TESLA

Decision Support Software

Evidence Support Logic: A Guide for TESLA Users

Version 3.0

Quintessa Limited

Quintessa

Long term environmental thinking...

...from data to decisions

Contents

Background

Complex decisions are typically informed by a wide range of factors, drawing on multiple information sources. This can involve having to assemble and assess substantial amounts of evidence. Decisions are then made on the basis of judgments about the quality and reliability of that evidence and the extent to which it provides confidence in a given interpretation. A familiar example is the way in which evidence is taken into account in a legal trial; the jury is required to examine all the evidence presented to them, and then to make judgments about the extent to which the evidence supports the case made by the trial lawyers.

Similar questions are faced in more technical disciplines, such as in the use of available data (and interpretations of such data) to support particular concepts or hypotheses. In the geological field, for example, a variety of sources of evidence may be collected together and analysed in determining mineral extraction prospects or, say, in developing a conceptual model for the hydrogeology of a given region. Evidence can come from a wide range of sources, including apparently "hard" quantitative data or the results of quantitative modelling, alongside analogue reasoning, expert judgments and the value judgments of different stakeholder groups.

Moreover, whilst there may be a large volume of information relating to the decision at hand, it may on the whole be only of partial relevance, incomplete and/or uncertain, or even conflicting in terms of the level of support it provides for a given interpretation. The range of available evidence may appear to give an indistinct picture, with no clear indication of how best to target resources to improve understanding. There may be disputed interpretations, perhaps because some practitioners appear to be biased by excessive reliance on a particular source of evidence in the face of contradictory, or seemingly more equivocal, evidence from elsewhere. Hence, in order to provide a justified interpretation of the available evidence, which can be audit-traced from start to finish, it is necessary to examine and make visible confidence judgments on both the quality of the data and the quality of the interpretation and modelling process.

The technique of Evidence Support Logic (ESL), implemented in Quintessa's TESLA software[1] is intended to support decision makers and modellers in their sense-making when faced with extensive information processing requirements. In summary, ESL involves systematically breaking down the question or hypothesis under consideration into a logical hypothesis model the elements of which expose basic judgments and opinions relating to the quality of evidence associated with a particular interpretation or proposition, in addition to establishing the level of confidence that can be placed in the relevant judgments. By independently evaluating confidence "for" and "against" propositions on the basis of evidence, uncertainty (and/or conflict) is captured and the sensitivity of the results to that uncertainty can be evaluated.

This document aims to describe the ESL methodology; how to construct a logical hypothesis model; how to quantify confidence on the basis of gathered evidence; the method that ESL uses to propagate confidence up the hierarchical structure of the model to produce a solution; various ways of visualising data within the model; and finally provides a brief overview of the implementation of ESL within the TESLA software.

---

[1] Further information about the implementation of ESL within the TESLA software can be found in the TESLA User Guide, the latest version of which can be downloaded via the Quintessa website: http://www.quintessa.org/software/TESLA/

A list of example applications of ESL (not necessarily using the TESLA software) is given towards the end of this document. Finally, a Glossary of Terms is also provided.

Benefits of ESL

Arguably, the greatest value of analysing uncertainty through developing an evidence-based logical hypothesis model is not so much in achieving a precise understanding of the knowledge inputs, or providing an absolute assessment of the degree of assurance supporting a particular proposition or hypothesis, but in identifying those uncertainties that have the greatest impact on overall confidence (Bowden, 2004).

Sensitivity analysis based on an ESL model helps to identify where future investigation and evidence gathering is likely to have the greatest impact in reducing uncertainty due to lack of knowledge. This may be particularly valuable, for instance, when limited resources are available and prioritisation judgements need to be made concerning the acquisition of further data to improve the quality of decision making. Where there is too much uncertainty for a particular hypothesis or model to be accepted or refuted, a systematic evidence-based evaluation of confidence therefore allows identification of the source(s) of uncertainty and hence possibilities for the targeting resources to improve confidence.

In addition, the outcome of parallel evidence-based analyses of alternative hypotheses (such as, for example, different conceptual models for the features and characteristics of a particular process system) can help in comparing relative levels of confidence in those alternatives. The impact of new understanding gained from additional evidence can then be evaluated in terms of its effect on the confidence that may be placed in each alternative, and hence on the extent to which one may be preferred over another.

As with all decision support tools and processes, ESL neither replaces the need for judgment nor eliminates subjectivity from the evaluation and interpretation of evidence. However, a systematic approach, breaking the problem down into a logical model whose elements expose essential judgments and opinions relating to that evidence, can lend a measure of structure and transparency to that appraisal. This, in turn, helps to ensure there is comprehensive coverage of relevant factors and uncertainties, and that an audit trail is established for the judgments that need to be made. Thus, for example, where there is potential for bias in a decision, the systematic evaluation of evidence can help to demonstrate how such bias has arisen and what its implications might be.

Key Steps in ESL

ESL has been developed from a methodology described by researchers at Bristol University (Cui and Blockley, 1990; Foley et al., 1997; Hall et al., 1998; Blockley and Godfrey, 2000; Davis and Hall, 2003) and subsequently adapted by Bowden (2004), primarily for application in the field of modelling interpretation. The technique seeks to build confidence in decisions based on a comprehensive and systematic identification of all potentially relevant evidence and the formal evaluation of that evidence.

At its core, ESL involves four main steps:

1. **Development of a hierarchical logical hypothesis model** to provide a common, coherent structure for assembling and assessing all the evidence that is relevant to an identified "root" (or top-level) hypothesis (or proposition);

2. **Parameterisation of the logical model** and identification and assessment of sources of evidence that contribute confidence for and/or against "child" hypotheses in the model;

3. **Propagation of confidence**, on the basis of judgments on evidence, through the logical model, representing uncertainty by separately propagating independent judgments on evidence for and against hypotheses using the principles of Interval Probability Theory (Cui & Blockley, 1990), to provide an assessment of confidence in the overall root hypothesis; and

4. **Analysis and visualisation** of the logical hypothesis model and its outcomes by, for example, examining the sensitivity of the confidence at the top level (root hypothesis) to confidence values at the bottom level ("leaf" hypotheses).

As noted by Davis and Hall (2003), the feasibility of such a systematic approach is greatly enhanced by the use of software to support model construction, knowledge recording and uncertainty handling. Quintessa has designed and developed the TESLA software tool for this purpose, with an emphasis on the ability to manipulate the hierarchical structure of a decision as it evolves and providing tools for efficiently analysing the model.

In what follows, the principles that underpin the steps outlined above are briefly described in turn.

Step 1: The Logical Hypothesis Model

ESL involves evaluating the support for an identified root (or top) hypothesis that is provided by a number of logical child hypotheses, each of which can be associated with some fraction of the underlying basic evidence. A logical hypothesis hierarchy (or tree) links the root hypothesis to data or information at the lowest level (that is, to leaf hypotheses), usually via a series of intermediate hypotheses, representing logical lines of reasoning. A chain of "parent" and supporting child hypotheses links the root hypothesis to each leaf hypothesis and is termed a "branch" of the tree. At each level of the tree, child hypotheses that have the same parent are termed sibling hypotheses.

The example shown in Figure 1 demonstrates the general principles associated with defining such a hypothesis model. In this case, the specified root hypothesis relates to testing the proposition that a particular location may be suitable for more detailed investigation as part of a technical evaluation protocol relating to the siting of a deep sub-surface radioactive-waste disposal facility.

Assessing suitability of site for detailed investigation

```
                        ┌──────────────────┐      ┌──────────────────────────┐
                        │ Future uplift and│      │ 1. Uplift will be less   │
                        │ erosion will not │──────│    than 300m in next     │
                        │ be prohibitive   │      │    100,000 years         │
                        └──────────────────┘      └──────────────────────────┘

                                                  ┌──────────────────────────┐
                                                  │ 2. Site is not within    │
                        ┌──────────────────┐      │    range of active fault │
                        │ Future fault     │──────│    zone                  │
                        │ activity will not│      └──────────────────────────┘
                        │ be prohibitive   │
                        └──────────────────┘      ┌──────────────────────────┐
                                                  │ 3. Site is not within    │
                                                  │    zone of potential     │
                                                  │    fault reactivation    │
                                                  └──────────────────────────┘

                                                  ┌──────────────────────────┐
   ┌──────────────┐                               │ 4. Site is not within    │
   │ The site is  │                               │    active volcanic area  │
   │ suitable for │                               └──────────────────────────┘
   │ detailed     │
   │ investigation│       ┌──────────────────┐    ┌──────────────────────────┐
   └──────────────┘       │ Future igneous   │    │ 5. Risk from direct      │
                          │ activity will not│────│    magma intrusion is low│
                          │ be prohibitive   │    └──────────────────────────┘
                          └──────────────────┘
                                                  ┌──────────────────────────┐
                                                  │ 6. Risk of hydrothermal  │
                                                  │    activity at depth due │
                                                  │    to magmatism is low   │
                                                  └──────────────────────────┘

                                                  ┌──────────────────────────┐
                                                  │ 7. The host rocks are all│
                        ┌──────────────────┐      │    consolidated          │
                        │ Host rock geology │─────└──────────────────────────┘
                        │ is suitable      │
                        └──────────────────┘      ┌──────────────────────────┐
                                                  │ 8. There are no viable   │
                                                  │    mineral resources     │
                                                  └──────────────────────────┘
```
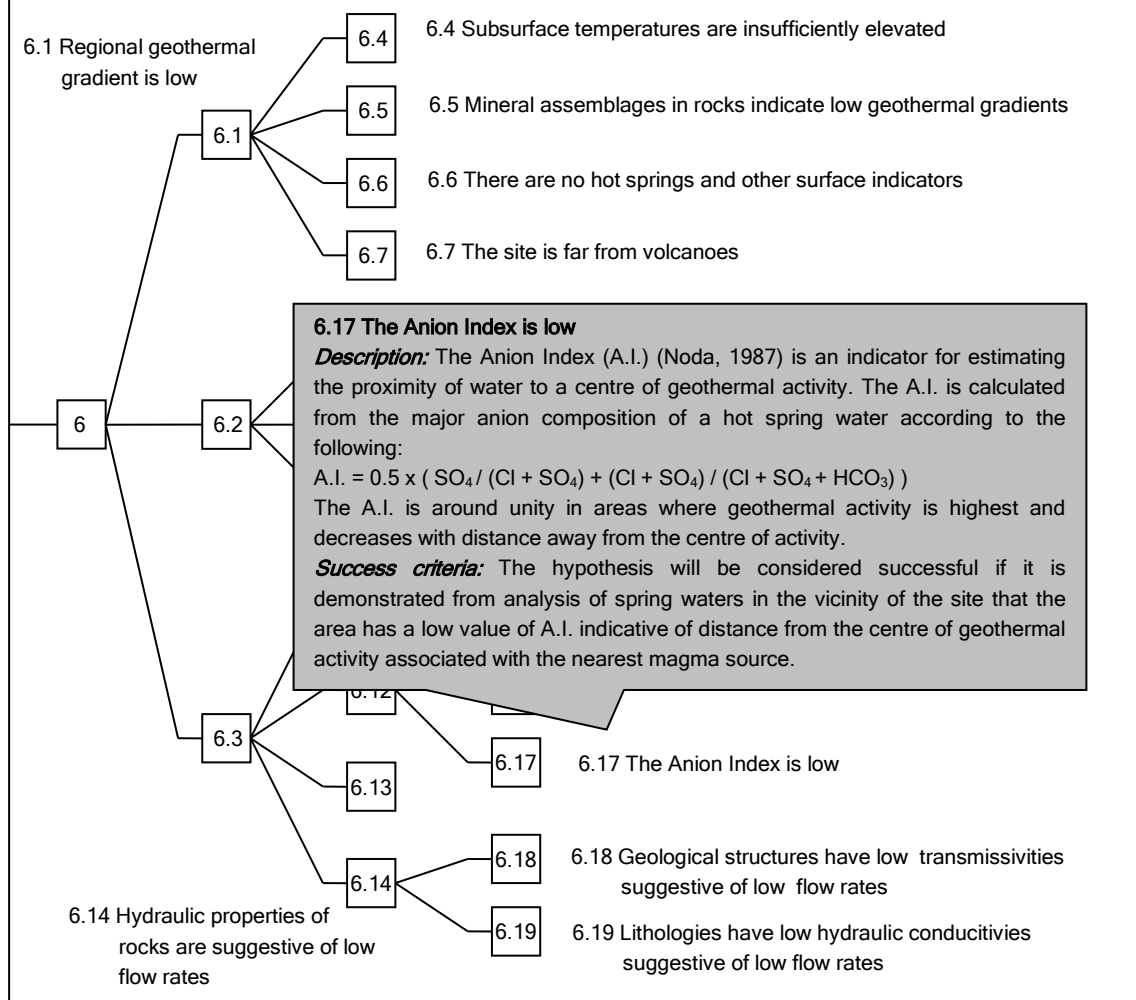
Root Hypothesis   *Children of the Root Hypothesis.*   *Leaf Hypotheses*
                           Parents of Leaf Hypotheses.
                           Sibling Hypotheses.

Figure 1: Simplified example of an evidence-based hypothesis model showing site-specific factors for assessing the suitability of a site for inclusion in a detailed investigation.

In this example, the root hypothesis ("The site is suitable for detailed investigation") is too substantial and vague to be answered directly and needs to be broken down in order to provide better definition of the decision problem and to understand the various contributing factors.

Figure 1 therefore illustrates an initial decomposition into eight contributory leaf hypotheses that together determine the confidence in the root hypothesis. The general principle at each step in developing the hierarchical model is to undertake a comprehensive top-down analysis of the various factors that contribute to a hypothesis, until a level of detail is reached at which people are comfortable in providing direct judgments about evidence in terms of the level of support that it provides for and against the child hypothesis in question. The example given in Figure 1 does not break down the hypotheses into sufficient detail for this purpose. Figure 2 shows how just one of the eight child hypotheses ("Risk of hydrothermal activity at depth due to magmatism is low") can be developed further to a level at which it may be judged that relevant evidence can be brought to bear.

Inevitably, the structure that is adopted in developing a logical hypothesis model is somewhat subjective, in so far as there may be a number of alternative ways of defining a comprehensive top-down hierarchy. Ideally, the model should therefore be determined by one or more specialists in the relevant field. Several approaches are available to do this, depending upon the nature of the information, the numbers of the experts and their specialities. A common approach is for a single person acting as a facilitator to lead the construction of a hypothesis model in a meeting or meetings involving the experts. At each stage, the structure can then be debated until a consensus is reached. Alternatively, a central process organiser can arrange for independent experts, or small groups of experts, to contribute to the development of particular aspects of the tree, with subsequent review, analysis and challenge by a second expert or set of experts. Sometimes, a combination of these two approaches can prove effective.

For each leaf hypothesis on which a judgment of confidence on the basis of evidence is required, it is important to guide that judgment by a clear description of the child hypothesis in question and the identification of criteria for its "success". The success criteria effectively define an agreed standard against which confidence in the child hypothesis can be assessed, based on the available evidence. This standard therefore defines the circumstances under which a hypothesis would be considered completely true. Within the example shown in Figure 2, the inset text box describes these attributes for one of the model child hypotheses ("The Anion Index is low"). As lines of reasoning establishing confidence "for" and "against" hypotheses are handled independently, in practice "failure" criteria are typically also specified; note that sometimes evidence for "failure" can be different to an absence of evidence for "success", as absence of evidence leading to a lack of confidence in success may imply uncertainty, not confidence in failure.

It can also be helpful at this stage to identify and record the relevant types and sources of information that can later be used to assess the confidence in the child hypothesis in question, according to the extent of its support for (or argument against) the success of the root hypothesis.

6. Risk of hydrothermal activity at depth due to magmatism is low

6.1 Regional geothermal gradient is low

6.4 Subsurface temperatures are insufficiently elevated

6.5 Mineral assemblages in rocks indicate low geothermal gradients

6.6 There are no hot springs and other surface indicators

6.7 The site is far from volcanoes

6.17 The Anion Index is low
Description: The Anion Index (A.I.) (Noda, 1987) is an indicator for estimating the proximity of water to a centre of geothermal activity. The A.I. is calculated from the major anion composition of a hot spring water according to the following:
A.I. = 0.5 x (SO_4 / ($Cl + SO_4$) + ($Cl + SO_4$) / ($Cl + SO_4 + HCO_3$))
The A.I. is around unity in areas where geothermal activity is highest and decreases with distance away from the centre of activity.
Success criteria: The hypothesis will be considered successful if it is demonstrated from analysis of spring waters in the vicinity of the site that the area has a low value of A.I. indicative of distance from the centre of geothermal activity associated with the nearest magma source.

6.17 The Anion Index is low

6.14 Hydraulic properties of rocks are suggestive of low flow rates

6.18 Geological structures have low transmissivities suggestive of low flow rates

6.19 Lithologies have low hydraulic conducitivies suggestive of low flow rates

Figure 2: Nested hypothesis model showing lower level of the tree for which evidence of support for and evidence against must be specified.

Step 2a: Model Parameterisation

Before evidence can be assessed and associated confidence can be input to the hierarchical hypothesis model, it is necessary to describe and parameterise the logic by which that confidence is propagated upwards through the model in order to assess the extent of support for the root hypothesis. As noted earlier, the mathematical basis of the ESL methodology is an approach known as Interval Probability Theory (IPT). A comprehensive discussion of IPT and the inference propagation calculus that has been developed from the theory is provided by Cui and Blockley (1990). For present purposes, the aim is simply to outline some major elements of the quantitative methodology, in order to enable a better understanding of the model parameterisation process. This covers three main aspects:

▲ the concept of three-value logic, which provides for an explicit recognition of uncertainty in evidential judgments;

▲ the parameterisation of relationships between hypotheses in the hypothesis model, expressed in terms of their *sufficiency*, *dependency* and *necessity,* or alternatively logical arguments that act on groups of hypotheses representing lines of reasoning; and

▲ the basis for defining evidential judgments to enable assessment of the confidence in child hypotheses at the base of the hierarchical model.

Three-value Logic

Evidential judgments based on classical probability theory follow two-value logic, whereby confidence that a hypothesis is true is associated with the probability that it is true. As a result, lack of confidence in favour of a hypothesis implicitly represents confidence in its falsehood, and cannot be set independently. This is sometimes described as a "closed world" perspective, in which confidence "for" and "against" are treated as complementary concepts (i.e. $p(A) + p(\text{not } A) = 1$, where $p(A)$ is the probability of event A occurring, or in other words the confidence supporting the occurrence of A). In this classical case, uncertainty is represented only by the fact that a probability has been assigned to the hypothesis, so that excluding extreme cases, it can be considered that there is confidence both for and against it.

Three-value logic, which is employed by ESL, extends this to allow the independent evaluation of the extent to which the evidence provides confidence "for" and "against" each hypothesis. Whilst the judgments "for" and "against" will be made on the basis of the same overall evidence base, they are made independently. This means that confidence "for" and confidence "against" need not sum to 1.

Representation of Uncertainty

In three-value logic, the sum of confidence "for" and confidence "against" does not have to equal 1. This is because three-value logic allows for a measure of residual uncertainty due to "uncommitted" or "overcommitted confidence". Uncertainty due to "uncommitted confidence" recognises that confidence in a hypothesis may be only partial and that some level of confidence may be assigned to an uncommitted state. In contrast, "overcommitted confidence" arises when the people making a judgment place too much confidence in evidence for and / or against a hypothesis. This means that the judged confidence for and confidence against, when represented on a scale of 0 to 1, actually sum to greater than 1.

"Residual uncertainty" due to uncommitted confidence is handled as an "interval" that enables the admission of a general level of uncertainty (Waltz, 1989), providing a recognition that information and judgments may be incomplete and possibly inconsistent (i.e. residual uncertainty due to uncommitted confidence = 1 - confidence for - confidence against, when confidence for + confidence against < 1). This is illustrated in Figure 3, which adopts the so-called "Italian flag" representation of three-value logic, in which confidence for a hypothesis is represented as green, confidence against as red, and residual uncertainty due to uncommitted confidence is white (Blockley and Godfrey, 2000).

Classical 2-value Probability Logic		Confidence-based 3-value Logic, with $C(A) + C(not\ A) \leq 1$		
Probability of truth	Probability of falsehood	Confidence in success on the basis of evidence	Residual uncertainty due to uncommitted confidence	Confidence in failure on the basis of evidence
0.28	0.72	0.28	0.42	0.30
$P(A)$	$P(not\ A) = 1 - P(A)$	$C(A)$	$1 - C(A) - C(not\ A)$	$C(not\ A)$
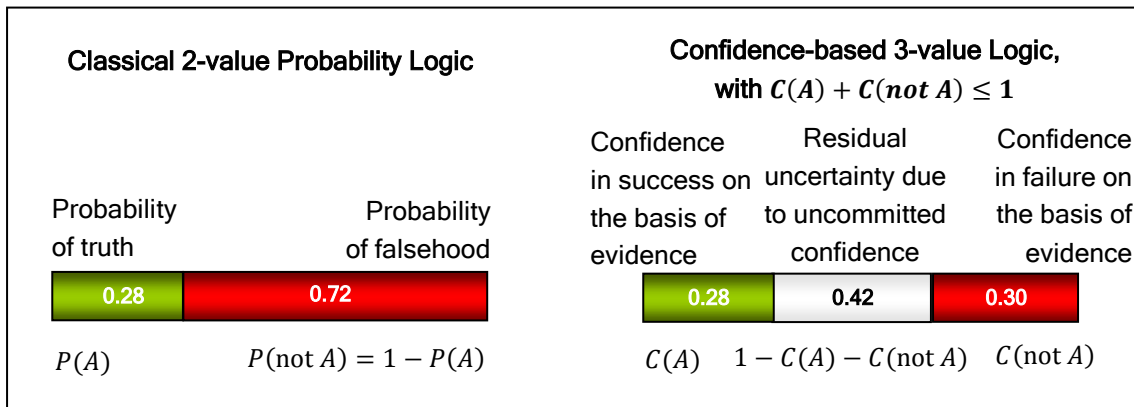
Figure 3: Classical two-value probability analysis compared with three-valued interval analysis.

As noted previously with the three-value formalism, confidence for and confidence against are evaluated independently, with confidence values relating to those judgments each ranging from 0 to 1, and with residual uncertainty due to uncommitted confidence taking a value from 0 to 1. Residual uncertainty due to uncommitted confidence of 1 implies that there is no evidence at all on which to base a judgment (or that the evidence that appears to exist is evaluated as being entirely unreliable). As an alternative to the Italian flag representation, the values may be simply represented in the triplet form [*confidence for, residual uncertainty due to uncommitted confidence, confidence against*], for example [0.28, 0.42, 0.30].

ESL thus represents two key aspects of uncertainty in confidence "for" a hypothesis; the elicited confidence against it, where the uncertainty can be allocated to evidence; and the residual uncertainty that cannot be allocated, represented by the "white space".

Positive values for residual uncertainty are typically termed "uncommitted confidence", recognising the potential for further improvement in confidence on the basis of, for example, new evidence. As noted above, the ESL approach also allows for "overcommitted confidence", which occurs when elicited or propagated values for confidence on the basis of evidence total more than 1 (i.e. residual uncertainty due to overcommitted confidence = confidence for + confidence against - 1). Overcommitted confidence typically represents conflict, either where the evidence points in two different directions, or more often where expert judgments on evidence are inconsistent. The importance of conflict and the relative priority that needs to be given to resolving it can be analysed by understanding its impact on the root hypothesis, and the sensitivity of the outcomes to the conflicting judgments made.

Within the TESLA software, a situation where there are overcommitted confidence judgments (say for example [0.78, -0.42, 0.64]) is indicated by a yellow central bar (Figure 4).

Overcommitted Confidence

$$C(A) + C(not\ A) > 1$$

Confidence in success on the basis of evidence	Residual uncertainty due to overcommitted confidence	Confidence in failure on the basis of evidence
0.70	0.30	0.60
$C(A)$	$C(A) - C(\text{not } A) - 1$	$C(\text{not } A)$
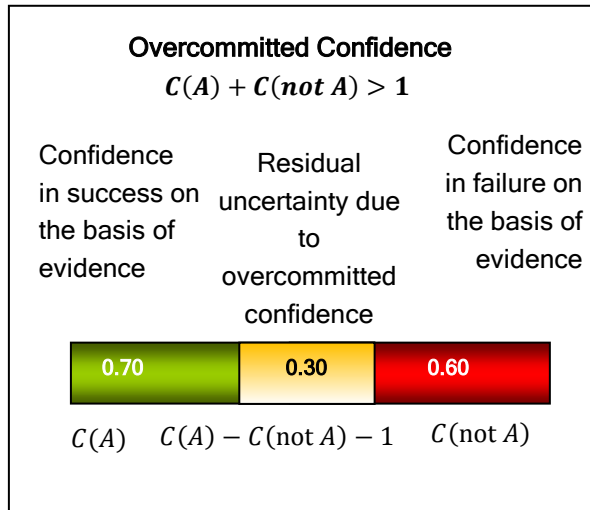
Figure 4: Example of overcommitted confidence.

Key Concepts in ESL Tree Parameterisation

The propagation of confidence through an ESL tree is achieved through the combination of two main classes of logical approach:

▲ The use of logical operators (or "heuristics") that select between lines of reasoning on the basis of certain rules, in order to establish the key contributions to confidence relevant to a particular decision.

▲ Application of the concept of sufficiency of judgments to understand the total contribution to confidence in a parent hypothesis imparted by its children.

These two aspects are described below. An illustration of the different concepts described is then provided.

Propagation by Operators (or "Heuristics") Acting on Lines of Reasoning

ALL child hypotheses required for confidence in the parent (the "Weakest Link" heuristic)

In many situations, confidence in the parent hypothesis is dependent on achieving confidence in all of the child hypotheses.

The ALL operator or heuristic is used to enforce a rule that confidence for (or against) the parent is the same as that for the weakest of its children. It represents a situation where all the child hypotheses are required collectively to ensure the success (or failure) of the parent, and where an absence of confidence for (or against) any one child alone would mean an equivalent absence of confidence in the parent. This situation is represented by propagating the minimum confidence value for (or against) from the child nodes (Figure 5). This form of logic is often (but by no means exclusively) used towards the top levels of a tree, to control integration of confidence arising from various lines of reasoning. It reflects an assessment that *confidence in the parent can be no better than that for the "weakest link" of the child hypotheses*, and that

there is no "mutual support" i.e. confidence in child nodes does not combine in providing confidence in the parent.

This form of logic is typically used when various lines of reasoning are critical to success of a parent, but it is not possible to determine the sufficiency of each to prove the parent (see below) because it is not meaningful to consider their impacts on the parent independently. For example, regulatory guidance might require that certain lines of reasoning have equal prominence in a decision, so that an independent assessment of their impact is not possible.

ANY child hypotheses sufficient for confidence in the parent (the "Strongest Link" heuristic)

In many situations, any and every single child hypothesis is sufficient on its own to ensure the success (or failure) of a given parent hypothesis and each child is independent of the others. The ANY operator or heuristic is used to enforce a rule that confidence for (or against) the parent is the same as that for the strongest of its children (Figure 5). *This is the "strongest link" argument whereby if any one of the children is true, then the parent must also be*, but confidence in the children does not combine to produce overall confidence in the parent higher than that for the node that represents the "strongest link". This latter aspect is important to identifying if ANY should be employed rather than just employing a sufficiency value of 1 for each of the sibling nodes in the group. ANY leads to the direct propagation of the maximum confidence value from among the child nodes. As for the "ALL" operator, this form of logic is often (but by no means exclusively) used towards the top levels of a tree, to control integration of confidence arising from various different lines of reasoning.

As for ALL, the ANY operator is often used where it is not meaningful to consider their impacts on the parent independently (for example if regulatory guidelines or other decision processes, govern the overall confidence that can be attributed to multiple lines of reasoning in supporting a decision).
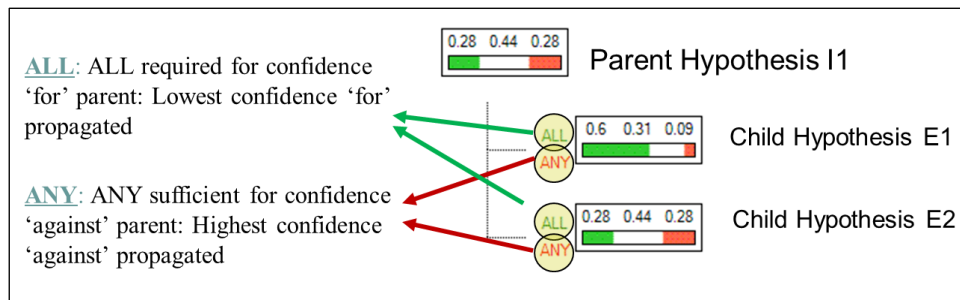
Figure 5: Illustration of use of ALL and ANY operators / heuristics.

Necessity Heuristic

In recent versions of TESLA (v2.2 and onwards), two necessity algorithms are implemented: symmetric and asymmetric. Here, only the newer symmetric algorithm is discussed, as this is the current default method.

The symmetric necessity heuristic allows hypotheses to be identified as *necessarily true* such that, irrespective of their siblings, confidence is required in them to have confidence that their parent is true. Alternatively, it can allow some hypotheses to be identified as *necessarily false*, such that irrespective of their siblings, confidence that they are false is required to have

confidence that their parent is false. For example, a site for a new supermarket must have planning permission. If there is sufficient confidence that this will not be achieved based on available evidence, then the site should not be chosen, even if it was perfect in every other sense. Therefore, it may be appropriate to mark a child hypothesis identifying that "The supermarket will acquire planning permission" as necessarily true.

The confidence values of necessary hypotheses are subject to a threshold test. If a necessarily true hypothesis fails by exceeding a threshold associated with confidence "against" then the confidence value against the parent will be set so that it is at least as large as that for the necessary hypothesis (Figure 6). If a necessarily true hypothesis fails a test by not meeting the confidence for threshold, then the confidence value for the parent will be set so that it is no greater than that of the necessary hypothesis (Figure 6). In each case, the mirror-image will apply for a necessarily false hypothesis.

For each of these tests, if confidence for / against does not cause failure compared to the threshold, then normal propagation rules for the confidence for / against apply.

For hypotheses that are necessarily true if its parent is to be true, confidence for must meet or exceed 0.5 to pass the confidence for necessity test and confidence against must not exceed 0.5 to pass the confidence against necessity test. For hypotheses that are necessarily false if its parent is to be false, confidence against must meet or exceed 0.5 to pass the confidence against necessity test and confidence for must not exceed 0.5 to pass the confidence for necessity test.
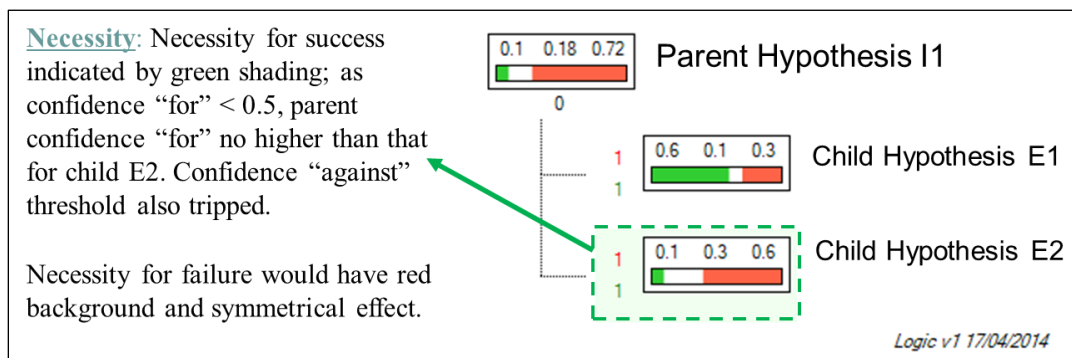
Figure 6: Illustration of use of the necessity heuristic.

Sufficiency and Dependency

Other than for aspects of a tree for which logical operators described above are appropriate, the fundamental approach to the propagation of confidence is based on Interval Probability Theory (IPT), which maps confidence values to a range of probabilities for which a hypothesis is true, and which then propagates the confidence values using probabilistic logic

The basic algorithm combines confidence from child hypotheses using such that the confidence in the parent is always greater than (or equal to) the contribution from any child on its own. The effect is additive, such that two separate lines of reasoning that both support (or refute) a proposition reinforce each other, with the combination being more supportive than either source of confidence taken individually. Even where uncertainty is present with respect to the individual items of contributing evidence, the application of fragments of knowledge from each element allows confidence in the proposition to be increased, thereby reducing uncertainty overall.

The basis of IPT is the assumption that confidence in the truth of a proposition can be modelled as a lower-bound on the probability that it is true (and, by symmetry, that one minus confidence in the falsehood of a proposition can be modelled as an upper bound on the probability that it is true). This establishes a link between confidence as assessed by the user, and the underlying probability theory.

For multiple lines of reasoning it is necessary to consider the how confidence in a hypothesis is impacted by contributions from each line of reasoning represented by its child hypotheses, while accounting for overlap in confidence due to shared evidence judgements to ensure there is no double counting. In practice, the combination of confidence is modified by two factors: *sufficiency*, and *dependency*.

At every branch within the hierarchical hypothesis model, each child hypothesis is assigned **sufficiency** values that dictate how important it is to determining the combined confidence of its parent hypothesis. Within TESLA, two separate values for this parameter may be assigned, representing the designated sufficiency of the child hypothesis both for and against the success of the parent hypothesis. In effect, when determining the sufficiency of a child hypothesis, consideration is given to its overall relevance to making judgments about confidence in the parent hypothesis. This is equivalent to asking the question:

If it were assumed that there is complete confidence for/against the child hypothesis (alone), how much confidence would there be for/against the parent hypothesis?

An equivalent question is asked concerning the sufficiency for "failure" – i.e. confidence that the higher level proposition is completely false.

The sufficiency parameter can take a value between 0 (insufficient) and 1 (completely sufficient) – a greater level of sufficiency results in confidence values associated with the child hypothesis being propagated more strongly up the hierarchy.

Within each set of sibling child hypotheses at a given level in the logical hypothesis model hierarchy, there is a chance that some of the contributing information may be overlapping, or shared, perhaps because of reliance on a common evidence source or argument. This is reflected in the **dependency** parameter, which describes the degree of commonality that is understood to exist between contributing hypotheses. The role of dependency in the quantitative propagation of confidence through the hypothesis model is to avoid double-counting of the support provided by mutually dependent items of evidence. It therefore provides a subtractive element to the propagation algorithm. In eliciting dependency, the question is asked:

How much shared information do the child hypotheses use in contributing to the confidence for/against the parent?

The value of dependency is a property of the parent hypothesis. It varies between 0 and 1, where 0 means that the confidence provided by the siblings is independent, and 1 means that the confidence provided by the siblings is maximally dependent – with, for example, one sibling hypothesis not providing any additional confidence because it is already provided by another. Subsets of sibling hypotheses may have different levels of dependency. Note that a subset of sibling hypotheses cannot have a smaller dependency value than the set that encompasses it. In other words, for 3 sibling hypotheses, if $D_{1,2,3} = 0.55$ then $D_{1,2}, D_{1,3}, D_{2,3} \geq 0.55$, where $D_{i,j,k}$ is the dependency between the sibling hypotheses i, j and k.

A visual representation of sufficiency and dependency is shown in Figure 7, in which the relationship between confidence for (or against) a parent hypothesis, H, and confidence for (or against) two sibling child hypotheses, C_1 and C_2, is illustrated in the form of a Venn diagram.

Here the sufficiency of a child hypothesis is demonstrated by the amount it overlaps its parent. If a hypothesis is completely sufficient (a value of 1), it will completely overlap its parent. Likewise the overlap between sibling hypotheses is linked to the dependency, D.
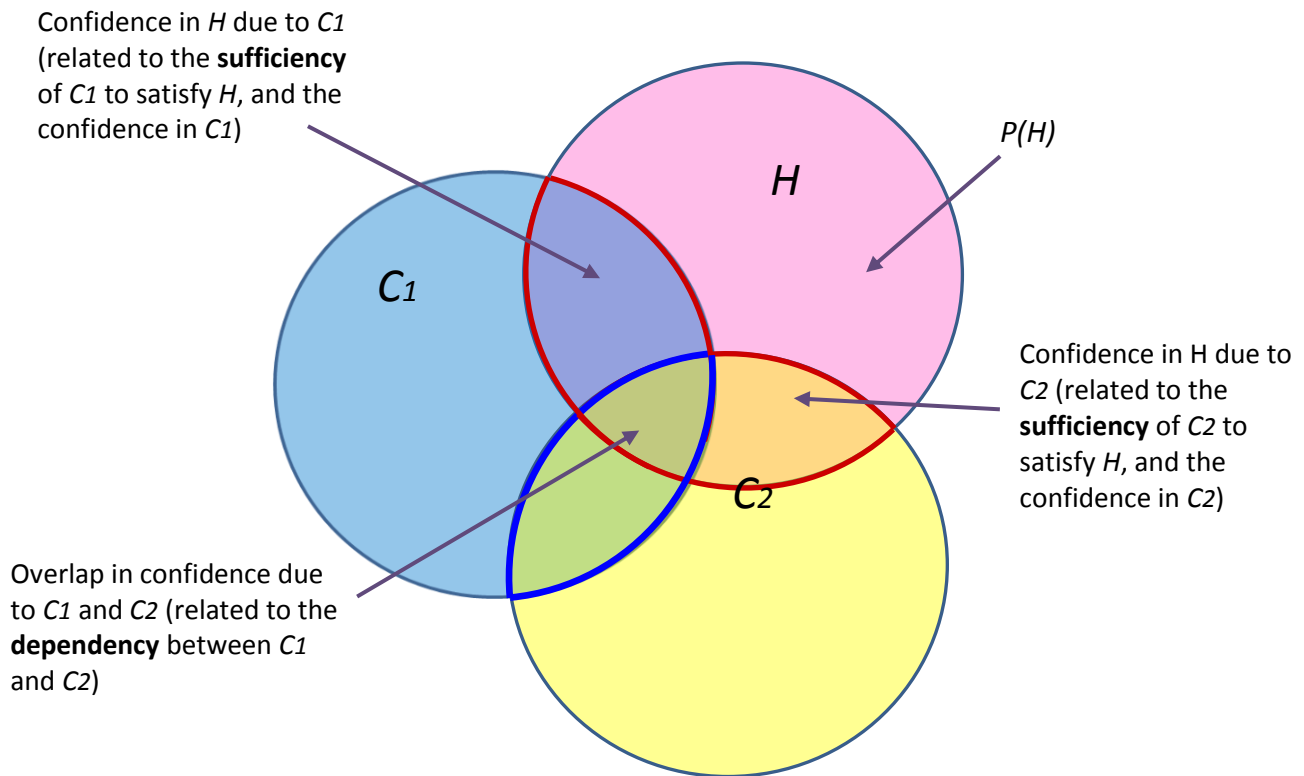
Confidence in H due to $C1$ (related to the **sufficiency** of $C1$ to satisfy H, and the confidence in $C1$)

$P(H)$

Confidence in H due to $C2$ (related to the **sufficiency** of $C2$ to satisfy H, and the confidence in $C2$)

Overlap in confidence due to $C1$ and $C2$ (related to the **dependency** between $C1$ and $C2$)

Figure 7: Schematic diagram of the ESL parameters sufficiency and dependency. The field outlined in red is a measure of confidence in H arising for confidence in C1 and C2, while not double counting dependency between C1 and C2.

Parameter values for sufficiency and dependency are necessarily assigned by expert judgment. Eliciting numerical values directly may be difficult where the hypothesis definition is imprecise or complex. In such cases an alternative process is to judge sufficiencies and dependencies using linguistic terms, such as "Very High", "High", "Intermediate", "Low" and "Very Low", which are subsequently mapped to numerical form for computational purposes, using a mapping scheme defined beforehand. As with all linguistic to numerical conversions, it is important that all contributing experts understand the conversion factors that have been used and have the opportunity to modify, revise or refine their initial judgment if the conversion appears to misrepresent their intention. The present version of TESLA (Version 2.2) does not allow users to define linguistic to numerical conversions directly for sufficiencies and dependences, but does allow such mapping to be defined for confidence judgments based on the face value of evidence.

Difficulties can arise in eliciting values of the dependency parameter, as there may be considerable uncertainty regarding the degree of shared information, or where (as noted above) there may be different degrees of dependency between different sibling hypotheses. For this reason, where alternative hierarchical model structures exist for a given root hypothesis, it can be preferable to adopt one in which levels of dependency are as low as possible.

Bowden (2004) provides some hypothetical examples to illustrate the impact of the different combinations of logical operators on the propagation of confidence. Examples of the use of the ANY and ALL heuristics in combination with sufficiencies are given in Figure 8.
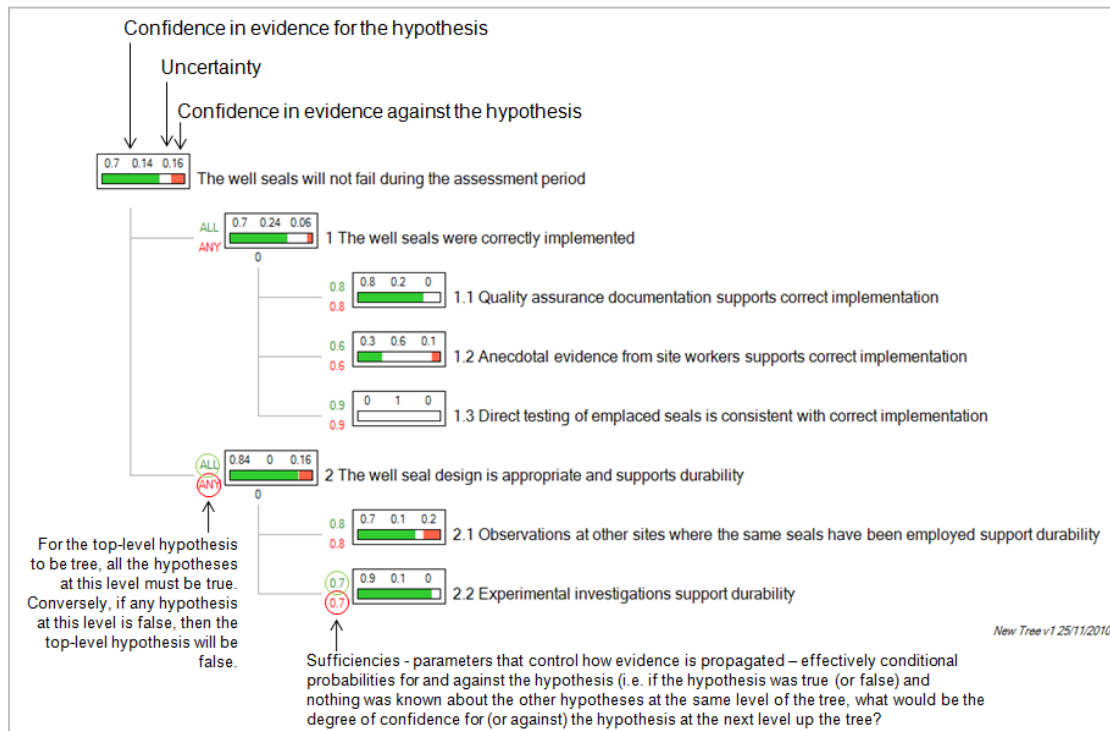
Figure 8: Example hypothesis model illustrating how degrees of confidence in hypotheses that closely relate to information or data (at the extreme right) are propagated to determine the degree of confidence in some root hypothesis of interest (at top left). An actual tree would typically be considerably larger than this example.

Step 2b: Assessing Confidence on the Basis of Evidence

Evaluating Evidence

Judgments on evidence can be elicited in several ways; the standard approach followed in TESLA involves separate elicitation of "confidence for" and "confidence against" each leaf hypothesis. Because this depends on expert judgments, a helpful approach can be to use qualitative linguistic judgments in a similar way to that described for sufficiency and dependency above. TESLA allows users to define a mapping scheme between linguistic terms describing confidence judgments based on the face values of evidence and numerical values.

There are many potential contributions to residual uncertainty in the treatment of evidence;

▲ *Incomplete knowledge* concerning the leaf hypotheses in the logical hierarchy – we don't understand all the issues involved;

▲ *Incomplete characterisation* of the system – we don't have all the data;

▲ *Uncertain quality* – we have the data but we're not sure of their reliability for use as evidence;

▲ *Uncertain meaning* - we have data but we're unsure what they mean;

▲ *Conflict* - we have relevant data from different sources which don't agree, or too much confidence has been ascribed to the evidence; and

▲ *Variability* - we have data but they don't give us a unique answer.

Even when there is no remaining uncommitted confidence or overcommitted confidence (i.e. when represented on scales of 0 to 1, confidence for and confidence against a hypothesis sum to 1) the balance of confidence may not be biased strongly towards either truth or falsehood of the hypothesis. This is because the odds may naturally not be stacked in favour of one outcome or another, and so obtaining further evidence may not change this. For example, if the hypothesis is that "Heads will face upwards after tossing a coin", clearly there will be confidence for 0.5 and confidence against of 0.5. Additional evidence (e.g. provided by more coin tosses) would not change these judgments. In such a case, restructuring the hypothesis model to ask a slightly different question, or breaking down the hypothesis further to give a finer grained assessment of the available evidence may help.

Bowden (2004) discusses a classification scheme for uncertainty to account for such contributions, based on the work of several authors (Foley et al., 1997; Funtowicz and Ravetz, 1990; Hoffman-Reim and Wynn, 2002). As a general rule, however, a fundamental distinction can be made between those uncertainties in confidence that are intrinsic to the system under investigation (aleatory uncertainty) and those that relate to the ignorance of those who wish to understand and model the system (epistemic uncertainty). In principle, both types of uncertainty can be expressed using the language of probability theory; in practice, it can be helpful in the elicitation process, where specialists are examining the evidence in support of each child hypothesis in the logical model, to recognise and separate these two categories.

In the application of ESL, when it is difficult to make a judgment of evidential support directly from the available evidence, it may be beneficial to split the judgment into two main steps. Firstly specialists are asked about the overall coverage of the knowledge base on which they are being asked to make judgments. That is, relative to the distribution of data (e.g. spatial or temporal) and "amount" of knowledge that they would ideally like to make a judgment, what is the distribution of data and "quantity" of knowledge that they actually have, recognising that this is a subjective judgment?

For example, for judgments relating to groundwater quality, the data might describe observed concentrations for a number of dissolved contaminant species, but might miss a couple of contaminants known to be potentially significant; in this case the coverage might be poor. Alternatively, the data might be associated with very large uncertainty ranges that suggest poor data collection approaches, and then the "quality" might be considered suspect.

Then, for the evidence that is available, a judgment is made about its "face value" in support of (or against) the hypothesis in question. This reflects the extent to which the evidence, if taken at face value and noting quality and coverage modifiers have already been taken into account, supports or refutes the hypothesis.

It is recommended that "confidence for" and "confidence against" values for a hypothesis are elicited separately (whether elicited directly or using the coverage / quality / face-value approach as above). This can be done using a linguistic confidence scale, such as "Very confident", "Confident" etc., which is then mapped to a numerical scale.

In seeking first to understand the coverage of the knowledge base, it is possible to invite expert judgment in relation to the following questions:

The ratio of the two answers gives an indication of coverage of the available evidence base, $k(E)$, on a scale of 0 to 100%. Hence it is possible to provide an estimate of the "residual uncertainty due to lack of knowledge", $1 - k(E)$, for the particular child hypothesis under consideration. Consideration of coverage of the evidence provides for the possibility of making firmer judgments based on data derived from a detailed investigation programme than that from a less mature knowledge base.

As noted previously, elicitation is carried out in two stages, in which judgments are made first of the face value of the evidence and then a further judgment is made of the quality of the evidence. In other words, two questions are asked:

Question (2a) is broadly equivalent to the evaluation of sufficiency in parameterisation of the hierarchical relationships between hypotheses in the logical model. Question (2b) extends the evaluation to an appraisal of the quality of the available evidence in order to modify its face value. The general procedure for using value functions to elicit quantitative measures of the degree of support provided by the evidence and its quantity and quality is illustrated by Figure 9.
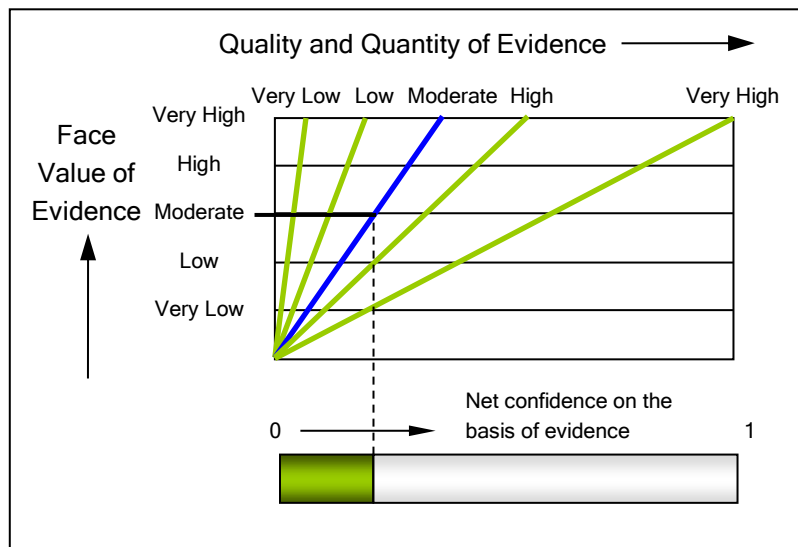
Figure 9: Use of value functions and linguistic variables to define membership categories for "face value" and "quality" of evidence (assuming complete coverage).

The net value of the confidence for or against is determined by first selecting a value function that reflects judgment of the face value of that evidence. The appropriate location on that value function is then determined by judging the quality of the available evidence. In the example shown in Figure 9, the chosen value function reflects a judgment that the face value of the

16

evidence provides a *moderate* level of support the hypothesis being true. The net confidence in favour ($n(E)$) is then determined based on the judgment that the evidence is of *moderate* quality.

The overall assessment of confidence in the evidence, $C(E)$, then needs to take into account both the net value of the evidence that exists, $n(E)$, and the previously estimated evidence coverage, $k(E)$. Thus:

$$C(E) = n(E)k(E)$$

Evaluation of the confidence against the hypothesis for a given leaf hypothesis is carried out separately but in essentially the same way. Assuming a normalised frame of reference for the assignment of confidence, then the "residual uncertainty" due to uncommitted confidence will be given by $1 - \big(C(E) + C(\text{not }E)\big)$, and the residual uncertainty due to overcommitted confidence will be given by $\big(C(E) + C(\text{not }E)\big) - 1$. (See Figure 3 and Figure 4, respectively).

Assessing the Quality of Evidence

A more detailed approach to assigning confidence values to evidence, whether quantitative or qualitative, can provide a more robust justification than is possible through the subjective linguistic response to a judgment of quality illustrated in Figure 10. As part of their NUSAP scheme for uncertainty analysis, Funtowicz and Ravetz introduced the concept of "Pedigree", in which the origin and trustworthiness of knowledge is based on some measure of who has it, how it was derived and what went into its derivation (Funtowicz and Ravetz, 1990). The form of the Pedigree evaluation is a rectangular array, in which the columns represent different quality indicators. The cells in each column describe the particular criteria against which judgments are made in rank order from top to bottom (Funtowicz and Ravetz, 1990; van der Sluijs et al., 2002). The quality indicators used may vary depending upon the subject of the analysis and there are various approaches to combining evaluations of different quality indicators to produce an overall quality score.

In a scientific context, the indicators given here (see Figure 10 and Text Box) are familiar confidence-building measures appropriate to the peer review process. These include, for example: theoretical basis, scientific method, auditability, calibration, validation and objectivity (Bowden, 2004). For each leaf hypothesis in the logical model for which evidence is assessed, confidence scores may be assigned to the evidence according to the various quality indicators. An overall quality score can then be calculated from the cumulative, normalised scores for the individual indicators. If so desired, the indicators themselves are weighted according to subjective judgment of their relative importance. Separate quality scores can be elicited for the supporting and refuting confidence.

In practice, for a given ESL application, a thorough analysis of the available evidence, taking into account its quality, can be a resource intensive and time-consuming process, requiring detailed evaluation of individual items of relevant information and the way in which they are combined to make evidential judgments.

Quality Indicators

Quality Score	Theoretical basis	Scientific method	Auditability	Calibration	Validation	Objectivity
Vhigh (1)	Well established theory	Best available practice; large sample; direct measure	Well documented trace to data	An exact fit to data	Independent measurement of same variable	No discernible bias
High	Accepted theory; high consensus	Accepted reliable method; small sample; direct measure	Poorly documented but traceable to data	Good fit to data	Independent measurement of high correlation variable	Weak bias
Mod	Accepted theory; low consensus	Accepted method; derived data; analogue; limited reliability	Traceable to data with difficulty	Moderately well correlated with data	Validation measure no truly independent	Moderate bias
Low	Preliminary theory	Preliminary method; unknown reliability	Weak, obscure link to data	Weak correlation to data	Weak, indirect validation	Strong bias
Vlow (0)	Crude speculation	No discernible rigour	No link back to data	No apparent correlation with data	No validation presented	Obvious bias

Figure 10: Judgment of the "quality" of the information on which evidence is given, based on a number of defined quality indicators.

QUALITY INDICATORS

Theoretical Basis
This indicator is used to indicate whether the evidence under consideration conforms to well-established theory at one extreme or is pure speculation at the other.

Scientific method
As a general rule the scientific method follows the sequence, Observation; Formulation of hypothesis; Test hypothesis; Reject or fail to reject the hypothesis. If rejected, a new hypothesis can be proposed and the process begun again. If not rejected, the hypothesis stands as a valid explanation of the observations. But, it is not necessarily the only valid explanation, there may be other equally valid alternative explanations of the observations and future observations or experiments could cause the hypothesis to be rejected.
This indicator provides a measure of how well the scientific method and best practice has been followed in the production of the information on which judgment of evidence has been made.

Auditability
No definitive, rule based methodology exists for data integration, interpretation and modelling. However, in order to demonstrate confidence in the models, it is essential to document and justify the sequence of conceptual and methodological steps contributing to specific interpretations such that an audit trail is discernible from the model to the data from it 'has' been conceived. This indicator provides a measure of how well the information presented can be traced back to the raw observations.

Calibration and validation
Models are abstractions of reality. Calibration may be used to confirm the legitimacy of the model as 'consistent with all the observational data', and validation may establish the model as permissible through some form of prediction testing but neither calibration nor validation can establish whether or not the model is correct. Nevertheless both calibration and validation are valuable and necessary confidence building measures for demonstrating quality in the modelling process.
These indicators are used to make judgments on whether the information is a) calibrated to data and b) validated through independent measurement or prediction testing.

Objectivity
Whilst the scientific method provides a logical framework for improving understanding it does not guarantee objectivity. The influence of entrenched values, motivational bias and peer and institutional pressures may obscure true objectivity. In order to maintain a check on the quality and objectivity of our interpretations we rely on peer review and exposure to critique through peer reviewed publication.

This indicator is used to give a judgment on the extent to which the information can be said to be objective and free from bias.

Text Box: Description of Data Quality Indicators.

Step 3: Propagation of Confidence

The mathematical algorithms used in ESL to propagate confidence on the basis of evidential judgments are founded on standard probability theory (Cui and Blockley, 1990). Within TESLA, the propagation of "confidence for" is treated independently from that for "confidence against", but using the same algorithm derived using the principle of inclusion and exclusion of probabilities:

$$C^{\pm}(h) = \sum_{k=1}^{|G_h|} (-1)^{k-1} \sum_{\substack{K \subset G_h \\ |K|=k}} C^{\pm}\left(\bigcap_{i \in K}[h \cap i]\right)$$

where a linear mapping between statistical independence and complete dependence is used, based on the sufficiency weighted confidence values:

$$C^{\pm}\left(\bigcap_{i \in K}[h \cap i]\right) = (1 - D_K) \prod_{i \in K} s_{hi}^{\pm} C^{\pm}(i) + D_K \min_{i \in K}\left(s_{hi}^{\pm} C^{\pm}(i)\right)$$

Here

▲ $C^{\pm}(h)$ is the calculated confidence value for (+) or against (-) the parent hypothesis, h;

▲ $C^{\pm}(i)$ is the calculated confidence value for (+) or against (-) the child hypothesis, i;

▲ G_h represents the set of all child hypotheses, i;

▲ $K \subset G_h, |K| = k$ represents all combinations of the subsets of the contributing hypotheses containing exactly k terms;

▲ $s_{hi}^{\pm}$ is the sufficiency of child hypothesis i for parent hypothesis h; and

▲ D_K is the dependency between the subset of child hypotheses given by K.

The number of operations required to calculate the confidence in a hypothesis grows rapidly with the total number of child hypotheses, increasing the emphasis on use of a supporting software tool. Nevertheless, a simple illustration of the parameterisation procedure, based on the above calculus for the propagation of confidence is shown in Figure 11.
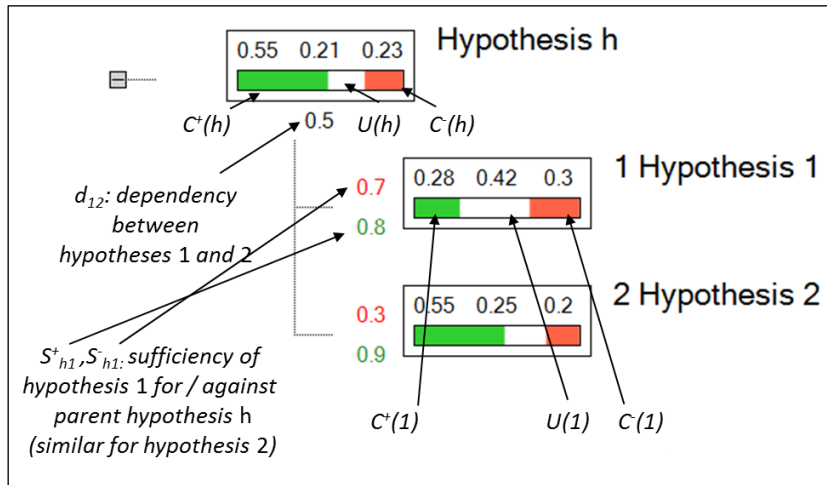
Figure 11: Propagation of confidence values in a hypothesis model. The confidence values in the supporting and refuting evidence are given by C^+ and C^- *respectively*, while the residual uncertainty due to uncommitted or overcommitted confidence is calculated as $U = 1 - C^+ - C^-$ (where positive U represents uncommitted confidence and negative U represents overcommitted confidence).

Heuristics act to override the default propagation algorithm and are applied after this algorithm. ALL is modelled by direct propagation of the minimum confidence values from among the child hypotheses, i.e.

$$C^{\pm}(h) = \min_{i \in G_h}(C^{\pm}(i))$$

ANY is modelled by direct propagation of the maximum confidence values from among the child hypotheses, i.e.

$$C^{\pm}(h) = \max_{i \in G_h}(C^{\pm}(i))$$

The symmetric necessity heuristic is applied using the following algorithm:

---

set the parent's confidence for and confidence against to the values obtained using the propagation algorithm

if the children are necessarily true, then
{
    for every necessary child
    {
        if its confidence for is less than a half **and** less than the parent's current confidence for
        {
            set this as the parent's confidence for
        }

        if its confidence against is greater than a half **and** greater than the parent's current confidence against
        {
            set this as the parent's confidence against
        }

    }
}
else if the children are necessarily false, then
{
    for every necessary child
    {
        if its confidence for is greater than a half **and** greater than the parent's current confidence for
        {
            set this as the parent's confidence for
        }
        if its confidence against is less than a half **and** less than the parent's current confidence against
        {
            set this as the parent's confidence against
        }
    }
}

Step 4: Analysis and Visualisation

In decision making that is based on an open-world perspective on the available evidence, it is not possible to deal with absolute truths or in mathematical terms of accuracy and precision. What we are searching for is some measure of our confidence in our models and hypotheses; or in the case say of a choice between alternative models, some measure of our relatively greater confidence in one model compared to the alternatives. The ESL methodology works in quantitative terms, and the top-level result is a measure of overall confidence in the model, or hypothesis, under evaluation. However, care is required in the interpretation of such output, to avoid the GIGO (garbage in/gospel-out) epithet that is traditionally associated with apparently numerically precise output from fuzzy inputs. The primary inputs to the ESL process (logical

model parameterisation and evidence evaluation to elicit confidence) are best understood in terms of linguistic or verbal expressions of subjective judgment, and hence are essentially "soft" or "fuzzy", terms. Ideally, the output should also be translated back into verbal qualifiers (in which case "FIFO" – fuzzy in-fuzzy out, might be a more appropriate acronym for the whole process).

Sense checking and sensitivity analyses are key components of the tree development process. If tree analysis and visualisation indicates that the outcomes of the process cannot be easily rationalised given real-world experience and expectations of the outcomes, or if the outcomes are shown to be unduly sensitive to individual judgements, then that provides important feedback to the tree development and/or decision-making process. In such cases either the tree logic and confidence values needs to be reviewed and iterated to improve and update the model, or the outcomes are providing important input directly to subsequent decision processes as they indicate a potential need for a changed perspective on some of the key aspects relevant to the decision. In either case the sensitivity analysis process will inform upon the impact of uncertainty and priorities for either next steps in the tree development process or more broadly (e.g. for research activities). The use of such analyses can be particularly helpful where evaluating confidence values or logical approaches has proved challenging given the nature of the evidence, but in practice all trees typically benefit from some form of iterative analysis and final sense-checking of the outputs.

Using TESLA, several complementary approaches to analysing and visualising an ESL analysis have been devised: the Evidence-Ratio Plot; the sensitivity (or Tornado) Plot; the Tree and Confidence Flow Line Plots; and the Portfolio Analysis Tool (Bowden, 2004; User Guide). These are described in turn below.

Ratio Plot

The Ratio Plot (Figure 12) provides a visual presentation of the levels of confidence of a chosen hypothesis (typically the root hypothesis) and all its descendants. The horizontal axis indicates the percentage uncertainty due to uncommitted confidence (white space in the Italian flag representation; Figure 3) or overcommitted confidence (yellow space in the Italian flag representation; Figure 4). The former plots to the right of the vertical axis (i.e. with positive values), while the latter plots to the left of the vertical axis (i.e. with negative values).
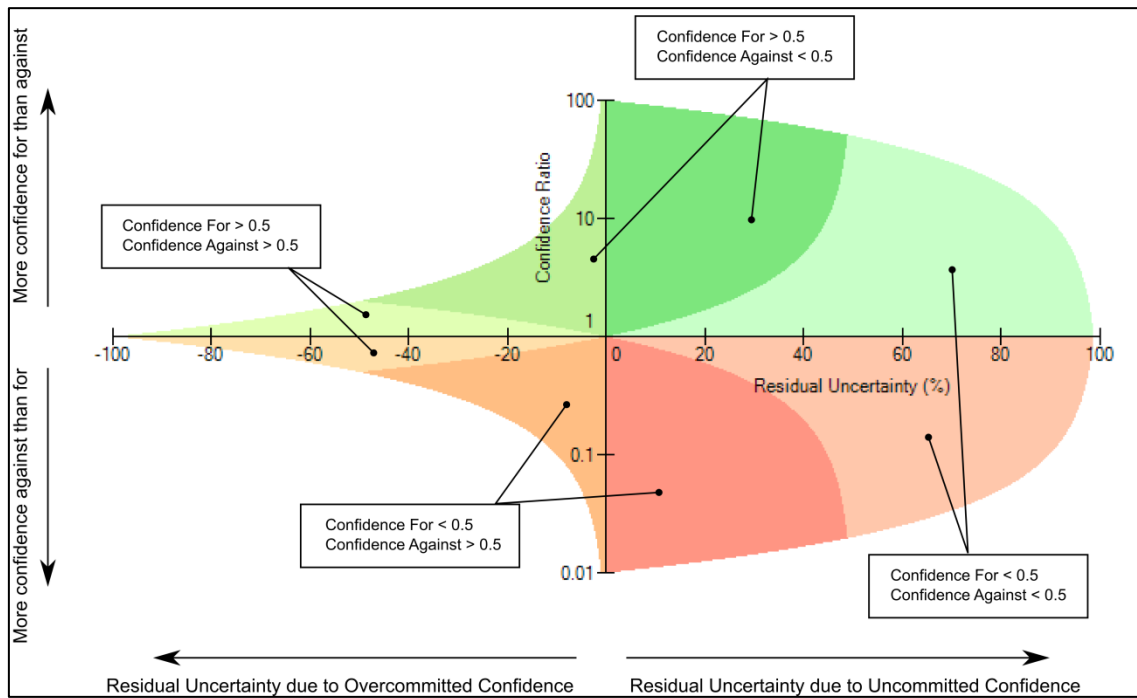
Figure 12: Regions of the Ratio Plot.

The vertical axis provides an indication of the balance of confidence. This is expressed as the ratio of confidence for to confidence against associated with each hypothesis, displayed on a logarithmic scale. The confidence values used in the calculation of the ratio are constrained to have a minimum value of 0.01, resulting in a possible range of between 0.01 and 100.

The location [0,1] on the ratio plot therefore represents a situation where there is a balance of confidence for and against (50% for and 50% against), with no residual uncertainty due to either overcommitted confidence or uncommitted confidence. This would be plotted at the intersection of the vertical and horizontal axes. Increasing residual uncertainty, while maintaining a balance of confidence for and against, results in movement either to the left or right along the horizontal axis, depending on whether or not there is conflict.

Values plotted above the horizontal axis represent a balance of confidence indicating support for the hypothesis under consideration; those below the line represent a balance of confidence against support for the hypothesis. Regions representing greater than 50% confidence for and against respectively are shaded on Figure 12, providing a visual guide to the extent of support that is judged to exist. In complex logical models with high levels of overcommitted confidence (to the left of the vertical axis) or uncommitted confidence (to the right of the vertical axis), the confidence ratios will still be an important confidence measure. In these cases, and in cases where hypotheses have confidence ratios near the middle of the range, conflict resolution will be important if the conflicts between confidence for and against are judged to have a significant impact on confidence in the root hypothesis.

It can be informative to plot confidence in the root hypothesis from the ESL model as a result of evidence evaluations at the leaf level on the same diagram, set against the background of the individual levels of confidence for each leaf hypothesis in the logical model. This can provide a strong visual indicator of the potential implications of bias. For example, in situations where an "outlier" piece of evidence is strongly (or even exclusively) relied upon in order to justify an overall conclusion, the root hypothesis will inevitably be skewed towards that location on the Ratio Plot. By contrast, where full account is taken of the balance of confidence, including the

possible weight of contradictory evidence and residual uncertainty due to under- and over-committed confidence, the "true" evidential support for the top-level hypothesis can be clearly visualised.

Tornado Plot

The Tornado Plot (also known as the Sensitivity Plot) indicates the sensitivity of confidence in the root hypothesis to small changes in confidence in each of the leaf hypotheses, referred to as "impact".

The derivation of the impact of each leaf hypothesis is a first-order differential calculation, and is implemented in TESLA by temporarily incrementing its confidence values by a marginal amount, noting the change in confidence values of the top hypothesis. The impact is thus defined by the ratio of the change in confidence for the root hypothesis to the change in confidence for the leaf hypothesis.

The impact for each leaf hypothesis is converted to a percentage and plotted as a horizontal bar, a green one to indicate sensitivity to confidence for, and a red one to indicate sensitivity to confidence against. The hypotheses are then plotted in descending order of total impact, thereby giving the whole plot its tornado-like appearance from which it takes its name. An example is illustrated in Figure 13.
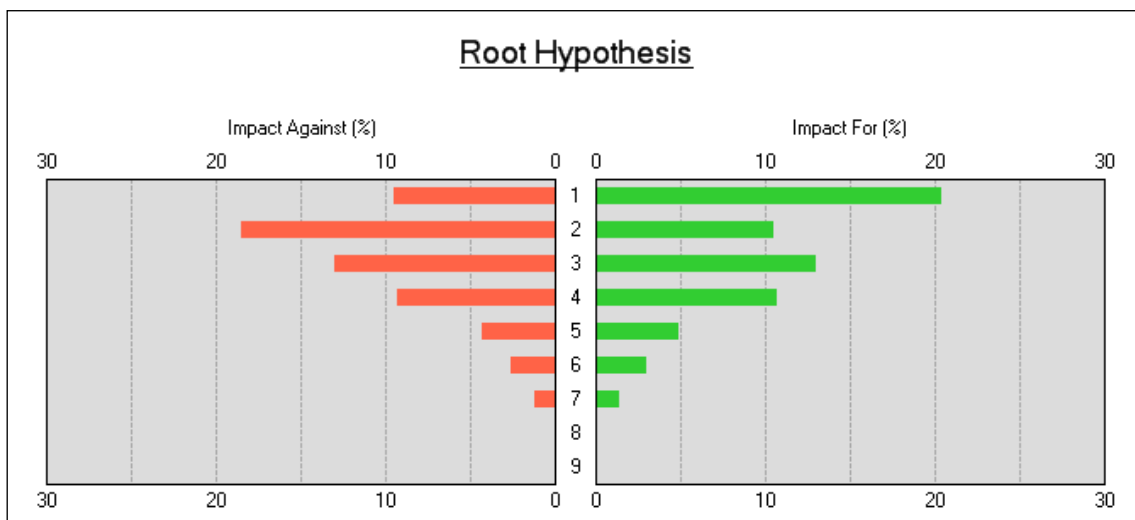
Figure 13: Example Tornado Plot.

In effect, such calculations provide a measure of the "value of information" associated with the ESL model under study. This, in turn, can be used to support decision making, through an analysis of the confidence-building implications of alternative data acquisition or conflict resolution strategies.

Note that it is important to consider the implications of logical "switches" such as ANY and ALL operators and necessary hypotheses within the tree in evaluating overall sensitivities. As the Tornado Plot is calculated on the basis of marginal differences at the leaf level, it does not take into account the implications of major switches in logic elsewhere. Therefore, it may be necessary to construct several Tornado Plots, for example one considering each major line of reasoning in the tree, to fully understand sensitivities by exploring confidence judgements that could trigger changes in lines of propagation.

Tree Display

TESLA has been designed as a flexible tool for decision modelling, with an emphasis on the ability to manipulate the hierarchical structure of the decision (as well as data input and parameterisation) as it evolves, coupled with a graphical user interface to promote group working.

An important aspect of the TESLA interface, from the perspective of describing its application, is the *Tree Display*. An example TESLA tree display is shown in Figure 14.

The hierarchical hypothesis model is displayed in TESLA in a collapsible tree view format. Lines connect associated hypotheses in the tree, with child hypotheses appearing below and to the right of their parent hypothesis. Thus, the root hypothesis is located to the far top left, and the leaf hypotheses to the far right of the display.
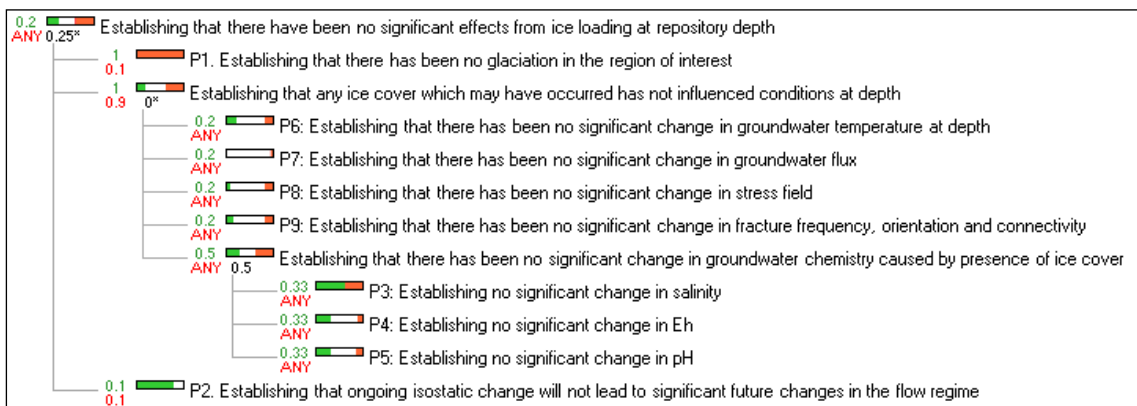
Figure 14: Example of ESL visualisation in the TESLA "tree display" (zoomed-out view).

Confidence in a hypothesis in the ESL model is displayed graphically using the "Italian flag" representation described earlier. Thus a bar is shown with a green portion stemming from the left (showing the confidence in favour), and a red portion stemming from the right (showing the confidence against). The section of the bar in between the red and green portions is white to indicate the uncommitted confidence, or coloured yellow when there is overlap in the case of overcommitted confidence.

Numeric values for *sufficiency* are shown to the left of the confidence display bar for each child hypothesis in the model. "Sufficiency for" and "sufficiency against" are coloured green and red, respectively. If logical operators are used, the words "ANY" or "ALL" are shown to the left of the confidence display, as required. If an individual child hypothesis is deemed to be a *necessity*, then the background to the display icon for that particular child hypothesis is shaded. A light green background indicates "necessarily true" and a light red background indicates "necessarily false" using the symmetric necessity heuristic. A light grey background indicates a necessary node using the asymmetric necessity heuristic.

The *dependency* parameter is displayed in the tree view beneath the Italian flag. It relates to the dependency between child nodes of the given hypothesis. If a "general dependency" value has been set then its numeric value is indicated. However, as noted previously, it is possible in situations where there are multiple sibling child hypotheses for the degree of dependency to differ between different groupings of siblings. If one or more alternative dependency values have been set for groups of child hypotheses this is indicated by an asterisk (*).

Confidence Flow Lines

Confidence flow lines and percentage contributions of confidence in the leaf hypotheses to confidence in the root hypothesis can be added to the tree plot to provide additional information about the flow of confidence through the tree. The flow lines can help to identify which leaf hypotheses (and ultimately, evidence sources) are contributing (or not contributing) to the root hypothesis outcome and by how much. They can also be used to check the influence of logic on the propagation of confidence through the tree.

Figure 15 demonstrates these features. The widths of the flow lines reflect the relative contributions to confidence. The percentage figures at the leaf level indicate the relative overall contribution of the leaf hypotheses to root hypothesis confidence.

In this example, the flow lines highlight where confidence in the root hypothesis "Material emplaced deep underground will not cause harm" originates. In particular, it shows that 79% of confidence "for" the root comes from the leaf hypothesis "Quality assurance documentation supports correct implementation" with a further significant contribution from "Anecdotal evidence from site workers supports correct implementation".

The figure also shows the effect of ANY and ALL logic with several leaf hypotheses providing zero contribution to confidence "for" the root hypothesis, because a hypothesis on another branch provides a lower level of confidence and that flow line is selected through he ALL switch.
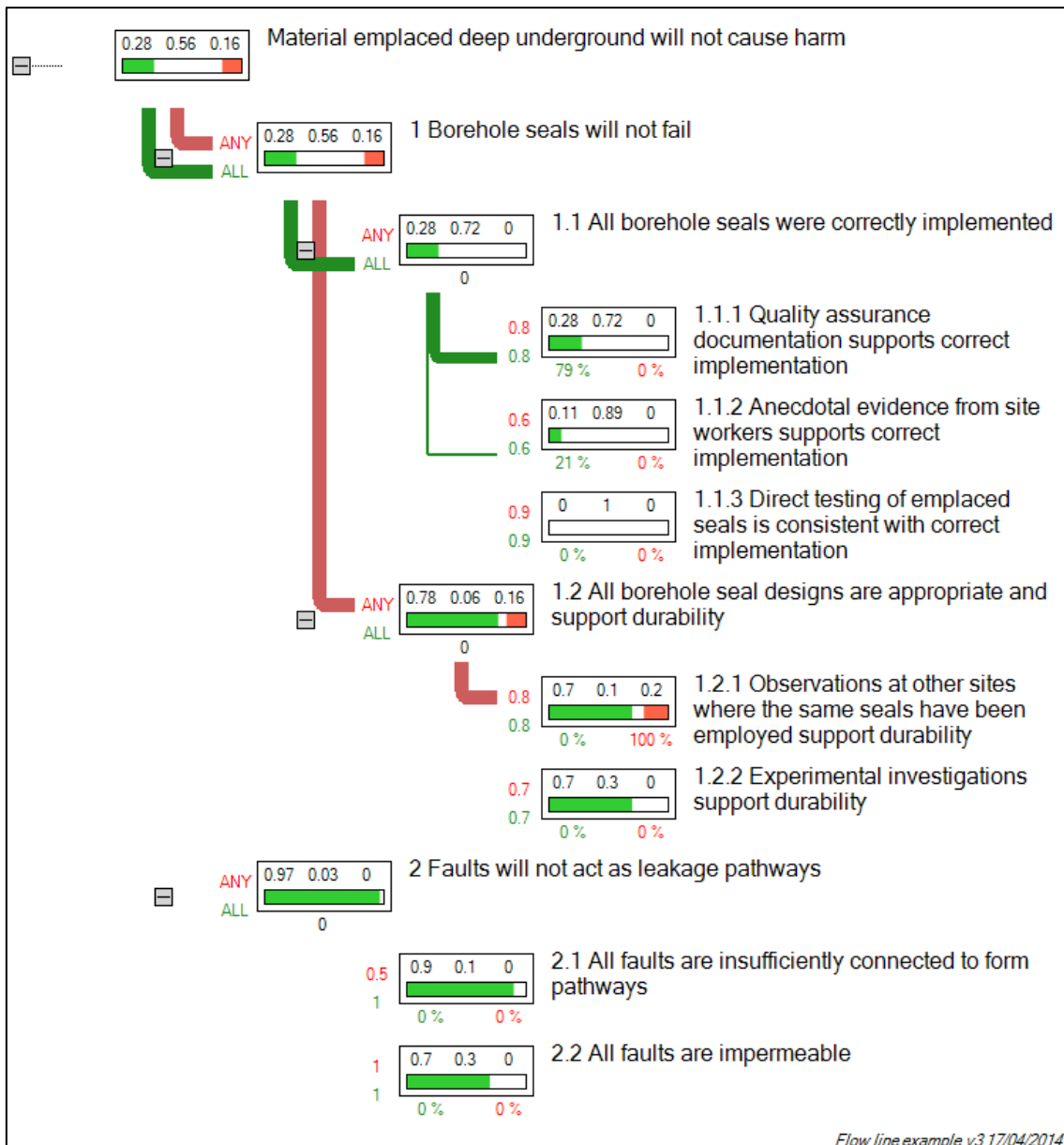
Figure 15: The Flow of Confidence for the Umbrella Example. The flow lines show the contribution of confidence in the leaf hypotheses to confidence in the root hypothesis.

Portfolio Tool

The "portfolio tool" is a development of the tree plot that allows a user to compare, in one diagram, multiple hypothesis models that have the same structure, but different confidence values (Figure 16). This plot is often used to compare the performance of different options with different associated evidence bases using a common framework. For example, the portfolio tool can be used to compare the judged suitability of different potential sites for storing CO_2 underground, where there are different datasets available for each site. Such a plot can also be used to show the evolution of performance with time, by recording confidence on the basis of evidence at different times during an evidence collection / collation process (e.g. site investigation or experiment).
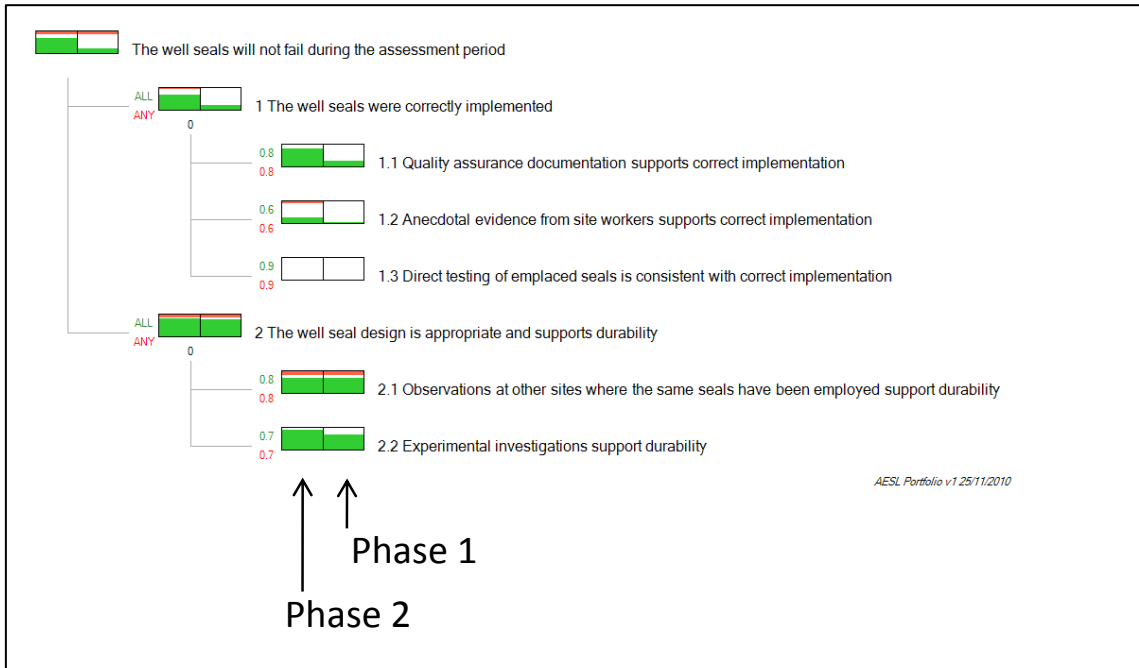
Figure 16: Example application of the portfolio tool to compare two instances ("Phase 1" and "Phase 2") of a hypothetical tree with different confidence values.

TESLA Downloads and Further Information

For full details of how to use the software please refer to the TESLA User Guide, which is shipped with the executable. An evaluation version of TESLA is available for download via the Quintessa website:

http://www.quintessa.org/software/TESLA/

The website also provides guidance on purchasing a license to unlock its full functionality.

References

Blockley, D. and Godfrey, P. (2000). "*Doing it Differently - Systems for Rethinking Construction*", Thomas Telford, London.

Bowden, R. A. (2004). "Building confidence in geological models". In: Curtis, A. and Wood, R. (Eds.) *Geological Prior Information*. Geological Society, London, Special Publications.

Cui, W. and Blockley, D. I. (1990). "Interval probability theory for evidential support." *International Journal of Intelligence Systems*, **5**, pp.183-192.

Davis, J. P. and Hall, J. W. (2003). "A software-supported process for assembling evidence and handling uncertainty in decision-making". *Decision Support Systems*, **35**, pp.415-433.

Foley, L., Ball, L., Hurst, A., Davis, J. and Blockley, D. (1997). "Fuzziness, incompleteness and randomness: classification of uncertainty in reservoir appraisal". *Petroleum Geoscience*, **3**, pp.203-209.

Funtowicz, S. O. and Ravetz, J. R. (1990). "*Uncertainty and Quality in Science for Policy*". Dordrecht: Kluwer.

Hall, J. W., Blockley, D. I. and Davis, J. P. (1998). "Uncertain inference using interval probability theory". *International Journal of Approximate Reasoning*, **19**, pp247-264.

Hoffman-Reim, H. and Wynn, B. (2002). "In risk assessment, one has to admit ignorance". *Nature*, **416** (123).

Noda, T. (1987). "Anion index as an indicator of geothermal activity". *Journal of the Geothermal Research Society of Japan*, **9**, pp.131-141 (in Japanese with an English abstract).

Quintessa, 2014. User Guide for Tesla version 2.2.

van der Sluijs, J. P., Potting, J., Risbey, J., van Vuuren, D., de Vries, B., Beusen, A., Heuberger, P., Corral Quintana, S., Funtowicz, S., Kloprogge, P., Nuijten, D., Petersen, A. and Ravetz, J. (2002). "*Uncertainty assessment of the IMAGE/TIMER B1 CO2 emissions scenario, using the NUSAP method*". Dutch National Research Program on Climate Change, Report no: 410 200 104, 227 pp.

Waltz, E. (1989). "*Multisensor Data Fusion*". Artech House, London.

Further Reading – Example Applications of ESL

Example applications in the public domain are listed below, the most recent first.

Metcalfe, R., Paulley, A., Suckling, P.M., and Watson, C.E., 2013. "*A tool for integrating and communicating performance-relevant information in CO_2 storage projects: description and application to In Salah*". Proceedings of the GHGT-11 Conference, 18th-22nd November 2012, Kyoto International Conference Center, Japan, Elsevier, Energy Procedia, **37**, 4741-4748.

Tucker, O., Holley, M., Metcalfe, R., Hurst, S., 2013. "*Containment risk management for CO_2 storage in the Goldeneye depleted gas field, UK North Sea.*" Proceedings of the GHGT-11 Conference, 18th-22nd November 2012, Kyoto International Conference Center, Japan, Elsevier, Energy Procedia, **37**, 4804-4817.

Paulley, A., Metcalfe, R., and Egan, M., 2012. "*Geological Disposal Programme Design and Prioritisation in the Face of Uncertainty: Use of Structured Evidence Support Logic Techniques.*" Mineralogical Magazine Special Addition: Proceedings of the 2012 Geological Disposal of Radioactive Waste: Underpinning Science and Technology Conference, December 2012, **76**(8), 3497-3507.

DNV, 2010. "*CO2QUALSTORE Workbook with examples of applications*". DNV Report No. 2010-0254. http://www.dnv.com/binaries/CO2QUALSTORE_Workbook_tcm4-436659.pdf

James, S., Garnett, A., Kumar, G., Kumar, V., Rao, N., Trivedi, B., Gupta, A., Salunka, S., Sarakar, S., Scrinivasan, A., Meen, P., Doran, S., Hall, N., and Barlar, P., 2010. "*What does it take to evaluate a potential CO_2 Storage Site? The ZeroGen example*". Proceedings of the Abu Dhabi International Petroleum Exhibition and Conference, UAE 1-4 November 2010. Society of Petroleum Engineers Reference SPE 137447.

Egan, M., Paulley, A., Lehman, L., Lowe, J., Rochette, E. and Baker, S., 2009. "*Assessing confidence in performance assessments using an Evidence Support Logic methodology: an application of TESLA*". Proceedings of the Waste Management 2009 Conference, Phoenix, AZ, March 1-5.

Toida, M., Suyama, Y., Seno, S., Atsumi, H. and Ogata, N. 2008. "*Assessment of uncertainties associated with characterisation of the geological environment in the Tono area*". JAEA Report JAEA-Research 2008-035 (in Japanese).

Mizuno, T., Metcalfe, R., Iwatsuki, T. and Mie H. 2007. "*Proposal for a quality assessment method of groundwater chemistry for hydrochemical investigation*". Journal of the Japanese Association of Groundwater Hydrology, 49, 139-152 (In Japanese).

Takase, H. and Kato, K., "Reasoning with imperfect information; a structured approach to handling subjective uncertainty", presented at the 10th International Conference on Environmental Remediation and Radioactive Waste Management (ICEM '05), September 2005.

Hall, J., Twyman, C. and Kay, A., "*Influence diagrams for representing uncertainty in climate-related propositions*", Climatic Change, **69**:2-3 pp 343-365, April 2005.

Egan, M. J. and Bowden, R. A., "*Application of evidence support logic to the role of palaeohydrogeology in long-term performance assessment*", Quintessa Report for UK Nirex Limited, QRS-1219A-1, October 2004.

Seo, T., Tsuchi, H., Metcalfe, R., Suyama, Y., Takase, H., Bowden, A., Toida, M., Furuichi, M., Matsumura, A., Yoshimura, M. and Horio, A., "*A decision making methodology taking into account uncertainties and its possible application for the selection of preliminary investigation*

areas", extended abstract for a presentation given at the Distec 2004 International Conference, Berlin, Germany, 26-28 April 2004.

Dawson, R., Hall, J. and Davis, J., "A performance-based, decision support process for flood defence asset management", *Proceedings of ICE, Water Management*, **157:**1 pp 35-44, March 2004.

Okubo, S., Suyama, Y., Bowden, A., Yoshimura, K., Toida, M. and Metcalfe, R., "Uncertainty assessment of geological modelling using Evidential Support Logic", *Butsuri-tansa*, **57**, pp 403-409, (in Japanese) 2004.

Tsuchi, H., Seo, T., Metcalfe, R., Kawano, K., Takase, H., Bowden, A., Suyama, Y., Toida, M., Furuichi, M., Matsumura, A. and Yoshimura, M., "*Study on a decision making methodology with uncertainties in the selection process of preliminary investigation areas*". Materials Research Society Symposium Proceedings, **807**, 21-26, 2004.

DEFRA/Environment Agency, "*Condition monitoring and asset management for flood and coastal defence systems*" R&D Technical Report W5A-042/TR, UK, August 2003.

Metcalfe, R., Takase, H., Toida, M., Suyama, Y., Kawano, K., Bowden, A. and Savage, D., "Using evidence-based logic to design investigations for clay barrier systems", Extended abstract for a presentation given at the international conference, '*Clays in natural and engineered barriers for radioactive waste confinement*', Reims, France, 9-12 December 2002.

Glossary of Terms

Term	Definition
Dependency	Value between 0 and 1 indicating the amount of confidence in two or more sibling hypotheses that is based upon the same evidence or judgments on evidence.
ESL	Evidence Support Logic, an approach to analysing confidence in hypotheses on the basis of available evidence, and supported by a mathematical algorithm founded on Interval Probability Theory.
Confidence Against	Value between 0 and 1 indicating the strength of confidence that a hypothesis is false, based on evidence.
Confidence For	Value between 0 and 1 indicating the strength of confidence that a hypothesis is true, based on evidence.
Child Hypothesis	A hypothesis that has a parent hypothesis (i.e. any but the root hypothesis). A child hypothesis occurs at the next level down the tree structure from its parent hypothesis.
Hierarchical Hypothesis Model	The decomposition of a single hypothesis (the root hypothesis) into a number of more specific hypotheses, which in turn may be broken down into even more specific hypotheses until a point is reached where the hypothesis is sufficiently well defined for factual information to be judged readily. This results in a tree-like hierarchy.
Hypothesis	An assertion to be evaluated, represented as a node in the tree. Supported or refuted by lines of reasoning on the basis of evidence.
Leaf Hypothesis	This is the only class of hypothesis for which confidence based on the available evidence can be input by the user; found at the lowest level of the tree. There can be many leaf hypotheses.
Necessity	Indicates that confidence in the truth of a hypothesis is necessary for confidence in its parent, irrespective of the confidence provided by its siblings, or alternatively that confidence in the falsehood of a hypothesis is necessary for confidence in its parent irrespective of the confidence provided by its siblings.
Parent Hypothesis	A hypothesis with one or more child hypotheses (i.e. any but the leaf hypotheses).
Propagation	The calculation of confidence in parent hypotheses by combining the confidence in its child hypotheses based on a user-specified ESL parameterisation.
Root Hypothesis	The root hypothesis is the one at the highest level of the tree and has no parent hypotheses. After confidence propagation, the root hypothesis indicates the confidence and residual uncertainty in the main hypothesis associated with a decision.
Sibling Hypotheses	Hypotheses that share the same parent.
Sufficiency	A numerical value that indicates how much judgments of confidence in a child hypothesis (for or against) contribute to confidence (for or against) in the parent hypothesis. A sufficiency answers the question "If this child hypothesis is true (or false), and I know nothing else, then what is my confidence that the parent is true (or false)?"
Tree	Graphical representation of a hierarchical hypothesis model

Term	Definition
Uncommitted confidence	Positive value ≤1, indicating the extent to which confidence in the truth or falsehood of a hypothesis cannot be assigned owing to lacking evidence. Numerically it is represented as: 1 - confidence for - confidence against, where confidence for and confidence against are each represented on a numerical scale of 0 to 1 and confidence for + confidence against <1.
Overcommitted confidence	Positive value ≤1 that measures the overlap between confidence that a hypothesis is true and confidence that the hypothesis is false. It is a measure of overconfidence in judgments of confidence for and / or confidence against a hypothesis. Numerically it is represented as: confidence for + confidence against - 1, where confidence for and confidence against are each represented on a numerical scale of 0 to 1 and confidence for + confidence against >1.

Contributors

M. J. Egan

A. Paulley

R. Metcalfe

P. Suckling

S. Stacey

Quintessa