Solid Earth Discuss., 7, C542–C549, 2015
www.solid-earth-discuss.net/7/C542/2015/

Solid Earth
Discussions
Open Access

Interactive comment on "Eliciting geologists' tacit model of the uncertainty of mapped geological boundaries" *by* R. M. Lark et al.

R. M. Lark et al.

mlark@nerc.ac.uk

Received and published: 29 April 2015

Reviewer 4

General comments

We are grateful for this reviewer's comments. They stimulate us to clarify the motivation for this study, and its precise objectives.

Our objective is *not* to 'extract the tacit understanding ... as to how to best draw lines on geological maps.' If it were then we would fully agree with this reviewer (and with reviewer 3, paragraph 3). *Inter alia* it would not be appropriate to use a group of surveyors from a single organization. Our position, which motivates this study, is as

follows.

i. We have a large body of linework at BGS, produced by BGS surveyors, and this underpins many of our products and outputs.

ii. We require, for these products and maps, some measure of the uncertainty attendant on the extant linework so it can be accounted for when it is used. As noted in the current paper, buffering is implemented to allow for cartographic uncertainty, so one approach would be to modify the buffer to incorporate the interpretation uncertainty. This is why the uncertainty expressed as a deviation of the true position of the boundary on a line perpendicular to the mapped boundary is pertinent. This is also why the title of the paper is '...uncertainty of *mapped* boundaries..' i.e. the position of the boundary as mapped by a BGS surveyor is a given, the question is where should it really be.

iii. Furthermore, the 'institutional memory' of field surveyors familiar with BGS procedures, is being lost through retirement of field surveyors at a faster rate than their replacement. This is a further motivation for aiming to elicit the tacit model of uncertainty that this particular group hold, without claiming that this is the only approach that should or could be taken.

Given this position, we want to elicit the tacit model of uncertainty of the BGS linework in the mind of experienced BGS surveyors, i.e. geologists who possess at least part of the institutional memory of the organization, and experience of its procedures and conventions. This awareness was important, for example, in scenario 1 (section 3.1, p 163 l 6, 'In practice the mapping...', and scenario 4 (section 3.4 , p 165 l 15 et seq.) where BGS procedures on mapping drift and practice on mapping brash was, at least considered, as relevant to the problem. Geologists from other survey organizations, or without experience of mapping themselves, would not have access to this understanding.

We acknowledge that a model of uncertainty elicited from geologists may be unduly optimistic, and state in the paper that some empirical validation would be desirable (page 168, l 20–23). However, for reasons given (page 153, l 17–26) it is unlikley that this could be done on anything but a very limited scale. These reservations granted, it seems to us that what one might call the 'BGS tacit model' of linework uncertainty: the mental model developed by the surveyors with experience of interpreting the range of evidence listed in the scenarios to draw boundaries, is a valid target for elicitation. It is not clear what a model elicited from a group of geologists with varied (and in some case no direct) experience of the organization's surveying approaches would mean. It is certainly not clear that it would provide an operationally meaningful basis for attaching buffers to published linework.

In revising the paper we propose the following:—

i. To change the title, e.g. to 'Eliciting the tacit model of uncertainty in mapped geological boundaries held by geologists from a national geological survey'.

ii. To clarify the motivation and scope of the study in line with the three points above and the paragraphs above beginning 'Given this position,...' and 'We acknowledge that a model...'.

iii. To emphasize the point already in the discussion (page 168, l 20–23) that some field-based study, perhaps using geophysical methods, might be used to estimate distributions of boundary error and compare these with the elicitation results to identify possible bias. It is necessary to note that this would not be possible over all scenarios.

The record of interests and expertise to which the reviewer refers is not circulated to the panel in the SHELF procedure but is retained as meta-data. Since the outside financial interests of BGS staff with any bearing on their work are a matter of record,

and the survey memoirs, reports and projects undertaken by each staff member are also known, this was not thought necessary, and it has no bearing on the conduct of the elicitation itself.

Specific responses to numbered comments

1. See comments above with respect to the scope of the study. Oakley and O'Hagan (2010) specifically state that there is no particular merit in a 'large sample' for elicitation, and the SHELF framework specifies a panel of five (as we had) as the ideal size since it ensures that effective discussion within the group is possible. As noted above, our objective is to elicit a tacit mental model of the uncertainty in existing linework on BGS maps, as perceived by surveyors, reflecting the uncertainty which, as geological surveyors, they recognize is inherent in the process of interpretation that they undertake in order to map boundaries. To achieve this we require geologists with substantial experience of doing geological survey according to BGS procedures and protocols, and in undertaking survey in contrasting settings so that they do not duplicate each other's expertise; see page 157, l 2–3.

   We do not accept that the group would inevitably not be able to engage in robust discussion because they were known to each other. In fact, it is clear from section 3.1 that this group was able to engage in such discussion. See, for example, scenario 1 (page 163, l 23 et seq.), scenario 4 (page 165 lines 11–20), scenario 5 (page 165 line 24 et seq.) and scenario 6 (which did not result in a consensus).

2. A scenario is not a simplification, it is a generalization, the panel being specifically invited to 'access', as the elicitation literature puts it, their own experience of the range of actual real cases which fall under the necessarily general description. This is true of any generalization of objects into classes, such as lithological units. A geologist knows what a 'mudstone' is, and the general term is useful without naively assuming that all rocks which can correctly be described as mudstone

are uniform. This is the essential idea in elictation of probability distributions, as noted by O'Hagan et al (2006) (see page 158 line 9, et seq. of the paper). We want the panel members mentally to access the range of conditions compatible with a generalized description consistent with their model of geological variability. A probability distribution is precisely what we require in order to make some generalization about uncertainty of linework without requiring a separate elicitation for every mapped boundary. This is illustrated particularly clearly at page 165, line 15. The panel recognized that linework in settings corresponding to scenario 4 may have been drawn with reference to observed surface brash in some instances of the setting and not in others. Similarly in scenario 1 (page 163, l 25) one panel member argued that, in some cases belonging to this scenario, the surveyor might be misled by isolated patches of terrace material, but that this does not occur every time. This variation between different instances which contributes to the breadth of the elicited distributions.

In revising the paper we will make some comments on the relation of individual cases to the scenarios to which they belong, specifically near page 158 lines 2–14 where the notion of 100 individual instances encompassing the variability of circumstances consistent with the scenario distribution is introduced.

3. This comment is closely linked to the preceding one. The objective of this study is not to extract some generalized tacit knowledge about the process of geological survey (interesting as that might be) but rather to elicit a very specific kind of statistical model reflecting uncertainty about where a contact between geological units can be expected to fall relative to a boundary mapped according to the procedures of the survey organization. That is the model which we will ultimately require in order to make practically useful statements about linework uncertainty in addition to the buffering already undertaken to reflect cartographic uncertainty.

4. This comment seems to be made without reference to the discussion in the elictation literature on the group elicitation approach used in SHELF (e.g. O'Hagan

et al, 2006) and its basis in 'behavioural aggregation'. O'Hagan et al (2006) note the existence of methods, such as Cooke's method, which use calibrating or 'seeding' variables to assign different weights to measure the degree of expertise of different panel members and to weight the distributions elicited from them to form the final elicitation output. The seeding variables are cases where the true value is known, and so can be compared with each expert's opinion. O'Hagan et al (2006) note that the testing of such weighting schemes has been somewhat limited for problems such as ours (probability distributions), and they have not consistently outperformed simple averaging in other elicitation problems— see section 9.2.4 on page 185 of O'Hagan et al. (2006).

Even if we wished to use weighting approaches we would face substantial difficulties. The first, already noted in the paper, is that it is far from trivial to obtain the sort of seeding information that would be required. The costs and challenges of coming up with direct verifications of mapped linework are precisely the motivation for our use of elicitation. Furthermore, not all experts are expected to be equally familiar with all scenarios, since they have varying experience in different geological settings. This means that different sets of seeding variables would be needed for each scenario.

This is why we chose to use the methods of 'behavioural aggregation' that form the basis of the SHELF procedure (Oakley and O'Hagan, 2010). O'Hagan et al. (2006) discuss this at length (section 9.3, page 186 et seq. of the book). The goal of behavioural aggregation is to arrive at an elicitation from a group working together. In the group elicitation each panel member starts with a personal position, derived by independent thought, and then participates in a general discussion to arrive at a consensus if that is possible. The SHELF procedure is designed to avoid well-known psychological issues in individual and group perception of uncertainty. For example, as Oakley and O'Hagan (2010) discuss, the sequence of discussion from limits to median to quartiles is designed to reduce

the effects of anchoring. As O'Hagan et al. (2006) state (page 187):

> ... the approach of bringing the experts together and eliciting a single distribution from the group as a whole is attractive. It should avoid the choice of a method of mathematical aggregation which, no matter how well argued each method may be, necessarily embodies an element of subjectivity and arbitrariness on the part of the decision maker. It should also have the merit of pooling knowledge and allowing the experts to bring their combined expertise to bear on the questions. In principle, this synthesis should lead to more informed 'aggregated' distributions than the mechanistic approaches of mathematical aggregation.'

We reiterate that this is how the SHELF procedure works, we have followed the SHELF procedure in all respects that have a bearing on the conduct of the elicitation itself, only deviating in terms of the approach to collecting metadata on expert experience and potential conflicts of interest.

As we note at several points in section 3, it happened on several occasions that a single panel member was able to convince the group that the elicited distribution should have particular features. This was based on specific technical arguments. We noted (page 168 l 3 et seq.) that

> 'The process does not necessarily entail convergence to a 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.'

We think that the descriptions in section 3 give the flavour of how these discussions proceded, and show that they comprise neither horse-trading to achieve an

uninformative 'average view', nor the domination by one individual. As we noted in the original paper, (page 168, lines 10–13) the fact that certain panel members were more experienced in particular scenarios than were others was explicitly recognized in the discussion. The outputs therefore do not reflect an assumption that all expert's opinions on any one scenario are of equal merit.

In revising the paper we will discuss the reasons for choosing behavioural aggregation in the SHELF framework overagainst alternatives (in section 2.1, near page 155).

Interactive comment on Solid Earth Discuss., 7, 147, 2015.