguous the precise position at which it should be described as a continuous line may depend on the spatial scale at which it is observed, and entails some degree of generalization of fine-scale variation. This is a consequence of fractal or quasi-fractal behaviour (Burrough, 1983). While “the coast of Britain” is a conceptually unambiguous boundary, its representation as a continuous line, and hence its measured length, depends on the scale of observation (Mandelbrot, 1967). Scale-dependent uncertainty is a consideration when

a boundary generalized at some scale of field survey is used to make decisions at a larger cartographic scale. It may be inappropriate, for example, to use certain mapped boundaries to make decisions about the location of a proposed structure at a resolution of tens of metres. Further investigation would be needed to improve the information. A survey organization may ensure that scale-dependent uncertainty is allowed for in the use of its products by attaching a scale-dependent “buffer” to published boundaries, or by giving written guidance on their proper usage, or both.

*Cartographic* uncertainty is introduced when the field-surveyor’s mapped boundaries are converted to a cartographic product. It encompasses scale-dependent uncertainty because a cartographer will usually generalize field-mapped boundaries to a smaller cartographic scale, and will do so more or less successfully. However, there are additional sources of uncertainty in the formation of a published map from field sheets. In addition to scale-dependent uncertainty, cartographic uncertainty includes other errors that are introduced in this process including errors arising from digitization (Gong et al., 1995). In this paper we do not consider scale-dependent or cartographic uncertainty, considering only the sources of error in boundaries as mapped on a field sheet at the typical UK mapping scale of 1 : 10 000.

The source of uncertainty that we consider here is *interpretation uncertainty*. This arises because, in many settings, the geological boundary of interest cannot be observed everywhere. Over most of the mapped length of a boundary, therefore, the position is based on the mapper’s interpretation of available information. Consider a simple case where the boundary position is constrained at two locations. The constraint may be strong (e.g. the contact of interest can be observed directly in a quarry or other exposure) or weak (e.g. it can be inferred that the crop line for a unit occurs somewhere between one borehole where the unit is seen to be in subcrop beneath superficial deposits and a second where the boundary is below other bedrock units). At intervening locations the possible position of the boundary is constrained by limited local direct observations, e.g. by topographic features such as breaks of slope, by spring-lines etc. and by geophysical measurements. The mapped position of the boundary is the geologist’s best expert interpretation of the available information. It is therefore subject to error because it is based, inevitably, on conceptual models (e.g. of the control of surface features by subsurface structure) which are themselves imperfect, which do not fully determine the position of boundaries even when good and dense observations are available (Brodaric et al., 2004) and which must be implemented with imperfect and partial information.

## 1.2 Past work on the uncertainty of geological boundaries

The uncertainty of linear features in geographical information has been the subject of considerable research. Much of

the research on *conceptual* uncertainty has been done in the context of soil mapping where mapped boundaries do not, in general, attempt to reproduce unambiguous boundaries between soils on the ground, but represent an interpretation of continuous variation. The utility of such boundaries is that they parcel up the landscape into regions which should be more internally homogeneous than the landscape as a whole, and so provide a basis for spatial prediction (by the regional mean). Webster and Beckett (1968) and successors such as Leenhardt et al. (1994) have examined the utility of such information by analysis of the variance components of terrain properties that one might predict from the delineated units.

There has been considerable interest in *scale-dependent* uncertainty, including the modelling of boundaries as fractal objects. The extent to which the generalization of a boundary at some scale introduces uncertainty into the resulting map can be measured by the proportion of sites within a delineated map unit which correspond to the notional class (soil, stratigraphic, etc.) to which the unit nominally corresponds. This proportion may also be affected by interpretation uncertainty, but Lark and Beckett (1998) presented a model for errors in soil maps which can be attributed to the generalization of the spatial pattern below some threshold scale.

*Cartographic* uncertainty is a large topic. Chrisman (1982) provided an early quantitative framework for its evaluation, and it has been the subject of empirical studies (e.g. Gong et al., 1995). At the BGS, all digital data products are provided with guidance for users concerning appropriate use at scale, given the cartographic uncertainty. Typically the advice uses the following form of words:

“The cartographic accuracy is nominally 1 mm which equates to 50 m on the ground at 1 : 50 000 scale. This is a measure of how faithfully the lines are captured; it is not a measure of the accuracy of the geological interpretation.”

*Interpretation* uncertainty is challenging to quantify. It arises from the imperfection of the conceptual models that the geologist uses to interpret available data, but also from the sparsity of those data. As noted by Brodaric et al. (2004), for some set of observations and a conceptual model for interpretation, the underlying distribution of boundaries is generally underdetermined, i.e. the rational interpreter is not constrained to a single interpretation. The interpretation may be expected to be more constrained the denser the data. For this reason one may think of the interpretation error in geological boundaries as a random process the variability of which depends on the density of available data, the complexity of the geological processes in the conceptual model and factors (experience, etc.) which may influence individual interpretation.

The parameterization of a model of boundary uncertainty is not straightforward. Most progress has been made in cases where boundaries are part of a statistical model for some

densely sampled or quasi-continuous measurements of some variables (e.g. geochemical data, geophysical variables). In this case a statistical model may be invoked for how the boundary uncertainty affects predictions from the model. Examples of this are given by Lillah and Boisvert (2013), Silan-Cárdenas et al. (2009) and Guillot et al. (2006) and an interesting stochastic model for uncertainty in geographical polygons is offered by Heuvelink et al. (2007). However, in the case of conventional geological survey, boundaries do not emerge from a statistical model for a response variable, but are the result of expert interpretation. Their uncertainty can therefore not be obtained directly from a statistical model. One way to examine the uncertainty would be to do so empirically.

Empirical assessments of interpretation error have been undertaken in the context of seismic interpretation (Bond et al., 2012), soil survey (Burrough et al., 1971) and 3-D geological modelling (Lark et al., 2013, 2014). These workers evaluated uncertainty in expert interpretation empirically, based on validation data. This allows one to examine the variability of interpretation errors, and the contribution of between-interpreter effects as well as differences between geological settings and the density of available observations. A similar empirical approach is reported by Albrecht et al. (2010) who examined between-interpreter variation of boundaries around objects in remotely sensed images.

The problem with the empirical approach is that it requires substantial effort. If one wishes to evaluate the uncertainty of geological boundaries empirically then one requires a number of geological maps of the same area, produced independently conditional on a (common) set of observations, and with sufficient local validation observations of the boundaries of interest, perhaps from geophysical data, boreholes, excavations or geological exposures. These validation data must not have been available to the surveyors. Such studies are very resource intensive, and provide information on uncertainty only for the geological setting of the particular study, and the nature and density of available supporting observations. For this reason we consider expert elicitation as an alternative approach.

### 1.3 Expert elicitation

#### 1.3.1 The objective of elicitation

Expert elicitation in this context is based on the assumption that the experienced geological mapper has a mental model of the uncertainty that is attached to mapped boundaries. This model comes from the geologist's awareness and experience of the variability of geological phenomena. It also reflects the geologist's awareness of how, in a particular setting, direct observations and the interpretative model of topographic features and other surface expression of geological structure and lithology constrain the possible distribution of boundaries. This model is almost certainly tacit rather than explicit,

still less can the geologist write it down in statistical terms. Nonetheless, the expert, through his or her experience, has an intuitive sense of the reliability of information. This fact is recognized in some survey procedures. For example, traditional geological mapping has always distinguished between boundaries that can be regarded as directly observed at the scale of survey and those inferred from other evidence. This expert assessment of uncertainty may be communicated on a conventional map by using solid lines for observed boundaries and dashed lines for those that are inferred. Expert elicitation methods have been used elsewhere in earth sciences – for example Martí et al. (2008), Truong et al. (2013), Polson and Curtis (2010), Bond et al. (2007) and Scourse et al. (2015).

### 1.3.2 What do we want to elicit?

Our objective is to elicit a model of uncertainty in mapped line work collected by the BGS in the UK according to its protocols and procedures. This requires careful definition. We do not want to elicit a model of the uncertainty in the position of a geological boundary, given available information, as one might do with a panel of experts at a particular field site. We require, rather, models of uncertainty which can be used to quantify that uncertainty in a range of instances where pre-existing BGS line work is used for inference and decision making. We therefore require a model of the uncertainty of the true position of a geological boundary given the position of the mapped boundary drawn by a geological surveyor following standard BGS protocols.

We chose to elicit the tacit model of uncertainty in geological boundaries in the context of a notional test of a mapped boundary along a 1-D line. Consider a transect perpendicular to a mapped geological boundary. The mapped boundary intersects the transect at a location  $x_m$  units from an arbitrary origin of the transect. We assume (see above) that the boundary is not subject to conceptual, scale-dependent or cartographic uncertainty, but only to interpretation uncertainty. This arises from the fact, for example, that the units separated by the boundary are largely covered by a thin, but possibly irregular blanket of concealing material including vegetation, soil and superficial deposits, so the interpretation is based on topographic features and some limited information from boreholes and exposure. This means that, if we were to excavate the overlying concealing material along the transect, we could identify the position where the actual boundary intersects the transect (true intersection) at a location  $x_t$  units from the arbitrary origin of the transect. Because of the interpretation uncertainty the difference between these positions,  $\varepsilon = x_t - x_m$ , is not, in general, equal to zero but is a variable with a distribution. The geological mapper's tacit model of boundary uncertainty implies some form for this distribution such that there exists a probability that  $\varepsilon \in [\varepsilon_l, \varepsilon_u]$  where  $\varepsilon_l$  and  $\varepsilon_u$  are real-valued limits and  $\varepsilon_l < \varepsilon_u$ .

We recognize that the uncertainty about the true position of a geological boundary, given its mapped position, is likely to vary between geological settings and the context of the original survey (e.g. the frequency of observations, whether the setting is urban or rural, whether land is under a crop, grass or woodland, etc.). For this reason we cannot elicit a single general model of uncertainty in boundaries. Rather, it is necessary to define a number of general scenarios for each of which it is plausible to make generalizations. A scenario is defined in terms of the nature of the geological boundary (the contrasting units, topography, etc.), land-cover and frequency of observations (auger borings, exposures, etc.). One example scenario could be a contact between two distinctive sedimentary rocks observed on a moderate slope under cultivated land (see the Appendix for scenario descriptions for this study). There is scope for variation in the magnitude of error of mapped boundaries within specific cases consistent with the scenario description, but generalizations can still be useful if the within-scenario variation is smaller than variation between scenarios. In the context of this elicitation procedure the expert is explicitly asked to consider the range of the within-scenario variation by considering the likely set of errors associated with a set of 100 instances of the scenario. This approach, a “frequency representation” of the scenario, is recognized in the elicitation literature as a way to encourage the expert to access the range of his or her experiences of a particular scenario rather than concentrating on particular single cases said to be more “available” (O’Hagan et al., 2006, and see Sect. 1.3.4).

### 1.3.3 Whose mental model do we want to elicit?

As noted in the previous section, the expert's tacit model of boundary uncertainty implies a probability that the boundary error in a particular transect across a boundary corresponding to a specified scenario falls within certain limits. This probability could be called the expert's personal or subjective probability that the difference between the true and mapped intersection falls in this interval. “Personal” or “subjective” imply that the tacit model depends on the particular expert's experience and understanding. Expert elicitation is the process of identifying the form of the statistical distribution implicit in the personal probabilities under an expert's tacit model of boundary uncertainty (O’Hagan et al., 2006). In planning this particular study we had to address the question of whose mental model of boundary uncertainty we wished to elicit.

Our objective (see previous section) was to elicit a model of the uncertainty of the true position of a geological boundary given the position of the mapped boundary drawn by a geological surveyor following standard BGS protocols. We are therefore interested in the tacit uncertainty model of geologists with considerable field experience of mapping in the UK and understanding of BGS protocols with access to experience of the scenarios for which our elicitation was con-

ducted. Such individuals are aware of the process by which boundaries are mapped by the interpretation of limited observations, surface features and other clues, and are aware from their field experience of the extent to which the position of the boundary is constrained by available information, and how this may vary for any scenario. For this reason we chose to elicit the tacit model of senior BGS geologists with field mapping experience.

### 1.3.4 What methods are appropriate for elicitation?

Expert elicitation is a structured approach to the formulation of a model based on expert knowledge. Different approaches can be taken to elicitation (e.g. O'Hagan et al., 2006). All methods take account of psychological research into the strategies or heuristics by which individuals form opinions about uncertain events. One common heuristic is “anchoring and adjustment” whereby individuals start with an initial estimate and then adjust it up or down. A common problem is that the adjustment made is not sufficient, given the evidence on which it is based. Elicitation methods need to reduce the “anchoring” effect whereby an elicited output is too strongly influenced by initial information. “Availability” is another important heuristic, in judging probabilities of outcomes individuals access cases from their own experience, and the availability of these may vary, possibly causing bias if factors other than frequency affect availability. For example, events of larger magnitude, or more recent events, may have greater availability than others.

There are two general approaches taken to elicitation from a group of two or more experts. The first is mathematical aggregation by which elicited outputs from different experts are combined, for example by a simple or a weighted averaging. The method of Cooke (1991) uses weights from a “seeding” elicitation where the target quantity is known and the success of each expert in reproducing the known information can be used to weight their opinions in succeeding elicitations. An alternative approach, behavioural aggregation, is based on the use of group discussion to arrive, if possible, at a consensus view starting from the results of separate individual elicitations. Behavioural aggregation can be effective when groups of experts recognize which members have most expertise in particular cases, where the group is guided by a facilitator to avoid problems such as anchoring and undue influence by dominant personalities and where carefully structured protocols are used, ideally with feedback which shows the implications of the expert judgements (Reagan-Cirincione, 1994).

In this paper we use established methods of expert elicitation to obtain statistical distributions for the variable  $\varepsilon$  in a number of scenarios by behavioural aggregation. The primary reason for doing this was the considerable difficulty in obtaining reasonable examples on which to base seeding elicitations for an application of Cooke's method (Cooke, 1991). To assess the elicited distribution for an expert in a particu-

lar scenario would require a sizable number of independent observations of boundary error in cases of that scenario, and we have already discussed why such empirical information on boundary error is extremely difficult to obtain.

The objective of this study was to evaluate the feasibility of running elicitations with behavioural aggregation according to established protocols with groups of experienced geological mappers. From this we aimed to assess whether larger-scale elicitations could be conducted to assess the uncertainty of mapped boundaries in a wider range of settings.

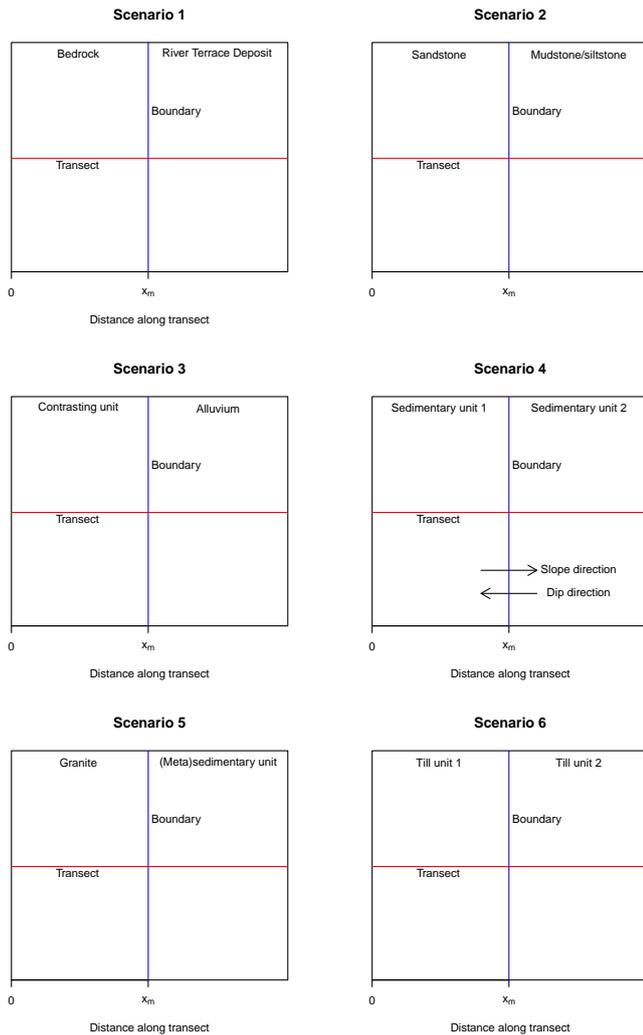
## 2 Methods

### 2.1 The elicitation framework

The Sheffield elicitation framework (SHELF) is described by Oakley and O'Hagan (2010). It uses the behavioural aggregation approach to group elicitation. The protocols are based on research into elicitation procedures reviewed by O'Hagan et al. (2006) and come with software which can provide feedback on the implications of group judgements during the elicitation. SHELF has been used for expert elicitation in various fields including veterinary medicine (Higgins et al., 2012), modelling of atmospheric processes (Lee et al., 2013), modelling of water distribution networks (Scholten et al., 2013), forecasting of energy demands (Usher and Strachan, 2013) and power analysis for clinical trials (Ren and Oakley, 2014). SHELF provided the basis for the elicitation procedure that we used. However, we cannot formally describe our elicitation as conducted according to the SHELF framework because we did not record personal interest and expertise statements from the participants. This is because all participants are current or recently retired members of staff at the BGS whose field experience and external interests are a matter of record. However, in SHELF these statements are recorded as metadata and are not used in the elicitation itself. Furthermore, we held a final feedback meeting after completion of the elicitation to give participants an overview of the outcomes and to allow them to register any concerns or change of opinion. In other respects we used the pro formas and software of the SHELF procedure.

In our elicitation procedure we followed SHELF guidelines, as described in detail in Sect. 2.3 below. We defined a set of scenarios for which we wanted to elicit probability distributions of  $\varepsilon$ . These were defined by an experienced geological surveyor (A. J. M. Barron) who did not serve as an expert for purposes of the elicitation, but rather as a geological facilitator. R. M. Lark served as statistical facilitator of the elicitation, having facilitated previous elicitations at the BGS using a framework based on SHELF.

In accordance with SHELF procedures, a briefing document setting out some principles of probability, elicitation and explaining the scenarios of interest was prepared and sent to all participants. There was then a briefing session to



**Figure 1.** Diagrams indicating dispositions of units in each scenario with the mapped boundary shown as a blue vertical line and the notional transect as a red line, perpendicular to the boundary and with an origin on the left.

explain this material and address any questions, and to conduct a practice elicitation to familiarize participants with the procedure. The main elicitation was then conducted in a single day; elicitation records were kept in line with SHELF protocols. After this a summary of results was presented to the participants, and a final feedback meeting was held to ensure that participants agreed that the outcomes reflected group opinions.

## 2.2 Selection of panel and definition of scenarios

The geological facilitator (A. J. M. Barron) and a BGS geologist with both field experience and specialist experience of geological product development (R. S. Lawley) met with R. M. Lark to agree on a common understanding of the goals of the project and to agree on a set of participants to consti-

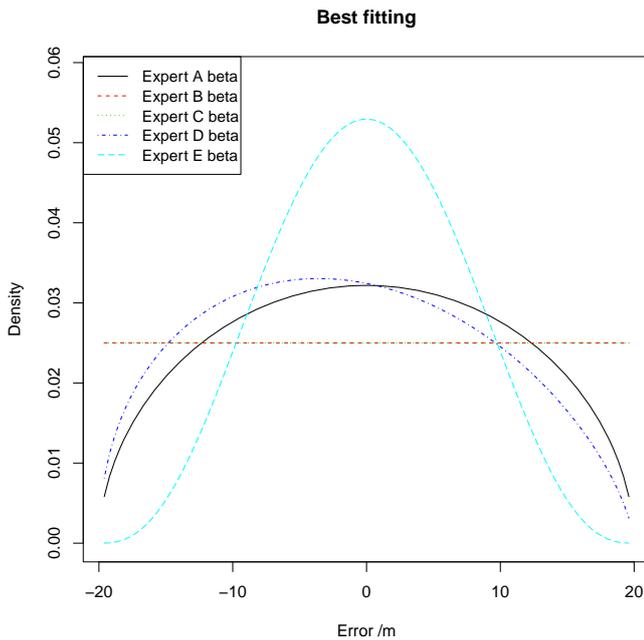
tute the panel (the panel members are the remaining authors of this paper). SHELF guidelines are to recruit a panel that is not too large (about five members) and who can work together rather than individually. A panel was identified which comprised five geologists with field experience in a range of settings. A. J. M. Barron then defined a set of scenarios, designed to encompass a range of conditions reflecting the mapped geological boundaries held by the BGS. A scenario was defined in terms of a general geological setting for a boundary. It was not defined with respect to particular stratigraphic units, but rather in terms of contrasting lithologies or deposits that would correspond to a common setting. The scenario was also defined in terms of land cover, any local exposure, and the frequency of augering in the case of superficial material. In some cases discussion of the scenario during the elicitation identified ways in which its definition required clarification. Since A. J. M. Barron was present as a facilitator, this could be done consistently, and any such modifications were recorded.

Scenario definitions are given in the Appendix along with modifications agreed during the elicitation. Figure 1 illustrates the mapped settings and the dispositions of the units relative to the notional transect. It is important that this is understood by all the group. For example, in scenario 1, Fig. 1 shows that a negative value of  $\varepsilon$ , which means that  $x_t < x_m$ , implies that the mapped boundary, indicated by the vertical blue line, is too far onto the river terrace deposit. Figures showing these dispositions were provided to participants during the elicitation.

## 2.3 Conduct of the elicitation

### 2.3.1 Briefing and practice elicitation

The SHELF guidelines (Oakley and O'Hagan, 2010) require an appropriate briefing for all participants. To this end a briefing document was produced. This explained why the elicitation was to be undertaken and what, in outline, an elicitation is. It gave a brief introduction to the model of errors in mapped boundaries, as set out in Sect. 1.3 above, and a reminder of the concepts of probability and of distributions and percentiles (specifically quartiles) of random variables. The elicitation task was then set out in terms of a frequency representation. That is to say the participants were told that they would be considering a notional set of 100 randomly and independently selected locations drawn from any one scenario. At each location a transect is considered, perpendicular to the mapped boundary as illustrated in Fig. 1. At each location the position  $x_t$  of the true intersection of the boundary is identified, and an error  $\varepsilon$  evaluated. The distribution to be elicited is the one realized in the histogram of the notional 100 observations of the error and, under the elicitation used, this entails making expert judgements about quartiles of the distribution. O'Hagan et al. (2006) note that this approach, in which a panel is required to visualize a range of instances



**Figure 2.** Individual best-fitting distributions for initial expert quartiles, scenario 2. Note that the lines for experts B and C coincide.

of one scenario, can be useful for ensuring that the experts consider a full range of possibilities under the scenario and not just those (most frequently or recently observed) which are more readily available (see Sect. 1.3.4). The scenario descriptions were also included in the briefing document.

The briefing document was circulated to participants a little over 2 weeks before a briefing meeting, and they were requested to read it in advance. In the briefing session, which took place the day before the main elicitation, the content of the document was reviewed, and participants had the opportunity to raise questions about any aspect of the procedure. In accordance with Oakley and O’Hagan (2010) the briefing session concluded with a practice elicitation to familiarize participants with the elicitation procedure. Oakley and O’Hagan (2010) stated that this practice elicitation should be to elicit the distribution for a variable unknown to the participants but known by the facilitator. In this case the distribution which was elicited was that of ages of delegates to the 2013 European Geosciences Union congress. This information is made available after the congress, but panel members had not seen it.

Ideally, more time would be available between the briefing and the main elicitation to allow agreement of any modifications to the scenarios or procedure, but this was not possible due to the participants’ availability. No difficulties of understanding or disagreements over the scenarios and their description emerged in the course of the briefing session.

### 2.3.2 Group elicitation

The main elicitation was conducted on 13 November 2013. The elicitation took place in a meeting room where all participants and facilitators could sit undisturbed around a large table. Hard copies of the scenario descriptions and associated figures (see Fig. 1) were provided to all participants. The room was equipped with a data projector which allowed elicited distributions and other feedback generated by the SHELF procedures to be seen by all participants. A flip chart was also used to record results from the individual elicitations so that these could be viewed by all participants. The geological facilitator (A. J. M. Barron) and the statistical facilitator (R. M. Lark) were present throughout the elicitation, as were all participants, the project administrator and a student who attended to gain experience of the elicitation method.

We used the Quartile method in the SHELF framework for both initial individual elicitations and the group elicitation Oakley and O’Hagan (2010). This was chosen because it had previously been successfully applied with a panel of geologists to elicit distributions pertaining to shallow geohazards. The method proceeded in three stages.

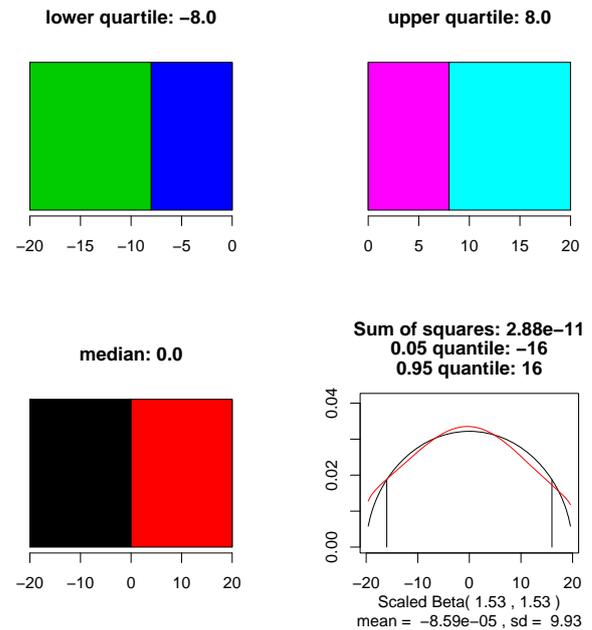
1. The scenario was presented. The group as a whole was then asked to provide upper and lower absolute bounds on the error variable,  $\varepsilon$ . This was done through a group discussion. The group was reminded that these bounds are minimum and maximum possible values of the variable, and the probability of a value of  $\varepsilon$  occurring in a range near these bounds may be very small. The group was reminded of the meaning of negative and positive values of  $\varepsilon$  in terms of the position of the mapped boundary on each unit that defines the scenario.
2. Each individual was then required independently to choose values of the median (second quartile) and the first and third quartiles of the distribution of  $\varepsilon$  which reflect their expectations. Since we were considering (see Sect. 2.3.1) a notional independent random sample of 100 intersections with boundaries corresponding to the scenario, this was framed in terms of, respectively, the value such that 50 locations had a larger value of  $\varepsilon$  and 50 a smaller; the value such that 25 locations had a smaller value of  $\varepsilon$  and 75 a larger value, and the value such that 25 locations had a larger value of  $\varepsilon$  and 75 a smaller value. Each participant recorded their values on a sheet with their name. Individual best-fitting distributions were then found for each set of quartiles, given the upper and lower bounds. This was done with the `elicit.group.values` procedure in the `SHELF2.R` code presented by Oakley and O’Hagan (2010) for use on the R platform (R development core team, 2013). Version 2.01 of the `SHELF2.R` source, modified on 11 November 2012 was used. This procedure generated a plot with the probability density function (PDF) for the dist for

each panel member. Figure 2 shows an illustrative plot for scenario 2 (although the axis labels and the legend have been somewhat modified from the original code). This plot was visible to all participants on the projector screen. The individual quartiles were also written on the flip-chart. Note that the participant code varied arbitrarily from one scenario to the next, so the distributions were anonymized, although participants in all cases chose to acknowledge their initial results in later discussion. The individual sheets with the initial values were retained at the end of the elicitation.

- The participants, as a group, were then asked to determine a group consensus set of quartiles. The discussion was allowed to proceed spontaneously, with the facilitators intervening when a particular question arose or, in the case of the statistical facilitator, if any comments made in the discussion indicated a misunderstanding of the nature of the probability model or the error variable. A visual display to facilitate this is generated by the `elicit.group.values` procedure, and this is illustrated in Fig. 3 for scenario 2. As values for the median, first and third quartiles are adjusted the values are displayed (panels in the top row and bottom right panel). A probability density function, the best fitting PDF of a set of distributions, to the quartiles, given the limits, was estimated and displayed (black line in bottom right panel) along with the mean and standard deviation (SD) and the 0.05 and 0.95 quantile, encompassing a 90 % probability interval. However, this feedback was generally consulted by the group at the end of the discussion.

### 2.3.3 Feedback

After the elicitation was completed a summary document was prepared. This contained the group-elicited quartiles and the lower 2.5th and upper 97.5th percentiles of the fitted distributions encompassing a range within which one would expect to find 95 % of boundary errors along the transect. These were also displayed graphically. The first output that we plotted displayed the elicited quartiles as a piecewise-uniform distribution, i.e. one in which the probability density is uniform over each of the four intervals defined by, respectively, the lower bound, first quartile, median, third quartile and upper bound. The density function for the best-fitting distribution among the set considered in the `elicit.group.values` procedure was also plotted on the same axes (see Fig. 4). The density function of a statistical distribution can be used to compute the probability that a random variable with that distribution falls in a particular interval. Different distributions have different properties, and so one may compare the fit of a range of distributions to find the most appropriate one in a particular case. A distribution's density function has parameters that define its shape, these are the mean and standard deviation in the case of the normal distribution. We used three



**Figure 3.** SHELF Feedback plot during group elicitation of the distribution for scenario 2.

distributions here. The most common was the Beta distribution. This has the density function

$$f_{\text{scaled}\beta}(x|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1} (1-y)^{b-1}, \quad (1)$$

where

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}},$$

$a$  and  $b$  are parameters that define the shape of the distribution and  $\Gamma(\cdot)$  denotes the Gamma function. This function defines the probability density for a variable which takes values in the range  $[0,1]$ . This variable is then rescaled to the range defined by the minimum and maximum values for the variable of interest, as set out in Table 1.

The Gamma distribution has the density function

$$f_{\text{Gamma}}(x|c, s) = \frac{1}{s^c \Gamma(c)} x^{c-1} e^{-\frac{x}{s}}, \quad (2)$$

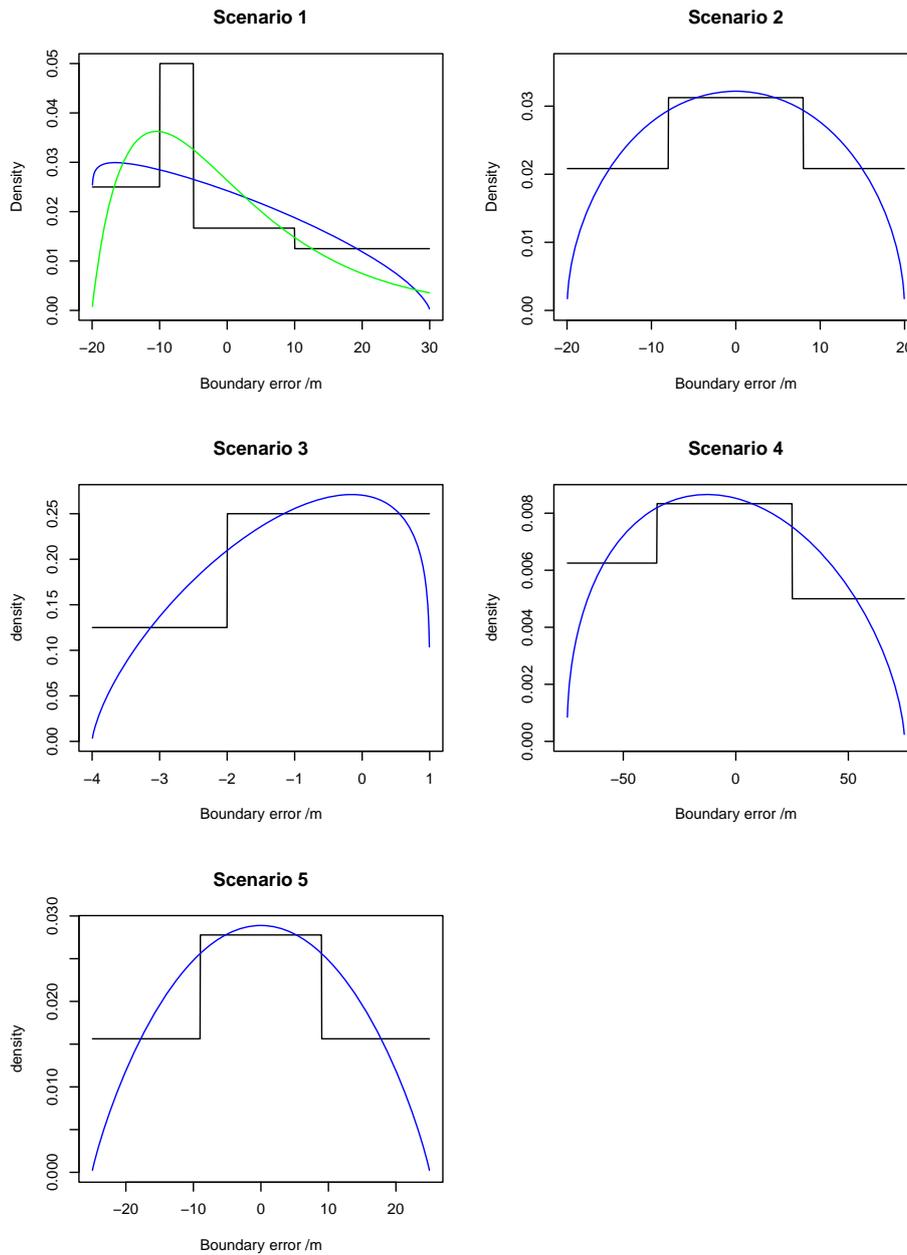
where  $s$  and  $c$  are parameters that define the shape of the distribution.

The normal distribution has the density function

$$f_{\text{Normal}}(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (3)$$

where  $\mu$  and  $\sigma$  are parameters, the mean and standard deviation respectively.

In one case (scenario 1) the goodness of fit of two competing distributions was very similar, so both were included



**Figure 4.** Best-fitting distributions for expert group quartiles, scenarios 1–5. Solid black line shows uniform density over elicited interquartile ranges, blue line show best-fitting distribution, green line shows closely competing alternative distribution.

in the summary document. The document was completed 15 days after the elicitation and circulated to all participants. They then participated in a discussion meeting after a further 12 days, at which they were asked whether they were still content with the group consensus statistics and, in the case of scenario 1, which of the two competing distributions, given the density plot and the 95 % interval, best represented their own expectation of the error distribution in the scenario.

### 3 Results

The initial group-agreed plausible range and the individually elicited quartiles for each scenario are presented in Table 2. Table 2 presents the group-elicited quartiles and the fitted distributions with parameters. Figures 4 and 5 are PDFs for fitted statistical distributions (blue lines) for the location of the boundary relative to the mapped position in each case. In addition, the elicited information on quartiles to which these distributions were fitted is represented by show-

ing corresponding piecewise-uniform density functions over the elicited interquartile ranges (black lines).

We now present brief summaries of key discussion points that arose in the course of each elicitation.

### 3.1 Scenario 1 – edge of river terrace deposit on bedrock

The first 15 min of the group discussion to agree on upper and lower bounds for this scenario was taken up with more general issues about the elicitation which had clearly occurred to participants since the briefing meeting, but these are reported here because they were raised only after the scenario had been introduced. One concern was whether results from this elicitation would be applied as quality measures or buffers to BGS's boundary-based products. Participants were assured that the present elicitation, about generalized scenarios, was an exploratory study, to inform any future use of elicitation for products. Some further issues to do with the kinds of uncertainty to be considered in this elicitation were clarified, specifically that effects of cartographic error or location error on the field map should be ignored, and that error at the scale of generalization of a field map sheet on a scale of 1 : 10 000 should be considered.

The discussion specifically to agree upper and lower bounds took 40 min. In the course of this discussion the principal issues were as follows:

1. In practice the mapping of superficial material has been influenced by the thickness of this deposit. The question was therefore raised of whether the boundary would be defined where the river terrace thinned to some minimum thickness rather than where the bedrock was at surface. After some discussion it was agreed that, in the particular setting (as opposed to a setting where superficial material is patchy), this consideration could be set aside.
2. Different surveyors would make different decisions as to whether to map head deposits arising from cryoturbation in this setting, which could lead to variation in the boundary location.
3. The extent to which the boundary is expressed as a sharp break of slope of the land surface will affect the variability of boundary error.

The geological facilitator indicated that it should be assumed that head is not mapped in this scenario and that the break of slope is a subtle feature. On this basis it was agreed that the surveyor would aim to map the break of slope as a feature indicating the boundary, but would not identify it precisely. Slightly asymmetric bounds were agreed, implying that the largest possible absolute error would be with the mapped boundary too far onto bedrock.

The individual and group elicitation of quartiles took 26 min in total. Three participants proposed a zero median

error, and the main difference was between one participant who argued for a slightly positive median, arguing that surveyors would tend to map the boundary too far onto bedrock, misled by isolated patches of terrace material, while another argued that there would be a tendency to map too far onto the terrace material due to problems identifying the edge as the deposit thins out. This illustrates how the panel were capable of accessing, in the sense of Sect. 1.3.2, the range of possible conditions consistent with a particular scenario. In some instances of scenario 1 the mapper may have encountered such patches, in others not. The result is a distribution of errors for the scenario, with a particular shape. This latter participant convinced the others that a negative median was appropriate, and agreed on a smaller absolute median error than in his individual elicitation, given the frequency of augering in the scenario description. Once this was agreed a consensus on the first and third quartiles was quickly achieved.

### 3.2 Scenario 2 – base of sandstone in mudstone/siltstone succession

The discussion to agree upper and lower bounds took 24 min. The principal consideration determining the interval in this setting was the scope to extrapolate from observations in the quarry, and the factors that would control the precision of this, specifically the urban setting. Once these bounds were agreed the individual and group elicitation of quartiles took 14 min. Again, the process of extrapolating from the quarry was critical in the group discussion. It was agreed that where this boundary was inferred solely from surface topography the first and third quartiles would be asymmetric about the median, with a tendency to map the boundary too far downslope, but that in the setting as described symmetrical quartiles were appropriate.

### 3.3 Scenario 3 – edge of alluvium/tidal river deposit against contrasting underlying geology

The discussion to agree upper and lower bounds took 10 min. It was agreed that this boundary should be relatively easy to identify in the field, so the interpretation uncertainty would be small relative to subsequent cartographic sources of error. There was some discussion as to whether a larger upper limit should be considered, because of the possibility in some circumstances of putting the boundary too far upslope (onto the bedrock) due to recent deposition of flood material, but it was agreed that cultivation, as indicated in the scenario description, made this unlikely. The individual and group elicitation of quartiles took 12 min. It was agreed that errors downslope (putting the boundary too far onto the alluvium) would be likely to predominate, and so it was appropriate to have a negative median and an upper quartile of zero.

**Table 1.** Agreed plausible range and initial individual quantiles for boundary error distributions under each scenario.

Scenario	Lower limit/m	Upper limit/m	Quantile/m	Expert				
				A	B	C	D	E
1	-20	30	Q1	-5	-5	-10	-5	-15
			Median	5	0	0	0	-10
			Q3	20	5	15	15	10
2	-20	20	Q1	-8	-10	-10	-10	-5
			Median	0	0	0	0	0
			Q3	8	10	10	5	5
3	-4	1	Q1	-2.5	-2	-2	-2	-1
			Median	-1.5	0	-1	-1	0
			Q3	0	0.5	0	0	0.25
4	-75	75	Q1	-30	-30	-35	-40	-40
			Median	0	0	0	0	-20
			Q3	30	30	35	40	10
5	-25	25	Q1	-5	-20	-15	-10	-8
			Median	5	0	0	0	0
			Q3	15	10	15	10	8
6	-1000	1000	Q1	-30	-50	-75	-250	-100
			Median	0	0	0	0	0
			Q3	50	50	75	250	100

**Table 2.** Group quantiles for each scenario and best-fitting distribution.

Scenario	Q1	Median/m	Q3	Distribution						
				Type	Parameters		Mean/m		Percentile	
							2.5th	97.5th		
1	-10	-5	10	Beta	$a = 0.97$	$b = 1.56$	-0.02	-19.2	24.9	
				Gamma <sup>a</sup>	$c = 1.90$	$s = 0.09$	0.0	-17.8	37.2	
2	-8	0	8	Beta	$a = 1.53$	$b = 1.53$	0.0	-17.5	17.5	
3	-2	-1	0	Beta	$a = 1.70$	$b = 1.21$	-1.07	-3.5	0.9	
4	-35	-5	25	Beta	$a = 1.44$	$b = 1.62$	-4.45	-67.4	63.5	
5	-9	0	9	Beta	$a = 1.87$	$b = 1.87$	0.0	-20.7	20.7	
6 (D)	-120 <sup>b</sup>	0	120 <sup>b</sup>	Normal	$\mu = 0$	$\sigma = 178$	0	-349	349	
6 (B)	-50	0	50	Normal	$\mu = 0$	$\sigma = 74$	0	-145	145	
6 (A, C, E)	-75	0	75	Normal	$\mu = 0$	$\sigma = 119$	0	-233	233	

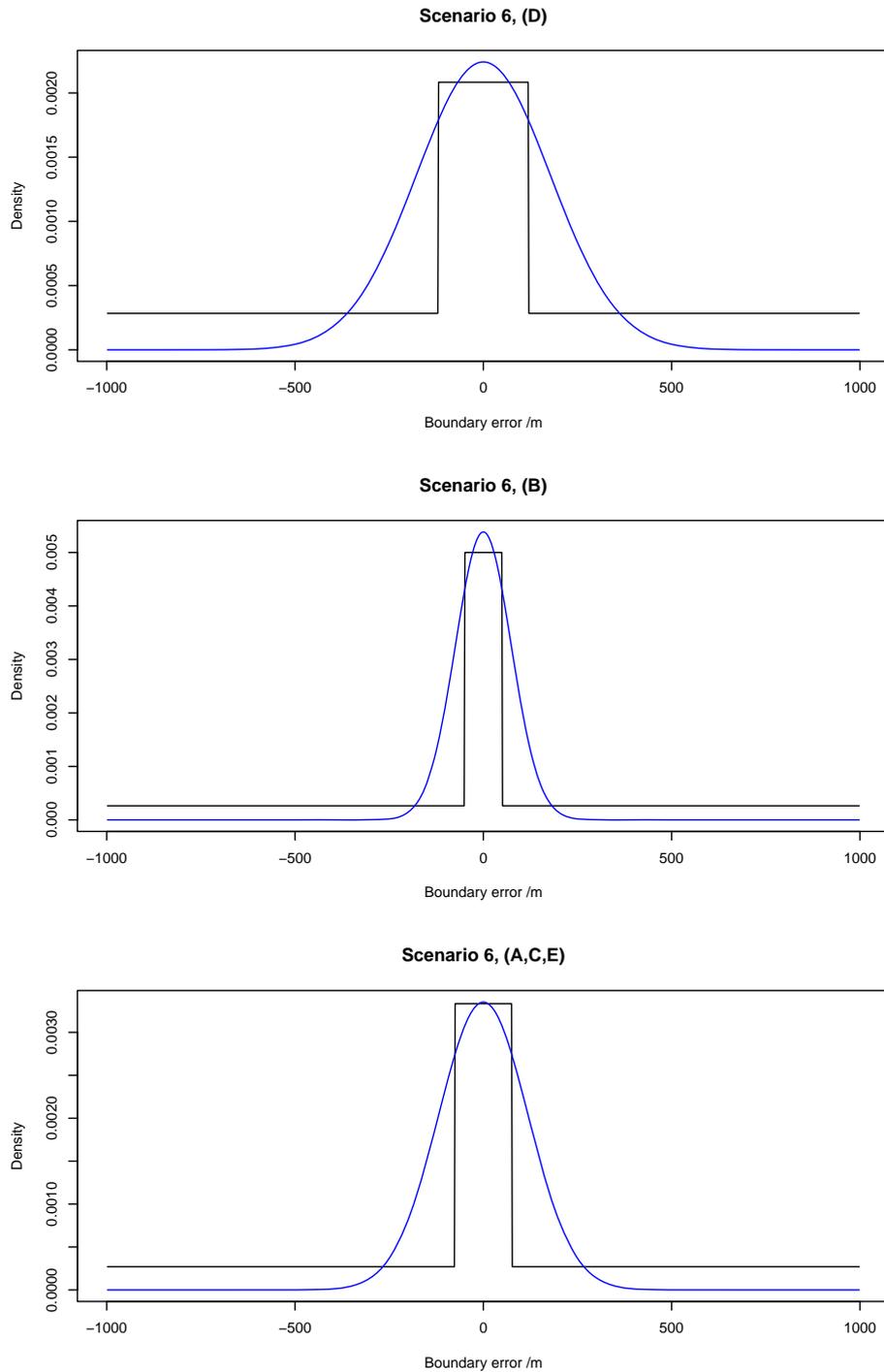
<sup>a</sup> The sum of squares for the fits of the Beta and Gamma distributions were very similar ( $6.3 \times 10^{-3}$  and  $6.5 \times 10^{-3}$  respectively) so both are reported here and were presented to the panel at the feedback. The panel agreed unanimously that the Gamma distribution best represented their perception of uncertainty of the line work for this scenario. <sup>b</sup> These are the only quantities that was adjusted during feedback. The expert in this group adjusted his quartiles to these values from -150 m and 150 m respectively.

### 3.4 Scenario 4 – stratigraphic boundary between two distinctive sedimentary rocks

The discussion to agree upper and lower bounds took 14 min. There was some initial disagreement as to whether this scenario was one in which field survey would be appropriate. One participant felt that it was not, but changed his view on this given the modification to the scenario that the superposition relationships of the units are assumed known, the sce-

nario is not approached “cold” but as part of a broader survey campaign in which this information would be developed.

The individual and group elicitation of quartiles took 10 min. One participant put asymmetrical quartiles in his individual elicitation, and argued in the group elicitation that this was necessary because downslope movement of surface brash (loose broken rock in soil) could result in larger errors in this direction. One participant, in response, queried whether the field surveyor would use brash in mapping.



**Figure 5.** Best-fitting distributions for quartiles of scenario 6 according to three expert subgroups (letters are expert codes). Solid black line shows uniform density over elicited interquartile ranges, blue line show best-fitting distribution.

A third participant suggested that the use of brush would depend on whether the particular survey was being undertaken rapidly or for a more detailed project so, over the population of BGS line work, some instances of this scenario would be cases where brush was used as information to identify the

boundary. As a result of this discussion the group arrived at a consensus agreeing to specify asymmetric quartiles.

### 3.5 Scenario 5 – faulted boundary between granite and hard non-igneous rock

This scenario was discussed after a 1-hour lunch break. The discussion to agree upper and lower bounds took 14 min. The individual and group elicitation of quartiles took 10 min. In both these discussions there was some debate as to whether the error distribution would be asymmetrical due to greater exposure of the country rock near the fault due to induration (the process by which the country rock is hardened as a consequence of recrystallization in the vicinity of the fault). However, the consensus agreed in the group elicitation was that the exposure would be primarily due to increased weathering near recent faulting, and so not, in general, greater over one unit than the other. The consensus quartiles were therefore symmetrical.

### 3.6 Scenario 6 – boundary between two distinctive tills, unknown relationship

The discussion to agree upper and lower bounds took 40 min. There was some disagreement as to whether such a scenario would be mapped in practice with units that are very similar, except in respect of colour and clast content. The modification of the scenario to specify the spacing between auger traverses allowed progress in the discussion. However, there remained disagreements. One participant, inclined to put wide bounds, thought that low-angle contact between the units could make the boundary very uncertain. While others accepted that the geometry of the contact is harder to visualize in this scenario than others, they thought that low-angle contacts would be a worse-case scenario rather than typical. On the basis of this discussion wide absolute bounds were agreed. However, in discussion after the individual elicitation, it was clear that a consensus was not possible. Three distributions are therefore presented, two reflecting strongly contrasting views of two participants (both with experience in superficial mapping), and the third a majority view.

The feedback session resulted in no substantial changes to the outcomes of the elicitation session. The participants agreed that the Gamma distribution for scenario 1 (Fig. 4) was most appropriate. As shown in Table 2 Participant D made a small modification to his quartiles for scenario 6 (individual distribution), but the basic disagreement over this scenario remained. It was agreed that the scenario was a difficult one, with many unknown factors that it would be hard to control with differences in approach between mappers, particularly over time.

## 4 Discussion

This exercise showed that it is possible to use a method based on the SHELF framework to elicit the tacit model of uncertainty that geologists employ when interpreting line work. The general framework of the elicitation was workable, and

the approach was accepted as meaningful by the five geologists from whom the distributions were elicited. Given this, expert elicitation provides a method to extract the tacit mental model of uncertainty that geological surveyors access when interpreting line work on boundaries. This could be useful as a step in developing methods to represent this uncertainty to map users (e.g. by adding buffers to boundaries, as is already done to represent cartographic uncertainty). It also allows us to retain the mental model of uncertainty held by experienced surveyors after their retirement. This could be usefully applied in cognate disciplines such as soil or vegetation survey where spatial phenomena are mapped in the field by experts.

The group voiced a reservation about the extent to which distributions elicited for a general scenario could be usefully applied to individual instances of that scenario. For practical purposes it was thought that elicitations should be undertaken for more tightly framed situations such as a boundary between specific units in a particular region, or a fault near a frack zone or proposed site for a development. Any model of uncertainty for objects such as mapped boundaries must be defined for some class of cases to which that model applies. The question underlying this view from the panel is how broad that class can be if the results are to be useful and practically meaningful. Further work is needed to compare elicited error distributions for geological boundaries in more or less narrowly defined sets of cases.

It was also thought that elicitation should include the field observation of settings of the problem. As the expert opinion on the valid application of the elicited tacit expert model this opinion must be considered carefully. However, it is also important to pay attention to the psychological research on the judgement heuristics which affect people's assessments of uncertain outcomes (O'Hagan et al., 2006). In particular the consideration of very specific settings, and even more so, of a necessarily limited number of field settings may serve to "anchor" expert judgement of particular statistics near values consistent with particular interpretations of a few boundaries and their field settings. It may also limit the range of possible conditions consistent with the elicited problem which the participants consider during the elicitation (accessibility judgement heuristic), which would result in elicited distributions which are too narrow. One might consider the possibility of considering substantial numbers of field locations in virtual field work in a 3-D visualization suite.

It is interesting and encouraging that the group of geologists, with experience in varied settings, were able to agree on consensus distributions for five out of six settings, the exception being a scenario in which two superficial units were mapped. In the elicitation one could see both the influence of individuals (e.g. expert E in scenario 4 who convinced the group that the distribution should be asymmetric), and the way in which initially contrasting views converged during discussion. The process does not necessarily entail convergence to what was initially a majority opinion, nor to some

linear pool of these opinions. In a complex problem such as this the process of discussion to agree a consensus may be more robust than attempts to weight contrasting individual distributions numerically.

At the same time the process of elicitation was not dominated by single voices. While E influenced the group significantly on scenario 4, the consensus was somewhat different from his original individual distribution. This shows how the structured discussion in the elicitation procedure can help with convergence to a consensus which reflects the variation of individual experience within the group. The fact that some experts had more experience in particular settings than did others was explicitly recognized in discussion. Nonetheless, the consensus distribution did not correspond directly to the initial individual distribution of the influential individual, but represented the outcome of a discussion in which all panel members contributed and modified their original positions to varying degrees. This is consistent with the findings of Polson and Curtis (2010) who distinguished what they called group “herding” to a dominant view from the formation of a genuine group consensus.

The one scenario in which a consensus was not achieved was a boundary between two contrasting superficial deposits. On reflection the group agreed that, in this case, the geometry of the contact represented by a boundary was harder to visualize than in the other cases with at least one solid geological unit. This may indicate that the approach is less applicable to superficial material, or that the scenario needs more careful description, perhaps with some visual examples. However, it is recognized in the literature on elicitation (e.g. Polson and Curtis, 2010; O’Hagan et al., 2006) that the pursuit of consensus “at all costs” can distort an elicitation, and alternative outcomes reflecting different opinions may be preferable where a consensus does not emerge.

We recognize that our expert panel was drawn from a narrow pool, in that all were BGS geologists. This was inevitable because, as discussed above, our objective is to elicit a tacit model of uncertainty from experts with close knowledge and experience of the procedures and protocols by which the line work in BGS mapping was obtained. There may be disadvantages in using a panel who know each other if this could impede robust discussion, although no panel member, when asked individually, thought this likely. Familiarity also had the advantage that panel members knew, in any scenario, which of their colleagues had the most pertinent experience. As observed by Reagan-Cirincione (1994) such awareness is important in the success of behavioural aggregation. The level of awareness of colleagues’ experience could be quite subtle – for example in scenario 5 in the course of discussion one panel member pointed out that a colleague’s argument, while based on field experience, was strongly influenced by experience in lower latitudes than the United Kingdom. That said, it is certainly possible that the elicited tacit model for BGS geologists is overoptimistic about the uncertainty of BGS line work. It would therefore be useful further research

to find a case study where new geophysical measurements allow the identification of a boundary belonging to one of these scenarios where it has been surveyed in the field. This would allow us to compare the elicited error distribution with an empirically estimated one. However, it would be difficult to do this convincingly in all scenarios, since independent identification of the position of a boundary on a test transect will not always be readily done from geophysical data.

The elicited tacit uncertainty model of BGS geologists may be useful for assessing uncertainty in BGS line work. It may also be useful as a way to capture expert understanding, not recordable in other forms such as survey reports or memoirs, particularly as field surveyors retire and, increasingly, are replaced at a declining rate. This may also be true for other disciplines in environmental science facing a numerical decline of field-experienced surveyors, such as soil survey (e.g. Anderson and Smith, 2011). However, the aim of this elicitation was rather narrowly focused on models of uncertainty of line work position. Elicitation specifically to extract, quantify and archive experienced geologists’ tacit knowledge should have a broader setting, based on a fuller consideration of the workflow of field mapping.

It is notable that there was considerable variation in the time taken for elicitation of each scenario. Not surprisingly, the first scenario took considerable time. In part this was because of complexities in the scenario itself, but it also reflects the time needed for familiarization with the process and associated concepts despite the briefing meeting and practice elicitation. Given this, there may be advantages in including a practice elicitation closer to the target problem. For example, in this case we might have undertaken a practice elicitation on the error distribution for a mapped fault. In each scenario the longest single component of the elicitation was the initial group discussion to agree limits for the boundary error. It was during this discussion that the group identified sources of uncertainty in the delineation of boundaries in the particular scenario.

The presence of the geological facilitator during the elicitation was important. The facilitator was able to make controlled changes to the scenarios during the elicitation, in particular adding elements to the description of the supporting field observations when participants raised queries, or adjusting elements of the initial description if participants thought these atypical. The statistical facilitator was also required, not just to operate the software but also to advise on questions such as interpretation of asymmetry in the distributions and to identify emerging confusions. For example, on occasions panel members needed to be reminded that they were considering a notional transect examined exhaustively to test an already-mapped boundary, not a traverse being examined according to normal procedures in order to map a boundary by interpretation. The meaning of errors of different sign required careful attention, and one topic for further work is whether it is better to consider the mapped boundary as fixed

and the true boundary as variable (as here) or to fix the true boundary.

For purposes of this elicitation we considered what is effectively a 1-D model for boundary errors, specifically in terms of the intersection of boundaries with a notional transect. Further work is needed to make this approach fully applicable to the error of delineations of geological units as 2-D objects on a map. The first issue is the uniformity of the error distribution along a boundary. Important details of the setting (such as the presence of exposures, or variations in land use) may vary along a single boundary. This increases the number of scenarios for which elicitation is required, and raises practical difficulties for how a distribution is selected for a particular problem from a set of available ones. Nonetheless, elicitation for many settings is more practically feasible than the empirical assessment of mapped boundaries. The more fundamental problem is how to treat the entire boundary of a polygon as an uncertain object. One possible approach would be based on the stochastic model for positional uncertainty in deformable objects proposed by Heuvelink et al. (2007).

## 5 Conclusions

In conclusion, expert elicitation using the SHELF-based methodology provides a method to extract the tacit model which geologists use when interpreting line work. In particular, the SHELF approach, based on a combination of individual and group elicitation, allowed our group to reach a consensus in five out of six scenarios. In several cases the final outcome was not the same as any one expert's initial distribution, indicating how the procedure allows us to arrive at a consensus through structured discussion. The elicitation process is most suitable for scenarios where the geometry of the contact represented by the boundary can be visualized by the experts. In our experience this precludes boundaries between superficial deposits.

Further work is needed to develop this approach. In particular we need to examine just how general a scenario can be used to elicit uncertainty models which are usable for the interpretation of specific boundaries. This could be examined by elicitation for scenarios of comparable generality to those reported here, and nested cases within each scenario which are more narrowly defined either in terms of lithology or specific units, or particular map sheets in which the target boundaries appear. The panel felt that clearer visualization of the scenario, ideally in the field, would help. It would be interesting to explore how far this can be achieved given the need to avoid "anchoring" and to ensure that the expert panel accesses a sufficiently wide distribution of cases for any scenario. This might be achieved by visualization in 3-D virtual reality using digital terrain models with overlaid air photography or satellite imagery. Associated validation of the elicited error model by geophysical inference of the location of a boundary at test locations would also be useful.

In addition to these general conclusions, we have drawn some practical conclusions for the use of elicitation. First, the use of a structured and transparent process is essential. The SHELF framework ensures that there is a combination of individual thought and group discussion. In this trial the procedure ensured that ideas were pooled and that individual voices were heard but not allowed to dominate. Our experience showed that some general issues in the elicitation may arise only when specific examples are being tackled (hence the long general discussion which took place during the elicitation for the first scenario). This is probably inevitable, but it may be good practice to use a practice elicitation which is closer to the main target elicitation in character. Both the statistical and geological facilitator were essential to the process, as were figures to keep the disposition of units in front of the panel at all times. Finally, many of the key issues in the understanding of boundary error in any scenario emerged in the initial discussion on the feasible range of error values. Sufficient time must therefore be allowed for this part of the discussion.

## Appendix

Here we detail the scenario descriptions as provided to participants. The uppercase codes in brackets indicate a range of possible lithologies for the units according to the BGS Rock Classification Scheme (Cooper et al., 2006). For some scenarios additional qualifications agreed during the elicitation are included, and are indicated as such.

### Scenario 1 – edge of river terrace deposit on bedrock

Back/uphill limit of a Quaternary river terrace deposit composed of brown sand and gravel (SV/XSV/SVZ/VS/XVS) resting on/against rock head on a bedrock unit of contrasting lithology (Mesozoic/Cenozoic sedimentary rocks).

- Grassland, short cropped grass, scattered molehills.
- Traverse parallel to clear but not freshly dug/cleaned ditch 1 m deep.
- Field is flat, with very subtle concave change/break-of-slope across the slope to gentle ( $< 2^\circ$ ) upward slope over 50 m distance.
- Field work included Dutch augering every 15 m.
- *During the elicitation it was agreed that no significant anthropogenic modification would be present at any instance of this scenario.*

### Scenario 2 – base of sandstone in mudstone/siltstone succession

Sharp or rapid-passage base of well-cemented grey-brown medium-grained silicate sandstone (SDST) bed 5 m thick within well-indurated grey mudstone  $\pm$  siltstone succession (Palaeozoic – e.g. Coal Measures/Millstone Grit).

- Urban street, approximately at right angles to strike.
- Beds dipping into slope.
- Sight of soil in 50 % of gardens.
- Quarry in the sandstone bed about 200 m away to one side exposes base, measurable dip, correctly shown on map with good contour information. No evidence of faulting in intervening ground.
- Street slopes up at about  $4^\circ$ , with subtle concave change/break-of-slope across the slope to steeper ( $7^\circ$ ) upward slope over a distance of 30 m.

### Scenario 3 – edge of alluvium/tidal river deposit against contrasting underlying geology

Lateral limit of Holocene/modern alluvium/tidal river deposits composed of dark brown clay and silt (CZ/XCZ) resting on/against any contrasting superficial deposit or bedrock lithology.

- Arable field, bare or short crops, soil easily visible.
- Field is flat, with conspicuous concave break-of-slope across the slope to moderate ( $< 5^\circ$ ) upward slope over 5 m distance.

### Scenario 4 – Stratigraphic boundary between two distinctive sedimentary rocks

Stratigraphic boundary between two distinctive (by colour, grain-size, grain type and weathering habit) limestone (LMST/LMOOL/SALMST), or two sandstone (SDST/CALSST) or two chalk (CHLK) lithologies (Mesozoic sedimentary rocks).

- Arable field, bare or short crops
- Soil easily visible with sparse to dense brash of dirty angular pieces, some of which can be inferred to have been derived (by ploughing/cryoturbation etc.) from underlying bedrock, easily broken by hammer. A fair scattering of other stones – e.g. pebbles from nearby superficial deposits, brash of other local bedrock units, possible exotics (may be natural anthropogenic substances, concrete etc.)
- Field is sloping  $2^\circ$  to northwest, but gently undulating with no clear linear features.
- Regional dip is about  $2^\circ$  to southeast.
- Three small quarries within 200 m radius show tabular beds with dips of  $0, 3^\circ$  to  $090$  and  $5^\circ$  to  $160$ .
- *During the elicitation it was agreed that that the mapper does know the superposition relationship between the units.*

### Scenario 5 – faulted boundary between granite and hard non-igneous rock

Fault between large granite body and well-indurated sedimentary or metasedimentary rock succession. Assume fault is high angle and there is a single plane of displacement.

- Moorland, long grass, heather etc. soil not generally visible, scattered large rock exposures spaced about 50 m apart – some may be ex situ. May detour up to 50 m to side.
- Uneven ground, some declivities, may form some alignments in various directions.

**Scenario 6 – boundary between two distinctive tills, unknown relationship**

Two juxtaposed Pleistocene tills of unknown superpositional relationship, with contrasting matrix colour/character – brown and smooth vs. grey and silty, and contrasting clast content – chalk, flint, quartz and quartzite pebbles and common igneous erratics vs. chalk, flint, underlying bedrock of mudstone and rare oyster fossils, very rare erratics.

- Arable field, bare or short crops.
- Soil easily visible with sparse to dense scattering of till clasts and ploughed-up subsoil (weathered till clay matrix). A fair scattering of other stones – pebbles from nearby superficial deposits and probable anthropogenically introduced stones.
- Field is flat with no linear features.
- Field work included Dutch augering every 10 m.
- *During the elicitation it was agreed that* interpretation would be based on the assumption that augering traverses are 250 m apart. The notional transect for the elicitation crosses the boundary at a random location so does not necessarily coincide with a traverse.

*Acknowledgements.* Robert Cooper administered this project, coordinating availability of participants for meetings. This paper is published with the permission of the executive director of the British Geological Survey (NERC).

Edited by: K. Zeigler

## References

- Albrecht, F., Lang, S., and Höbbling, D.: Spatial accuracy assessment of object boundaries for object-based image analysis, in: Proceedings of GEOBIA 2010-Geographic Object-Based Image Analysis, edited by: Addink, E. A., and van Coillie, F. M. B., vol. XXXVIII-4/C7, Ghent University, Ghent, Belgium, 2010.
- Anderson, D. W. and Smith C. A. S.: A history of soil classification and soil survey in Canada: Personal perspectives, 91, 675–694, 2011.
- Bond, C. E., Gibbs, A., Shipton, Z. K. and Jones, S.: What do you think this is? “conceptual uncertainty” in geoscience interpretation, *GSA Today*, 17, 4–10, 2007.
- Bond, C. E., Lunn, R. J., Shipton, Z. K., and Lunn, A. D.: What makes an expert effective at interpreting seismic images?, *Geology*, 40, 75–78, 2012.
- Brodaric, B., Gahegan, M., and Harrap, R.: The art and science of mapping: computing geological categories from field data, *Comput. Geosci.*, 30, 719–740, 2004.
- Burrough, P. A.: Multiscale sources of spatial variation in soil, I: The application of fractal concepts to nested levels of soil variation, *J. Soil Sci.*, 34, 577–597, 1983.
- Burrough, P. A., Beckett, P. H. T., and Jarvis, M. G.: The relation between cost and utility in soil survey, *J. Soil Sci.*, 22, 359–394, 1971.
- Chrisman, N.: Theory of cartographic error and its measurement in digital data basesm, *Proceedings of AUTO-CARTO*, 5, 159–168, 1982.
- Cooke, R. M.: *Experts in uncertainty: opinion and subjective probability in science*, Oxford University Press, Oxford, 1991.
- Cooper, A. H., Kessler, H., and Ford, J.: A revised scheme for coding unlithified deposits (also applicable to engineering soils), British Geological Survey Internal Report IR/05/123, British Geological Survey, Nottingham, 2006.
- Gong, P., Zheng, X., and Chen, J.: Boundary errors on digitized maps: an experiment on digitization errors, *Lect. Notes Comput. Sc.*, 1, 65–72, 1995.
- Guillot, G., Kan-King-yu, D., Michelin, J., and Huet, P.: Inference of a hidden spatial tessellation from multivariate data: application to the delineation of homogeneous regions in an agricultural field, *J. Roy. Stat. Soc. C-App.*, 55, 407–430, 2006.
- Heuvelink, G. B. M., Brown, J. D., and Van Loon, E. E.: A probabilistic framework for representing and simulating uncertain environmental variables, *Int. J. Geogr. Inf. Sci.*, 21, 497–513, 2007.
- Higgins, H. M., Dryden, I. L., and Green, M. J.: A Bayesian elicitation of veterinary beliefs regarding systemic dry cow therapy: Variation and importance for clinical trial design, *Prev. Vet. Med.*, 106, 87–96, 2012.
- Lark, R. M. and Beckett, P. H. T.: A geostatistical descriptor of the spatial distribution of soil classes, and its use in predicting the purity of possible soil map units, *Geoderma*, 83, 243–267, 1998.
- Lark, R. M., Mathers, S. J., Thorpe, S., Arkley, S. L. B., Morgan, D. J., and Lawrence, D. J. D.: A statistical assessment of the uncertainty in a 3-D geological framework model, *P. Geologist. Assoc.*, 124, 946–958, 2013.
- Lark, R. M., Thorpe, S., Kessler, H., and Mathers, S. J.: Interpretative modelling of a geological cross section from boreholes: sources of uncertainty and their quantification, *Solid Earth*, 5, 1189–1203, doi:10.5194/se-5-1189-2014, 2014.
- Lee, L. A., Pringle, K. J., Reddington, C. L., Mann, G. W., Stier, P., Spracklen, D. V., Pierce, J. R., and Carslaw, K. S.: The magnitude and causes of uncertainty in global model simulations of cloud condensation nuclei, *Atmos. Chem. Phys.*, 13, 8879–8914, doi:10.5194/acp-13-8879-2013, 2013.
- Leenhardt, D., Voltz, M., Bornand, M., and Webster, R.: Evaluating soil maps for prediction of soil water properties, *Eur. J. Soil Sci.*, 45, 293–301, 1994.
- Lillah, M. and Boisvert, J. B.: Stochastic distance based geological boundary modeling with curvilinear features, *Math. Geosci.*, 45, 651–665, 2013.
- Mandelbrot, B. B.: How long is the coast of Britain? Statistical self-similarity and fractional dimension, *Science*, 156, 636–638, 1967.
- Martí, J., Aspinall, W. P., Sobradello, R., Felpeto, A., Geyer, A., Ortiz, R., Baxter, P., Cole, P., Pacheco, J., Blanco, M. J., and Lopez, C.: A long-term volcanic hazard event tree for Teide-Pico Viejo stratovolcanoes (Tenerife, Canary Islands), *J. Volcanol. Geoth. Res.*, 178, 543–552, 2008.
- Oakley, J. E. and O’Hagan, A.: SHELF: the Sheffield Elicitation Framework (version 2.0), School of Mathematics and Statistics, University of Sheffield, UK, available at: <http://tonyohagan.co.uk/shelf>, 2010.
- O’Hagan, A., Buck, C. E., Daneshkhan, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., and Rakow, T.: *Uncertain Judgements: Eliciting Experts’ Probabilities*, John Wiley and Sons, Chichester, 2006.
- Polson, D. and Curtis, A.: Dynamics of uncertainty in geological interpretation, *J. Geol. Soc.*, London, 167, 5–10, 2010.
- R Development Core Team: *R: a language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, available at: <http://www.R-project.org/>, 2013.
- Reagan-Cirincione, P.: Improving the accuracy of group judgement: a process intervention combining group facilitation, social judgement analysis and information technology, *Organizational Behavior and Human Decision processes*, 58, 246–270, 1994.
- Ren, S. and Oakley, J. E.: Assurance calculations for planning clinical trials with time-to-event outcomes, *Stat. Med.*, 33, 31–45, 2014.
- Scholten, L., Scheidegger, A., Reichert, P., and Maurer, M.: Combining expert knowledge and local data for improved service life modeling of water supply networks, *Environ. Modell. Softw.*, 42, 1–16, 2013.
- Scourse, E., Aspinall, W. P., and Chapman, N.: Using expert elicitation to characterise long-term tectonic risks to radioactive waste repositories in Japan, *J. Risk Res.*, 18, 364–377, 2015.
- Silan-Cárdenas, J. L., Wang, L., and Zhan, F. B.: Representing geographical objects with scale-induced indeterminate boundaries: a neural network-based data model, *Int. J. Geogr. Inf. Sci.*, 23, 295–318, 2009.

Truong, P. N., Heuvelink, G. B. M., and Gosling, J. P. Web-based tool for expert elicitation of the variogram, *Comput. Geosci.*, 51, 390–399, 2013.

Usher, W. and Strachan, N.: An expert elicitation of climate, energy and economic uncertainties, *Energ. Policy*, 61, 811–821, 2013.

Webster, R. and Beckett, P. H. T.: Quality and usefulness of soil maps, *Nature*, 219, 680–682, 1968.