uncertainty-enabled model web

Proposal / Contract no.: 248488
Project start: February 1, 2010
Project end: January 31, 2013

UncertWeb

The *Uncert*ainty Enabled Model *Web*

SEVENTH FRAMEWORK PROGRAMME

THEME FP7-ICT-2009-4

ICT for Environmental Services and Climate Change Adaptation

Deliverable 3.1

Expert elicitation of input and model uncertainties

Title of Deliverable	UncertWeb Web site, mailing list, forums and wiki
Deliverable reference number	UncertWeb D3.1
Related WP and Tasks	WP3, Task 3.1
Type of Document	Public
Authors	Gerard Heuvelink, Lucy Bastin, Dan Cornford, John Paul Gosling, Phuong Truong, Matt Williams
Date	7/7/2011
Version	1.0

Project coordinator

Dr. Dan Cornford

Aston University, United Kingdom

E-mail: d.cornford@aston.ac.uk

http://www.uncertweb.org

Revision History

Version	Date	Changes	Authors
1.0	07/7/2011	Created document	Gerard Heuvelink
1.1	22/7/2011	Chapter 4	Phuong Truong
1.2	22/8/2011	Chapter 5	Gerard Heuvelink
1.3	26/8/2011	Chapters 3 and 4 added	Dan Cornford
1.4	07/09/2011	Formatting corrected	C. Doering-Saad
1.5	09/09/2011	Revision based on comments Harry	Gerard Heuvelink
1.6	13/09/2011	Error corrections by all	Gerard Heuvelink

Legal Notices

Executive Summary

This report describes the tool that was developed within UncertWeb to assess uncertainties in inputs and parameters of the web model chain through the elicitation of experts.

Expert elicitation is a multi-disciplinary process that can inform decision-making by characterising uncertainty about data and models. It must be performed using appropriate methods and quality standards, including peer review and transparency, for which the literature provides many examples. This report describes how the most appropriate approaches in the context of UncertWeb were selected from the literature and how these were implemented in a web service that guides the experts through the process of specifying their uncertainty. The tool is implemented as an UncertWeb service, with a thin client interface that produces uncertainty specifications in full compatibility with the requirements and standards of the UncertML encoding.

The elicitation tool distinguishes between uncertainty assessment of non-spatial and spatial variables. The latter makes use of the former but also employs additional tailored approaches. The 'Elicitator' tool that is developed for non-spatial variables can handle both continuous and categorical variables and can be accessed at http://elicitator.uncertweb.org/. Problem owners define an elicitation problem and invite experts to participate, who subsequently login to the website and complete a list of questions that make up the elicitation process. This results in the quantification of the uncertainty about the continuous or categorical variable by means of a probability distribution. The spatial case is restricted to continuous variables, whose spatial correlation function is elicited from experts in addition to the (marginal) probability distribution. The tool can be accessed at http://www.variogramelicitation.org/. Once this tool has been fully tested and approved, it will be merged with the Elicitator.

Contents

1 Introduction

This report describes the expert elicitation process and the associated tool that was developed within UncertWeb to assess uncertainties in inputs and parameters of the model web.

We regard elicitation as being the process of translating someone's beliefs about some quantities of interest into probability distributions. Typically, this is achieved by an expert being asked for summaries of their distributions and then an appropriate parametric distribution is fitted to those judgements. Elicitation is an important process: it has a part to play in every application where data and observations do not provide enough information by themselves. Elicitation cannot replace data-driven scientific studies, but it can help us to make rigorous use of available knowledge when time or resources to gather new data are limited, and to take stock of uncertainty about the quantities of interest. Although elicitation is far from being a precise science, elicitation methodologies capture experts' current knowledge in coherent (and, hopefully, transparent) ways. It can be difficult for experts to articulate their beliefs. There are also complications due to the biases of the individual experts and the biases created by the questioning process.

There are several roles in the elicitation exercise. The **problem owner** is the person who desires to quantify the likely values of the quantity of interest. The **experts** are the people who are likely to have informed beliefs about the quantity of interest and the **facilitator** is the person or mechanism that permits the problem owner to interact appropriately with the experts to obtain the desired information about the quantities of interest.

An elicitation exercise is not just about asking an expert a number of questions about the quantities of interest and fitting distributions. There are several stages the facilitator of the elicitation exercise should go through:

- problem set-up and training,
- eliciting beliefs about the quantities of interest,
- fitting of an appropriate distribution,
- feedback of implications of fitted distribution,
- revision of judgements.

The final three steps should be repeated until the expert(s) are happy that the fitted distribution reflects their beliefs about the quantity of interest.

There are several resources available over the internet that are designed to help capture expert knowledge and encode it in a probabilistic way. James et al. (2010) describes the Elicitator software that is used for eliciting judgements about prior structures in regression models, although this is not easily available on the web. The Elicitator is built for applications in ecological modelling and uses maps to give experts instantaneous feedback about the consequences of their judgements on the model in question. We have used the name "The Elicitator" for our software as the software described in James et al. (2010) is not found by a web search, and this name reflects the functionality of our implementation. The SHeffield ELicitation Framework (SHELF) is a package of documents, templates and software to carry out elicitation of probability distributions for uncertain quantities from a group of experts. SHELF advocates the elicitation method that is described in O'Hagan (1998). EXCALIBUR is a stand-alone piece of software that implements Cooke's classical method for elicitation and opinion pooling (see later for a description of this).

The first stage consists of setting up the exercise: the selection and training of experts, and the identification (and definition) of parameters to elicit judgements about. The training of the experts is carried out to help familiarise the experts with statistical procedures that may be unknown to them. As pointed out by Kadane and Wolfson (1998), being an expert in your area does not make you an expert in statistics and giving probability judgements. In The Elicitator we face additional issues since it is necessary to communicate with the experts remotely, where traditional methods are based on workshops where all experts are physically present in the same location. There is significant benefit in being more flexible in terms of the demands on the experts' time, as provided by The Elicitator. The web based nature of the solution means experts can access the system at their convenience.

Once the expert has read the briefing document, which explains the problem setup and the issues that might be encountered in elicitation, they move onto the second stage that is the actual elicitation of judgements from the expert. It is assumed that the expert cannot state their distribution explicitly. They can only state certain summaries of their distribution such as the mean or various percentiles. These elicited summaries do not identify the expert's distribution uniquely. Most of the applications of elicitation employ parametric techniques. In a parametric technique, the facilitator fits the elicited summaries to a distribution that is a member of some specified parametric family. The resulting distribution or summaries from that distribution are then presented to the expert to verify that the proposed distribution actually fits their beliefs. It is conceivable that any number of distributions would be accepted by the expert as their distribution. This is the satisficing prior distribution of Winkler (1967): the expert will be content to adopt that prior distribution at that moment in time. As the expert cannot differentiate between well-fitted distributions, there are an infinite number of distributions that the expert would accept as their own distribution.

The questioning stage of the process has been the focus of a lot of work in the psychological literature. As the Elicitator acts as the facilitator of the elicitation process, we must be aware of some of the caveats given in cognitive psychology literature. Tversky and Kahnemann (1974) listed four aspects of making judgements that facilitators of an elicitation exercise should keep in mind: representativeness, availability, adjustment and anchoring. A fifth effect that Kadane and Wolfson (1998) considered was hindsight bias: if the expert had already seen some data about a parameter, then judgements made about the parameter may be based on these data. Kadane and Wolfson (1998) and O'Hagan et al. (2006) give guidance on precautions that can be taken to avoid these sources of bias.

The third stage of the elicitation process is to fit a probability distribution to the judgements given in the previous stage. The summaries elicited from the expert were traditionally selected to make the fitting of the selected distribution easier. However, increased computing power has made this much less of a restriction.

Finally, the results of the fitting should be presented to the expert so that they have the final say on whether the results fit their beliefs well enough; this is called the feedback stage. This is where the assumption of the fitted distribution being a valid representation of the expert's beliefs is judged. If the expert does not agree with the results, then they can return to the second stage and more information can be elicited from them or a different distribution can be fitted. The structure of the elicitation process is considered in detail in Garthwaite et al. (2005) where the importance of feedback is stressed.

Feedback of the results is important to confirm that the fitted distributions are representative of the experts' beliefs. In the Elicitator, we feedback graphical representations of the judgements in terms of fitted probability distributions for continuous variables and other distribution summaries, which have not been asked for from the expert. Of course, after the results have been displayed to the expert, they are free to alter their judgements as often as they wish until they are completely happy with the implications of their judgements. Further details are given in subsequent chapters.

This report is organised as follows. Chapter 2 provides background information about the theory behind The Elicitator and describes some technical implementation aspects. Chapter 3 presents a tutorial to The Elicitator that guides the reader through the various steps needed to complete an elicitation process. Readers are expected to start the tool at http://elicitator.uncertweb.org/ and work their way through the tutorial and tool simultaneously. An example elicitation problem is provided as a case study. Chapter 4 provides background documentation for the tool developed for the elicitation of spatial variables. The tool itself is presented in Chapter 5. Again a tutorial is provided that the reader is expected to complete while running the tool at http://www.variogramelicitation.org/.

2 Background material for expert elicitation of continuous and categorical variables

2.1 Conducting expert elicitation

As explained in the Introduction, we regard elicitation as the process of translating someone's beliefs about some quantities of interest into probability distributions. An elicitation exercise involves several stages:

- problem set-up and training,
- eliciting beliefs about the quantities of interest,
- fitting of an appropriate distribution,
- feedback of implications of fitted distribution,
- revision of judgements.

This chapter provides a detailed description of the methodologies behind these steps. We distinguish between expert elicitation of continuous and categorical variables. It also describes the software tool in which these methodologies are implemented and for which a tutorial is given in the next chapter.

2.1.1 Quantifying beliefs about continuous variables

In this section, we consider the problem of eliciting one expert’s beliefs about a continuous quantity that may or may not be bounded. Many strategies have been proposed over the past fifty years for capturing such information and to fit appropriate probability distributions.

In The Elicitator, we used a variation of the bisection method of Raiffa (1968). This method involves dividing the range of possible values into areas of equal probability. This leads to judgements being made on the median of the expected distribution and then the lower and upper quartiles. Part of this task is to question the experts about extreme events to prevent overconfidence in their median estimate before asking them about the quartiles. The bisection method was selected as it does not require the experts to have a good understanding of statistics. A variation is that we could allow the experts to make judgements about quantiles other than the quartiles: this allows the experts to make judgements about the quantities that they are most familiar with. For example, they might prefer to give a 90% credible interval by specifying the 5th and 95th percentiles.

In order to fit representative distributions to the experts’ judgements, a least-squares fitting procedure was used to select appropriate parameters for the representative distributions. In The Elicitator, fitting is done for several parametric families and the best match is reported (this is the same as in the SHELF software). For a least-squares fitting procedure, an optimisation routine is used to find the distribution parameters that minimise the squared difference between the experts’ judgements and the corresponding summaries from the fitted distribution. This is close to the procedure for fitting beta distributions described in O’Hagan (1998). The best match is then displayed for the expert in both probability density and cumulative distribution form. Changes can be made in real time by the expert to find a shape that matches their beliefs.

Once the experts are happy with their choices, they are given two statements about the implications of their judgements and the fitted distribution:

> “Your responses suggest that a value of X is highly unlikely.”
> “According to your judgements, there is a 1 in 5 chance of the variable falling outside the range (A, B)”

The first statement refers to the value of the 99th percentile of the fitted distribution and the second gives an 80% credible interval. Both are different from the judgements that were given originally and give the expert a different angle on their choices. If the expert is happy with the consequences of their judgements, then the elicitation task is complete for that variable. If not, the expert can go back to their original judgements and alter them to correct the error spotted in the feedback process. This feedback process may require several iterations before the expert is satisfied with the resulting fitted distribution.

2.1.2 Quantifying beliefs about categorical variables

Categorical variables differ from continuous variables in that the aim of an expert elicitation exercise is to get probabilities for a variable being in different states, which may or may not have some ordering to them. In this case, the idea of fitting a distribution is not useful unless the quantity of interest is on a discrete numerical scale. Categorical variables include binary outcomes, class membership and discrete variables. The principles of elicitation remain the same, but, instead of trying to fit a parametric distribution to the experts' judgements, the individual probabilities are elicited. This puts a constraint on the number of categories we can reasonably ask an expert to consider: it is difficult to imagine a situation where more than 10 probabilities are assigned to different categories, but this will depend on the skill and the patience of the individual expert.

For categorical variables, we could follow a similar approach to the continuous case if the discrete classes could be approximated on a continuous scale. However, if we have a variable that has a number of classes that may or may not be ordered and are not related to a continuous scale, we need to use a different approach. In this case, it is important for the individual classes to be well defined because ambiguity will only add the experts' uncertainty.

The following questions can be used to elicit information about categorical variables and the required probabilities:

> "For each class, can you give a description of the circumstances when the variable will fall inside that class?"
> "Can you give a probability for the variable falling in this class?"

Here we check that the probabilities sum to one. The success of this method depends on the ability of the expert to assign probabilities. An alternative approach would be to allocate a number of counters to the expert and ask them to divide these between the classes in proportion to their relative probability.

The second question is an example of what would be called a direct method of eliciting probabilities (O'Hagan, et al., 2006). It has been noted that experts who may not have experience of subjective probability benefit from using graphical methods when assigning probabilities to several classes (Johnson et al., 2010). We employ a form of graphical method in The Elicitator that has been called the "roulette" method or the "chip and bin" method.

The basic idea is to allocate each expert a number of identical items that symbolise blocks of probability, they assign the items to bins that represent the categories and they stop when they are happy that the proportion of items in each bin over all the items in the bins reflects their probability of that category occurring for the variable in question. When the expert is happy with the number of items placed in each bin, they are then given two feedback statements. The first is a table showing the implied probabilities for each category given the expert's item allocation. The second is a statement of the form:

> "This means there is a x% chance of being in the category 'A' or 'B', is this correct?"

If the expert is happy with the consequences of their judgements, then the elicitation task is complete for that variable. If not, the expert can go back to their original judgements and alter them to correct the error spotted in the feedback process. Again, this feedback process may require several iterations before the expert is satisfied with the implied probabilities.

2.1.3 Pooling of expert opinion

The problem of combining any number of experts' beliefs is another area of elicitation where a great deal of research has been carried out. Kadane (1986) recommends that prior distributions used in medical applications are representative of the community of experts; this may lead to any number of expert's beliefs being combined to form one prior distribution. Opinion pooling offers us a mathematical framework to form a prior distribution for the whole group. The combining of beliefs means that each individual's opinion must be pooled together in a way in which everyone is satisfied. French (2011) gives a review of the problem of putting opinion pooling techniques into practice and the problem of distinguishing between the experts and the decision makers. French commented that this is an extremely difficult task and a simplistic, democratic technique for constructing consensus probabilities is the only viable method — assuming one does exist.

The problem was first addressed mathematically in Stone (1961) where a linear pooling strategy is proposed. In SHELF, the simple linear opinion pooling technique is used. This was followed by Bacharach (1972) who proposed a logarithmic combination of distributions fitted to each expert's judgements. A more recent alternative is Cooke's classical method (as described in Cooke, 1991) where experts' distributions are weighted based on their ability to make judgements about uncertainty and the amount of information in the judgements they have provided. This method requires a number of seed (or test) questions to be answered where the true answer is available, but not known by the expert. The seed questions are used to judge the calibration of the experts. Typically, we will need more than ten questions on variables that are comparable with the variable of interest. This requires a lot of thought by the problem owner, and the completion of the questionnaire will require a lot of time from the experts. This method is implemented in the EXCALIBUR software.

Another mathematical aggregation method described in Thomas and Ross (1980) is *Vincentization*. Vincentization consists of calculating a weighted average of quantiles from the experts' fitted distributions to use as consensus quantiles. The problem of selecting weights in the averaging procedure remains a problem and the most democratic method is to set equal weights. In The Elicitator, this averaging is done for 100 quantiles and the weights are set to be equal. If all of the expert's distributions share the same parametric form, the distribution resulting from the Vincentization process will follow that same distributional form. Of course, this will often not be the case; therefore, the original least-squares fitting procedure is then used on the averaged quantiles to fit an appropriate parametric distribution to the group consensus.

Mathematical pooling of experts' opinions needs to be done with care. If we use a process like the linear or logarithmic opinion pools, we can get an 'average' distribution that does not reflect the opinions of any experts in the group and it is possible to fix your own judgements to have a massive impact on the resulting 'group' distribution. Reaching a group consensus about the judgements needed to fit a distribution through discussion is not really viable when the exercise is being carried out remotely. The combining of beliefs means that each individual's opinion must be pooled together in a way in which the ultimate decision maker is satisfied. The democratic way is to give all the experts equal weight in the pooling scheme; however, in doing this, we might be giving more weight to a particular school of thought. Also, we might lose information about there being several distinct groups of opinions. However, it is not clear, in the context of a web-based tool, who should be judging the adequacy of the ultimate ‘group’ distribution in the light of the individual’s judgements and the extra qualitative information.

The difficulties of actually implementing a fair mathematical opinion pool have led to a search for alternatives. Lindley (1985) suggests implementing a multivariate normal model that brings together all the differing opinions. By using this method, the decision maker can state their beliefs about the experts’ knowledge. After all, it is the decision maker who has the final say not the panel of experts. De la Fuente et al (1993) offers a possible solution to the problem of pooling opinions by considering the situation from a psychological standpoint. It is clear from all the papers considered in this section that the formulation of a representative opinion is contentious at best.

A comprehensive literature review of techniques for opinion pooling is given in Genest and Zidek (1986) and French (2011).

2.2 Description of the tool

In this section, we outline the structure of the expert elicitation tool that is developed in UncertWeb. The idea is to produce a generic tool for capturing experts’ beliefs, based on the principles highlighted in the previous section. The tool consists of the following sequence of steps:

- experts are provided with briefing documents that explain the goals of the exercise and what is expected in the elicitation process,
- experts fill in a well-designed questionnaire. This yields initial estimates of appropriate summaries of the variables of interest,
- the summaries are used to fit a probability distribution that captures each expert’s beliefs.
- Summaries of the fitted distribution are shown to the expert (possibly with some visualisation of the associated probability density function),
- experts can then modify their judgement, which will, in turn, change the fitted distribution. When the experts are happy that the distribution adequately captures their beliefs, the results are stored,
- given independent distributions for each expert’s beliefs, mathematical aggregation of the distributions is performed to hopefully get a distribution that encapsulates the group of experts’ beliefs,
- the results are reported in a way that ensures transparency and allows the information from the process to be carried through the model chain.

Each one of these steps has a number of difficulties that we must overcome in the tool, and there are a number of times where the R software (http://www.r-project.org/) is called. Further details are given in the sections below.

The Elicitator is built around a PHP Model-View-Controller architecture framework called Symfony (http://www.symfony-project.org/). The loose coupling of components enforced by the framework ensures that the project is maintainable and scalable. The client-side work is handled using a number of JavaScript frameworks. All the animation and user interface interactions are handled via the jQuery (http://www.jquery.com) and jQuery UI (http://www.jqueryui.com) frameworks. The statistical analysis is provided by an R package and rendered using jStat (http://www.jstat.org), a JavaScript statistical library. Data persistence is handled by a remote MySQL database.

2.2.1 Setting the scene (step 1)

The first step is to provide the experts with information about the process and to clearly spell out the variable that is of interest in the elicitation. In the context of UncertWeb, we will be using the definitions of the inputs to the models. However, as shown in the tutorial section, any variable or parameter can be elicited. It is important that the definitions are unambiguous and written using the experts' language.

In The Elicitator, the problem owner sets up a briefing document that serves to familiarise the expert with the elicitation process and to clearly define the quantities of interest. The briefing document provides experts with:

- a statement of the nature of the problem and usage of the elicitation result,
- a description of the elicitation procedure,
- a description of the target variables and related conditions,
- an explanation of elicitation techniques (e.g. bisection method, Roulette method, …),
- an explanation of the probabilistic summaries involved (e.g. median, quartiles, …),
- a statement of requirements from experts (e.g. providing personal data, information about their expertise, participating times, …),
- notices for experts about biased judgments,
- recommended literature about the problem (if available).

A set of passages like the ones given in the SHELF documentation are a good starting point for this step:

> "The purpose of the elicitation meeting is to obtain probability distributions to represent your uncertainty about various quantities of interest."
>
> "It is important to note that you will not be asked to provide single estimates of any of these quantities. The elicitation process will instead involve considerations such as what a plausible range of values would be for each unknown quantity, and whether, in your opinion, some values are more likely than others. You may have considerable uncertainty about some of these quantities (though less than that of a lay person). This will not be of concern during the elicitation itself, as the outputs from the elicitation will reflect large uncertainty when it is present."
>
> "Due to the subjective nature of elicited probability distributions, it is important to make the elicitation process as transparent as possible. A written record will be kept of the meeting, which will include details of experts present at the meeting, a summary of each expert's relevant expertise, and any declarations of interest."

Also, at this stage, we clarify the use of the judgements (and the supplementary information that will be collected) in the model chain and what will be accessible (and attributable) once the group distribution has been fitted.

2.2.2 Collecting information (step 2)

Before quantitative judgements are made about the variable(s) of interest, we need to collect information about the experts and their expertise. This serves two purposes: first, the ultimate decision maker needs to know who the opinions belong to so they can have faith in the results from the model chain, and, secondly, this stage helps the expert to focus on the problem in hand and to remember relevant information sources.

Some of the questions suggested in SHELF and in Gosling (2005) are:

> "Have you got any interests that are related to the variable under consideration?"

This is asking for a declaration of interests. This is important as experts are often stakeholders in the wider process. They may be employees who will benefit from success in the enterprise to which the elicitation contributes. They may be invited specifically to represent a stakeholder group or point of view. Recognising the potential vested interests of themselves and other participants helps the experts to report their beliefs openly and in an informed way. It is also important for the decision maker to be aware of possible biases.

> "What is your expertise in relation to the variable under consideration?"

The purpose of this question is self-explanatory. As a decision maker, we would want to know about the expert's experience in areas related to the variable(s).

> "What facts are important when making judgements about the variable under consideration?"

This could lead to a list of other influencing factors that the expert needs to think about when making the judgement. We also use this question to check for ambiguities in the definition of the variable (for example, has the scenario been defined clearly enough?).

> "What quantitative or qualitative evidence have you seen relating to the variable under consideration?"

This could be a list of key publications and reports. It could also be evidence that they have seen in their careers and research.

When the first stage of information collecting is completed, we can move our focus to getting quantitative estimates of uncertainty about the variable of interest. The way this is done is different depending on the variable(s) and the level of statistical training the experts have had. The common case for elicitation considers a continuous one-dimensional variable. Below, we outline two approaches to getting the relevant information.

1. Range estimation (see Gosling, 2005)

> "Can you give some examples of scenarios when the variable would be much higher than usual?"
> "Make a high estimate such that you feel there is only a 5% probability that the true value of the variable would exceed your estimate."
> "Can you give some examples of scenarios when the variable would be much lower than usual?"
> "Make a low estimate such that you feel there is only a 5% probability that the true value of the variable would fall below your estimate."
> "Your answers give a 90% range for the variable; do you agree that there is a one in ten chance of the variable being outside this range?"
> "Would you like to change your estimates for the 5th or 95th percentiles?"
> "What is the most likely value of the variable?"

This method gives a 90% credible interval and the mode of the distribution. It relies on the experts having some knowledge of probability. The questions are done in this order to help prevent the experts anchoring on their best guess for the variable.

2. The bisection method (see Raiffa, 1968)

> "Can you determine a value such that the variable is equally likely to be less than or greater than this point?" (a)
> "Can you give some examples of scenarios when the variable would be much lower than usual?"

“Suppose that the variable is definitely below your answer to (a). Can you now determine a new value such that the variable is equally likely to be less than or greater than this point?” (c)
“Can you give some examples of scenarios when the variable would be much higher than usual?”
“Suppose that the variable is definitely above your answer to (a). Can you now determine a new value such that the variable is equally likely to be less than or greater than this point?” (e)
“Your answers to (c) and (e) give a 50% range for the variable; do you agree that there is a 50-50 chance of the variable being outside this range?”
“Would you like to change your estimates for 25^{th} or 75^{th} percentiles?”

This method gives the median and interquartile range of the distribution. It has the advantage of the experts not needing to know much about probabilities. However, the hypothetical nature of two of the questions can be confusing, and we do not get information about the tails of the distribution.

Of course, for both methods, the tool should check that the judgements made are coherent.

2.2.3 Fitting appropriate distributions (step 3)

In order to fit a distribution to the judgements, we employ a least-squares fitting procedure (for example, see O’Hagan, 1998). In such a procedure, the elicited judgements are compared against the corresponding theoretical quantities from a fully specified probability distribution. We select the parameters of that distribution by finding the parameters that minimise the squared difference between the elicited and the theoretical quantities.

Minimising this measure does not necessarily result in an appropriate distribution being fitted. For example, if the expert’s judgements indicate that the distribution is likely to be heavily skewed, we will not be able to find a set of parameters for a normal distribution such that that distribution is an adequate representation of the expert’s beliefs. In the SHELF package, several different distributions can be considered in the fitting process. Often, we have information on the likely shape of the distribution prior to the judgements being made (for instance, we may know that the distribution is bounded between 0 and 1); we can use this information to choose appropriate distributions on which to attempt the fit.

For a categorical variable, this type of fitting is unnecessary because the full distribution will have been defined by the expert in the previous step.

2.2.4 Feedback of the initial fits (step 4)

In the feedback stage, the expert judges the assumption that the fitted distribution is an adequate representation of their beliefs. As the experts are often not experts in interpreting plots of density functions, there is little point to just showing a graph and asking if that conforms with what they had in mind. Even if the experts could do this, we could not expect the expert to differentiate between several distributions that have similar characteristics. Often, the expert will find it useful to be given statements or summaries about the fitted distribution that are in a similar format to the original questions. For instance, we might produce a credible interval based on the fitted distribution.

It is possible that our reported distribution for the expert’s density may not agree with what the expert really thinks: the distributions we have chosen might be inappropriate and/or the expert may have given us probability judgements that do not really match their beliefs. In this case, the expert could give us different or more information to help update the fitted distribution.

Appropriate questions at this stage include:

> “Your responses suggest that a value of X (maybe using the 99^{th} percentile here) is highly unlikely; do you agree with this?”
> “According to your judgements, there is a 1 in 5 chance of the variable falling outside the range (u,v)?”

For a categorical variable, the feedback of the results in this way is not necessary, but it is advisable to challenge the judgements. For instance, the following type of question could be used:

> “According to your judgements, there is over a 70% chance of the variable falling in category A or B. Does this fit with your beliefs?”

2.2.5 Modification in the light of the feedback (step 5)

One of the key stages in the elicitation process is the sharing of experience. At this stage, it might be worthwhile reporting back all of the answers to the earlier question “What quantitative or qualitative evidence have you seen relating to the variable under consideration?” that we have got so far. There is obviously an issue of timing here.

At this stage, it would be appropriate to allow the experts to give information about the most likely value they expect and other percentiles of their distributions. The experts would also be able to change their original judgements. These two options are particularly important if they are not satisfied with the distribution that has been fitted. We could then fit another distribution using the methods of step 3.

This is part of a feedback “loop” where steps 3 to 5 are repeated until the expert is satisfied that their beliefs have been captured adequately. In an automated procedure, we must be careful about the fact that no distributions might be judged to be adequate and that an expert might just get bored with the process and accept anything after a few iterations of the feedback loop.

2.2.6 Pooling of the experts’ beliefs (step 6)

As discussed above, the pooling of expert beliefs is contentious. The Elicitator adopts a mathematical opinion pool, which weights experts equally. The experts’ judgements are combined using the Vicentization method and the best fitting parametric distribution is used to represent the combination of the experts’ beliefs. We may explore further options in the future, including allowing the problem owner to weight experts, before seeing their elicitation results, on the basis of self-assessed or problem-owner-assessed expertise.

2.2.7 Reporting of results (step 7)

It is important to be as transparent as possible. Therefore, whenever the expert elicitation tool has been used, we must try to carry all the information about the process along with the elicited distribution. Sometimes the experts must remain anonymous, and this will have to be stated in the supporting information. The final pooled estimate is supplied as UncertML (see UncertWeb Deliverable D1.2) for further usage within UncertWeb.

2.3 Conclusions about the tool

The Elicitator is the first truly web-based tool that allows a pooled estimate of uncertainty of a particular variable of interest to be obtained from expert beliefs. The emphasis has been put on usability and simplicity. Further enhancements can be envisaged over time, allowing the system to manage a large pool of experts and problems, and work to achieve this is on-going.

3 Tutorial expert elicitation of continuous and categorical variables

The Elicitator (http://elicitator.uncertweb.org) is a web-based application that facilitates the elicitation of continuous and categorical variables. The system acts as a facilitator between the problem owner (a user who requires more knowledge about a particular variable) and domain experts (users who are 'experts' in a particular domain). This section provides a step-by-step tutorial on how a user can register, outline a problem and construct a number of continuous or categorical variables to be elicited

3.1 User registration

In order to effectively facilitate the elicitation process, it is mandatory that all users (both problem owners and experts) are registered in the application. The registration form is simple and only requires a few pieces of information to function. Figure 1 illustrates the registration form with a user's details filled in.

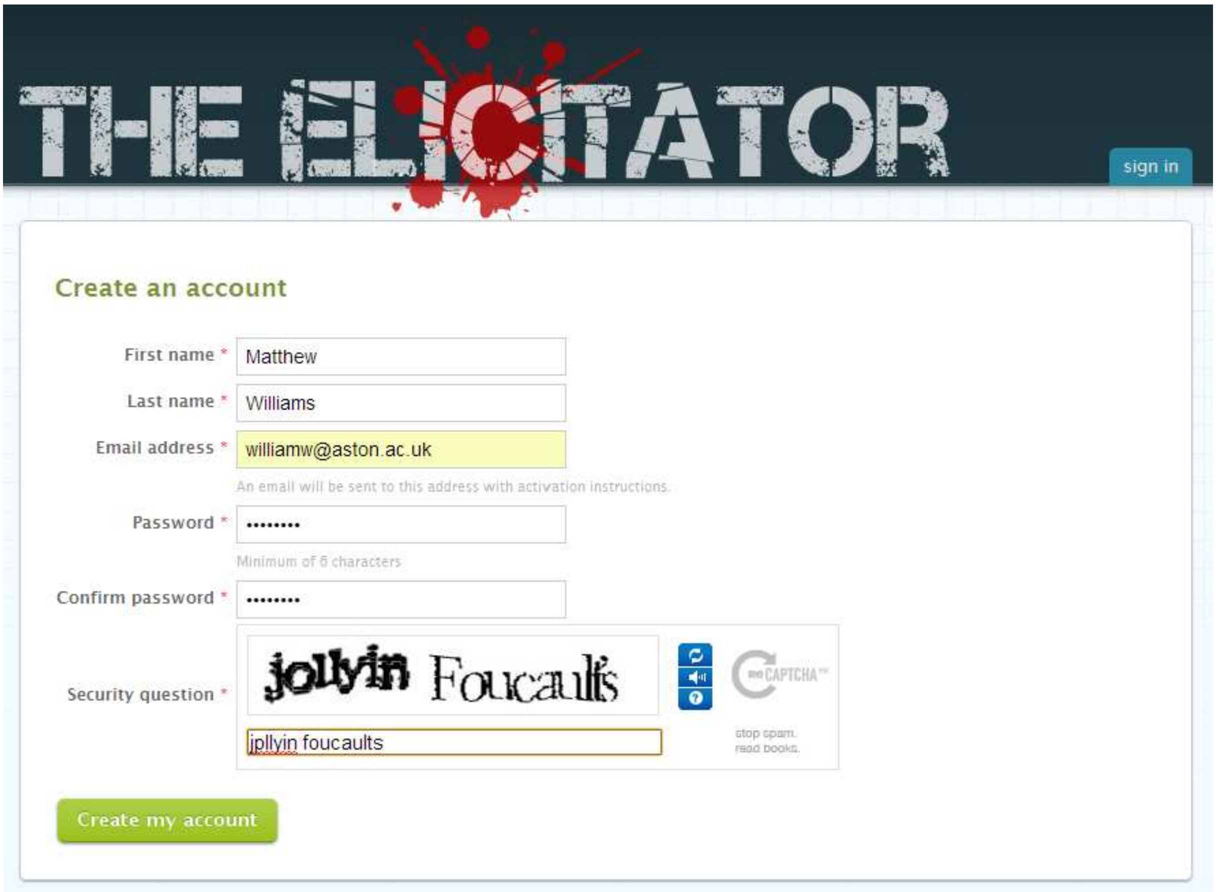


Figure 1: User registration form for The Elicitator.

Once the registration has been completed, the user is redirected to their control panel.

3.1.1 The control panel

The control panel is the central hub for The Elicitator. Everything a user is required to do can be accessed from here.

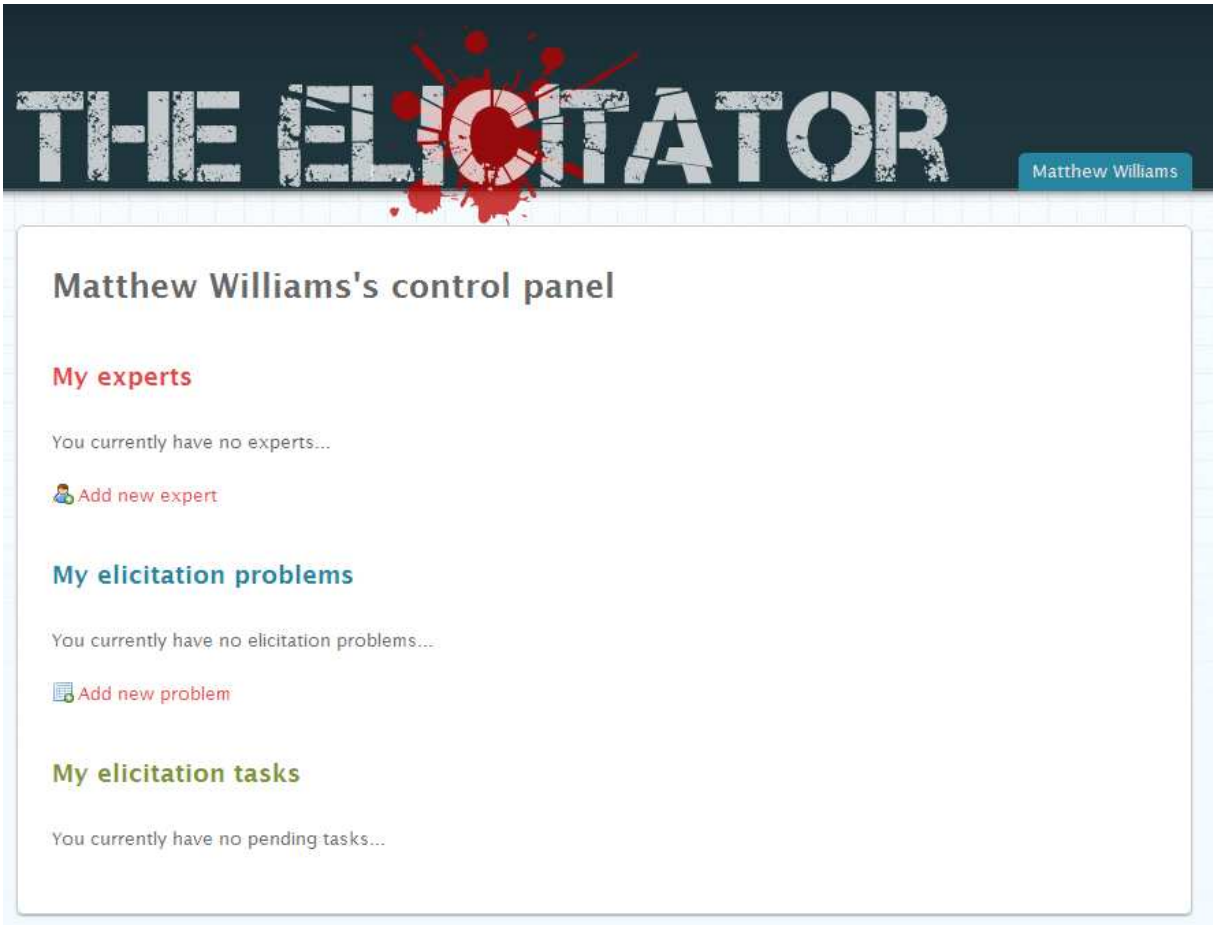


Figure 2: The control panel is the central hub of The Elicitator.

Figure 2 illustrates a user's control panel immediately after registration. It shows that the current user has no experts, no elicitation problems and no elicitation tasks. The following sections will explain what each of these sections is and how they are populated.

3.2 Constructing a list of experts

A successful elicitation requires a number of experts in a particular domain. The more experts are willing to share their knowledge of a problem the greater the chance of a reliable elicitation. After a user has registered, and before they create the problem they wish to elicit, they should take time to register a number of experts. The registration of experts requires only their names and email addresses. If an expert is already registered in the system, perhaps added by another user, their account will not be duplicated.

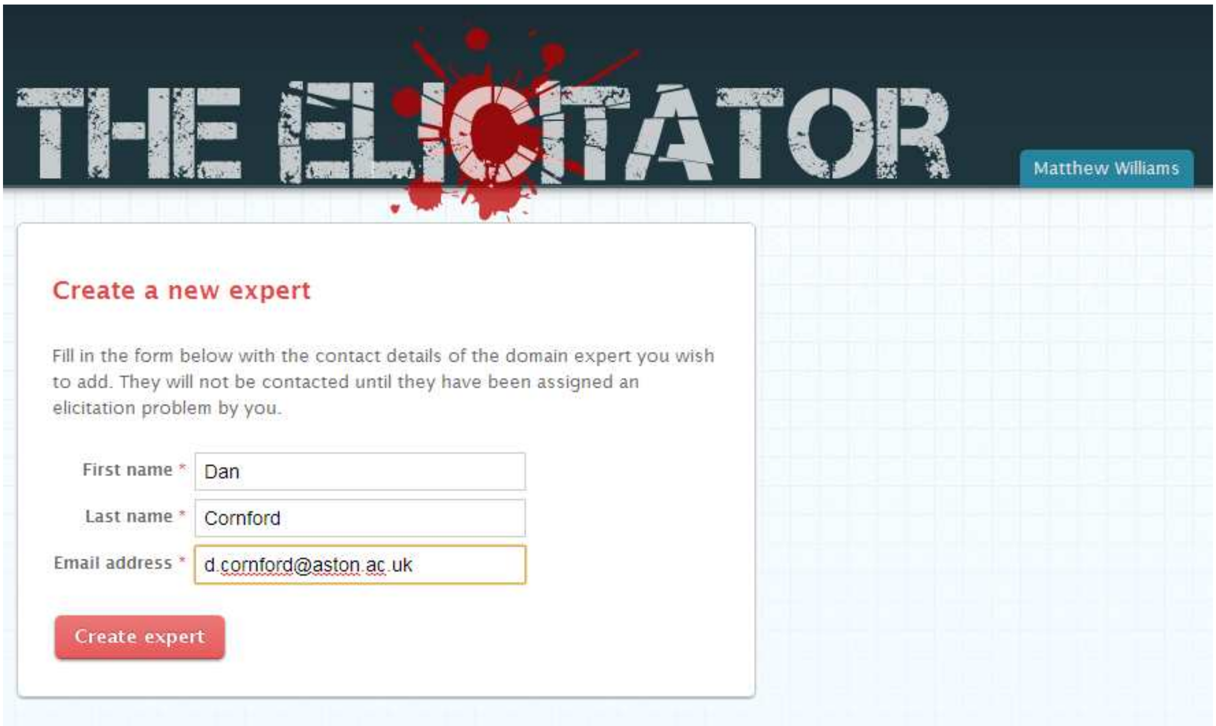


Figure 3: The form to create and add experts to your list of contacts.

Figure 3 shows the simple form for adding experts to the list of contacts. In the final system, each expert will be contacted by email requesting the confirmation of their registration. However, during the testing process the email notifications have been disabled.

A problem owner should strive to add as many experts as they can – the more experts available the more reliable an elicitation will be. In future versions the pool of experts will contain much more detail, including the domain that the person is an expert in, thus allowing a full search capability so a problem owner can discover experts that they do not know personally.

Once a list of experts has been compiled they are available to the problem owner for any elicitation problem. The current list of experts can be reviewed at any time by visiting the control panel (Figure 4). A problem owner can add more experts at any time.

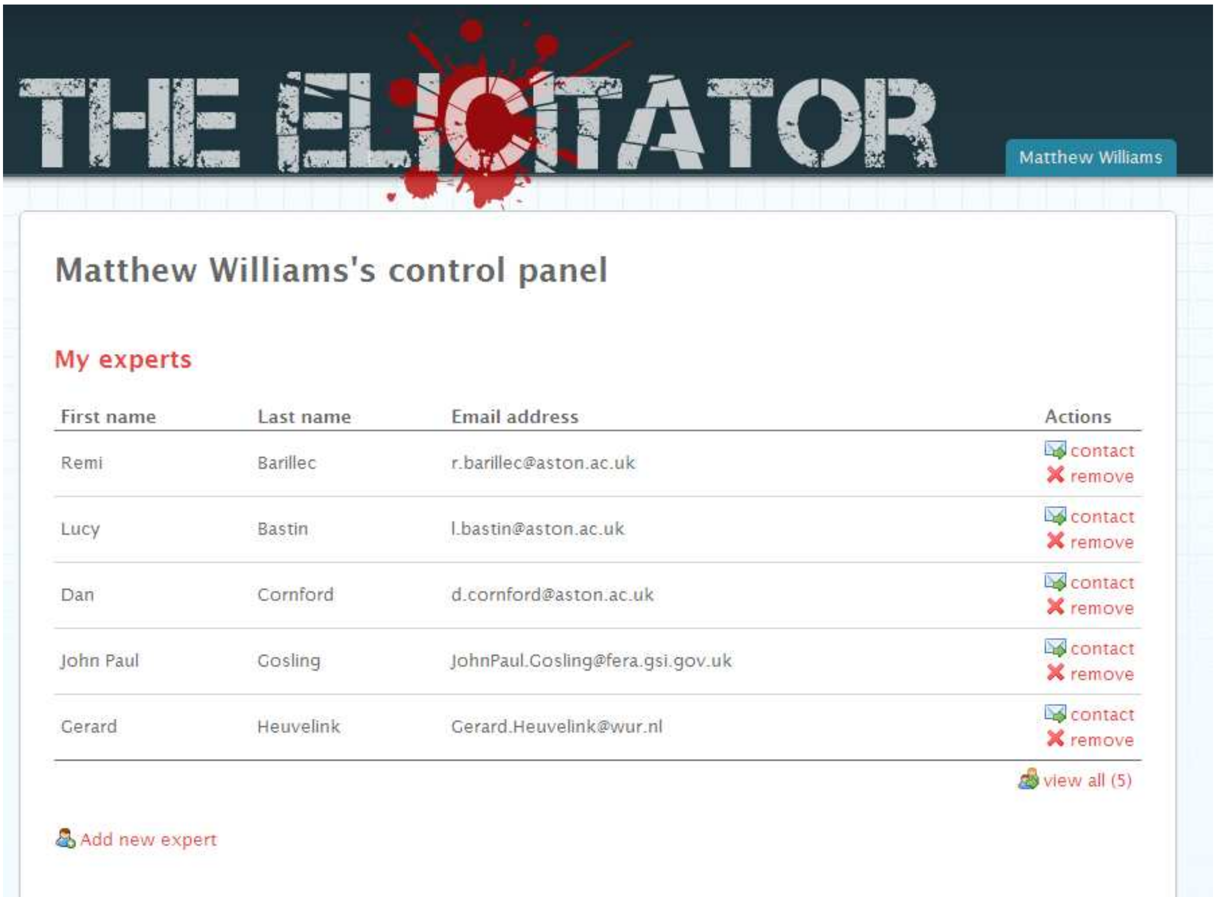


Figure 4: All experts can be navigated via the control panel.

3.3 Adding an elicitation problem

On the control panel are two other lists: elicitation problems and elicitation tasks. Elicitation problems are problems that have been defined by the problem owner and that contain a number of variables to be elicited. Elicitation tasks are tasks that have been assigned to a user by other problem owners. In effect, every user of the system is both a problem owner and a domain expert. This section details the process of creating a new elicitation problem and creating some variables to be elicited.

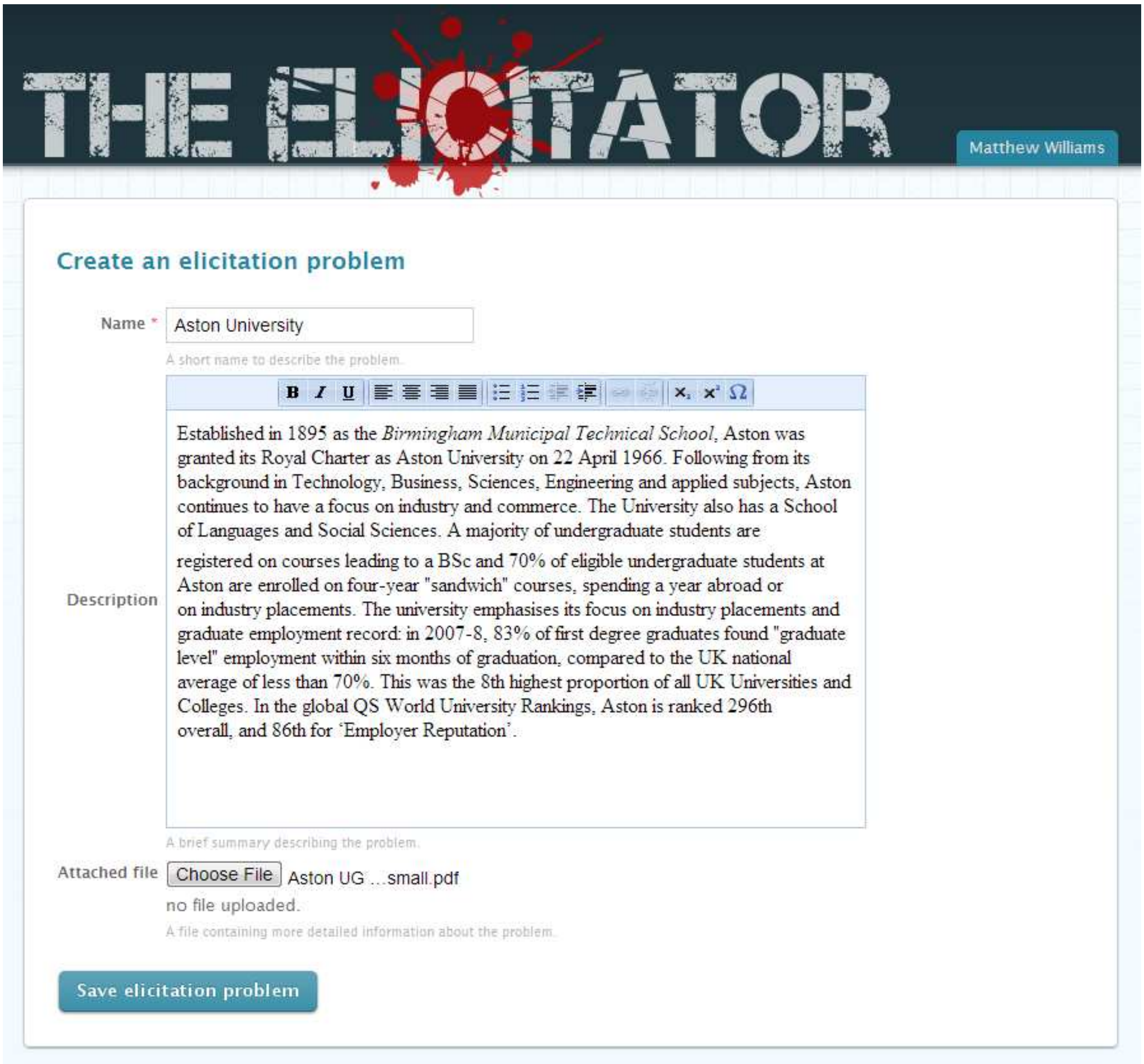


Figure 5: Creating a problem to contain several variables for elicitation.

The first stage is to create an elicitation problem. In The Elicitator, all variables that need eliciting belong to a problem. It is a convenient way of grouping variables into related concepts. Figure 5 displays the form for creating a new problem. In this example the problem is 'Aston University'; a number of variables related to Aston University will be added later for elicitation. The description field contains a brief piece of text about the problem. This is used to give context to the variables that are to be elicited. A file can be attached to the problem that provides greater detail about the problem. Both description and file fields are optional, only the problem name is mandatory.

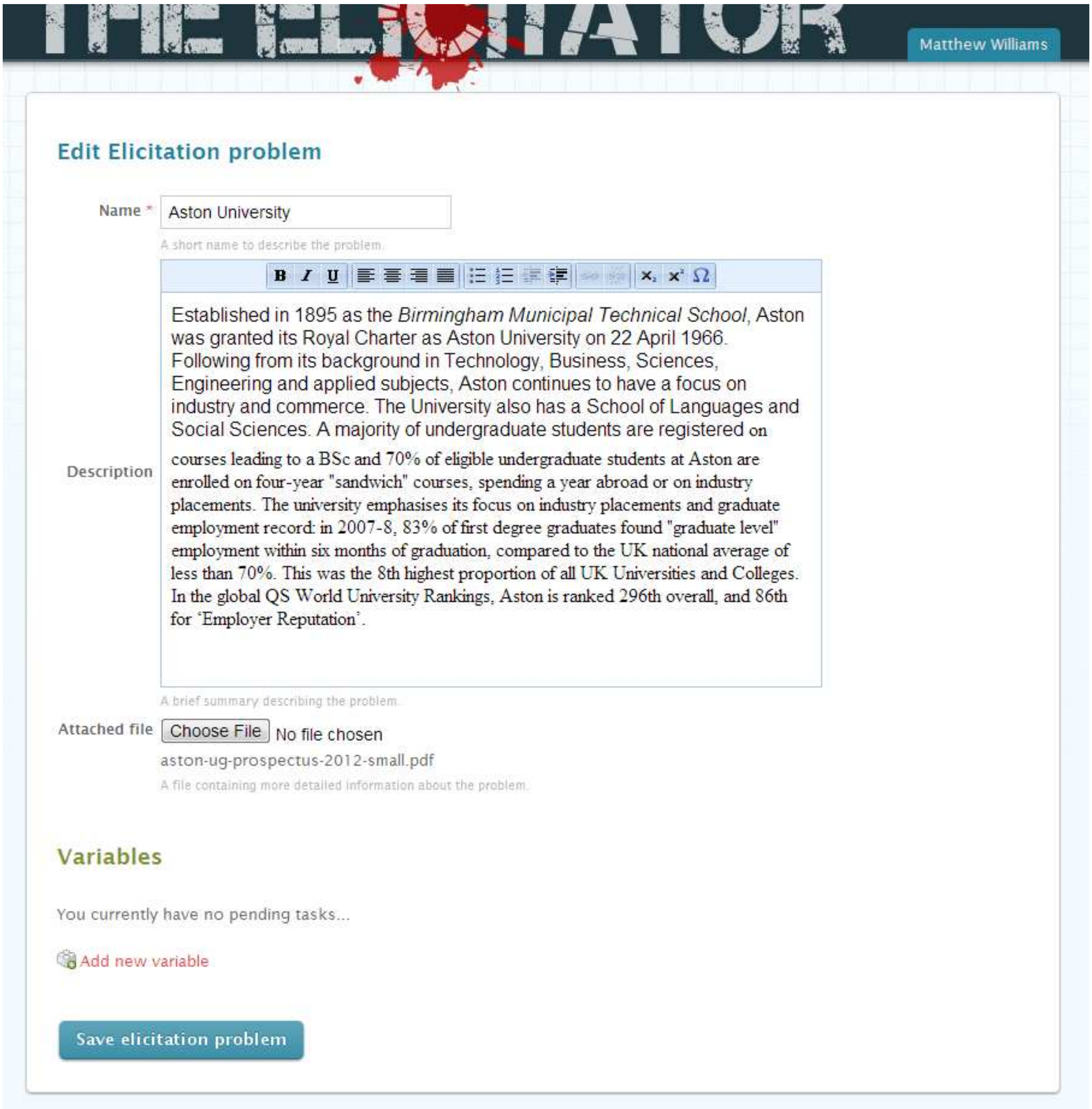


Figure 6: Adding variables to a problem can be done once the problem has been saved.

Once the elicitation problem has been saved more options are available for adding specific variables (Figure 6). Below the description of the problem is now a list of variables assigned to this problem. Initially it is empty, but more variables can be added at any time by clicking the 'add new variable' link.

3.3.1 Categorical variables

Currently The Elicitator supports both categorical and continuous variables. Future versions will incorporate the elicitation of spatial variables (see Chapters 4 and 5). Adding any type of variable to a problem is done via the 'add new variable' link on the elicitation problem page.

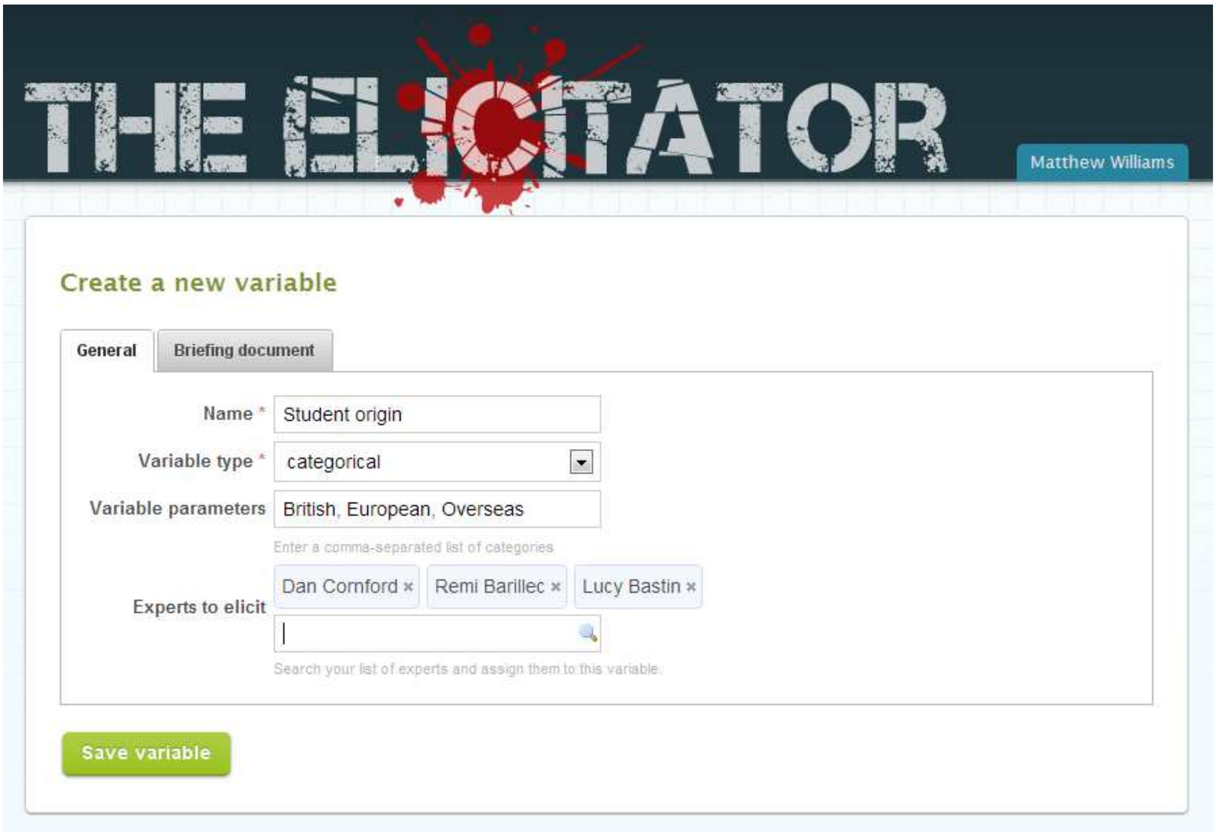


Figure 7: Form for creating a new categorical variable.

Figure 7 illustrates the form for creating a new categorical variable. A descriptive name of the variable is required, as well as the type (either categorical or continuous). If the variable is categorical an additional field 'variable parameters' is required. The variable parameters field should contain a comma-separated list of possible categories that are provided to the experts during their elicitation task. The final field is a list of experts that will be invited to participate in the elicitation of this variable. In the example in Figure 7, three experts have been selected based on their suitability to the problem.

As an optional step a problem owner can outline a briefing document that will be presented to each expert before they start the elicitation task. If no briefing document is provided, only the general elicitation problem description will be available. A problem owner can supply the briefing document by selecting the tab at the top of the page.

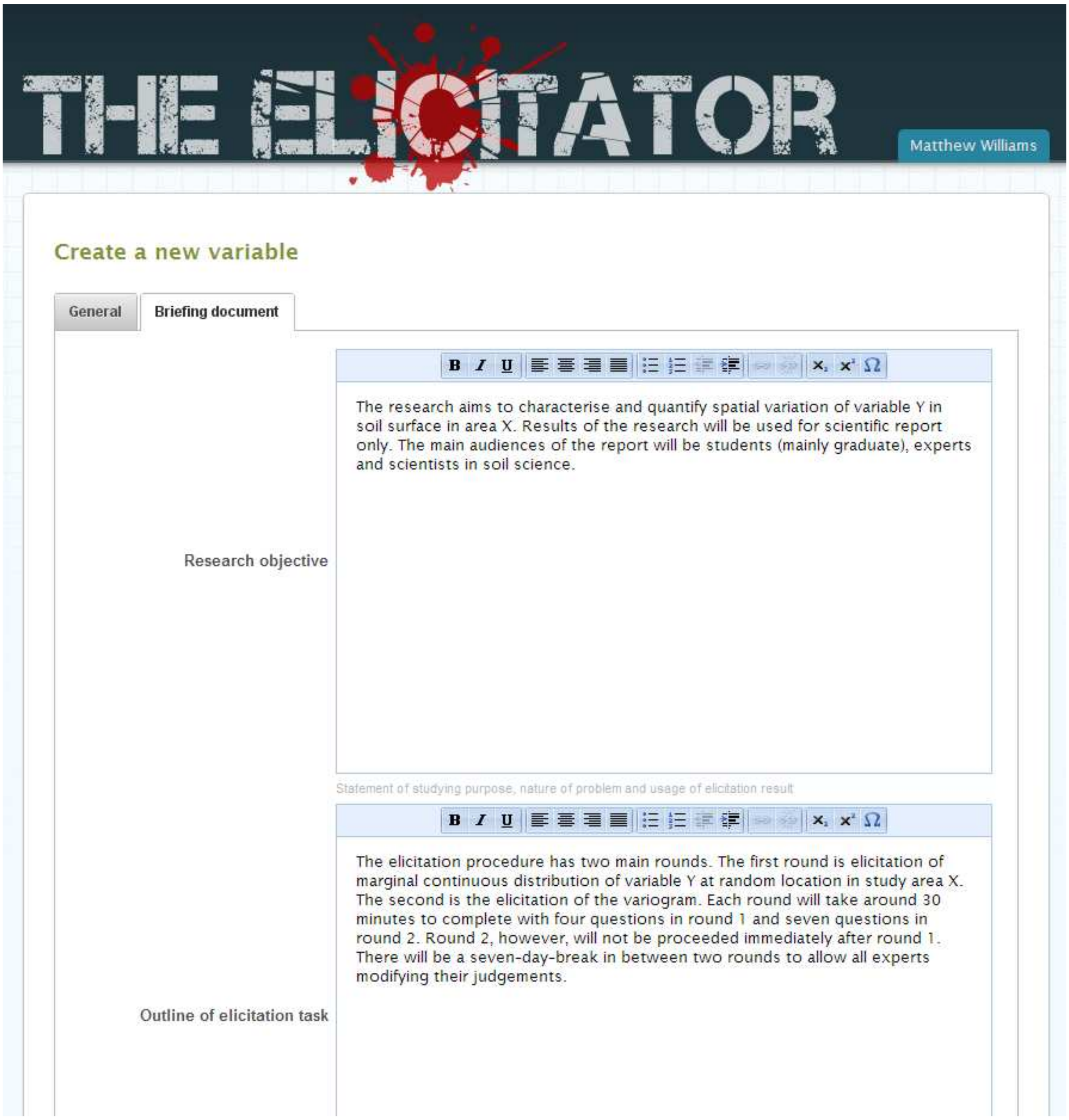


Figure 8: Form for creating a briefing document for a given variable.

Figure 8 displays a section of the briefing document form. There are nine sections that can be filled in that will aid the expert in understanding the context of the problem and what is required of them. A rich text format box is provided to allow a problem owner to format the briefing document to provide emphasis to important concepts. All nine sections are optional; those that are emitted will not be compiled into the final document. Once the briefing document has been completed, the variable can be saved.

3.3.2 Continuous variables

Continuous variables are implemented in a similar way to categorical variables. The primary difference in their creation is that the variable parameters field has been omitted (Figure 9).

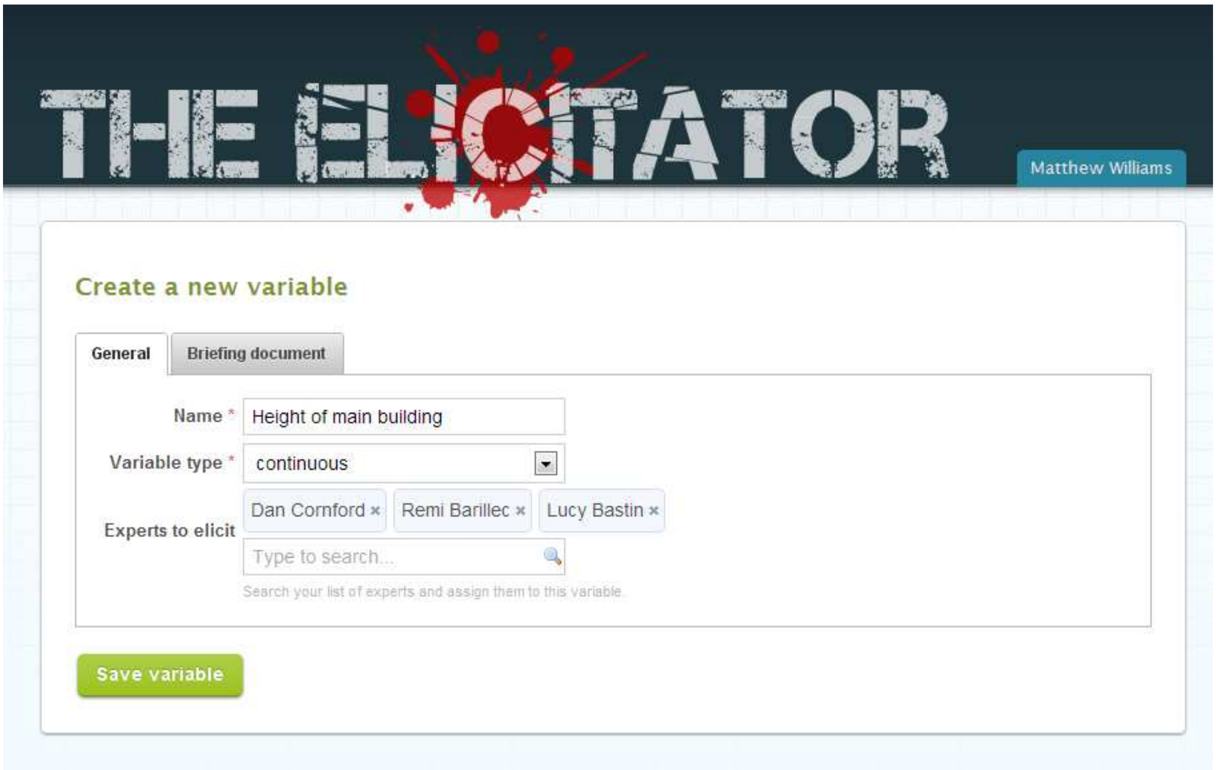


Figure 9: Form for creating a new continuous variable.

The same options are still available to the problem owner for creating a briefing document. It should also be stated that the briefing document is on a *per variable* basis, i.e. a document must be created for each variable, or propagated up to the elicitation problem description.

Once all variables have been created for an elicitation problem, the control panel should resemble the one in Figure 10. From here a problem owner can review all their elicitation problems with options to view the current problem, including who has completed the elicitation tasks, edit the problem, perhaps to add more variables or to delete the problem and stop the elicitation task. A convenient progress bar is situated next to the problem name to provide the problem owner an overview of how far the elicitation task is going at a glance.

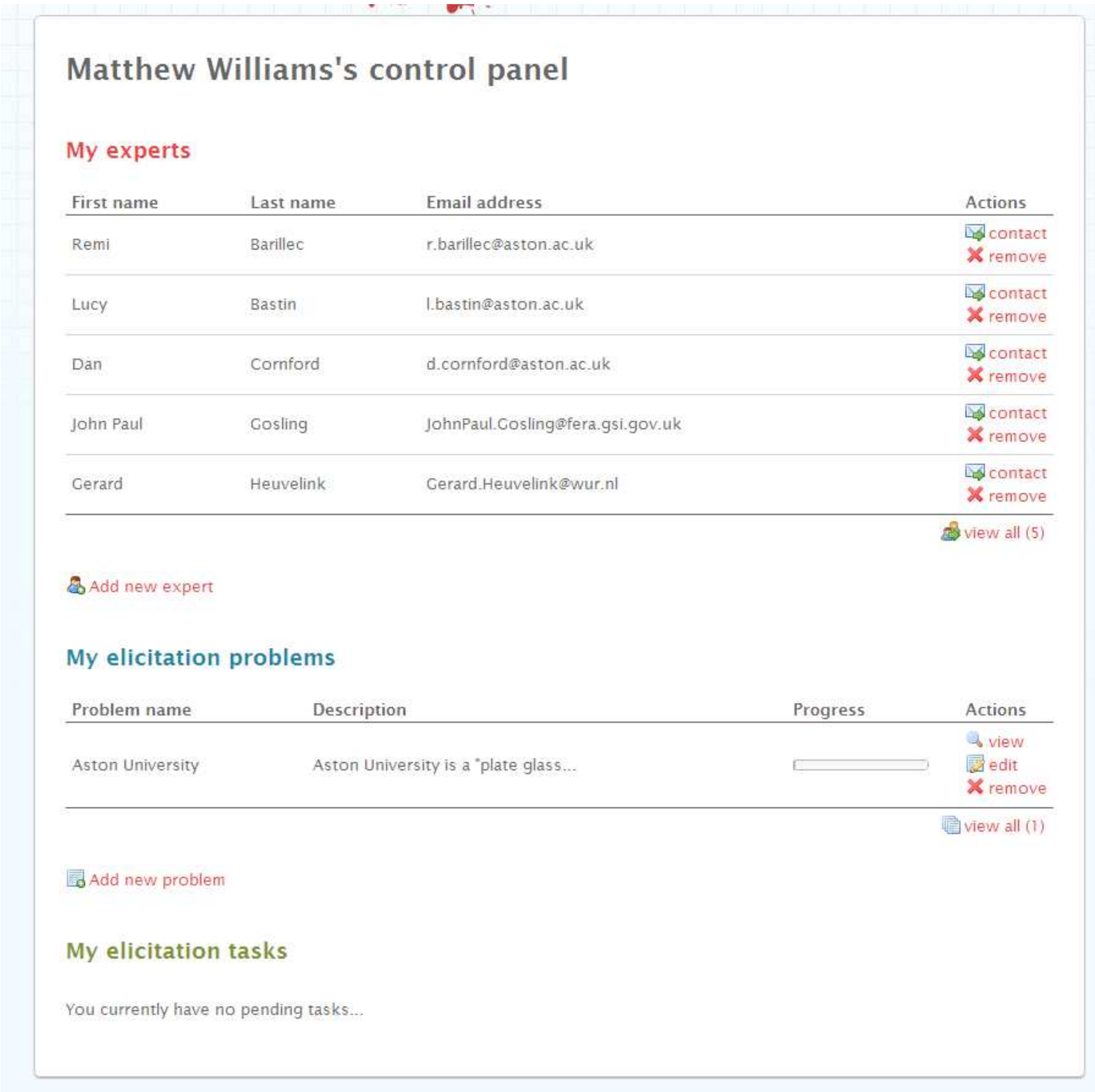


Figure 10: The control panel once an elicitation problem has been created.

At this stage in the final system an email will have been sent out to all experts assigned to the variables inviting them to take part in the elicitation task, providing a convenient URL for them to follow. The perspective of an expert can be seen in the following section.

3.4 Logging in as an expert

The previous sections outlined how a user can act as a domain expert to add elicitation problems and create both categorical and continuous variables to be elicited. It also showed how domain experts could be assigned to any number of variables. This section will outline how The Elicitator works from an expert's perspective, how they can keep track of their tasks and how the actual elicitation process is achieved.

3.4.1 The control panel

Once an expert has logged in they, like the problem owner, are faced with the control panel. However, if they are a new user, they will have no experts or problems but rather a number of elicitation tasks (Figure 11).

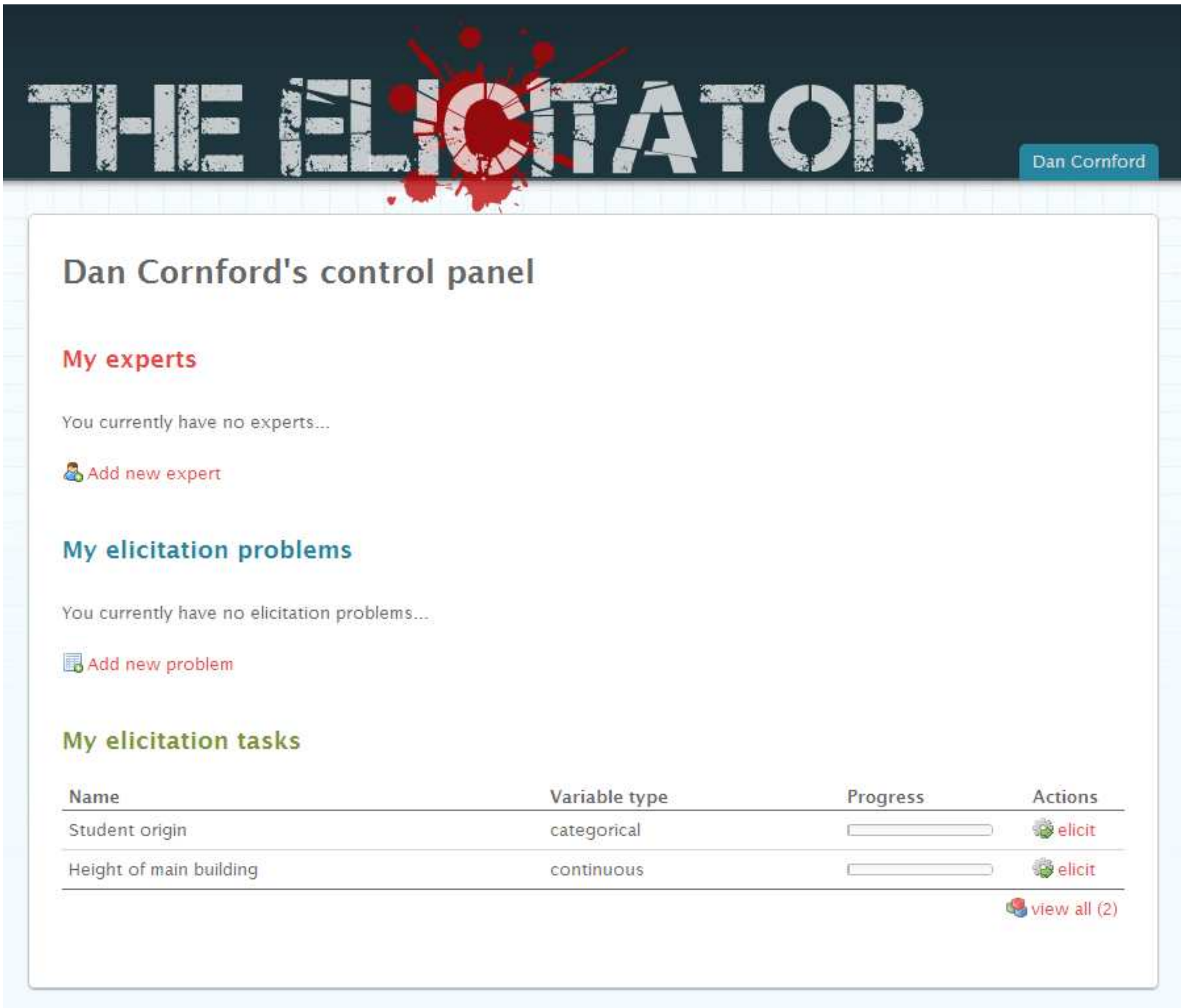


Figure 11: The control panel of a domain expert.

Figure 11 shows the control panel of one of the experts invited to the variables created in the previous section. In the 'My elicitation tasks' section of the control panel are the two variables. As every user of The Elicitator is both an expert and a problem owner, this expert could at this point create their own list of experts and problems that they want eliciting. However, for the purpose of this tutorial it is assumed that this user is assuming the role of an expert.

The table of elicitation tasks is similar to the table of elicitation problems seen in Figure 10. A progress bar is again present, which shows the current progress of each task as it is possible to save progress at any stage of the process. The variable type column displays the type of elicitation that will be performed for each variable.

To commence an elicitation task an expert clicks on the 'elicit' link.

3.4.2 Elicitation of a categorical variable

When an expert is faced with the elicitation of a categorical variable they are presented with the screen seen in Figure 12.

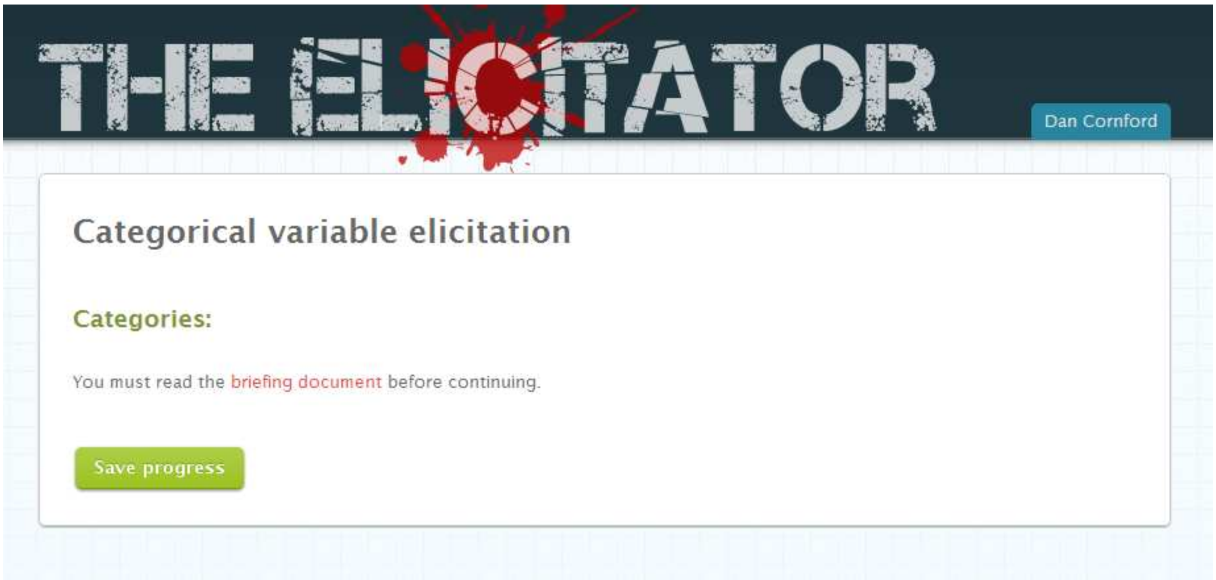


Figure 12: Initial screen of a categorical elicitation.

It is mandatory that before the elicitation can commence the expert must read and confirm they understand the contents of the briefing document. This may be uniquely tailored to the particular variable, or a more generic problem overview as discussed in previous sections. The briefing document is accessed by clicking on the 'briefing document' link (Figure 13).

The briefing document is formatted according to the rich text format applied by the problem owner, and can include concepts such as numbered or bulleted lists. The example in Figure 13 is purely illustrative and does not provide context to this elicitation task.

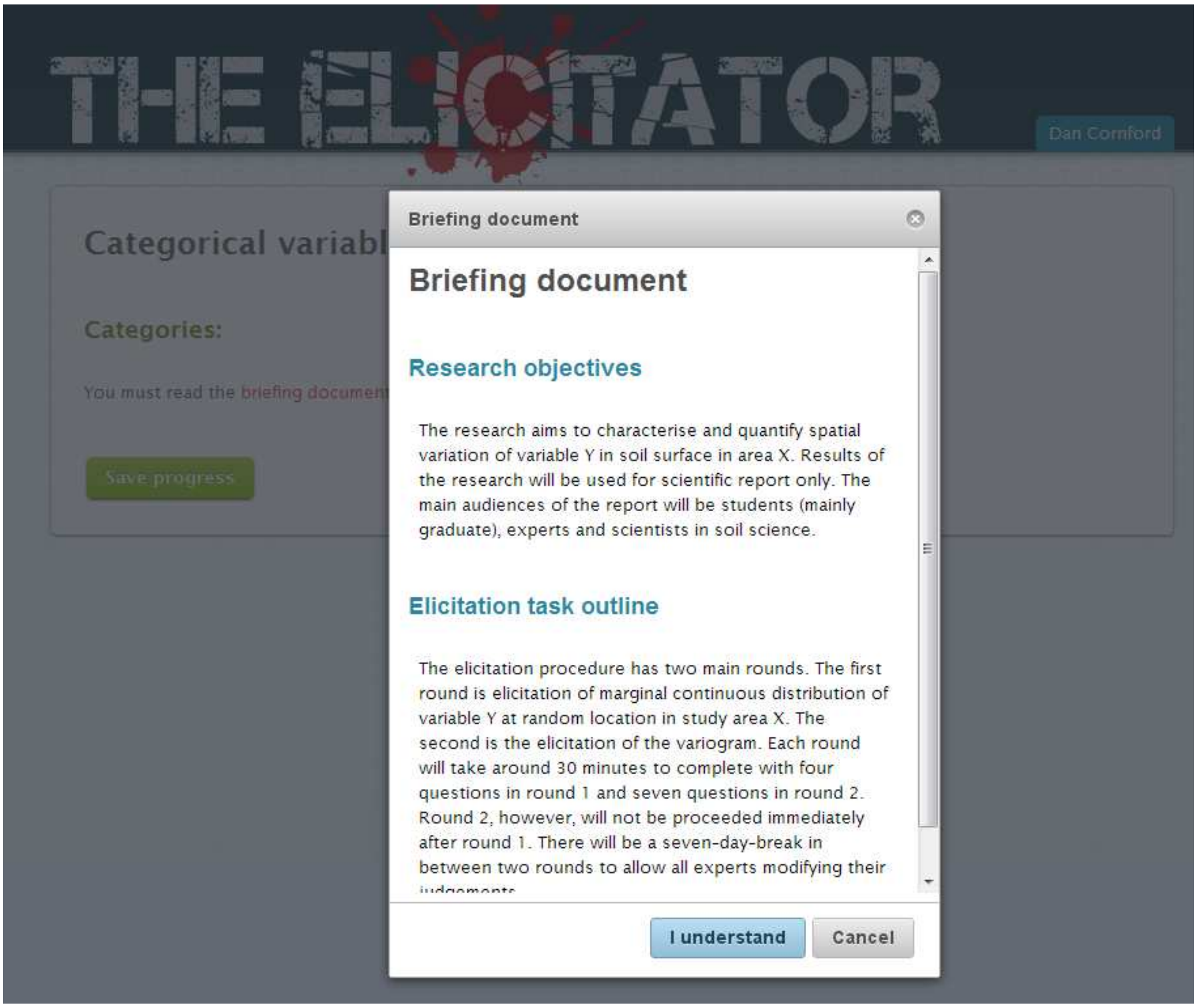


Figure 13: A briefing document for a given variable.

Once the expert has read and is confident they understand the briefing document, they can click ‘I understand’ and commence with the elicitation.

Elicitation of categorical variables is achieved using the idea of placing beans into buckets. The more beans a bucket, or category, has the more likely this particular variable is of that type (see Johnson et al., 2010).

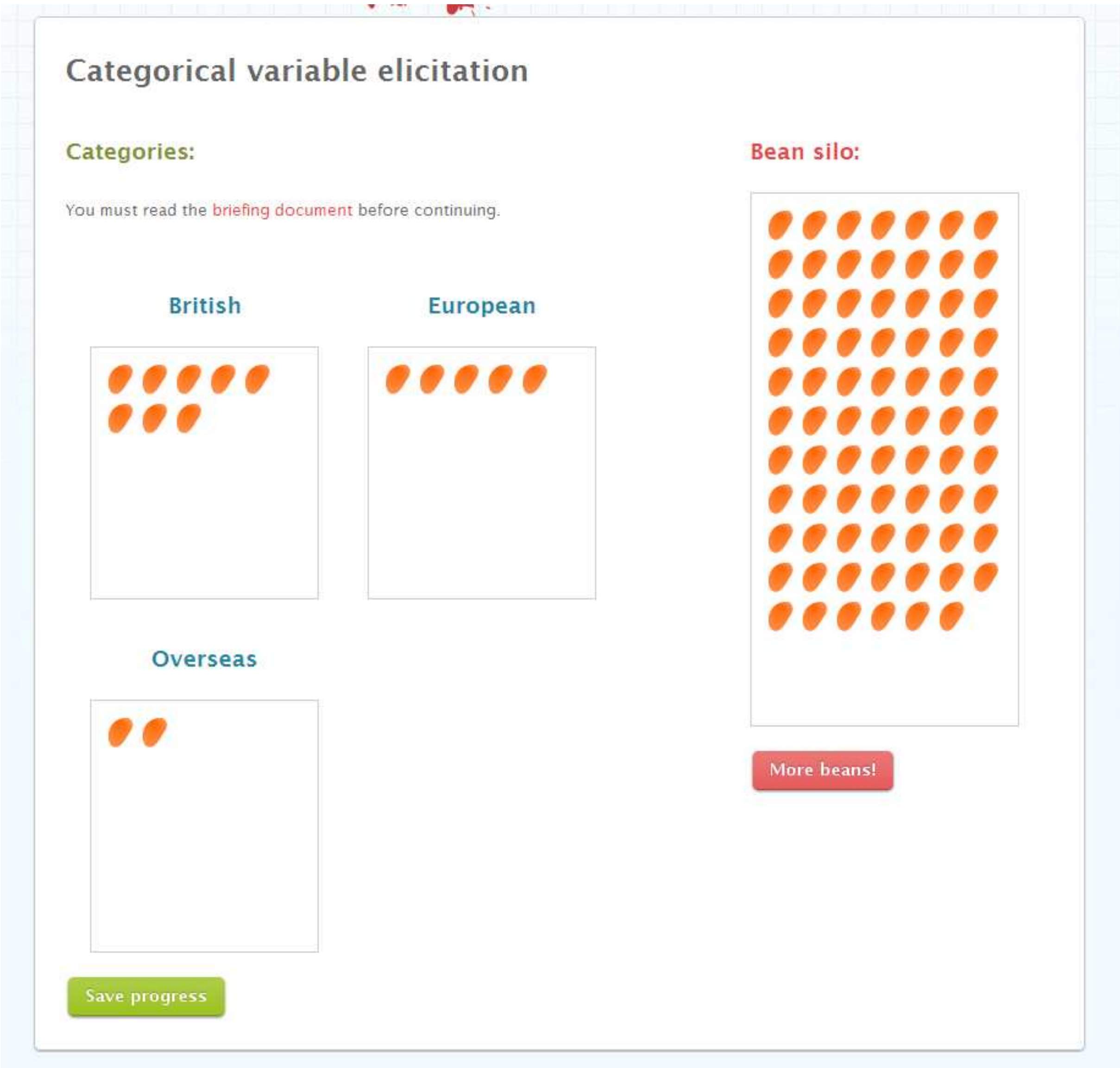


Figure 14: Categorical elicitation uses the beans and buckets method.

The image in Figure 14: Categorical elicitation uses the beans and buckets method.Figure 14 shows a completed categorical elicitation. It can be seen that the expert believes the variable (a student chosen at random from the student population at Aston) is most likely to be British, closely followed by European and a small chance of them being Overseas. The expert is free to use as many beans as they need to provide an accurate assessment. If more beans are required they can click the 'More beans!' button which will refill the bean silo. Once an expert is satisfied with their judgment they must click the 'Save progress' button to upload the results to the problem owner. However, before the results are uploaded the expert must read and confirm a statement about their answers. These confirmatory questions explain the consequences of the expert's decisions in a simple to understand phrase. The purpose of the confirmatory questions is to try and catch any mistakes the expert may have made with their judgment. Figure 15 displays a typical confirmatory question.

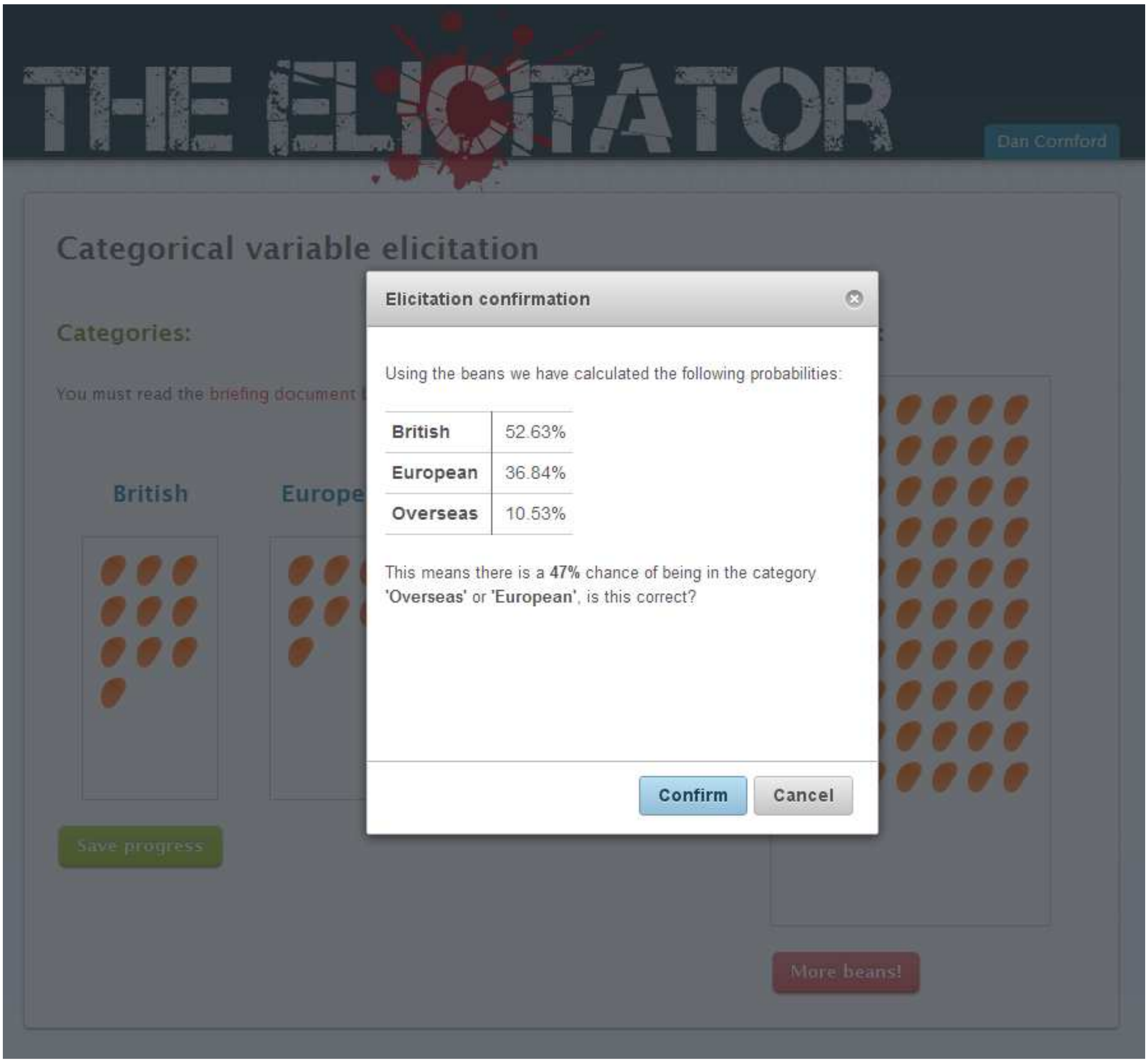


Figure 15: Question asked to the expert before the results are uploaded.

3.4.3 Elicitation of a continuous variable

Elicitation of a continuous variable is more complicated than the elicitation of a categorical variable. This is reflected by the progress indicator on the right of the screen (Figure 16).

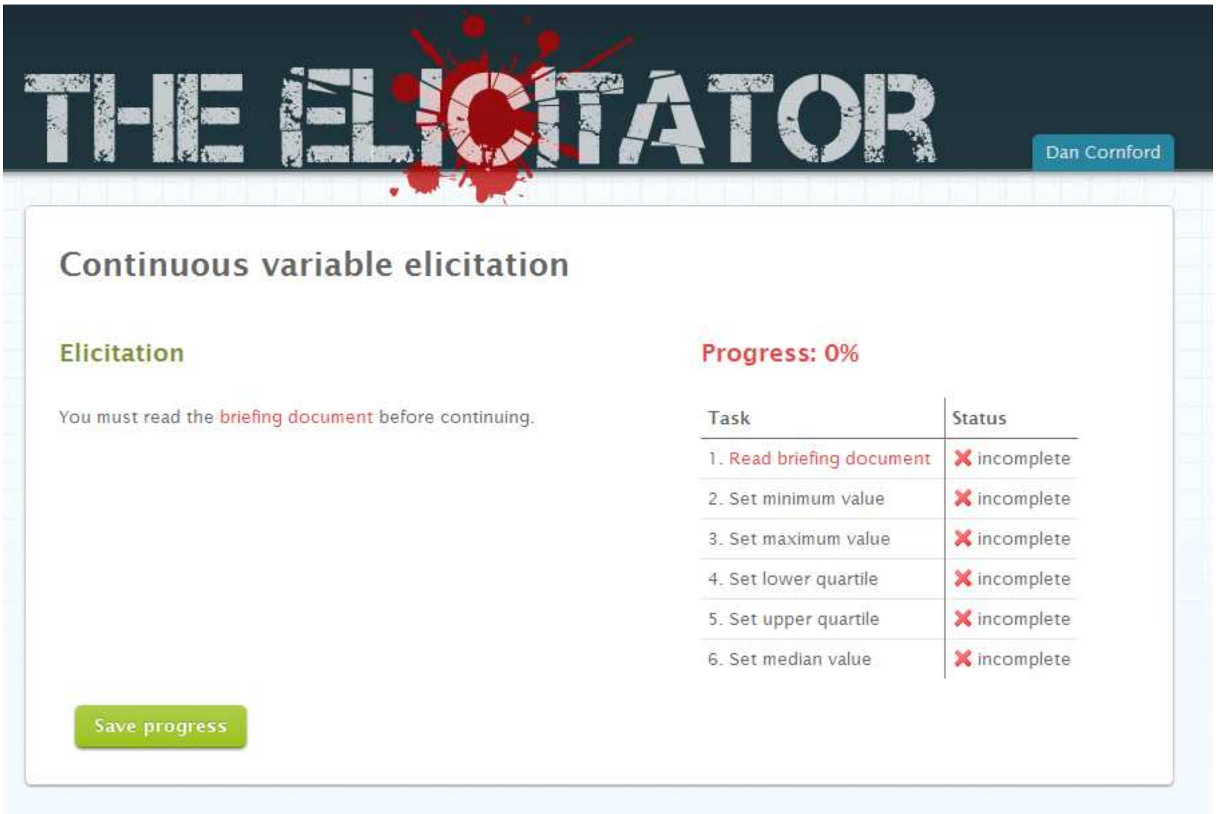


Figure 16: Continuous elicitation of a variable.

The first stage is identical to the categorical variable example; the expert must read and understand the briefing document. Once they have done that options become available to them which must be completed in a carefully prescribed order. They must set the lower, or minimum, value they believe the variable can take, followed by the maximum. Once they have entered these values three sliders become available that allow them to set the lower and upper quartiles and the median value for the variable. After all values have been elicited The Elicitator calculates a 'best-fit' probability distribution and returns it to the expert.

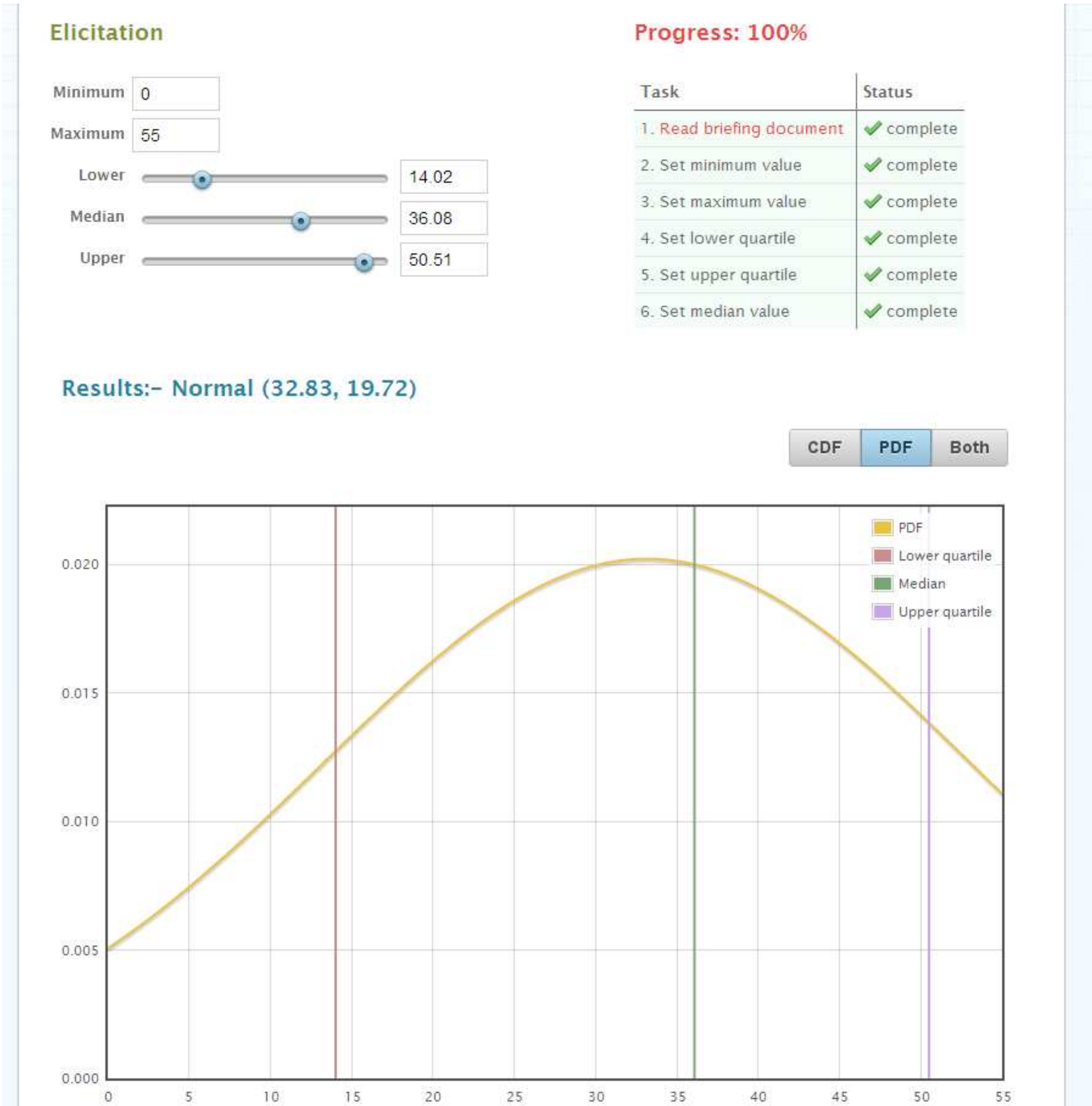


Figure 17: A completed elicitation of a continuous variable.

Figure 17 shows a completed elicitation of a continuous variable. In this example a Normal distribution with a mean of 32.93 and standard deviation of 19.72 was chosen. If the expert is not satisfied with this result they can of course adjust the values they provided. The expert has the ability to view the PDF and CDF of the distribution to see how close the distribution fits their quartiles.

Once the expert is satisfied with the elicitation they can click the save button to upload the results to the problem owner. However, as with the categorical variables they must first confirm a couple of questions.

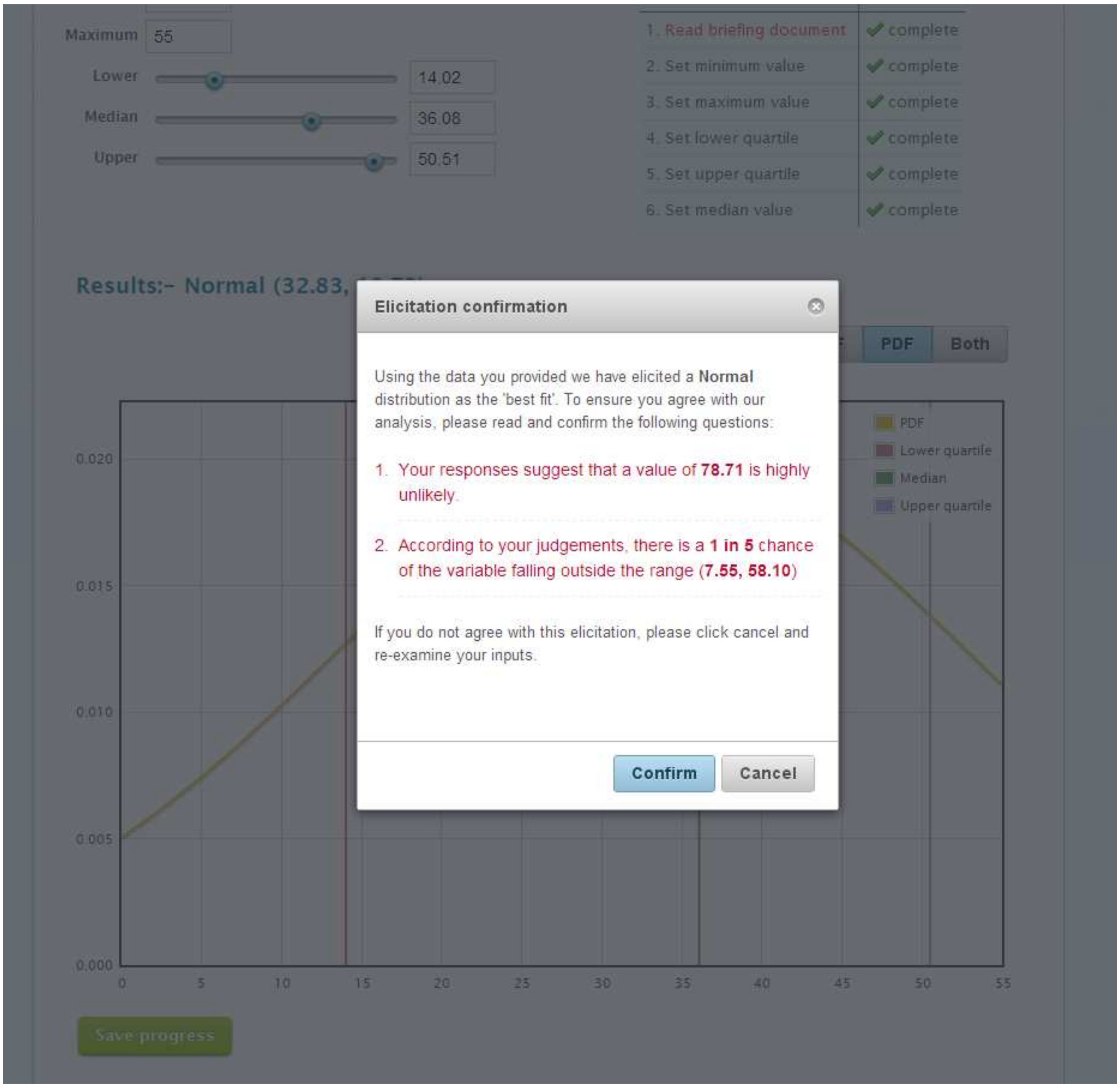


Figure 18: Confirmatory questions for a continuous variable.

Figure 18 shows an example set of confirmatory questions as a result of a continuous elicitation. These questions are phrased in easy to understand sentences that provide the expert with an alternative view of the probability distribution they were shown. If they disagree with the confirmatory questions and press cancel they are given an opportunity to adjust the values once more. Once they are satisfied they can click confirm and the results will be uploaded to the problem owner.

3.5 Reviewing the results of an elicitation problem

Once one or more experts have finished their elicitation tasks the problem owner can review the results from their control panel.

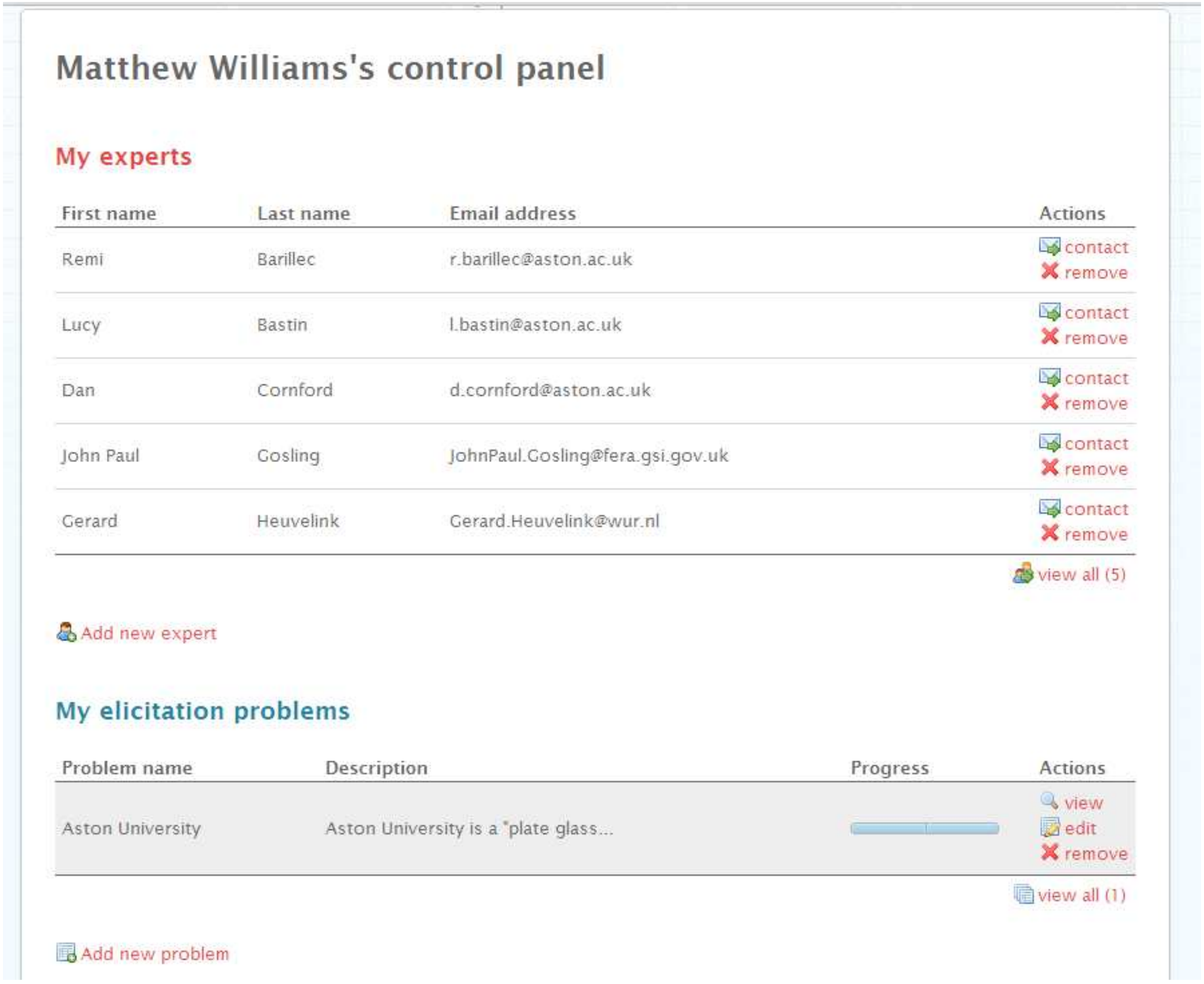


Figure 19: A completed elicitation problem shown on the control panel.

Figure 19 shows the control panel of the problem owner once all elicitation tasks have been completed for the Aston University problem. The progress bar clearly indicates that all tasks have been completed. At this stage the problem owner can view the results by clicking the 'view' link. This then provides a quick overview of each variable for this elicitation problem. Again, the progress bars clearly show that all experts have successfully completed the elicitation tasks for both variables (Figure 20).

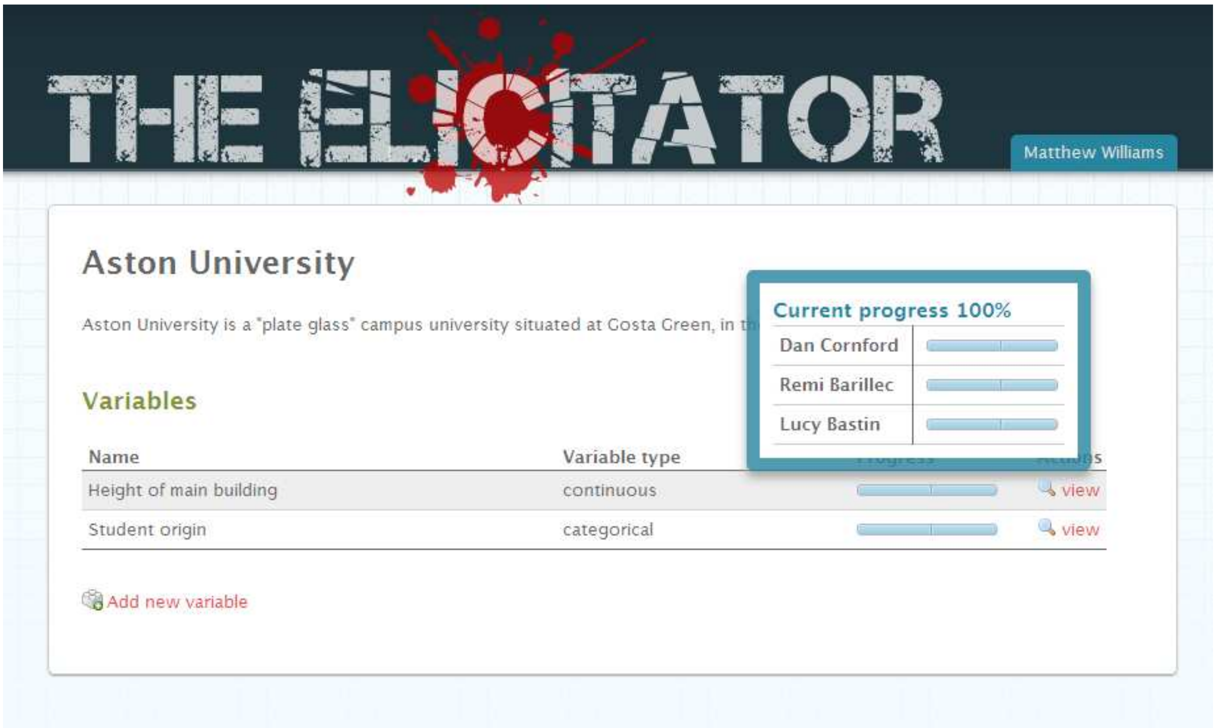


Figure 20: Progress of each elicitation task shown using progress bars.

3.5.1 View categorical elicitation results

The results of the categorical elicitation are calculated by averaging all experts' proportions for each category. These average proportions are shown to the problem owner via a bar chart. Figure 21 shows an example pooled result where the consensus is that the variable is most likely to be British. A problem owner can also view each individual expert's results below. These are illustrated as the proportion of beans for each category, and can be seen in Figure 22.

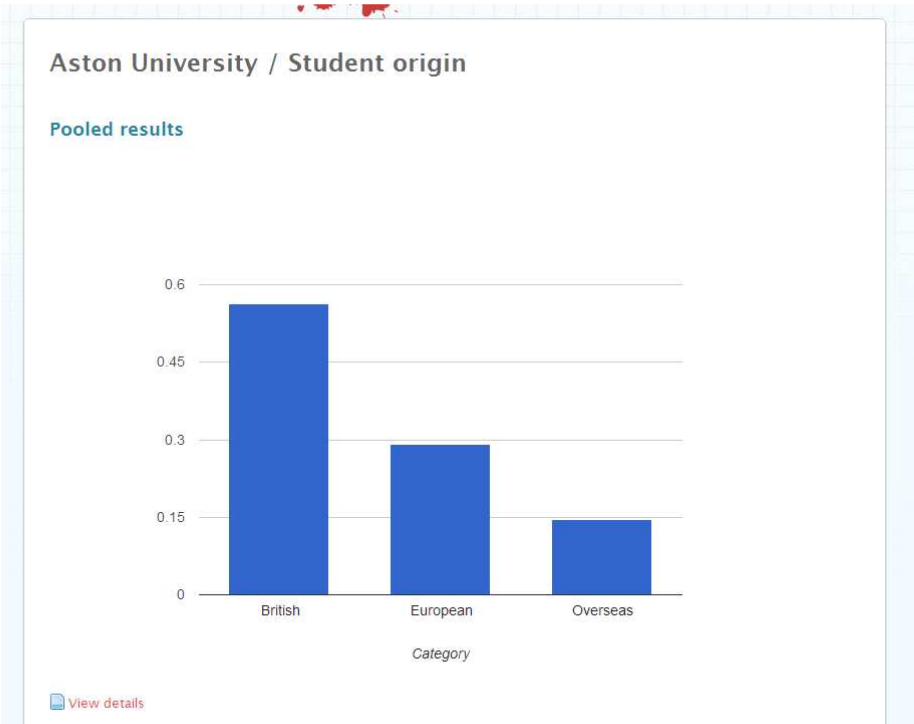


Figure 21: Pooled results for a categorical variable.

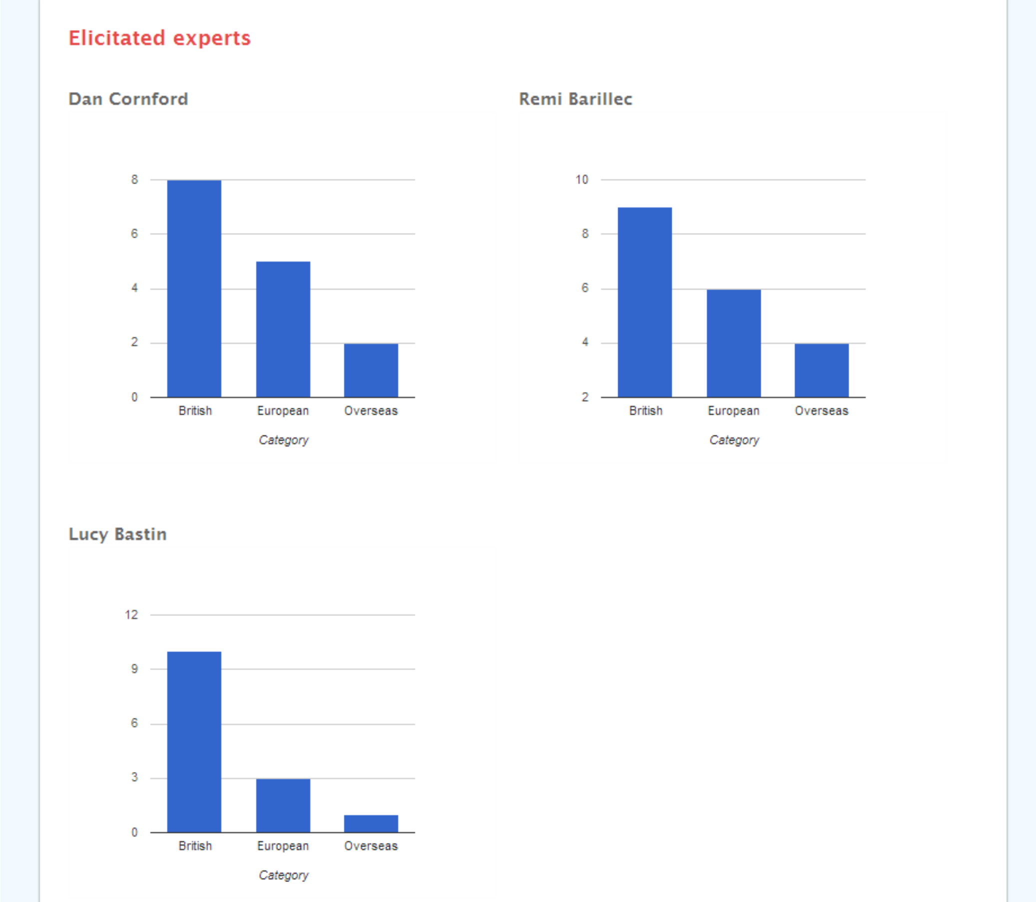


Figure 22: Each expert's results are shown as proportions.

3.5.2 View continuous elicitation results

The results of a continuous elicitation are slightly more complex. Each expert’s elicited distribution is evaluated at 100 quantiles and then these quantiles are averaged. The average of these quantiles are then used to elicit an overall distribution.

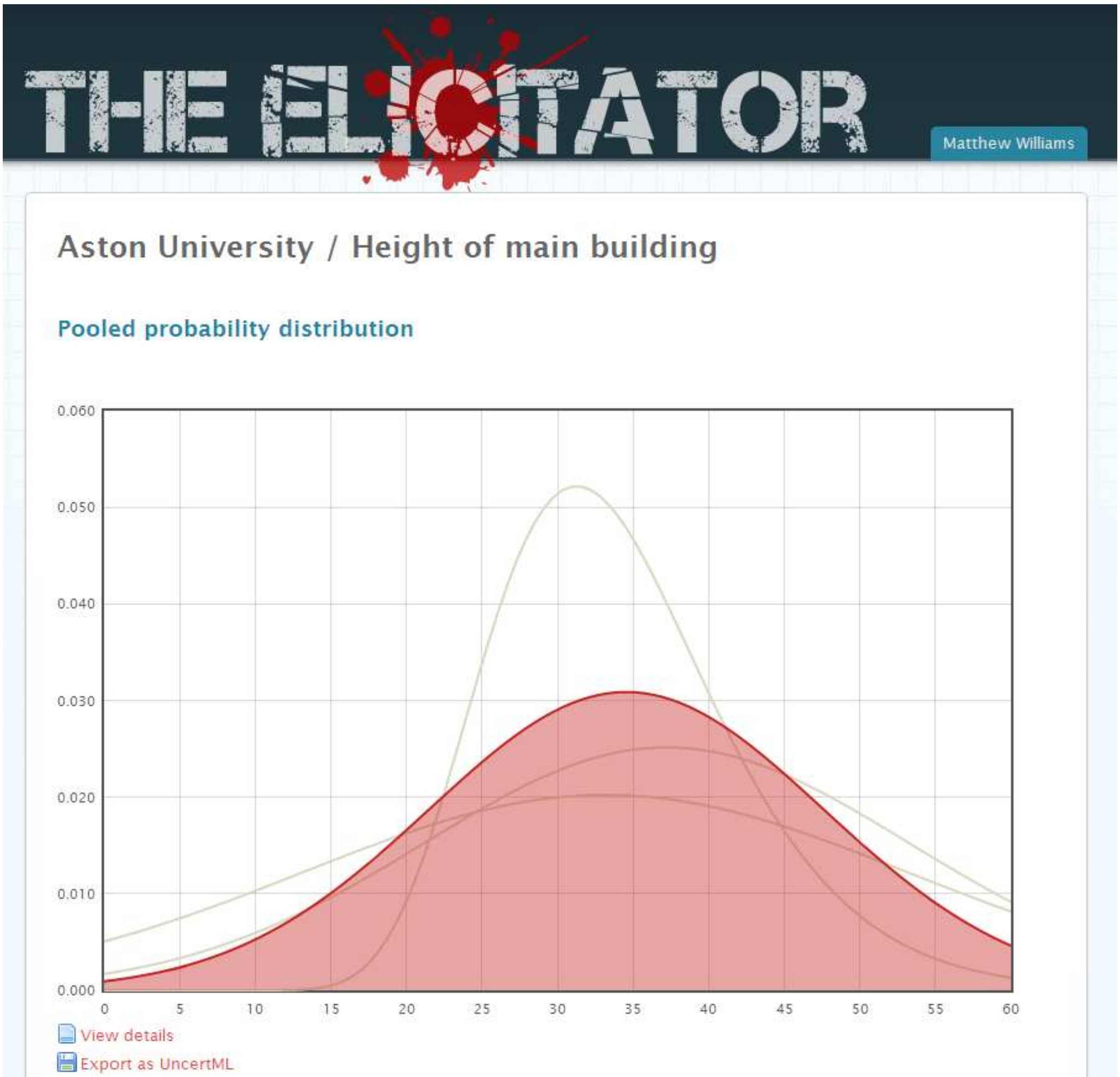


Figure 23: Pooled probability distribution for continuous variables.

Figure 23 illustrates a pooled distribution for the example illustrated in the previous section. The red distribution represents the overall pooled distribution and the red-grey lines are the distributions of each individual expert. To view the exact details of the distribution, i.e., what type of distribution it is and its parameters the problem owner can click the 'view details' link. The 'Export as UncertML' link provides a unique URL to an UncertML fragment that encodes the resulting distribution. This is useful for sharing the elicitation result with other users or applications.

Similar to the categorical elicitation results, the continuous elicitation results also show each individual experts elicited distributions. These can be seen in Figure 24.

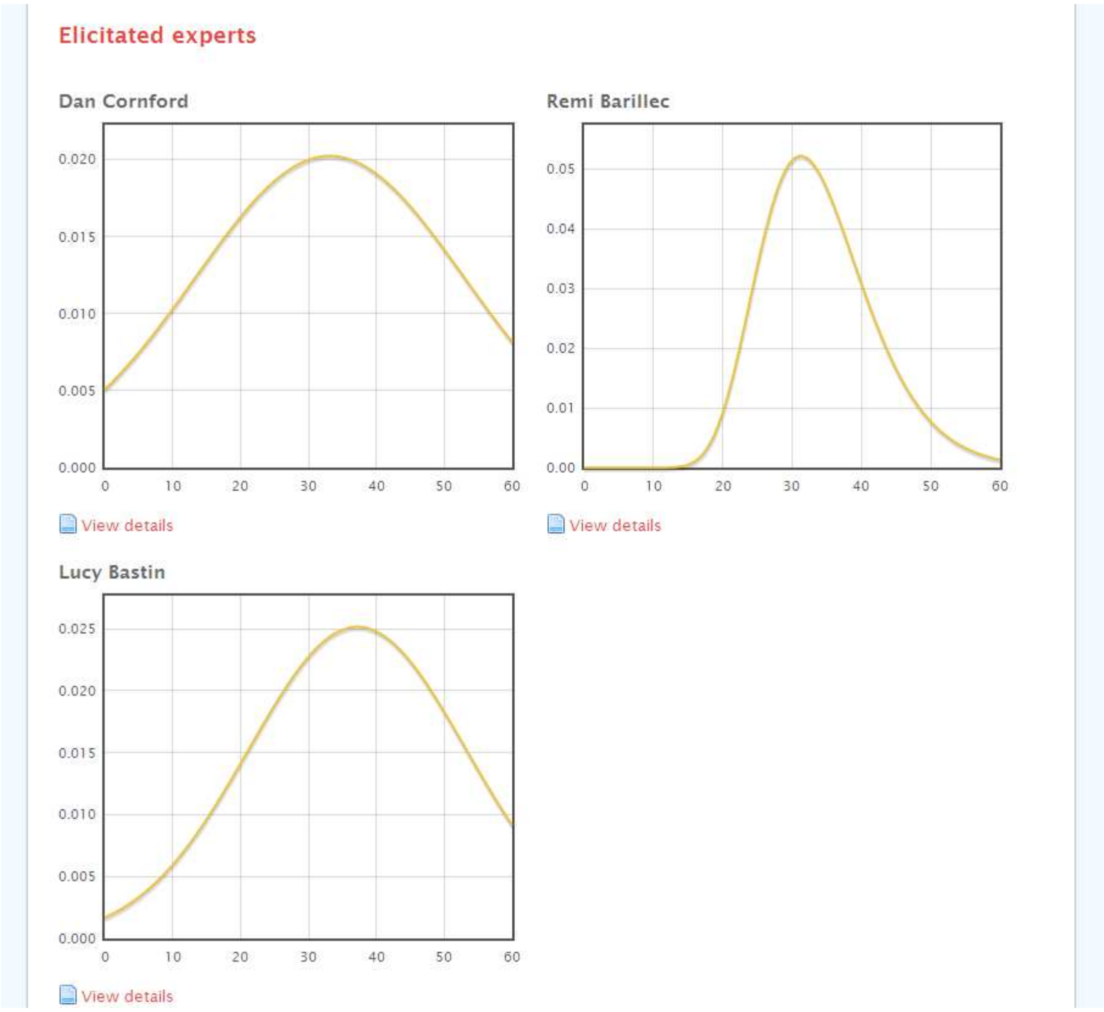


Figure 24: Each expert's distribution is displayed below the pooled result.

For each distribution you can view the exact details again by clicking on the ‘View details’ link.

3.6 Future work

The tutorial outlined in this section has given a good indication of the functionality of The Elicitator. However, there are still several features that could be implemented to improve functionality. Below is a list summarising future improvements:

- provide more detailed information about experts and allow a user to search the current collection of experts,
- problem owners should be able to remove an expert, or weight an expert for a given elicitation task. This could be done based on the expert’s perceived expertise for the given variable,
- deadlines could be added to each task to help expert’s manage their time,
- more qualitative information about the expert’s knowledge, background and thought process should be captured during the elicitation process. This would help guard against some of the biases that arise through the questioning process,
- the email system should be implemented to keep problem owners and experts informed of the current state of the elicitation process. For example, a problem owner could be informed when an expert has successfully completed an elicitation in their problem.

4 Background material for expert elicitation of spatial variables

4.1 Introduction

This chapter gives a detailed explanation of the web-based tool which executes the elicitation procedure of characterising the distribution and correlation structure of continuous spatial variables. Both the elicitation procedure to characterise the marginal probability distribution and that of the variogram of a stationary random function will be described. The technical description of the tool together with a detailed explanation about how the tool works is also given.

The chapter includes three main sections:

- Section 4.2: detailed elaboration of the elicitation procedure,
- Section 4.3: description of the web-based tool,
- Section 4.4: how to carry out expert elicitation by the tool (note that this is explained in more detail in Chapter 5).

4.2 General explanation of the elicitation procedure

The elicitation procedure is an iterative process which includes two main rounds: (1) elicitation of the marginal distribution (Figure 25), and (2) elicitation of the variogram (Figure 26). Figure 27 shows how both rounds are executed consecutively.

4.2.1 Elicitation of the marginal distribution

In this round, each expert will be asked to judge the marginal distribution of the stationary random function. The procedure used is similar to that described in Chapter 2. Each expert separately provides judgements by individually answering the questionnaires provided as web-forms. Figure 25 outlines the process.

Each expert will judge the marginal probability distribution of the target variable at a randomly chosen location in the study area. Note that the marginal distribution is spatially invariant because of the stationarity assumption. Each expert will be first asked to estimate the minimum and the maximum value of the variable and next the three quartiles (the upper quartile, the lower quartile and the median) that characterize its probability distribution (see section 4.3.1 for the questionnaires). These probability judgements from each expert about the marginal probability distribution will be checked on whether they satisfy the rule that the quartiles are increasing, greater than the minimum and smaller than the maximum. Also, the probability mass corresponding with each of the intervals between quartiles must be equal. Questions for an internal consistency check are also given (see section 4.3.1). If the check fails then the expert is asked to revise one or more judged values until all conditions are satisfied. The proper quartiles then go through a process of fitting the marginal probability distribution function. Under the assumption of stationarity, the random function has the same marginal probability distribution everywhere within the research area, with constant mean and variance. We further assume that the probability distribution of the variable at any location can be sufficiently defined by either the normal or lognormal distribution. The test for normality is based on a diagnosis of the skewness of the distribution. We used the Bowley coefficient of skewness γ (Bowley, 1920) as a diagnosis of normality:

$$\gamma = \frac{Z_{0.75} + Z_{0.25} - 2 \times Z_{\mathrm{med}}}{Z_{0.75} - Z_{0.25}} \tag{1}$$

where $Z_{0.25}, Z_{0.75}, Z_{med}$ are the lower quartile, the upper quartile and the median, respectively.

The normal distribution is a symmetric distribution for which the coefficient of skewness equals zero. However, in expert elicitation, judged values of experts can only reach a certain level of precision of about 0.05 (O'Hagan et al., 2006). Therefore, we define a threshold for γ, called γ_t. When $|\gamma| < \gamma_t$, the distribution is assumed normally distributed. When $|\gamma| > \gamma_t$, the distribution is either a positive skew distribution or negative skew distribution. In this work, we chose $\gamma_t = 0.05$. When the threshold is exceeded the lognormal distribution is assumed, i.e. it is assumed that the natural log of the target variable is normally distributed. Three-parameter lognormal distributions with a location or shift parameter (Cohen and Whitten, 1980) are not included.

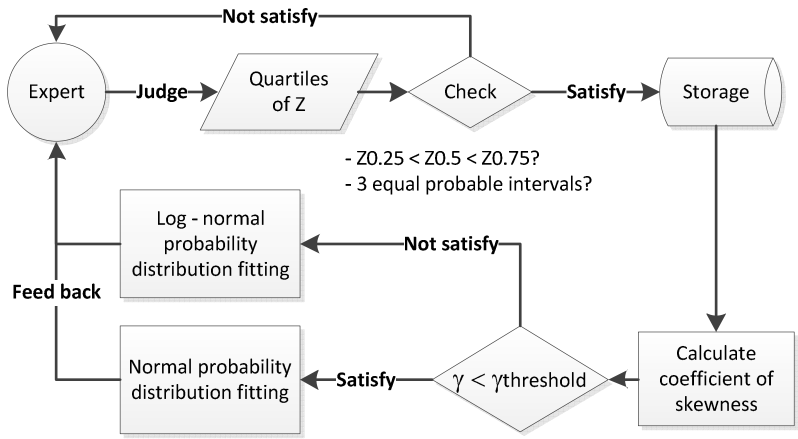


Figure 25: Flowchart of iterative procedure for elicitation of the marginal distribution.

The fitted marginal probability distribution will be reported back to each expert. Experts can reflect and revise their judgements until they satisfy that their judgements are correctly conveyed in the given feedbacks.

As the marginal distribution of the random function (normal or lognormal) will influence the next round of variogram elicitation, experts have to reach a consensus about the distribution of the random variable at this point, before proceeding to the next round. Section 4.2.3 details how consensus amongst experts is obtained.

4.2.2 Elicitation of the variogram

Depending on the result of the first round, that is whether the random variable has a normal or lognormal distribution, the next step of eliciting the variogram will be different. Figure 26 outlines the process.

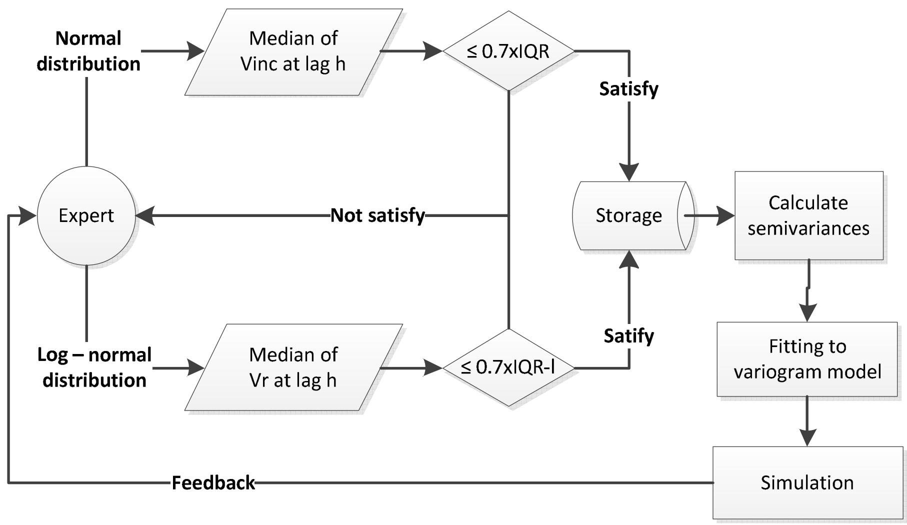


Figure 26: Flowchart of iterative elicitation procedure for the variogram.

Multivariate normal distribution

Each expert will be asked to judge the medians of the absolute first increments: $V_{inc}(|h|) = |Z(x+|h|) - Z(x)|$ of the random variable's value for several distance lags in the given area. The distance lags are defined by estimating seven proportions of the central axis through the area. The maximum distance for which we elicit the median of the difference is defined as half of the diagonal (D) of the area:

$$D = \sqrt{(x_{max} - x_{min})^2 + (y_{max} - y_{min})^2} \tag{2}$$

where x_{max} ,x_{min}, y_{max}, y_{min} define the extent of the area.

For convenience of experts when considering the distance, we round the maximum distance to the nearest number of type $k \times 10^x$ with k=1, 2 or 5 and an integer x. Next the distance is (approximately) halved by choosing the nearest smaller distance of the same type (e.g. if the initial distance was 5×10^3 then the next is 2×10^3, if it was 1×10^{-2} then the next is 5×10^{-3}). This is repeated to form smaller lags from the maximum lag until seven values of |h| are obtained.

The medians elicited from each expert will be used to calculate the semivariances, using the robust estimation procedure defined by Cressie and Hawkins (1980) and Cressie (1991, p. 75):

$$2\hat{\gamma}(|h|) = [\text{median}|Z(x+h) - Z(x)|]^2/0.457 \tag{3}$$

The constant 0.457 corrects for bias when using the median as robust estimation of the variogram instead of the mean.

Moreover, external checking is also required. For the stationary Gaussian stochastic process, the random variable at each location has the same marginal normal distribution with constant mean and variance. The semivariance cannot be larger than the variance (σ^2) of the marginal distribution since $\gamma(h) = \sigma^2(1-\rho(h))$, $0 < \rho(h) < 1$. Therefore, the medians judged by experts must satisfy the following condition:

$$V_{inc\_med} \leq 0.709 \times (Z_{0.75} - Z_{0.25}) \tag{4}$$

Proof:

$$\hat{\gamma} = \frac{V_{inc\_med}{}^{2}}{2\times 0.457} \leq \sigma^2 \Leftrightarrow V_{inc\_med} \leq \sqrt{2\times 0.457}\times\sigma$$

$$\Leftrightarrow V_{inc\_med} \leq \sqrt{2\times 0.457}\times\frac{Z_{0.75}-Z_{0.25}}{1.349} = 0.708698\ \times(Z_{0.75}-Z_{0.25})$$

After the variogram values at the seven distance lags are accepted, the procedure continues with fitting a variogram model (Section 4.3.3). The fitted variogram model is then used to sample from the spatial distribution of the variable, using unconditional sequential Gaussian simulation (Goovaerts, 1997). The variation in simulated values of the random variable along a transect within the study area are shown to the expert (without defining the location of the transect, since this is immaterial given that the random function is stationary). The values of the maximum and the minimum are used to replace all simulated values outside these bounds (i.e. larger than the maximum or smaller than the minimum). Experts can reconsider whether the spatial structure shown along the transect conveys what they think it should be. If not, they can revise their judgements about the medians of increments at lags and the variogram elicitation is reiterated.

Multivariate lognormal distribution

When the multivariate distribution of the random variable is lognormal, each expert will be asked to judge the median of the absolute ratio of change in its value at two locations at distances $|\mathrm{h}|$:

$$V_r = \left|\frac{\mathrm{Z(x+h)}}{\mathrm{Z(x)}}\right|, \text{assuming that } |Z(x+h)| > |Z(x)|$$

The median of the log – transformed difference is calculated by:

$$\mathrm{median}\{|\log(\mathrm{Z(x+h)}) - \log(\mathrm{Z(h)})|\} = \mathrm{median}\left(\left|\log\left[\frac{\mathrm{Z(x+h)}}{\mathrm{Z(x)}}\right]\right|\right) = \log\left(\mathrm{median}\left|\frac{\mathrm{Z(x+h)}}{\mathrm{Z(x)}}\right|\right) \quad (5)$$

This equation is satisfied as $\left|\frac{\boldsymbol{Z(x+h)}}{\boldsymbol{Z(x)}}\right| > 0$**.** The lag distances are defined in the same way as in the multivariate normal case. Externally checking, the elicited medians need to satisfy the condition:

$$\log(\mathrm{V_{r\_med}}) \leq 0.709 \times [\log(\mathrm{Z_{0.75}}) - \log(\mathrm{Z_{0.25}})] \quad (6)$$

Or:

$$\mathrm{V_{r\_med}} \leq \left(\frac{\mathrm{Z_{0.75}}}{\mathrm{Z_{0.25}}}\right)^{0.709} \quad (7)$$

The semivariances are estimated by Eq. (3) using the median of absolute increment of log–transformed values in Eq. (5). The remaining steps are the same as in the case of the multivariate normal distribution.

4.2.3 Pooling multiple experts' judgements

The purpose of judgement pool or opinion pool is to combine individual judgements about the quantity of target variable to form a single pooled judgements about that quantity. We follow the linear opinion pooling method in which after individually giving judgements, all experts' judgements are combined by applying a linear opinion pool (e.g. equal weighted average). This has already been explained in Chapter 2. Amongst available mathematical combination methods, the average pooling method is the simplest. In most practical cases, this method performs as well as more complex methods (Clemen and Winkler, 1999). Figure 27 indicates where in the elicitation procedure the judgement pooling needs to be done.

Pooling of the marginal distribution

As explained in Section 4.2.1, experts need to reach agreement about the marginal probability distribution of the random function before they continue with the next elicitation round. Therefore, after initial individual elicitation of the marginal distribution, the distributions from all experts are pooled. The task is done by first generating (many) quantiles from the fitted probability distributions of all experts. Next, all of these are linearly combined, in other words an equal weighted average is taken. The coefficient of skewness of the combined distribution is calculated to again diagnose normality. The resulting quantiles after mathematical combination are then fitted again to the chosen probability distribution (i.e. the normal or lognormal). The minimum value of the pooled probability distribution is taken as the minimum of all minimum values from all experts. The maximum value is the maximum of all experts' maximum values.

The consensual result is reported back to each expert, giving them a chance to reconsider and revise their individual judgements of quartiles. The process continues until all experts are satisfied with the final consensual results and stop changing their judgements. At this point, the elicitation process of the marginal distribution is ended. In practice, it may be sensible to allow just a single round of revision.

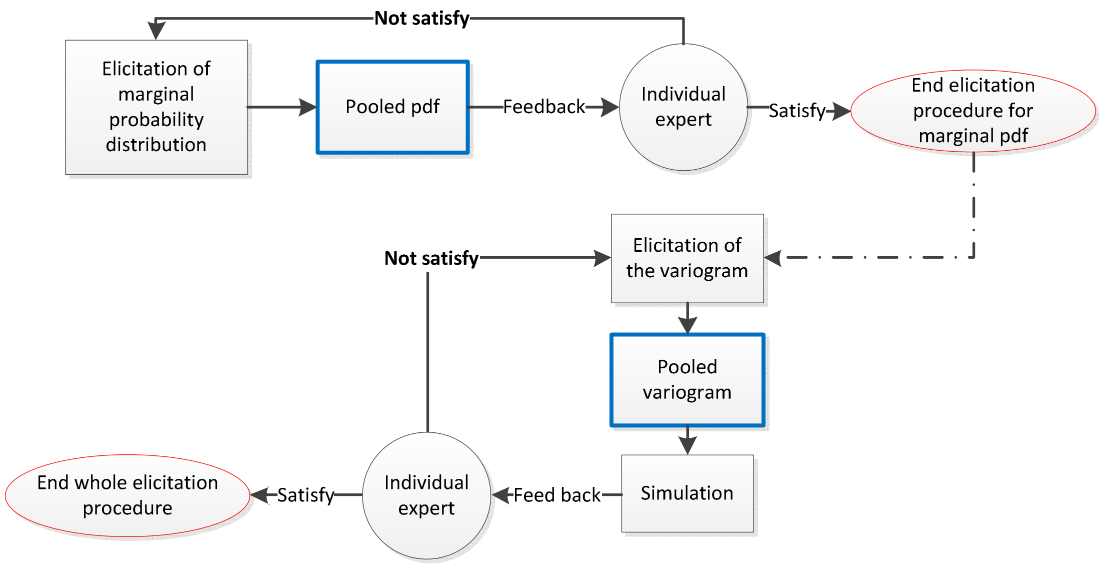


Figure 27: Flowchart of the whole iterative elicitation procedure.

Pooling of the variogram

According to the type of distribution of the random variable, which is determined by the consensual combination of all experts' distributions as explained above, all experts proceed to the next step of variogram elicitation (Section 4.2.2). After the individual variogram elicitation process is done, all judgements of experts again are pooled using equal weight averaging. The medians elicited at the seven lags of all experts are combined to obtain the average for each lag:

$$Med_c(|h_m|) = \frac{1}{k}\sum_{i=1}^{k} median_i(|h_m|) \tag{8}$$

where the lag number m = 1 ... 7 and k is the number of experts.

Next the seven consensual values of the medians are used to calculate the semivariances with Eq. (3). These semivariances are used to fit the variogram model that characterizes the spatial correlation of the target variable and used for stochastic simulation. The pooled medians at seven lags and the transect of simulated values are fed back to experts. They can make changes to the medians of the increments (or ratios – in case of the lognormal distribution) at the seven lags but they cannot change the quartiles of the marginal distribution. The whole elicitation procedure ends when all experts stop revising their judged values for the medians. Again, in practice perhaps only a single round of revision may be used.

4.3 Description of the web–based tool

The general structure of the computer tool for expert elicitation of the variogram has four main components (Figure 28), besides complementary documents. These components are discussed hereafter.

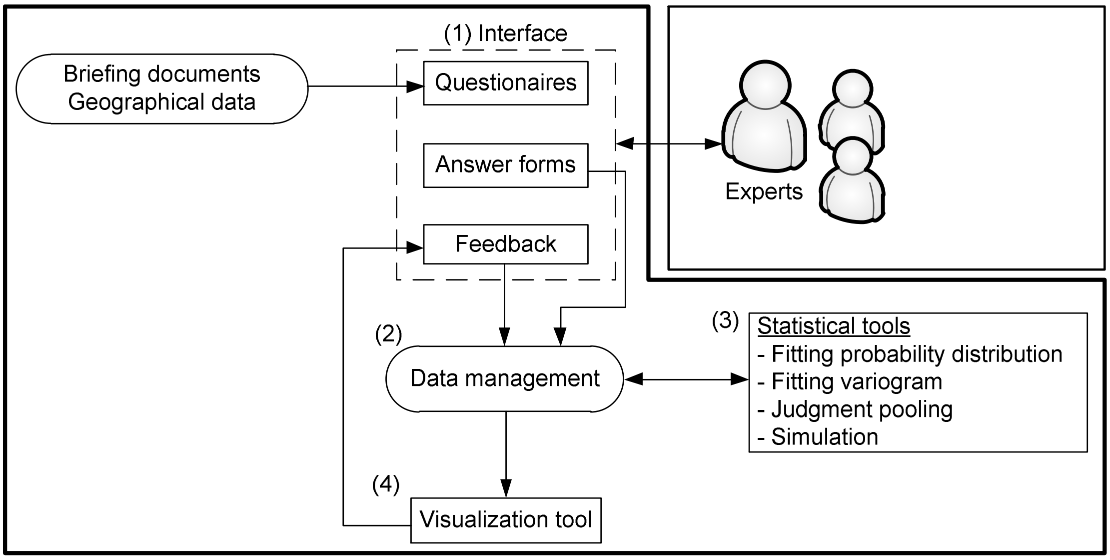


Figure 28: Schema of the main components of the variogram elicitation tool.

4.3.1 Web elicitation interface

Each expert interacts with the web elicitation interface individually to proceed through the elicitation procedure in an automatic manner. The web interface provides information about context and target object, the elicitation procedure, useful guidelines for properly giving probabilistic judgements and all other constraints or conditions of the elicitation procedure (see Chapter 2, also).

Several main pages include a description of the study area, a briefing document and questionnaires.

Description of study area

A map and information about the study area and information about the target variable are given on this pages. Experts have to carefully go through these pages.

Briefing documents

These documents give supplementary information, including:

- statement of studying purpose, nature of problem and usage of elicitation result,
- description of elicitation procedure,
- explanation of eliciting techniques,
- explanation of the probabilistic summaries,
- statement of requirements from the experts,
- causes of biased judgements and illustrating examples,
- recommended literature about the problem (if available).

Questionnaires

The aim of the questionnaire is to extract from the expert the plausible range (maximum and minimum) and quartiles (first, second and third quartile) of the marginal distribution of the target variable. This questionnaire was developed based on the quartile elicitation framework of SHELF (2010) for individual elicitation using the bisection method, as described in Chapter 2. However, we modified these slightly to develop questionnaires by using difference expressions in questioning probability. Different ways of formulating questions to different groups of experts likely yields better results (Pat and Schrage, 2003). Thus, we distinguish between a quantitative and qualitative form. The quantitative form uses a numerical expression of probability, while the qualitative expression uses a verbal description of probability. The latter is expected to be more suitable for non–mathematicians and non–statisticians. Tables 1, 2 and 3 provide both types of questions.

Table 1: Questionnaire for eliciting the marginal distribution of the target variable

Qualitative expression	**Quantitative expression**
Purpose of question: Eliciting the plausible range [A_{min} , M_{ax}] of spatial random function Z. Must allow experts to answer -∞ or +∞ (i.e. there is no minimum or maximum). Question 1 – Which is the smallest possible value (A_{min}) of Z? Question 2 – Which is the largest possible value (M_{ax}) of Z?	
Purpose of question: Eliciting from experts the median Z_{ed}	
Question 3 – Could you determine a value such that the value of Z is equally likely to be less than or greater than this value?	Question 3 – What is the value of Z_{ed} such that there is 0.5 probability that the value of Z is equal or less than this value? $\Pr(Z \leq Z_{med}) = 0.5$
Purpose of question: Eliciting from experts the lower quartile $Z_{0.25}$	
Question 4 – If $Z_{0.25}$ is smaller than $Z_{med,}$ could you determine $Z_{0.25}$ that within the interval [Z_{min}, Z_{med}] the value of Z is equally likely to be less than or greater than this point?	Question 4 – What is the value of $Z_{0.25}$ such that the probability that the value of Z is equal or less than this value within the interval [Z_{min}, Z_{med}] is 0.5? $\Pr(Z \leq Z_{0.25}) = 0.25$
Purpose of question: Eliciting from the experts the upper quartile $Z_{0.75}$	
Question 5 – If $Z_{0.75}$ is larger than $Z_{med,}$ can you determine its value such that within the interval [Z_{med}, Z_{max}] the value of Z is equally likely to be less than or greater than $Z_{0.75}$?	Question 5– What is the value of $Z_{0.75}$ such that the probability that the value of Z is equal or less than this value within the interval [Z_{med}, Z_{max}] is 0.5? $\Pr(Z \leq Z_{0.75}) = 0.75$

Qualitative expression	**Quantitative expression**
Purpose of question: Checking for experts' internal consistency Question 6 – Is there any interval amongst the four [Z_{min}, $Z_{0.25}$], [$Z_{0.25}$, Z_{med}], [Z_{med}, $Z_{0.75}$], and [$Z_{0.75}$, Z_{max}] more likely or probable than any other?	

Table 2: Questionnaire for eliciting the median of the absolute first increment

Qualitative expression	**Quantitative expression**
Purpose of question: Eliciting from the experts the median of the absolute first increments (V_{inc_med})	
Question – Could you determine a value such that the absolute difference V_{inc} of attribute Z at two locations that are a distance $\lvert h \rvert$ apart is equally likely to be less than or greater than this value?	Question – What is the value of V_{inc_med} such that the probability that the value of $V_{inc}=\lvert Z(x)-Z(x+h)\rvert$ is equal or less than this value is 0.5? $\Pr\left(V_{inc}(h) \le V_{inc\_med}\right) = 0.5$

Table 3: Questionnaire for eliciting the median of the ratio of differences

Qualitative expression	**Quantitative expression**
Purpose of question: Eliciting from the experts the median (V_{r_med}) of the ratio of change	
Question – Suppose that we observe Z at some location and another location a distance |h| from it. Next we take the ratio of these two observations, while putting the largest of the two in the numerator. Could you determine a value such that the ratio V_r is equally likely to be less than or greater than this value?	Question – Suppose that we observe Z at some location and another location a distance |h| from it. Next we take the ratio of these two observations, while putting the larger of the two in the numerator. What is the value of V_{r_med} such that the probability that the ratio V_r is equal or less than this value is 0.5? $\Pr\left(V_r(h) \le V_{r\_med}\right) = 0.5$

4.3.2 Database

In order to allow experts to reassess their judgements during the elicitation procedure and for executing other tasks related to the analysis of judgements, the information extracted from experts is stored in a database. The database includes the final individually judged values of the minimum, maximum, the three quartiles and the medians at the seven lags. The aggregated values of these parameters are also stored.

4.3.3 Statistical tools

Three main statistical tools are required. The first concerns fitting functions used to fit the probability distribution and the variogram. Second, an aggregation tool combines multiple judgements from experts. The third tool is used to simulate realizations of the target spatial variable along a transect. These statistical tools or functions were all built in R and R-gstat (Pebesma, 2004), partly by reuse and adaptation of available functions in SHELF (2010) and the Automap R – library: Automatic interpolation package (Hiemstra et al., 2009). All statistical tools are programmed in R and assembled in the library *eevariogram*. The

Appendix provides the code of all functions. The fitting and pooling tools are also used by The Elicitator for elicitation of univariate distributions of continuous and categorical variables (Chapters 2 and 3).

Fitting the probability distribution function is a parametric fitting to an underlying normal or lognormal distribution. The ordinary least squares approach was used to optimally fit the parameters of the probability distribution from given experts' judgements (the quartiles). The mean (μ) and the variance (σ^2) are chosen by numerically minimizing:

$$[F(Z_{0.25};\mu,\sigma^2)-0.25]^2+[F(Z_{med};\mu,\sigma^2)-0.5]^2+[F(Z_{0.75};\mu,\sigma^2)-0.75]^2 \qquad (9)$$

where F is the normal or lognormal cumulative distribution function. The fitting function used in this tool is adapted from the one used in SHELF (2010).

Ordinary least squares optimization with equal weights is also used to fit experts' judgements to valid variogram models. We utilize the automatic fitting function provided in the Automap library, named *autofitvariogram* which automatically chooses the best fitting model based on minimizing the sum of squared errors. The chosen variogram models include the Nugget, Exponential, Spherical, Gaussian and Matèrn models. No nested models are fitted, except for the combination of the Nugget model with any of the others. The initial parameters of the variogram model are defined as the default in the *autofitvariogram* function. The fitting process is automatic so that experts cannot interfere in this stage.

Unconditional simulation is used to plot spatial variation of the random variable along a transect. The *gstat* function of the Gstat package is used to simulate realizations of the target variable along the transect. The length of the transect is chosen equal to the central axis through the study area. The transect is divided into a sufficiently large number of intervals and simulations are generated at the interval boundaries. Several simulations are generated and experts can toggle between these to get an impression of the whole range of possible realities. Since experts have specified the maximum and minimum value of the random variable, simulated values exceeding the bound will be replaced by one of the bound values.

4.3.4 Visualization tool

Feedback to experts are given through a web interface by means of graphs of probability density function and simulations along a transect. The visualization tool must provide these graphs through the web. Some other maps of the area are also needed to give support to experts to explore the area.

5 Tutorial expert elicitation of spatial variables

In this tutorial, we will be learning to use the web-based tool for expert elicitation of the spatial structure of environmental variables over a specified geographical area. In order to characterize the spatial structure of a continuous variable, the spatial probability distribution of the variable must be specified. As explained in the previous chapter, two measures must be elicited from the experts: the marginal probability distribution of the variable at an arbitrary location and the spatial autocorrelation (the variogram). To this end, the tool has two main rounds corresponding to elicitation procedures for each of these two measures.

The tool can be accessed through the website http://www.variogramelicitation.org/. The first page (Figure 29) provides a general introduction about the purpose of the tool and the tasks that the experts need to do. To complete an elicitation exercise, the experts need to complete four main tasks:

1. read the case study description,
2. read the briefing document,
3. elicit the marginal distribution of the environmental variable,
4. elicit the variogram of the environmental variable.

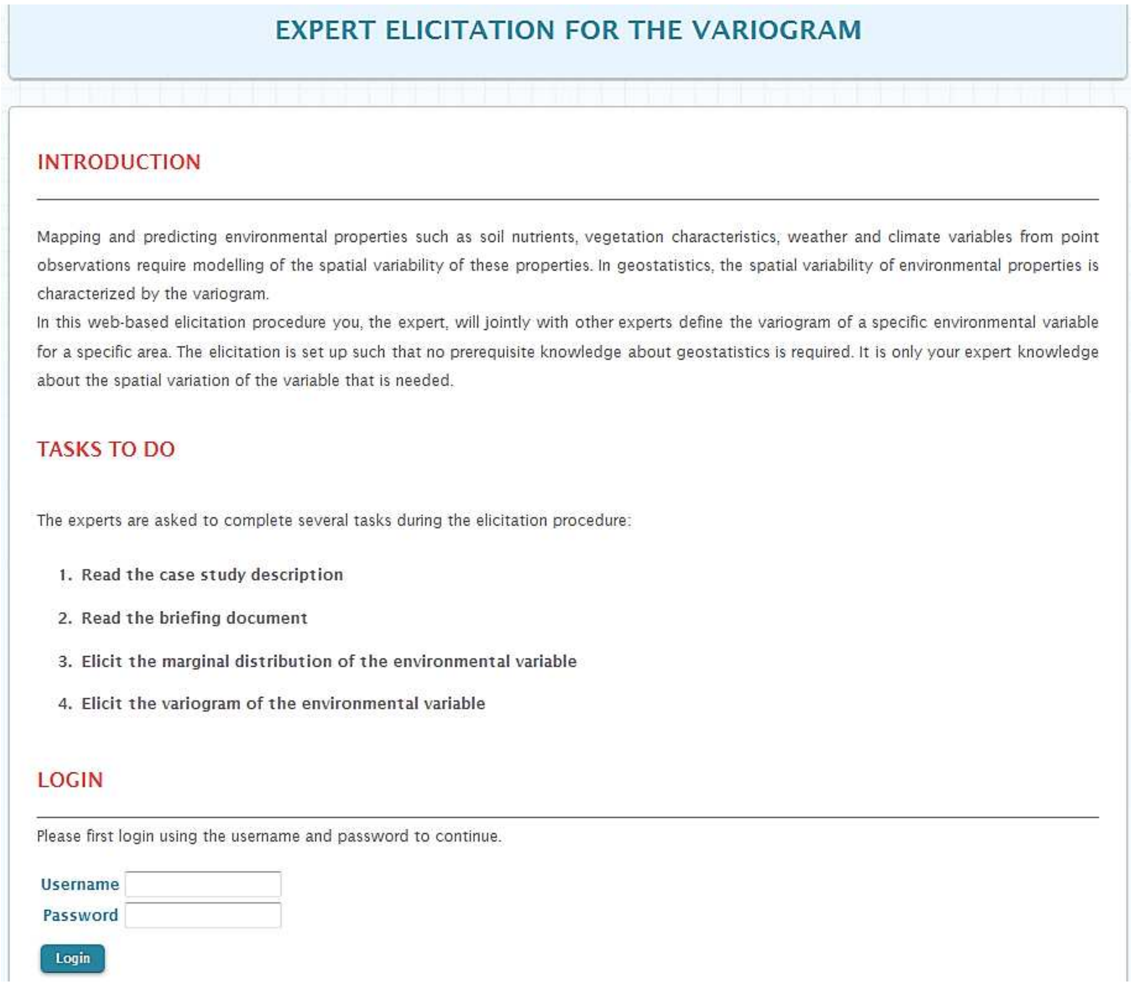


Figure 29: Login page of the variogram expert elicitation tool.

The tool was designed to guide the experts through the elicitation process without requirement of much knowledge of statistical probability. As a first prototype, the tool can be used only by experts to provide their judgements about the spatial pattern of the continuous variables. The ability to administer the tool by problem owners is expected to be the same as the Elicitator, i.e. the tool use for elicitation of continuous and categorical variables (see chapters 2 and 3). In a next phase the variogram elicitation tool will be integrated into The Elicitator (http://elicitator.uncertweb.org/).

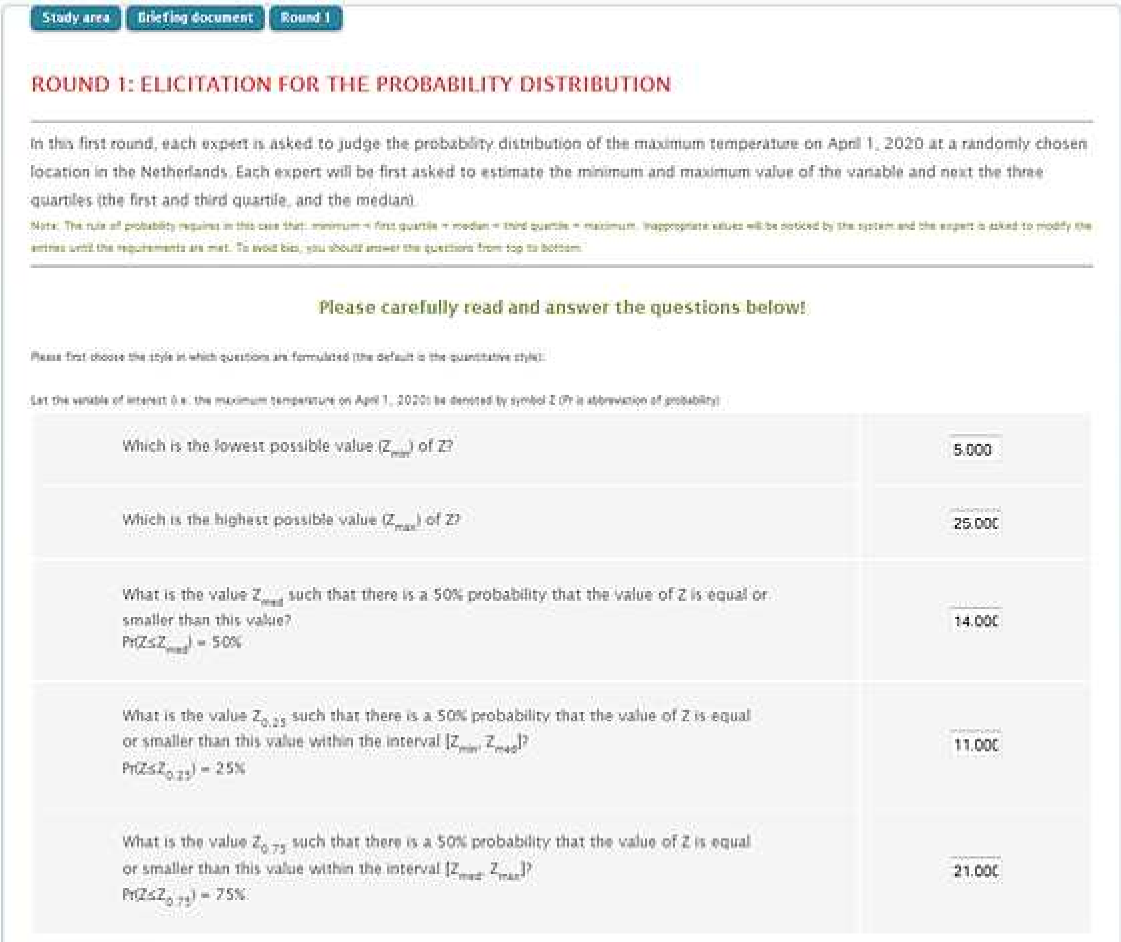
Using the tool to give expert judgements is quite simple. The experts first have to login using their username and password. After login, the experts will choose the case study that they take part in. The case studies can be restricted to a group of experts. As mentioned above, there are two main rounds. However, before getting started with these rounds, the experts have to carefully go through two preliminary sections Study area Briefing document by clicking on the two buttons at the top of the main page or can click on Click to continue to go forward.

5.1 First round: elicitation of the marginal probability distribution

Before proceeding to answer questions in round 1, the experts are asked to carefully read the case study description, including a description of the study area and the target variable. Because the experts will judge the spatial correlation of the variable over a certain area, it is important that the experts get familiar to the study area, its geographical location and extent. An embedded Google Map (http://maps.google.com/), which is located at the study area, allows the experts to interactively explore the study area. In addition, the experts should be well aware of the causes that can lead them to give biased judgements. They should read the briefing document. The briefing document provides the experts with useful information about:

- statement of studying purpose, nature of problem and usage of elicitation result,
- description of the elicitation procedure,
- explanation of elicitation techniques,
- explanation of the probabilistic summaries,
- statement of requirements from the experts,
- causes of biased judgements and illustrating examples,
- recommended literature about the problem (if available).

In round 1, the question form (Figure 30) is used to elicit quantitative judgements from the experts. The experts are asked to fairly judge the possible maximum and minimum values, and the three quartiles (the first quartile, the median and the third quartile) of the probability distribution of the variable. Quantitative judgements of the experts need not be more precise than three digits. Useful explanation is given to guide the experts to answer the questions. The experts' judgements have to satisfy some conditions (see Chapter 4); otherwise, notices of inappropriate judged values will be given. Unless the experts change their judgements, they cannot go forward to the feedback page.



Study area | Briefing document | Round 1

ROUND 1: ELICITATION FOR THE PROBABILITY DISTRIBUTION

In this first round, each expert is asked to judge the probability distribution of the maximum temperature on April 1, 2020 at a randomly chosen location in the Netherlands. Each expert will be first asked to estimate the minimum and maximum value of the variable and next the three quartiles (the first and third quartile, and the median).

Note: The rule of probability requires in this case that: minimum < first quartile < median < third quartile < maximum. Inappropriate values will be noticed by the system and the expert is asked to modify the entries until the requirements are met. To avoid bias, you should answer the questions from top to bottom.

Please carefully read and answer the questions below!

Please first choose the style in which questions are formulated (the default is the quantitative style):

Let the variable of interest (i.e. the maximum temperature on April 1, 2020) be denoted by symbol Z (Pr is abbreviation of probability)

Which is the lowest possible value (Z_{min}) of Z?	5.000
Which is the highest possible value (Z_{max}) of Z?	25.000
What is the value Z_{med} such that there is a 50% probability that the value of Z is equal or smaller than this value? $Pr(Z \leq Z_{med}) = 50\%$	14.000
What is the value $Z_{0.25}$ such that there is a 50% probability that the value of Z is equal or smaller than this value within the interval $[Z_{min}, Z_{med}]$? $Pr(Z \leq Z_{0.25}) = 25\%$	11.000
What is the value $Z_{0.75}$ such that there is a 50% probability that the value of Z is equal or smaller than this value within the interval $[Z_{med}, Z_{max}]$? $Pr(Z \leq Z_{0.75}) = 75\%$	21.000

Figure 30: Question form to elicit the probability distribution at round 1.

When all of the experts' judgements are coherent, the experts can proceed to see the probability distribution fitted based on their judged values. The experts' judgements are fitted to one of two types of probability density function: normal or lognormal probability density function. Feedback to the experts is a graph of the fitted probability density function and information about the probability distribution type and the three fitted quartiles (Figure 31). Their own judged values of the three quartiles that help the experts to compare the difference between their own judgements and the fitted values are shown as well. Moreover, the 10^{th} and 90^{th} percentiles are also provided as part of the feedback.

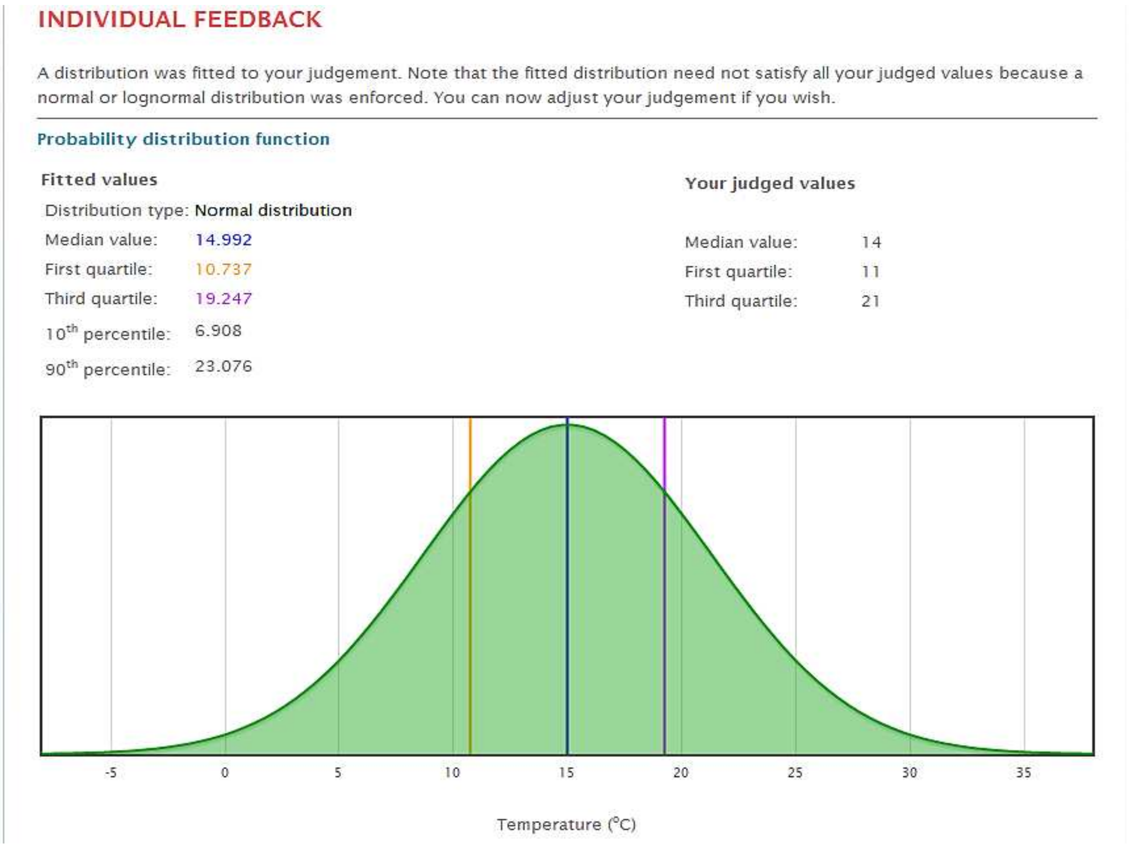


Figure 31: Feedback for round 1.

The experts can change their judgements if they think that the feedback does not correctly reflect their belief on the marginal probability distribution. The experts can review their previous judgements and change them in the revision form (Figure 32) below the feedback. Changes in the experts' judgements are interactively incorporated with a change in feedback graph and fitted values when the experts confirm their change by clicking the Save changes button at the bottom of the revision form. Some advice is given for the experts to change their judgements. Additional buttons enable the experts to continue to the next step in the process or to go back and revise entries in earlier steps.

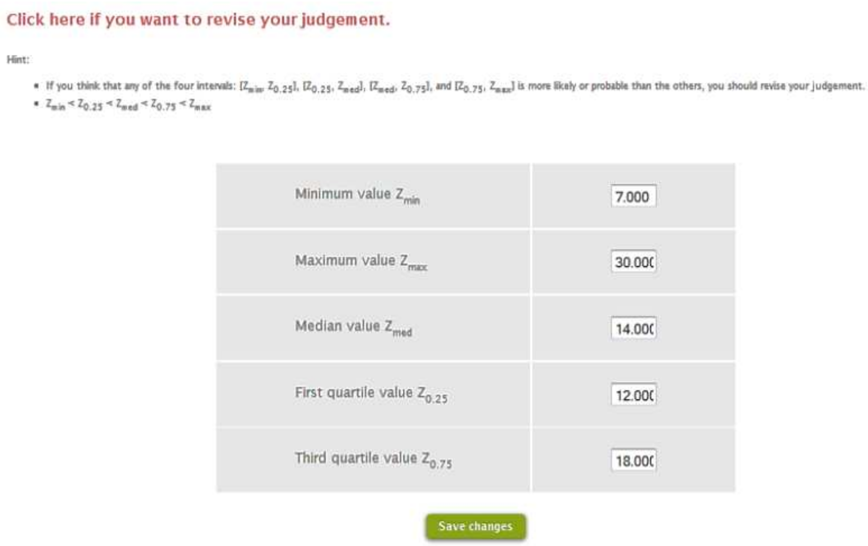


Figure 32: Feedback for round 1.

When the experts are satisfied that the feedback probability density function truly reflects their belief, they can go on to see the combined result of all other experts who finished the first round. By clicking on Click to see pooling result the experts can now see the pooled probability density function (Figure 33), that is the probabilistic average of all other experts' fitted probability density function together with their own function. At this stage, the experts can still go back to change their judged values which can influence the pooled result.

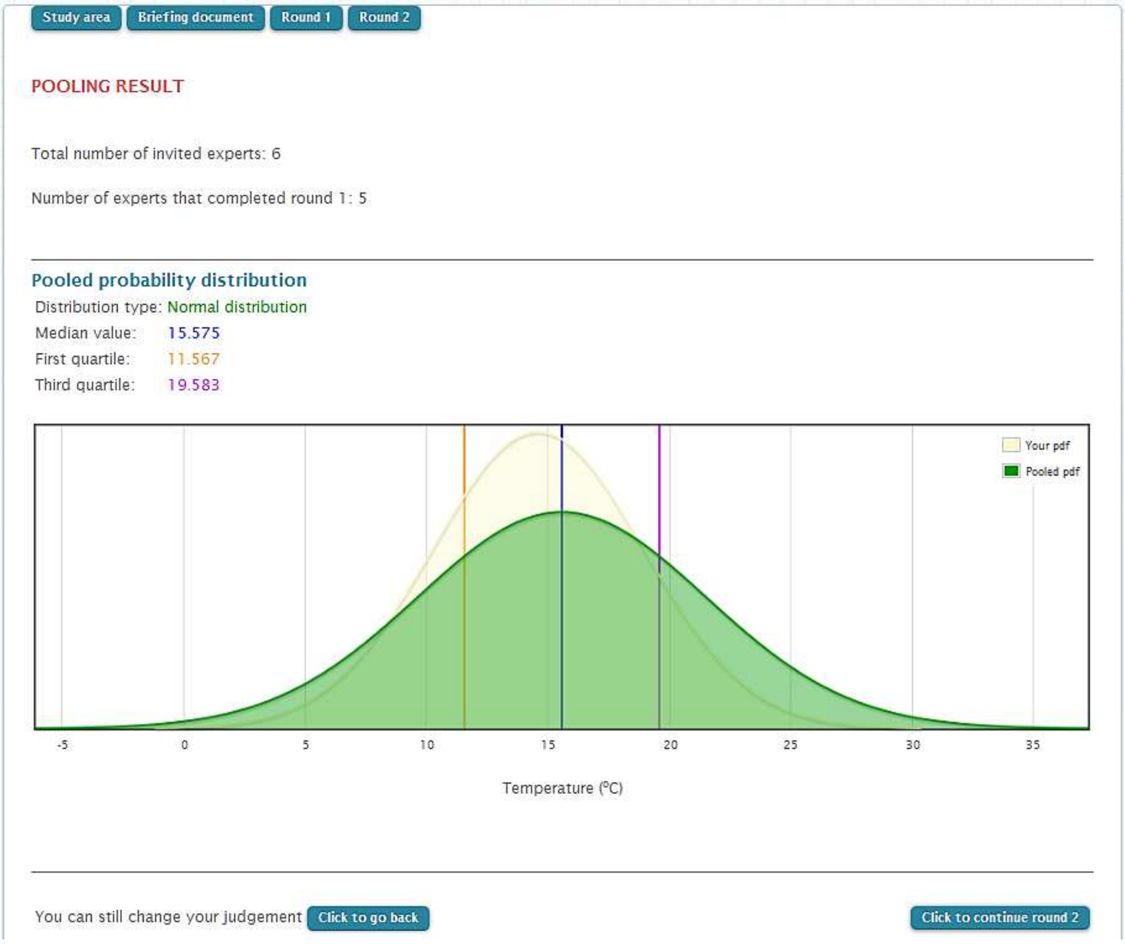


Figure 33: Feedback for the pooled result of round 1.

After the experts finish revising their judged values and leave round 1, they can relogin another day within a given duration period to change their judgements. After the limited duration for round 1 ends, they are no longer able to change their judgements for the marginal probability distribution. Instead, they will be invited to proceed to round 2.

5.2 Second round: elicitation of the variogram

Round 2 was designed to facilitate the experts giving judgements even when experts have no knowledge of geostatistics in general or the variogram specifically. To remind the experts about the pooled probability distribution resulting from the first ground, the second round starts with a short summary of the pooled probability distribution (Figure 34). The type of the pooled probability distribution, either normal or lognormal, defines the quantities that will be elicited from the experts in the second round. There is also a brief definition of these quantities. The experts now have a general idea of the quantitative measures that they are expected to be asked to judge. See Chapter 4 for a more detailed explanation of these quantitative measures.

ROUND 2: ELICITATION FOR THE VARIOGRAM

Depending on the result of the first round, that is whether the random variable has a normal or lognormal distribution, the next step of eliciting the variogram will be different.

1. If the pooled probability distribution is the normal distribution: Each expert will be asked to judge the medians of the absolute first increments.
   Again, let the variable of interest (i.e. the maximum temperature on April 1, 2020) be denoted by symbol Z. The absolute first increment is the absolute difference V_{inc} of the value of Z at two points located a certain distance h apart, regardless of its direction: $V_{inc}=|Z(x) - Z(x+h)|$.
2. If the pooled probability distribution is the lognormal distrbution: Each expert will be asked to judge the median of the absolute ratio of change V_r in the value of Z at two locations: $V_r=|Z(x)/Z(x+h)|$, assuming that $|Z(x)|>|Z(x+h)|$.

Information about pooled marginal probability distribution

Probability distribution:	Normal Distribution
Maximum value:	35.000
Minimum value:	5.000
Mean value:	15.706
Standard deviation:	5.507

Figure 34: Starting page of round 2.

Again, the question form was designed to elicit the defined quantities. Figure 35 shows the question form in case the pooled probability distribution is normal. The quantity to be elicited is the absolute difference between values at two locations a given distance apart.

Could you determine a value such that the absolute difference $V_{inc}=|Z(x) - Z(x+h)|$ of attribute Z at two locations that are a distance h apart is equally likely to be less than or greater than this value?

At distance 2000 meter apart	0.300
At distance 5000 meter apart	0.600
At distance 10000 meter apart	1.000
At distance 20000 meter apart	1.400
At distance 50000 meter apart	2.500
At distance 100000 meter apart	4.000
At distance 200000 meter apart	4.700

Figure 35: Question form for normal marginal probability distribution.

In cases where the pooled probability distribution is lognormal, the question is slightly more complicated (Figure 36). The quantity is now the ratio of the values at two locations (with the assumption that the ratio is always greater than one). Again see Chapter 4 for a more detailed elaboration of the differences between the two question forms.

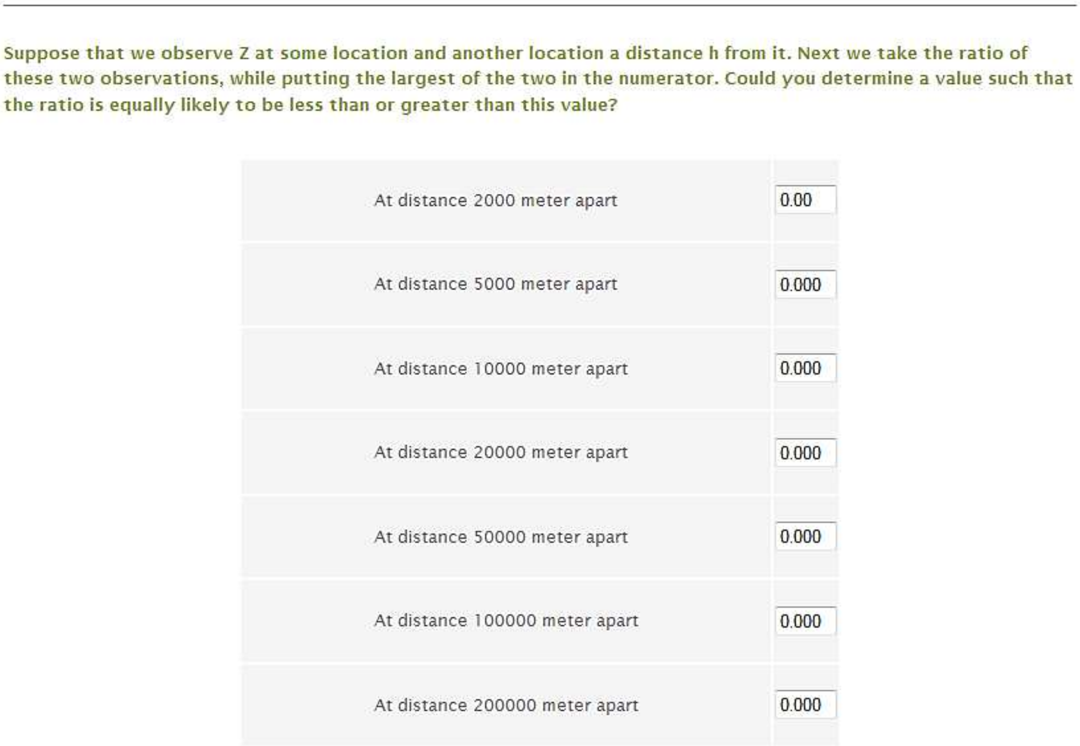


Figure 36: Question form for lognormal marginal probability distribution.

In both cases, the experts are required to carefully read the question forms. They should reconsider the study area and supplementary information about it and the target variable. Quantitative judgements of the experts need not be more precise than three digits. The question form also has a validation function to check whether the judged values satisfy some conditions (see Chapter 4). If all experts' judged values meet the conditions, the experts can proceed to the feedback page.

At this stage, the experts are recommended to go back to the briefing document page again, using the button at the top of the page. This helps the experts always to be aware of the unexpected causes of biased judgements.

Figure 37 shows a graph of simulated values along a transect within the study area. In order to produce this graph, the operations behind it include fitting the experts' judgements to a valid variogram model. The variogram model characterizes the level of autocorrelation of the variable over any distance in space. The fitting process automatically sets the initial values for the variogram parameters based on all experts' judged values. Next, it automatically fits the experts' judgements to different variogram models and selects the model that is considered as the best-fit in terms of having the smallest square error. In some cases the automatic fitting procedure may not be able to converge to a valid best-fit variogram. In this case, the experts receive a notice and a suggestion for a possible solution shows up as in Figure 38.

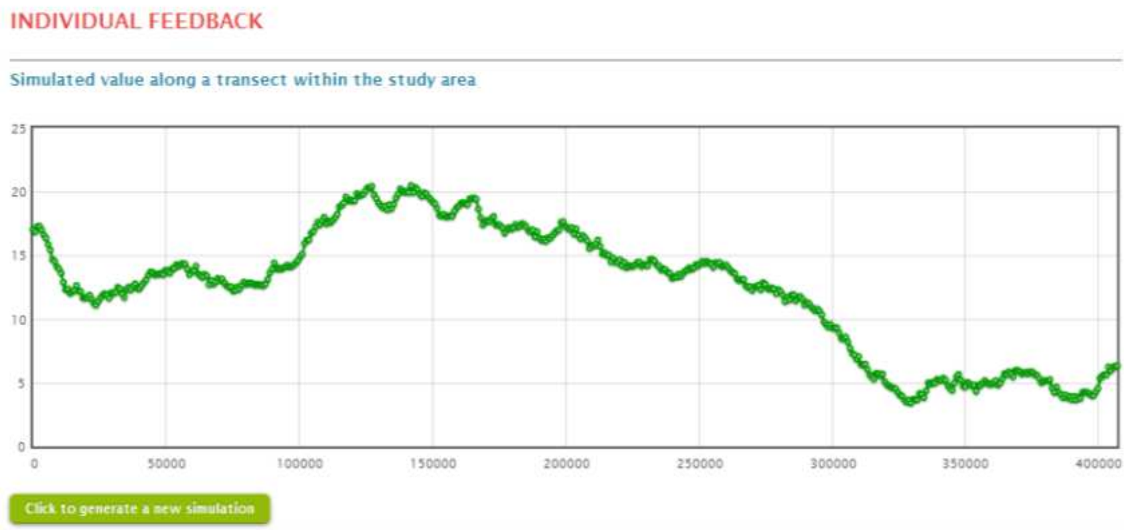


Figure 37: Individual feedback of round 2.

The variable values are unconditionally simulated (i.e. without conditioning to observations) along the transect by sampling from the multivariate probability distribution. This multivariate probability distribution is parameterized by the marginal probability distribution and the autocorrelation function (i.e. variogram) defined from the experts' judged values in round 1 and 2, respectively.

INDIVIDUAL FEEDBACK

Your judged values can not be automatically fitted to a valid variogram model. Please carefully consider again your judgments. You can change your judgements by clicking to go back. If you believe that your judgements are true, please send an email to phuong.truong@wur.nl, we will try to get back to you very soon with feedback .

Figure 38: Notice for failure of fitting expert judgements to a valid variogram in round 2.

From the graph of the transect (Figure 37), the experts can see how strongly the values of the target variable are spatially correlated. The differences of the values versus the distance shows the strength of the autocorrelation of target variables within the study area. In general, the greater the distance between the two locations, the greater the difference in the values. In other word, the further apart two locations in geographical space, the smaller the correlation. Two locations next two each other may still have difference in their value due to stochastic nature of the phenomenon. If the zero-distance-difference in the values is small, the transect will smoothly go up or down.

By looking at the feedback graph along the transect, the experts can recognize how the target variable varies in the study area. The experts should carefully observe the feedback graph to verify if the spatial variation of the target variable is in accordance with their belief.

There is always a possibility for the experts to revise their judgements. The revision form can be found by clicking on the link below the feedback graph. Every time the experts finish a revision, they should save their changes. By clicking on saving changes, the feedback graph is also updated and shows up.

Proceeding to the feedback page of the pooled result (Figure 39), the experts can also see again the results of their own judgements. The pooled feedback from all other experts who finished the round can be useful information for all other experts who did achieve a valid variogram fitted. For this reason, the experts, even those without a contribution to the pooled result can go forward to the pooled feedback page.

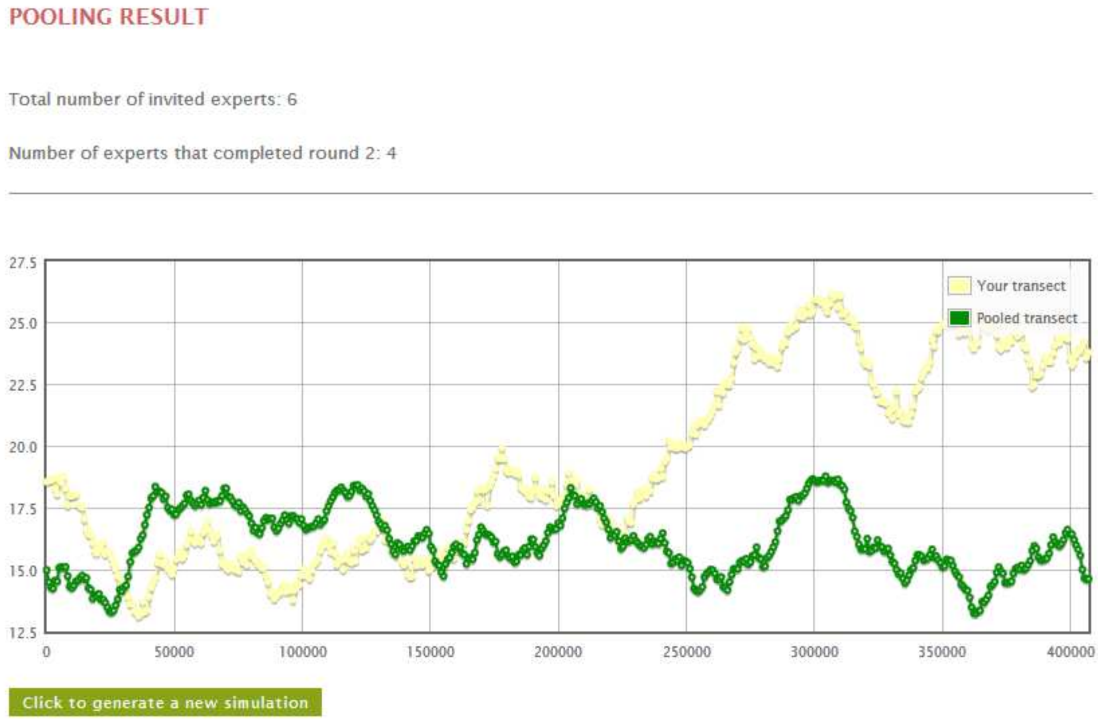


Figure 39: Pooled feedback of round 2.

When the experts decide to leave the tool, they can click on the button at the bottom of the pooling page to close the session.

6 References

The Sheffield Elicitation Framework (SHELF) and associated R routines are available from http://www.tonyohagan.co.uk/shelf/.

EXCALIBUR is the software produced by Delft that implements Cooke's classical method; it is freely available from http://ewi.tudelft.nl/index.php?id=23155&L=1.

Bacharach, M.,1972. Scientific disagreement. *Unpublished manuscript*, Christ Church, Oxford.

Bordley, R.F. and Kadane, J.B.,1999. Experiment-dependent priors in psychology and physics, *Theory and Decision*, **47**, 213–227.

Bowley, A.L., 1920. Elements of Statistics (London).

Clemen R., Winkler R., 1999. Combining probability distributions from experts in risk analysis, *Risk Analysis*, **19**, 187 – 203.

Cohen, A.C., Whitten B.J., 1980. Estimation in the Three-Parameter Lognormal Distribution, *Journal of the American Statistical Association*, 75, 399-404.

Cooke, R.M., 1991. Experts in Uncertainty: Opinion and Subjective Probability in Science. New York: Oxford University Press.

Cressie, N., Hawkins D., 1980. Robust estimation of the variogram, *Mathematical Geology*, **12**, 115-125.

Cressie, N.A.C., 1991, Statistics for Spatial Data (Wiley, New York)

De la Fuente, E., Trujillo, H., Ortega, A., Martin, I. and Estarelles, R., 1993. Assessment in probability judgement: a model for combining sources.,*The Statistician*, **42**, 561–570.

French, S., 2011. Aggregating expert judgement, *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales*, **105**, 181–206.

Garthwaite, P.H., Kadane, J.B. and O'Hagan, A., 2005. Statistical methods for eliciting probability distributions, *J. Amer. Statist. Ass.*, **100**, 680–701.

Genest, C. and Zidek, J.V., 1986. Combining probability distributions: a critique and an annotated bibliography, *Statistical Science*, **1**, 114-148.

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

Gosling, J.P., 2005. Elicitation: a nonparametric view. *PhD Thesis*. Department of Probability and Statistics, the University of Sheffield.

Hiemstra P.H., Pebesma E.J., Twenhöfel C.J.W., Heuvelink G.B.M., 2009. Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network, *Computers & Geosciences*, **35**, 1711-1721.

James, A., Low Choy, S. and Mengersen, K.L., 2010. Elicitator : an expert elicitation tool for regression in ecology, *Environmental Modelling & Software*, **25**, 129-145.

Johnson, S.R., Tomlinson, G.A., Hawker, G.A., Granton, J.T., Grosbein, H.A. and Feldman, B., 2010. A valid and reliable belief elicitation method for Bayesian priors, *Journal of Clinical Epidemiology*, **63**, 370-383.

Kadane, J.B., 1986. Progress toward a more ethical method for clinical trials, *Journal of Medicine and Philosophy*, **11**, 385–404.

Kadane, J.B. and Wolfson, L.J., 1998. Experiences in elicitation, *The Statistician*, **47**, 3–19.

Lindley, D.V., 1985. Reconciliation of discrete probability distributions, In: *Bayesian Statistics 2* (eds. J.M. Bernardo et al.), 375–390. Oxford: Oxford University Press.

O’Hagan A., Buck C., Daneshkhah A., Eiser J., Garthwaite P., Jenkinson D., Oakley J., Rakow T., 2006. Uncertain Judgements : Eliciting Experts' Probabilities (John Wiley & Sons, Ltd., Chichester)

O'Hagan, A., 1998. Eliciting expert beliefs in substantial practical applications, *The Statistician*, **47**, 21-35.

Patt A.G., Schrag D.P., 2003. Using Specific Language to Describe Risk and Probability, *Climatic Change*, **61**, 17-30.

Pebesma E.J., 2004. Multivariable geostatistics in S: the gstat package, *Computers & Geosciences*, **30**, 683-691.

Raiffa, H., 1968. Decision Analysis: Introductory Lectures on Choices under Uncertainty, 161–165. London: Addison-Wesley

Stone, M., 1961. The opinion pool, *Ann. Math. Statist.*, **32**, 1339-1342.

Thomas, E.A.C. and Ross, B.H., 1980. On appropriate procedures for combining probability distributions within the same family, *J. Math. Psych*., **21**, 136-152.

Tversky, A. and Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases, *Science*, **185**, 1124-1131.

Winkler, R. L., 1967. The assessment of prior distributions in Bayesian analysis, *Journal of the American Statistical Association*, **62**, 776-800.

7 Appendix: R code for expert elicitation of spatial variables

Fitting probability distribution function

```
### Modified from SHELF 2.0
# Main variables
# Lower bound: Lo
# Upper bound: Up
# Lower quartile: l
# Median: m
# Upper quartile: u

pdf.fitting <- function(Lo, Up, l, m, u){

#*Calculate coefficient of
skewness*##########################################################
 SK = abs((u+l-(2*m))/(u-l))

###########################*Fitting probability
distribution*##########################################################

### Initial values for optimisation, based on normal approximation
 s<-qnorm(1-(0.5/2))# getting upper quartile of standard normal
distribution (with mean=0 and std=1)
 v<-((u-l)/(2*s))^2 # initial with normal distribution assumption: Q3-
Q1=2.s.sigma with s=0.6744898
 if (SK <= 0.5){

### Fit a normal distribution
  normalerror<-function(x,elicited,probabilities){
   a<-x[1]
   b<-exp(x[2])
   sum((pnorm(elicited,a,b)-probabilities)^2)
  }
  normalfit<-
optim(c(m,0.5*log(v)),normalerror,elicited=c(Lo,l,m,u,Up),probabilities=c(0
,0.25,0.5,0.75,1))
  normalparameters<-c(normalfit$par[1],exp(normalfit$par[2]))
  normallsq<-normalfit$value
  re = list (dist = "Normal",pars = normalparameters, lsq = normallsq)
  return(re)
 } else {

### Fit a log normal distribution
  lnormerror<-function(x,elicited,probabilities){
   a<-x[1]
   b<-exp(x[2])
  sum((plnorm(elicited,a,b)-probabilities)^2)
  }
  std<-((log(u)-log(l))/(2*s))
  lognormalfit<-
optim(c(log(m),log(std)),lnormerror,elicited=c(Lo,l,m,u,Up),probabilities=c
(0,0.25,0.5,0.75,1))
  lnormparameters<-lognormalfit$par
  lnormparameters[2]<-exp(lnormparameters[2])
  lognormallsq<-lognormalfit$value
  re = list (dist = "Lognormal", pars = lnormparameters, lsq =
lognormallsq)
  return(re)
 }
}
```

Automatic fitting variogram function

```
### Modified from autofitVariogram - Automap R Library v1.0-7
# Main variables
# Semivariance: semivariance data organised in defined struture
# Study area bound: input_data
# Valid variogram model: model

# This function automatically fits and optimally choose a variogram model
to elicited semivariances
m_autofitVariogram = function(semivariance, input_data, model = c("Sph",
"Exp", "Gau", "Ste", "Mat"), kappa = c(0.05, seq(0.2, 2, 0.1), 5, 10),
GLS.model = NA, verbose = FALSE)
{
 x = input_data@bbox[1,]
 y = input_data@bbox[2,]
 experimental_variogram = semivariance

 # Automatically choosing the initial guess for fit.variogram
 # initial_sill = mean(max(semi_variance) + median(semi-vararariance))
 # initial_range = 0.10 * central axis of the area.
 # initial_nugget = minimum semi-variance value
 initial_nugget = min(experimental_variogram$gamma)
 initial_range = (0.1*sqrt((max(x) - min(x))^2 + (max(y) - min(y))^2) )
 initial_sill = mean(c(max(experimental_variogram$gamma),
median(experimental_variogram$gamma)))

 # Determine nugget, range and sill to be automatically fitted
 fit_nugget = TRUE
 fit_range = TRUE
 fit_sill = TRUE

 getModel <- function(psill, model, range, kappa, nugget, fit_range,
fit_sill, fit_nugget, verbose)
 {
  if(verbose) debug.level = 1 else debug.level = 0
  obj = try(fit.variogram(experimental_variogram,
        model = vgm(psill=psill, model=model, range=range,
        nugget=nugget,kappa = kappa),
        fit.ranges = c(fit_range), fit.sills = c(fit_nugget, fit_sill),
        debug.level = 0), TRUE)
  if("try-error" %in% class(obj)) {
  #print(traceback())
  warning("An error has occured during variogram fitting. Used:\n",
  "\tnugget:\t", nugget,
  "\n\tmodel:\t", model,
  "\n\tpsill:\t", psill,
  "\n\trange:\t", range,
  "\n\tkappa:\t",ifelse(kappa == 0, NA, kappa),
  "\n  as initial guess. This particular variogram fit is not taken into
account. \nGstat error:\n", obj)
  return(NULL)
  } else return(obj)
 }

 # Automatically testing different models, the one with the smallest sums-
of-squares and not singular is chosen
 test_models = model
 SSerr_list = c()
 sing_list = c()
 vgm_list = list()
 counter = 1
 for(m in test_models) {
  if(m != "Mat" && m != "Ste") {        # If not Matern and not Stein
```

```
  model_fit = getModel(initial_sill - initial_nugget, m, initial_range,
kappa = 0, initial_nugget, fit_range, fit_sill, fit_nugget, verbose =
verbose)
  if(!is.null(model_fit)) {   # skip models that failed
   vgm_list[[counter]] = model_fit
   SSerr_list = c(SSerr_list, attr(model_fit, "SSErr"))
   sing_list  = c(sing_list, attr(model_fit, "singular"))}
   counter = counter + 1
  } else {                 # Else loop also over kappa values
   for(k in kappa) {
    model_fit = getModel(initial_sill - initial_nugget, m, initial_range,
k, initial_nugget, fit_range, fit_sill, fit_nugget, verbose = verbose)
    if(!is.null(model_fit)) {
    vgm_list[[counter]] = model_fit
    SSerr_list = c(SSerr_list, attr(model_fit, "SSErr"))
    sing_list  = c(sing_list, attr(model_fit, "singular"))}
    counter = counter + 1
   }
  }
 }
 # Check for negative values in sill or range coming from fit.variogram
 # and NULL values in vgm_list, and remove those with a warning
 strange_entries = sapply(vgm_list, function(v) any(c(v$psill, v$range) <
0) | is.null(v))
 if(any(strange_entries)) {
  if(verbose) {
   print(vgm_list[strange_entries])
   cat("^^^ ABOVE MODELS WERE REMOVED ^^^\n\n")
  }
  warning("Some models where removed for being either NULL or having a
negative sill/range/nugget, \n\tset verbose == TRUE for more information")
  SSerr_list = SSerr_list[!strange_entries]
  vgm_list = vgm_list[!strange_entries]
  sing_list = sing_list[!strange_entries]
 }
 countertwo = 1
 prevgm_list = list()
 preSSerr_list = c()
 non_singular = which(!sing_list)
if (length(non_singular) != 0){
 for (n in non_singular){
 prevgm_list[[countertwo]] = vgm_list[[n]] #list of variogram with
non_singular attribute.
 preSSerr_list = c(preSSerr_list, SSerr_list[n]) #list of error of fitting
to non_singular variogram
 countertwo = countertwo + 1
 }
 if(verbose) {
  cat("Selected:\n")
  print(prevgm_list[[which.min(preSSerr_list)]])
  cat("\nTested models, best first:\n")
  tested = data.frame("Tested models" = sapply(vgm_list, function(x)
as.character(x[2,1])),
                      kappa = sapply(vgm_list, function(x)
as.character(x[2,4])),
                      "SSerror" = SSerr_list, "singular" = sing_list)
  tested = tested[order(tested$SSerror),]
  print(tested)
 }
 result = list(exp_var = experimental_variogram, var_model =
prevgm_list[[which.min(preSSerr_list)]], sserr = min(preSSerr_list))
 class(result) = c("autofitVariogram","list")}else{ result = NULL }
 return(result)
}
```

Making feedback function

```
# Visualizing  the pdf

### Modified from SHELF 2.0
# Main variables
# Type of probability distribution: Distribution
# Lower bound: Lo
# Upper bound: Up
# Mean (mu) and variance (sigma): parameters
# Sum of square error: lsq

#################################Probability density function
feedback#########################################################################
###

pdf.feedback <- function (Distribution, par1, par2, lsq, sig.fig = 3){
if (Distribution == "Normal"){
Lo = qnorm(0.0001, par1, par2)
Up = qnorm(0.9999, par1, par2)
}else{
if (Distribution == "Lognormal"){
Lo = qlnorm(0.0001, par1, par2)
Up = qlnorm(0.9999, par1, par2)}
}
x<-seq(from=Up,to=Lo,length=102)[2:101]
 fq1    = 0.25
 fq2    = 0.75
 median = 0.5

 if (Distribution == "Normal"){
  f.normal<-dnorm(x, par1, par2)
  q1      <-qnorm(fq1, par1, par2)
  q2      <-qnorm(fq2, par1, par2)
  med     <-qnorm(median, par1, par2)
  fb      = list(f.normal, q1, q2, med)
  return (fb)
  }else{

 if (Distribution == "Lognormal"){
  f.lognormal<-dlnorm(x, par1, par2)
  q1         <-qlnorm(fq1, par1, par2)
  q2         <-qlnorm(fq2, par1, par2)
  med        <-qlnorm(median, par1, par2)
  fb         = list(f.lognormal, q1, q2, med)
  return (fb)
 }
}
}
```

Define lag intervals from the extent of study area

```
# calculate lag interval, maximum lag <=1/2 diagonal of the area rounded to
k*10^j, k in , j interger
# smaller lag is either 0.5 or 0.4 of the previous larger lag

cal.lag.int <- function(area){
 x = area@bbox[1,]
 y = area@bbox[2,]
 D = sqrt((max(x) - min(x))^2 + (max(y) - min(y))^2)
 Diag = D
 k_int = 0
 while (Diag>=1){
  k_int = k_int + 1
  Diag = Diag%/%10
```

```
}
if (k_int >0){
 int.D = D%/%(10^(k_int-1))
 flo.D = (D%/%(10^(k_int-2)))/10
 if (flo.D < 1.5) {
  int.D = 1
  D.round = int.D*10^(k_int-1)
  lag1 = D.round*0.5
  lag2 = lag1*0.4
  lag3 = lag2*0.5
  lag4 = lag3*0.5
  lag5 = lag4*0.4
  lag6 = lag5*0.5
  lag7 = lag6*0.4
  }
 if ((1.5 <= flo.D) & (flo.D < 3.5)) {
  int.D = 2
  D.round = int.D*10^(k_int-1)
  lag1 = D.round*0.5
  lag2 = lag1*0.5
  lag3 = lag2*0.4
  lag4 = lag3*0.5
  lag5 = lag4*0.5
  lag6 = lag5*0.4
  lag7 = lag6*0.5
 }
 if ((3.5 <= flo.D) & (flo.D < 7.5)) {
  int.D = 5
  D.round = int.D*10^(k_int-1)
  lag1 = D.round*0.4
  lag2 = lag1*0.5
  lag3 = lag2*0.5
  lag4 = lag3*0.4
  lag5 = lag4*0.5
  lag6 = lag5*0.5
  lag7 = lag6*0.4
 }
 if (flo.D >= 7.5) {
  int.D = 10
  D.round = int.D*10^(k_int-1)
  lag1 = D.round*0.5
  lag2 = lag1*0.4
  lag3 = lag2*0.5
  lag4 = lag3*0.5
  lag5 = lag4*0.4
  lag6 = lag5*0.5
  lag7 = lag6*0.5
 }
}
if (k_int<=0){
 int.D = 0
 while (int.D<=0){
  k_int = k_int + 1
  int.D = (D*10^k_int)%/%1
 }
 flo.D = (trunc(D*10^(k_int+1)))/10
 if (flo.D < 1.5) {
  int.D = 1
  D.round = int.D/(10^k_int)
  lag1 = D.round*0.5
  lag2 = lag1*0.4
  lag3 = lag2*0.5
  lag4 = lag3*0.5
  lag5 = lag4*0.4
```

```
   lag6 = lag5*0.5
   lag7 = lag6*0.4
   }
  if ((1.5 <= flo.D) & (flo.D < 3.5)) {
   int.D = 2
   D.round = int.D/(10^k_int)
   lag1 = D.round*0.5
   lag2 = lag1*0.5
   lag3 = lag2*0.4
   lag4 = lag3*0.5
   lag5 = lag4*0.5
   lag6 = lag5*0.4
   lag7 = lag6*0.5
  }
  if ((3.5 <= flo.D) & (flo.D < 7.5)) {
   int.D = 5
   D.round = int.D/(10^k_int)
   lag1 = D.round*0.4
   lag2 = lag1*0.5
   lag3 = lag2*0.5
   lag4 = lag3*0.4
   lag5 = lag4*0.5
   lag6 = lag5*0.5
   lag7 = lag6*0.4
  }
  if (flo.D >= 7.5) {
   int.D = 10
   D.round = int.D/(10^k_int)
   lag1 = D.round*0.5
   lag2 = lag1*0.4
   lag3 = lag2*0.5
   lag4 = lag3*0.5
   lag5 = lag4*0.4
   lag6 = lag5*0.5
   lag7 = lag6*0.5
  }
 }
 lags = matrix(c(lag7, lag6, lag5, lag4, lag3, lag2, lag1), nrow = 7, ncol
= 1)
 return(lags)
}
```

Unconditional simulation

```
#doing uncondistional simulation
uncondition_sim <- function(mu, sigma, N, grid_map, vari_model)
{
 gridcord = coordinates(grid_map)
 vx = gridcord[,1]
 X =sample(vx,1)
 for (i in 1:length(vx))
  if (gridcord[i,1] == X) Y = gridcord[i,2]
 names(Y) <- NULL
 in_sample = data.frame(x=X,y=Y,sample_value=rnorm(1,mean=mu,sd=sigma))
 coordinates(in_sample)=~x+y
 fsim =
gstat(id=c("sample_value"),formula=sample_value~1,data=in_sample,beta=mu,nm
ax=24,model=vari_model,dummy=TRUE)
 unconditional_sim = predict.gstat(fsim,grid_map,nsim=N,BLUE=FALSE)
 return (unconditional_sim)
}
```